

Special Issue Reprint

Industrial Process Operation State Sensing and Performance Optimization

Edited by
Sheng Du, Li Jin, Xiongbo Wan and Zixin Huang

mdpi.com/journal/processes

Industrial Process Operation State Sensing and Performance Optimization

Industrial Process Operation State Sensing and Performance Optimization

Editors

Sheng Du

Li Jin

Xiongbo Wan

Zixin Huang



Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Editors

Sheng Du
School of Automation
China University of Geosciences
Wuhan
China

Li Jin
School of Automation
China University of
Geosciences
Wuhan
China

Xiongbo Wan
School of Automation
China University of
Geosciences
Wuhan
China

Zixin Huang
School of Electrical and
Information Engineering
Wuhan Institute of
Technology
Wuhan
China

Editorial Office

MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Processes* (ISSN 2227-9717) (available at: https://www.mdpi.com/journal/processes/special_issues/7R1O67UA2V).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. <i>Journal Name</i> Year , <i>Volume Number</i> , Page Range.
--

ISBN 978-3-7258-2439-7 (Hbk)

ISBN 978-3-7258-2440-3 (PDF)

doi.org/10.3390/books978-3-7258-2440-3

© 2024 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license.

Contents

About the Editors	ix
Biyu Yang, Songli Yang, Zhoujing Ye, Xiaohua Zhou and Linbing Wang Optimization of Anti-Skid and Noise Reduction Performance of Cement Concrete Pavement with Different Grooved and Dragged Textures Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 800, doi:10.3390/pr12040800	1
Daniele Naviglio, Paolo Trucillo, Angela Perrone, Domenico Montesano and Monica Gallo Comparison between Conventional Ageing Process in Barrels and a New Rapid Aging Process Based on RSLDE: Analysis of Bioactive Compounds in Spirit Drinks Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 829, doi:10.3390/pr12040829	20
Hongli Li, Zhiqi Yi, Liye Mei, Jia Duan, Kaimin Sun, Mengcheng Li, et al. SCFNet: Lightweight Steel Defect Detection Network Based on Spatial Channel Reorganization and Weighted Jump Fusion Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 931, doi:10.3390/pr12050931	36
Samuel C. Olisa and Muhammad A. Khan Confluence Effect of Debris-Filled Damage and Temperature Variations on Guided-Wave Ultrasonic Testing (GWUT) Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 957, doi:10.3390/pr12050957	55
Viorel Mînză, Iulian Arama and Eugen Rusu Machine Learning Algorithms That Emulate Controllers Based on Particle Swarm Optimization—An Application to a Photobioreactor for Algal Growth Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 991, doi:10.3390/pr12050991	77
Nodir Yunusov, Bappy MD Siful Islam, Akmalbek Abdusalomov and Wooseong Kim Robust Forest Fire Detection Method for Surveillance Systems Based on You Only Look Once Version 8 and Transfer Learning Approaches Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1039, doi:10.3390/pr12051039	105
Georgina Elizabeth Riosvelasco-Monroy, Iván Juan Carlos Pérez-Olguín, Salvador Noriega-Morales, Luis Asunción Pérez-Domínguez, Luis Carlos Méndez-González and Luis Alberto Rodríguez-Picón CODAS–Hamming–Mahalanobis Method for Hierarchizing Green Energy Indicators and a Linearity Factor for Relevant Factors’ Prediction through Enterprises’ Opinions Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1070, doi:10.3390/pr12061070	123
Iakov A. Lyashenko, Thao H. Pham and Valentin L. Popov Controlling the Friction Coefficient and Adhesive Properties of a Contact by Varying the Indenter Geometry Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1209, doi:10.3390/pr12061209	145
Ionut Nicolae, Mihaela Bojan and Cristian Viespe Acoustic Effects of Uneven Polymeric Layers on Tunable SAW Oscillators Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1217, doi:10.3390/pr12061217	161
Rúben D. F. S. Costa, Marta L. S. Barbosa, Filipe G. A. Silva, Tiago E. F. Silva, Abílio M. P. de Jesus, Francisco J. G. Silva, et al. Image Analysis Techniques Applied in the Drilling of a Carbon Fibre Reinforced Polymer and Aluminium Multi-Material to Assess the Delamination Damage Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1258, doi:10.3390/pr12061258	172

Dong Wang, Lili Jiang, Måns Kjellander, Eva Weidemann, Johan Trygg and Mats Tysklind A Novel Data Mining Framework to Investigate Causes of Boiler Failures in Waste-to-Energy Plants Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1346, doi:10.3390/pr12071346	187
Paul Arévalo and Danny Ochoa-Correa Toward Enhanced Efficiency: Soft Sensing and Intelligent Modeling in Industrial Electrical Systems Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1365, doi:10.3390/pr12071365	206
Alaa Agala, Muhammad A. Khan, Feiyang He and Abdulaaziz Alnuman Leakage Quantification in Metallic Pipes under Different Corrosion Exposure Times Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1529, doi:10.3390/pr12071529	228
Xin Wan Zone Model Predictive Control with Ellipsoid Softing Target in Chemical Processes Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1611, doi:10.3390/pr12081611	242
Ngoc-Tien Tran, Van-Long Trinh and Chen-Kuei Chung An Integrated Approach of Fuzzy AHP-TOPSIS for Multi-Criteria Decision-Making in Industrial Robot Selection Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1723, doi:10.3390/pr12081723	262
Inga Burke, Sven Salzer, Sebastian Stein, Tom Olatomiwa Olakunle Olusanya, Ole Fabian Thiel and Norbert Kockmann AI-Based Integrated Smart Process Sensor for Emulsion Control in Industrial Application Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1821, doi:10.3390/pr12091821	282
Khursheed Ahmad and Tae Hwan Oh Advanced-Functional-Material-Modified Electrodes for the Monitoring of Nitrobenzene: Progress in Nitrobenzene Electrochemical Sensing Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1884, doi:10.3390/pr12091884	301
Anna Kamińska-Dwórznicza and Anna Kot The Physical Properties and Crystal Structure Changes of Stabilized Ice Cream Affected by Ultrasound-Assisted Freezing Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1957, doi:10.3390/pr12091957	322
Van-Long Trinh and Chen-Kuei Chung Use of Triboelectric Nanogenerators in Advanced Hybrid Renewable Energy Systems for High Efficiency in Sustainable Energy Production: A Review Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1964, doi:10.3390/pr12091964	336
Sergio Taraglio, Stefano Chiesa, Saverio De Vito, Marco Paoloni, Gabriele Piantadosi, Andrea Zanela and Girolamo Di Francia Robots for the Energy Transition: A Review Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1982, doi:10.3390/pr12091982	366
Difen Shi, Kai Bodemann, Yao Wang, Changliang Xu, Lulu Liu and Chungui Feng A Novel Nonsingular Fast Terminal Sliding Mode Control with Sliding Mode Disturbance Observer for Permanent Magnet Synchronous Motor Servo Control Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 1986, doi:10.3390/pr12091986	401
Mingdi Xu, Zhaoyang Jin, Shengjie Ye and Haipeng Fan Characteristic Canonical Analysis-Based Attack Detection of Industrial Control Systems in the Geological Drilling Process Reprinted from: <i>Processes</i> 2024 , <i>12</i> , 2053, doi:10.3390/pr12092053	414

Liangyu Li, Yaoyao Shi, Ye Tian, Wenyan Liu and Run Zou
Research on Dimension Reduction Method for Combustion Chamber Structure Parameters of
Wankel Engine Based on Active Subspace
Reprinted from: *Processes* **2024**, *12*, 2238, doi:10.3390/pr12102238 **433**

About the Editors

Sheng Du

Sheng Du received a B.S. degree in Measurement and Control Technology and Instrumentation and a Ph.D. degree in Control Science and Engineering from China University of Geosciences, Wuhan, China, in 2016 and 2021, respectively. He was a joint Ph.D. student with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada, from 2019 to 2021. He was then a Postdoctoral Fellow in Control Science and Engineering with China University of Geosciences, Wuhan, China, from 2021 to 2023. He is currently a professor with the School of Automation, China University of Geosciences, Wuhan, China. His research interests include process control, intelligent control, intelligent optimization, computational intelligence, and artificial intelligence. Dr. Du is a Senior Member of the Chinese Association of Automation (CAA) and a winner of the Outstanding Doctoral Thesis Nomination Award of the CAA. He is also an Associate Editor of the journal *Measurement and Control*.

Li Jin

Li Jin received a B.S. degree in Automation and a Ph.D. degree in Control Science and Engineering from China University of Geosciences, Wuhan, China, in 2016 and 2021, respectively. She was a joint Ph.D. student with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, U.K., from 2018 to 2020. She was then a Postdoctoral Fellow in Control Science and Engineering with China University of Geosciences, Wuhan, China, from 2022 to 2024. She is currently a professor in the School of Automation, China University of Geosciences, Wuhan, China. Her current research interests include power system stability analysis and control, time-delay systems, and robust theory and its applications.

Xiongbo Wan

Xiongbo Wan received a B.S. degree in Information and Computing Science from Huazhong Agricultural University, Wuhan, China, in 2006 and an M.S. degree in Probability Theory and Mathematical Statistics and a Ph.D. degree in Control Theory and Engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2011, respectively. From 2012 to 2013, he was a visiting scholar with Akita Prefectural University, Akita, Japan. From 2011 to 2014, he was a lecturer and then an associate professor with the College of Engineering, Huazhong Agricultural University. From 2016 to 2017, he was a visiting scholar with the Department of Computer Science, Brunel University London, Uxbridge, U.K. Since 2015, he has been with China University of Geosciences, Wuhan, where he is currently a professor with the School of Automation. His research interests include fault diagnosis, state estimation, and model predictive control of complex industrial systems.

Zixin Huang

Zixin Huang received B.S. and M.S. degrees in Engineering from Wuhan Textile University, Wuhan, China, in 2013 and 2016, respectively, and a Ph.D. degree from the School of Automation, China University of Geosciences, Wuhan, China, in 2020. He was a postdoctoral fellow with the Nankai University, Tianjin, China, in 2023. He is currently a postdoctoral fellow with the Shanghai Jiao Tong University, Shanghai, China, a senior visiting scholar at the University of Science and Technology of China, Anhui, China, and an associate professor with the School of Electrical and

Information Engineering, Wuhan Institute of Technology. His current research interests include nonlinear system control and intelligent control. Dr. Huang is also a senior member of the Chinese Association of Automation.



Article

Optimization of Anti-Skid and Noise Reduction Performance of Cement Concrete Pavement with Different Grooved and Dragged Textures

Biyu Yang^{1,2}, Songli Yang³, Zhoujing Ye³, Xiaohua Zhou⁴ and Linbing Wang^{5,*}

¹ Highway School, Chang'an University, Xi'an 710054, China; yang.yby@gmail.com

² Zhaotong Highway Investment Development Co., Ltd., Zhaotong 657099, China

³ National Center for Materials Service Safety, University of Science and Technology Beijing, Beijing 100083, China; songliyang96@gmail.com (S.Y.); yezhoujing@ustb.edu.cn (Z.Y.)

⁴ Yunnan Research Institute of Highway Science and Technology, Panlong District, Kunming 650051, China; hnzhehua@163.com

⁵ School of Environmental, Civil, Agricultural and Mechanical Engineering, University of Georgia, Athens, GA 30602, USA

* Correspondence: linbing.wang@uga.edu

Abstract: Cement concrete pavements are crucial to urban infrastructure, significantly influencing road safety and environmental sustainability with their anti-skid and noise reduction properties. However, while texturing techniques like transverse grooving have been widely adopted to enhance skid resistance, they may inadvertently increase road noise. This study addressed the critical need to optimize pavement textures to balance improved skid resistance with noise reduction. Tests were conducted to assess the influence of surface texture on skid resistance and noise, exploring the relationship between texture attributes and their performance in these areas. The investigation examined the effects of texture representation methods, mean profile depth, and the high-speed sideways force coefficient (SFC) on noise intensity and pavement skid resistance. The findings revealed that transverse grooves significantly improved the SFC, enhancing skid resistance. In contrast, longitudinal burlap drag, through its micro- and macro-texture adjustments, effectively reduced vibration frequencies between the tire and pavement, thus mitigating noise. Utilizing the TOPSIS multi-objective optimization framework, an optimization model for pavement textures was developed to augment skid resistance and noise reduction at varying speeds. The results indicated that at 60 km/h, an optimal balance of groove width, depth, and spacing yielded superior skid resistance with a minimal noise increase. At 80 km/h, increased groove spacing and depth were shown to effectively decrease noise while maintaining efficient water evacuation. The optimal pavement texture design must consider the specific context, including traffic volume, vehicle types, and operating speeds. This study provides essential guidance for optimizing urban cement concrete pavement textures, aiming to diminish traffic noise and bolster road safety.

Keywords: cement concrete pavement; texture; grooving; dragging; tire/pavement noise; skid resistance

Citation: Yang, B.; Yang, S.; Ye, Z.; Zhou, X.; Wang, L. Optimization of Anti-Skid and Noise Reduction Performance of Cement Concrete Pavement with Different Grooved and Dragged Textures. *Processes* **2024**, *12*, 800. <https://doi.org/10.3390/pr12040800>

Academic Editor: Francisco Vazquez

Received: 11 March 2024

Revised: 9 April 2024

Accepted: 12 April 2024

Published: 16 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As urbanization intensifies, the functionality and growth of cities increasingly hinge on road traffic efficiency. The surge in traffic volume, coupled with stringent environmental standards, has accentuated the need for diligent urban road functional health monitoring. This includes a thorough assessment and maintenance of road surface conditions, with a keen focus on safety and environmental attributes. Cement concrete pavement, a predominant choice for urban roads, plays a crucial role in this context, particularly due to its anti-skid and noise dampening properties [1–4]. Optimal skid resistance is vital for ensuring vehicular safety in adverse weather conditions like rain and snow [5,6].

Ensuring optimal skid resistance is paramount to vehicular safety, especially under adverse weather conditions such as rain or snow. Concurrently, the mitigation of road traffic noise, a significant source of urban pollution, is essential for enhancing the urban living experience [7,8].

Grooving techniques, esteemed for their cost-efficiency and simplicity, are prevalently employed to augment the performance of cement concrete pavements. Yet, the efficacy of these techniques in enhancing skid resistance and reducing tire/pavement noise is profoundly influenced by a multitude of factors including groove dimensions, spacing, and orientation, and vehicular attributes like tire patterns, weight, and velocity.

The relationship between pavement performance and its texture characteristics—both macro and micro—is well-established in the realm of skid resistance. Extensive studies leveraging traditional and advanced methodologies have scrutinized the influence of texture dimensions on skid resistance, utilizing tools ranging from the British Pendulum Tester (BPT) and Surface Friction Tester (SFT) to high-speed sideway force coefficient testing vehicles and high-resolution laser scanners. These investigations have elucidated that nuanced adjustments to texture parameters can markedly affect key metrics like the high-speed sideway force coefficient (SFC), mean profile depth (MPD), and British Pendulum Number (BPN). For instance, a texture depth of 2 mm and adequate surface area are pivotal for optimal pavement friction performance and longevity, with a texture wear rate beyond 40% significantly impairing braking skid resistance [9–14]. Innovative studies integrating optical scanning, image analyses, and advanced modeling techniques have further refined the precision in predicting skid resistance, offering novel insights into the intricate interplay between texture characteristics and skid resistance [1,15–20]. These advancements provide invaluable technical and theoretical underpinnings for the design of effective anti-skid pavements. Globally, research institutions have rigorously evaluated a spectrum of surface textures on experimental road sections, striving to pinpoint the ideal anti-skid texture configuration. In the USA, the focus is on optimizing transverse grooves with parameters like a groove width of 2.3–3.2 mm, depth of 3.2–4 mm, and spacing of 12–25 mm that significantly boost friction coefficients and safety [8,21–23]. Denmark employs grooving or embossing techniques coupled with seasonal treatments to elevate skid resistance [24]. Japan advocates for specific groove configurations on airport pavements to ensure safety, while in China, specific groove patterns are recommended to bolster lateral anti-skid performance [25–27]. These comprehensive studies offer substantial practical insights and technical recommendations for the crafting of anti-skid pavement designs.

In noise reduction performance, the On-Board Sound Intensity (OBSI) method is commonly used for noise level assessments. Cement pavement noise levels increase with macro-texture depth, but under specific conditions, a shallower texture depth (e.g., 0.8 mm) can effectively reduce noise. Additionally, increased pavement texture complexity and International Roughness Index (IRI) also raise noise levels [28–30]. New three-dimensional pavement texture metrics have proven superior in noise evaluation for unevenly textured surfaces. Moreover, texture-based noise models, such as the Statistical Energy Analysis (SEA) and Vibration Model (VM) algorithms, have enhanced our understanding of the relationship between texture level and noise frequency, with digital image processing revealing the critical role of micro-texture in noise reduction [31–33]. Material-wise, rubberized pavements show significant potential in reducing high-frequency noise. Comparisons indicate that small-particle exposed-aggregate concrete pavements outperform transverse grooved surfaces in noise reduction, demonstrating that precise texture spacing and depth adjustments can effectively lower noise in specific frequency bands [27,34–38]. Regarding transverse grooves' impact on noise, the spacing between grooves has the most significant correlation with noise levels. Uniformly spaced transverse grooves create harsh tonal peaks, whereas non-uniformly spaced grooves can effectively eliminate these peaks, although the overall noise may be the same or increased. Combining angled and varied spacing grooves can effectively eliminate tonal peaks and reduce overall noise, but angled grooves introduce construction challenges [39–44].

There remains a relatively narrow focus within the research community on the collective impact of pavement texture on anti-skid efficiency and noise reduction. Recent researches have utilized clustering analyses to classify pavement characteristics, advocating for benchmarks such as a minimum high-speed sideway force coefficient (SFC) of 50 for tunnel pavements to safeguard against skidding, and a noise threshold of 105 dB to preserve a comfortable environment [45]. Investigations into assorted pavement designs have highlighted the benefits of wider groove spacing—extending up to 25 cm—not only for its efficacy in curtailing pumping noise but also in averting lateral vehicular slippage, thus endorsing it as a sound methodological choice for cement concrete pavements in tunnel sections of highways [46]. Although there has been a concerted effort to unravel the interplay between pavement textures and their skid-resistant and noise-dampening properties, research that meticulously examines the balance between enhancing skid resistance and suppressing noise is sparse. This observation points to an existing gap in research—an oversight in the comprehensive optimization of pavement texture design, which is pivotal for advancing the field.

Leveraging the existing research, this study conducted an in-depth analysis of how textures affect pavement anti-skid properties and noise reduction capabilities. Section 2 outlines the field experiments that examined both the grooving and burlap drag texture parameters, in addition to key pavement performance metrics such as mean profile depth, high-speed sideway force coefficient, and tire/pavement noise levels. In Section 3, the focus shifts to delineating texture characteristics, followed by an exploration of the relationship between grooved and burlap drag textures and the pavement's skid resistance. Section 4 ventures into the realm of tire–pavement noise characteristics, analyzing the impact of various texture characteristics on noise levels associated with pavements. Utilizing the TOPSIS multi-objective optimization approach in Section 5, the study assessed eight experimental scenarios on the test sections, balancing considerations of skid resistance and noise reduction, to identify optimal texture configurations for speeds of 60 km/h and 80 km/h. This research aimed to pinpoint the ideal pavement texture design that maximizes overall performance, thereby enhancing driving safety and minimizing noise pollution.

2. Tire/Pavement Noise and Skid Resistance Assessment

2.1. Experimental Setup

Field evaluations were conducted on a cement concrete pavement in Zhaotong City, Yunnan Province, spanning a 28.7 km stretch of open road. The surface texture was engineered using two distinct methods: transverse grooving and longitudinal burlap dragging. The grooving involved the use of an electric machine to cut into the concrete, creating grooves with predefined dimensions, thereby modifying the surface texture. The dragging technique employed a steel frame to pull double-layered coarse burlap over the fresh concrete, with the fabric's friction against the soft surface generating a unique texture. The processes are depicted in Figure 1.

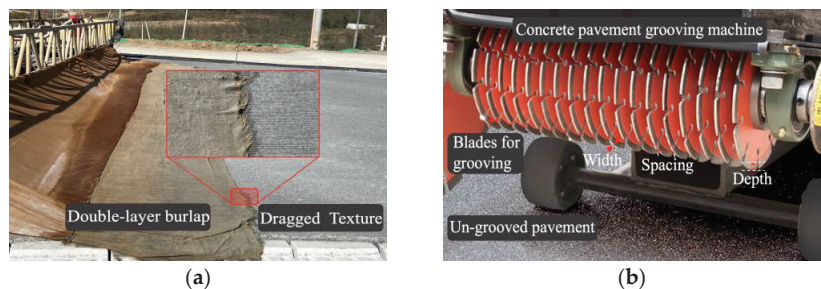


Figure 1. Pavement texture creation at test sites: (a) burlap drag and (b) grooved textures.

Research by the Federal Highway Administration in the United States indicates that the depth, width, spacing, and orientation of transverse grooves are crucial factors influencing tire/pavement noise. Among these, groove spacing exhibits the strongest correlation with noise intensity levels [39,40]. To corroborate these findings, the experiment included sections with three variations of transverse grooves at uniform spacings of 13/19/25 mm, two variations with non-uniform spacings averaging 17 mm and 28.8 mm, and several conditions for longitudinal burlap drag depths. Each design spanned a 200 m section. Figure 2 presents the test site surface conditions.

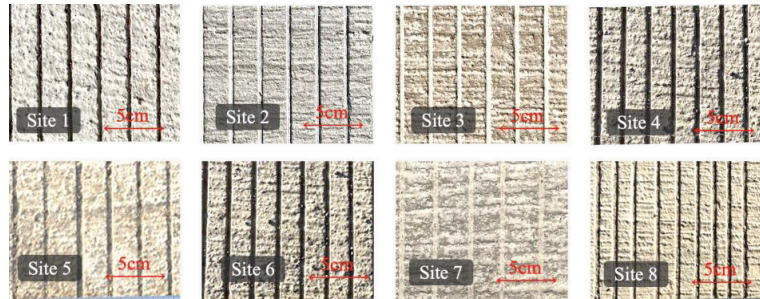


Figure 2. Test site surface conditions.

2.2. Test Methodology

2.2.1. Pavement Texture

To assess and characterize the surface texture of cement concrete pavements, this study employed the HC-CK103 crack measuring instrument from HICHANCE (Beijing, China), which can capture intricate texture images, as illustrated in Figure 3a. It uses an optical lens to magnify the crack, and combines built-in image processing technology to display the crack width on the instrument's scale dial. This device boasts a precision of 0.01 mm and a measurement range from 0 to 10 mm, as shown in Figure 3b. For attributes exceeding the HC-CK103's measurement capabilities, particularly concerning macro-texture such as groove spacing, a millimeter-scale ruler was utilized for the measurements.

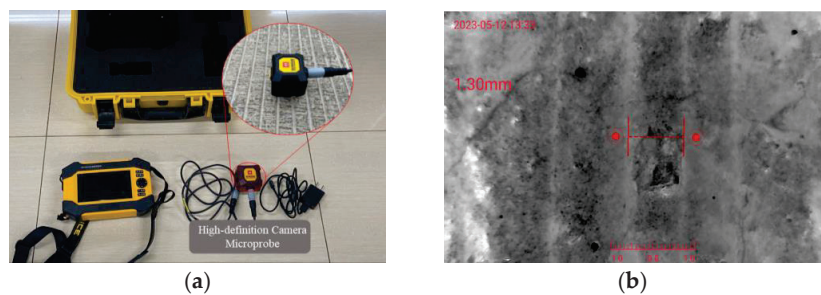


Figure 3. Texture measurement systems. (a) Texture captured by a high-definition camera microprobe; (b) measurement of the dragged texture.

2.2.2. Mean Profile Depth

An RTM-type vehicle-mounted intelligent road detection system, developed by Wuhan University, was utilized to evaluate the road's evenness and mean profile depth, as depicted in Figure 4. Testing was conducted at a speed of 50 km/h, with each 100 m segment considered as a separate detection unit for mean profile depth calculations.



Figure 4. RTM-type vehicle-mounted intelligent road detection system.

2.2.3. Tire/Pavement Noise

To ascertain tire/pavement noise on cement concrete surfaces, the experiment harnessed the OBSI noise testing system developed by the AVEC Corporation in the USA. This system comprises two arrays, each with two GRAS 26CA CCP Intensity Probes, strategically placed near the tire/pavement interface to precisely capture noise produced during movement, while isolating other noise sources. These probes can accurately gather noise data across a frequency spectrum of 2.5 Hz–200 kHz, featuring a noise level of 1.8 μ V Gain and a gain of -0.30 dB. The system's configuration is portrayed in Figure 5, with the captured noise signals being recorded at 26,500 Hz by the NI cDAQ-9171 and instantaneously relayed to the AVEC's OBSI Software (<https://www.avec-engineering.com/OBSI.html>, accessed on 9 March 2024).

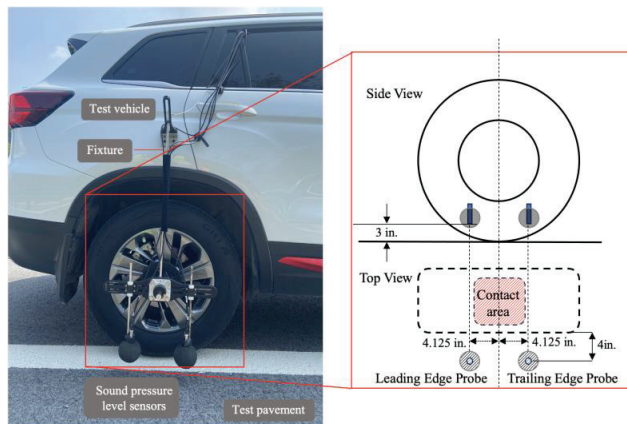


Figure 5. Noise testing system schematic.

For this study, a Changan CS75 vehicle equipped with YOKOHAMA 225/65R17 tires was chosen, conforming to ASTM F 2493 specifications. Testing encompassed various pavement texture types, striving to maintain speeds as close to 60 km/h and 80 km/h as possible over a 200 m travel distance. The procedures and conditions were aligned with the AASHTO T 360-16 standard [47], ensuring data reliability through 2–3 repeated tests on most pavement surfaces at consistent speeds.

2.2.4. High-Speed Sideway Force Coefficient

The MCY-1-type pavement friction coefficient testing system was deployed to gauge the lateral friction coefficient of the road sections under study, as displayed in Figure 6. Throughout the testing, the test tire's static vertical load was consistently held at 2000 ± 20 N, with the standard tire pressure maintained at 3.5 ± 0.2 kg/cm². The vehicle traversed the pavement at a steady 50 km/h, performing continuous measurements over the pavement where each 100 m segment was treated as an individual evaluation unit.

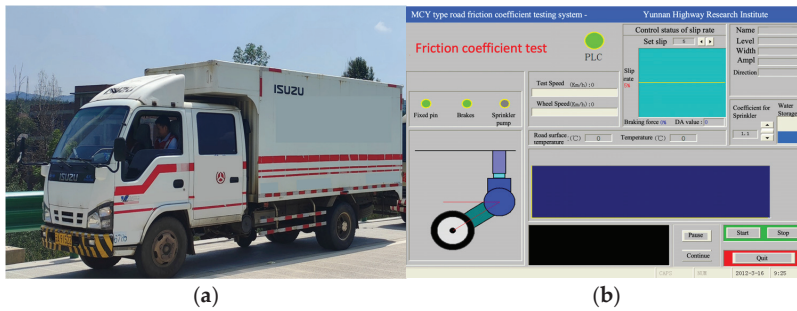


Figure 6. MCY-1-type pavement friction coefficient testing system and system operating interface. (a) The vehicle-mounted MCY-1-type testing system; (b) testing system interface.

3. Pavement Texture Characteristics and Skid Resistance Analysis

3.1. Texture Characteristic Description

Different texture techniques generate distinct pavement surface profiles. Grooving results in undulations perpendicular to the travel direction, creating longitudinal texture contour curves. Conversely, burlap dragging yields textures parallel to the travel direction, resulting in transverse contour curves, as illustrated in Figure 7. The interweaving of textures from burlap dragging and grooving sketches a grid-like design on the pavement. To conduct a comprehensive evaluation, this study identified three critical parameters, groove width, groove depth/height, and groove spacing, which were measured for various grooved texture types.

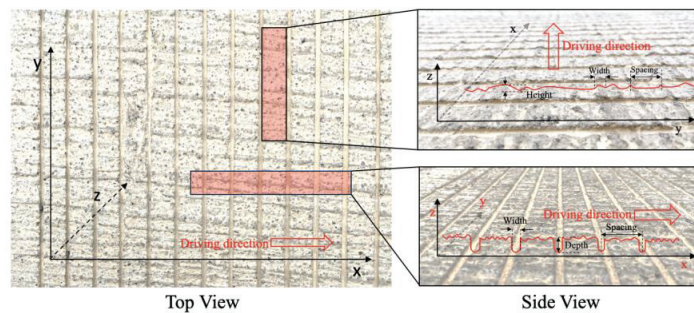


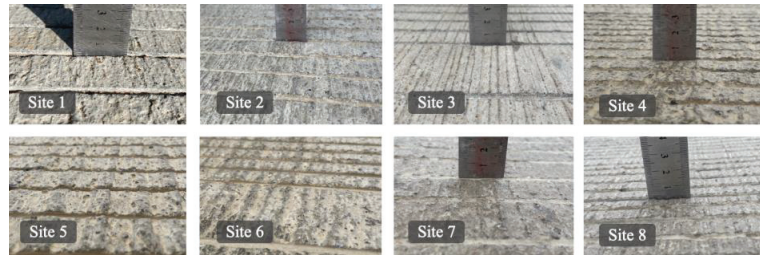
Figure 7. Description of texture characteristic parameters.

Construction process variability leads to fluctuations in actual groove width, depth, and spacing. Groove textures, produced through mechanical cutting, exhibit minimal variations. Therefore, after multiple point measurements, mean values were computed to represent each condition's texture characteristic values. Table 1 details the groove texture parameters for the evaluated pavement sites.

Table 1. Grooved texture parameters at various sites.

Site No.	Method	Width (mm)	Depth (mm)	Spacing (mm)
1	Transverse Grooves with Equal Spacing	4.0	2	25
2		3.9	1	25
3		4.0	0.75	25
4		3.5	1	19
5		3.7	2.25	25
6		4.0	1.5	13
7	Transverse Grooves with Unequal Spacing	4.0	0.5	28-30-22-31-27-32-28-30-27-33-24-34-29 (375/13 = 28.8)
8				13-17-19-17-13-16-18-20-17-13-24-17-13-17-19- 16-15-17-18-20-22-17-16 (390/23 = 17.0)

Due to the predominant formation of raised grooves in burlap dragging, this study utilized the groove height, width, and spacing as evaluation metrics, akin to the system used for grooved textures. Repeated measurements of the burlap drag texture in specific zones gathered extensive data to capture the variability of burlap drag textures under differing conditions. Figure 8 displays the burlap drag textures on eight pavement types, with Table 2 listing the burlap drag texture parameters at these sites.

**Figure 8.** Burlap drag textures at various sites.**Table 2.** Burlap dragged texture parameters at various sites.

Site No.	Method	Width (mm)	Height (mm)	Spacing (mm)
1	Longitudinal burlap drag	1.5	1.0	12.5
2		1.6	0.5	3.2
3		1.0	0.9	6.5
4		3.1	1.5	8.4
5		3.1	1.8	9.1
6		2.2	1.1	4.9
7		1.7	0.8	8.8
8		2.0	1.0	7.4

3.2. Analysis of Mean Texture Depth

The Mean Texture Depth (MTD) serves as a critical measure of the pavement's texture depth, illustrating the surface's roughness and its substantial impact on the tire–pavement interactions and traction. Generally, a higher MTD signifies enhanced water drainage and improved skid resistance. The MTD values for different pavement texture sites are listed in Table 3.

Table 3. Mean texture depth at 8 test sites.

Site No.	1	2	3	4	5	6	7	8
MTD (mm)	0.955	0.77	0.79	0.84	0.775	0.82	0.535	0.87

The observed variation in MTD values across test sites 1 to 8 (ranging from 0.535 mm to 0.955 mm) underscores the influence of texture parameters on the pavement's water drainage and grip. Notably, broader and deeper grooves contributed to superior water displacement, likely resulting in increased MTD values. For instance, site 1, characterized by a deeper groove (2 mm) and wider width (4.0 mm), aligns with the highest MTD value (0.955 mm), suggesting effective water drainage and potential for enhanced skid resistance. Conversely, site 7 exhibited a lower MTD value (0.535 mm) due to its shallower groove depth (0.5 mm), despite a comparable groove width (4.0 mm), indicating reduced drainage efficiency. Additionally, the texturing technique, including equal and unequal transverse spacing, plays a role in determining MTD values, with uneven spacing potentially affecting water flow and, consequently, MTD values.

3.3. Analysis of High-Speed Sideway Force Coefficient

A greater SFC value denotes superior skid resistance, providing more effective lateral grip, minimizing vehicle lateral movement, and thereby boosting driving stability and safety. The SFC values were continuously assessed at each test site, where each segment of 100 m was delineated as an individual unit. Two such units were established, and their respective SFC values were computed as an average. The SFC values for the various sites are presented in Table 4.

Table 4. High-speed sideway force coefficient at 8 test sites.

Site No.	1	2	3	4	5	6	7	8
SFC	64	62.9	60.2	65.9	61.3	57.1	59.8	63.3

The findings suggest that, in grooved textures, an increased groove width or depth typically correlates with enhanced friction points and improved water removal capability, contributing to higher SFC values. Sites 1 and 5, with their wider and deeper grooves, exhibited higher SFC values, indicating strong lateral traction. Interestingly, site 4 showed the highest SFC despite site 6 having the densest texture, indicating that optimal groove spacing can enhance the contact area between the tires and pavement, thus boosting the SFC. Comparatively, sites with uniform texture spacing, like sites 4 and 8 or 3 and 7, demonstrated higher SFC values, presumably due to the consistent grip provided by even spacing, which helps stabilize the SFC values. Variable spacing may result in inconsistent lateral grip, potentially impacting SFC stability.

In the context of burlap drag textures, variations in texture parameters significantly affected the contact nuances between the pavement and tires. While wider grooves might reduce local contact areas, an optimal width can elevate friction edges, thus improving the SFC. Sites 4 and 5, with their broader grooves, are likely to offer enhanced edge friction, influencing their SFC values. Groove spacing impacts texture density and continuity; tighter spacing in burlap drag textures could provide more continuous friction edges, aiding in lateral grip. However, excessively close spacing might lead to friction saturation, failing to provide additional lateral traction.

3.4. Correlation Analysis

To quantify the strength of the association between the parameters of grooved and burlap drag textures and the skid resistance indicators of pavement, a correlation analysis was performed, relating these textures to MTD and SFC values. Table 5 displays the

correlation coefficients, and Figure 9 illustrates the relationships between the various texture parameters and MTD and the SFC at 8 test sites.

The analysis indicated that groove textures significantly influenced MTD and the SFC, underscoring the importance of texture dimensions on pavement performance. The depth of the groove emerged as a critical factor affecting MTD, whereas the width predominantly influenced the SFC. The burlap drag textures exhibited a subtler impact, likely affecting the micro-texture characteristics rather than macro-texture attributes.

Table 5. Correlation coefficients between groove, burlap drag textures, and pavement MTD and SFC.

Performance Index	Burlap Drag Texture			Groove Texture		
	Width	Depth	Spacing	Width	Height	Spacing
MTD	0.0717	0.2161	0.2128	−0.0912	0.5397	−0.4835
SFC	0.2516	0.1497	0.3669	−0.5453	0.0504	0.1210

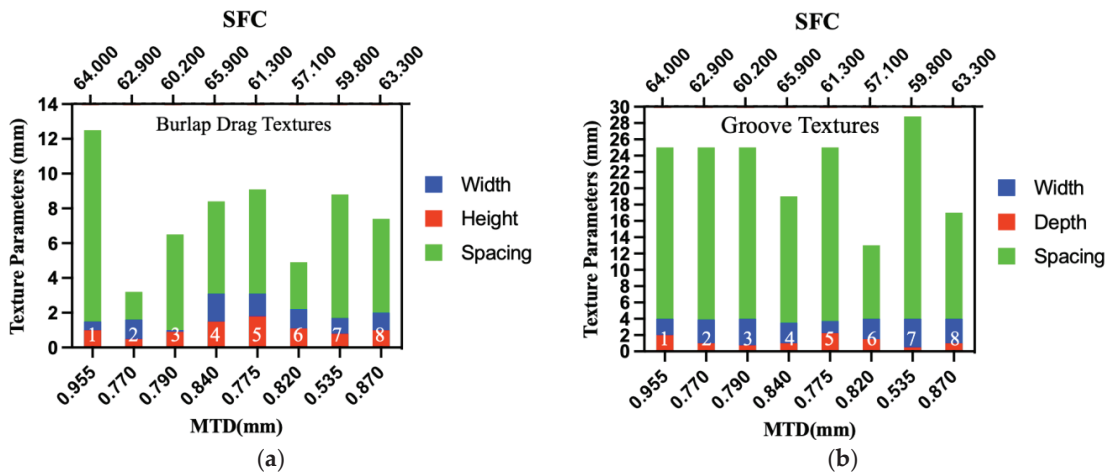


Figure 9. Relationship between groove and burlap drag textures and MTD and SFC at test sites 1 to 8. (a) Burlap drag with MTD and SFC; (b) groove textures with MTD and SFC.

A strong positive correlation was observed between groove depth and MTD (0.5397), suggesting that deeper grooves result in greater texture depth, as depicted in Figure 9b. Conversely, the relationship between groove depth and the SFC was minimal (0.0504), indicating that depth alone does not directly impact surface friction. Groove width negatively correlated with MTD (−0.0912), likely because wider grooves reduce the per-unit-area texture, diminishing the texture depth. Moreover, wider grooves decreased the effective contact area between pavement and tire, resulting in a notable negative correlation with the SFC (−0.5453). Groove spacing influenced texture continuity and density, leading to a negative correlation with MTD (−0.4835). As the spacing increased, the SFC first rose and then declined, suggesting that optimal spacing can enhance tire–pavement contact but after becoming excessively large, it causes uneven contact areas.

In burlap drag textures, the positive correlations of groove width, height, and spacing with MTD (0.0717, 0.2161, 0.2128) were relatively weak, as illustrated in Table 5, reflecting the limited contribution of these smaller characteristic sizes to the overall texture depth. The positive correlations of groove width, height, and spacing with the SFC (0.2516, 0.1497, 0.3669) indicate that these parameters notably influence the SFC, possibly by affecting the texture density and distribution, thereby enhancing the micro-contact points between the tires and pavement.

4. Pavement Texture Characteristics and Noise Reduction Performance Analysis

4.1. Tire–Pavement Noise Characterization

Tire–pavement noise was recorded at 60 km/h and 80 km/h using the On-Board Sound Intensity (OBSI) noise testing system, which provided sound pressure amplitude–time curves for the tire–pavement interface. Time-domain signals were subjected to Fast Fourier Transform, as delineated in Equation (1) [48], to calculate the A-weighted sound level (overall level OA) for the test duration. The OA value was determined following the sound pressure level definition, yielding the aggregate sound pressure level for the time-domain signal, as expressed in Equation (2) [48], indicative of the peak noise intensity level for the roadway segment under study, as detailed in Table 6. For safety considerations, data at the higher speed of 80 km/h were not collected.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \quad k = 0, \dots, N-1 \quad (1)$$

$$L_{Aeq} = 10 \lg \left[\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \left(\frac{P_A^2(t)}{P_0^2} \right) dt \right] (dBA) \quad (2)$$

where $P_A(t)$ —instantaneous A-weighted sound pressure level of the noise model in pascals (Pa); P_0 —reference sound pressure in micropascals (μ Pa), with a reference sound pressure of 20 μ Pa; and $t_2 - t_1$ —interval of measurement time period T in seconds (s).

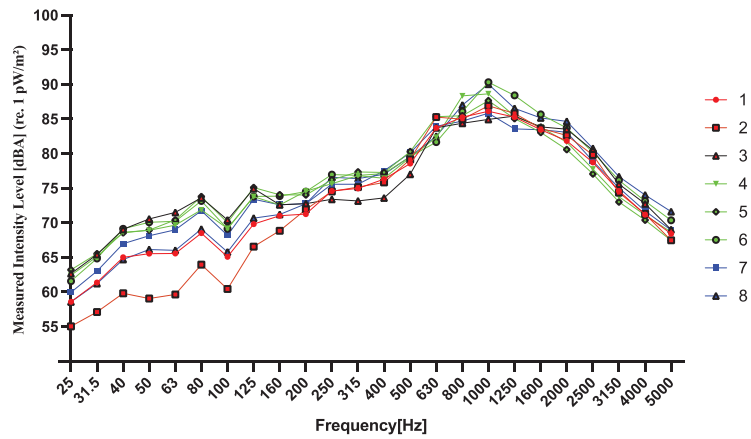
Table 6. A-weighted sound levels (dBA) of tire–pavement noise at 8 test sites.

Test Speed	1	2	3	4	5	6	7	8
60 km/h	91.0	92.1	91.1	92.5	91.7	92.4	92.7	93.3
80 km/h	98.1	97.4	97.7	97.0	96.9	96.8	-	-

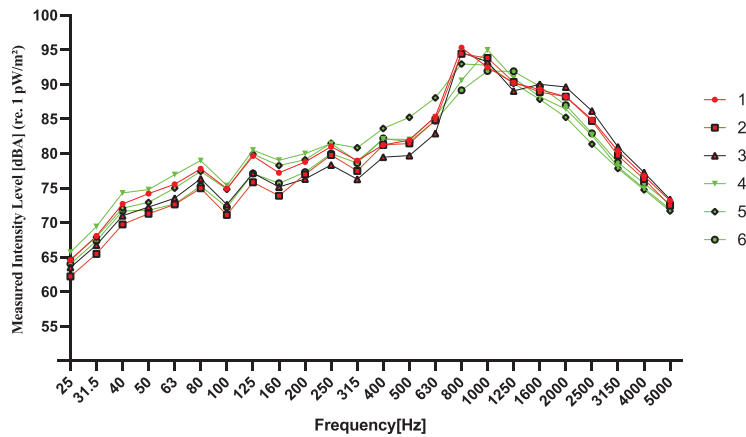
To dissect the influence of various frequency components on the overall noise, Fast Fourier Transform (FFT) and 1/3 octave band analysis were performed on the gathered time-domain signals. This analysis elucidated the distribution of tire–pavement noise across the frequency spectrum under different conditions, as shown by curves 1 to 8 in Figure 10a for 60 km/h and 1 to 6 in Figure 10b for 80 km/h.

The test findings revealed that at 60 km/h, sites 7 and 8 exhibited the highest tire–pavement noise levels, while sites 1 and 3 had the lowest levels. An analysis of their burlap drag and grooving parameters showed that sites 7 and 8 featured non-uniformly spaced grooves, whereas sites 1 and 3 had uniformly spaced burlap textures, which were more consistent overall, suggesting that such uniformity in texture can mitigate random vibrations induced by irregular textures, thus diminishing noise levels. At 80 km/h, the lowest noise levels were recorded at sites 4 and 5 (groove heights over 3 mm and widths over 1.5 mm), where the groove textures were notably narrower, which might have lessened air vibrations and compression during tire rolling. Moreover, the more pronounced groove heights and widths at these sites suggest that rougher textures can reduce noise generated by smoother surfaces, especially at increased speeds.

The 1/3 octave band analysis indicated an upward trend in noise levels with frequency, with all eight test conditions exhibiting tonal peaks around 1000 Hz. At higher speeds, the noise contribution from frequencies above 1000 Hz became more pronounced. Upon comparing texture profile levels with tire–pavement noise trends across the conditions, varying patterns were noted across different frequency ranges, implying that the interplay between pavement texture and tire–pavement noise may shift across different frequency bands, underscoring the need to assess the influence of texture parameters on noise across the spectral range.



(a)



(b)

Figure 10. The 1/3 octave frequency spectrum of tire–pavement noise at (a) 60 km/h; (b) 80 km/h.

4.2. Analysis of Pavement Texture's Influence on Noise Intensity

4.2.1. Spacing

At a speed of 60 km/h, site 1 exhibited superior noise levels compared to sites 4 and 6. However, at 80 km/h, site 6 demonstrated the best noise performance. Figure 11 provides a noise spectrum analysis for the various evenly spaced grooves and burlap drag textures at both 60 km/h and 80 km/h, aiming to assess the effect of groove spacing on noise at differing velocities.

The data in Figure 12 indicate that noise intensities above 85 dBA are predominantly located between 630 Hz and 2000 Hz, with all three sites showcasing tonal peaks at 1000 Hz. At 60 km/h, site 1, featuring 25 mm evenly spaced grooves, displayed consistently lower noise levels across the frequency spectrum compared to sites 4 (19 mm spacing) and 6 (13 mm spacing). However, at 80 km/h, site 6 presented reduced noise levels within the 25 to 1000 Hz range, while site 1 had elevated levels within the 1000 to 5000 Hz range. This suggests that larger spacings can diminish air vibrations and compression as tires traverse texture gaps at lower speeds, thereby reducing air pumping noise. Consequently, the smallest spacing at site 6 (13 mm) may lead to increased air pumping noise, culminating in the highest noise level among the sites.

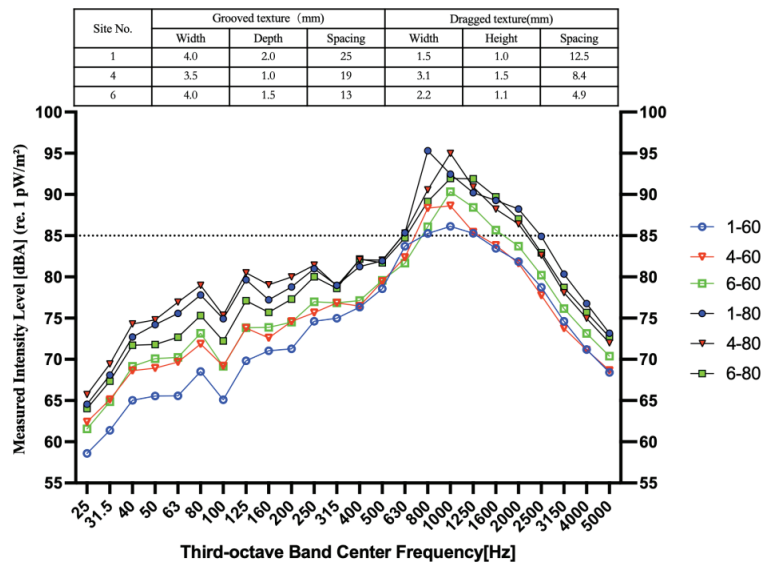


Figure 11. Noise spectrum analysis for sites 1, 4, and 6.

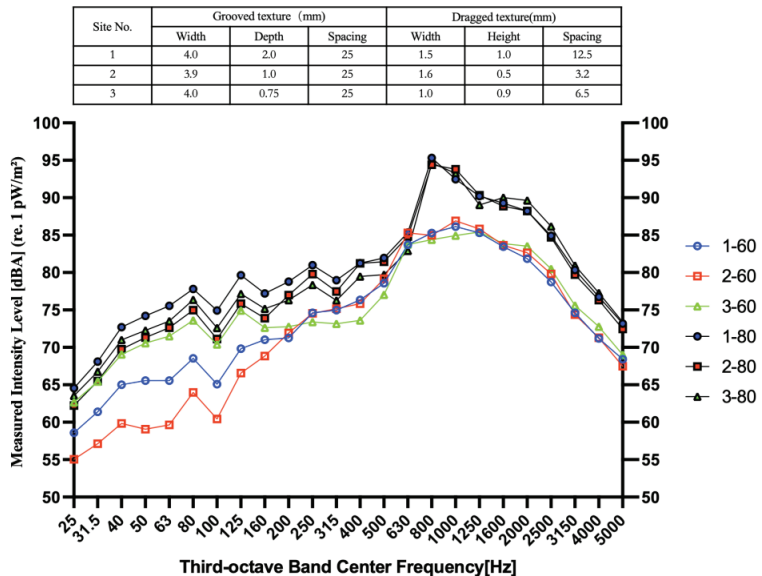


Figure 12. Noise spectrum analysis for sites 1, 2, and 3.

Upon acceleration from 60 km/h to 80 km/h, sites 1 and 4 showed amplified tonal peaks, with the peak frequency transitioning from 1000 Hz to 1250 Hz, indicating an uptick in high-frequency noise components. This frequency-specific noise could be attributed to the air pumping effect, where rapid air compression and release occurred as the tires engaged with the grooved textures. Smaller spacings could mitigate noise by intensifying the frequency of this air pumping phenomenon, whereas larger spacings may not disperse this effect as effectively. Therefore, site 1 with a 25 mm spacing exhibited a tonal peak at an 800 Hz center frequency, whereas site 4 with 19 mm spacings demonstrated a peak at

1000 Hz. Additionally, the increase in speed likely accentuated tire–pavement friction and vibration, influencing the noise levels. Site 6, with the narrowest spacing in burlap drag texture, could enhance tire–pavement interactions at elevated speeds, reducing the noise generated by friction and vibrations.

4.2.2. Depth/Height

At 60 km/h, test sites 1 and 3 demonstrated superior noise levels compared to site 2, but at 80 km/h, site 2 outperformed in terms of noise levels. To investigate the effect of groove depth on noise at different velocities, Figure 12 provides a noise spectrum analysis for various groove depths at 60 km/h and 80 km/h.

At 60 km/h within the 25 Hz to 250 Hz frequency band, site 2, with a 1 mm groove depth, showed consistently lower noise intensities across all frequencies compared to sites 1 (2 mm depth) and 3 (0.75 mm depth). Yet, within the 500 Hz to 1600 Hz range, site 2 recorded the highest noise intensity, indicating a higher overall noise level. This phenomenon suggests that a moderate groove depth at lower frequencies can disrupt the air pressure vibrations occurring during tire–pavement compression, which is not as effectively achieved by excessively deep or shallow grooves. At mid to high frequencies, the groove depth may amplify resonances at specific frequencies, particularly those aligning with tire vibration modes, thus increasing the noise. Site 2, exhibiting the smallest groove height and narrower texture spacing, has more intimate tire–pavement interactions, minimizing the voids created by texture. This enhanced contact contributed to the reduction of vibrations stemming from the tire compression of air within the pavement textures, an effect that was more noticeable at lower frequencies.

At 80 km/h, site 2 displayed lower noise levels in the 25–630 Hz range, whereas sites 1, 2, and 3 showed comparable noise levels in the 1000–5000 Hz range. This pattern suggests that at higher speeds, the influence of groove depth on high-frequency noise is mitigated. High-frequency noise appears to correlate more closely with tire vibration modes and aerodynamic noise rather than groove depth, diminishing the latter’s relative significance.

5. Optimization of Skid Resistance and Noise Reduction Texture Using TOPSIS

5.1. Method Overview

This study’s evaluation of the high-speed sideway force coefficient and tire–pavement noise levels provide insights into the performance of various pavement textures. To devise an optimal strategy that boosts pavement skid resistance while mitigating tire/road noise, a multi-objective optimization analysis was undertaken. Given the small sample size, the TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method was employed for a comprehensive assessment of each texture design [48]. This technique, which does not rely on the distribution pattern of the data, is appropriate for the limited empirical data available. The core principle involves determining the performance of each option across multiple criteria, and then identifying the optimal and least optimal solutions. The TOPSIS approach selects the most suitable option by calculating the distance of each alternative from the optimal and least optimal solutions, favoring the choice nearest to the optimal while the farthest is identified as the least optimal.

TOPSIS Calculation Steps: (1) Construct the decision matrix: Assuming there are m evaluation objects and n criteria, form decision matrix X :

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

where x_{ij} represents the performance value of the i -th option under the j -th criterion.

(2) Normalize to eliminate the impact of different units and magnitudes, with the normalized r_{ij} as shown in Equation (3):

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}} \tag{3}$$

(3) Introduce weights w_j for each criterion, forming the weighted normalized decision matrix:

$$v_{ij} = w_j \times r_{ij} \tag{4}$$

where w_j is the weight of the j -th criterion.

(4) Determine the ideal best (A^*) and worst (A^-) solutions:

$$A^* = \{ \max v_{ij} | j \in J^+, \min v_{ij} | j \in J^- \} \tag{5}$$

$$A^- = \{ \min v_{ij} | j \in J^+, \max v_{ij} | j \in J^- \} \tag{6}$$

where J^+ is the set of beneficial criteria, and J^- is the set of non-beneficial criteria.

(5) Calculate the Euclidean distance of each option from A^* and A^- :

$$D_i^* = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^*)^2} \tag{7}$$

$$D_i^- = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^-)^2} \tag{8}$$

(6) Compute the relative closeness to the ideal solution for each option, C_i :

$$C_i = \frac{D_i^-}{D_i^* + D_i^-} \tag{9}$$

The closer C_i is to 1, the closer the option is to the ideal solution.

5.2. Data Preprocessing

The dataset comprises texture parameters from 8 test sites and the corresponding performance metrics. The texture parameters were categorized into grooving and burlap drag, measured by width, depth, and spacing, as detailed in Tables 1 and 2. The performance metrics included the pavement’s high-speed sideway force coefficient and noise levels at 60 km/h and 80 km/h, as listed in Tables 4 and 6.

To mitigate the influence of diverse scales and magnitudes, all parameters and results were normalized using Z-score normalization, resulting in a dataset with a mean of 0 and a standard deviation of 1, as illustrated in Table 7.

Table 7. The normalized dataset.

Site No.	Burlap Drag Texture			Groove Texture			Noise		SFC
	Width	Depth	Spacing	Width	Height	Spacing	60 km/h	80 km/h	
1	-0.6984	-0.1846	1.7287	0.5968	1.2247	0.5233	-1.3843	1.5313	0.7875
2	-0.5654	-1.4156	-1.5523	0.0063	-0.4082	0.5233	0	0.1629	0.3915
3	-1.3636	-0.4308	-0.3881	0.5968	-0.8165	0.5233	-1.2585	0.7494	-0.5805
4	1.4302	1.0463	0.2822	-2.0556	-0.4082	-0.6082	0.5034	-0.6191	1.4715
5	1.4302	1.7848	0.5292	-0.9946	1.633	0.5233	-0.5034	-0.8145	-0.1845
6	0.2328	0.0615	-0.9526	0.5968	0.4082	-1.7397	0.3775	-1.01	-1.6966
7	-0.4324	-0.677	0.4234	0.5968	-1.2247	1.2399	0.7551	-	-0.7245
8	-0.0333	-0.1846	-0.0706	0.5968	-0.4082	-0.9853	1.5101	-	0.5355

Given the potential variation in pavement texture's impact on noise at different speeds, the data for 60 km/h and 80 km/h were analyzed separately to gauge the speed's influence on the optimization outcomes. The absence of noise data for test sites 7 and 8 at 80 km/h necessitated their exclusion from the dataset for this speed segment.

5.3. Determination of Weights

5.3.1. Top-Level Weight

To elucidate the causal relationship between performance objectives and texture design, top-level weights were assigned to pavement performance indicators (anti-skid and noise reduction capabilities). The parameters for burlap drag and grooving textures were established as secondary weights, with their internal distribution reflecting the influence of various texture parameters on performance metrics.

The objective of this study was to identify an optimal texture design that enhances pavement skid resistance while minimizing tire/pavement noise. Consequently, equal importance was attributed to skid resistance and noise reduction, assigning a weight of 0.5 to each at the top level.

5.3.2. Internal Weight Setting for Grooved and Burlap Drag Textures

The behavior of textures in noise reduction varies with speed, necessitating a distinct discussion on internal weight assignment at 60 km/h and 80 km/h.

At 60 km/h, the grooved textures' width negatively correlated with the SFC and has a smaller impact on noise, warranting a 15% weight. Groove depth, which was positively correlated with the SFC and significant in reducing low-frequency noise, was allocated a 30% weight, emphasizing its crucial role. Groove spacing received a 55% weight due to its significant noise control contribution, highlighting its pivotal role in noise reduction. The burlap drag textures' width was assigned a 20% weight for its minor noise control role at low speeds. Groove depth was given a 30% weight for its impact on texture roughness and indirect effects on the SFC and noise, while spacing, crucial for noise control, was given a 50% weight to underline its central function.

At 80 km/h, the weight for groove depth decreased to 20% as its noise impact diminishes at higher speeds. Groove width was assigned a 15% weight, and spacing, still a key noise and SFC influencer, especially for high-frequency noise, was given a 65% weight. For burlap drag textures, the groove depth's weight was reduced to 25%, width was given a 20% weight, and spacing, maintaining its significance for noise and the SFC, was allocated a 55% weight. Table 8 presents the weight distribution across the different conditions and speeds.

Table 8. Weight distribution of pavement performance and texture parameters at different speeds.

Speed	Burlap Drag Texture Width	Texture Depth	Spacing	Groove Texture Width	Texture Height	Spacing	Anti-Skid	Noise Reduction
60 km/h	0.3	0.35	0.35	0.3	0.4	0.3	0.5	0.5
80 km/h	0.2	0.25	0.55	0.15	0.2	0.65	0.5	0.5

5.4. Results

Utilizing normalized texture parameters and performance data, the relative closeness index was calculated for each test site, considering anti-skid and noise reduction. Table 9 displays these indices across the varying speeds.

Table 9. Relative closeness indices at different speeds.

Speed	1	2	3	4	5	6	7	8
60 km/h	0.5603	0.4297	0.3521	0.6130	0.6069	0.4024	0.4836	0.5846
80 km/h	0.8023	0.4635	0.5157	0.5720	0.5881	0.2025	-	-

The results indicate that at 60 km/h, test sites 4 and 5 showed superior overall performance. Test site 4, with a groove width of 3.5 mm, depth of 1 mm, and spacing of 19 mm, offers a balanced configuration that may facilitate adequate water drainage while minimizing noise induced by excessively deep or narrow textures. For the burlap drag texture, test site 4's larger groove width (3.1 mm) and relatively higher groove height (1.5 mm) may improve the contact pressure distribution, reducing high-frequency vibrations and noise.

At 80 km/h, the texture design of test site 1 aligns most closely with the ideal solution. Featuring a groove width of 4.0 mm, a depth of 2 mm, and a spacing of 25 mm, this configuration is likely to enhance water drainage and provide adequate friction at high speeds, while the greater spacing assists in mitigating noise generated by air vibration. The burlap drag texture at test site 1, characterized by a groove width of 1.5 mm, a height of 1.0 mm, and a spacing of 12.5 mm, is expected to maintain effective contact stability and contribute to noise reduction at elevated speeds.

In essence, improving pavement skid resistance typically necessitates an increase in surface roughness or texture depth, facilitating enhanced friction between the tires and pavement, thereby boosting vehicle stability and safety. Nevertheless, such modifications can result in heightened noise levels due to the increased surface vibrations and air compression fluctuations associated with rougher textures. Consequently, the design of pavement texture should be tailored to the specific conditions of road use and the requirements of the intended users. For example, in urban or heavy traffic areas, skid resistance may be a priority. Additionally, the design should take into account the diversity of vehicle types, traffic volumes, speed ranges, and load categories, as the needs for skid resistance and noise reduction can vary significantly between heavy-duty vehicles and light passenger cars. Along this vein, the objective of optimal pavement texture design is to strike an ideal balance between skid resistance and noise reduction, tailored to the particular application context.

6. Conclusions

This research provided a detailed evaluation of the texture characteristics of cement concrete pavements and their effects on pavement's skid resistance and the levels of tire/pavement noise. Utilizing the TOPSIS multi-objective optimization method, various texture designs were appraised, culminating in a texture design scheme that optimally balances skid resistance and noise reduction. The key findings are as follows:

- (1) Grooved textures significantly enhance the high-speed sideway force coefficient (SFC), thereby markedly improving the pavement's skid resistance. The groove's width and depth are critical, particularly at lower speeds, where optimal dimensions support water displacement and deliver adequate traction without markedly heightening noise;
- (2) Dragged (burlap drag) textures directly influence noise levels, with their micro- and macro-textures diminishing vibration frequencies between the tires and pavement, thereby lowering noise. Within these textures, groove height and spacing are pivotal in managing noise levels and ensuring adequate skid resistance;
- (3) At 60 km/h, optimal skid resistance and noise reduction are achieved with a moderate groove width and depth alongside suitable spacing, exemplified by test site 4's configuration of transverse grooves (3.5 mm width, 1 mm depth, 19 mm spacing) and longitudinal dragged texture (3.1 mm width, 1.5 mm height, 8.4 mm spacing). At 80 km/h, wider grooves, more considerable spacing, and deeper cuts enhance drainage while reducing noise, as seen in test site 1's design featuring transverse grooves (4 mm width, 2 mm depth, 25 mm spacing) and longitudinal dragged texture (1.5 mm width, 1.0 mm height, 12.5 mm spacing);
- (4) Pavement texture design should be tailored to actual road use and user requirements. It is crucial to consider the diverse performance requirements of different vehicles, such as heavy versus light vehicles, and the distinct needs of urban versus intercity highways.

Future research should expand the TOPSIS method by including additional environmental and operational factors like tire types, weather conditions, and vehicle speed distribution to refine the model's precision and dependability. Moreover, establishing ongoing pavement performance monitoring is recommended to evaluate the real-world efficacy of texture designs, providing valuable empirical data to guide future texture design decisions.

Author Contributions: All authors made significant contributions to the work presented in this manuscript. Conceptualization, B.Y. and L.W.; methodology, S.Y., X.Z. and Z.Y.; investigation, X.Z. and Z.Y.; resources, B.Y.; writing—original draft preparation, B.Y. and S.Y.; writing—review and editing, Z.Y. and L.W.; supervision, L.W.; project administration, B.Y.; funding acquisition, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Technology Innovation and Demonstration Project of the Department of Transport of Yunnan Province (2021) No. 25. This work was also supported by the National Natural Science Foundation of China (No. 52208414) and Major Science and Technology Projects of Yunnan Province Science and Technology Department (No. 202302AD080007).

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: Author Biyu Yang was employed by the company Zhaotong Highway Investment Development Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Jiang, T.H.; Ren, W.Y.; Dong, Y.S.; Hou, Y.; Yuan, J.M. Precise representation of macro-texture of pavement and effect on anti-skidding performance. *Munic. Eng. Technol.* **2022**, *40*, 1–7, 24.
- Liu, J.; Li, J.; Li, X.Y.; Li, M.L. Determination of elastic component amounts of perforated elastic road surface and analysis of the noise reduction effect. *J. Munic. Technol.* **2023**, *41*, 7–11.
- Yang, E.H.; Cheng, Q.; Li, J.; Di, H.; Huang, B.; Qiu, Y. Surface Texture Reconstruction and Mean Texture Depth Prediction Model of Asphalt Pavement. *China J. Highw. Transp.* **2023**, *36*, 14–23.
- Zhang, Y.F. Measurement and Analysis of Skid-resistance and Noise Reduction Characteristics of the Cement Concrete Pavement. *Munic. Eng. Technol.* **2015**, *33*, 41–44.
- Fwa, T. Skid resistance determination for pavement management and wet-weather road safety. *Int. J. Transp. Sci. Technol.* **2017**, *6*, 217–227. [CrossRef]
- Wu, J.; Zhang, C.; Wang, Y.; Su, B.; Gond, B. Investigation on wet skid resistance of tread rubber. *Exp. Tech.* **2019**, *43*, 81–89. [CrossRef]
- Zhang, Y.C.; Zhou, X.X. Application of Recovery Technology for Anti-sliding Capability of Cement Concrete Pavement of Tunnel on Expressway. *Highway* **2021**, *66*, 297–299.
- Li, Y.X.; Dong, Y.S.; Yuan, J.M.; Jiang, T.H.; Ren, W.Y.; Yang, Y.Q. Research on Attenuation Law of Anti-skidding Performance of Asphalt Pavement. *Munic. Eng. Technol.* **2022**, *40*, 54–57+61.
- Zhou, W.; Tian, G.J.; Niu, Y.D.; Liu, Z.; Meng, F.; Liu, W. Design of Surface Friction Tester Measuring System and Experiment Analysis. *J. Electron. Meas. Instrum.* **2021**, *35*, 205–211.
- Wang, Y.X.; Miao, N.; Jing, G.Q.; Chen, N.; Liu, L. Modeling and error analysis of digital pendulum friction tester. *Chin. J. Sci. Instrum.* **2018**, *39*, 42–48.
- Liu, Y.; Tian, B.; Niu, K.M. Skid-resistance and Denoising Properties of Cement Concrete Pavement with Different Surface Texture. *J. Highw. Transp. Res. Dev.* **2012**, *29*, 28–33.
- Ran, M.P.; Yang, Y.M.; Huang, L.Y.; Zhou, X.-L. Analysis on Influence of Pavement Texture on Tire-road Rolling Resistance. *J. Highw. Transp. Res. Dev.* **2021**, *38*, 23–29.
- Ding, S.; Wang, K.C.P.; Yang, E.; Zhan, Y. Influence of effective texture depth on pavement friction based on 3D texture area. *Constr. Build. Mater.* **2021**, *287*, 123002. [CrossRef]
- Fan, Y.H.; Zhou, X.Y.; Yang, H.L.; Wang, L.-B. Analysis of Effects of Worn-out Surface Texture and Wet Surface on Asphalt Pavement Skid Resistance Through FEM-VDM Modeling. *China J. Highw. Transp.* **2023**, *36*, 318–332.
- Lu, X.L. Method of characterizing pavement texture wear attenuation based on entropy. *China Sci.* **2023**, *18*, 897–904.
- Dai, G.; Luo, Z.; Chen, M.; Zhan, Y.; Ai, C. Reconstruction and Intelligent Evaluation of Three-Dimensional Texture of Stone Matrix Asphalt-13 Pavement for Skid Resistance. *Lubricants* **2023**, *11*, 535. [CrossRef]
- Xu, G.; Lin, X.; Wang, S.; Zhan, Y.; Liu, J.; Huang, H. Incep-FrictionNet-Based Pavement Texture Friction Level Classification Prediction Method. *Lubricants* **2023**, *12*, 8. [CrossRef]

18. Fang, T.; Guo, R.; Si, Y. Ensemble Mode and Singular Value Based Road Texture-Skid Study. *Coatings* **2021**, *11*, 947. [CrossRef]
19. Yu, M.; Zhang, Q.; Shi, L.; Tang, O.D. Prediction Model of Skid Resistance of Asphalt Pavement Based on Dynamic Friction Behavior of Tire/Road. *J. Chongqing Jiaotong Univ. (Nat. Sci.)* **2023**, *42*, 40–46.
20. Li, X.Y.; He, H.; Hu, J.B. Three-dimensional asphalt pavement texture study based on structured light measurement. *J. China Foreign Highw.* **2022**, *42*, 83–89.
21. Smith, R.D. *Longitudinal Grinding and Transverse Grooving to Improve Frictional and Profile Properties*; Office of Materials Highway Division, Iowa Department of Transportation: Ames, IA, USA, 1983.
22. Rasmussen, R.O.; Whirlledge, R.P.; Turner, D.J.; Light, R.J.; Ferragut, T.R.; Wiegand, P. Identifying quieter concrete pavements using on-board sound intensity. In Proceedings of the Transportation Research Board 87th Annual Meeting, Washington, DC, USA, 13–17 January 2008.
23. Ong, G.P.; Fwa, T.F. Effectiveness of transverse and longitudinal pavement grooving in wet-skidding control. *Transp. Res. Rec.* **2007**, *2005*, 172–182. [CrossRef]
24. Kuemmel, D.A.; Sontag, R.C.; Crovetto, J.A.; Becker, Y.; Jaeckel, J.R.; Satanovsky, A. *Noise and Texture on PCC Pavements-Results of a Multi-State Study*; The National Academies of Sciences, Engineering, and Medicine: Washington, DC, USA, 2000.
25. Ghafoori, N.; Diawara, H. Abrasion resistance of fine aggregate-replaced silica fume concrete. *Mater. J.* **1999**, *96*, 559–569.
26. Ling, J.Q.; Han, S.; Guo, Z.T. Skid Resistance and Noise Reduction Texture Technique for Cement Concrete Pavement. *Transp. Stand.* **2009**, *11*, 62–66.
27. Xu, J.; Yang, S.L.; Ye, Z.J.; Wang, L.B.; Wei, Y.; Yang, B.Y.; Meng, W.; Cui, T.; He, C.W. Analysis of Driving Noise Characteristics on Cement Concrete Pavement with Different Groove Textures. *J. Munic. Technol.* **2023**, *41*, 27–35.
28. Chen, L.; Cong, L.; Dong, Y.; Yang, G.; Tang, B.; Wang, X.; Gong, H. Investigation of influential factors of tire/pavement noise: A multilevel Bayesian analysis of full-scale track testing data. *Constr. Build. Mater.* **2021**, *270*, 121484. [CrossRef]
29. Gardziejczyk, W.; Gierasimiuk, P.; Motylewicz, M.; Wasilewska, M. Evaluation of noisiness of exposed aggregate cement concrete pavement. *Road Mater. Pavement Des.* **2021**, *22*, 2352–2368. [CrossRef]
30. Kamil, B.A.; AlJameel, H.A.E. Impact of traffic characteristics and pavement surface condition on noise level for rigid pavement[C]/ IOP conference series: Earth and environmental science. *IOP Publ.* **2022**, *961*, 012101.
31. Dong, S.; Han, S.; Luo, Y.; Han, X.; Xu, O. Evaluation of tire-pavement noise based on three-dimensional pavement texture characteristics. *Constr. Build. Mater.* **2021**, *306*, 124935. [CrossRef]
32. Del Pizzo, L.G.; Teti, L.; Moro, A.; Bianco, F.; Fredianelli, L.; Licitra, G. Influence of texture on tyre road noise spectra in rubberized pavements. *Appl. Acoust.* **2020**, *159*, 107080. [CrossRef]
33. de León, G.; Del Pizzo, L.G.; Teti, L.; Moro, A.; Bianco, F.; Fredianelli, L.; Licitra, G. Evaluation of tyre/road noise and texture interaction on rubberised and conventional pavements using CPX and profiling measurements. *Road Mater. Pavement Des.* **2020**, *21* (Supp. 1), S91–S102. [CrossRef]
34. Chen, W.; Zheng, M.; Wang, H. Evaluating the tire/pavement noise and surface texture of low-noise micro-surface using 3D digital image technology. *Front. Mater.* **2021**, *8*, 683947. [CrossRef]
35. Cai, Z.S.; He, Z.; Qiu, Z.X.; Hou, Z.; Liu, C.; Yan, L. Study on skid-resistance and noise reduction performance of different texture cement concrete pavement. *Concrete* **2018**, *4*, 5–7+12.
36. Ginevra, D.P.L.; Bianco, F.; Moro, A.; Schiaffino, G.; Licitra, G. Relationship between tyre cavity noise and road surface characteristics on low-noise pavements. *Transp. Res. Part D* **2021**, *98*, 102971.
37. Zheng, M.L.; Chen, W.; Wang, H.Y. Evaluation of texture and noise of low-noise micro-surface based on image processing technology. *J. Traffic Transp. Eng.* **2023**, *23*, 80–92.
38. Lu, Z.F.; Li, B.; Zhang, Z.W.; Yang, X.L.; Ma, W.-Z. Statistical Analysis of Correlation between Surface Texture and Noise for Grooved Concrete Pavement. *J. Highw. Transp. Res. Dev.* **2015**, *32*, 14–19+26.
39. Hall, J.W.; Smith, K.L.; Titus-Glover, L.; Wambold, J.C.; Yager, T.J.; Rado, Z. *Guide for Pavement Friction*; NCHRP Web Document; The National Academies Press: Washington, DC, USA, 2009.
40. Hall, J.W.; Smith, K.L.; Littleton, P. *Texturing of Concrete Pavements*; NCHRP Report; The National Academies Press: Washington, DC, USA, 2009.
41. Li, B.; Kang, H.W.; Zhang, Z.W. Comparison of skid resistance and noise between transverse and longitudinal grooving pavements in newly constructed concrete pavement. *Adv. Mater. Res.* **2012**, *446*, 2637–2640. [CrossRef]
42. Cackler, E.T.; Harrington, D.S.; Ferragut, T. *Evaluation of US and European Concrete Pavement Noise Reduction Methods*; National Concrete Pavement Technology Center: Ames, IA, USA, 2006.
43. National Academies of Sciences, Engineering, and Medicine; Transportation Research Board; National Cooperative Highway Research Program. *Texturing of Concrete Pavements*; National Academies Press: Washington, DC, USA, 2009.
44. Behzadian, M.; Otagsara, S.K.; Yazdani, M.; Ignatius, J. A state-of-the-art survey of TOPSIS applications. *Expert Syst. Appl.* **2012**, *39*, 13051–13069. [CrossRef]
45. Fang, J.; Tu, J.; Wu, K. Analysis of Skid Resistance and Noise Characteristics for Varieties of Concrete Pavement. *Adv. Mater. Sci. Eng.* **2020**, *2020*, 7427314. [CrossRef]
46. Zhang, Z.; Luan, B.; Liu, X.; Zhang, M. Effects of surface texture on tire-pavement noise and skid resistance in long freeway tunnels: From field investigation to technical practice. *Appl. Acoust.* **2020**, *160*, 107120. [CrossRef]

47. TP 76-2009; Standard Method of Test for Measurement of Tire. Pavement Noise Using the On-Board Sound Intensity (OBSI) Method. AASHTO—American Association of State Highway and Transportation Officials: Washington, DC, USA, 2009.
48. Hwang, C.L.; Yoon, K. *Multiple Attribute Decision Making: Methods and Applications a State-of-the-Art Survey*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 186.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Comparison between Conventional Ageing Process in Barrels and a New Rapid Aging Process Based on RSLDE: Analysis of Bioactive Compounds in Spirit Drinks

Daniele Naviglio ¹, Paolo Trucillo ^{2,*}, Angela Perrone ¹, Domenico Montesano ³ and Monica Gallo ^{4,*}

- ¹ Department of Chemical Sciences, University of Naples Federico II, Via Cintia, 4, 80126 Naples, Italy; naviglio@unina.it (D.N.); angela.perrone@unina.it (A.P.)
- ² Dipartimento di Ingegneria Chimica, dei Materiali e della Produzione Industriale, University of Naples Federico II, Piazzale V. Tecchio 80, 80125 Naples, Italy
- ³ Department of Research & Development, Erbagil s.r.l., Via L. Settembrini 13, 82034 Telesse Terme, Italy; d.montesano@erbagil.com
- ⁴ Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II, Via Pansini 5, 80131 Naples, Italy
- * Correspondence: paolo.trucillo@unina.it (P.T.); mongallo@unina.it (M.G.); Tel.: +39-3296566043 (P.T.); +39-3441342574 (M.G.)

Abstract: “Aging” is a practice that allows alcoholic beverages to mature and gives them particular flavors and colors. In this context, oak or durmast wooden barrels are used in this process, thus providing different types of aging. This conventional process produces a slow enrichment of organic compounds in the spirit inside the barrels. Organic substances present in the internal part of the barrels slowly undergo the phenomenon of extraction by the liquid phase (solid–liquid extraction). In this work, a new procedure based on rapid solid–liquid dynamic extraction (RSLDE) was used to evaluate the potential of obtaining the effects of aging in spirits in shorter times than conventional methods. For this purpose, a comparison between two solid–liquid extraction techniques, RSLDE and conventional maceration, was made. Four water/ethanol 60:40 (*v/v*) model solutions were prepared and put in contact with medium-toasted chips using the two extraction procedures (conventional and non-conventional) and determining dry residue and total polyphenol content. Reversed phase high-performance liquid chromatography (RP-HPLC) analyses allowed the identification and quantification of furfural, ellagic acid and phenolic aldehydes (vanillin, syringaldehyde, coniferaldehyde and sinapaldehyde). The aging procedure with medium-toasted chips was tested on a young commercial *grappa* using maceration and RSLDE.

Keywords: bioactive compounds; chips; alcoholic beverages; maceration; RSLDE; grappa; distillates; aged beverages

Citation: Naviglio, D.; Trucillo, P.; Perrone, A.; Montesano, D.; Gallo, M. Comparison between Conventional Ageing Process in Barrels and a New Rapid Aging Process Based on RSLDE: Analysis of Bioactive Compounds in Spirit Drinks. *Processes* **2024**, *12*, 829. <https://doi.org/10.3390/pr12040829>

Academic Editors: Sheng Du, Li Jin, Xiongbo Wan, Zixin Huang and Dariusz Dziki

Received: 11 March 2024

Revised: 11 April 2024

Accepted: 17 April 2024

Published: 19 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Spirits and liqueurs define a category of alcoholic beverages produced by the distillation of grains, fruits or vegetables that have first been fermented through alcoholic fermentation [1]. During the distillation process, the liquid is concentrated to increase its alcohol grade by volume (vol/vol%). Distilled spirits and liqueurs contain much more ethyl alcohol than other alcoholic beverages, such as wine, beer or mixed alcoholic beverages with a grade of around 40% (*v/v*). The aromatic characteristics of distillates are obtained mainly from agricultural raw materials, which are processed during various treatment stages, including fermentation, distillation, and barrel aging [2,3].

Generally, a spirit drink is characterized by an alcoholic strength equal to or greater than 15% (*v/v*), produced by distillation or maceration, followed by a flavoring step. Freshly distilled spirits have a harsh taste, which can be attributed to the lack of various chemical reactions. For this reason, liqueurs require time to mature after distillation [4,5].

Aging is indeed a process used in the production of alcoholic beverages, such as wine, whiskey, and some types of beer. During aging, the beverage interacts with the container (such as wooden barrels), which can impart specific flavors, aromas, and colors to the drink, thus contributing unique characteristics to the beverage. However, it is important to note that the aging process and its effects can vary depending on the type of beverage and the specific production methods used. Therefore, many high-quality alcoholic beverages require aging in barrels for long periods of time [6]. Indeed, the objective of the aging process is to impart the aromas of the barrel's wood, such as oak, maple, acacia, chestnut and cherry. During the aging period, numerous chemical reactions occur that significantly change the taste and aroma of the alcoholic beverage [7–9]. However, the procedure for aging alcoholic beverages is characterized by a slow process of solid–liquid extraction that is based on the phenomena of diffusion and osmosis. Compounds contained in the wood of barrels are more slowly extracted because the process occurs at room temperature. To obtain faster aging, smaller barrels can be used, since in this case the amount of liquid in contact with wood is greater than in larger barrels (larger contact surface). However, the procedure is slow, and the overall process is particularly expensive.

In order to overcome the above-described drawbacks, novel and inexpensive techniques have been developed to simplify the aging process while ensuring that wood-bound volatiles are released into drinks and have similar sensory properties to those aged in barrels. These alternatives typically include oak chips or larger oak pieces [10]. The use of chips for the aging of wines is permitted in several countries, such as the United States, Australia and Chile. In Italy, it is only permitted on table wines, but the statement “aged in barrique” nowadays is forbidden on the label. Moreover, to enhance the final taste of aged spirits, the effect of toasting oak chips on the sensory characteristics and chemical composition of the wine was recognized to be greater than the oak type used [11,12]. Toasting oak increases the quantities of compounds deriving from the thermal degradation of lignin (vanillin, eugenol, guaiacol and its derivatives) and from the pyrolysis of cellulose and hemicellulose (furfural and 5-methylfurfural) and decreases the concentration of the two isomers of whiskey lactone. Similarly, during the aging of a distillate in a barrel, a series of chemical and physical interactions occur, involving the surrounding atmosphere and the maturing distillate, transforming the composition and, therefore, the final taste [13]. Recently, a growing interest has emerged in methods of predicting, controlling, and simulating the effects of maturation, given that the aging of distillates is one of the most important and expensive factors influencing their quality and their price. On the other hand, a new winemaking practice to rationalize work and reduce production costs for wine aging involves the use of wood chips that are 100 to 300 times cheaper than traditional barrels. Due to these favorable production conditions, newly emerging wine-producing countries such as New Zealand, California, Argentina, South Africa and Hungary are conquering the mass market with wines produced with this technique that have a good quality/price ratio and greater flexibility with respect to the tastes of the modern consumer. Among cheaper and faster alternatives than the conventional barrel aging method, maceration of oak pieces in wines and spirits is a well-known technique [14–20]. The use of chips, slats or strips immersed in the spirits has the aim of increasing the ratio between wood surface and volume liquid; this procedure promotes a faster solid–liquid extraction of the compounds from the wood. In this way, a larger surface is exposed to the liquid than 40%, as in the case of barrels. The products thus obtained are cheaper than those aged in barrels, but they have similar chemical characteristics and sensory properties. Therefore, even water, which is a highly polar solvent, can extract organic compounds in a heterogeneous phase. In traditional maceration, the effects of diffusion and osmosis are predominant; therefore, substances not chemically linked to the internal structure of the solid matrix tend to dissolve in the liquid. For this reason, a necessary condition for the traditional maceration process is that the substances to be extracted are soluble in the extractant liquid. Furthermore, to increase the extraction efficiency, it is necessary to increase the temperature, on which diffusion and osmosis depend, according to Fick's Law. However, a possible

drawback consists in the degradation of thermolabile compounds extracted or in favoring undesirable reactions [21–25].

Starting from these premises, the aim of this work was to evaluate a new method for obtaining the aging of spirits as an alternative method to the traditional one. This procedure is based on rapid solid–liquid dynamic extraction (RSLDE). The RSLDE occurs through the generation of a negative pressure gradient from the inside to the outside of the solid matrix. Therefore, it can be conducted at room or even sub-room temperature, avoiding the alteration of temperature-sensitive compounds, but also with other advantages. In fact, extraction requires much shorter times than maceration (2 h vs. 20 days) and allows the recovery and reuse of solvents with less environmental impact, according to the principle of the circular economy [26].

On the other hand, as reported in the literature, the areas of application of RSLDE are numerous and include the pharmaceutical, cosmetic, herbal, food and beverage sectors [27,28]. RSLDE can even be used for the extraction of bioactive compounds from agricultural and food waste, guaranteeing their possible reuse and reintroduction in the market as industrial by-products, according to the current eco-sustainable model of the circular economy [29].

To evaluate the efficiency of the alternative aging procedures, model water/ethanol solutions (60:40, *v/v*), with the addition of medium-toasted chips, were subjected to maceration and RLSDE techniques. Subsequently, the extracts obtained were subjected to various determinations, such as the dry residue, total polyphenols and the content of phenolic acids, furfuran derivatives and phenolic aldehydes deriving from wood, and identified as influencing the aroma and taste of aged spirits and drinks. Finally, a comparison between the two methods was carried out on a young commercial Italian spirit (named as *grappa*) subjected to aging using toasted chips.

2. Materials and Methods

2.1. Chemicals, Reagents, and Instrumentation

All solvents and reagents were of analytical grade and were purchased from Merck (Darmstadt, Germany), Carlo Erba (Milan, Italy), and Sigma-Aldrich Co. (Buchs, Switzerland); Folin Ciocalteu, anhydrous sodium carbonate, reagent and 2,4,6-tri(2-pyridyl-s-triazine) (TPTZ) were purchased from Sigma (Milan, Italy); Furfural, ellagic acid, vanillin, syringaldehyde, coniferaldehyde and sinapaldehyde standards used to identify the antioxidant compounds of the extracts were obtained from Sigma (Milan, Italy). All the reactives and reagents were used without previous alterations or pre-treatments.

2.2. Materials and Preparation of Samples

The chips used belonged to the Nobile[®] line, Laffort Italia S.R.L. (Tortona Alessandria, Italy), a company specializing in oenological products. Both Nobile[®] Fresh and Nobile[®] Sweet chips used in this work were qualified as suitable for the development of products for direct human consumption according to Regulation (CE) 606/2009. Nobile[®] Fresh and Nobile[®] Sweet are oak fragments of 7 to 20 mm in diameter that derive from the hardwood of French oak *Quercus Petraea* dried in air at room temperature. The Nobile[®] Fresh were unroasted, while the Nobile[®] Sweet underwent a medium roasting in the oven (Figure 1).

To determine the effectiveness of the extraction process and the stability of the extracted compounds, solid–liquid chip extraction was carried out in a water/ethanol solution of 60:40 (*v/v*) with Millipore water and 96% (*v/v*) Ethyl alcohol from Carlo Erba.



Figure 1. Nobile® Sweet chips (left) and Nobile® Fresh chips (right).

2.3. Maceration vs. RSLDE Techniques

Maceration was carried out for 15 days, as suggested in the technical data sheet of the commercial chips (Laffort Italia S.R.L., Tortona Alessandria, Italy). Vice versa, the duration of the extraction tests carried out with RSLDE was 3 h and 16 min (49 extraction cycles) and 17 h and 52 min (268 extraction cycles), respectively. Each extractive cycle consisted of a 2 min static phase and a dynamic phase of 2 min for a total cycle time of 4 min. The duration determined with these RSLDE experiments was optimized in order to obtain a comparable quality of maceration, which lasted for 15 days. The maximum cycle number defined for RSLDE was 1000, corresponding to approximately 3 days of extraction.

The extraction tests were carried out on a young commercial *grappa* using 10 g of Nobile® Sweet chips in 600 mL of *grappa* for 3 days. The same amount of chips and *grappa* were used for maceration for 15 days.

Preliminarily, the solid–liquid extraction was carried out in model solutions of water/ethanol (60:40, *v/v*), using unroasted and medium-toasted chips. In order to follow the extraction, solutions were stored in bottles (under reducing conditions) and sampled over time to reveal and compare only the extraction process from the wood and the stability of the extracted compounds.

2.4. Analysis of the Extracted Samples

2.4.1. Determination of Dry Residue

For the determination of the dry residue, 10 mL of the extracted sample was dried in the oven. The oven temperature was first set at 75 °C for the removal of the alcohol, and when the volume of the liquid became negligible, the temperature was raised to 105 °C to aid the evaporation of traces of water. The dry extract was recovered from the oven and, having cooled to room temperature, was weighed with an analytical balance with a sensibility of 0.1 mg. This operation was repeated until reaching a constant weight. Each determination was repeated 3 times, and the mean result was reported.

2.4.2. Determination of Total Polyphenols Using the Folin–Ciocalteu Method

Total phenol concentrations in plant extracts were determined spectrophotometrically by the Folin–Ciocalteu assay using gallic acid as a standard (mg GAE/g) [30]. A 125 µL intake of the methanolic extract was mixed with 500 µL of distilled water and 125 µL of the reagent of Folin–Ciocalteu were added. After vigorous stirring of the mixture followed by standing for 3 min, an aliquot of 1250 µL of 7% Na₂CO₃ was added. The solution was placed for 90 min at room temperature in a dark place. Lastly, absorbance was measured at 760 nm using a spectrophotometer UV-Vis (SmartSpec 3000, Bio-Rad Laboratories, Inc., Hercules, CA, USA). A calibration curve of gallic acid was prepared, and the results, determined from the regression equation of the calibration curve, were expressed as the mass of equivalent gallic acid over the mass of the sample. The same procedure was followed for the samples.

From the sample absorbance values at 760 nm, the total polyphenols expressed as mg gallic acid equivalent per gram were determined.

2.5. HPLC Analysis

The chromatographic measurements were carried out with a Waters 1525 binary HPLC and a Waters 2996 photo diode array (PDA) detector; this analysis was used to support the determination of the UV spectra (230–400 nm) of the samples. The analytical column was a reverse phase C18 (250 mm × 4.6 mm; 5 µm) (Phenomenex, Torrance, CA, USA); The mobile phase was eluent A water/formic acid (98:2, *v/v*) and eluent B methanol-water-formic acid (70:28:2, *v/v/v*). All the HPLC eluents used were purchased from Sigma Aldrich, Milan, Italy. The flow rate was set at 1 mL/min, with an injected volume of 20 µL. Before injections, the samples were filtered with nylon syringe filters of 0.45 µm with an external diameter of 25 mm from Millipore, Merck (Milan, Italy). The Limit Of Detection (LOD) determined for all compounds was in the range of 1 to 5 ppm. The elution program is described in Table 1.

Table 1. High performance liquid chromatography method for analysis of phenolic acids.

Time, min	Reservoir of Water-Formic Acid (98:2, <i>v/v</i>)	Reservoir of Methanol-Water-Formic Acid (70:28:2 <i>v/v/v</i>)
0	90%	10%
3	90%	10%
25	40%	60%
43	40%	60%
55	0%	100%
65	0%	100%

2.6. Analysis of a Young Commercial Grappa

A total of 10 g of Nobile[®] Sweet chips was put in contact with 600 mL of young commercial grappa. The sample was subjected to two extraction methods: maceration for 15 days and RSLDE at a different number of cycles.

2.7. Statistical Analyses

Results are the mean values ± SD (standard deviation) of 3 replicates. The significance level was set at *p* = 0.05.

3. Results and Discussion

In the present work, the potential of an alternative procedure to traditional aging in alcoholic beverage barrels was evaluated. Initially, solid–liquid extraction was conducted in water/ethanol 60:40 (*v/v*) model solutions using toasted and untoasted chips. Following the extraction, the solutions were stored in bottles (reductive environment) and sampled over time with the aim of revealing and comparing only the extraction process from the wood and the stability of the extracted compounds. Dry residue, total polyphenol content and RP-HPLC analysis were determined on the different solutions obtained, which allowed the identification and quantification of furfural, ellagic acid and phenolic aldehydes (vanillin, syringaldehyde, coniferaldehyde and sinapaldehyde). The RP-HPLC analysis shows that untoasted chips (Nobile[®] Fresh) have a lower content of bioactive compounds than medium-roasted chips (Nobile[®] Sweet). On the other hand, furfural and phenolic aldehydes are essentially formed following roasting from the degradation of hemicelluloses and lignin, respectively. While ellagic acid is also present in the solutions in which the untoasted chips have been extracted, it can derive both from the degradation of the ellagitannins during the roasting treatment and from the hydrolysis of the ellagitannins during the aging process. This is probably linked to the fact that ellagitannins, from which ellagic acid can derive, are present in greater quantities in untoasted chips, as they degrade

with roasting [31]. This greater quantity of ellagitannins is probably also responsible for the higher values of total polyphenols and dry residue for the solutions obtained by extraction of the unroasted chips. This may seem to contradict the lesser variation in color that occurred during the extraction of the Nobile[®] Fresh chips compared to the Nobile[®] Sweet ones, as the partially oxidized and polymerized ellagitannins are considered responsible for a large part of the color of the heartwood [32].

On the other hand, a study by Canas et al., 2013 reported that other compounds, not quantified by the total amount of polyphenols or the dry extract, can determine the color of brandy [33]. During the roasting process, various substances are produced following the interaction between sugars and amino acids, known as the Maillard reaction. Therefore, the color is influenced by melanoidins and probably other colored compounds that form during this reaction [34].

Based on this hypothesis, higher roasting intensities can favor the formation and accumulation of these compounds; such concentrations do not significantly influence the dry extract but contribute to the evolution of color [35]. Once the number of cycles necessary to equalize the extraction yield of the maceration had been estimated, an unaged (young) commercial *grappa* was subjected to extraction with Nobile[®] Sweet chips using the two methods to verify the effects of the interaction of the compounds extracted with the drink compounds. The data obtained show that the extraction yield is greater for maceration, as reported in Table 2.

Table 2. Dry residue of model solutions with Nobile[®] Fresh chips extracted using the two methods.

Time, Days	Maceration		RSLDE 49 Cycles (196 min)		RSLDE 268 Cycles (1072 min)		RSLDE 360 Cycles (1440 min)	
	Dry residue (g/L)	Non-volatile amount extracted (%)	Dry residue (g/L)	Non-volatile amount extracted (%)	Dry residue (g/L)	Non-volatile amount extracted (%)	Dry residue (g/L)	Non-volatile amount extracted (%)
0.005	---	---	0.62 ± 0.03	3.5	---	---	---	---
0.734	---	---	---	---	0.75 ± 0.02	4.5	---	---
1	---	---	---	---	---	---	1.51 ± 0.06	8.6
6	0.63 ± 0.07	---	---	---	0.85 ± 0.03	5.5	1.75 ± 0.05	10.5
15	1.37 ± 0.02	8.2	0.61 ± 0.02	3.7	0.83 ± 0.02	5	1.77 ± 0.03	11.1
80	1.45 ± 0.02	8.7	0.56 ± 0.04	3.4	0.98 ± 0.03	5.9	1.81 ± 0.04	12.5
210	1.71 ± 0.03	10.3	0.68 ± 0.5	4.2	---	---	---	---

3.1. Extraction of Model Solutions

Furthermore, the comparison with an aged commercial *grappa* of the same brand, which was aged in barrels for 12 months, showed a significantly higher content of total polyphenols and phenolic and furan compounds deriving from wood compared to the two alternative aging techniques.

3.1.1. Extraction of Nobile[®] Fresh Chips

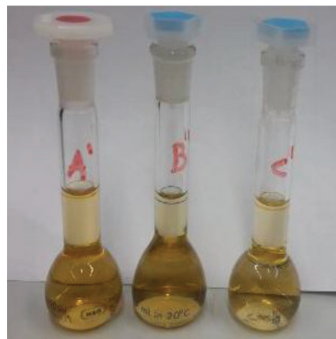
The dry residue and the total polyphenol content of the extractions obtained by maceration and by RSLDE containing Nobile[®] Fresh chips are shown in Tables 2 and 3.

Table 3. Total polyphenols (GAE/L) of the model solutions with Nobile® Fresh chips extracted with the two methods (GAE: mg of gallic acid equivalent).

Time, Days	Maceration	RSLDE 49 Cycles (196 min)	RSLDE 268 Cycles (1072 min)	RSLDE 360 Cycles (1440 min)
0.005	---	170 ± 5	---	---
0.734	---	---	368 ± 5	---
1	---	---	---	440 ± 6
6	257 ± 9	---	380 ± 6	480 ± 2
15	510 ± 12	176 ± 7	402 ± 3	510 ± 9
80	426 ± 8	168 ± 6	446 ± 2	553 ± 4
210	593 ± 8	187 ± 2	---	---

The maceration process has been performed for 210 days, during which the contact between chips and solvent was constant under discontinuous stirring. Different was the process of RLSDE. In RLSDE 50 cycles extraction, the total extraction time is 200 min, corresponding to 0.005 days; instead, in RLSDE 268 cycles extraction, the extraction time is 1072 min, corresponding to 0.744 days. After RLSDE extraction, the liquid extract was withdrawn, separated from chips and stocked at room temperature in the dark. Then, measurements were performed on these samples at 6, 15, 80 and 210 days, respectively.

The solutions obtained through maceration yielded higher values for both dry residue and total polyphenols, indicating an increased extraction yield after 6 and 15 days of observation. Nevertheless, in the solutions resulting from RSLDE, the values of dry residue and total polyphenols have the ability to increase by increasing the number of cycles. This suggests that it might be sufficient to increase the number of extraction cycles to achieve extraction yields comparable to those of a 15-day maceration. Furthermore, the greater yield of maceration was confirmed by the evolution of the color of the solutions and the decrease in pH. In fact, as can be seen in Figure 2, the solution obtained from the maceration of the chips developed a more intense color than that resulting from 268 extraction cycles, which is in turn more intense than that obtained from 49 extraction cycles.

**Figure 2.** Model solutions with Nobile® Fresh chips obtained from: A': Maceration; B': RSLDE 268 cycles; C': RSLDE 49 cycles.

Data reported in Tables 2 and 3 show that, over time, the values of the dry residue and polyphenols tend to increase, suggesting an evolution of the extracted compounds.

Additionally, the decrease in pH, characteristic of the aging process [36] and indicative of the extraction of phenolic compounds and/or other acidic compounds, was larger in the case of maceration (Table 4).

Table 4. Measurement of the pH of the solutions extracted with the two methods using Nobile[®] Fresh chips.

Solutions	pH
Hydroalcoholic solution	6.39
Maceration after 210 days	3.66
RSLDE 49 cycles after 210 days	4.44
RSLDE 268 cycles after 80 days	4.22

3.1.2. Extraction of Nobile[®] Sweet Chips

Similarly, in Tables 5 and 6, the dry residue and total polyphenol content of the model solutions containing Nobile[®] Sweet chips by maceration and RSLDE are reported.

Table 5. Dry residue of model solutions with Nobile[®] Sweet chips extracted using the two methods.

Time, Days	Maceration		RSLDE 49 Cycles (196 min)		RSLDE 268 Cycles (1072 min)		RSLDE 360 Cycles (1440 min)	
	Dry residue (g/L)	Non-volatile amount extracted (%)	Dry residue (g/L)	Non-volatile amount extracted (%)	Dry residue (g/L)	Non-volatile amount extracted (%)	Dry residue (g/L)	Non-volatile amount extracted (%)
0.005	---	---	0.30 ± 0.03	1.6	---	---	---	---
0.734	---	---	---	---	0.34 ± 0.02	2	---	---
1	---	---	---	---	---	---	1.2 ± 0.05	6.5
6	---	---	---	---	0.34 ± 0.03	2	1.3 ± 0.07	7.8
15	1.10 ± 0.02	6.6	0.29 ± 0.03	1.7	0.58 ± 0.03	1.7	1.45 ± 0.10	8.5
80	1.31 ± 0.02	7.9	0.40 ± 0.02	2.4	0.75 ± 0.05	4.5	1.51 ± 0.09	8.9
210	1.38 ± 0.02	8.1	1.11 ± 0.07	6.3	---	---	---	---

Table 6. Total polyphenols (GAE/L) of the model solutions with Nobile[®] Sweet chips extracted with the two methods (GAE: mg of gallic acid equivalent).

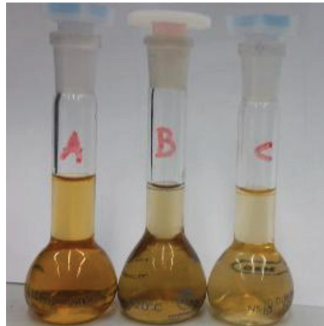
Time, Days	Maceration	RSLDE 49 Cycles (196 min)	RSLDE 268 Cycles (1072 min)	RSLDE 360 Cycles (1440 min)
0.005	---	110 ± 5	---	---
0.734	---	---	198 ± 3	---
1	---	---	228 ± 4	403 ± 8
6	---	---	220 ± 5	410 ± 5
15	330 ± 10	116 ± 6	212 ± 4	440 ± 4
80	304 ± 6	132 ± 8	±	445 ± 7
210	379 ± 8	149 ± 3	---	---

Also, in this case, the solutions resulting from the maceration of RSLDE gave higher values of both dry residue and total polyphenols, suggesting a greater extraction yield with values that approached those of the maceration with the increase in the number of cycles. This was confirmed by the evolution of the color of the solutions and by the decrease in pH (Table 7). More intense colors and a lower pH were obtained for the solution obtained by macerating the chips, followed by that resulting from the extraction of RSLDE 268 cycles (Figure 3).

Table 7. Measurement of the pH of the solutions extracted with the two methods using Nobile[®] Sweet chips.

Solutions	pH
Hydroalcoholic solution	6.39
Maceration after 210 days	3.63
RSLDE 49 cycles after 210 days	4.58
RSLDE 268 cycles after 80 days	4.30

Even in the case of extractions with Nobile[®] Sweet chips, the same trend was respected for the values of the dry residue and polyphenols reported previously for extractions with Nobile[®] Fresh chips.

**Figure 3.** Solutions obtained from: A: Maceration; B: RSLDE 268 cycles; C: RSLDE 49 cycles.

3.2. HPLC Analysis

The available standards were analyzed by chromatographic analysis first individually and then in mixture at wavelengths between 200 and 400 nm (see Table 8). Therefore, each standard was injected into the HPLC-DAD (diode array detector) to define the retention time and absorption peak of each compound. Table 8 shows the identification of the standard compounds, their retention times and the respective absorption maxima of each compound.

Table 8. Identification of chromatographic peaks, retention times and characteristic wavelengths of each compound of standard mixture.

Compounds	Retention Time	Wavelengths	Wavelengths
Furfural	13.40 ± 1.11	233	276
Vanillin	26.14 ± 0.49	238	280–309
Syringaldehyde	28.55 ± 0.45	238	309
Coniferaldehyde	33.69 ± 0.51	243	306–342
Sinapaldehyde	34.89 ± 0.33	246	347
Ellagic acid	42.40 ± 2.48	252	365

Figure 4 shows the HPLC analysis of the same standard mixture with the identification of peaks corresponding to the standards at the two selected wavelengths.

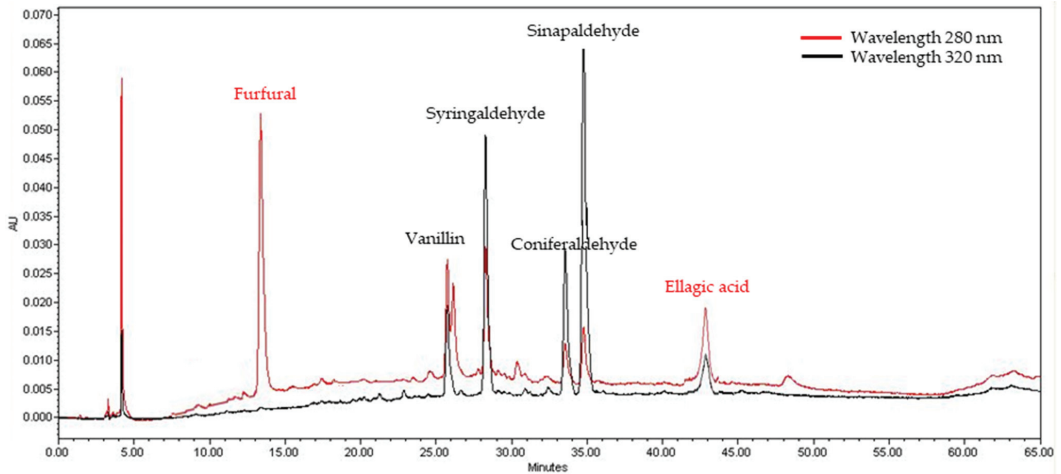


Figure 4. RP-HPLC analysis at 280 nm (red) and 320 nm (black) of the mixture of standard compounds.

Subsequently, under the same chromatographic conditions, HPLC analyses were carried out on the solutions extracted with both methods, which allowed the identification of furfural, ellagic acid and phenolic aldehydes (vanillin, syringaldehyde, coniferaldehyde and sinapaldehyde) based on the standards (Table 8). However, the chromatographic analysis of the solutions obtained by maceration and RSLDE revealed the evident presence of only ellagic acid, while the other compounds were present only in traces (see Figure 4). Therefore, it could be hypothesized that the phenolic and furan compounds present in the samples were not extracted under conditions more suitable for their production. In all cases, the extractions of Nobile[®] Sweet chips were richer in compounds than those of Nobile[®] Fresh chips. Furthermore, for all the solutions resulting from the extraction of Nobile[®] Sweet chips, a higher percentage of cinnamic aldehydes compared to benzoic ones was obtained. This may be due either to the degree of roasting, given that in the degradation of lignin, cinnamic aldehydes are formed, which then evolve to phenolic aldehydes, to a greater ease of extraction of cinnamic aldehydes [37], or to a different response factor of the compounds detected by HPLC.

The comparison of average values of total areas also enabled us to estimate the percentage decrease in extraction yield compared to maceration. This decrease was found to be 64% for 49 cycles of extraction and 48% for 268 cycles of extraction.

From the comparison of the individual peak areas, the extracted quantities of each individually identified analyte were also estimated in terms of percentages compared to maceration (Table 9).

Table 9. Comparison of the percentages of bioactive compounds extracted by RSLDE versus maceration.

Compounds	RSLDE 49 Cycles, %	RSLDE 268 Cycles, %
Furfural	36	48
Vanillin	46	58
Syringaldehyde	36	54
Coniferaldehyde	30	47
Sinapaldehyde	48	81
Ellagic acid	30	36

3.3. Aging Tests on Commercial Grappa

Once the required number of cycles for RSLDE (1030 cycles) to equalize the extraction yield with maceration had been determined, extractions were performed on a young

commercial *grappa* using the toasted Nobile[®] Sweet chips. These chips were found to be richer in compounds compared to the Nobile[®] Fresh chips. The goal was to examine the effects of the interaction between the extracted compounds and those present in the drink and to assess the proximity of the results to a commercially aged *grappa* in barrels.

Comparing the data in Table 10, it is evident that the dry residue is higher in the aged commercial *grappa*, whereas the total polyphenol content is greater in the young commercial *grappa* subjected to extraction with Nobile[®] Sweet chips. Additionally, consistent with previous findings, maceration exhibits higher dry residue and polyphenol values than RSLDE. This is further supported by the increased coloration of the *grappa* resulting from maceration (Figure 5).

Table 10. Dry residue and total polyphenols in young commercial *grappa*, young commercial *grappa* with Nobile[®] Sweet chips, aged commercial *grappa* extracted by maceration and RSLDE 1030 cycles. GAE: mg of gallic acid equivalent.

Sample	Dry Residue g/L	Polyphenols GAE/L
Young commercial <i>grappa</i>	1.69 ± 0.02	16 ± 2
RSLDE 1030 cycles with Nobile [®] Sweet chips	2.84 ± 0.05	483 ± 7
Aged commercial <i>grappa</i>	6.99 ± 0.07	32 ± 3



Figure 5. Extracted solutions obtained using the two extraction methods: M: Maceration; NE: RSLDE 1030 cycles.

In the chromatographic conditions used, the HPLC analysis of young commercial *grappa* and that of the same brand aged 12 months did not highlight the presence of peaks of interest, such as phenolic aldehydes, but only the presence of a peak corresponding to furfural (Figure 6a,b).

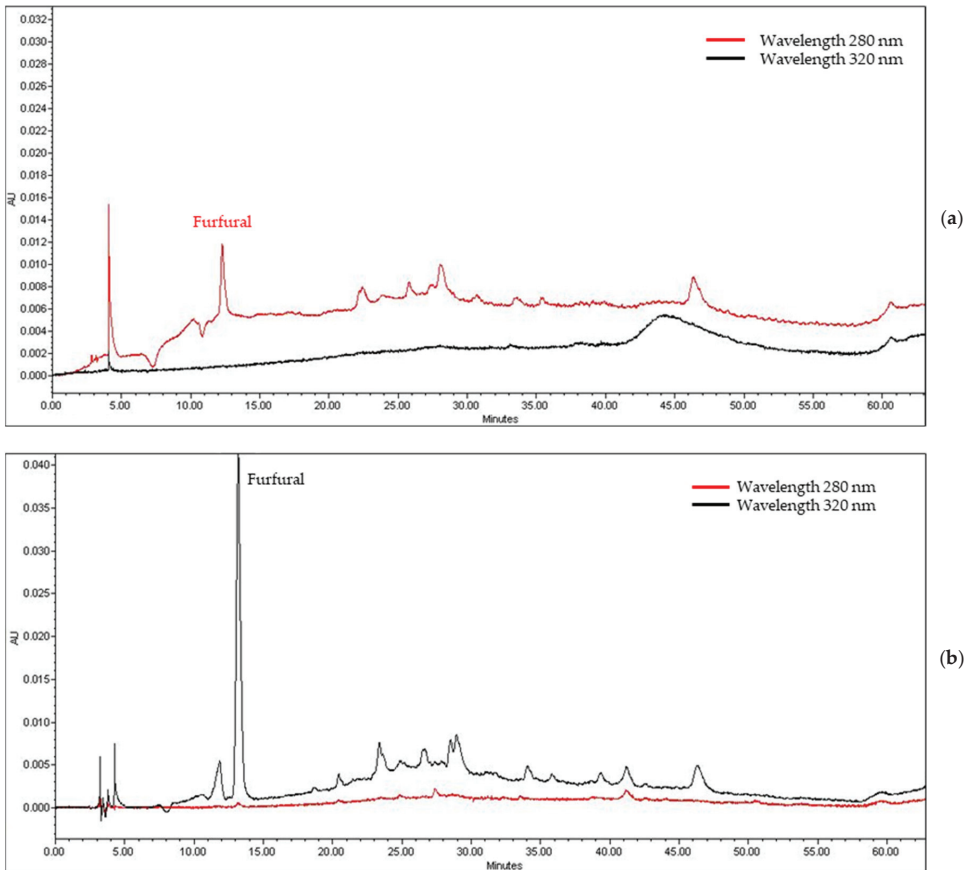


Figure 6. Chromatographic analysis of young commercial *grappa* (a) and aged commercial *grappa* (b) at 280 and 320 nm wavelengths.

Subsequently, the young commercial *grappa* was subjected to aging with Nobile[®] Sweet chips and extraction by maceration (15 days) and RSLDE 1030 cycles.

The HPLC analysis of young commercial *grappa* aged with Nobile[®] Sweet chips and extracted by maceration for 15 days showed a greater extraction yield compared to RSLDE. However, analysis of the sample extracted using RSLDE revealed the presence of a peak that was not present in the analysis of the sample extracted by maceration. The retention time is 18.60 min and could correspond to 4-hydroxybenzaldehyde, based on comparison with the literature [38], but the lack of standards did not allow the compound to be identified with certainty (Figure 7a,b).

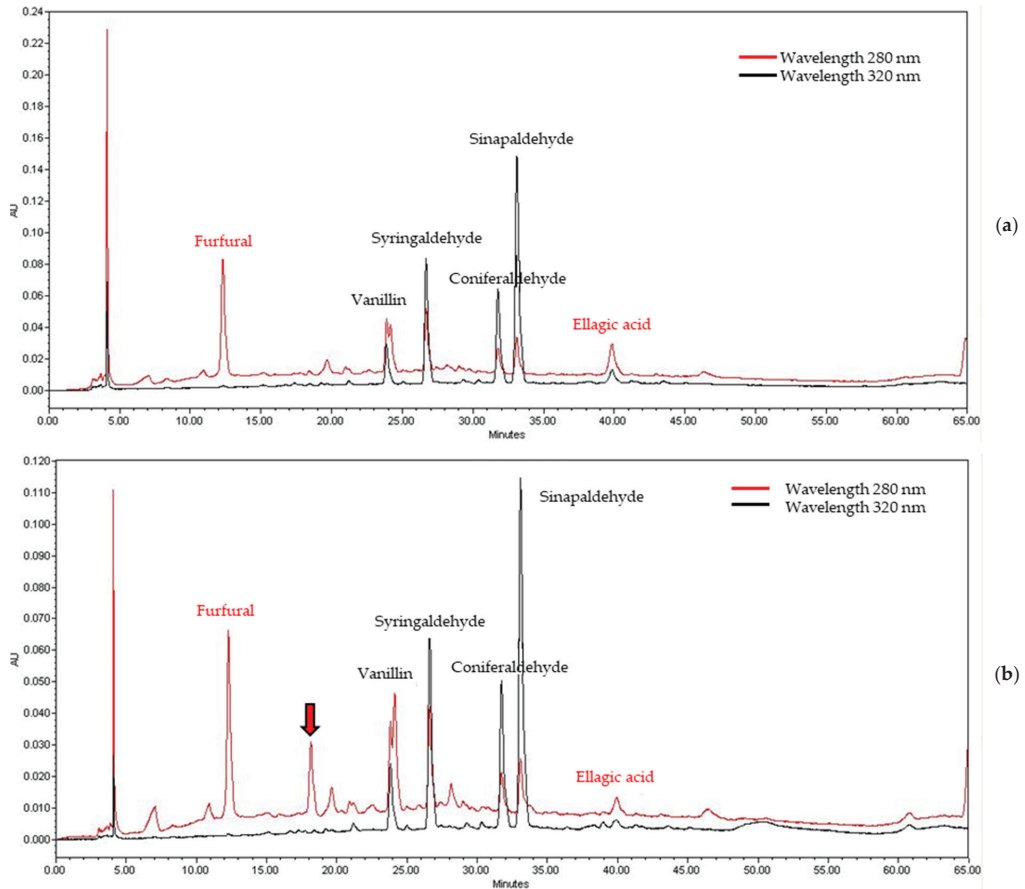


Figure 7. Chromatographic analysis of young commercial *grappa* aged with Nobile[®] Sweet chips extracted by maceration for 15 days (a) and RSLDE 1030 cycles (b) at 280 and 320 nm wavelengths. The arrow (b) indicates a peak at the retention time of 18.60 min, probably corresponding to 4-hydroxybenzaldehyde.

Also, in this case, a higher percentage of cinnamic aldehydes was observed compared to benzoic ones. The comparison of total areas and individual peaks confirms the greater extraction yield of maceration. Looking at results for *grappa* and for the water/ethanol 60:40% *v/v* solution (obtained by extraction with the two methods), results highlight that with the same extraction technique, the total areas appear greater for *grappa* compared to the water/ethanol 60:40% *v/v* model solution. This is probably due to the lower pH of *grappa*, which favors the extraction of polyphenolic compounds from the wood [39,40].

Finally, tests were conducted to assess the stability of bioactive compounds over time using the two methods. While these tests are preliminary and only cover the time interval between 0 and 6 days, the compounds demonstrated stability over this period.

4. Conclusions

According to the data obtained by the shown experiments, it is possible to conclude that in all the extracted solutions containing Nobile[®] Sweet chips, a higher percentage of cinnamic aldehydes was detected compared to benzoic ones. This could be attributed either to the degree of roasting, as the degradation of lignin results in the formation of cinnamic aldehydes that subsequently evolve into phenolic aldehydes. Alternatively, it could be due

to a higher ease of extraction of cinnamic aldehydes or a variation in the response factor of the compounds identified during the HPLC analysis.

However, other extraction tests will be necessary to be able to study and reproduce the aging of alcoholic beverages by verifying the effects of all possible process parameters. Future studies will be directed to follow the temporal evolution of *grappa* samples to verify the possibility of interactions between compounds deriving from wood and the drink in a reducing environment. After this, further research will also aim to analyze the volatile extract components. Furthermore, the aging process will be simulated, considering oxygen concentration. In any case, the effectiveness demonstrated by RSLDE in reducing aging times compared to conventional maceration stimulates research interest to encourage the application of this method in the alcoholic beverages sector.

This work was focused on the chemistry of the extraction and on process optimization. However, another issue to be addressed concerns the sensorial analysis, which plays a key role and requires the involvement of a team of tasters (panels). Therefore, based on this, a sensory evaluation requires significant planning and experimentation, which was not the subject of this manuscript but will certainly be carried out in future work.

Author Contributions: Conceptualization, D.N., P.T., D.M. and M.G.; methodology, A.P. and P.T.; software, A.P. and P.T.; validation, D.N., P.T., D.M. and M.G.; formal analysis, A.P. and D.M.; investigation, D.N. and M.G.; resources, D.N. and M.G.; data curation, A.P. and P.T.; writing—original draft preparation, P.T. and M.G.; writing—review and editing, P.T. and D.M.; visualization, D.N., P.T., D.M. and M.G.; supervision, D.N., P.T., D.M. and M.G.; project administration, D.N. and M.G.; funding acquisition, D.N. and M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest. The research was conducted in the absence of commercial or financial relationships that could be interpreted as a potential conflict of interest with the company Erbagil s.r.l. where the co-author Domenico Montesano works.

References

- Śliwińska, M.; Wiśniewska, P.; Dymerski, T.; Wardencki, W.; Namieśnik, J. The flavour of fruit spirits and fruit liqueurs: A review. *Flavour Fragr. J.* **2015**, *30*, 197–207. [CrossRef]
- Tonutti, I.; Liddle, P. Aromatic plants in alcoholic beverages. A review. *Flavour Fragr. J.* **2010**, *25*, 341–350. [CrossRef]
- Wang, L.; Chen, S.; Xu, Y. Distilled beverage aging: A review on aroma characteristics, maturation mechanisms, and artificial aging techniques. *Compr. Rev. Food Sci. Food Saf.* **2023**, *22*, 502–534. [CrossRef]
- Christoph, N.; Bauer-Christoph, C. Flavour of spirit drinks: Raw materials, fermentation, distillation, and ageing. In *Flavours and Fragrances: Chemistry, Bioprocessing and Sustainability*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 219–239.
- Marianski, S.; Marianski, A. *Home Production of Vodkas, Infusions & Liqueurs*; Bookmagic LLC: Seminole, FL, USA, 2012.
- Conner, J. Maturation. In *Whisky and Other Spirits*; Academic Press: Cambridge, MA, USA; Elsevier Ltd.: Amsterdam, The Netherlands, 2022; pp. 291–311.
- Carpena, M.; Pereira, A.G.; Prieto, M.A.; Simal-Gandara, J. Wine aging technology: Fundamental role of wood barrels. *Foods* **2020**, *9*, 1160. [CrossRef] [PubMed]
- Bortoletto, A.M.; Silvello, G.C.; Alcarde, A.R. Aromatic profiling of flavor active compounds in sugarcane spirits aged in tropical wooden barrels. *Braz. J. Food Technol.* **2021**, *24*, e2019071. [CrossRef]
- Krstić, J.D.; Kostić, S.M.M.; Veljović, S.P. Traditional and innovative aging technologies of distilled beverages: The influence on the quality and consumer preferences of aged spirit drinks. *J. Agric. Sci. Belgrade* **2021**, *66*, 209–230. [CrossRef]
- Petrozziello, M.; Nardi, T.; Asproudi, A.; Cravero, M.C.; Bonello, F. Chemistry and technology of wine aging with oak chips. In *Chemistry and Biochemistry of Winemaking, Wine Stabilization and Aging*; BoD—Books on Demand: Norderstedt, Germany, 2021; Volume 9.
- Rubio-Bretón, P.; Lorenzo, C.; Salinas, M.R.; Martínez, J.; Garde-Cerdán, T. Influence of Oak Barrel Aging on the Quality of Red Wines. In *Oak: Ecology, Types and Management*; Chuteira, C.A., Grao, A.B., Eds.; Nova Science Publishers: New York, NY, USA, 2013; Volume 2, pp. 59–86, ISBN 978-1-61942-493-7.
- Rubio-Bretón, P.; Garde-Cerdán, T.; Martínez, J. Use of oak fragments during the aging of red wines. Effect on the phenolic, aromatic, and sensory composition of wines as a function of the contact time with the wood. *Beverages* **2018**, *4*, 102. [CrossRef]

13. Li, Q.; Xu, H.; Yu, Y.; Zheng, Q. Why does distilled liquor has a soft and harmonious flavor after long-time ageing? A thermodynamic analysis. *J. Food Compos. Anal.* **2023**, *123*, 105609. [CrossRef]
14. Arapitsas, P.; Antonopoulos, A.; Stefanou, E.; Dourtoglou, V.G. Artificial aging of wines using oak chips. *Food Chem.* **2004**, *86*, 563–570. [CrossRef]
15. Caldeira, I.; Anjos, O.; Portal, V.; Belchior, A.P.; Canas, S. Sensory and chemical modifications of wine-brandy aged with chestnut and oak wood fragments in comparison to wooden barrels. *Anal. Chim. Acta* **2010**, *660*, 43–52. [CrossRef]
16. Madrera, R.R.; Valles, B.S.; García, Y.D.; del Valle Argüelles, P.; Lobo, A.P. Alternative woods for aging distillates—an insight into their phenolic profiles and antioxidant activities. *Food Sci. Biotechnol.* **2010**, *19*, 1129–1134. [CrossRef]
17. Jordão, A.M.; Correia, A.C.; DelCampo, R.; González SanJosé, M.L. Antioxidant capacity, scavenger activity, and ellagitannins content from commercial oak pieces used in winemaking. *Eur. Food Res. Technol.* **2012**, *235*, 817–825. [CrossRef]
18. Madrera, R.R.; Hevia, A.G.; Valles, B.S. Comparative study of two aging systems for cider brandy making. Changes in chemical composition. *LWT-Food Sci. Technol.* **2013**, *54*, 513–520. [CrossRef]
19. Taloumi, T.; Makris, D.P. Accelerated aging of the traditional greek distillate tsipouro using wooden chips. part i: Effect of static maceration vs. ultrasonication on the polyphenol extraction and antioxidant activity. *Beverages* **2017**, *3*, 5. [CrossRef]
20. Alañón, M.E.; Marchante, L.; Alarcón, M.; Díaz-Maroto, I.J.; Pérez-Coello, S.; Díaz-Maroto, M.C. Fingerprints of acacia aging treatments by barrels or chips based on volatile profile, sensorial properties, and multivariate analysis. *J. Sci. Food Agric.* **2018**, *98*, 5795–5806. [CrossRef] [PubMed]
21. Dumitriu, G.D.; Teodosiu, C.; Gabur, I.; Cotea, V.V.; Peinado, R.A.; López de Lerma, N. Evaluation of aroma compounds in the process of wine ageing with oak chips. *Foods* **2019**, *8*, 662. [CrossRef]
22. Laqui-Estaña, J.; López-Solís, R.; Peña-Neira, Á.; Medel-Marabolí, M.; Obrique-Slier, E. Wines in contact with oak wood: The impact of the variety (Carménère and Cabernet Sauvignon), format (barrels, chips and staves), and aging time on the phenolic composition. *J. Sci. Food Agric.* **2019**, *99*, 436–448. [CrossRef]
23. Cerdán, T.G.; Mozaz, S.R.; Azpilicueta, C.A. Volatile composition of aged wine in used barrels of French oak and of American oak. *Food Res. Int.* **2022**, *35*, 603–610. [CrossRef]
24. Krüger, R.T.; Alberti, A.; Nogueira, A. Current technologies to accelerate the aging process of alcoholic beverages: A review. *Beverages* **2022**, *8*, 65. [CrossRef]
25. Solar, S.; Castro, R.; Guerrero, E.D. New accelerating techniques applied to the ageing of oenological products. *Food Rev. Int.* **2023**, *39*, 1526–1546. [CrossRef]
26. Naviglio, D. Naviglio’s principle and presentation of an innovative solid–liquid extraction technology: Extractor Naviglio®. *Anal. Lett.* **2003**, *36*, 1647–1659. [CrossRef]
27. Formato, A.; Gallo, M.; Ianniello, D.; Montesano, D.; Naviglio, D. Supercritical fluid extraction of α - and β -acids from hops compared to cyclically pressurized solid–liquid extraction. *J. Supercrit. Fluids* **2013**, *84*, 113–120. [CrossRef]
28. Naviglio, D.; Formato, A.; Vitulano, M.; Cozzolino, I.; Ferrara, L.; Zanoelo, E.F.; Gallo, M. Comparison between the kinetics of conventional maceration and a cyclic pressurization extraction process for the production of lemon liqueur using a numerical model. *J. Food Process Eng.* **2017**, *40*, e12350. [CrossRef]
29. Naviglio, D.; Scarano, P.; Ciaravolo, M.; Gallo, M. Rapid Solid-Liquid Dynamic Extraction (RSLDE): A powerful and greener alternative to the latest solid-liquid extraction techniques. *Foods* **2019**, *8*, 245. [CrossRef] [PubMed]
30. Ouerghemmi, I.; Rebey, I.B.; Rahali, F.Z.; Bourgou, S.; Pistelli, L.; Ksouri, R.; Marzouk, B.; Tounsi, M.S. Antioxidant and antimicrobial phenolic compounds from extracts of cultivated and wild-grown Tunisian *Ruta chalepensis*. *J. Food Drug Anal.* **2017**, *25*, 350–359. [CrossRef] [PubMed]
31. Cabrita, M.J.; Dias, C.B.; Freitas, A.M.C. Phenolic acids, phenolic aldehydes and furanic derivatives in oak composition of white wine. *S. Afr. J. Enol. Vitic.* **2011**, *32*, 204–210.
32. Vivas, N.; Bourden-Nonier, M.F.; de Gaulejac, N.V.; Mouche, C.; Rossy, C. Origin and characterisation of the extractable colour of oak heartwood used for ageing spirits. *J. Wood Sci.* **2020**, *66*, 21. [CrossRef]
33. Canas, S.; Caldeira, I.; Belchior, A.P. Extraction/oxidation kinetics of low molecular weight compounds in wine brandy resulting from different ageing technologies. *Food Chem.* **2013**, *138*, 2460–2467. [CrossRef] [PubMed]
34. Alañón, M.E.; Rubio, H.; Díaz-Maroto, M.C.; Pérez-Coello, M.S. Monosaccharide anhydrides, new markers of toasted oak wood used for ageing wines and distillates. *Food Chem.* **2010**, *119*, 505–512. [CrossRef]
35. Zamora, F. Barrel aging; types of wood. In *Red Wine Technology*; Academic Press: Cambridge, MA, USA; Elsevier Inc.: Amsterdam, The Netherlands, 2019; pp. 125–147.
36. Comuzzo, P.; Battistutta, F. Acidification and pH control in red wines. In *Red Wine Technology*; Academic Press: Cambridge, MA, USA; Elsevier Inc.: Amsterdam, The Netherlands, 2019; pp. 17–34.
37. Nunes, I.; Correia, A.C.; Jordão, A.M.; MRicardo-da-Silva, J. Use of oak and cherry wood chips during alcoholic fermentation and the maturation process of rosé wines: Impact on phenolic composition and sensory profile. *Molecules* **2020**, *25*, 1236. [CrossRef]
38. Canas, S.; Belchior, A.P.; Spranger, M.I.; Bruno-de-Sousa, R. High-performance liquid chromatography method for analysis of phenolic acids, phenolic aldehydes, and furanic derivatives in brandies. Development and validation. *J. Sep. Sci.* **2003**, *26*, 496–502. [CrossRef]

39. De Rosso, M.; Cancian, D.; Panighel, A.; Dalla Vedova, A.; Flamini, R. Chemical compounds released from five different woods used to make barrels for aging wines and spirits: Volatile compounds and polyphenols. *Wood Sci. Technol.* **2009**, *43*, 375–385. [CrossRef]
40. Iacumin, L.; Manzano, M.; Cecchini, F.; Orlic, S.; Zironi, R.; Comi, G. Influence of specific fermentation conditions on natural microflora of pomace in “Grappa” production. *World J. Microbiol. Biotechnol.* **2012**, *28*, 1747–1759. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

SCFNet: Lightweight Steel Defect Detection Network Based on Spatial Channel Reorganization and Weighted Jump Fusion

Hongli Li ^{1,2,†}, Zhiqi Yi ^{1,2,†}, Liye Mei ^{3,4,†}, Jia Duan ⁵, Kaimin Sun ⁶, Mengcheng Li ¹, Wei Yang ⁵
and Ying Wang ^{5,*}

¹ School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China

² Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430205, China

³ The Institute of Technological Sciences, Wuhan University, Wuhan 430072, China

⁴ School of Computer Science, Hubei University of Technology, Wuhan 430068, China

⁵ School of Information Science and Engineering, Wuchang Shouyi University, Wuhan 430064, China

⁶ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China

* Correspondence: wangying@wsyu.edu.cn

† These authors contributed equally to this work.

Abstract: The goal of steel defect detection is to enhance the recognition accuracy and accelerate the detection speed with fewer parameters. However, challenges arise in steel sample detection due to issues such as feature ambiguity, low contrast, and similarity among inter-class features. Moreover, limited computing capability makes it difficult for small and medium-sized enterprises to deploy and utilize networks effectively. Therefore, we propose a novel lightweight steel detection network (SCFNet), which is based on spatial channel reconstruction and deep feature fusion. The network adopts a lightweight and efficient feature extraction module (LEM) for multi-scale feature extraction, enhancing the capability to extract blurry features. Simultaneously, we adopt spatial and channel reconstruction convolution (ScConv) to reconstruct the spatial and channel features of the feature maps, enhancing the spatial localization and semantic representation of defects. Additionally, we adopt the Weighted Bidirectional Feature Pyramid Network (BiFPN) for defect feature fusion, thereby enhancing the capability of the model in detecting low-contrast defects. Finally, we discuss the impact of different data augmentation methods on the model accuracy. Extensive experiments are conducted on the NEU-DET dataset, resulting in a final model achieving an mAP of 81.2%. Remarkably, this model only required 2.01 M parameters and 5.9 GFLOPs of computation. Compared to state-of-the-art object detection algorithms, our approach achieves a higher detection accuracy while requiring fewer computational resources, effectively balancing the model size and detection accuracy.

Keywords: surface defect detection; feature reconstruction; lightweight network; feature fusion

Citation: Li, H.; Yi, Z.; Mei, L.; Duan, J.; Sun, K.; Li, M.; Yang, W.; Wang, Y. SCFNet: Lightweight Steel Defect Detection Network Based on Spatial Channel Reorganization and Weighted Jump Fusion. *Processes* **2024**, *12*, 931. <https://doi.org/10.3390/pr12050931>

Academic Editor: Chin-Hyung Lee

Received: 8 April 2024

Revised: 27 April 2024

Accepted: 29 April 2024

Published: 2 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Steel is one of the most commonly used metals in manufacturing and is used widely in a variety of applications including construction, bridges, automobiles, and machinery. Due to its excellent performance, steel plays a crucial role in large industries such as metallurgy, geological drilling, and marine exploration. However, quality issues in steel often precipitate safety incidents, significantly compromising engineering integrity and personal safety [1]. As steel production increases, the possibility of defective steel entering the market increases, resulting in increasingly strict quality standards. In industrial manufacturing, the production environment for steel is complex and susceptible to various factors such as temperature and impact [2]. This results in surface defects such as cracks, patches, scratches, and inclusions [3,4]. Steel surface defect detection algorithms are essential for ensuring product quality, steel safety, and controlling production costs.

Typically, different types of defects on steel surfaces exhibit significant differences in terms of shape, size, and distribution. Examples include the following: (A) indistinctive features of defects: defect textures and grayscale are similar to the background [see Figure 1A]; (B) similar defects of different categories: different defects have similar distributions in shape and texture [see Figure 1B]; (C) low-contrast defects: defects have low color contrast with the background [see Figure 1C]; and (D) varied defects within the same category: defects within the same category exhibit significant differences in shape and texture [see Figure 1D]. This presents a considerable challenge to the feature extraction capacity of detectors. Early defect detection relied heavily on manual identification. However, manual identification is often costly, slow, and highly dependent on the experience and working conditions of the identification personnel. With advancements in the computer industry driving the automation sector forward, there arises an urgent demand across various industries for lightweight defect detection algorithms that enable automation while ensuring high-speed and high-precision performance [5–9].

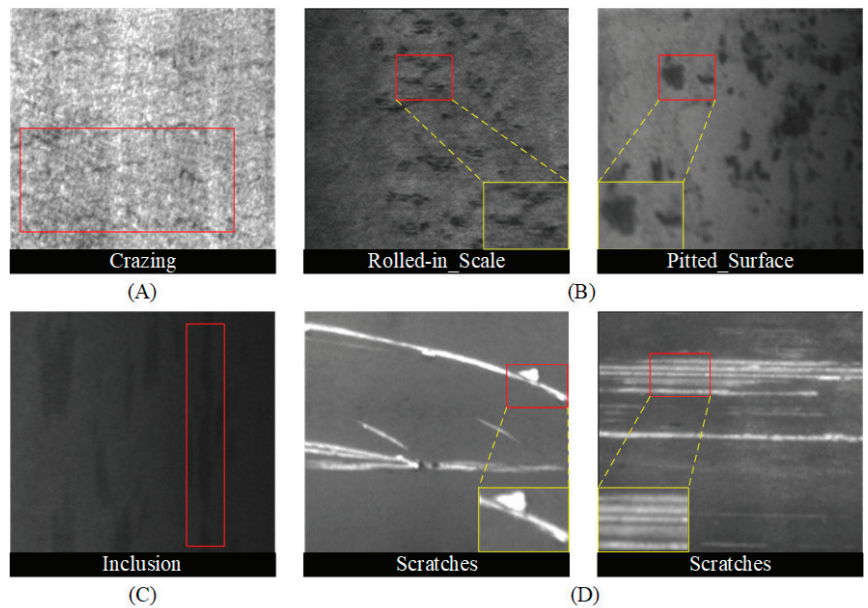


Figure 1. Steel surface defect detection faces a number of challenges, including (from the NEU-DET [10] dataset. The red box represents the defect location, and the yellow box shows the enlarged result.): (A) Defects with indistinct features. (B) Similar defects from different categories. (C) Defects with low contrast. (D) Defects within the same category exhibit significant variations.

The technique of detecting and classifying steel defects automatically is called computer vision-based steel defects detection. Typically, this approach involves extracting the shape, color, and texture information from images to describe and differentiate different types of defects. Techniques such as edge detection, corner detection, and texture analysis are utilized to extract features from images. In order to classify the features once they have been extracted, methods such as Support Vector Machines (SVMs) [11], clustering, Adaboost classifiers, or naive Bayes classifiers are used. However, since feature extractors often rely on manually designed features, this leads to lower model robustness. This makes it highly susceptible to factors such as lighting conditions, shooting angles, and the proportion of the target area [12].

In recent years, deep learning has undergone rapid development, significantly advancing object detection [13–15]. Images are transformed into feature maps through convolu-

tional neural networks. These feature maps typically contain higher-dimensional abstract features that are more targeted than manually designed features. Presently, mainstream object detection algorithms include two-stage algorithms such as R-CNN series [16–19], as well as single-stage algorithms such as SSD [20], You Only Look Once (YOLO) series [21–25], and Transformer-based algorithms such as DETR [26]. However, within the realm of steel defect detection, these deep learning-based object detection models face constraints due to the computational capabilities of terminal devices. Addressing how to optimize these object detection models with large parameters and computational overhead, while meeting task accuracy requirements, to enable deployment on devices with limited computing resources, remains a focal point in the current research on steel defect detection.

For detecting and recognizing defects on industrial steel surfaces, traditional machine learning methods have played a vital role in the early stages, usually involving image preprocessing, thresholding, and feature extraction. Traditional algorithms include LBP [10], HOG [27], and GLCM [28]. A number of studies [29,30] have developed more complex feature extractors by combining other methods in order to extract more accurate features of steel surface defects. Zhao et al. [31] utilized vector regularized kernel approximation and SVM for defect detection. Gong et al. [32] proposed developing a new Multi-Hypersphere SVM (MHSVM+) algorithm to provide additional information for detection tasks. Chu et al. [33] developed Multi-Information Siamese SVMs (MTSVMs), which are based on binary Siamese SVMs. Zhang et al. [34] proposed a method for identifying and diagnosing defects by merging Gaussian functions fitted to histograms of test images with membership matrices. However, traditional machine learning methods have significant limitations. The features used in these methods are manually designed, making them susceptible to changes in imaging environments and exhibiting poor robustness. Additionally, these methods often require extensive computational resources, resulting in slow processing speeds and difficulty in real-time detection.

Neural networks possess the capability to automatically extract features, fit models, and dynamically update parameters through learning processes, thereby allowing deep learning methods to excel across a multitude of tasks [35–39]. Upon entering samples into the network, it is capable of automatically classifying defect types and predicting defect locations. In practical steel surface defect detection, defects vary in size and shape, and the complex background makes them difficult to detect. Furthermore, smaller defects exhibit relatively minor changes in texture and color, making it difficult to distinguish between them. Using RDD-YOLO, Zhao et al. [40] integrated Res2Net blocks into the backbone network in order to enhance neck feature extraction and reuse shallow feature maps. Additionally, this method separates regression and classification with decoupled detection heads, improving detection accuracy. According to Wang et al. [41], YOLOv7 can be improved by integrating ConvNeXt modules into the backbone network and incorporating attention mechanisms in the pooling layers. The Diagonal Feature Fusion Network (DFN) strategy introduced by Yu et al. [42] matches multi-scale feature information without sacrificing speed, thereby significantly reducing the model size. Liu et al. [43] proposed DLF-YOLOF, using anchor-free detectors to reduce hyperparameters and expand contextual information in feature maps using deformable convolution networks and local spatial attention modules. Using a multi-scale lightweight network, Shao et al. [44] proposed a steel defect detection model that reduces the parameter count while improving the model accuracy. The aforementioned algorithms have made significant contributions in terms of both accuracy and speed. Nonetheless, these methods do not take into account the loss of information during the layer-by-layer feature extraction and spatial transformation of data, which is crucial for the detection of steel defects.

In order to further improve the detection accuracy while ensuring the lightweight of the model, we propose a lightweight and efficient steel defect detection algorithm called SCFNet. Specifically, we adopt an efficient and lightweight feature extraction module, LEM, to deeply excavate the defect information within the steel. And ScConv is applied in the deep network to reconstruct the spatial and channel information of feature maps, enhanc-

ing the representation of the defect features while reducing the generation of redundant information. Additionally, this article utilizes BiFPN for feature fusion, integrating deep semantic information and shallow spatial textures into one feature map, thereby preserving more complete spatial details and richer semantic features of the defect targets. We outline the contributions of this article as follows.

1. We propose a lightweight and efficient steel defect detection network, namely SCFNet. This network utilizes an LEM to extract feature information. The LEM, based on Depth-Wise convolution and channel-weighted fusion, can better extract ambiguous features.
2. Considering the low-contrast defects present in steel materials, we introduce the ScConv module into the LEM. By reconstructing the spatial information and channel features of the feature map, ScConv effectively reduces redundant features while enhancing key features in steel, thus making the defect area more clearly and accurately represented in the feature map.
3. We introduce the BiFPN module for feature fusion, leveraging its unique skip connection structure to minimize feature information loss during the convolution process. BiFPN ensures the preservation of crucial texture features, making it easier for the network to identify low-contrast defects.
4. We apply data augmentation techniques on the steel defect dataset and discuss the impact of different data augmentation methods on the detection accuracy. Ultimately, the proposed SCFNet demonstrates strong detection performance, outperforming state-of-the-art detectors in steel defect detection.

2. Materials and Methods

In practical steel defect detection, owing to complex backgrounds and the indistinct features of certain defects, detectors are susceptible to false positives and false negatives. We have noted that existing mainstream object detection networks lack sufficient capability in extracting ambiguous and low-contrast features. To optimize the defect detection performance in steel materials, we propose the SCFNet network model, as illustrated in Figure 2, which consists of three main components: the feature extraction module, neck fusion module, and detection head module. An image's deep features are extracted using the feature extraction module. Next, these features are forwarded to the neck fusion module. The neck fusion module is capable of constructing a feature pyramid network from top to bottom, transmitting the semantic information features of the fused feature maps, and then propagating the fused texture features from bottom to top. The neck network generates three feature maps with different spatial sizes, which are then fed into the detection heads separately. This allows the model to better detect objects on large, medium, and small scales, thereby alleviating the issue of inconsistent target scales. Specifically, when the image is input into the LEM consisting of three convolutional layers, 16 Mobile Inverted Bottleneck Convolution (MBConv) modules, 1 spatial channel recombination convolution module (comprising spatial recombination module SRU and channel recombination module CRU), and 3 feature maps with different spatial sizes and channel numbers, C3, C4, and C5, are obtained. Among them, C3 represents the shallow feature map with more texture information, C4 represents the middle feature map with certain texture information and semantic information, and C5 represents the deep feature map with more semantic information. A neck fusion network integrates the information from three different depths to coordinate and enrich the semantic and texture information of the three feature maps. Finally, the detection heads operate on the three feature maps separately to obtain the output information.

In the SCFNet network architecture, the feature extraction module is the LEM, which is extremely lightweight yet possesses strong feature extraction capabilities. As a result, it is able to extract deeper features from steel materials and adapt to defects that are not readily apparent. The neck module adopts a BiFPN for feature fusion. Compared to mainstream fusion networks like PANet [45], this fusion network features a unique skip

connection structure, minimizing the loss of spatial information and thereby enhancing the detector's performance.

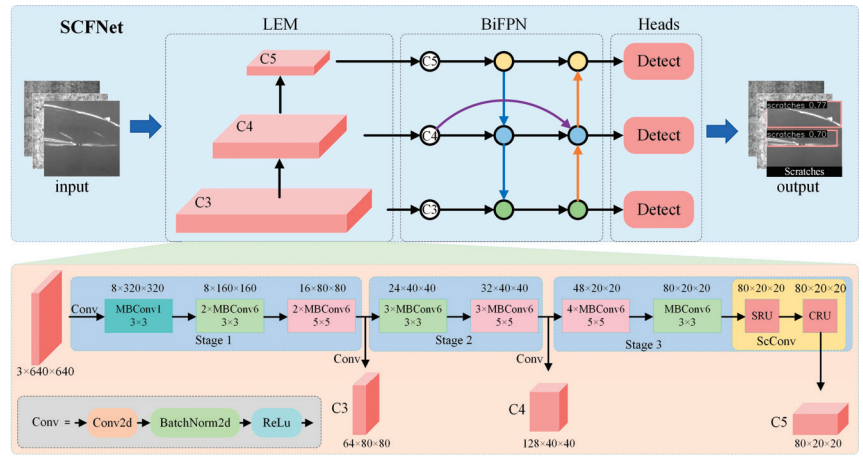


Figure 2. Architecture of the proposed SCFNet. The yellow, blue, and green circles represent feature maps of different scales.

2.1. Lightweight and Efficient Feature Extraction Module

Given the challenge of ambiguously extracting the target features of steel defects and the computational constraints imposed by terminal devices, deploying and utilizing networks present significant challenges. To address this issue, inspired by previous works [46,47], we propose a lightweight feature extraction network. In the past, convolutional neural network models typically optimized the model by adjusting one of three parameters: the input image resolution, network channel width, or depth. Tan's [46] study demonstrates that all three factors significantly impact the final accuracy and proposes a composite scaling method to uniformly adjust the network width, depth, and input image resolution, as shown in Equation (1) [46].

$$\begin{cases} \text{depth} : d = \alpha^\varphi \\ \text{width} : w = \beta^\varphi \\ \text{resolution} : r = \gamma^\varphi \\ \text{s.t. } \beta^2 \gamma^2 \alpha \approx 2 \\ \beta \geq 1, \gamma \geq 1, \alpha \geq 1 \end{cases} \quad (1)$$

where α , β , and γ are constants that can be determined by a small grid search. φ is an intuitively defined coefficient used to determine how many extra resources are available to scale the model.

Setting $\varphi = 1$ and based on the constraints in Equation (1), a grid search was performed. This led to $\alpha = 1.2$, $\beta = 1.1$, and $\gamma = 1.15$, resulting in the basic feature extraction module, LEM. The LEM has a relatively small parameter count and operates at a faster speed, making it highly suitable for lightweight detection tasks.

Figure 2 illustrates the LEM model structure composed of 3 convolutional layers, 16 MBConv modules, and 1 ScConv module. This model possesses strong feature extraction capabilities. Upon putting images into the network, the dimensions of the output feature maps increase gradually while the image size decreases. The deep feature maps harbor abundant semantic information, enabling the network to extract a broader range of classification features. In contrast, shallow feature maps contain a high level of texture information, which allows the network to retain certain texture characteristics and, as a result, place bounding boxes around the target objects in a more accurate manner. Similar

to other mainstream single-stage object detection models, SCFNet’s feature extraction module outputs three layers of feature maps. These three layers of feature maps undergo interaction in the neck network, complementing each other’s feature information, before being separately input into the detection heads for detection.

Figure 3 shows the MBConv module structure. This module mainly consists of two regular convolutions, one Depth-Wise convolution, one Squeeze-and-Excitation (SE) module, and a Dropout layer. The first convolution aims to increase the dimensionality, which helps in extracting features from deeper layers. In this context, MBConv1 signifies that the first convolution does not augment the dimensionality, whereas MBConv6 denotes a six-fold increase in the dimensionality. Depth-Wise convolution performs grouped convolutions, where each channel of the input is convolved separately without altering the number of channels in the feature map. A convolution following the SE module is a pointwise convolution, which uses only 1×1 convolutional kernels, operates on all channels, and can change the number of channels. By using Depth-Wise convolutions and pointwise convolutions, it is possible to construct deeper networks with smaller convolutional kernels and fewer parameters. This makes the model more lightweight without sacrificing accuracy. As a learnable attention mechanism, the SE module determines the importance of each channel by learning weights, thus guiding the model attention to more significant features.

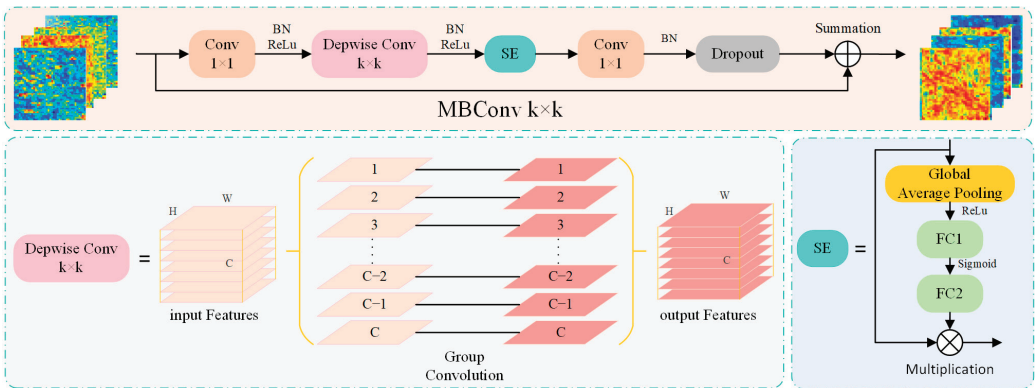


Figure 3. Structure diagram of the MBConv. The input and output feature maps are represented using heatmaps, where warm colors indicate strong features and cool colors indicate weak features.

2.2. Spatial and Channel Reconstruction Convolution

Due to the existence of similar features between different defect categories and defects with low contrast in steel defects, this poses a challenge to the feature expression capability of detectors. The ability of the feature extraction module to obtain representative features directly impacts the final results of the entire network. To enhance the representational capacity of the network, we propagate deep feature maps through the spatial and channel reconstruction convolution (ScConv) module. The ScConv module structure, as shown in Figure 4, consists of two units: the Spatial Reconstruction Unit (SRU) and the Channel Reconstruction Unit (CRU), which are sequentially placed in the module.

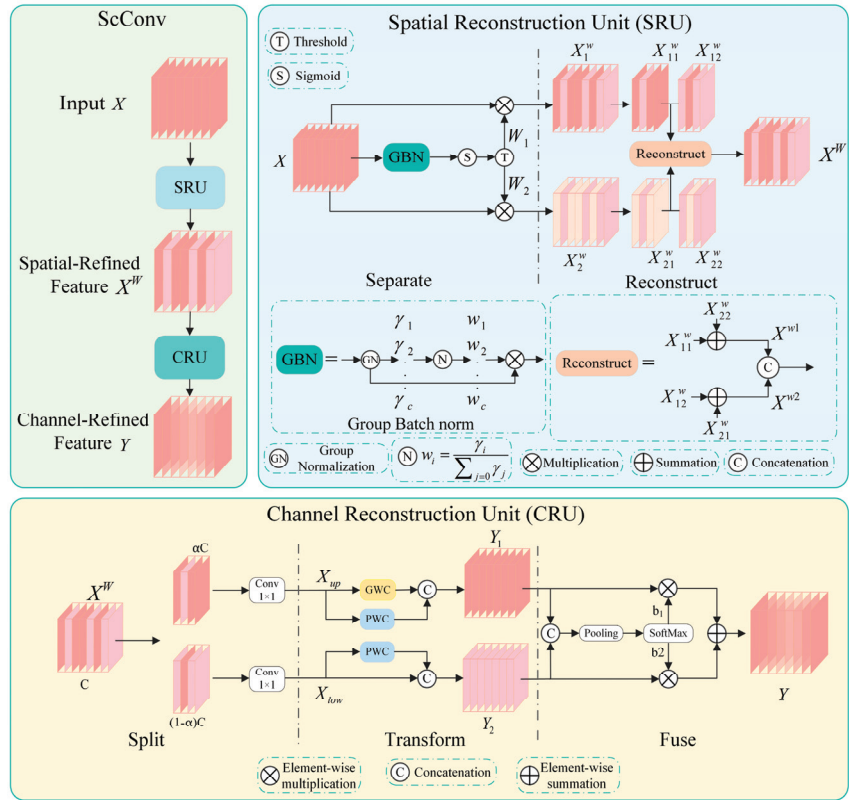


Figure 4. The ScConv module structure diagram. Dark red indicates feature maps with rich information, while light red indicates feature maps with less information.

The ScConv module can utilize spatial and channel redundancies between features to enhance feature map feature representation. The output feature map of the last MBConv6 in the feature extraction module serves as the input to the ScConv module. Firstly, it passes through the SRU to obtain spatial-refined features X^W , then it utilizes the CRU to obtain channel-refined features Y . The SRU separates parts of the feature map that contain rich spatial information from those with relatively less spatial content. Specifically, it evaluates the information content of different feature maps using the Group Batch Normalization (GBN) module. Given an input feature map $X \in R^{N \times C \times H \times W}$, where N is the batch axis, C is the channel axis, and H and W are the height and width axes of the feature map, the operation of the Group Normalization (GN) module is as shown in Equation (2) [47]:

$$X_{out} = GN(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \tag{2}$$

where μ and σ are the mean and standard deviation of X , ϵ is a small natural number, and X and β are trainable affine transformations. Subsequently, the normalized correlation weights of w_c are calculated, which represent the importance of spatial information at different positions in the feature map. Then, the weight coefficients are multiplied by the feature map, normalized using the Sigmoid function, and thresholded to separate them. Those weights normalized above the threshold are set to 1 to obtain the information-rich weight W_1 , while those below the threshold are set to 0 to obtain the weight W_2 with less information. Then, W_1 and W_2 are, respectively, multiplied with feature map X , resulting in feature map X_1^W rich in information and feature map X_2^W with less information. To further

compress the spatial redundancy, the two feature maps are cross-reconstructed by fully combining their information through addition before being connected, resulting in the spatially refined feature map X^W . This approach, superior to directly adding the feature maps, enables a tighter interaction of spatial information between the two feature maps.

The CRU plays a pivotal role in harnessing channel information redundancy to further refine and enhance the channel features of the feature maps. The CRU primarily comprises three modules: Split, Transform, and Fuse. The Split module first divides the given spatially refined feature map into two feature maps with channel numbers denoted as αC and $(1 - \alpha)C$, respectively. Then, both feature maps undergo 1×1 convolutions to adjust the channel numbers to half of the original input feature map, resulting in outputs X_{up} and X_{low} . The Transform module takes X_{up} as the input and processes it through a "strong feature extractor". The "strong feature extractor" employs Group-wise Convolution (GWC) and pointwise convolution (PWC) instead of regular convolutions. GWC has fewer parameters and computations compared to conventional convolutions but lacks inter-channel information flow, while PWC supplements the inter-channel information flow. The outputs of both operations are summed to obtain Y_1 . Meanwhile, X_{low} is passed into the "weak feature extractor", which only employs 1×1 PWC to extract some detailed features. Afterwards, it undergoes residual connections to yield Y_2 . The Fuse module combines the two feature maps by concatenating Y_1 and Y_2 along the channel dimension. To extract the global spatial information, the concatenated feature map undergoes global average pooling. This information is utilized to generate feature vectors β_1 and β_2 using SoftMax. These vectors are then multiplied and added to Y_1 and Y_2 , respectively, to obtain the channel-refined feature map Y . The feature maps processed through the ScConv module are enhanced in their representation of important features, significantly improving the detection of steel defects with less prominent characteristics. The overall computation formula for global spatial information $S_m \in R^{c \times 1 \times 1}$ is described in Equation (3) [47].

$$\left\{ \begin{array}{l} S_m = Pooling(Y_m) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W Y_c(i, j), m = 1, 2 \\ \beta_1 = \frac{e^{s_1}}{e^{s_1} + e^{s_2}}, \beta_2 = \frac{e^{s_2}}{e^{s_1} + e^{s_2}}, \beta_1 + \beta_2 = 1 \\ Y = \beta_1 Y_1 + \beta_2 Y_2 \end{array} \right. \quad (3)$$

where S_1 and S_2 represent global channel descriptors, while β_1 and β_2 denote feature vectors. After passing through the ScConv module, the feature representation is enhanced. At this point, the LEM sends the last layer feature map C_5 along with C_4 and C_3 to the neck network for feature fusion. In summary, the proposed LEM is lightweight yet possesses strong feature extraction capabilities. Additionally, the ScConv module utilizes spatial and channel redundancies to enhance feature representation, thereby improving the model learning capability and detection accuracy.

2.3. Feature Pyramid Fusion with a Weighted Bidirectional Approach

Considering the significant scale variations and indistinct features of defects in steel, to enable the model to address the issue of large-scale variations in objects within images, we separately input three feature maps into the detection heads to detect objects at large, medium, and small scales. Generally, shallow feature maps possess higher spatial resolution and carry abundant spatial and positional information but lack distinct semantic features. Conversely, deep feature maps contain rich semantic information but lack sufficient spatial details. Deep feature maps provide the model with abundant semantic information that is used to categorize objects, while shallow feature maps provide the model with abundant texture information that is used to locate objects. Both are crucial for object detection tasks. To further compensate for the resulting accuracy loss, inspired by [48], we employ a BiFPN based on weighted fusion to interactively fuse the three feature maps. Through

weighted fusion, local details, spatial positions, and semantic information are amalgamated, bolstering the representational capacity of semantic features.

As shown in Figure 5, the Bifpn module comprises a set of learnable weight parameters. After receiving feature maps with the same spatial channel size, the module performs weighted summation on each feature map, followed by activation processing using SiLu, and finally convolutional output. BiFPN utilizes a feature propagation structure similar to the Path Aggregation Network (PAN) [45], sequentially transmitting feature information from deep feature maps to shallow ones, and then propagating the fused shallow feature maps back to the deep feature maps. Specifically, BiFPN first processes the deep feature map C_5 through convolution and upsampling to match the shape of C_4 , then performs weighted fusion. Taking the intermediate feature maps C_4 and P_4 as an example, the fusion process is as described by Equation (4) [48].

$$\begin{cases} P_4^{td} = \text{Conv}\left(\frac{w_1 \cdot C_4 + w_2 \cdot \text{Resize}(C_5)}{w_1 + w_2 + \epsilon}\right) \\ P_4 = \text{Conv}\left(\frac{w'_1 \cdot C_4 + w'_2 \cdot P_4^{td} + w'_3 \cdot \text{Resize}(P_3)}{w'_1 + w'_2 + w'_3 + \epsilon}\right) \end{cases} \quad (4)$$

where P_4^{td} is the intermediate feature from the fourth level of the top-down path, P_4 is the output feature from the fourth level of the bottom-up path, w represents the learnable feature fusion coefficient, and ϵ is a very small constant (in this experiment, this coefficient is 0.0001) to prevent division by zero errors. (\cdot) denotes the SiLu activation function. This fusion method allows feature fusion with minimal feature loss and fewer parameters, enabling the network to fully integrate the feature map information while ensuring a lightweight design, which is beneficial for detecting subtle defects in steel materials. The fused three-layer feature maps are then summed through the Cross Stage Partial (CSP) module. The CSP module divides the input into two parts, where one part undergoes two convolution operations and is then concatenated with the other part. This structure amplifies the CNN learning capability and diminishes computational bottlenecks, making it suitable for industrial applications. After enhancing the features through the CSP, the three feature maps are used as inputs to the detection head module. In summary, SCFNet utilizes a weighted BiFPN for feature fusion, carefully controlling the parameter count increases to maintain a lightweight model structure. Furthermore, the experimental results validate the feasibility and efficacy of this approach.

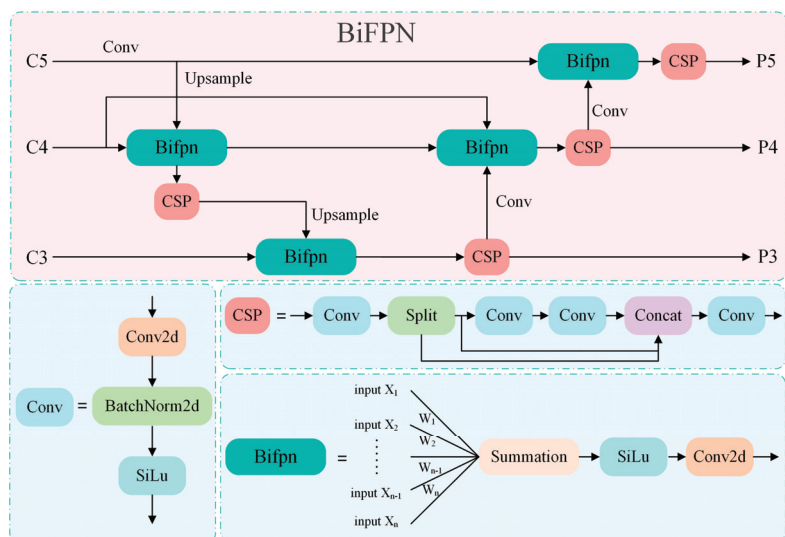


Figure 5. Structure diagram of the BiFPN module.

2.4. Loss Function

The network loss consists of two components [49]: object classification loss L_{cls} and bounding box loss L_{cls} .

$$L_{total} = c_1 L_{cls} + c_2 L_{bbox} \quad (5)$$

where c_1 and c_2 represent the weights of the loss functions. A Binary Cross-Entropy Loss (BCE) is used to calculate the classification loss, while CIoU and distribution focal losses (DFLs) are used to compute the bounding box loss. The formulas for calculation are as follows [49]:

$$L_{cls}(y, p) = y \log(1 - p) - y \log(p) - \log(1 - p) \quad (6)$$

where y represents the actual class of the target, taking values of 0 or 1, and p represents the predicted class of the target, ranging from 0 to 1.

$$L_{bbox} = \lambda_1 L_{CIoU} + \lambda_2 L_{DFL} \quad (7)$$

$$\begin{cases} L_{CIoU} = IoU - \left(\frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \right) \\ v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \\ \alpha = \frac{v}{(1 - IoU) + v} \end{cases} \quad (8)$$

$$L_{DFL}(y_i, y_{i+1}) = -(i + 1 - y) \log(y_i) - (y - i) \log(y_{i+1}) \quad (9)$$

where λ_1 and λ_2 represent the weighting coefficients of the loss. IoU stands for Intersection over Union, ρ denotes the Euclidean distance between the centers of the predicted bounding box and the ground truth bounding box, while c represents the distance between the predicted bounding box and the closest point to the ground truth bounding box's enclosing rectangle. αv stands for the aspect ratio, which is the ratio of width to height, between the predicted bounding box and the ground truth bounding box. (b, b^{gt}) represent the center coordinates of both the predicted and ground truth bounding boxes, while w, h, w^{gt}, h^{gt} denote their respective widths and heights. y denotes the actual label.

3. Experiments

3.1. Datasets

Our proposed defect detection method is evaluated using the NEU-DET [10] dataset to assess its accuracy, robustness, and generalizability. Developed by Northeastern University researchers, the NEU-DET dataset includes six common surface defects in steel. During the manufacturing process of steel plates, six different types of surface defects are commonly encountered. These defects include Scratches (Sc), Inclusion (In), Cracking (Cr), Pitted Surface (PS), Patches (Pa), and Rolled-in Scales (RS). There are 300 images of each defect type, each with a resolution of 200×200 pixels, adding up to 1800 images in total.

3.2. Implementation Details

In this article, we conducted experiments using a 16 GB Nvidia RTX 4060 Ti GPU. The deep learning framework utilized was PyTorch 2.0.1. The ratio of the training data, validation data, and testing data was set to 8:1:1. We employed the SGD optimizer with a momentum of 0.937 and a learning rate of 0.01. There was a BatchSize of 32, and the training was conducted for 400 epochs. The code has been open-sourced at <https://github.com/LazyShark2001/SCFNet> (accessed on 25 April 2024).

3.3. Evaluation Metrics

Selecting appropriate evaluation metrics is crucial for assessing the algorithm performance in defect detection. Evaluation metrics should be chosen in a way that objectively measure the algorithm's accuracy and robustness. In practical industrial production, both the accuracy of defect detection and the size of the model are crucial. When the detection

accuracy of defects is too low, machines may make incorrect judgments, failing to accurately identify defective workpieces. Additionally, large model sizes can pose deployment challenges on terminal devices. Precision (P), Recall (R), and Mean Average Precision (mAP) are commonly used as metrics to evaluate algorithm performance [4]. Furthermore, to evaluate the complexity and size of the model, we can consider the number of Floating-point Operations (FLOPs) and the number of parameters (Params). FLOPs quantify the computational workload required for inference, while Params represent the number of trainable parameters in the model. These metrics offer insights into the computational efficiency and model complexity, which are essential considerations for deployment on terminal devices and real-world applications.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (11)$$

$$mAP = \frac{\sum_{f=1}^c \int_0^1 P(R) dR}{c} \quad (12)$$

where TP represents the number of correctly classified as positive samples; FP represents the number of incorrectly classified as positive samples; and FN represents the number of incorrectly classified as negative samples. Precision and recall are, respectively, denoted as P and R .

3.4. Comparison with State-of-the-Art Models

We conducted comparative experiments with several mainstream detection algorithms to validate the superiority of our proposed model, including two-stage algorithms such as Faster R-CNN, as well as one-stage algorithms such as YOLOv5s, YOLOv7-tiny, YOLOv8s, CG-Net, and FCCv5s.

In Figure 6, we visually compare our SCFNet (right) with other models on the NEU-DET [10] dataset. Specifically, in the “Crazing” category, our model accurately detects defects. Due to the indistinct features of the targets, other models such as Faster R-CNN and YOLOv5s often lose significant texture information during feature extraction and transformation. This can result in unreliable feature learning and lead to false alarms. SSD and CenterNet models have weak feature extraction capabilities, resulting in missed detections. In the “Inclusion” category, our model accurately detects two defects with high confidence. Our algorithm achieves good visual results in “Patches,” “Pitted Surface,” “Rolled-in Scale,” and “Scratches” without missing detections or false alarms. Compared to other networks, our model successfully identifies defects with ambiguous features (Crazing) and detects low-contrast defects (Inclusion in the sixth row of Figure 6) better, demonstrating its outstanding capability in defect detection.

Table 1 presents the results. In our experimental results, it has been demonstrated that our proposed lightweight and highly efficient steel surface defect detection network, SCFNet, performs better on the NEU-DET dataset when analyzing the P, mAP_{50} , $mAP_{50:95}$, model parameter count, and model computation complexity for the NEU-DET dataset, with values of 0.876, 0.812, 0.469, 5.9, and 2, respectively. Among them, metrics P, mAP_{50} , and $mAP_{50:95}$ perform the best, while the model parameter count and model computational complexity rank second. Compared to the current mainstream detectors, our proposed model achieves a balance between lightweight design and high accuracy in steel defect detection, achieving optimal precision with minimal model parameters and computational complexity.

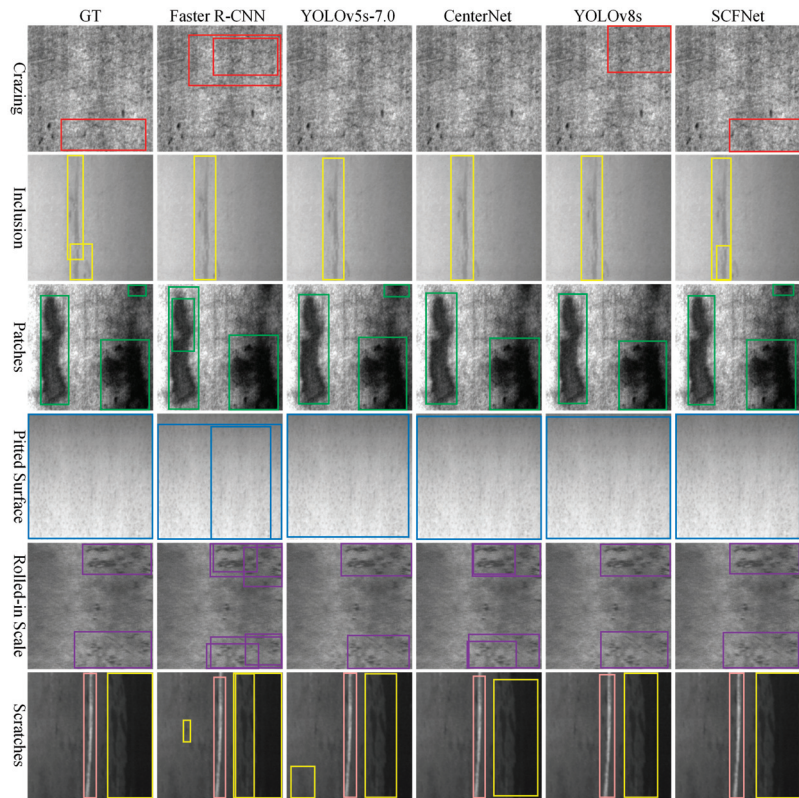


Figure 6. Results of different models compared to SCFNet on NEU-DET [10] dataset. In the picture, the red box represents “Crazing”, the yellow box represents “Inclusion”, the green box represents “Patches”, the blue box represents “Pitted Surface”, the purple box represents “Rolled-in Scales”, and the pink box represents “Scratches”.

Table 1. SCFNet algorithm performance comparison with other object detection algorithms on NEU-DET [10] dataset.

Methods	P	R	mAP ₅₀	mAP _{50:95}	GFLOPs	Params/M
Faster R-CNN [18]	0.615	0.865	0.76	0.377	135	41.75
CenterNet [50]	0.712	0.749	0.764	0.412	123	32.12
YOLOv5n-7.0 [51]	0.694	0.694	0.746	0.422	4.2	1.77
YOLOv5s-7.0 [51]	0.745	0.719	0.761	0.429	15.8	7.03
YOLOv7-tiny [25]	0.645	0.775	0.753	0.399	13.1	6.02
YOLOv8s [49]	0.768	0.726	0.795	0.467	28.4	11.13
YOLOX-tiny [52]	0.746	0.768	0.76	0.357	7.58	5.03
MRF-YOLO [53]	0.761	0.707	0.768	-	29.7	14.9
YOLOv5s-FCC [54]	-	-	0.795	-	-	13.35
WFRE-YOLOv8s [55]	0.759	0.736	0.794	0.425	32.6	13.78
CG-Net [56]	0.734	0.687	0.759	0.399	6.5	2.3
ACD-YOLO [57]	-	-	0.793	-	21.3	-
YOLOv5-ESS [58]	-	0.764	0.788	-	-	7.07
PMSA-DyTr [2]	-	-	0.812	-	-	-
MED-YOLO [4]	-	-	0.731	0.376	18	9.54
MAR-YOLO [15]	-	-	0.785	-	20.1	-
SCFNet	0.786	0.715	0.812	0.469	5.9	2

Red bold indicates the top-ranking performance, while **blue** bold indicates the second-ranking performance.

Further validating our proposed SCFNet across different defect categories, we conducted comparison experiments with mainstream detection algorithms on the GC10-DET dataset [59]. The specific experimental results and performance are shown in Table 2.

Table 2. Performance comparison of SCFNet algorithm and other object detection algorithms on GC10-DET dataset [59].

Methods	P	R	mAP ₅₀	mAP _{50:95}	GFLOPs	Params/M
Faster R-CNN [18]	0.579	0.656	0.652	0.293	135	41.75
YOLOv5n-7.0 [51]	0.729	0.666	0.699	0.366	4.2	1.77
YOLOv7-tiny [25]	0.707	0.657	0.697	0.344	13.1	6.02
CenterNet [50]	0.726	0.619	0.665	0.308	78.66	32.12
YOLOv8n [49]	0.704	0.65	0.684	0.365	8.1	3.01
YOLOX-tiny [52]	0.659	0.546	0.611	0.259	7.58	5.03
MAR-YOLO [15]	-	-	0.673	-	20.1	-
SCFNet	0.713	0.68	0.704	0.366	5.9	2

Red bold indicates the top-ranking performance, while blue bold indicates the second-ranking performance.

GC10-DET is a dataset of steel surface defects obtained from real industrial environments. This dataset contains 3570 grayscale images of defects in steel plates. The experimental setup is consistent with Section 3.2. According to Table 2, our proposed SCFNet achieves high performance on the GC10-DET dataset, with the model parameter count and computational cost only second to YOLOv5n. The SCFNet upholds detection accuracy while possessing a smaller model size and lower computational cost, rendering it well suited for deployment on terminal detection devices with limited computing capability.

3.5. Data Augmentation Module Discussion

Considering the limited availability and scale of publicly available datasets on industrial steel surface defects, training networks with limited data may result in lower robustness and difficulty in detecting blurry samples. In order to investigate the impact of various augmentation techniques on the accuracy of the model, we conducted data augmentation on the steel surface defect dataset. The data are augmented by six different techniques, as illustrated in Table 3, including flipping transformation, shifting transformation, adding noise transformation, adjusting brightness transformation, rotating transformation, and combining the above techniques. Each augmentation technique doubled the dataset, increasing the original training set of 1440 images to 2880 images.

Table 3. Data augmentation results.

Methods	Augment	mAP ₅₀	mAP _{50:95}
SCFNet	Original	0.778	0.448
SCFNet	Shift	0.785	0.45
SCFNet	Noise	0.781	0.441
SCFNet	Brightness	0.785	0.45
SCFNet	Rotation	0.767	0.454
SCFNet	Flip	0.812	0.469
SCFNet	All	0.797	0.458

Red bold indicates the top-ranking performance, while blue bold indicates the second-ranking performance.

In Table 3, most data augmentation techniques resulted in varying degrees of improvements in the model performance, whereas rotation augmentation reduced the model accuracy. This discrepancy could arise from inconsistencies in size ratios between the rotated images and the original ones, resulting in the distortion of targets when forcibly resized to a consistent size during network preprocessing. However, other data augmentation methods showed improvements in results. Among them, flipping augmentation achieved the highest accuracy improvement, with an mAP₅₀ reaching 0.812. This might be

because in steel defect detection, where defect features may not be prominent, techniques like adding noise, adjusting brightness, and others might make it challenging for the model to propagate gradients correctly; shift could alter image sizes, potentially causing feature loss around the targets. However, flipping augmentation does not cause these issues. Therefore, flip augmentation appears to maximize the detection performance of models on the NEU-DET [10] dataset.

3.6. Ablation Study

To confirm the roles of each module, we conducted ablation studies on the NEU-DET [10] dataset. Using YOLOv8n as a baseline, we replaced the backbone network for feature extraction with the LEM to reduce the model complexity. As a final layer in the feature extraction module, we introduce the ScConv module to enhance the ability to extract features. We also employed BiFPN as a feature fusion network, retaining more original information. As this network is a lightweight detector, ablation studies on the BiFPN and ScConv modules are conducted on the LEM. Table 4 shows the experimental results.

Table 4. Ablation experiment results on the NEU-DET [10] dataset.

Model	LEM	ScConv	BiFPN	mAP ₅₀	mAP _{50:95}	GFLOPs	Params/M
Baseline	-	-	-	0.773	0.444	8.1	3.01
Baseline	✓	-	-	0.783	0.457	5.7	1.9
Baseline	✓	✓	-	0.787	0.455	5.7	1.91
Baseline	✓	-	✓	0.793	0.455	5.9	1.99
Baseline	✓	✓	✓	0.8	0.456	5.9	2

Red bold indicates the top-ranking performance, while **blue** bold indicates the second-ranking performance.

LEM: By replacing the feature extraction module of YOLOv8n with the LEM, the number of model parameters decreased from 3.01 M to 1.9 M, while the gigaflops (GFLOPs) decreased from 8.1 to 5.7. Additionally, mAP₅₀ increased from 0.773 to 0.783, and mAP_{50:95} increased from 0.444 to 0.457. The LEM utilizes Depth-Wise convolution and SE modules for feature extraction, with fewer connections between different blocks and the avoidance of branching structures. A replacement of the backbone network of YOLOv8n with the LEM improves the model detection accuracy while maintaining a lightweight design and reducing the computational requirements.

ScConv Module: ScConv operates on the deepest layer of feature maps, removing redundant spatial and channel information from feature maps and enhancing their representational capacity. Steel surface defect features are not prominent, leading to potential false positives or negatives. By strengthening the representational capacity of the feature maps through the ScConv module, the model can more easily detect steel surface defects. Figure 7 illustrates a comparison of heatmaps with and without the ScConv module. Heatmaps depict the model prediction results for each pixel, typically using colors to indicate the level of confidence associated with each pixel. Warmer tones, such as red, are used to represent pixels with higher confidence, while cooler tones, such as blue, are used to represent pixels with lower confidence. Additionally, heatmaps aid in analyzing model detection results, highlighting areas that are easier to detect or overlook. Features of defects such as crazing and patches are not prominent, making them difficult for the model to recognize. With the addition of the ScConv module, however, the representational capacity of the feature maps is enhanced, thereby improving the model detection ability. In the ablation experiments, adding the ScConv module increased the model mAP₅₀ from 0.783 to 0.787, with a minimal increase in the model parameters and computational load. As a result, the ScConv module has a low number of model parameters and a low computational load, but significantly enhances the network's feature representation capability, resulting in a more accurate model.

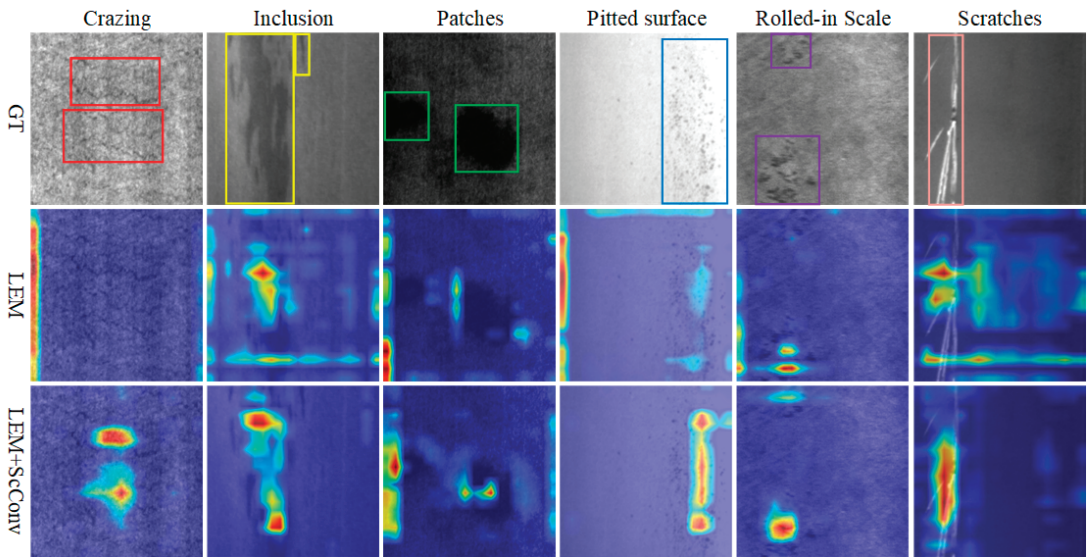


Figure 7. Comparison of heatmaps under ablation of the ScConv module (from the NEU-DET [10] dataset). In the picture, the red box represents “Crazing”, the yellow box represents “Inclusion”, the green box represents “Patches”, the blue box represents “Pitted Surface”, the purple box represents “Rolled-in Scales”, and the pink box represents “Scratches”.

BiFPN: In Table 4, the comparison between the second and fifth rows clearly demonstrates the effectiveness of using BiFPN. The mAP_{50} increased from 0.783 to 0.793 (from the second to the fourth row with BiFPN) and from 0.787 to 0.8 (from the third to the fifth row with BiFPN), while the increase in the model parameters and computational load is minimal. The BiFPN uses unique skip connections and weighted feature fusion mechanisms, allowing the neck network to reuse feature maps and better combine semantic and texture features. This improvement enhances the detection accuracy. Using fewer parameters, BiFPN significantly improves the accuracy by slightly increasing the computational load and parameter count, resulting in a better balance between lightweight design and accuracy.

4. Discussion

Some defective images restrict the detection performance, as depicted in Figure 8 showing cases of detection failure. Defects with low contrast and unclear features in steel materials can lead to missed detections (see Case 1 and 2 of Figure 8). Additionally, there exist defects in steel materials that are highly similar to the background, which can result in false detections (see Case 3 and 4 of Figure 8). In our future work, we intend to incorporate a learnable image enhancement module into the model to improve the detection accuracy of defects with low contrast. Furthermore, we plan to continue researching more effective feature extraction modules to enhance the effectiveness of our approach.

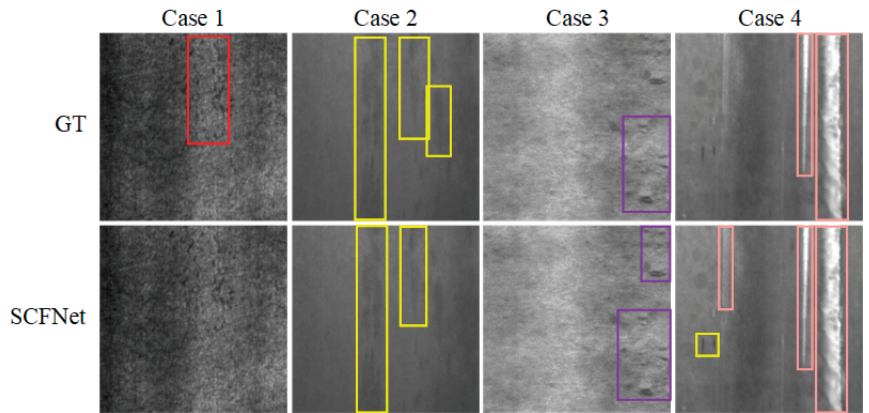


Figure 8. Some failure cases of SCFNet (from the NEU-DET [10] dataset). In the picture, the red box represents “Crazing”, the yellow box represents “Inclusion”, the purple box represents “Rolled-in Scales”, and the pink box represents “Scratches”.

5. Conclusions

Addressing ambiguous defects and low-contrast defects in steel, while accurately identifying defects with similar features but different categories, is crucial for modern industrial production. This article proposes a lightweight steel defect detection algorithm called SCFNet to tackle the aforementioned challenges. To achieve a lightweight defect detection model, SCFNet utilizes the LEM as a feature extraction module. This module is based on Depth-Wise convolution with channel weighting, resulting in stronger capabilities in extracting ambiguous features. We use convolutional structures based on spatial and channel recombination to process the deepest layer feature maps, reducing redundancy and enhancing the model feature representation capability. This module facilitates effective feature representation while disregarding noise information. To preserve more defect texture information, a weighted bidirectional feature pyramid fusion structure is adopted in the neck of the network for feature fusion. In addition, it retains more original content by employing a more effective information propagation mechanism. The experimental results show that on the NEU-DET dataset, compared with most deep learning detection methods, the SCFNet algorithm achieves the highest mAP_{50} metric of 81.2%, the highest $mAP_{50:95}$ metric of 46.9%, the smallest model parameters of 2 M, and the least model computation of 5.9 GFLOPs. SCFNet also achieves the highest accuracy and the smallest computation and model parameters on the GC10-DET dataset. SCFNet demonstrates excellent performance, making it more suitable for practical applications in industrial production.

Author Contributions: Conceptualization, H.L., Z.Y. and L.M.; methodology, W.Y.; software, Y.W.; validation, Y.W., J.D. and K.S.; formal analysis, Y.W.; investigation, M.L.; resources, W.Y.; data curation, Y.W.; writing—original draft preparation, Z.Y.; writing—review and editing, H.L.; visualization, J.D.; supervision, Y.W., L.M. and W.Y.; project administration, H.L.; funding acquisition, H.L. and Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by Hubei Province Young Science and Technology Talent Morning High Lift Project (202319); Open Research Fund Program of LIESMARS (Grant No. 21E02); Doctoral Starting up Foundation of Hubei University of Technology (XJ2023007301); Natural Science Foundation of Hubei Province (2022CFB501); University Student Innovation and Entrepreneurship Training Program Project (202210500028).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Guo, B.; Wang, Y.; Zhen, S.; Yu, R.; Su, Z. SPEED: Semantic prior and extremely efficient dilated convolution network for real-time metal surface defects detection. *IEEE Trans. Ind. Inform.* **2023**, *19*, 11380–11390. [CrossRef]
- Su, J.; Luo, Q.; Yang, C.; Gui, W.; Silvén, O.; Liu, L. PMSA-DyTr: Prior-Modulated and Semantic-Aligned Dynamic Transformer for Strip Steel Defect Detection. *IEEE Trans. Ind. Inform.* **2024**, *20*, 6684–6695. [CrossRef]
- Luo, Q.; Fang, X.; Liu, L.; Yang, C.; Sun, Y. Automated visual defect detection for flat steel surface: A survey. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 626–644. [CrossRef]
- Li, Z.; Wei, X.; Hassaballah, M.; Li, Y.; Jiang, X. A deep learning model for steel surface defect detection. *Complex Intell. Syst.* **2024**, *10*, 885–897. [CrossRef]
- Sampath, V.; Murtua, I.; Martín, J.J.A.; Rivera, A.; Molina, J.; Gutierrez, A. Attention guided multi-task learning for surface defect identification. *IEEE Trans. Ind. Inform.* **2023**, *19*, 9713–9721. [CrossRef]
- Wen, L.; Wang, Y.; Li, X. A new cycle-consistent adversarial networks with attention mechanism for surface defect classification with small samples. *IEEE Trans. Ind. Inform.* **2022**, *18*, 8988–8998. [CrossRef]
- Lian, J.; Jia, W.; Zareapoor, M.; Zheng, Y.; Luo, R.; Jain, D.K.; Kumar, N. Deep-learning-based small surface defect detection via an exaggerated local variation-based generative adversarial network. *IEEE Trans. Ind. Inform.* **2019**, *16*, 1343–1351. [CrossRef]
- Zhang, D.; Song, K.; Wang, Q.; He, Y.; Wen, X.; Yan, Y. Two deep learning networks for rail surface defect inspection of limited samples with line-level label. *IEEE Trans. Ind. Inform.* **2020**, *17*, 6731–6741. [CrossRef]
- Shen, K.; Zhou, X.; Liu, Z. MINet: Multiscale Interactive Network for Real-Time Salient Object Detection of Strip Steel Surface Defects. *IEEE Trans. Ind. Inform.* **2024**, 1–11. [CrossRef]
- Song, K.; Yan, Y. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Appl. Surf. Sci.* **2013**, *285*, 858–864. [CrossRef]
- Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
- Luo, Q.; Su, J.; Yang, C.; Silvén, O.; Liu, L. Scale-selective and noise-robust extended local binary pattern for texture classification. *Pattern Recognit.* **2022**, *132*, 108901. [CrossRef]
- Lu, H.-P.; Su, C.-T. CNNs combined with a conditional GAN for Mura defect classification in TFT-LCDs. *IEEE Trans. Semicond. Manuf.* **2021**, *34*, 25–33. [CrossRef]
- Wen, L.; Li, X.; Gao, L. A new reinforcement learning based learning rate scheduler for convolutional neural network in fault classification. *IEEE Trans. Ind. Electron.* **2020**, *68*, 12890–12900. [CrossRef]
- Zhang, H.; Li, S.; Miao, Q.; Fang, R.; Xue, S.; Hu, Q.; Hu, J.; Chan, S. Surface defect detection of hot rolled steel based on multi-scale feature fusion and attention mechanism residual block. *Sci. Rep.* **2024**, *14*, 7671. [CrossRef]
- Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 580–587.
- Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Washington, DC, USA, 7–13 December 2015; pp. 1440–1448.
- Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems 28 (NIPS 2015), Montreal, QC, Canada, 7–12 December 2015; Volume 28.
- He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
- Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
- Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
- Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
- Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
- Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
- Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 7464–7475.
- Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 213–229.
- Shumin, D.; Zhoufeng, L.; Chunlei, L. AdaBoost learning for fabric defect detection based on HOG and SVM. In Proceedings of the 2011 International Conference on Multimedia Technology, Hangzhou, China, 26–28 July 2011; pp. 2903–2906.
- Chondronasios, A.; Popov, I.; Jordanov, I. Feature selection for surface defect classification of extruded aluminum profiles. *Int. J. Adv. Manuf. Technol.* **2016**, *83*, 33–41. [CrossRef]

29. Liang, Y.; Xu, K.; Zhou, P. Mask gradient response-based threshold segmentation for surface defect detection of milled aluminum ingot. *Sensors* **2020**, *20*, 4519. [CrossRef]
30. Wu, X.-Y.; Xu, K.; Xu, J.-W. Application of undecimated wavelet transform to surface defect detection of hot rolled steel plates. In Proceedings of the 2008 Congress on Image and Signal Processing, Sanya, China, 27–30 May 2008; pp. 528–532.
31. Zhao, T.; Chen, X.; Yang, L. IPCA-SVM based real-time wrinkling detection approaches for strip steel production process. *Int. J. Wirel. Mob. Comput.* **2019**, *16*, 160–165. [CrossRef]
32. Gong, R.; Wu, C.; Chu, M. Steel surface defect classification using multiple hyper-spheres support vector machine with additional information. *Chemom. Intell. Lab. Syst.* **2018**, *172*, 109–117. [CrossRef]
33. Chu, M.; Liu, X.; Gong, R.; Liu, L. Multi-class classification method using twin support vector machines with multi-information for steel surface defects. *Chemom. Intell. Lab. Syst.* **2018**, *176*, 108–118. [CrossRef]
34. Zhang, J.; Wang, H.; Tian, Y.; Liu, K. An accurate fuzzy measure-based detection method for various types of defects on strip steel surfaces. *Comput. Ind.* **2020**, *122*, 103231. [CrossRef]
35. Mei, L.; Hu, X.; Ye, Z.; Tang, L.; Wang, Y.; Li, D.; Liu, Y.; Hao, X.; Lei, C.; Xu, C. GTMFuse: Group-Attention Transformer-Driven Multiscale Dense Feature-Enhanced Network for Infrared and Visible Image Fusion. *Knowl. Based Syst.* **2024**, *293*, 111658.
36. Xu, C.; Ye, Z.; Mei, L.; Yu, H.; Liu, J.; Yalikul, Y.; Jin, S.; Liu, S.; Yang, W.; Lei, C. Hybrid Attention-Aware Transformer Network Collaborative Multiscale Feature Alignment for Building Change Detection. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–14. [CrossRef]
37. Yang, W.; Mei, L.; Ye, Z.; Wang, Y.; Hu, X.; Zhang, Y.; Yao, Y. Adjacent Self-Similarity Three-dimensional Convolution for Multi-modal Image Registration. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 6002505. [CrossRef]
38. Xu, C.; Ye, Z.; Mei, L.; Shen, S.; Sun, S.; Wang, Y.; Yang, W. Cross-Attention Guided Group Aggregation Network for Cropland Change Detection. *IEEE Sens. J.* **2023**, *23*, 13680–13691. [CrossRef]
39. Yang, W.; Shen, P.; Ye, Z.; Zhu, Z.; Xu, C.; Liu, Y.; Mei, L. Adversarial Training Collaborating Multi-Path Context Feature Aggregation Network for Maize Disease Density Prediction. *Processes* **2023**, *11*, 1132. [CrossRef]
40. Zhao, C.; Shu, X.; Yan, X.; Zuo, X.; Zhu, F. RDD-YOLO: A modified YOLO for detection of steel surface defects. *Measurement* **2023**, *214*, 112776. [CrossRef]
41. Wang, R.; Liang, F.; Mou, X.; Chen, L.; Yu, X.; Peng, Z.; Chen, H. Development of an improved yolov7-based model for detecting defects on strip steel surfaces. *Coatings* **2023**, *13*, 536. [CrossRef]
42. Yu, Z.; Wu, Y.; Wei, B.; Ding, Z.; Luo, F. A lightweight and efficient model for surface tiny defect detection. *Appl. Intell.* **2023**, *53*, 6344–6353. [CrossRef]
43. Liu, G.-H.; Chu, M.-X.; Gong, R.-F.; Zheng, Z.-H. DLF-YOLOF: An improved YOLOF-based surface defect detection for steel plate. *J. Iron Steel Res. Int.* **2023**, *31*, 442–451. [CrossRef]
44. Shao, Y.; Fan, S.; Sun, H.; Tan, Z.; Cai, Y.; Zhang, C.; Zhang, L. Multi-Scale Lightweight Neural Network for Steel Surface Defect Detection. *Coatings* **2023**, *13*, 1202. [CrossRef]
45. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.
46. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
47. Li, J.; Wen, Y.; He, L. Scconv: Spatial and channel reconstruction convolution for feature redundancy. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 6153–6162.
48. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
49. Ultralytics/Ultralytics. 2023. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 29 October 2023).
50. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as points. *arXiv* **2019**, arXiv:1904.07850.
51. Jocher, G. Stoken Yolov5. Available online: <https://github.com/ultralytics/yolov5/releases/tag/v7.0> (accessed on 20 November 2023).
52. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
53. Xia, K.; Lv, Z.; Zhou, C.; Gu, G.; Zhao, Z.; Liu, K.; Li, Z. Mixed receptive fields augmented YOLO with multi-path spatial pyramid pooling for steel surface defect detection. *Sensors* **2023**, *23*, 5114. [CrossRef] [PubMed]
54. Yu, J.; Wang, C.; Xi, T.; Ju, H.; Qu, Y.; Kong, Y.; Chen, X. Development of an Algorithm for Detecting Real-Time Defects in Steel. *Electronics* **2023**, *12*, 4422. [CrossRef]
55. Huang, Y.; Tan, W.; Li, L.; Wu, L. WFRE-YOLOv8s: A New Type of Defect Detector for Steel Surfaces. *Coatings* **2023**, *13*, 2011. [CrossRef]
56. Wang, H.; Yang, X.; Zhou, B.; Shi, Z.; Zhan, D.; Huang, R.; Lin, J.; Wu, Z.; Long, D. Strip surface defect detection algorithm based on YOLOV5. *Materials* **2023**, *16*, 2811. [CrossRef]
57. Fan, J.; Wang, M.; Li, B.; Liu, M.; Shen, D. ACD-YOLO: Improved YOLOv5-based method for steel surface defects detection. *IET Image Process.* **2024**, *18*, 761–771. [CrossRef]

58. Ren, F.; Fei, J.; Li, H.; Doma, B.T. Steel Surface Defect Detection Using Improved Deep Learning Algorithm: ECA-SimSPPF-SIoU-Yolov5. *IEEE Access* **2024**, *12*, 32545–32553. [CrossRef]
59. Lv, X.; Duan, F.; Jiang, J.-J.; Fu, X.; Gan, L. Deep metallic surface defect detection: The new benchmark and detection network. *Sensors* **2020**, *20*, 1562. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Confluence Effect of Debris-Filled Damage and Temperature Variations on Guided-Wave Ultrasonic Testing (GWUT)

Samuel C. Olisa and Muhammad A. Khan *

School of Aerospace, Transport and Manufacturing, Cranfield University, Bedford MK43 0AL, UK;
s.c.olisa@cranfield.ac.uk

* Correspondence: muhammad.a.khan@cranfield.ac.uk

Abstract: Continuous monitoring of structural health is essential for the timely detection of damage and avoidance of structural failure. Guided-wave ultrasonic testing (GWUT) assesses structural damages by correlating its sensitive features with the damage parameter of interest. However, few or no studies have been performed on the detection and influence of debris-filled damage on GWUT under environmental conditions. This paper used the pitch-catch technique of GWUT, signal cross-correlation, statistical root mean square (RMS) and root mean square deviation (RMSD) to study the combined influence of varying debris-filled damage percentages and temperatures on damage detection. Through experimental result analysis, a predictive model with an R2 of about 78% and RMSE values of about 7.5×10^{-5} was established. When validated, the model proved effective, with a comparable relative error of less than 10%.

Keywords: ultrasonic; guided wave; RMS; damage detection

1. Introduction

High economic structural assets in aerospace, shipbuilding, oil, and gas are constructed using thin-wall metallic materials. The structural assets convey products prone to environmental hazards that could happen through structural failure [1,2]. Failures in structures are deeply traced to undetected damage. Cracks and corrosion are forms of damage that are mainly studied individually using guided-wave ultrasonics [3]. Damage is a state of structure that differs from its pristine state [4]. It is an indispensable factor affecting the structure's service lifecycle and its reliability for intended operational services. Damage may be initiated in structural material during the manufacturing process at the micro-level and progresses into the macro-level during the service time of the structure. Formed damage, such as a crack in structures, could accumulate debris over time and accelerate structure deterioration, paving the way for catastrophic failure. The accumulated debris causes crack closure, affects crack propagation and enhances pitting corrosion activities, as explained in [5,6]. The prime effects of corrosion and crack on metallic structures are thickness thinning, rigidity reduction, integrity and load-carrying-capacity degradation [7], shortening the structure's service lifecycle and making it unreliable. Hence, continuous monitoring of the structural assets with cost-effective damage-sensitive NDT devices and techniques is a priority to avert catastrophic failure, suggest timely maintenance, and save the economy. Most of the previous studies have been on the monitoring and prediction of empty cracks (cracks without debris or dry cracks) in structures using various structural health-monitoring techniques [8–14]. In recent decades, the guided-wave ultrasonic technique has received more attention than other NDT techniques for SHM [15]. The high interest in GWUT is due to its potential for long-range inspection, non-invasive inspection, the lightweight and small size of the ceramic PZT, and cost-effectiveness. It has been successfully used for the detection and localisation of various damages in structures [9,15–18]. The effect of environmental and operational conditions on the ultrasonic guided wave,

Citation: Olisa, S.C.; Khan, M.A. Confluence Effect of Debris-Filled Damage and Temperature Variations on Guided-Wave Ultrasonic Testing (GWUT). *Processes* **2024**, *12*, 957. <https://doi.org/10.3390/pr12050957>

Academic Editor: Zhibin Lin

Received: 30 March 2024

Revised: 16 April 2024

Accepted: 27 April 2024

Published: 8 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

especially temperature, has been well studied [19,20]. However, the influence of debris-filled damage cum temperature variation on the guided wave has not been studied. The influence of debris-filled damage of different percentages under temperature variation is meticulously studied in this paper using statistical RMS and RMSD of the captured response signals from the pitch-catch ultrasonic guided-wave configuration. The suitability of RMS is because the guided-wave ultrasonic used in SHM is a continuous propagation of energy from one point of the structure to another. And, the sensitivity of the signal's energy to damage is significantly higher than the amplitude, since it is the square of amplitude by value.

2. Theory

2.1. Structural Health Monitoring (SHM)

Structural health monitoring (SHM) via an ultrasonic guided wave uses an embedded ceramic piezoelectric (PZT) transducer to transform an oscillating excitation signal into an actuating signal that would cause a stress wave to propagate in the host structure through surface pinching. The propagating wave interacts with the structure and becomes captured as a response signal by another ceramic PZT called a sensor [21]. The PZT operates through a linear transduction mechanism as expressed in the linear constitutive equations in [9,15].

$$\text{Actuation : } S_{ij} = s_{ijkl}^E T_{kl} + d_{kij} E_k \quad (1)$$

$$\text{Sensing : } D_i = d_{ikl} T_{kl} + \epsilon_{ik}^T E_k \quad (2)$$

where S_{ij} and T_{kl} , are the strain and stress mechanical variables, respectively. D_i and E_k , are electrical displacement and electrical field variables, respectively. s_{ijkl}^E , is the mechanical compliance at zero electric fields, ϵ_{ik}^T , is the dielectric constant at zero mechanical stress, and d_{ikl} and d_{kij} , denote a coupling between the electrical and mechanical variables. Equations (1) and (2) are actuation and sensing equations. Hence, the sensor PZT captures the strain effect of the propagating stress wave and transforms it into a signal voltage. The sensor's output voltage is proportional to the amount of the strain effect caused by the propagating wave and the thickness of the sensor.

2.2. Wave Motion in Thin Plate

The ultrasonic wave propagates as a distributed stress on a thin plate. The stress causes material particles to vibrate and transfer energy from one point to another as waves. The waves are guided to propagate within the boundaries of the structure. The motion of the vibrating particle in the plate is governed by Equation (3). The boundary condition of surface traction force is in Equation (4).

$$\mu \cdot u_{i,jj} + (\lambda + \mu) u_{j,ji} + \rho f_i = \rho \cdot \ddot{u}_i \quad (i, j = 1, 2, 3) \quad (3)$$

$$t_i = \sigma_{ji} n_j \quad (4)$$

where u is the displacement of the plate via the particle motion, λ and μ are Lamé's constants, ρ is the density of the plate particles, \ddot{u}_i is the acceleration of the vibrating particles, f_i is the force body on the material, t_i is surface tractions at the normal n_j , and σ_{ji} , is the stress on the plate surface. It is considered zero for a traction-free body condition. The solution to Equation (3) by the method of potential separation, as in [22], results in characteristics frequency equations of lamb wave grouped into two modes as in Equations (5) and (6). The terms of the equations are as expanded in Equations (7) through (9).

$$\text{Symmetric modes : } \frac{\tan(qb)}{\tan pb} = \frac{4k^2 pq}{(p^2 - k^2)^2} \quad (5)$$

$$\text{Anti-symmetric modes : } \frac{\tan(qb)}{\tan(pb)} = -\frac{(p^2 - k^2)^2}{4k^2pq} \quad (6)$$

where

$$k = \frac{2\pi}{\lambda_w}$$

$$p = \sqrt{c_L^2 - \left(\frac{\omega}{C_p}\right)^2} \quad (7)$$

$$q = \sqrt{c_T^2 - \left(\frac{\omega}{C_p}\right)^2} \quad (8)$$

$$C_g = \frac{\partial\omega}{\partial k} = \frac{C_p}{1 - \frac{\omega \partial C_p}{k \partial \omega}} \quad (9)$$

The velocity of the propagating wave is related to the lame's constant and mechanical properties of the material, as expressed in Equations (10) and (11).

$$C_T = \sqrt{\frac{\mu}{\rho}} = \sqrt{\frac{E}{2\rho(1+\nu)}} \quad (10)$$

$$C_L = \sqrt{\frac{\lambda + 2\mu}{\rho}} = \sqrt{\frac{(1-\nu)E}{(1+\nu)(1-2\nu)\rho}} \quad (11)$$

where C_T is the transverse velocity, and C_L is the longitudinal or radial velocity of the wave in a solid medium. C_p is the phase velocity of lamb wave, and ρ , E , and ν are the material density, elastic modulus, and Poisson ratio, respectively. It is pertinent to observe that the velocity of the propagating wave depends on the mechanical properties of the waveguide. Hence, any consequential effect on the material would affect sensitive parameters of the propagating wave, especially its energy being transferred as the wave propagates.

2.3. Wave Dispersion Curve

In an active structural health-monitoring (SHM) mode, the actuator pinches the structure to create probing waves when excited with an oscillating electrical voltage. The frequency of the probing signal is often selected from the dispersion curve. The dispersion curves are a set of curves that represent the propagation of wave modes in each waveguide. It describes the relationship between wave velocity (group or phase) and the excitation frequency [18]. The dispersion curve is determined using Disperse software or a Dispersion calculator version V2.3 by providing the necessary properties of the plate material [23]. The dispersion curves of the 3.00 mm mild carbon steel plate with properties listed in Table 1 are in Figure 1.

Table 1. Mechanical properties of mild carbon steel.

Material	Young Modulus, E (N/m ²)	Poisson's Ratio, ν	Density, ρ (kg/m ³)	Length, L (mm)	Width, W (mm)	Thickness, Th (mm)
Mild Carbon steel	2×10^{11}	0.289	7800	500	300	3

From the dispersion curves, only the fundamental modes exist when the frequency is below 500 KHz, but beyond it, multi-modes exist. Also, the fundamental symmetric mode possesses higher velocity, which is relatively more constant at the early frequency range

than the fundamental asymmetric mode. The selection of probing frequency is critical for damage detection. There are some criteria used to select an appropriate probing frequency. One of the criteria is that the frequency must be sensitive to the targeted damage and possess enough energy to propagate the needed distance. An excitation frequency with the largest RMS is selected through experimental frequency sweep because it possesses more energy than any other and could travel the longest distance. Recall the rule of thumb for damage detection, which states that damage would stand a higher chance of detection if it is larger than $\frac{1}{2}\lambda$ of the probing wave [24]. Hence, it is necessary to compare the probing wave's wavelength with the targeted damage's length. The wavelength of the excitation frequency can be determined using Equation (13).

$$v = f \times \lambda \quad (12)$$

$$\lambda = v/f \quad (13)$$

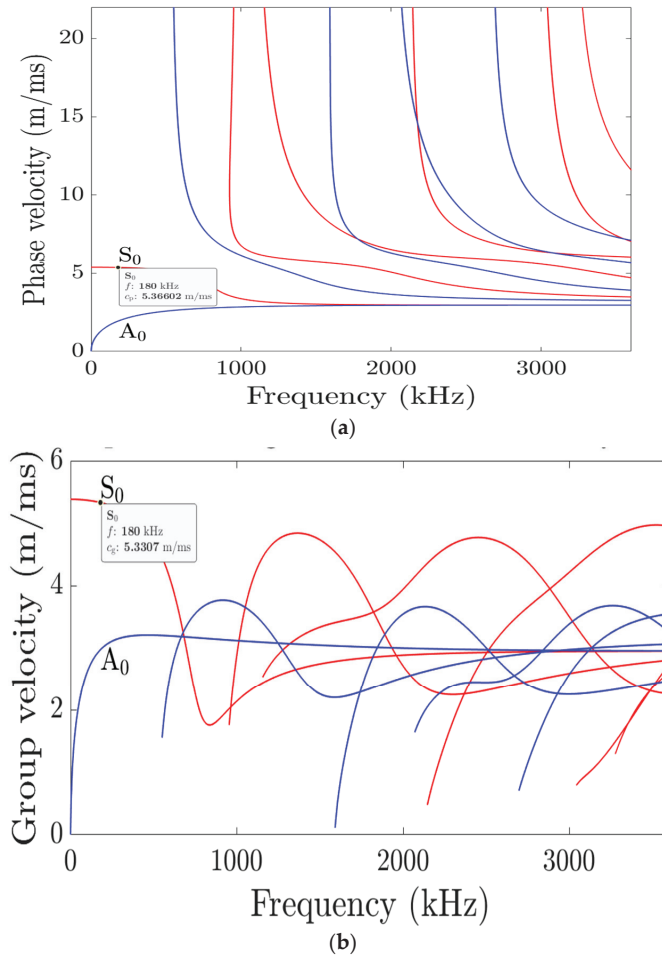


Figure 1. (a) Phase dispersion curve, (b) Group dispersion curve.

2.4. Wave–Damage Interaction Effects and Inspection-Transducer Configuration

Guided-wave ultrasonic inspection could be performed using active or passive ultrasonic configuration techniques. The exciter is vital in guided-wave inspection, which determines if a configuration is active or passive. When damage generates the propagating wave, a passive mode is configured [21], and all the embedded transducers would function like human ears and listen to pick up the wave. The active mode is configured when the propagating wave is due to the excitation of the actuator with an oscillating electrical voltage of a given frequency [21,25–27]. The vibration of the actuator causes atoms of the structure around its installation to start displacing their positions temporarily. This mechanism transfers wave energy from one point of the structure to another. The propagating stress wave is constrained by the boundaries of the structure (waveguide), hence the guided wave. The wave interaction with the structure, especially with damage or discontinuity, causes scattering, reflections, mode conversion, and some energy absorption at the damaged spot [28]. Figure 2b depicts the effect of wave–damage interaction. Figure 2b shows that the through-transmitted wave is much smaller than the edge-scattered wave. Hence, capturing the edge-scattered wave would be more valuable than the through-transmitted wave. Also, depending on the nature of the damage, the through transmitted is the resultant wave after absorption, back reflection, and diffraction of the incident wave, which is usually small in value.

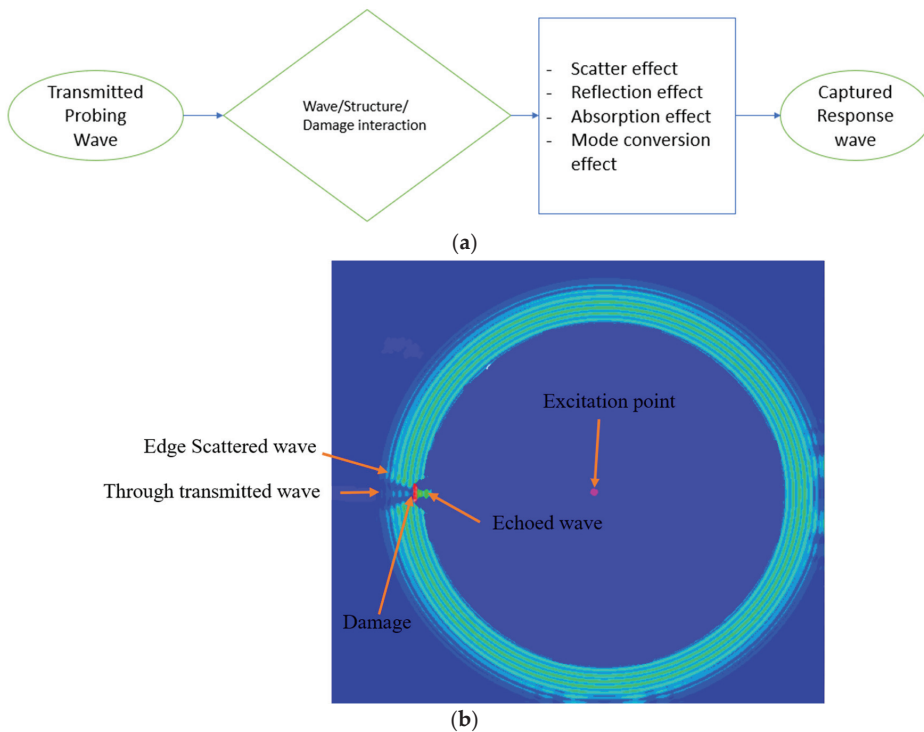


Figure 2. (a) Wave–damage interaction effects. (b) Typical wave–damage interaction effect.

All the effects mentioned above make it possible for damage to be detected and localised using guided-wave techniques and a sensitive wave parameter called the signature or feature. The extent of wave reflection or scattering depends on the encountered damage's nature, shape, size, and orientation. Pitch–catch and Pulse–Echo transducer configuration topology are often used in the active-mode inspection technique, as shown in Figure 3. The

choice of either topology depends on the nature of the damage. This study uses pitch-catch configuration topology to acquire wave signals interacting with the damaged and scattered.

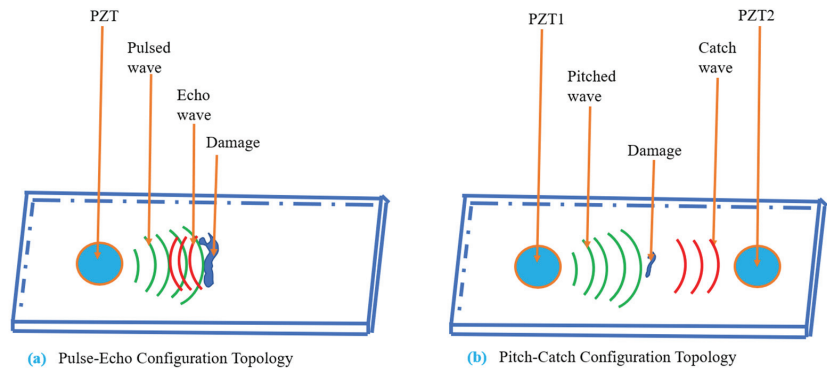


Figure 3. Active-mode transducers configuration topology.

2.5. Temperature Effect on Guided Wave

Damage and temperature affect mechanical structures differently but collectively influence structural degradation [29,30]. The core influence of temperature on a mechanical structure is thermal expansion and assisted degradation of the structure's strength and stiffness. In particular, an increase in temperature causes a decrease in the elastic modulus of a metallic structure, and features of the propagating wave depend directly on most of the structure's mechanical properties, especially the elastic modulus, as in Equations (10) and (11). Assuming a linear dependence of the wave characteristics on temperature, the equation below would be suitable for the study.

$$S(T) = S(T_0) + \frac{\partial S(T)}{\partial T} \Delta T \quad (14)$$

where

S = any characteristics parameter of the wave.

$S(T_0)$ = the value of the wave feature at a reference temperature.

T = current-state temperature of the structure.

T_0 = reference temperature of the structure.

$$\Delta T = (T - T_0)$$

$\frac{\partial S(T)}{\partial T}$ = the parameter sensitivity to temperature.

2.6. Signature/Wave Feature Extraction and Processing

Damage detection and identification largely depend on the approach and signal-processing technique adopted. The captured response signal carries sufficient information about the structure. Wave amplitude is one of the sensitive parameters of the wave and could be easily extracted from the signal. The damage in the structure could modulate the signal amplitude to either higher or lower values. However, the signal energy is more sensitive to damage and is proportional to the square of its amplitude. Hence, we adopted root mean square (RMS) and root mean square deviation (RMSD) to extract the energy feature of the response wave and determine the damage index. A model for predicting the debris-filled damage under temperature variation will be established through the extracted

feature of the wave. The average RMS of the two sensors' signals used in the design is given in Equation (15), and the relative RMS is expressed in Equation (16).

$$X_{ARMS} = \frac{1}{2} \left(\left(\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \right)_{PZT2} + \left(\sqrt{\frac{1}{N} \sum_{j=1}^N x_j^2} \right)_{PZT4} \right) \quad (15)$$

$$R_{RMS} = \frac{X_{ARMS}}{E_{RMS}} \quad (16)$$

where

E_{rms} = the RMS of the excitation signal.

x_i^2 and x_j^2 are the captured response signals from PZT2 and PZT4, respectively.

N is the number of samples of the signal. Hence, the R_{RMS} is computed for each study case and used to establish the predictive regression model for each damage situation. The damage index is computed by determining the RMSD between the healthy state of the structure and its damaged state. Equation (17) expresses the damage index.

$$DI_{SC} = \sum_{i=1}^M \sqrt{\frac{\sum (R_{RMS_i^D} - R_{RMS_i^H})^2}{\sum (R_{RMS_i^H})^2}} \quad (17)$$

where

$R_{RMS_i^D}$ is the relative RMS of the sensors' captured response signals in the damaged state of the plate, while the $R_{RMS_i^H}$ is the healthy counterpart of the sensors' response signals.

3. Materials and Methods

3.1. Material Preparation

The materials used in the study are mild carbon steel plates. Mild carbon steel has a low carbon percentage. It is tough and ductile, making it a predominantly used material to construct essential and critical structures of high economic value, such as oil and gas pipelines, rail tracks, and bridges. The dimensions of the mild carbon steel plate used in this study are 500.00 mm × 300.00 mm × 3.00 mm. Damage of 40.00 mm length, 5.00 mm width, and varying depth of 1.00 mm to 2.50 mm at an interval of 0.50 mm, as depicted in Figure 4, was carefully machined at the mid-distance between the actuator, PZT1 and sensor, and PZT3. Circular PZT transducers were used due to their omnidirectional capability of radiating and receiving response signals, unlike rectangular PZTs, which are orientation dependent [31]. The variation in the damage depth was used to simulate the material-thinning thickness that corrosion activities would cause. Table 1 reveals the mechanical properties of the used plate material. Wave reflection from the edges of the plate is known to contribute to the complexity of the response wave. In particular, the edge-reflected wave would cause mutual interference with the response wave of the actuated signal. Hence, DAS modelling air-dried clay [32] is installed around the edges of the plate to absorb any wave reflection from it. Also, the transition of random vibration between the plate and the working table was minimised by placing soft foams between the two.

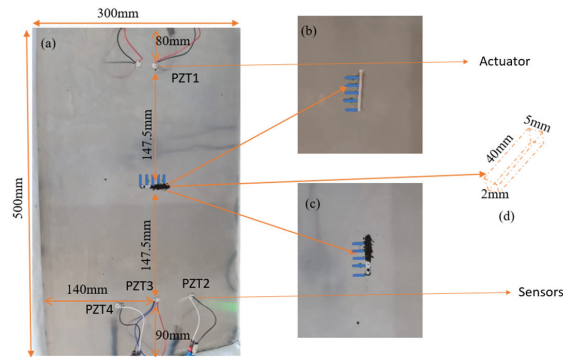


Figure 4. (a) The implemented damaged plate with bonded PZTs, (b) the calibrated empty damage, and (c) the progressively filled damage.

3.2. Experimental Setup

Pitch–catch configuration topology was adopted and implemented using four ceramic PZT transducers from Mouser Electronics, Wycombe, UK (PZT: 7.000 mm × 0.195 mm) as shown in Figure 4. A circular type of PZT transducer was used due to its omnidirectional capability in capturing response signals. Before transducers were installed, acetone was used to clean the plate’s surfaces, and it was allowed to dry up for about 10 min. The transducers were embedded on the top surface of the plate using epoxy adhesive. The curing time of the epoxy adhesive is 150 min and has an operating temperature range of $-55\text{ }^{\circ}\text{C}$ and $120\text{ }^{\circ}\text{C}$ [33]. However, we allowed 180 min to ensure the proper curing of the adhesive. One transducer, PZT1, was wired as the actuator to generate the probing wave, while the remaining three transducers were used as sensors. The sensor PZTs were arranged and named as left sensor (PZT2), middle sensor (PZT3), and right sensor (PZT4) when viewed from the actuator’s position. The epoxy adhesive was lightly applied to bond the PZTs on the top surface of the plate and to eliminate the high impedance that air would have introduced. The properties of the ceramic PZTs are in Table 2. The PZTs are installed 102.00 mm from the end edges of the plate and at least 100.00 mm from the side edges. The PZT1 is 300.00 mm separated from the PZT3. The distance between the middle sensor, PZT3, and either of the two sensors is 50.00 mm. The excitation signal was generated using an arbitrary function generator (TG550 Function Generator), while the response signals were visualised and registered using an Agilent Technologies Mixed Signal Oscilloscope (MSO-X 3024A), with a sampling frequency of 1 MHz. The experiment study was conducted in 3 phases.

Table 2. The properties of PZT used in the study are PZT-5A.

Parameter	Unit	Min. Value	Typical Value	Max. Value
Diameter of ceramics	mm	6.80	7.00	7.20
Thickness of ceramics	μm	175	195	215
Curie temperature	Tc		340	
Piezoelectric constant	pC/N		420	
Elastic compliance	m^2/N		19.6×10^{-12}	
Serial resonance frequency fs	kHz	-5%	285	+5%

3.3. Study Phases

Phase I: To ensure that the testing rig works and proper excitation frequency is selected and used for the study, PZT1 was excited with different centre frequencies of an amplitude-modulated wave. The modulating frequency is internal 400 Hz of the arbitrary function generator, and the different centre frequencies are 100 KHz, 180 KHz, 280 KHz,

and 360 KHz. Each modulated wave was used to excite the PZT1 5 times and averaged to ensure repeatability and minimise the random-noise effect. The signals captured by PZT2 and PZT4 were processed using Equations (15) and (16) to determine which excitation frequency offered the maximum signal energy. The signal captured by PZT3 was processed using the cross-correlation method to determine the time of flight (ToF_E) of the wave packets and compared it with the theoretically determined time of flight, (ToF_T).

Phase II: The selected excitation frequency was used to acquire baseline signals by PZT2 and PZT4 under the influence and no influence of temperature on the wave signal at the healthy state of the plate. A silicone heat mat and K-type thermocouple temperature sensor were integrated into the testing rig to generate heat and acquire the plate's temperature, respectively. A control circuitry was designed using Arduino Mega and solid-state relay (SSL) to regulate the heating rate of the silicone heat mat and ensure that the targeted temperature value was achieved and maintained before baseline response signals were acquired and recorded. The temperature was varied from 30 °C to 70 °C at an interval of 5 °C.

Phase III: After baseline signal acquisition, the healthy plate was replaced with an unhealthy plate that had damage dimensions, as depicted in Table 3. The response signals were captured by PZT2 and PZT4 after they had interacted with the following:

Empty damage at influence and no influence of temperature.

Damage filled with different percentages of corrosion debris at influence and no influence of temperature. The testing rig used is as in Figures 4 and 5.

Table 3. The damage dimension.

S/N	Damage Depth (mm)	Damage Length (mm)	Damage Width (mm)
1	1.00	40.00	5.00
2	1.50	40.00	5.00
3	2.00	40.00	5.00
4	2.50	40.00	5.00

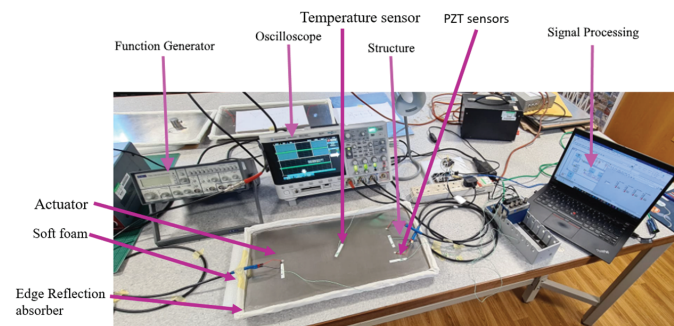


Figure 5. The experimental setup for data acquisition.

4. Results

4.1. Excitation Frequency Selection

In structural health monitoring, SHM, inspection frequency is essential for detecting damage. As stated in Phase I, two approaches were used to determine the excitation frequency for the studies. In the first approach, PZT1 was excited with different centre frequency signals, and the corresponding response signals for each frequency were captured using the PZT2 and PZT4 sensors. The following four frequencies, 100 KHz, 180 KHz, 280 KHz, and 360 KHz, were used to excite the actuator. The response signals captured by the sensors were processed for the relative average RMS value as in Equation (16). The

response signals were sampled at 1 MHz to avoid signal aliasing due to under-sampling, according to Nyquist's theory in Equation (18).

$$f_s \geq 2 \times f_{max} \quad (18)$$

where f_{max} and f_s are the maximum frequency of the signal and the sampling frequency, respectively. Figure 6 is the plot of the computed result. The response signal of excited 180 KHz exhibited the maximum relative RMS value compared to others. The RMS value starts decreasing shortly after 180 kHz. The second approach computed the response-signal time of flight (ToF) generated using 180 KHz and captured by PZT3. ToF is a crucial feature of guided-wave ultrasonic for damage localisation. In a pitch-catch configuration, ToF is the time taken for the excited wave to propagate from the actuator to the sensor. Different methods are used to determine ToF, but cross-correlation ensures a better result because it compares the similarity between two wave signals of the same length and shape and eliminates spurious reflections caused by the structure boundaries [34,35]. Also, velocity is a needful feature of the wave that depends on and changes with variations in the structure parameters. The knowledge of the propagating wave velocity is essential, especially for damage localisation. From the group-velocity dispersion curve generated using the mechanical parameters of the plate material and a dispersion calculator [36], the fundamental symmetric mode was observed to be the fastest and non-dispersive wave up to 500 KHz, with relatively stable group velocity, as shown in Figure 1b. The group velocity of the wave at 180 KHz is 5330.7 m/s. The signal's theoretical time of flight (ToF) is computed using the group velocity and the actuator-sensor distance. Experimentally, the excitation signal is cross-correlated with the response signal captured by the PZT3 sensor to obtain the ToF of the signal using Equation (19). Figure 7 shows the cross-correlation between the excitation and captured response signals.

$$R_{i-j} = \frac{\sum(x_i(t) - \bar{x})(y_j(t) - \bar{y})}{\sqrt{\sum(x_i(t) - \bar{x})^2 \sum(y_j(t) - \bar{y})^2}} \quad (19)$$

where

$x_i(t)$ = excitation signal of the i -sensor.

$y_i(t)$ = response signal of the j -sensor.

\bar{x} = mean of the excitation signal.

\bar{y} = mean of the captured response signal.

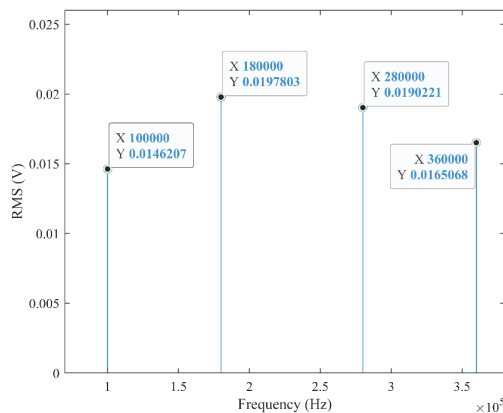


Figure 6. The relative average RMS of response signals captured by PZT2 and PZT4 against the excitation centre frequency.

The relationship between the propagating wave velocity, distance covered, and ToF is expressed in Equation (20). The equation defines the theoretical time of flight, ToF_T , as

$$ToF_T = \frac{d_{a-s}}{V} \quad (20)$$

where d_{a-s} is the distance between the actuator, PZT1, and the middle sensor, PZT3. The distance is 300.00 mm. The group velocity of the propagating wave packet is V . From the group-velocity dispersion curve of Figure 1b, the nearly stable group velocity at 180 KHz frequency is 5330.7 m/s. Using Equation (20), the theoretical ToF of the signal is

$$ToF_T = \frac{0.3}{5330.7} = 5.6278 \times 10^{-5} \text{s} \quad (21)$$

By cross-correlating the excitation signal with the response signals captured by the PZT3 sensor, we observed that the excitation signal lags the response signal by 261 matching index values, as shown in the resultant cross-correlated signal of Figure 7c. The 261 matching index value is the point at which the two signals indicate the maximum peak of similarity. The matching index value is then converted into time of flight using Equation (22).

$$ToF_E = \left| \frac{(I_{max} - I_{total})}{I_{total}} \right| \times t \quad (22)$$

where

I_{max} = matching index value of the maximum peak of the correlated signals.

I_{total} = total matching index value of the correlated signal.

t = sampling time in seconds.

$$ToF_E = \left| \frac{(261 - 5000)}{5000} \right| \times 0.001 = 5.22 \times 10^{-5} \text{s}. \quad (23)$$

$$Relative\ Error, R.E = \left| \frac{ToF_E - ToF_T}{ToF_T} \right| = 0.072 \text{ or } 7.2\% \quad (24)$$

Hence, the relative error percentage between the theoretical and experimental ToF is 7.2%. The result signifies that the system setup works well, and the stress wave propagates as designed. Therefore, the S_0 probing frequency of 180 KHz is selected to further the studies. The choice of the S_0 mode also encompasses its in-plane displacement that contains energy for long coverage inspection, unlike the A_0 mode, which would leak energy to surroundings through out-of-plane displacement [36].

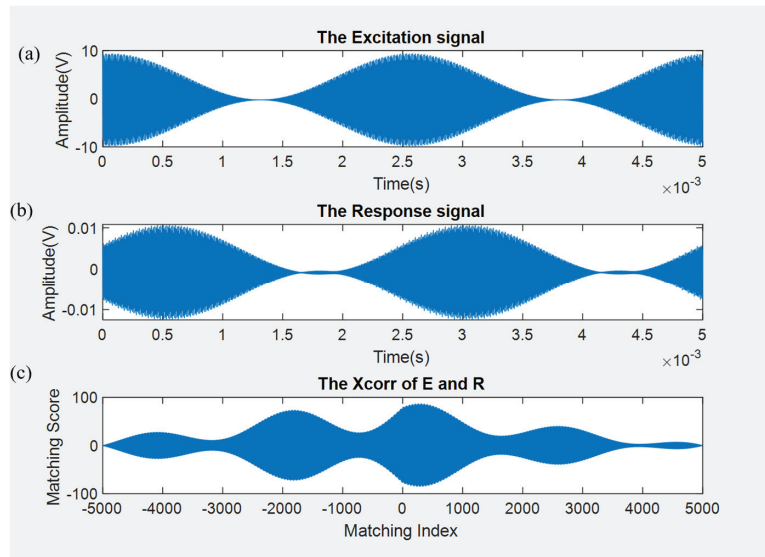


Figure 7. (a) The excitation signal, (b) the response signal, and (c) the cross-correlation of the excitation signal with the response signal captured by PZT3.

4.2. Baseline Signal Acquisition

Acquisition of the baseline signal is vital in structural health monitoring, as it helps to quickly detect a deviant behaviour of the structure at an early stage. The PZT2 and PZT4 captured the response signals that were excited using PZT1. In the absence of temperature variation, 10,000 samples were captured and processed for their RMS values. The relative average RMS of the two sensors' response signals is 2.236×10^{-3} . The experiment was repeated when the heat source was activated. The plate temperature was raised from 30 °C to 70 °C at an interval of 5 °C. At each temperature level, response signals were captured by the sensors PZT2 and PZT4. The relative average RMS value of the captured response signals was computed and plotted against temperature variation, as shown in Figure 8.

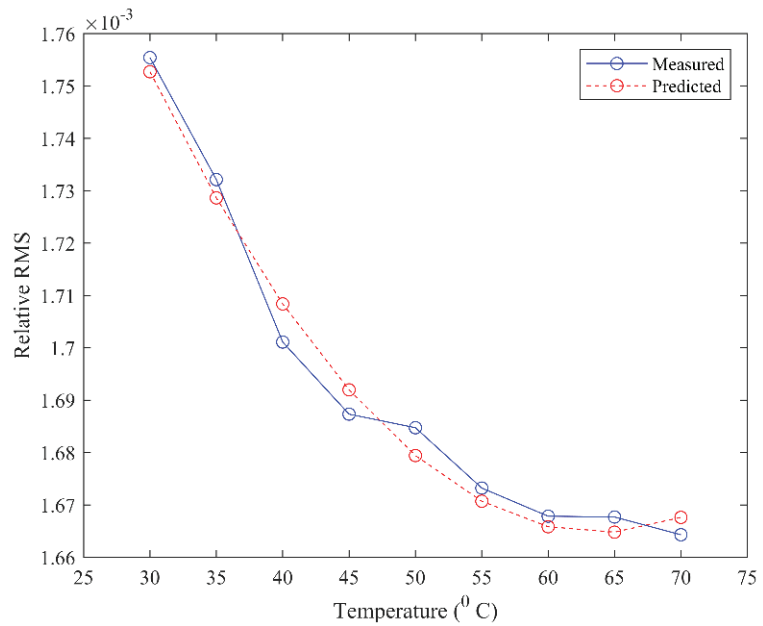


Figure 8. The relative RMS response of the plate under varying temperatures.

Figure 8 shows that, as the temperature of the plate increases, the corresponding computed relative average RMS of the sensors decreases. This could be attributed to an expansion effect of heat on the material and variation of the mechanical properties of the plate as its body temperature increases, especially the elastic modulus that is very sensitive to temperature. The elastic modulus of steel decreases with an increase in temperature, and lamb wave features depend on it for propagation [37]. The thermal differential effect between two response intervals decreases as the temperature increases, leading to a general polynomial response behaviour, as depicted in Figures 8 and 9. Figure 9 is the Power Spectrum of the PZT2 sensor, showing that the signal power decreases as the plate temperature increases. It could be said that guided-wave amplitude is thermally sensitive and decreases due to the superposition of the guided wave with the thermally induced stress wave in the structure. An empirical predictive model for a thermally influenced guided wave is deduced by fitting the measured values. The predictive model is Equation (25) with an RMSE of 5.012×10^{-6} and an R2 of 0.9814 when compared with the measured signal. The empirical predictive model is good since it could explain about 98% of the variation in the measured signal due to temperature increase.

$$f(T) = A_1 T^2 - A_2 T + C \quad (25)$$

where T is the body temperature of the plate,

$$A_1 = 7.687 \times 10^{-8}, A_2 = 9.815 \times 10^{-8} \text{ and } C = 0.001978$$

$$f(T) = 7.687 \times 10^{-8} T^2 - 9.815 \times 10^{-8} T + 0.001978$$

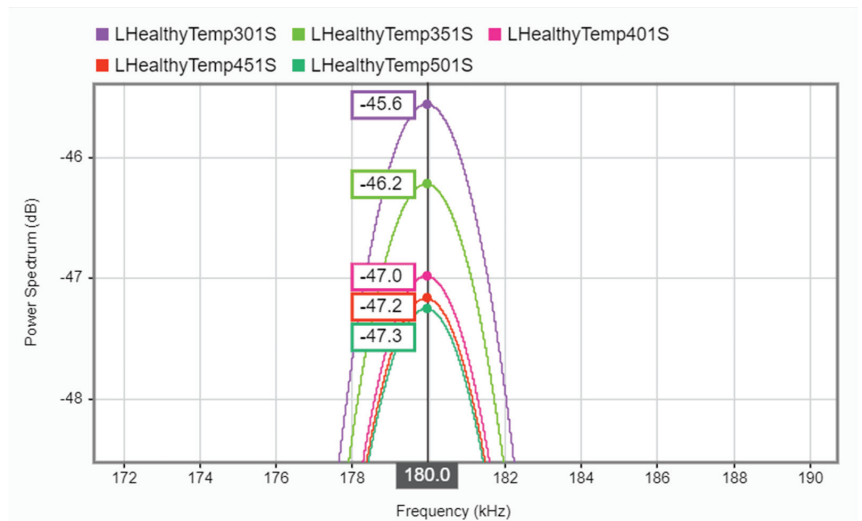


Figure 9. Power spectrum of the captured response signals at different temperatures by PZT2.

4.3. Depth Variation Response

The influence of damage with varying depths on the guided wave was studied. As in Table 3, crack damage of varying depths was machined mid-distance between the PZT1 and PZT3 positions. The distance between the actuator and the sensors is 300.00 mm. The selected inspection frequency is 180 KHz of 29.615 mm wavelength and propagates with a group velocity of 5330.7 m/s². The wavelength of the inspecting wave can detect the damage sufficiently since the damage has the possibility of detection if its length is at least greater than $\frac{1}{2}\lambda$ of the inspecting wave [24]. The response signals captured by PZT2 and PZT4 for each damage depth were analysed for relative average RMS values. From Figure 10, we observed that the average relative RMS value of the response signals increases as the depth of the damage increases. The increase in RMS values implies that the scattering effect of the response signal increases with damage depth. The captured scattered waves are probably from the tips of the damage, as seen in Figure 2b. And, the two sensors were installed to capture scattered and diffracted wavefields from the tips of the damaged. The result suggests that more incident waves are diffracted as the damage depth increases. Although, it is observed that the response value is less than the healthy state value when the damage depth is less than half the thickness of the plate, suggesting a combined effect of attenuation and scattering being more pronounced than wave scattering at the damage tip. It is noted that the scattering effect due to the shallowest depth is about 0.26% of the incident wave, while it is about 0.48% for the deepest depth.

The relationship between the damage depth and the response signal is linear, as in the deduced empirical predictive model in Equation (26), with an R-squared value of 0.9381 and an error value of 1.574×10^{-4} when compared with the measured data.

$$f(d) = B_1 \times d + B_0 \quad (26)$$

where d is the damage-depth value in mm, $B_1 = 7.747 \times 10^{-4}$, and $B_0 = 4.3 \times 10^{-4}$.

It was observed that, when the damage depth is about half the structure thickness or less, the relative RMS of the responses is less than that of the healthy state of the structure. But, in the case of the damage depth being greater than half the thickness of the structure, the relative RMS response signal is greater than that for the healthy state of the structure. Hence, the model could predict damage depth greater than half of the structure's thickness more accurately than damage depth less than half the structural thickness.

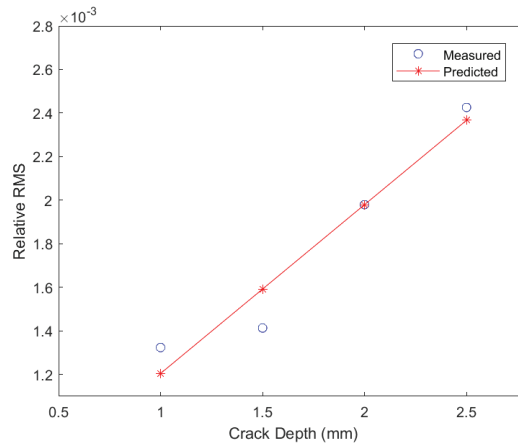


Figure 10. The relative RMS response of the incident wave at varying crack depth.

Recall that the response value of the system in a healthy condition is 2.236×10^{-3} , which is the baseline signal value. Unhealthy responses are noted as the responses captured when damage exists in the plate. The root mean squared deviation (RMSD) between the healthy and unhealthy state of the plate is computed and plotted in Figure 11. It was observed that the percentage of the RMSD value decreases as the damage depth increases. As the guided wave interacts with the damaged area, some of its energy is dissipated through various mechanisms, such as scattering, reflection, and mode conversion. The deeper the damage, the more energy is absorbed or redirected away from the sensor. This is manifested in the decreased value of RMSD as the depth increases. An empirical predictive model for detecting damage depth due to deviation from the baseline signal value is deduced and expressed in Equation (27) with an R-squared value of 0.8866. From Equation (27), it implies that, beyond the damage depth of 2.748 mm, the damage-response deviation becomes negative. This change in value sign could serve as a great deal of alarm for system stoppage to avoid abrupt failure of the system, as it signifies that the structure is in the worst health state when compared with its pristine state of health.

$$f(\dot{d}) = C_1 \times \dot{d} + C_0 \quad (27)$$

where \dot{d} = damage-depth value in mm. $C_1 = -24.44$ and $C_0 = 67.16$

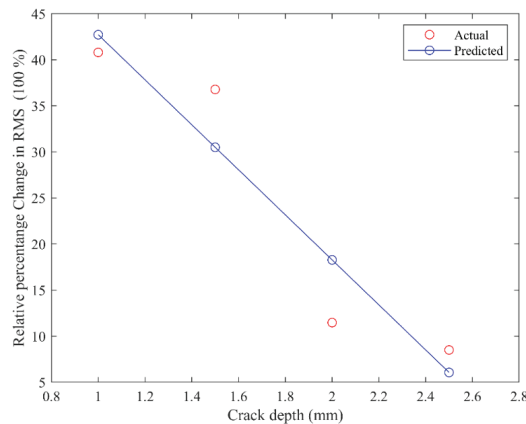


Figure 11. The percentage change in the relative RMS as crack depth increases.

4.4. Temperature Influence on Empty Damage Response of Guided Wave

Two damage depths, 1.50 mm and 2.00 mm were selected. The plate temperature varied from 35 °C to 65 °C at an interval of 10 °C. At each targeted temperature stage, the response signals were captured by PZT2 and PZT4 sensors, and RMS values were processed. Comparing the results in Figures 12 and 13 showed a variation in the RMS signal value as the temperature increased. The RMS value of the response signal decreases as the temperature increases due to the dependence of the wave on the material parameter for propagation, especially on the material elastic modulus. The percentage of RMSD was computed and plotted against the temperature variation and depth, as shown in Figure 14. A linear predictive model of Equation (28) was deduced through curve fitting in Figure 14. An R2 value of 0.8614 and an RMSE value of 1.377 were obtained when the model result was compared with the actual RMSD. This implies that the model could explain more than 86% variation in the response signal caused by temperature and damage depth. Also, the decrease in the RMS value still complies with the trend of elastic material modulus variation, when the temperature of the material is increased. This still suggests the dependence of guided waves on material parameters and responsiveness to its variations.

$$f(T, d) = p_1 * T + p_2 * d + p_0 \quad (28)$$

where d is damage depth, T is the temperature of the material, and p_1 , p_2 and p_0 are -0.2669 , 12.69 and 21.96 , respectively.

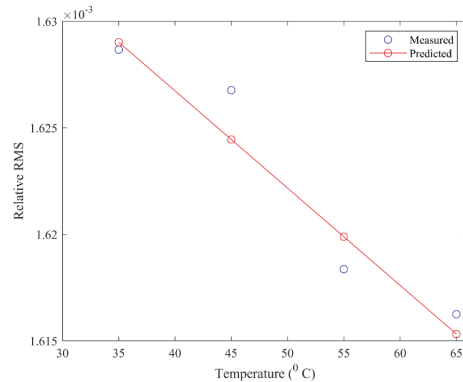


Figure 12. The effect of temperature on 1.5 mm damage depth.

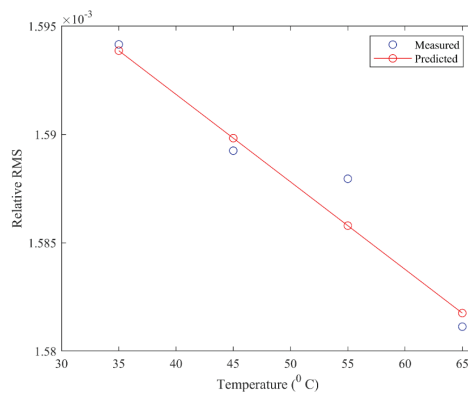


Figure 13. Temperature effect on 2 mm damage depth.

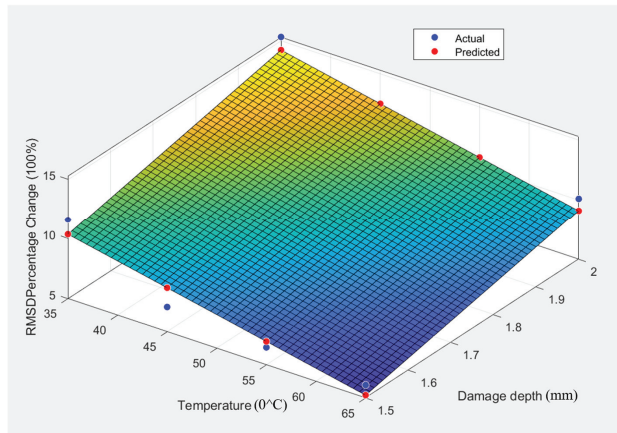


Figure 14. Surface plot of the combined influence of damage depth and temperature.

4.5. Effect of Debris-Filled Damage on the Response Signal

The testing rig remains the same, but the damage is gradually filled with corrosion debris. The debris filled the damage lengthwise at an interval of 20%, starting from 20% to 80% of the damage length. At each increased interval of the debris, the response wave is captured and used to compute the relative RMS value and RMSD. Figure 15 shows that debris-filled damage could cause the response signal to either increase or decrease. When the damage depth is less than or equal to half of the plate thickness, the debris damage index decreases with the increasing debris percentage. But, when the damage depth is greater than half of the plate thickness, the debris damage index increases, as the filled debris percentage increases. This suggests that response wave scattering increases as the damage depth filled with debris increases. Table 4 and Figure 16 compares the DI of a specified empty damage depth with its debris filled to obtain the debris-fill factor. This factor depicts the intensity of debris accumulated by a given damage depth. Hence, as the damage depth increases, the debris factor increases, which suggests more debris accumulation in the damage. From the result, debris in the damage tends to amplify the response wave signals, suggesting that incident wave scattering increases as the debris percentage and the damage depth increase simultaneously.

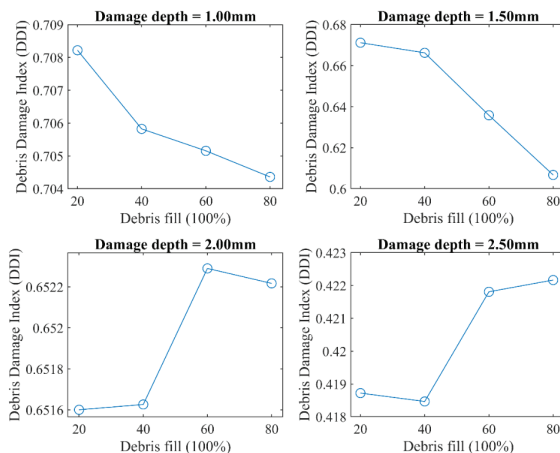
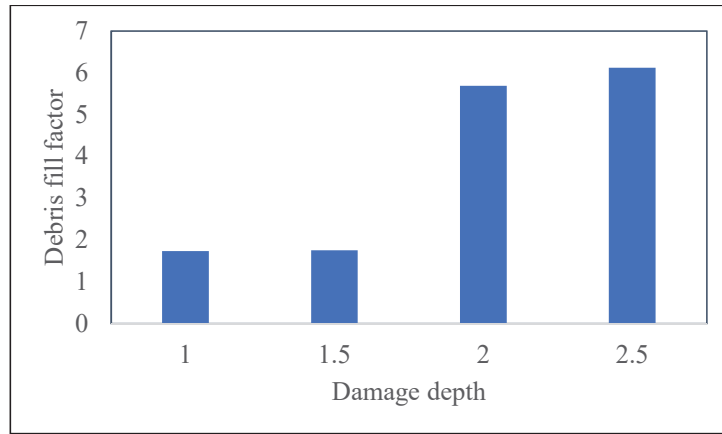


Figure 15. The effect of debris-filled damage on the guided wave.

Table 4. Comparing empty and debris-filled DI.

Crack Depth (mm)	Empty Damage DI	Debris-Filled Damage DI	Debris Factor
1.00	0.40800000	0.7058892961	1.730120824
1.50	0.36780000	0.6449988693	1.753667399
2.00	0.11470000	0.6519333363	5.683812871
2.50	0.08500000	0.5202911002	6.121071767

**Figure 16.** Debris-filled damage-depth factor.

4.6. Influence of Temperature and Debris-Filled Damage on Guided-Wave Response

The testing rig for 1.5 mm damage depth was maintained, but the temperature of the plate was varied gradually from 25 °C to 65 °C to study the collective influence of temperature and debris-filled damage on the response wave signal. From the early study, it was observed that, in the healthy state of the structure, an increase in temperature decreases the RMS of the captured response signals, while an increase in the damage depth increases the response of the captured signal. The effect of temperature increase on the guided wave that had interacted with empty damage is a decrease in the intensity of the measured RMS, as shown in the cases of 1.5 mm and 2.00 mm in Figures 12 and 13, respectively. In furtherance, 1.50 mm crack depth damage was used to understand the effect of temperature on a guided wave that had interacted with debris-filled damage. The crack was filled with different corrosion-debris percentages from 20% to 100% of the damage length. The damage is filled with debris from 20% to 100% at each temperature stage, and PZT2 and PZT4 capture the response signals. The relative average RMS is computed and used to make a surface plot, as shown in Figure 17. An empirical predictive model was deduced through curve fitting, as shown in Equation (29). The model result was compared with the measured response signal, as shown in Figure 18. The goodness of fit, R2 value of 0.7879, and RMSE of 7.5×10^{-5} , derived from the predictive model, shows that it could explain more than 78% of the variation in the measured responses of the structure caused by temperature and debris filled. Also, the relative error in Table 5 shows that a high error is recorded when the temperature is very high, and the ratio of unfilled damage length to the probing wavelength is less than 0.5. This is because, as the debris fills the damage, its length decreases through the closure, consequently decreasing the damage size ratio to probing wavelength. As earlier said, damage stands a high probability of detection if the ratio of its size is greater than or equal to one-half of the probing wavelength [24].

$$f(D, T) = A_1 \times D + A_2 \times T + A_0 \quad (29)$$

where T = temperature, and D = percentage debris. A_1 , A_2 , and A_0 are 4.77×10^{-6} , 1.17×10^{-6} , and 1.252×10^{-3} .

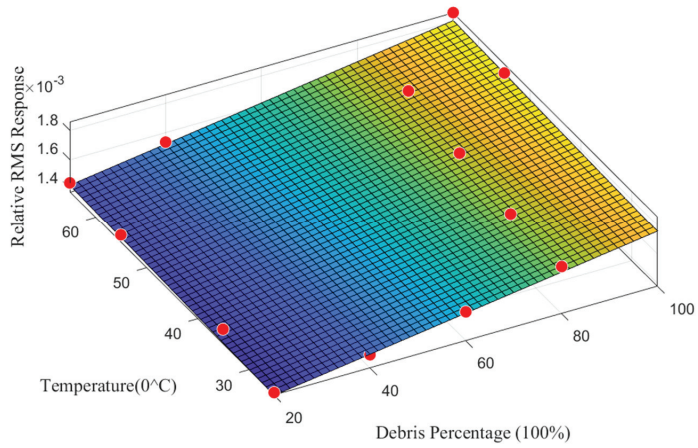


Figure 17. Surface plot of the combined influence of debris-filled damage and temperature.

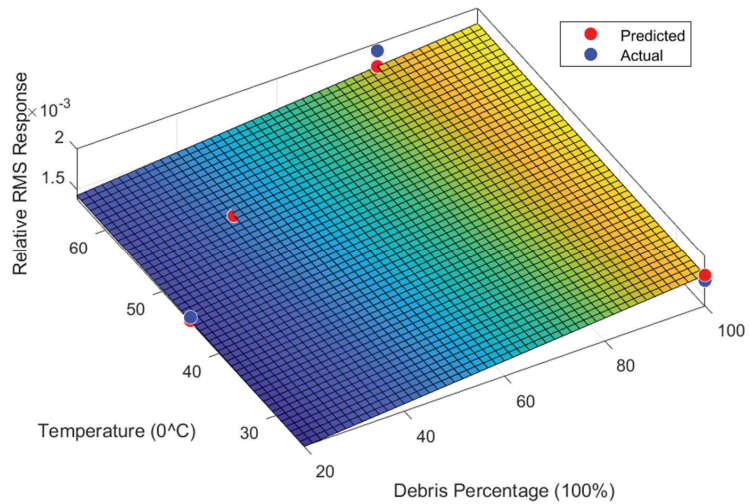


Figure 18. Predictive model response and measured test response.

Table 5. Comparing the predictive model response with the measured sensor response.

Debris	Temp.	Predicted Response	Measured Response	Relative Percentage Error
20	45	0.001400	0.001442	3.01
40	55	0.001507	0.001495	0.78
80	65	0.001709	0.001868	9.31
100	25	0.001758	0.001692	3.75

5. Conclusions

As damage is an inevitable part of structures, continuous monitoring of structural health status using cost-effective technology and processing algorithms of less computational power is needed to avert the catastrophic failure of high-valued structures. This has

sparked a high interest in using guided-wave ultrasonic testing (GWUT) for SHM. GWUT has been used to inspect many forms of damage in structures. However, not much work has been done to account for debris-filled damage, especially under environmental conditions such as temperature. Also, the choice of postprocessing technique for the response-signal feature extraction captured is a crucial aspect of SHM. In this work, three transducers were used to design a pitch-catch configuration topology to study the confluence effect of temperature and debris-filled damage meticulously. By relying on the wave scattering effect from the tip edges of the damage, as in Figure 2, only two sensors were used to capture and process the damage's impact on the propagating signals. Due to the high sensitivity of the propagating wave energy to the damage when compared to other features of the response wave signal, RMS and RMSD were used to analyse the captured response signals. Monitoring the damage-depth increase suggests that more incident waves are diffracted as the damage depth increases. Also, the influence of shallow depth differs from that of the deepest depth as the latter causes more wave energy diffraction than the former. The relationship between the damage depth and the response signal was linear, establishing the empirical predictive model of Equation (25) with an R2 value of 0.9381.

The temperature significantly influenced the response signals, especially on the wave that had interacted with the damage. The major influence is a decrease in the intensity of the measured RMS in the cases of 1.5 mm and 2.00 mm damage depths in Figures 12 and 13, respectively. Also, the combined influence of debris-filled damage and temperature was studied, and a good predictive model was established. The validation of the model with arbitrary values shows a relative error of less than 10% while its R2 value is about 78% and has an RMSE value of about 7.5×10^{-5} . In summary, this study's results are useful for continuously monitoring structures for possible damage detection in an oil and gas facility where debris-filled damage with the influence of temperature is highly possible.

Author Contributions: S.C.O.: Conceptualization, methodology, original draft preparation. M.A.K.: Writing-review and editing, supervision, and project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Petroleum Technology Development Fund (PTDF) of the Federal Government of Nigeria, Grant Number: PTDF/ED/OSS/PHD/SCO/1520/19.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available because it forms part of an ongoing study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Qi, C.; Weixi, Y.; Jun, L.; Heming, G.; Yao, M. A research on fatigue crack growth monitoring based on multi-sensor and data fusion. *Struct. Health Monit.* **2021**, *20*, 848–860. [CrossRef]
2. De La Ree, J.; Liu, Y.; Mili, L.; Phadke, A.G.; Dasilva, L. Catastrophic failures in power systems: Causes, analyses, and countermeasures. *Proc. IEEE* **2005**, *93*, 956–964. [CrossRef]
3. Marcantonio, V.; Monarca, D.; Colantoni, A.; Cecchini, M. Ultrasonic waves for materials evaluation in fatigue, thermal and corrosion damage: A review. *Mech. Syst. Signal Process.* **2019**, *120*, 32–42. [CrossRef]
4. Kralovec, C.; Schagerl, M. Review of Structural Health Monitoring Methods Regarding a Multi-Sensor Approach for Damage Assessment of Metal and Composite Structures. *Sensors* **2020**, *20*, 826. [CrossRef] [PubMed]
5. Olisa, S.C.; Starr, A.; Khan, M.A. Monitoring evolution of debris-filled damage using pre-modulated wave and guided wave ultrasonic testing (GWUT). *Measurement* **2022**, *199*, 111558. [CrossRef]
6. Khan, M.A.; Cooper, D.; Starr, A. BS-ISO helical gear fatigue life estimation and wear quantitative feature analysis. *Strain* **2009**, *45*, 358–363. [CrossRef]
7. Liu, Y.; Feng, X. Monitoring corrosion-induced thickness loss of stainless steel plates using the electromechanical impedance technique. *Meas. Sci. Technol.* **2020**, *32*, 025104. [CrossRef]
8. Guan, R.; Lu, Y.; Wang, K.; Su, Z. Fatigue crack detection in pipes with multiple mode nonlinear guided waves. *Struct. Health Monit.* **2019**, *18*, 180–192. [CrossRef]
9. Giurgiutiu, V.; Zagrai, A.; Bao, J. Damage identification in aging aircraft structures with piezoelectric wafer active sensors. *J. Intell. Mater. Syst. Struct.* **2004**, *15*, 673–687. [CrossRef]

10. Lambinet, F.; Khodaei, Z.S. Damage detection & localisation on composite patch repair under different environmental effects. *Eng. Res. Express* **2020**, *2*, 045032. [CrossRef]
11. Fleet, T.; Kamei, K.; He, F.; Khan, M.A.; Khan, K.A.; Starr, A. A machine learning approach to model interdependencies between dynamic response and crack propagation. *Sensors* **2020**, *20*, 6847. [CrossRef] [PubMed]
12. Kannusamy, M.; Kapuria, S.; Sasmal, S. Accurate baseline-free damage localisation in plates using refined Lamb wave time-reversal method. *Smart Mater. Struct.* **2020**, *29*, 055044. [CrossRef]
13. Azuara, G.; Barrera, E. Influence and compensation of temperature effects for damage detection and localisation in aerospace composites. *Sensors* **2020**, *20*, 4153. [CrossRef] [PubMed]
14. Wei, D.; Liu, X.; Wang, B.; Tang, Z.; Bo, L. Damage quantification of aluminum plates using SC-DTW method based on Lamb waves. *Meas. Sci. Technol.* **2022**, *33*, 045001. [CrossRef]
15. Olisa, S.C.; Khan, M.A.; Starr, A. Review of current guided wave ultrasonic testing (GWUT) limitations and future directions. *Sensors* **2021**, *21*, 811. [CrossRef]
16. Huang, X.; Wang, P.; Zhang, S.; Zhao, X.; Zhang, Y. Structural health monitoring and material safety with multispectral technique: A review. *J. Saf. Sci. Resil.* **2022**, *3*, 48–60. [CrossRef]
17. Kudela, P.; Radzienski, M.; Ostachowicz, W.; Yang, Z. Structural Health Monitoring system based on a concept of Lamb wave focusing by the piezoelectric array. *Mech. Syst. Signal Process.* **2018**, *108*, 21–32. [CrossRef]
18. Samaitis, V.; Jasiūnienė, E.; Packo, P.; Smagulova, D. Ultrasonic Methods. In *Springer Aerospace Technology*; Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2021; pp. 87–131. Available online: https://link.springer.com/chapter/10.1007/978-3-030-72192-3_5#citeas (accessed on 4 January 2023).
19. Yan, X.; Qiu, L.; Yuan, S. Experimental study of guided waves propagation characteristics under the changing temperatures. *Vibroeng. Procedia* **2018**, *20*, 208–212. [CrossRef]
20. Zai, B.A.; Khan, M.A.; Khan, K.A.; Mansoor, A. A novel approach for damage quantification using the dynamic response of a metallic beam under thermo-mechanical loads. *J. Sound Vib.* **2020**, *469*, 115134. [CrossRef]
21. Giurgiutiu, V. *Structural Health Monitoring with Piezoelectric Wafer Active Sensors*; Elsevier: Amsterdam, The Netherlands, 2006. [CrossRef]
22. Šofer, M.; Ferfecki, P.; Šofer, P. Numerical solution of Rayleigh-Lamb frequency equation for real, imaginary and complex wavenumbers. *MATEC Web Conf.* **2018**, *157*, 08011. [CrossRef]
23. Center for Lightweight-Production-Technology. The Dispersion Calculator: An Open Source Software for Calculating Dispersion Curves and Mode Shapes of Guided Waves. Available online: https://www.dlr.de/zlp/en/desktopdefault.aspx/tabid-14332/24874_read-61142/ (accessed on 4 January 2023).
24. Nondestructive Evaluation Physics: Waves. Available online: <https://www.nde-ed.org/Physics/Waves/defectdetect.shtml> (accessed on 5 January 2023).
25. Rojas, O.E.; Khan, M.A. A review on electrical and mechanical performance parameters in lithium-ion battery packs. *J. Clean. Prod.* **2022**, *378*, 134381. [CrossRef]
26. Lu, Y.; Lu, M.; Ye, L.; Wang, D.; Zhou, L.; Su, Z. Lamb wave based monitoring of fatigue crack growth using principal component analysis. *Key Eng. Mater.* **2013**, *558*, 260–267. [CrossRef]
27. Rizvi, S.; Khan, S.; Tariq, M.; Khan, M.A. Propagation of Gaussian-Modulated Lamb Wave in Healthy and Damaged Plate for Structural Health Monitoring. *Am. Soc. Nondestruct. Test.* **2017**, 24–30.
28. Yu, T.H. Plate waves scattering analysis and active damage detection. *Sensors* **2021**, *21*, 5458. [CrossRef] [PubMed]
29. Gorgin, R.; Luo, Y.; Wu, Z. Environmental and operational conditions effects on Lamb wave based structural health monitoring systems: A review. *Ultrasonics* **2020**, *105*, 106114. [CrossRef] [PubMed]
30. Salmanpour, M.; Khodaei, Z.S.; Aliabadi, M.H. Damage detection with ultrasonic guided wave under operational conditions. In Proceedings of the 9th European Workshop on Structural Health Monitoring, EWSHM 2018, Manchester, UK, 10–13 July 2018.
31. IPfeiffer, F.; Wriggers, I.P. *Identification of Damage Using Lamb Waves*; Springer: Berlin/Heidelberg, Germany.
32. Amazon UK. DAS Modelling Air Dry Clay White. Available online: https://www.amazon.co.uk/DAS-Modelling-Air-Clay-White/dp/B003P8RRU0/ref=sr_1_13?crid=3MN56983W60HG&keywords=paper+clay&qid=1643611314&srefix=paper+clay,aps,63&sr=8-13 (accessed on 31 January 2023).
33. RS. Loctite HYSOL 3421 Transparent Yellow 50 mL Epoxy Adhesive Dual Cartridge for Various Materials. Available online: <https://uk.rs-online.com/web/p/resins/4589478> (accessed on 18 January 2023).
34. Garnier, J.; Papanicolaou, G. Travel time estimation by cross correlation of noisy signals. *Esaim Proc.* **2009**, *27*, 122–137. [CrossRef]
35. Gibbs, G.; Jia, H.; Madani, I. Obstacle Detection with Ultrasonic Sensors and Signal Analysis Metrics. *Transp. Res. Procedia* **2017**, *28*, 173–182. [CrossRef]
36. Konstantinidis, G.; Drinkwater, B.W.; Wilcox, P.D. The temperature stability of guided wave structural health monitoring systems. *Smart Mater. Struct.* **2006**, *15*, 967–976. [CrossRef]
37. Kim, B.G.; Remppe, J.L.; Knudson, D.L.; Condie, K.G.; Sencer, B.H. In-situ creep testing capability for the advanced test reactor. *Nucl. Technol.* **2012**, *179*, 417–428. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Machine Learning Algorithms That Emulate Controllers Based on Particle Swarm Optimization—An Application to a Photobioreactor for Algal Growth

Viorel Minzu ^{1,*}, Iulian Arama ² and Eugen Rusu ³¹ Control and Electrical Engineering Department, “Dunarea de Jos” University, 800008 Galati, Romania² Informatics Department, “Danubius” University, 800654 Galati, Romania; iulian.arama@univ-danubius.ro³ Mechanical Engineering Department, “Dunarea de Jos” University, 800008 Galati, Romania; eugen.rusu@ugal.ro

* Correspondence: viorel.minzu@ugal.ro

Abstract: Particle Swarm Optimization (PSO) algorithms within control structures are a realistic approach; their task is often to predict the optimal control values working with a process model (PM). Owing to numerous numerical integrations of the PM, there is a big computational effort that leads to a large controller execution time. The main motivation of this work is to decrease the computational effort and, consequently, the controller execution time. This paper proposes to replace the PSO predictor with a machine learning model that has “learned” the quasi-optimal behavior of the couple (PSO and PM); the training data are obtained through closed-loop simulations over the control horizon. The new controller should preserve the process’s quasi-optimal control. In identical conditions, the process evolutions must also be quasi-optimal. The multiple linear regression and the regression neural networks were considered the predicting models. This paper first proposes algorithms for collecting and aggregating data sets for the learning process. Algorithms for constructing the machine learning models and implementing the controllers and closed-loop simulations are also proposed. The simulations prove that the two machine learning predictors have learned the PSO predictor’s behavior, such that the process evolves almost identically. The resulting controllers’ execution time have decreased hundreds of times while keeping their optimality; the performance index has even slightly increased.

Keywords: particle swarm optimization; machine learning; optimal control; simulation

Citation: Minzu, V.; Arama, I.; Rusu, E. Machine Learning Algorithms That Emulate Controllers Based on Particle Swarm Optimization—An Application to a Photobioreactor for Algal Growth. *Processes* **2024**, *12*, 991. <https://doi.org/10.3390/pr12050991>

Academic Editor: Xiong Luo

Received: 21 April 2024

Revised: 8 May 2024

Accepted: 10 May 2024

Published: 13 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A common task in process engineering is to control processes whose quality is evaluated through a performance index. When the process model has certain mathematical properties, theoretical control laws can be adopted for implementation; on the contrary, when the process model is uncertain, incomplete, and imprecise or has profound nonlinearities, metaheuristic algorithms (MAs) like Evolutionary Algorithms, Particle Swarm Optimization, etc., within a suitable control structure could be successfully used [1–3]. Control engineering has afforded numerous examples where metaheuristics were used [4–8] owing to their robustness and capacity to deal with complex processes.

The role of an MA within a controller is usually to predict the optimal control values within each sampling period, but first, it searches for the optimal value. For example, the PSO algorithm follows its optimization mechanism using particles and the internal PM.

A control structure fitting this type of controller is Receding Horizon Control (RHC) [9–11]. This structure is suitable for implementing solutions to Optimal Control Problems (OCs); it includes an internal process model (PM) [8,12,13].

Another facet of the prediction process also used within a control problem is described, for example, in [14]. The study proposes a physics-assisted transfer learning metamodeling

framework to predict laser butt welding bead geometry and carbon emissions. This time, the learning process updates the process model.

A predilect research topic for the authors was implementing the prediction module within an RHC structure employing MAs. The results are partially reflected in previous works [8,15,16]. The robustness, efficiency, and usability of MAs inside a controller have a price to pay: the controller's execution time. The optimization mechanism and the PM's numerous numerical integrations take a relatively long time to find the optimal value after a convergence process. That is why this approach is mainly suitable for slow processes when the predictions' computation time is smaller than the sampling period [17,18]. Decreasing the predictor's execution time is a challenge [15,16] because it could extend the applicability of controllers using MAs. This work goes in the same direction but involves a new technique: using machine learning (ML) to emulate predictors based on MAs. Recently, we have proposed linear regression (LR) predictors that are "equivalent" in a certain sense to predictors based on Evolutionary Algorithms (Eas) [19]. This paper deals with OCPs having final costs and solutions involving PSO predictions.

For the reader, who is a newcomer equally in the fields of control systems, computational intelligence, and machine learning, we have to answer why our objective is to replace the PSO predictor with an ML predictor (inside the RHC structure) when solving an OCP.

- The PSO predictor predicts an optimal control value, but first, it searches for the optimal control sequence following its optimization mechanism using a swarm of particles and the PM. That is why it takes a relatively long time to find this value after a convergence process.
- The ML predictor (LR or RNN) predicts using an already-known regression function. Being an ML model, it reproduces what it has learned, the PSO predictor's behavior. It does not search for anything. Moreover, it does not make numerical integrations of the PM. That is why it takes a much shorter time to calculate the predicted value.
- The ML predictor replaces the PSO predictor only in execution when the controller achieves the control action. Intrinsicly, the solution is given by the PSO algorithm. The solutions are "learned" by the ML predictor; that is, the ML model emulates the PSO algorithm.

To continue the work presented in [19], we shall consider the equivalence mentioned above and implement Regression Neural Network (RNN) predictors besides the LR ones.

Throughout this paper, we have recourse to a specific OCP to make the explanations easy to follow for the reader. In [11], for the optimal control of a specific photobioreactor (PBR) lighted for algal growth, we have presented a solution in the same context, RHC and predictions based on PSO. We shall adopt the PSO predictor already constructed in [11]; by employing this one, we shall generate new ML predictors. The data generated by simulation modules, already developed previously, are stored or recorded. These data will be needed to train and test the ML models. Some results from [11] will be reported in this paper for comparison (Section 7.1).

Section 2 recalls the general approach developed in previous work [8,20,21] to solve such problems using PSO algorithms. Besides the recall of the PBR problem's statement, Section 2 also introduces the notations and formulas that keep the presentation self-contained.

Section 3 answers the following two main questions:

- What data are needed to capture the optimal (quasi-optimal) behavior of the couple (PSO and PM)?
- How are the data sets for the ML algorithms generated?

The PSO prediction module, included in the controller, is available from our previous work, which has already solved the PBR problem. Section 3 presents an algorithm carrying out the closed-loop simulation over the control horizon using the PSO predictor. A mandatory hypothesis is that the real process and the internal PM are considered identical because the data recorded should capture only the behavior of the couple (PSO and PM).

At the end of the simulation, the sequence of optimal control values (optimal control profile) and the sequence of state variables (optimal trajectory) can be recorded. The optimal CP and trajectory can be seen as the “signature” data of the optimal solution. Considering together the two sequences, we obtain a sequence of couples (state; control value), one couple for each sampling period. The simulation program is executed M times (e.g., two hundred times); the two sequences are collected each time and aggregated into a data structure. This data structure expresses the PSO predictor’s experience as a decision maker; it will be used to obtain the ML models [22–26].

Section 4 presents the general approach to learning the optimal behavior of the couple (PSO and MP). The learning process is split at the level of each sampling period, and consequently, new data structures are derived for each of them. With each new data structure, which is a collection of couples (state; control value), a generic regression function [22,27,28] is associated. The latter is materialized through an ML regression model devoted to the sampling period at hand, which must be capable of giving accurate predictions.

An ML controller’s systematic design procedure is also proposed. We have to emphasize that the entire design procedure of the ML controller needs only simulations and offline program executions.

In this paper, we consider as regression models only two kinds of models: multiple linear regression [29–31] and Regression Neural Network [22,32]. Other regression models (trees, support vector machines, and Gaussian processes) were considered in our studies as well. Still, only the models LR and RNN are relevant to this presentation. Implementing an ML controller, in our context, involves determining a regression model for each sampling period.

Section 5 deals with constructing a set of linear regression (LR) models [29,31] that are trained with the data sets constructed in Section 4. A general construction algorithm using the stepwise regression [30] strategy is proposed. A table with the regressions’ coefficients is extracted from the models. The LR controller is implemented using the LR predictors; it is also integrated into a proposed closed-loop simulation program, allowing us to evaluate the entire approach. Some simulation results are given.

Section 6 proposes a general algorithm for constructing the models using Regression Neural Networks [32]. The training and testing data sets are already determined in Section 4, and they are saved in an external file. A specific closed-loop simulation program is also proposed; it includes the RNN controller using the RNN predictors. The simulation results are presented for further analysis.

The Discussion section first answers the following question: Did the two kinds of predictors “learn” the behavior of the couple (PSO and PM), such that the new process’s evolutions would also be quasi-optimal? To do this, we depict the new process’s evolutions and display some numerical information using the closed-loop simulation programs proposed in Sections 5 and 6. The simulation results are compared with those already available concerning the PSO predictor. Owing to their generalization ability, both ML controllers make accurate predictions of the control value sent to the process, and the state evolutions are practically identical.

The second question that this section has to answer is as follows: Did the controller’s execution time decrease significantly?

The positive answer to both questions proves that the ML controllers are an effective way to avoid the large execution time of the controllers based on PSO while maintaining the optimality of the control. This result is important because it extends the possibility of using PSO (or other MA) controllers for processes with smaller time constants.

Special attention was paid to the implementation aspects such that the interested reader could find support to understand and, eventually, reproduce parts of this work or use it in their projects. With this aim in view, all algorithms used in this presentation are implemented, the associated scripts are attached in the Supplementary Materials, and all of the necessary details are given in the Appendices A–E.

2. Controllers with Predictions Based on PSO: Connection with Machine Learning Algorithms

Many controlled processes, such as biochemical processes, are repetitive, like those organized in batches. For efficiency reasons, they generate Optimal Control Problems involving three components:

- The process model can include nonlinearities, imprecise, incomplete, and uncertain knowledge, correspond to a distributed-parameter system, etc.
- There are constraints, such as initial conditions, bound constraints, final constraints, etc.
- The cost function, which should be optimized, leads to a performance index.

To solve such a control problem, we need an adequate control structure which will define the optimal controller. The latter includes a prediction module that calculates the optimal control sequence and the optimal trajectory over the prediction horizon or even until the end of the control horizon. For its work, the predictor uses the PM for a huge number of numerical integrations. In this context, the predictor has a very complex numerical task; that is why a metaheuristic algorithm is often a realistic solution to fulfil this task.

The Receding Horizon Control (RHC) [8,10,11] is a very simple control strategy that can easily integrate a metaheuristic algorithm as a predictor (Figure 1). The authors have studied and simulated the RHC in solving different OCPs in conjunction with an EA or PSO. The solutions are realistic, they can be used in real-time control, and several techniques can be used to decrease the numerical complexity of the predictor. Nevertheless, the inconvenience is that the control action takes up a big part of the sampling period.

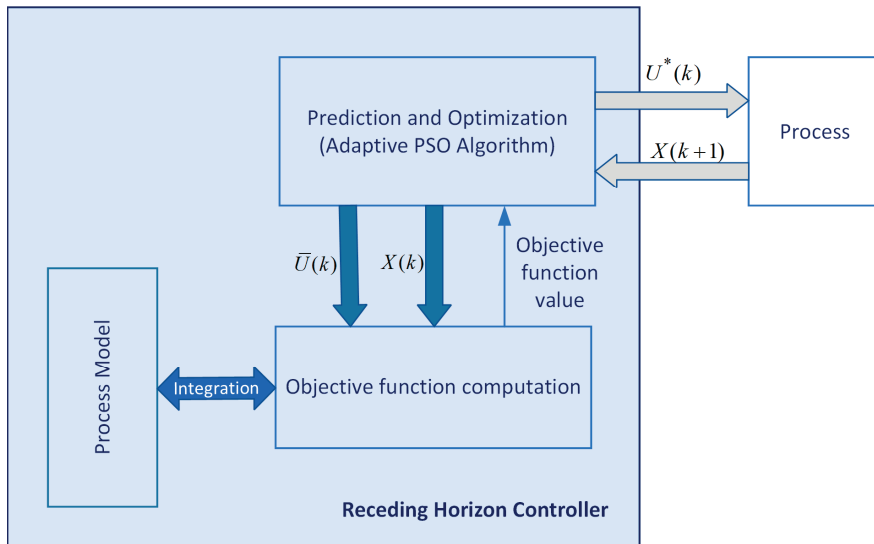


Figure 1. Receding Horizon Control using adaptive PSO algorithm.

An interesting and practical approach, in this context, is to replace the predictor with a machine learning algorithm inside the controller. The ML algorithm must emulate the predictor of the RHC structure following its training offline. To emphasize its role, we shall refer to this algorithm as the *ML controller*. In this work, we have to answer the following questions:

- What does it mean that the ML algorithm must emulate the predictor?
- What data sets are used in training the ML, and how are they obtained?
- What kind of ML model can be used to achieve an appropriate controller?

In [11], we have presented the optimal control of a continuously stirred flat-plate photobioreactor (PBR) lighted on a single side for algal growth in the same context: the RHC that uses predictions based on PSO.

In this presentation, all of the essential tasks concerning the ML do not need the PM; only the final simulations, which allow us to validate the entire approach, employ the PM. That is why the reader can find in Appendix A the equations modeling the PBR, the constraints, and the cost function of the OCP. The PBR is a distributed parameter system, but the PM is converted through discretization into a lumped parameter process. We have solved this problem in [11], adding the discretization constraint, which refers to the input variables:

$$U(t) = U(kT) \triangleq U(k), \text{ for } k \cdot T \leq t < (k + 1) \cdot T; k = 0, \dots, H - 1.$$

where T is the sampling period, and the final time of the batch equals $H \cdot T$. In our example, the input vector has a single component, i.e.,

$$U(k) = q(k), k = 0, \dots, H - 1. \tag{1}$$

The variable $q(k)$ is the intensity of incident light throughout the k th sampling period.

At every moment $0 \leq k < H - 1$, the predictor calculates the optimal control sequence (1) using the usual version of the APSOA (adaptive PSO algorithm) and the PM, which is integrated a large number of times. The optimal control sequence minimizes the cost function $J(k, X(k))$ over the current prediction horizon $[k, H]$; in our example, the following holds:

$$J(k, X(k)) = \min_{\text{predicted sequence}} \left\{ w_1 \cdot A \cdot C \sum_{i=k}^{H-1} U(i) + w_2 \cdot [V \cdot x_1(H) - m_0] \right\},$$

$$X(k) = [x_1(k) \ x_2(k)].$$

The vector $X(k)$ is the current state of the process. A predicted sequence is a control sequence with the following structure:

$$\bar{U}(k) = \langle U(k), \dots, U(H - 1) \rangle, \tag{2}$$

Using the PM and Equation (2), the APSOA calculates the corresponding state sequence.

$$\bar{X}(k) = \langle X(k), \dots, X(H) \rangle.$$

The latter has $H - k + 1$ elements.

When the APSOA converges, it supplies the best sequence $\bar{U}(k)$ for the current prediction horizon, denoted by $\bar{V}(k)$:

$$\bar{V}(k) \triangleq \arg \min_{\bar{U}(k)} J(\bar{U}(k), X(k)) = \langle V(k), \dots, V(H - 1) \rangle \tag{3}$$

The controller's best output, denoted by $U^*(k)$, is the first value of this sequence, i.e.,

$$U^*(k) \triangleq V(k). \tag{4}$$

Applying Equations (3) and (4) is, in fact, the control strategy "Receding Horizon Control".

A sequence of H control vectors, $U(0), U(1), \dots, U(H - 1)$, will be referred to as a "control profile" (CP). The latter will produce a state transition like in Figure 2.

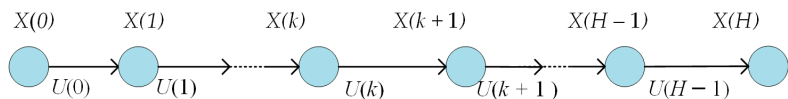


Figure 2. The state trajectory and its CP.

Finally, the controller following the RHC strategy using predictions implemented by a PSO algorithm achieves the optimal CP for the given initial state and control hori-

zon. The optimal CP denoted by $\Omega(X_0)$, which represents our problem's solution, is the concatenation of the optimal controls $U^*(k)$, $k = 0, \dots, H - 1$:

$$\Omega(X_0) \triangleq \langle U^*(0), U^*(1), \dots, U^*(H - 1) \rangle \quad (5)$$

Forced by this CP, the process will follow an "optimal trajectory" $\Gamma(X_0)$:

$$\Gamma(X_0) \triangleq \langle X_0, X^*(1), \dots, X^*(H) \rangle. \quad (6)$$

A closed-loop simulation is the simulation of the controller, which includes the APSOA and PM, connected to the (real) process (see Figure 1). Our study requires only the situation when the real process and the PM are identical.

Remark 1. *The two sequences (5) and (6) can fully characterize the process's optimal behavior in the context of closed-loop simulation over the control horizon when the process and its model are identical.*

Supposing the convergence of the APSOA, the value $J_0 = J(0, X(0))$ theoretically equals the optimal cost function. Practically, at the end of a closed-loop simulation, the two values will be very close to each other, so $\Omega(X_0)$ would be a quasi-optimal solution for the problem at hand.

The predictor's behavior depends on two factors: the metaheuristic algorithm (APSOA) and the PM. In this work, the main objective is to capture the predictor's behavior through an ML algorithm. Hence, the latter has to "learn" the optimal behavior of the couple (APSOA and PM).

Remark 2. *Our purpose is to capture the predictor's behavior using an ML algorithm, that is, to "learn" the optimal behavior of the couple (APSOA and PM). The final objective is to replace the predictor with the new ML algorithm, such that the process's state evolution and the performance index would be maintained. In this situation, we can state that the ML algorithm emulates the predictor.*

The two sequences ($\Omega(X_0)$ and $\Gamma(X_0)$) are the data results of a closed-loop simulation and can be considered the "signature" data of the couple (APSOA and PM); there is a correspondence between the values $X^*(k)$ and $U^*(k)$ signifying that "when the process is in the state $X^*(k)$ at the moment k , the APSOA will predict the best control value $U^*(k)$ ".

So, the source of the data used in a potential learning process can be a closed-loop simulation, considering the (real) process and the PM identical. Of course, the data produced by a single simulation over the entire control horizon cannot be sufficient for the learning process.

3. Data Generation Using Closed-Loop Simulation over Control Horizon

For any OCP like the PBR problem, the designed controller must be validated by closed-loop simulation, considering the (real) process and the PM identical. This validation must be carried out before using the implemented controller in real time, connected to the (real) process. Hence, we must have a simulation program that fulfils this task of closed-loop simulating over the control horizon, with a given initial state, and considering the process and the PM identical.

A simulation can be carried out in more realistic situations, for example, when the process takes into consideration, besides the PM, unmodeled dynamics and noises. But we do not need such simulations.

Remark 3. *The fact that the process and the PM are identical is not a simplification to render our study's conclusion more favorable but is a necessity. The ML algorithm has to learn the behavior of the couple (APSOA and PM); otherwise, it will "learn", besides the APSOA and PM, the influence of other perturbing factors.*

Figure 3 shows the closed-loop simulation program’s flowchart in the conditions mentioned above. This program is generically called “ContrLoop_PSO”. The function “Predictor_PSO” returns the predicted sequence $\bar{V}(k)$, whose first element will give the optimal control value $U^*(k)$. Sending the latter value to the PM and integrating the process over a sampling time, the function “ProcessStep” will determine the process’s next state, that is, at the next moment, $k + 1$.

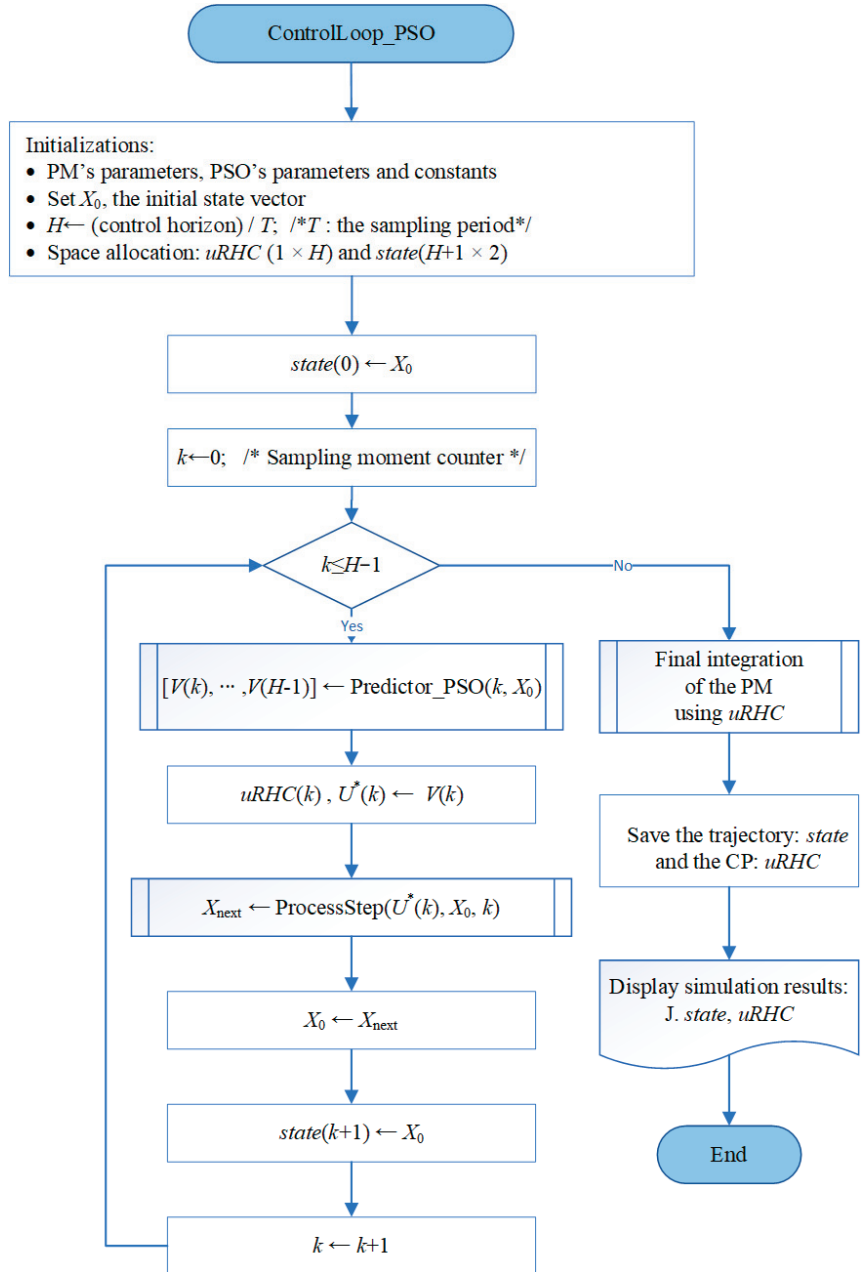


Figure 3. Closed-loop simulation using predictions based on adaptive PSO algorithm.

When the controller designer decides to use an ML algorithm to replace the couple (APSOA and PM), the functions “Predictor_PSO” and “ProcessStep” are already written as a part of the PSO controller’s construction. This is also the case with the PBR problem; we have already accomplished the entire design procedure (for more details, see [11]).

The reader can understand and execute the “ContrlLoop_PSO” program using the script `ControlLoop_PSO_RHC.m`. Details are also given in Appendix B, concerning the generic functions “Predictor_PSO” and “ProcessStep”.

As we already mentioned, after the closed-loop simulation, the data generated are a couple of sequences ($\Omega(X_0)$ and $\Gamma(X_0)$), which can be renamed (control profile—trajectory). To prepare the data for the ML process, we shall repeat M times (e.g., $M = 200$) the closed-loop simulation and produce M different quasi-optimal couples (CP—trajectory). There are two reasons why data couples are different:

- The PSO has a stochastic character, and the convergence process is imperfect. So, the optimal control values are different (and so are the state vector’s values), even if the initial state is strictly the same.
- The initial state values are not the same. A standard initial state (of the standard batch) could be perturbed to simulate different initial conditions (the standard ones are imprecisely achieved).

The optimal control value and the optimal states are stored in the matrices $uRHC$ ($H \times m$) and $state$ ($H + 1 \times n$), respectively, with their structure presented in Figure 4.

$state (H+1 \times n)$	$uRHC (H \times m)$
$(X(0))^T$	$(U^*(0))^T$
$(X^*(1))^T$	$(U^*(1))^T$
\vdots	\vdots
$(X^*(H-1))^T$	$(U^*(H-1))^T$
$(X^*(H))^T$	

Figure 4. The matrices that store the quasi-optimal trajectory and its CP.

Hence, the optimal CP and trajectory are described by the matrices $uRHC$ and $state$, respectively, which are the images of $\Omega(X_0)$ and $\Gamma(X_0)$ sequences (see (5) and (6)). For each of the M simulations, the two matrices are saved in the cell array `STATE` and the matrix `UstarRHC` ($M \times H$), as suggested in Figure A1 (Appendix B) for our case study.

The script `L00P_M_ControlLoop_PSO.m` collects the data from M executions of the closed-loop simulation. The data structures presented in Figure A1 are created and loaded. A concrete example of the data collected in the first simulation is given in Appendix B.

4. The ML Controller: The Design Procedure and the General Algorithm

The M simulations can be collectively presented in Figure 5, where the state variables and control output are regrouped by sampling periods.

At each step k , $0 \leq k \leq H - 1$, of the control horizon, the controller predicts the optimal control output relying on the couple (APSOA and PM). The state vectors considered at the same step have some common characteristics:

- The same PSO algorithm generates the M state inside a group.
- The M simulations work with the same PM.
- Each state $X_i^*(k)$, $1 \leq i \leq M$, is transferred as the initial state to the predictor.
- The prediction horizon has $H - k$ sampling periods.

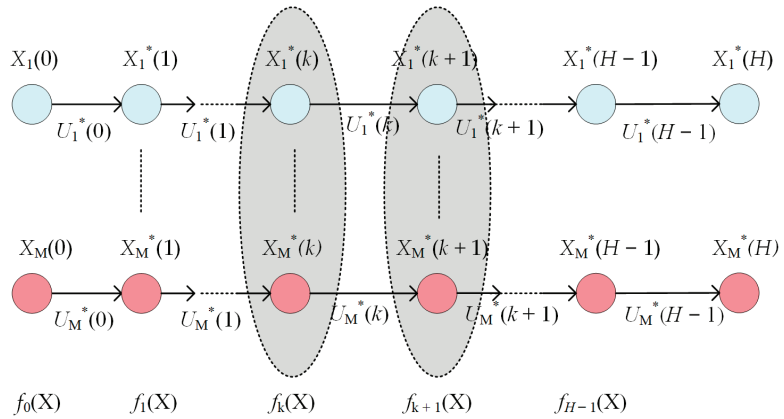


Figure 5. The quasi-optimal trajectories produced by M executions of “ControlLoop_PSO.”.

The APSOA calculates the prediction $\bar{V}(k)$, and the controller extracts only the optimal control values $U_i^*(k)$, $1 \leq i \leq M$.

The simulation results for step k can be organized as a data set, and a table can be constructed, as shown below.

We have considered, as usual, that the state and control vector are column vectors. In our case study, the state vector ($n = 2$) is generated like a line vector to avoid transposition.

When M has a big enough value, the data set from Table 1 represents to some extent the ability of the couple (APSOA and PM) to predict optimal control values at step k . Our desideratum is to generalize this ability to predict the optimal control when the process accesses states other than those from Table 1; this can be achieved using a machine learning algorithm.

Table 1. Data set for step k .

X^T	U^T
$(X_1^*(k))^T$	$(U_1^*(k))^T$
.....
$(X_M^*(k))^T$	$(U_M^*(k))^T$

Remark 4. The four characteristics enumerated before are the reasons making us adopt the hypothesis that the examples (data points) of Table 1 belong to the same data-generating process; that is, they are independently identically distributed.

With each group of states presented in Figure 5, equivalent to a table like Table 1, a regression function $f_k(X)$ can be associated:

$$f_k : \mathbf{R}^n \rightarrow \mathbf{R}^m, k = 0, 1, \dots, H - 1.$$

When these functions exist, they can be used successively within the controller to replace the predictor at each control step.

Remark 5. The regression function f_k models how the APSOA determines the optimal prediction at step k . The entire set of functions $\Phi = \{f_k | k = 0, 1, \dots, H - 1\}$ is the couple (APSOA–PM) machine learning model. The behavior of the PSO algorithm, which, in turn, depends on the PM, is captured by the set of functions Φ .

To be systematic, at this point of our presentation, we propose a design procedure for the ML controller that the interested reader could use in their implementation.

Design Procedure

1. Write the "ControlLoop_PSO" program simulating the closed-loop functioning of the controller based on the PSO algorithm over the control horizon. The output data are the quasi-optimal trajectory and its associated control profile ($\Omega(X_0)$ and $\Gamma(X_0)$).
2. Repeat M times the "ControlLoop_PSO" program's execution to produce the sequences ($\Omega(X_0)$ and $\Gamma(X_0)$) and save them in data structures similar to those in Figure A1 (Appendix B).
3. For each sampling period k , derive data sets similar to those in Table 1 from the data saved in step 2.
4. Determine the set of functions Φ using the data sets derived in step 3 and an ML model; a function f_k is associated with each sampling period k .
5. Implement the new controller based on the ML model, i.e., the set of functions Φ determined in step 4.
6. Write the "CONTROL_loop" program to simulate the closed-loop functioning equipped with the ML controller. The proposed method's feasibility, performance index, solution quality, and execution time will be evaluated.

Remark 6. *The entire design procedure of the ML controller needs only simulations and offline program executions. The ML models for each sampling period are determined offline ahead of using the ML controller in real time.*

Steps 1–2 are already covered by the details given in the anterior section.

Step 3 Implementation

This step yields the data sets that the ML model would use for training and testing.

Remark 7. *The controller's optimal behavior is specific to each sampling period, whose prediction horizon is specific to $H - k$. So, optimal behavioral learning will be performed for each sampling period.*

For each $k, k = 0, \dots, H - 1$, we construct a matrix SOCSK ($M \times (n + m)$) (SOCSK stands for "States and Optimal Control values concerning Step K"), which is Table 1's image. Line $i, 1 \leq i \leq M$, is devoted to experience i as follows:

$$SOCSK_i \leftarrow \left[(X_i^*(k))^T (U_i^*(k))^T \right].$$

Using the data structures proposed before, the following holds:

$$SOCSK_i \leftarrow [STATE_i(k, 1 : n) \quad UstarRHC(i, k)].$$

$STATE_i$ is the i th element of the $STATE$ cell array. In the PBR case ($n = 2; m = 1$), the data set for the current step will be as follows:

$$SOCSK = \begin{bmatrix} x_1(k)^1 & x_2(k)^1 & u(k)^1 \\ \dots & \dots & \dots \\ x_1(k)^i & x_2(k)^i & u(k)^i \\ \dots & \dots & \dots \\ x_1(k)^M & x_2(k)^M & u(k)^M \end{bmatrix}.$$

A fragment of the SOCSK matrix produced by a MATLAB script is presented in Appendix C for step $k = 1$. Only when $k = 0$ does the value of $x_2(0)$ equal 0 for any observation.

Owing to Remark 7, step 3 should establish for each k the data sets for training and testing; these sets are stored in the cell arrays `DATAkTest` and `DATAkTrain`. Table 2 presents the pseudocode of the script preparing all of the sets needed by the learning algorithm.

Table 2. The pseudocode preparing the data sets for the ML models' training and testing.

/*This pseudocode describes the construction of the data sets needed by the ML models at the level of each sampling period*/	
Inputs: cell array <code>STATE</code> , matrix <code>UstarRHC</code> ;	
Outputs: matrix <code>SOCSK</code> , table <code>datak</code> , cell arrays <code>DATAkTest</code> , <code>DATAkTrain</code>	
1.	#Load the file containing the data structure <code>STATE</code> and <code>UstarRHC</code> (Figure A1)
2.	$k \leftarrow 0$
3.	while $k \leq H-1$
4.	for $i = 1, \dots, M$
5.	$SOCSK_i \leftarrow [STATE_i(k, 1:n) \ UstarRHC(i, k)]$
6.	end
7.	#Convert the matrix <code>SOCSK</code> into the table <code>datak</code> .
8.	<code>datakTest</code> \leftarrow lines #1—60 of <code>datak</code>
9.	<code>datakTrain</code> \leftarrow lines #61—120 of <code>datak</code>
10.	<code>DATAkTest</code> { k } \leftarrow <code>datakTest</code>
11.	<code>DATAkTrain</code> { k } \leftarrow <code>datakTrain</code>
12.	$k \leftarrow k + 1$
13.	end
14.	#Save the cell array <code>DATAkTrain</code> and <code>DATAkTest</code> in a file.

Step 4's implementation will determine the set of ML models $\Phi = \{f_k | k = 0, 1, \dots, H-1\}$ and will be addressed in the next section.

Step 5 aims to implement the ML controller. Once the set of regression models is determined in step 4, the controller can be written following the algorithm in Table 3.

Table 3. The structure of the ML controller's algorithm.

The General Algorithm of the ML Controller	
/*The controller program is called at each sampling period, k */	
1	Get the current value of the state vector, $X(k)$; /* Initialize k and $X(k)$ */
2	Predict the optimal control value $U^*(k)$ using the regression model $f_k(X(k))$ /* whatever is the regression model's type */
3	Send the optimal control value $U^*(k)$ towards the process.
4	Wait for the next sampling period.

Notice that the cumulative effect of calling the controller at each sampling period is to achieve the following sequence of predictions using the regression models and the current states that the process accesses:

$$U_1^* = f_0(X_0); U_2^* = f_1(X_1); \dots U_{H-1}^* = f_{H-1}(X_{H-1}).$$

In the sequel, the controllers based on ML models will be called LR controller (from linear regression) or RNN controller (from Regression Neural Network).

5. Linear Regression Controller

5.1. General Algorithm

The first approach that the authors considered was to use multiple linear regression for the function set Φ . Such a model contains an intercept, linear terms for each state variable, squared terms, products of features (interactions), etc. Hence, as functions of state variables, the regression functions could be nonlinear.

For our example, the *stepwise regression* strategy [30], which adds or removes terms starting from a constant model, was also applied. We consider in this presentation only models with an intercept, linear terms, and an interaction:

$$f_k(X(k)) = C^k_0 + C^k_1 \cdot x_1(k) + C^k_2 \cdot x_2(k) + C^k_{12} \cdot x_1(k) \cdot x_2(k). \quad (7)$$

Remark 8. Our goal is not to find the best sequence of linear regression models but to validate our approach, i.e., the ML model can successfully replace the couple (APSOA and PM) inside a new controller.

As we shall see, the model (7) is largely sufficient for our goal.

Table 4 presents the construction of the H models representing linear regressions in a general manner, that is, not only for our example. This pseudocode describes the linear models' training and testing using the sets generated in step 3.

Table 4. The pseudocode of the linear regressions' construction.

Construction of the linear regression models	
Input:	cell arrays DATAKTrain, DATAKTest
Output:	matrix KOEF ($H \times (n + 1)$), /* the regression coefficients for each sampling period */ cell array MODELSPW ($H \times 1$) /* cell array storing objects that are the linear models f_k */
1	for $k = 0 \dots H-1$.
2	datakTrain ← DATAKTrain[k]; /* Recover the data set for training */
3	datakTest ← DATAKTest[k]; /* Recover the data set for testing */
4	mdlsw ← fitting_to_data(datakTrain); /* Training the linear regression */
5	#display mdlsw; /* mdlsw is the linear regression model */
6	coef(:) ← get_the_coefficients(mdl)
7	KOEF(k,:) ← coef(:); /* The k^{th} line of KOEF receives the coefficients */
8	MODELSPW[k,1] ← mdlsw;
9	uPred ← fpredict(mdlsw, datakTest) /* The vector uPred stores the predicted control values */
10	# Make the comparison between uPred and the real control values;
11	end.

The script in Table 4 uses the generic functions *fitting_to_data*, *get_the_coefficients*, and *fpredict*, which make actions suggested by the comments.

The implementation of this algorithm is included in the script GENERATE_ModelSPW; some details are given in Appendix D.

5.2. Simulation Results

Although we have determined the usual linear regressions that contain the two linear terms (for x_1 and x_2) and an intercept, we present hereafter the *stepwise* version as it is implemented in the MATLAB system. Table 5 displays a listing's fragment obtained during the script GENERATE_ModelSW's execution; this one presents the model for a single sampling period.

Table 5. The actions and results of the stepwise linear regression for the 14th sampling period.

&&&&kp1 = 14				
1. Adding x1, FStat = 12.7755, p Value = 0.000484491				
Linear regression model: u ~ 1 + x1				
Estimated Coefficients:				
	<u>Estimate</u>	<u>SE</u>	<u>tStat</u>	<u>p_Value</u>
(Intercept)	−985.91	437.95	−2.2512	0.025952
x1	2147.7	600.87	3.5743	0.00048449
Number of observations: 140; Error degrees of freedom: 138				
Root mean square error: 117				
R-squared: 0.0847; Adjusted R-squared: 0.0781				
F-statistic vs. constant model: 12.8; p-value = 0.000484				

The procedure begins with only an intercept, and after that, it tries and succeeds in adding the term corresponding to x_1 . Statistical parameters do not allow us to add another term. So, the prediction (control value) will be $f_{13}([x_1, x_2]) = -985.91 + 2147.7 \cdot x_1$.

We notice that the training time for all 120 linear regressions is 6.166654 s.

Following the algorithm presented before, the resulting coefficients of the H regression are given in Table 6.

Table 6. The coefficients of the H linear regressions determined with a stepwise strategy.

k	C_0	C_1	C_2
0	564.18		0
1	585.53	0	0
2	−20.368	1441.5	0
:	:	:	:
9	591.85	0	0
10	119.74	0	620.84
11	591.48	0	0
12	590.95	0	0
13	−985.91	2147.7	0
14	−328.16	1205.6	0
:	:	:	:
111	4055.1	−1476.6	0
112	5482.6	−2067.6	0
113	597.8	0	0
114	3300.2	0	−308.67
115	587.66	0	0
116	6446.3	−2451.2	0
117	7144	−2732.8	0
118	4410.7	0	−419.32
119	565.2	0	0

There are sampling periods for which the regression model has only the intercept C_0 ; this situation will be discussed in Section 7. These coefficients will be used directly by the controller as a control law.

The comparison achieved in lines #9-10 of the construction algorithm is summarized in Figure 6.

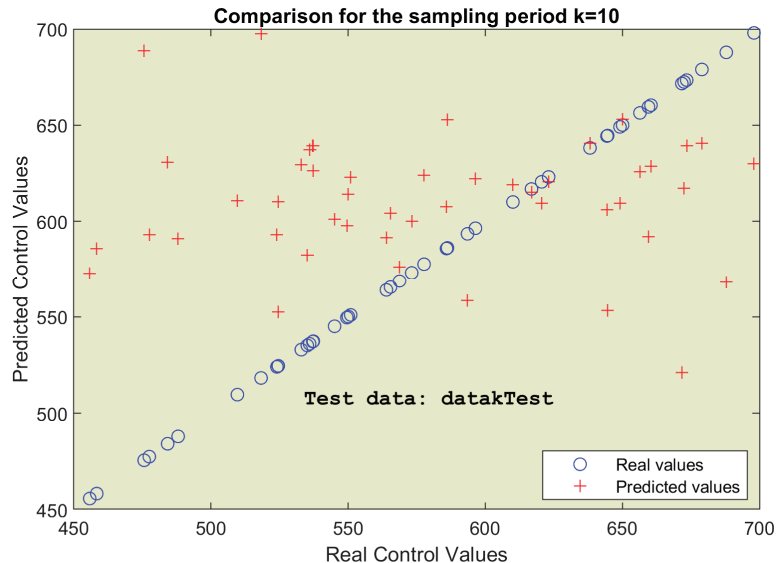


Figure 6. Comparison: predicted versus real control values for specific sampling period.

We have to mention that the predicted values were calculated directly using the formulas, not using the generic function *fpredict*. The table *datakTest* supplied the 60 examples, states—control value, for testing the linear regressions.

The quality of the predictions will be evaluated at the time of using the regression models within the controller, that is, inside the closed-loop simulation. The ultimate evaluation of predictions would be the optimality of the process evolution.

To prepare this evaluation, we need the simulation program for the control loop working with the LR controller. The flowchart in Figure 7 describes this program, *CONTROL_loopLINREG*, which is step #6 of the design procedure.

Although there are similarities with Figure 3, there actually are big differences in execution:

- The state variable has two elements.
- The coefficients' matrix must be loaded from an existing file.
- The gray instructions make the predictions, avoiding any numerical integration.
- The green instruction updates the next state, which has two components. The amount of light irradiated in the current sampling period is added to $x_2(k)$.

Only the orange column of the flowchart has big similarities because it is about the simulation results needed to depict the process evolution and the performance index.

The script *CONTROL_loopLINREG.m* included in the attached folder implements the presented algorithm.

The closed-loop simulation program produces a listing, a fragment of which is presented in Figure A3 (Appendix D), and two drawings reproduced in Section 7.

We notice the very short time period used to control the process over the entire control horizon, 0.6401 s (the simulation processor is Intel(R) Core(TM) i7-6700HQ CPU @ 2.60 GHz).

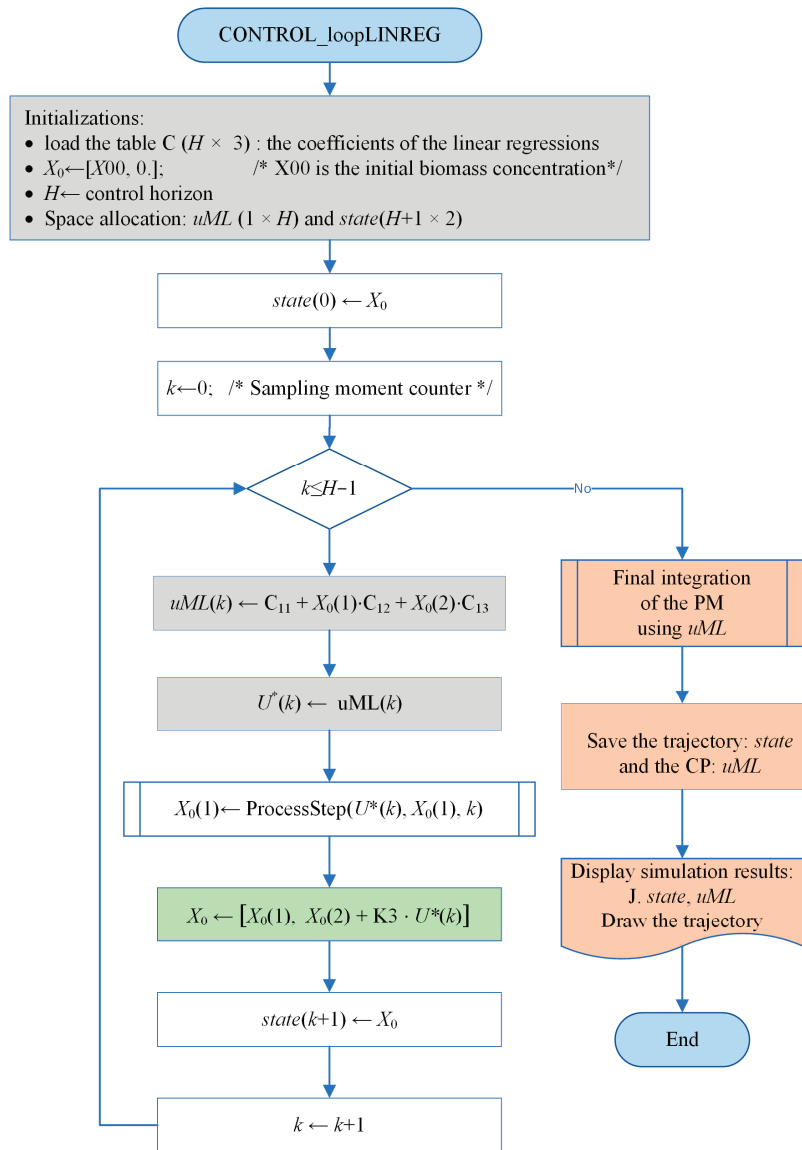


Figure 7. The simulation program for the ML controller based on linear regression models.

6. Controller Based on Regression Neural Networks

6.1. General Approach

Because the linear regression could seem much too simple, we have studied other types of models (trees, support vector machines, and Gaussian processes) trying to improve capturing the optimality of the couple (APSOA and PM), the final target being that the designed ML controller would better approach the optimal behavior. The obtained models have performances weaker than those of the LR and RNN models.

Better predictions than those obtained with other types of ML models are produced by Regression Neural Networks (RNNs), of course, with a possible penalty concerning the model's size. The decision to choose between these models and the linear regressions in implementing the controller will be analyzed in Section 7.

As in the case of linear regressions, the RNN models must be obtained offline, and their construction must be organized in a loop because the number of sampling periods could be large, like in our case. The pseudocode of the RNNs' construction is presented in Table 7.

Table 7. The pseudocode of the regression NNs' construction.

Construction of the RNN models	
Input:	cell arrays DATAkTrain, DATAkTest
Output:	cell array MODELNN $\{H \times 1\}$
/* cell array storing objects that are RNN */	
1	for $k = 0 \dots H-1$.
2	datakTrain ← DATAkTrain[k, 1]; /* Recover the data set for training */
3	datakTest ← DATAkTest[k, 1]; /* Recover the data set for testing */
4	mdlINN ← trainRegNN(datakTrain); /* Training the RNN */
5	MODELNN[k, 1] ← mdlINN /* Store the object mdlINN into the cell array MODELNN */
6	predictionNN ← mdlINN.predictFcn(datakTest) /* Make predictions and store them into the table predictionNN */
7	# Comparison between predictionNN and datakTest
8	end

In our case study, for $k = 0$, i.e., the first sampling period, we have a special situation because $x_2 \equiv 0$ (the light amount equals 0 through initialization) for all examples. Hence, this variable cannot be a prediction variable. For this situation, the tables datakTrain and datakTest have different structures, and the RNN model has a single predictor variable x_1 . To maintain the general structure of the algorithm in Table 7, we did not treat the first sampling period distinctly.

Most data structures were presented before, except for the cell array MODELNN that collects the model for each sampling period, called "mdlINN". The function trainRegNN trains the current RNN using its specific data set [32].

In line #6, the predictions made by the method "mdlINN.predictFcn" are stored in the local table predictionNN, which can be compared to datakTest or saved for further utilization.

The implementation of this algorithm is achieved by the script GENERATE_ModelINN; some details are given in Appendix E.

6.2. Simulation Results

The execution of the script GENERATE_ModelINN gives us an indication of the RNNs' construction complexity (training and testing). A fragment of its listing is given in Table 8.

The RMSEValid is the root mean square error (RMSE) between the predictions and datakTest vectors. The vrmse value is the RMSE calculated in the training process (the phase of validation). The program displays the k value, the prediction uNN, and the control value (valreal) for the states included in record #10 of the data set (as an example).

We notice that all 120 RNNs are trained in 233 s (offline, as we mentioned before).

The comparison mentioned in line #7 of the algorithm can be achieved by calculating the root mean square error (RMSE) between the predicted and observed values. For graphical analysis, Figure 8 plots the predicted values yielded by the RNN model versus the real control values from the datakTest table.

Table 8. The execution of RNN training (a fragment of the listing).

```

>> GENERATE_ModelINN
-----
&&&& vrmse = 108.4187 RMSEValid = 111.9601
kplus1 = 1 uNN = 564.1789 valreal = 487.0276

&&&& vrmse = 104.5341 RMSEValid = 130.6362
kplus1 = 2 uNN = 574.4657 valreal = 672.5530

&&&& vrmse = 106.0061 RMSEValid = 127.3596
kplus1 = 3 uNN = 559.1601 valreal = 676.6160
-----
&&&& vrmse = 112.0696 RMSEValid = 91.4825
kplus1 = 119 uNN = 588.0487 valreal = 634.9497

&&&& vrmse=130.1694 RMSEValid = 125.2698
kplus1 = 120 uNN = 576.2782 valreal = 714.3906

Elapsed time is 232.829571 s.
    
```

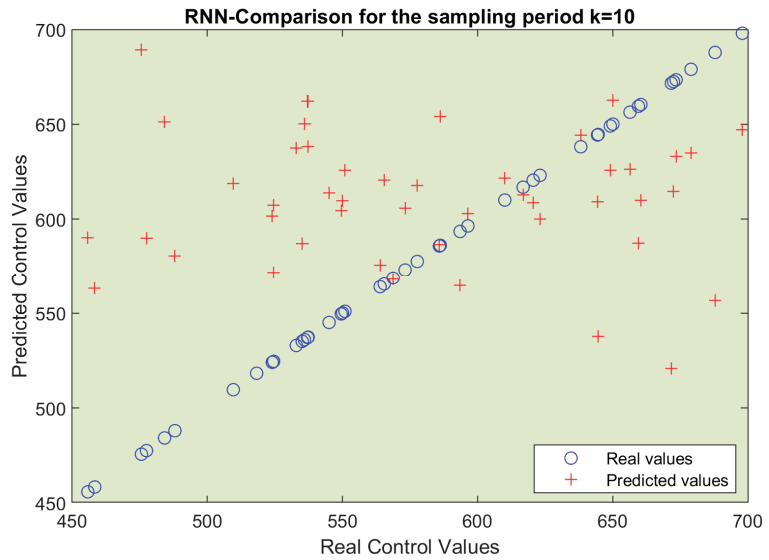


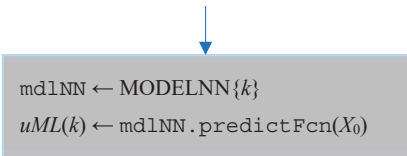
Figure 8. The predicted values yielded by the RNN model versus the real control values for a specific sampling period.

As in Section 5.2, to evaluate the controller’s performances, we need the simulation program for the closed-loop functioning with the RNN models. Its algorithm’s flowchart would be very similar to that of Figure 7, except for two instructions. That is why we do not redraw the flowchart, as we are content with indicating only the changes.

The gray instruction

$$u_{ML}(k) \leftarrow C_{11} + X_0(1) \cdot C_{12} + X_0(2) \cdot C_{13}$$

will be replaced by this block



This block means that the neural network model `mdlNN` is selected as the current RNN from the cell array of objects `MODELNN`. Its method `predictFcn` will calculate the predicted control value as a function of the current state.

The second change is inside the block “Initializations”. Instead of loading the coefficients’ table, `C`, it will load the cell array `MODELNN`; the latter is already saved in a file created by the script constructing the RNN models (see `GENERATE_ModelNN.m`).

The script `CONTROL_loopNN.m` included in the attached folder implements the above algorithm.

The simulation program `CONTROL_loopNN` produces a listing, a fragment of which is presented in Figure A4 (Appendix E), and two drawings, presented in Section 7.

We notice, as in the case of the linear regression controller, the very short time period used to control the process over the entire control horizon, 1.35 s.

7. Discussion

7.1. Comparison between PSO and ML Predictors

In this section, we have to answer the following questions:

- Did the ML predictors succeed in “learning” the behavior of the couple (APSOA and PM), such that the process’s evolution would be quasi-optimal?
- Did the controller’s execution time decrease significantly?

As we mentioned, we have already solved the PBR problem using RHC and a predictor based on PSO. For the sake of simplicity, we shall refer to its controller as the *PSO Controller*. Using the `ControlLoop_PSO` script, the closed-loop simulation produces the typical evolution depicted in Figure 9. *This simulation is one among the $M = 200$ evolutions that contributed to our big data set with CPs and trajectories.* The final lines of the simulation’s listing summarize its performances, as given in Table 9.

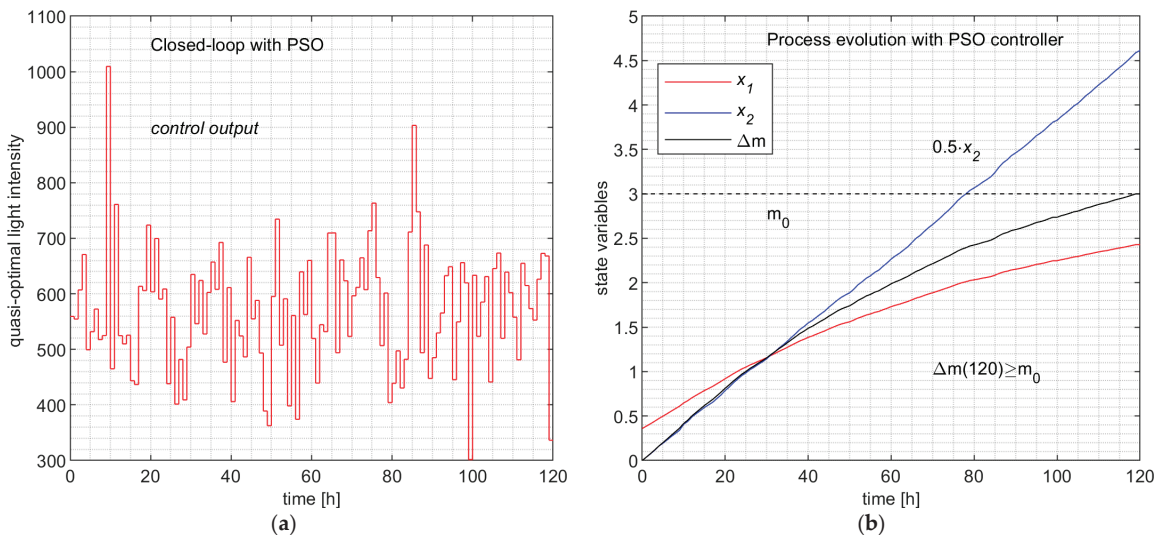


Figure 9. Closed-loop evolution with the PSO controller. (a) The control output values over the control horizon; (b) the process and the biomass evolution.

The simulation programs `CONTROL_loopLINREG` and `CONTROL_loopNN` produce, besides the data in Figures A3 and A4, the evolutions depicted in Figures 10 and 11, respectively.

Table 9. Execution of closed-loop simulation program (ControlLoop_PSO).

>>ControlLoop_PSO_RHC
x00 = 0.3660
Yield mass = 3.0000
Light = 9.2474
Perf index = 9.2474
Elapsed time is 447.281879 s.

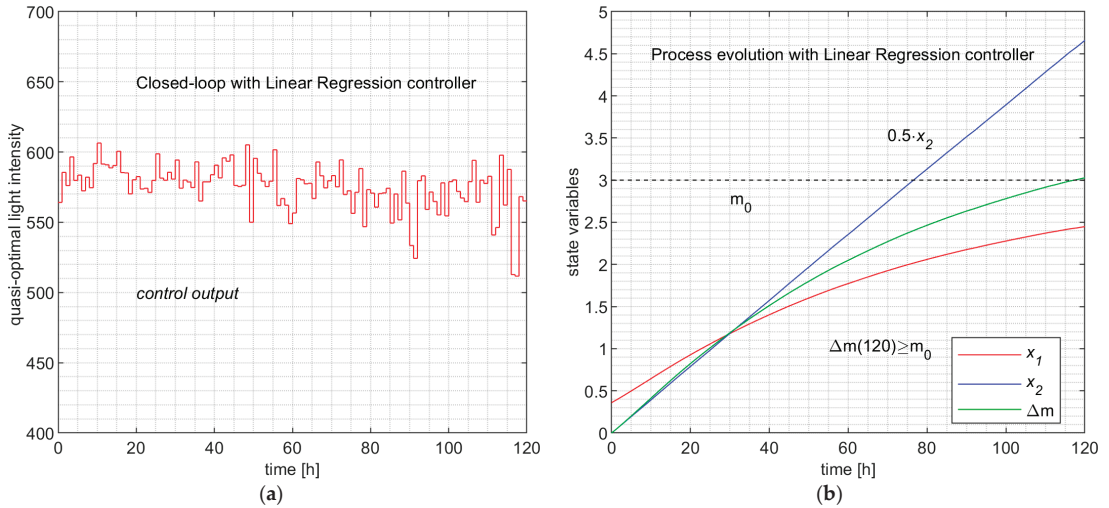


Figure 10. Closed-loop evolution with the linear regression controller. (a) The control output values over the control horizon; (b) the process and the biomass evolution. The mass m_0 labels the dotted line, the final value of the green curve.

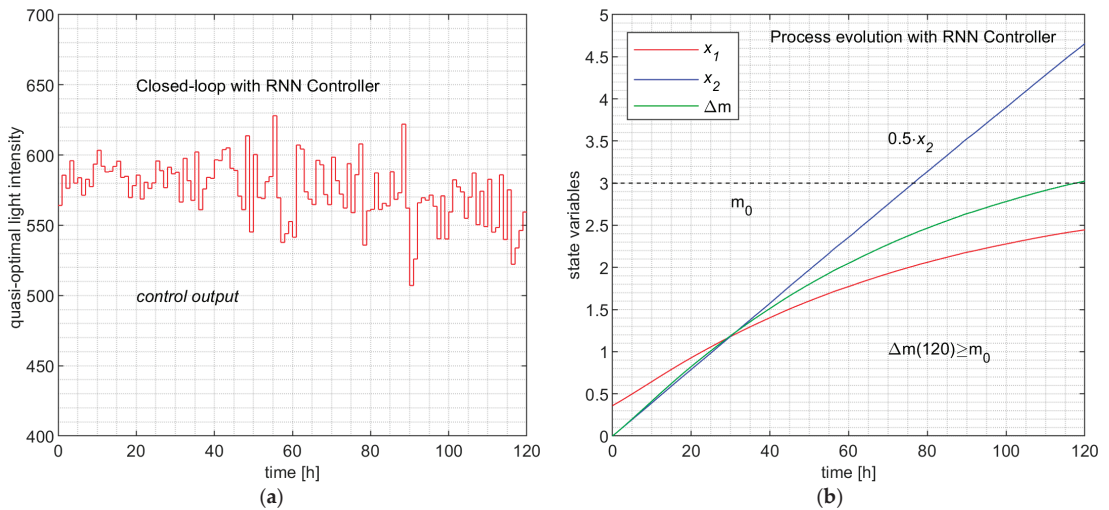


Figure 11. Closed-loop evolution with the Regression Neural Network controller. (a) The control output values over the control horizon; (b) the process and the biomass evolution. The mass m_0 labels the dotted line, the final value of the green curve.

Remark 9. The state evolutions (x_1 and x_2) and the evolution of the biomass, which can be considered the output of the process, are practically identical. Hence, both ML controllers emulate the PSO controller.

The LR controller and the RNN controller are facing situations when the current state is totally new (states unobserved in the training or testing data sets). In this situation, the generalization ability of the ML model is exploited but also verified. That is why the simulation programs are adequate tests for the predictions' testing.

Remark 10. Using the controller inside a closed-loop simulation program over the control horizon will be a test in which the predictor experiences new process states unobserved in the training and testing phase of the ML model's construction.

The fact that Figures 9a, 10a and 11a, describing the control value's evolution, are very different has no relevance to the matter at hand; the following aspects uphold this:

- The similarity at this level would imply the same sequences of states, but we just stated that the ML controllers could experience new unobserved states. So, the three processes do not pass through the same set of states (the sets of accessible states are different).
- The learning is made at the level of each sampling period and the model "learns" couples (state; control value), not globally, but at the control profile level; our method is not based on learning CPs. On the other hand, the PSO predictor is very "noisy" due to its stochastic character and produces outliers among the 200 control values from time to time.

To make a quantitative comparison among the three controllers, Table 10 displays some pieces of information from Figures A3 and A4, and Table 9.

Table 10. Quantitative comparison among the three controllers.

	PSO Controller	LR Controller	RNN Controller
x00	0.360	0.360	0.360
Yield mass	3.0000	3.0282	3.0249
Light	9.2474	9.3166	9.3115
Perf index	9.2474	9.5981	9.5604
Control time [s]	447.28	0.64	1.35
Training time [s]	-	5.4	232.83
Model size	-	3 kB	17 kB
Root mean square error (RMSE for k = 10)	-	113.73	116.63

The first four lines of Table 10 are evidence that the three controllers can be considered equally performant, although the ML controllers have slightly better parameters. However, the time devoted to controlling the process over the control horizon (all of the H sampling period) has values much inferior to that of the PSO controller (hundreds of times smaller). This analysis and Remark 9 allow us to state the following conclusions:

- The two ML controllers have predictors that have learned the behavior of the couple (APSOA and PM), such that in closed-loop functioning, the process evolves almost identically.
- The resulting controllers have execution times hundreds of times smaller than that of the PSO Controller.

7.2. Comparison between the LR and RNN Controllers

In Table 10, all of the parameters in the second column are superior to those in the third column. The differences are not relevant for some of them, but the control time, training time, and model size make the LR controller preferable to the RNN controller. However, we have trained all its RNNs using hyperparameter optimization.

The usual comparison between the predicted and real (observed) values is illustrated in Figures 6 and 8 for the LR and RNN predictors, respectively. This comparison is made for $k = 10$ (as an example) and its testing table `datakTest`. For the other values of k , the situation is the same.

At first sight, both predictors seem to be similar, but the values of RMSE from Table 10 show that the LR predictor is slightly better than the RNN predictor. This remark goes in the same direction as the superiority of the LR controller. However, because the `tatakTest` has 60 data points, we can consider the difference between RMSEs irrelevant and that they are similar from an accuracy point of view.

Considering Remark 6, the time to train offline 120 RNNs (for the RNN controller) of 4 min is absolutely acceptable. So, even this controller can be considered a good solution for the PBR problem or another OCP.

For the reader, who is a newcomer equally in the fields of control systems, computational intelligence, and machine learning, we must compare the role of the PSO predictor versus the role of the ML predictor when solving an OCP.

- *The ML predictor replaces the PSO predictor only in execution* when the controller achieves the control action. So, the controller's execution time is hundreds of times smaller compared to initially. That was our desideratum.
- When we solve a new OCP, sometimes we need a metaheuristic (PSO, EA, etc.) that searches for the optimal solution inside of a control structure. If the controller's execution time is not acceptable, we can use the approach presented in this paper to create an ML controller. However, initially, we need the MA to search for the optimal solution.

8. Conclusions

In this paper, we have proposed two ML controllers (LR controller and RNN controller), including the linear regression and Regression Neural Network predictors, that can replace the controller using a PSO algorithm; the optimal control structure works with an internal process model.

The main conclusions are given as follows:

- The machine learning models succeed in "learning" the quasi-optimal behavior of the couple (PSO and PM) using data capturing the PSO predictor's behavior. The training data are the optimal control profiles and trajectories recorded during M offline simulations of the closed-loop over the control horizon.
- The current paper proposes algorithms for collecting data and aggregating data sets for the learning process. The learning process is split according to the level of each sampling period so that a predictor model is trained for each one. The multiple linear regression and the Regression Neural Networks are considered the predicting models.
- For each case, we propose algorithms for constructing the set of ML models and the controller (LR or RNN controller). Algorithms for the closed-loop simulations using the two controllers are also proposed; they allow us to compare the process evolutions involved by the three controllers, the PSO, LR, and RNN controllers.
- The final simulations show that the new controllers preserve the quasi-optimality of the process evolution. In the same conditions, the process evolutions are almost identical.
- An advantage of our approach refers to data collection, data set preparation for the training process, and the construction of ML models; all of these activities need only simulations (using the PSO controller) and offline program executions (Remark 6). The ML models for each sampling period are determined offline ahead of using the ML controller in real time.

- We emphasize that during the final closed-loop simulations, the ML controller encounters new process states unobserved in the training and testing of its predictor (Remark 10). Owing to its generalization ability, the controller makes accurate predictions of the control value sent to the process.

The PSO predictor first searches for the optimal control value, following its optimization mechanism using particles and the PM, before predicting it. This search sometimes means a big computational effort and a large controller execution time. The ML controller (LR or RNN) predicts using an already-known regression function, which can emulate the PSO predictor. In other words, the ML predictor replaces the PSO predictor only in execution when the controller achieves the control action. That is why the controller's execution time decreases drastically. However, the solution belongs intrinsically to the PSO predictor.

When we solve a new OCP, sometimes, for different reasons, we shall need a meta-heuristic (PSO, EA, etc.) searching for the optimal solution inside of a control structure. Finally, the implemented controller integrated into the control structure can be one of the two ML controllers.

In our opinion, this work goes beyond the controller's execution time decrease and opens a perspective to emulate and replace in a general manner optimization structures.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pr12050991/s1>, File S1: The archive "Processes_PSO_ML.zip" contains the files mentioned in Appendices A–E.

Author Contributions: Conceptualization, V.M.; methodology, V.M.; software, V.M. and I.A.; validation, V.M. and E.R.; formal analysis, V.M.; investigation, I.A.; resources, E.R.; data curation, I.A.; writing—original draft preparation, V.M.; writing—review and editing, V.M. and E.R.; visualization, E.R.; supervision, I.A.; project administration, I.A.; funding acquisition, E.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research study was funded by the Executive Agency for Higher Education, Research, Development and Innovation Funding (UEFISCDI—Roumania), project code COFUND-LEAP-RE-D3T4H2S; Europe Horizon—LEAP-RE program. The APC received no external funding.

Data Availability Statement: Data are contained within the article and Supplementary Materials.

Acknowledgments: This study benefited from the administrative support of the Doctoral School of "Dunarea de Jos" University of Galati, Romania.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

The controlled physical system is a flat-plate photobioreactor (PBR) lighted on a single side for algal growth. The constructive and physical parameters are presented in Table A1. Their definitions are irrelevant to this work.

The PBR is a distributed parameter system because the light is attenuated inside. To convert it into a lumped-parameter system, its depth, $L = 0.04$ m, is discretized in $k_L = 100$ points equally spaced ($z_i \in [0, L]$). After discretization, the process model (PM) is as follows:

$$\dot{x}_1(t) = \left[\mu_{\max} \cdot \frac{1}{100} \cdot \sum_{i=1}^{100} \frac{G_i(t)}{k_S + G_i(t) + \frac{1}{k_I} \cdot G_i(t)^2} - \mu_d \right] \cdot x_1(t)$$

$$\dot{x}_2(t) = A \cdot C \cdot q(t)$$

$$G_i(t) = q(t) \cdot k_i^{x_1(t)}, \quad i = 1, \dots, k_L$$

$$k_i = e^{-\frac{1+\alpha}{2\alpha} \cdot E_a \cdot z_i}, \quad i = 1, \dots, k_L$$

$$m(t) = V \cdot x_1(t)$$

The control input $u(t)$ is considered the intensity of incident light:

$$u(t) = q(t).$$

The state variables are as follows:

$x_1(t)$: the biomass concentration (in g/L);

$x_2(t)$: the light amount which, up to moment t , has illuminated the PBR (in $\mu\text{mol}/\text{m}^2/\text{s}$).

Table A1. The constants of the PBR model.

$E_a = 172 \text{ m}^2 \cdot \text{kg}^{-1}$	absorption coefficient
$E_s = 870 \text{ m}^2 \cdot \text{kg}^{-1}$	scattering coefficient
$b = 0.0008$	backward scattering fraction
$\mu_{max} = 0.16 \text{ h}^{-1}$	specific growth rate
$\mu_d = 0.013 \text{ h}^{-1}$	specific decay rate
$K_S = 120 \mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$	saturation constant
$K_I = 2500 \mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$	inhibition constant
$V = 1.45 \cdot 10^{-3} \text{ m}^3$	volume of PBR
$L = 0.04 \text{ m}$	depth of PBR
$A = 3.75 \cdot 10^{-2} \text{ m}^2$	lighted surface
$x_0 = 0.36 \text{ g/L}$	initial biomass concentration
$C = 3600 \cdot 10^{-2}$	light-intensity conversion constant
$k_L = 100$	number of discretization points
$q_m = 50 \mu\text{mol}/\text{m}^2/\text{s}$	lower technological light intensity
$q_M = 2000 \mu\text{mol}/\text{m}^2/\text{s}$	upper technological light intensity
$m_0 = 3 \text{ g}$	minimal final biomass
$t_{\text{final}} = 120 \text{ h}$	control horizon
$T = 1 \text{ h}$	sampling period

The output variable, the biomass $m(t)$ calculated by Equation (A1), is the PBR's product. As a productivity constraint, it must hold that

$$m(t_f) \geq m_0. \quad (\text{A1})$$

The cost function (A2) represents the amount of light used for the current batch while constraint (A1) is fulfilled. The two weight factors (w_1 and w_2) are established through simulation.

$$J(q(\cdot), x_0) = w_1 \cdot \int_{t_0}^{t_f} q(t) \cdot dt + w_2 (m(t_f) - m_0). \quad (\text{A2})$$

Appendix B

Our implementations are based on the MATLAB system and language. The reader can find inside the folder `Processes_PSO_ML`, supplied in the Supplementary Materials, the guide "README.txt". The following scripts can be used to carry out the closed-loop simulation:

- `ControlLoop_PSO_RHC.m` that implements the "ControlLoop_PSO" program;
- `INV_PSO_Predictor1.m` that implements the "Predictor_PSO" function;
- `INV_RealProcessStep.m` that implements the "ProcessStep" function.

The functions called recursively are also present inside the folder.

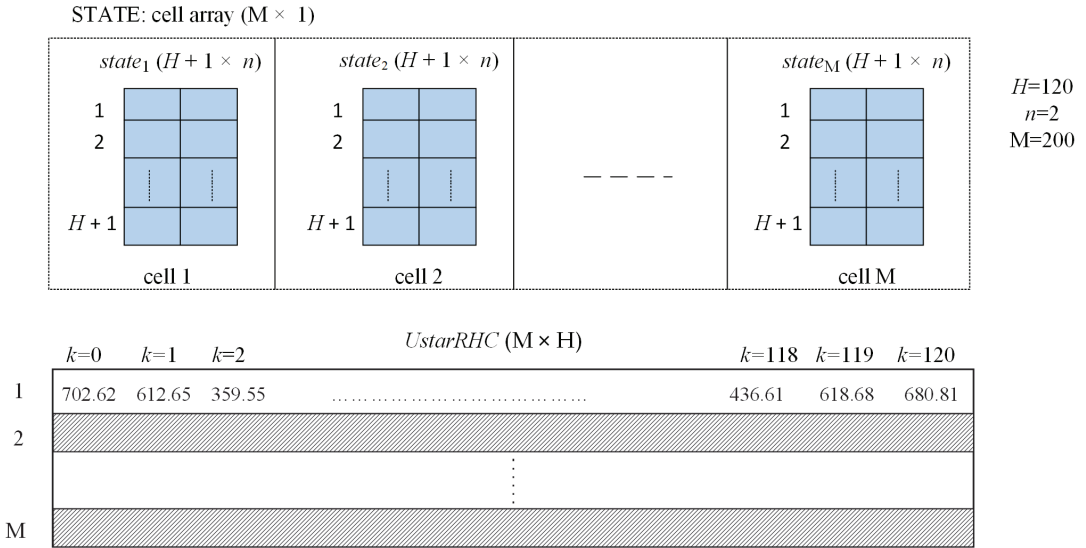


Figure A1. The data collected following M executions of the closed-loop simulation.

The script `LOOP_M_ControlLoop_PSO.m`, also included in the Supplementary Materials, gathers data from all of the 200 simulations and saves them in the file `WS_data200.mat`.

A fragment of the matrices describing the closed-loop simulation’s data, that is, the quasi-optimal evolution, is given in Figure A2. Notice that we have a single control variable and the 121st state is the final one.

$state_1(H + 1, n)$			$uRHC$
k	x_1	x_2	$u^*(k)$
0	0.3502	0	702.62
1	0.3783	0.0948	612.65
2	0.4060	0.1775	359.55
3	0.4281	0.2261	535.81
⋮	⋮	⋮	⋮
117	2.3947	9.0497	436.61
118	2.3988	9.1087	618.68
119	2.4083	9.1922	680.81
120	2.4192	9.2841	-

Figure A2. The matrices for the optimal trajectory and its CP (first simulation).

Appendix C

A fragment of the SOCSK matrix produced by a MATLAB script is presented hereafter. The matrices like this one, presented in Table A2, are the data sets that allow for the construction of the ML models for each sampling period.

Table A2. The matrix SOCSK for the second sampling period ($k = 1$).

	x1	x2	u
i = 1	0.37831	0.094853	612.65
i = 2	0.40708	0.077925	557.37
i = 3	0.40122	0.066811	802.69
i = 4	0.40359	0.079127	387.99
i = 5	0.37865	0.078882	560.77
i = 6	0.37801	0.079205	510.57
...
...
i = 197	0.40607	0.074351	547.26
i = 198	0.39419	0.069684	485.64
i = 199	0.38625	0.082921	558.9
i = 200	0.39531	0.071325	539.92

Appendix D

The Linear Regression Models' Construction

This construction of the $H = 120$ linear regressions is achieved by using the script GENERATE_ModelSW. The latter opens the file WS_Modelsv1.mat (see below) to load the needed data sets.

The generic functions *fitting_to_data* and *get_the_coefficients* from Table 4 correspond to MATLAB functions *stepwiselm* and *mdlsw.Coefficients.Estimate*. The *fpredict* function is directly implemented using the regression formula and the coefficients.

The reader can also examine the script Model_ConstructionLINREG.m, which does not use the *stepwise regression* strategy; the regression models contain only the two linear terms and an intercept. The coefficients for all 120 regression functions are stored in the file WS_3coeff.mat. The cell arrays, MODEL—which stores the 120 objects of type linear regression—DATAKTest, and DATAKTrain, are saved in the file WS_Modelsv1.mat.

In every step, the current state, the predicted control value u_{ML} , and the process's next state are displayed. The final lines display the biomass produced, $\Delta m = 3.0282$ g, and the performance index $J = 9.5981$.

```
>> CONTROL_loopLINREG
k= 1  x1= 0.3600 x2= 0.0000 uML= 564.1787
      next state X01= 0.3863 X02= 0.0762
k= 2  x1= 0.3863 x2= 0.0762 uML= 585.5286
      next state X01= 0.4138 X02= 0.1552
           :
           :
           :
k= 118 x1= 2.4269 x2= 9.0945 uML= 511.6694
       2.4269      9.0945
       next state X01= 2.4331 X02= 9.1636
k= 119 x1= 2.4331 x2= 9.1636 uML= 568.2736
```

Figure A3. Cont.

```

2.4331      9.1636
      next state X01= 2.4408 X02= 9.2403

k= 120  x1= 2.4408 x2= 9.2403 uML= 565.2016
      2.4408      9.2403
      next state X01= 2.4484 X02= 9.3166
#####
final state:
      2.4484      9.3166
&&& x00= 0.3600 yield mass= 3.0282 Light= 9.3166 Perf Index= 9.5981
Elapsed time is 0.640188 seconds.

```

Figure A3. Execution of the closed-loop simulation program (CONTROL_loopLINREG).

Appendix E

The script GENERATE_ModelNN.m implements the algorithm presented in Table 7. We recall that it is carried out offline in step 5 of the design procedure.

The function *trainRegNN* is implemented in two versions by the script *trainRegNNK0.m* for the first sampling period and *trainRegNN.m* for the others; it trains the RNN and can be generated automatically using the regression application (eventually with hyperparameter optimization) or written ad hoc using another training function. As an orientation, we give here a few RNN parameters:

```

RNN = fitrnet(predictors, response...,
              'LayerSizes', [14 1 7], ...
              'Activations', 'none', ...
              'Lambda', 0.00015, ...
              'IterationLimit', 1000, ...
              'Standardize', true);

```

(see [32]).

```

>> CONTROL_loopNN

kplus1=  1  x1= 0.3600 x2= 0.0000 uML= 564.1785
      Next state:      X01= 0.3863 X02= 0.0762

kplus1=  2  x1= 0.3863 x2= 0.0762 uML= 585.6589
      Next state:      X01= 0.4138 X02= 0.1552
                        :
                        :
kplus1= 118 x1= 2.4247 x2= 9.0902 uML= 533.8785
      Next state:      X01= 2.4316 X02= 9.1622

kplus1= 119 x1= 2.4316 x2= 9.1622 uML= 546.2667
      Next state:      X01= 2.4387 X02= 9.2360

```

Figure A4. Cont.

```

kplus1= 120  x1= 2.4387  x2= 9.2360  uML= 559.4318
Next state:      X01= 2.4461  X02= 9.3115

Elapsed time is 1.349449 seconds.
#####
final state:
      2.4461      9.3115
&&&  x00= 0.3600  yield mass= 3.0249  Light= 9.3115  Perf Index= 9.5604

```

Figure A4. Execution of the closed-loop simulation program (CONTROL_loopNN).

In every step, the current state, the predicted control value u_{ML} , and the process's next state are displayed. The final lines of Figure A4 display the biomass produced, $\Delta m = 3.0249$ g, and the performance index $J = 9.5604$.

References

1. Siarry, P. *Metaheuristics*; Springer: Berlin/Heidelberg, Germany, 2016; ISBN 978-3-319-45403-0.
2. Talbi, E.G. *Metaheuristics—From Design to Implementation*; Wiley: Hoboken, NJ, USA, 2009; ISBN 978-0-470-27858-1.
3. Kruse, R.; Borgelt, C.; Braune, C.; Mostaghim, S.; Steinbrecher, M. *Computational Intelligence—A Methodological Introduction*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2016.
4. Tian, G.; Zhang, L.; Fathollahi-Fard, A.M.; Kang, Q.; Li, Z.; Wong, K.Y. Addressing a Collaborative Maintenance Planning Using Multiple Operators by a Multi-Objective Metaheuristic Algorithm. *IEEE Trans. Autom. Sci. Eng.* **2023**, *7*, 1–13. [CrossRef]
5. Onwubolu, G.; Babu, B.V. *New Optimization Techniques in Engineering*; Springer: Berlin/Heidelberg, Germany, 2004.
6. Valadi, J.; Siarry, P. *Applications of Metaheuristics in Process Engineering*; Springer International Publishing: Berlin/Heidelberg, Germany, 2014; pp. 1–39. [CrossRef]
7. Abraham, A.; Jain, L.; Goldberg, R. *Evolutionary Multi-objective Optimization—Theoretical Advances and Applications*; Springer: Berlin/Heidelberg, Germany, 2005; ISBN 1-85233-787-7.
8. Minzu, V.; Serbencu, A. Systematic procedure for optimal controller implementation using metaheuristic algorithms. *Intell. Autom. Soft Comput.* **2020**, *26*, 663–677. [CrossRef]
9. Hu, X.B.; Chen, W.H. Genetic algorithm based on receding horizon control for arrival sequencing and scheduling. *Eng. Appl. Artif. Intell.* **2005**, *18*, 633–642. [CrossRef]
10. Mayne, D.Q.; Michalska, H. Receding Horizon Control of Nonlinear Systems. *IEEE Trans. Autom. Control* **1990**, *35*, 814–824. [CrossRef]
11. Minzu, V.; Rusu, E.; Arama, I. Execution Time Decrease for Controllers Based on Adaptive Particle Swarm Optimization. *Inventions* **2023**, *8*, 9. [CrossRef]
12. Goggos, V.; King, R. Evolutionary predictive control. *Comput. Chem. Eng.* **1996**, *20* (Suppl. S2), S817–S822. [CrossRef]
13. Chiang, P.-K.; Willems, P. Combine Evolutionary Optimization with Model Predictive Control in Real-time Flood Control of a River System. *Water Resour. Manag.* **2015**, *29*, 2527–2542. [CrossRef]
14. Wu, J.; Zhang, C.; Giam, A.; Chia, H.Y.; Cao, H.; Ge, W.; Yan, W. Physics-assisted transfer learning metamodels to predict bead geometry and carbon emission in laser butt welding. *Appl. Energy* **2024**, *359*, 122682. [CrossRef]
15. Minzu, V.; Riahi, S.; Rusu, E. Implementation aspects regarding closed-loop control systems using evolutionary algorithms. *Inventions* **2021**, *6*, 53. [CrossRef]
16. Minzu, V.; Georgescu, L.; Rusu, E. Predictions Based on Evolutionary Algorithms Using Predefined Control Profiles. *Electronics* **2022**, *11*, 1682. [CrossRef]
17. Banga, J.R.; Balsa-Canto, E.; Moles, C.G.; Alonso, A. Dynamic optimization of bioprocesses: Efficient and robust numerical strategies. *J. Biotechnol.* **2005**, *117*, 407–419. [CrossRef] [PubMed]
18. Balsa-Canto, E.; Banga, J.R.; Alosa, A.V. Vassiliadis. Dynamic optimization of chemical and biochemical processes using restricted second-order information 2001. *Comput. Chem. Eng.* **2001**, *25*, 539–546. [CrossRef]
19. Minzu, V.; Arama, I. A Machine Learning Algorithm That Experiences the Evolutionary Algorithm's Predictions—An Application to Optimal Control. *Mathematics* **2024**, *12*, 187. [CrossRef]
20. Minzu, V.; Riahi, S.; Rusu, E. Optimal control of an ultraviolet water disinfection system. *Appl. Sci.* **2021**, *11*, 2638. [CrossRef]
21. Minzu, V.; Ifrim, G.; Arama, I. Control of Microalgae Growth in Artificially Lighted Photobioreactors Using Metaheuristic-Based Predictions. *Sensors* **2021**, *21*, 8065. [CrossRef] [PubMed]

22. Goodfellow, I.; Bengio, Y.; Courville, A. Machine Learning Basics. In *Deep Learning*; The MIT Press: Cambridge, MA, USA, 2016; pp. 95–161. ISBN 978-0262035613.
23. Zou, S.; Chu, C.; Shen, N.; Ren, J. Healthcare Cost Prediction Based on Hybrid Machine Learning Algorithms. *Mathematics* **2023**, *11*, 4778. [CrossRef]
24. Cuadrado, D.; Valls, A.; Riaño, D. Predicting Intensive Care Unit Patients' Discharge Date with a Hybrid Machine Learning Model That Combines Length of Stay and Days to Discharge. *Mathematics* **2023**, *11*, 4773. [CrossRef]
25. Albahli, S.; Irtaza, A.; Nazir, T.; Mehmood, A.; Alkhalifah, A.; Albattah, W. A Machine Learning Method for Prediction of Stock Market Using Real-Time Twitter Data. *Electronics* **2022**, *11*, 3414. [CrossRef]
26. Wilson, C.; Marchetti, F.; Di Carlo, M.; Riccardi, A.; Minisci, E. Classifying Intelligence in Machines: A Taxonomy of Intelligent Control. *Robotics* **2020**, *9*, 64. [CrossRef]
27. Alatefi, S.; Abdel Azim, R.; Alkouh, A.; Hamada, G. Integration of Multiple Bayesian Optimized Machine Learning Techniques and Conventional Well Logs for Accurate Prediction of Porosity in Carbonate Reservoirs. *Processes* **2023**, *11*, 1339. [CrossRef]
28. Guo, R.; Zhao, Z.; Huo, S.; Jin, Z.; Zhao, J.; Gao, D. Research on State Recognition and Failure Prediction of Axial Piston Pump Based on Performance Degradation Data. *Processes* **2020**, *8*, 609. [CrossRef]
29. Newbold, P.; Carlson, W.L.; Thorne, B. Multiple Regression. In *Statistics for Business and Economics*, 6th ed.; Pfaltzgraff, M., Bradley, A., Eds.; Pearson Education, Inc.: Upper Saddle River, NJ, USA, 2007; pp. 454–537.
30. The MathWorks Inc. *Stepwise Regression Toolbox Documentation*; The MathWorks Inc.: Natick, MA, USA, 2024; Available online: <https://www.mathworks.com/help/stats/stepwise-regression.html> (accessed on 2 September 2023).
31. Goodfellow, I.; Bengio, Y.; Courville, A. Example: Linear Regression. In *Deep Learning*; The MIT Press: Cambridge, MA, USA, 2016; pp. 104–113. ISBN 978-0262035613.
32. The MathWorks Inc. *Regression Neural Network Toolbox Documentation*; The MathWorks Inc.: Natick, MA, USA, 2024; Available online: <https://www.mathworks.com/help/stats/regressionneuralnetwork.html> (accessed on 2 September 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Robust Forest Fire Detection Method for Surveillance Systems Based on You Only Look Once Version 8 and Transfer Learning Approaches

Nodir Yunusov ¹, Bappy MD Siful Islam ¹, Akmalbek Abdusalomov ^{1,2} and Wooseong Kim ^{1,*}

¹ Department of Computer Engineering, Gachon University, Sujeong-gu, Seongnam-si 461-701, Gyeonggi-do, Republic of Korea; bappy.ac.bd@gmail.com (B.M.S.I.); a.abdusalomov@tsue.uz (A.A.)

² Department of Information Systems and Technologies, Tashkent State University of Economics, Tashkent 100066, Uzbekistan

* Correspondence: wooseong@gachon.ac.kr

Abstract: Forest fires have emerged as a significant global concern, exacerbated by both global warming and the expanding human population. Several adverse outcomes can result from this, including climatic shifts and greenhouse effects. The ramifications of fire incidents extend widely, impacting human communities, financial resources, the natural environment, and global warming. Therefore, timely fire detection is essential for quick and effective response and not to endanger forest resources, animal life, and the human economy. This study introduces a forest fire detection approach utilizing transfer learning with the YOLOv8 (You Only Look Once version 8) pretraining model and the TranSDet model, which integrates an improved deep learning algorithm. Transfer Learning based on pre-trained YoloV8 enhances a fast and accurate object detection aggregate with the TranSDet structure to detect small fires. Furthermore, to train the model, we collected 5200 images and performed augmentation techniques for data, such as rotation, scaling, and changing hue and saturation. Small fires can be detected from a distance by our suggested model both during the day and at night. Objects with similarities can lead to false predictions. However, the dataset augmentation technique reduces the feasibility. The experimental results prove that our proposed model can successfully achieve 98% accuracy to minimize catastrophic incidents. In recent years, the advancement of deep learning techniques has enhanced safety and secure environments. Lastly, we conducted a comparative analysis of our method's performance based on widely used evaluation metrics to validate the achieved results.

Keywords: forest fire; fire detection; YOLOv8; deep learning; TranSDet; wildfire incidents; brushfire spread

Citation: Yunusov, N.; Islam, B.M.S.; Abdusalomov, A.; Kim, W. Robust Forest Fire Detection Method for Surveillance Systems Based on You Only Look Once Version 8 and Transfer Learning Approaches. *Processes* **2024**, *12*, 1039. <https://doi.org/10.3390/pr12051039>

Academic Editors: Xiong Luo, Sheng Du, Li Jin, Xiongbo Wan and Zixin Huang

Received: 25 March 2024

Revised: 1 May 2024

Accepted: 15 May 2024

Published: 20 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Forest fires are catastrophic events that result in widespread economic, ecological, and environmental damage all over the world. High temperatures can ignite dry fuels, such as sawdust, leaves, and lightning, or they can be sparked by human activities, such as unextinguished fires, arson, or improperly burned debris [1]. Between 2002 and 2016, an estimated 4,225,000 km³ of forest burned by fire [2]. Forest fires can arise from both natural phenomena and human activities. Natural causes include factors such as heat, dry weather, lightning strikes, volcanic eruptions, coal-seam fires, and smoking. On the other hand, human-induced causes encompass activities like cooking, accidental ignition, and deliberate fire lighting. Either natural or human-created fires have a severe impact on wild as well as human life. Human activity contributes 90%, and lightning provides the rest of the 10% of forest fire sources [3]. Both people and wildlife are affected by wildfire toxic gases in the troposphere [4].

Forest fires were previously identified by watchtowers, which are ineffective, and human surveillance is expensive [4,5]. Implementing automation offers a significantly improved and more precise approach to forest fire detection. Additionally, weather conditions, rainy days, and high temperatures constrain the fire detection process. Therefore, a real-time fire detection technique is much better and has a low cost [6].

To prevent expanding fire, commonly there are two methods used, and are vision-based fire detection (VFDs) and sensors that exhibit sensitivity to sound, flames, temperature, gases, or solid materials [7]. Sensors trigger chemical characteristics of smoke and the variance of the environment. Once smoke is in the range, the sensor alarm turns on. In certain scenarios, sensor-based detection systems [8] may not be feasible, particularly in expansive coverage areas, forested regions, and environments with elevated temperatures, as they may generate numerous false alarms [9]. Moreover, the operational range of sensors is constrained [10], leading to a reduced lifespan.

The inception of object detection leveraging AI (artificial intelligence) traces back to 1986 [11]. Nonetheless, the substantive contributions of AI and machine learning models were hindered during that period due to technological constraints. Furthermore, the introduction of Facebook and Google big data forced deep learning (DL) models to have technical advantages. Perceptron was the first stepping stone towards deep learning. Time-different deep learning techniques, such as Alexnet [12], VGG16 [13], Faster R-CNN [14], Detectron2 [15], and YOLOv1 [16], were introduced by the need for scaling tasks, speed, and so on. In this study, the following problems were identified in detecting forest fires.

1. Collection and labeling images of forest fires pose significant challenges, primarily attributed to the absence of readily available open-access datasets containing fire images.
2. Given the absence of standardized shapes or sizes of fires, detecting objects of varying dimensions in real-time poses a considerable challenge, particularly in achieving high levels of accuracy.
3. Fire and fire-like object detection as fire is a real problem in forest fire identification and classification.

The integration of artificial intelligence (AI) with mathematical models for fire detection and prediction has been a burgeoning area of research. In [17–19], researchers provide a comprehensive overview of various machine-learning techniques applied to fire detection and prediction. It discusses the integration of AI algorithms with mathematical models for more accurate predictions. These papers explore the application of deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), for wildfire detection and prediction. They discuss how these AI methods can be integrated with mathematical models to improve prediction accuracy [20,21].

Some papers discuss wildfire probability modeling and resilience in wildland fire management, exploring wildfire risk assessment methodologies and resilience planning strategies [22]. The mentioned approaches introduced probabilistic risk modeling, scenario analysis, and community-based resilience planning. It also provides case studies illustrating the application of these methods in wildfire-prone regions. Scientists also present an integrated modeling framework that combines wildfire probability modeling with resilience assessment. It discusses how environmental factors, land use patterns, and social dynamics influence both wildfire occurrence and community resilience [23,24].

This paper is organized into the following sections: Section 2 presents an overview of the relevant literature. Section 3 details our dataset and explains our proposed fire detection method. Section 4 provides a comprehensive examination of experimental results, accompanied by a detailed analysis of performance, and Section 5 discusses limitations and outlines future work. In the end, Section 6 provides the findings and final summaries to overcome challenges.

2. Literature Review

Forest flame recognition techniques generally fall into the following two main categories: artificial intelligence/computer vision approaches and sensor-based techniques. There are some limitations to the sensor-based method. To overcome these limitations, we upgraded the deep learning method based on transfer learning on pre-trained Yolov8 and TranSDet. CNNs [25–29] (convolutional neural networks) and DNN (deep neural network) are the most popular methods in the field of object detection. As sensor-based systems have some limitations and object detection techniques cover these limitations, deep learning has gained more popularity [30,31].

2.1. Detection of Forest Fires Utilizing Machine Learning and Deep Learning Methodologies

With AI advancing day by day, numerous research techniques have been innovated in the field of deep learning on fire detection. Among all deep learning models, CNNs commonly use techniques in computer vision. Toulouse et al. [32] proposed a system to recognize geometrical observation of flame based on length and position. The model's foundation is the classification of pixels based on the non-refractory pictures' average intensity. Jian et al. [33] upgraded the model with a boundary detection operator with multistep operation. However, the model only obtains better performance in simple and stable fire and flame images. For the first time in the history of deep learning, using a combination of foreground and background color frames, Turgay [34] generated a real-time fire detector. Though the model produced better output for fire images only, for fire-like images, and in the presence of smoke and shadow, the model was outperformed. With the improvement in deep learning, LDS is capable of identifying dynamic smoke and flame textures [35].

With technological improvement, researchers introduced the YOLOFM algorithm based on YOLOv5n approach to classifying small objects, which makes forest fire detection more accurate [36]. Furthermore, in [37–41], more improved fire detection techniques were innovated. To improve fire detection precision accuracy, a DL-based approach named DTA (Detection and Temporal Accusations) was proposed in [42]. It tries to identify a person by imitating the identification process. This model increases the accuracy of flame identification while accurately interpreting the temporal SRoF characteristics. The authors of [43] designed an early flame recognition technique using a lightweight CNN.

2.2. Detection of Forest Fires Utilizing YOLO and Transformers Methodologies

In [44], the Yolov3 network was employed for small-scale object classification. This model leverages the K-means clustering technique to distinguish flames. In [45], the authors introduce a fire depth separable convolution to mitigate the computational costs of the method and enhance the perceptual feature layer using the cavity convolution method. Furthermore, in [46], the authors propose ELASTIC-YOLOv3 as an enhancement over YOLOv2 to enhance performance without increasing the number of parameters. Initially, fire detection algorithms encountered challenges, such as high light intensity, limited color information, and variations in flame shapes and sizes, which prompted the development of enhanced technologies for real-time flame classification and recognition, manifested in modulated YOLO networks (v4, v5, v6, v7, v8) as introduced in [47–53].

Far object detection using a deep learning model did not perform well before introducing transformer learning. Transformer-based learning shows superior object prediction performance in various advanced vision areas, including image/video analyzing [54], image super-resolution [55], object recognition [56], segmentation [57], and ViT [58] (image classification). This advancement is facilitated through the utilization of Vision Transformer [59], DeiT [60] (Data-Efficient Image Transformers), and MedT [61] (Medical Transformers). Vision Transformer (ViT) is a deep learning model that applies the transformer architecture, originally designed for sequence processing tasks like natural language processing, to image data. Instead of using convolutional layers like traditional convolutional neural networks (CNNs), ViT represents images as sequences of patches, which are then processed

by a transformer encoder. ViT divides the input image into fixed-size, non-overlapping patches and flattens them into sequences, which are then fed into the transformer encoder. By leveraging self-attention mechanisms, ViT captures long-range dependencies within the image and learns representations that are effective for image classification tasks. Data-Efficient Image Transformers (DeiT) is a variant of ViT designed to achieve better performance with smaller amounts of labeled training data. DeiT introduces techniques such as distillation, where a larger pre-trained model (teacher model) is used to distill knowledge into a smaller student model. Medical Transformers (MedT) are transformer-based models specifically tailored for medical imaging tasks. These models adapt the transformer architecture to handle medical image data, which often differ from natural images due to factors like different modalities (e.g., MRI, CT scans) and specialized structures (e.g., anatomical features). MedT models are trained in medical imaging datasets and learn to perform tasks, such as image classification, segmentation, and detection, within the medical domain. By leveraging the power of transformers, MedT models can capture complex patterns and relationships in medical images, leading to improved performance in various medical image analysis tasks [62–64].

3. Proposed Method and Model Architecture

3.1. Forest Fire Dataset

For the model to be trained, the first task begins with collecting a diverse range of datasets. We collected fire images and videos from the internet. Due to a wide range of datasets helping to generalize the model, our obtained images collected distinct angles, focal lengths, and brightening conditions. Additionally, on the internet, there are some popular platforms such as Roboflow, Bing, Flickr images, and Kaggle to arrange images. To achieve more accurate results, we divided our dataset into two classes: fire and non-fire images. To train our dataset, we standardized the dimensions of all images by resizing them to the same height and width, thus minimizing potential errors. Our dataset comprised 7000 images captured during both the daytime and nighttime, which were subsequently compressed, as shown in Table 1.

Table 1. Images of the forest fire scenes from the custom dataset.

Dataset	Google, Bing, Kaggle, Flickr Images	Video Frames	Total
Forest Fire Images	4136	2864	7000

After collecting the dataset, we applied custom image pre-processing, dropped 1000 images from the dataset and left 6000 images. Figure 1 shows how we increased our dataset by rotating each image at 90° angles to 270° using a computer vision algorithm. After applying the augmentation technique, our dataset increased four times, and the total number of images extended to 24,000 images. Our dataset is divided as follows: 70% for training images, 10% for test images, and 20% for validation images, as shown in Table 2.

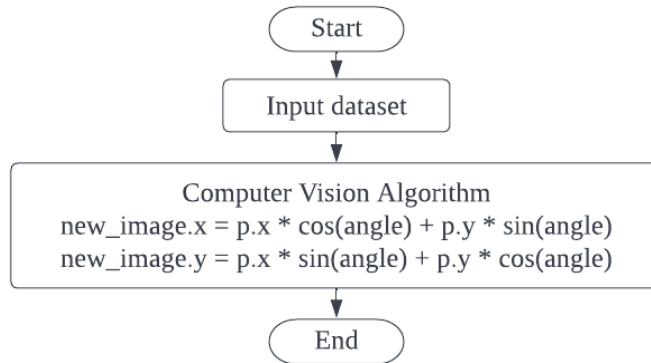


Figure 1. Sample images of forest fires rotated from various perspectives. (a) 90° rotation, (b) 180° rotation, (c) 270° rotation, and (d) the original image.

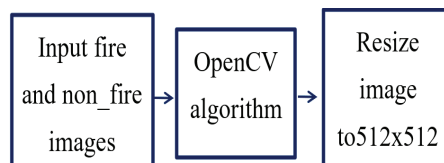
Table 2. Distribution of flame frames within the dataset.

Dataset	Training Images	Testing Images	Validation Images	Total
Fire	11,760	1680	3360	16,800
Non-Fire	5040	720	1440	7200

Scheme 1 shows the convergence of image sizes where x and y are the new image dimensions. P is the input image matrix, and using the multiplication of the angle, the expected image rotation takes place.

**Scheme 1.** Image processing (rotation).

After completing image augmentation, we labeled our images into two classes: fire and non-fire. According to image annotation, a JSON file was created for each image. Earlier in this section, we mentioned that we resized all the images to overcome unexpected errors. In our dataset, we resized our images to 512×512 using OpenCV2, as shown in Figure 2.

**Figure 2.** The overall process of resizing images.

3.2. Model Selection

YOLOv8 is the latest edition of the CNN dynamic object model with high accuracy in real-time. The entire image region is processed by a single neural network, after which it is divided into multiple components, and the potential bounding boxes and probability estimates are estimated for each component. The YOLOv8 network is the continuous improvement of YOLOv1 to YOLOv7. The YOLO model has a backbone that is a series of convolution layers pulling the pixels down at the different resolution sizes, and these features pass through a neck where they are pulled together and put into a head, which leads to an object detection process base loss matrix. The YOLOv8 model architecture is shown in Figure 3.

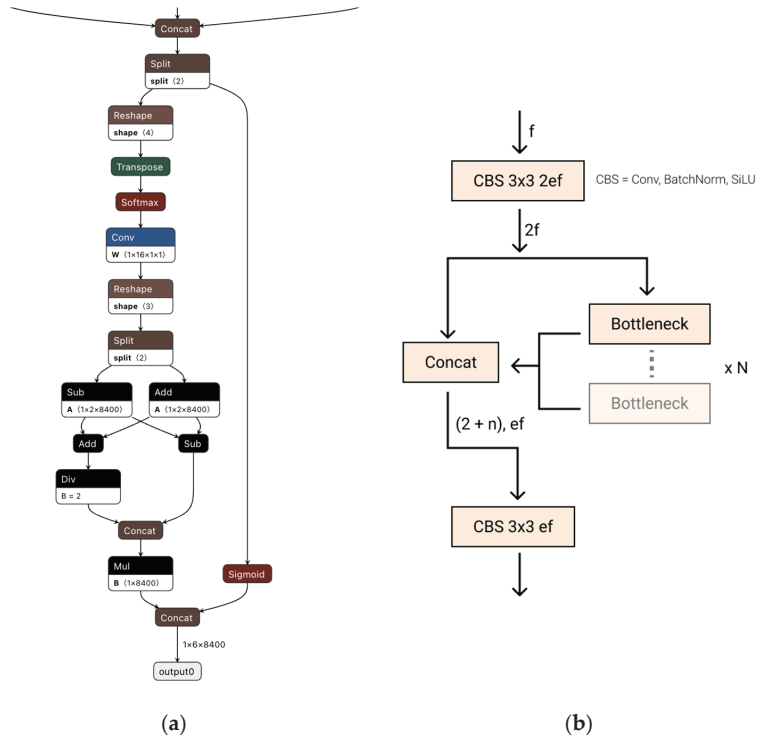


Figure 3. (a,b) are the overall architecture of the YOLOv8 model.

As shown in Figure 3, YOLOv8 is an anchor-free model. Instead of predicting the offset of an anchor box, it predicts the center of an object directly. With anchor-free detection, fewer box predictions are made, which speeds up Non-Maximum Suppression (NMS), which is a post-processing step used to sift through candidate detections. YOLOv8 uses RasNet as its head. The final detection stage used C2F; here, f is the number of features. CBS block consists of convolution, batch norm, and silu function. The final detection stage uses anchor boxes to show the probability of object detection.

The YOLOv8 network contains five different models. YOLOv8n is a good solution for mobile phone applications. YOLOv8s (small model) is compatible with the CPU. YOLOv8m is a medium-sized model that has 25.9 million parameters, which are balanced between speed and accuracy. YOLOv8l has 43.7 million parameters, and this model works best with numerous databases and training. Out of the five models, YOLOv8x is the largest and has the highest mean average precision; nevertheless, it performs slower than YOLOv8l. The relation between YOLOv8n and YOLOv8x is shown in Table 3.

Table 3. Relation between the YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x model [65].

Network	Size (Pixels)	aMP ^{val} (50–95)	Speed CPU (ms)	Speed T4 GPU (ms)	Params (M)	Flop _s (B)
YOLOv8n	640	37.3	-	-	3.2	8.7
YOLOv8s	640	44.9	-	-	11.2	28.6
YOLOv8m	640	50.2	-	-	25.9	78.9
YOLOv8l	640	52.9	-	-	43.7	165.2
YOLOv8x	640	53.9	-	-	68.2	257.8

3.3. Proposed Forest Fire Model

In this subsection, we elaborate on the methodologies employed in computer vision for forest detection, focusing on deep learning approaches, transfer learning techniques, and the aggregation of models.

3.3.1. Transfer Learning

Deep learning algorithms learn features from data, which provide additional support to a deep learning model to detect objects more accurately and fast. However, collecting data is expensive, may not be available, and annotation is a tedious task with high costs. Transfer learning allows one to learn from features and transfer model weight to another model for further learning.

First, we review traditional methods of object detection for transfer learning [66]. This paper proposes using the YOLO pre-trained weights and transfer learning features. As we mentioned earlier, we used 24,000 forest fire images in our dataset, and for the train, 16,800 images were used. We used the default YOLOv8 in all five models to train our dataset and show the result after 50,000 iterations in Table 4. In addition, image hue 0.1, saturation 1.5, and exposure 1.5 were used.

Table 4. Pre-trained weights obtained using a limited dataset.

Models	Input Size	Training Accuracy (AP50)	Testing Accuracy (AP50)	Weight Size	Iteration Number	Training Time
YOLOv8n		83.8%	81.8%			27 h
YOLOv8s		84.1%	82.9%			34 h
YOLOv8m	512 × 512	86.4%	84.8%	186 MB	50,000	38 h
YOLOv8l		91.7%	90.7%			43 h
YOLOv8x		87.1%	85.5%			48 h

Table 4 shows the results obtained for the training and testing accuracies with different indicators. YOLOv8l showed the highest training and test accuracy of 91.7% and 90.7% in 43 h, respectively. The following results were obtained from YOLOv8x with 87.1% and 85.5% accuracy. YOLOv8m and YOLOv8s had 86.4% and 84.1% training accuracy and 84.8% and 82.9% testing accuracy, respectively. YOLOv8n showed the lowest training accuracy at 83.8%; on the other hand, 81.8% was the testing accuracy with 27 h of training time. Moreover, our fire detection approach using YOLOv8 showed better accuracy with large-size forest fire detection but was not efficient with small-size forest fires. Images of forest fires, small-scale fires, and no fires may all be distinguished with the human eye. For deep learning, the method needs more information to improve prediction accuracy. Figure 4 shows the overall fire detection using the YOLOv8 object detection model. Here, the model shows insufficient accuracy in terms of small images but is compatible with large-size forest fire detection.



Figure 4. Big and small-size forest fire detection.

3.3.2. Detect Small-Size Image

From the previous section, detecting small fire images has some limitations in our model. To solve the accuracy factor, we drove forward the concept of the TranSDet [67] model. This model proposes a meta-learning-based dynamic resolution adaption transfer learning (DRAT) schema to adapt the pre-trained general model to detect small objects. The model includes an additional stage on the pre-trained model using DRAT and then transfers the model modified for the intended database. The pre-trained model only generates training on small objects, and to generalize for small objects, the data augmentation (resizing the input images to small images) technique is used to train the model. Fine-tune follow-up is used after the augmentation technique. Figure 5 shows the TranSDet architecture.

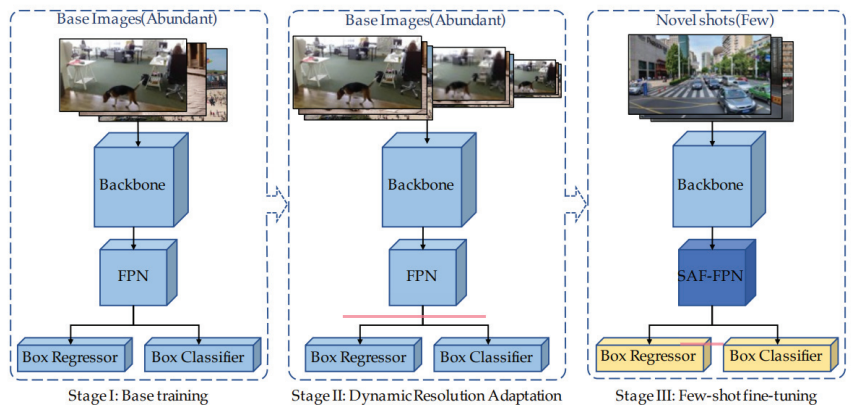


Figure 5. The overall architecture of the TranSDet model.

TranSDet model directly transfers conventional stage 1 and stage 3 pre-trained models using fine-tuning and implementing the dynamic resolution adaptation technique. Stage 2 adopts the pre-trained model to improve transfer learning to detect small objects.

$$\theta = \arg \min_{\theta} E_{R_i} L(D, R_i, M) \quad (1)$$

Here, R is a set of resolutions, and M is the model, and θ pre is the weight to generalize the dataset. In the equation, D represents the database, and L stands for the loss function. To address the meta-learning problem, we employed the widely used MAML (Model Agnostic Meta-Learning) model [68].

$$\theta'_i = \theta - \alpha \nabla_{\theta} L_{R_i}(M(\theta; X_i)) \quad (2)$$

$$\theta_{new} = \theta - \beta \nabla_{\theta} \sum_{R_i} L_{R_i}(M(\theta'_i; X_i)) \quad (3)$$

Here, Equation (2) θ'_i is updated by θ , X_i is the mini batch of images, R_i is the resolution of set R and α is the step size. From Equation (2), DRAT performs an inner update. In Equation (3), θ_{new} is the next iteration. When the model is complete, the final epoch of the model training is complete.

After understanding the mathematical equation, we started building the model and obtained the final output, as shown in Figure 6. After implementing the TranSDet model, we obtained a maximum accuracy of 95%.



Figure 6. Prediction after implementing the TranSDet model.

3.3.3. Model Aggregation

Model aggregation in deep learning refers back to the method of combining the predictions or outputs of a couple of common neural networks to make a final prediction or choice. This technique is regularly used to enhance the overall performance and robustness of gadget learning. There are several techniques for model aggregation, such as boosting, bagging, stacking, averaging, voting, and so on. Our proposed model used the boosting technique (the amalgamation of multiple weak models, each exhibiting slightly superior performance to random guessing, culminating in the formation of a robust learner).

Boosting is an ensemble learning technique that merges multiple weak learners, often decision trees, to construct a formidable learner. Its fundamental concept revolves around assigning greater weight to misclassified instances from the preceding iteration, which helps the ensemble focus on the difficult instances. The equations involved in boosting include the following:

The weighted error (ϵ_t) for the t -th weak learner is calculated as follows:

$$\epsilon_t = \text{Sum of weights of misclassified examples} / \text{Total weight of all examples.}$$

The weight (α_t) for the t-th weak learner is calculated as follows:

$$\alpha_t = 0.5 * \ln((1 - \epsilon_t) / \epsilon_t)$$

The weights are updated based on whether they are correctly or incorrectly classified by the t-th weak learner:

$$\text{If correctly classified: } w_{i,t}(t + 1) = w_{i,t} * e^{(-\alpha_t)}$$

$$\text{If misclassified: } w_{i,t}(t + 1) = w_{i,t} * e^{(\alpha_t)}$$

The final prediction is a weighted combination of the weak learners' outputs as follows: Final Prediction = Sign($\sum \alpha_t * \text{Output of Weak Learner}_t$).

These equations are used iteratively in boosting algorithms like AdaBoost and Gradient Boosting to create a strong ensemble model from weak learners.

Figure 7 shows our proposed simple model diagram. As mentioned earlier, we prepared our dataset and used YOLOv8l pre-train model transfer model learning and TranSDet model learning transfer and aggregated both modes to detect small and large forest fires using the boosting technique.

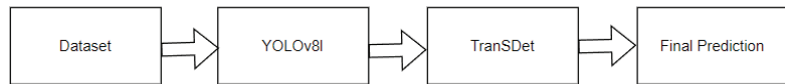


Figure 7. Our proposed model workflow.

After applying the boosting method to our dataset, the accuracy increased to 97% and a small fire to 96%, respectively. Furthermore, the inclusion of fire-like images enhanced our model's accuracy, as shown in Figure 8. Finally, our proposed model was implemented on Raspberry PI 3B+, as shown in Figure 9. The suggested approach employed two different models and used the transfer learning technique, and the model achieved 97% accuracy performance.



Figure 8. Cont.



Figure 8. Prediction after implementing boosting technique on the YOLOv8l and TranSDet model. The first row displays the result of detecting a small fire, the second row shows the fire-like image level to represent no fire, and the third row shows the detection of a large fire.

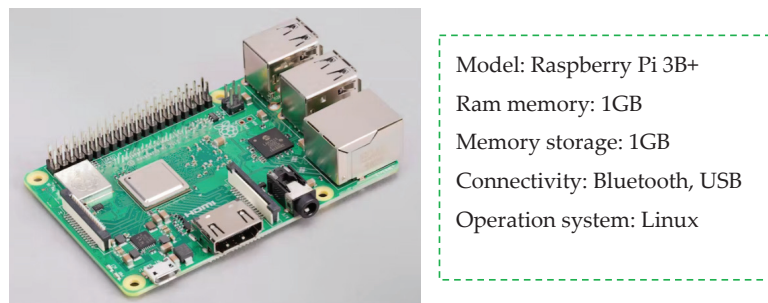


Figure 9. Characteristics of Raspberry Pi 3B+ [12].

To evaluate the performance of our suggested model, we juxtapose it with an established model in Table 5. However, the Results and Discussion sections show an explanation of these observations in detail.

Table 5. Comparison between different models.

Features	YOLOv8l	TranSDet	Our Method (Model Aggregation)
Test speed/s	2 s	2.3 s	4.5 s
Real-time implementation	Possible	Possible	Possible
Small object detection	Possible (but not sufficient)	possible(shows better output)	Possible (highly accurate)
Algorithm	Selective search	Selective search	Selective search

4. Results and Discussions from the Experimentations

Test with Fire and Non-Fire Image

We evaluated our model implementing the Visual Studio code on our MSI GS66 laptop (MSI, Taipei, Taiwan), equipped with a CPU speed of 5.3 GHz, 64 GB of RAM, and 6 GPUs. In the previous section, we discussed and implemented our proposed model and aggregated the YOLOv8l and TranSDet models. In this subsection, we review and discuss our model's advantages and drawbacks. Traditionally, the YOLO model is known for real-time fire detection with high accuracy. However, when we applied the model to our custom dataset, small fire images were provided with insufficient accuracy. To improve our model, we applied another model called TranSDet, which provides a high accuracy of up to 96%. In our proposed model, the boosting technique was used to aggregate both models,

and our model provided accuracy of up to 97% and 96% with large and small forest fires, respectively, as shown in Figures 10 and 11.

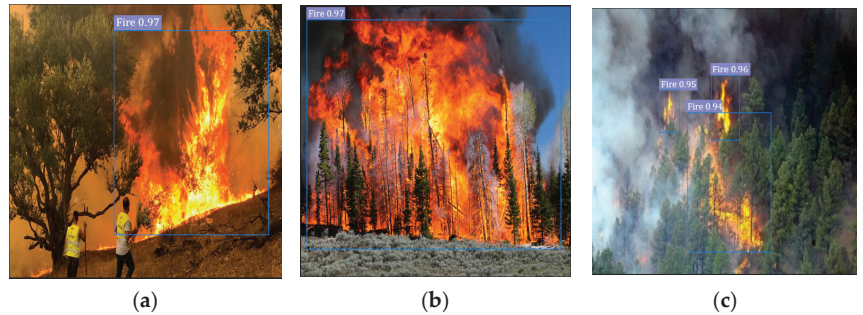


Figure 10. (a–c) Outcomes of daytime image detection accuracies of forest fires.

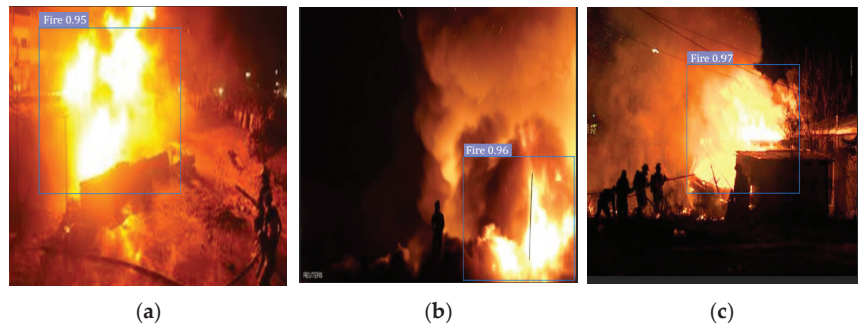


Figure 11. (a–c) Outcomes of nighttime image detection accuracies of forest fires.

Figure 12 shows the training and testing accuracy with loss throughout passing the epoch to train the model. At the beginning of training, our model loss was at its maximum at 0.5. However, after completing the model training process, the loss was minimal at 0.11. On the other hand, for initial testing, it was 0.9, and at the very end, it was 0.1. As we mentioned earlier, for training accuracy, the model reached 96.7% training accuracy and a testing accuracy of 97%.

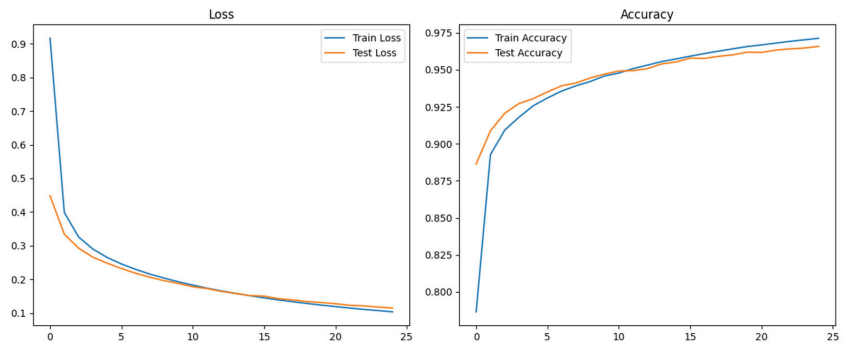


Figure 12. Model training and testing loss and accuracy visualization with epoch.

In this section, the compression of our proposed approach is discussed utilizing various parameters and models. Our model was developed in three stages. First, we used

pre-train YOLOv8l and then the TransDet model to detect small forest fires, and we reached 97% accuracy. F-measures (FM) were employed to compute the weighted average, which balances precision and recall. The true-positive and false-negative rates were taken into consideration when calculating this score. In order to detect an object, the FM parameter was most commonly used since measuring the accuracy rate was difficult. True positives and false negatives were more effective in a detection model with identical weights. If true positives and false negatives differ, precision and recall need to be taken into account. True positive observations are measured by precision.

Recall, on the other hand, is the ratio of false positives to true positives, as described in previous studies [69,70]. Our developed system had a precision of 97% and a false detection rate of 0.7%. As shown in Equations (4) and (5), our proposed model had an average precision and recall rate of 97% and 3%, respectively. TP refers to the accurate detection of a forest fire, while FP refers to a false negative detection (Figure 13).

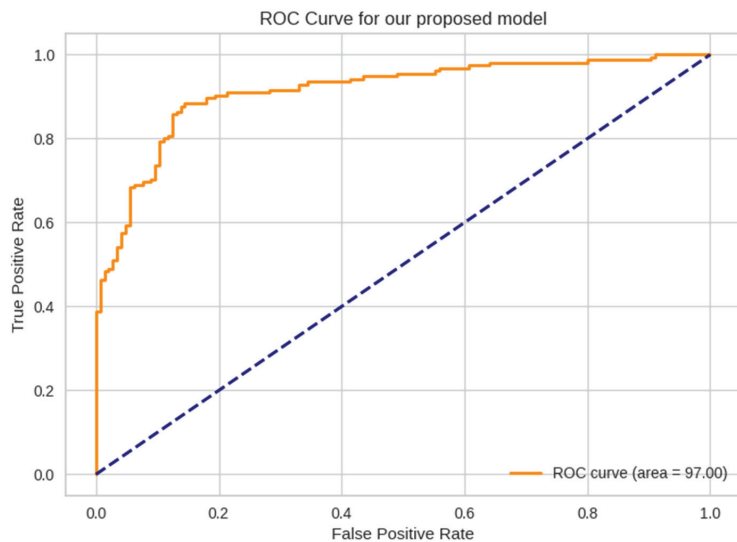


Figure 13. ROC curve of our proposed model.

Equation (6) shows the relation between precision and recall.

$$\text{Precision} = \frac{TP}{TP + FP'} \quad (4)$$

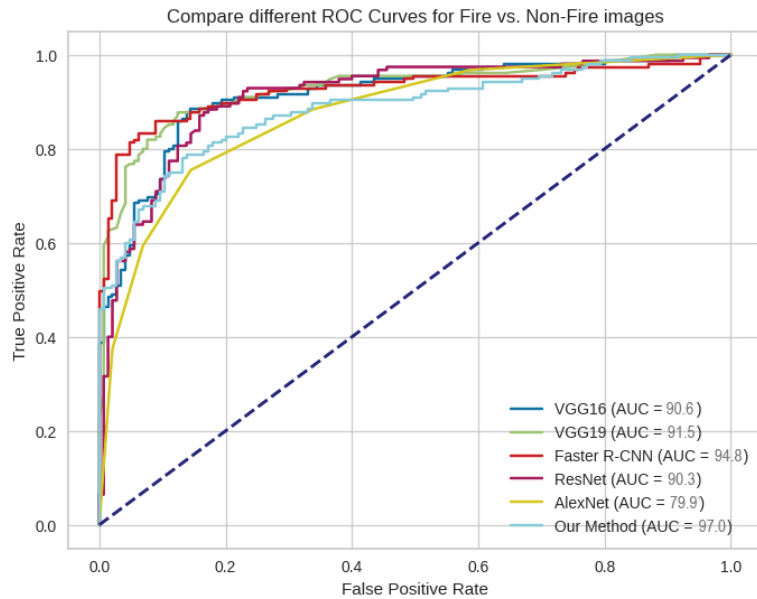
$$\text{Recall} = \frac{TP}{TP + FN'} \quad (5)$$

$$\text{FM} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

Forest fire detection is a complex task in the field of deep learning to achieve high accuracy. Table 6 shows recently published fire detection models with precision, recall, and FM. Here, our model reached the highest precision, recall, and FM with 97%, 96.1%, and 96.5, respectively. This was followed by VGG16 and RsNet with a precision of 92.5% and 90.8%, recall of 82.9% and 98.6%, and finally 90.8% and 90.2%, respectively. However, AlexNet was the poorest in terms of precision, recall, and FM at 73.3%, 61.3%, and 75.1%, respectively (Figure 14).

Table 6. Numerical outcomes for the detection of fire.

Algorithm	<i>P</i> (%)	<i>R</i> (%)	<i>FM</i> (%)	Average (%)
VGG16	92.5	82.9	90.8	90.6
VGG19	93.1	84.5	91.7	91.5
Faster R-CNN [55]	81.7	94.5	87.2	97.8
ResNet [56]	90.8	89.6	90.2	90.3
AlexNet	73.3	61.3	75.1	79.9
Our Method	97	96.1	96.3	96.5

**Figure 14.** Comparison of different models using ROC for the fire and non-fire images.

5. Discussion

As stated in Table 6, the categorization of a model as good or bad is contingent upon specific criteria rather than its overall performance. Furthermore, a model shows high accuracy in some specific tasks, but depending on the complexity model can have some limitations. Our proposed model has a couple of limitations. First, our dataset did not include smoke images. Therefore, if only smoke was visible at the initial stage of the forest fire, the model did not detect this as fire. Next, occasionally, the model considered sun and electric light as forest fires when testing the method in various scenes. In the future, we intend to enhance the developed system by incorporating a database encompassing additional classes from diverse environments pertinent to this challenge [71]. Our methodology involved utilizing extensive datasets, like JFT-300M [72], comprising 300 million annotated images.

Despite the aforementioned challenges, the main contributions of this study are as follows:

1. The pre-trained YOLOv8 model and transferring the learning can detect large-size forest fires. The YOLOv8 algorithm is known for its speed and ability to perform object detection in real-time.
2. To detect small-size fires, the TransDet model and transfer learning approaches can be applied. Utilizing deep learning to acquire fire-specific features, the presented

methodology has the potential to mitigate the prevalent issue of false alarms in conventional fire detection methods. Such an advancement stands to not only prevent unwarranted emergency responses but also to alleviate the financial burden attributed to false alarms.

3. Both models can be aggregated with boosting techniques to detect forest fires. The goal of this research was to apply deep learning models in the field of forest fire prevention. Early detection with high accuracy is beneficial for environmental safety.
4. In contrast to alternative approaches that rely on limited datasets, our method leverages a substantial dataset encompassing fire, fire-like, and standard scenes. This dataset comprises authentic imagery and videos sourced from diverse origins, thereby encapsulating a broad spectrum of fire scenarios. These scenarios encompass both day and night fire incidents, spanning variations in fire scale and accounting for varying lighting conditions, including low-light and high-light environments.

Future efforts will aim to overcome the model's limitation of yielding a high number of false positives, particularly in challenging scenarios like low-altitude cloud cover and haze. Enhancements could involve integrating historical fire record data on fire location, date, and weather conditions, as fires often occur in similar contexts during specific months, thereby improving prediction accuracy. Additionally, the current approach's incompatibility with edge devices presents a drawback. Nonetheless, we plan to address this issue in upcoming research by optimizing the model size while maintaining prediction accuracy. One potential avenue is to employ distillation techniques to train a smaller deep network, such as YOLOv9, which is capable of achieving comparable performance to our current model, thus making it more suitable for edge computing environments.

6. Conclusions

Day-to-day life unicorn tech companies work with big data. Aligning AI to mimic human operations can secure human life in such a way as preventing life-threatening tasks and better service. For object detection, numerous models have been developed based on deep-learning CNN models. Forest fire detection using YOLOv is nothing new and provides high accuracy. In the first proposed model, we collected the dataset and conducted some preprocessing tasks. Our model architecture is based on three stages. First, use the transfer technique on the YOLOv8 pre-train model to detect large-size forest fires. Next, to detect small-size forest fires in real-time, the TransDet model technique was transferred to the learning of our model. Lastly, both transfers were fed learning to the boosting algorithm to train weak learners and show high accuracy.

Several investigations have focused on enhancing forest flame identification and classification tasks through CNN-based AI networks. Nevertheless, the potential of the Detectron2 network remains unexplored in forest fire detection. The acquisition of ample image data for training forest fire detection models presents challenges, often resulting in issues such as data imbalance or overfitting, which can hinder the efficacy of the model. In this study, we introduce a method for forest fire detection utilizing the enhanced Detectron2 model and develop a dataset to address these challenges.

After achieving 97% tested accuracy, our model was implemented in Raspberry Pi 3B+ to run on GPU mode. Furthermore, after testing our model in the variant environment, we detected some limitations in real-time frameworks, such as no data related to smoke frames; therefore, fume detection was not included in our model. In the future, we aim to develop a model that can concentrate on healthcare using 3D CNN/U-Net [73–75].

Author Contributions: Conceptualization, N.Y. and B.M.S.I.; Formal analysis, W.K.; Algorithms: N.Y.; Funding acquisition, W.K.; Investigation, B.M.S.I. and N.Y.; Methodology, A.A. and N.Y.; Project administration, W.K.; Resources, A.A. and N.Y.; Software, B.M.S.I.; Supervision, W.K.; Validation, W.K.; Writing—original draft, N.Y. and B.M.S.I.; Writing—review and editing W.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Gachon University Research Fund (GCU-202300740001), the Ministry of Education of the Republic of Korea, and the National Research Foundation of Korea (NRF-2022S1A5C2A07090938).

Data Availability Statement: The data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Nelson, R. Untamedscience.com. April 2019. Available online: <https://untamedscience.com/blog/the-environmental-impact-of-forest-fires/> (accessed on 30 December 2023).
- Jain, P.; Coogan, S.C.; Subramanian, S.G.; Crowley, M.; Taylor, S.; Flannigan, M.D. A review of machine learning applications in wildfire science and management. *Environ. Rev.* **2020**, *28*, 478–505. [CrossRef]
- Milne, M.; Clayton, H.; Dovers, S.; Cary, G.J. Evaluating benefits and costs of wildland fires: Critical review and future applications. *Environ. Hazards* **2014**, *13*, 114–132. [CrossRef]
- Varma, S.; Sreeraj, M. Object detection and classification in surveillance system. In Proceedings of the 2013 IEEE Recent Advances in Intelligent Computational Systems (RAICS), Trivandrum, India, 19–21 December 2013; pp. 299–303. [CrossRef]
- Terradas, J.; Pinol, J.; Lloret, F. Climate warming, wildfire hazard, and wildfire occurrence in coastal eastern Spain. *Clim. Chang.* **1998**, *38*, 345–357.
- Alkhatib, A.A. A review on forest-fire detection techniques. *Int. J. Distrib. Sens. Netw.* **2014**, *10*, 597368. [CrossRef]
- Xavier, K.L.B.L.; Nanayakkara, V.K. Development of an Early Fire Detection Technique Using a Passive Infrared Sensor and Deep Neural Networks. *Fire Technol.* **2022**, *58*, 3529–3552. [CrossRef]
- Zhang, F.; Zhao, P.; Xu, S.; Wu, Y.; Yang, X.; Zhang, Y. Integrating multiple factors to optimize watchtower deployment for wildfire detection. *Sci. Total Environ.* **2020**, *737*, 139561. [CrossRef] [PubMed]
- Karthi, M.; Priscilla, R.; Subhashini, G.; Abijith, G.R.; Vinisha, J. Forest fire detection: A comparative analysis of deep learning algorithms. In Proceedings of the 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 5–7 January 2023.
- Kaur, P.; Kaur, K.; Singh, K.; Kim, S. Early Forest Fire Detection Using a Protocol for Energy-Efficient Clustering with Weighted-Based Optimization in Wireless Sensor Networks. *Appl. Sci.* **2023**, *13*, 3048. [CrossRef]
- Mijwil, M.M. *History of Artificial Intelligence*; 2015; Volume 3, pp. 1–8. [CrossRef]
- Xiao, L.; Yan, Q.; Deng, S. Scene classification with improved AlexNet model. In Proceedings of the 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Nanjing, China, 24–26 November 2017; pp. 1–6. [CrossRef]
- Tammina, S. Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. *Int. J. Sci. Res. Publ. (IJSRP)* **2019**, *9*, 143–150. [CrossRef]
- Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [CrossRef]
- Abdusalomov, A.B.; Islam, B.M.S.; Nasimov, R.; Mukhiddinov, M.; Whangbo, T.K. An Improved Forest Fire Detection Method Based on the Detectron2 Model and a Deep Learning Approach. *Sensors* **2023**, *23*, 1512. [CrossRef]
- Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2015**, arXiv:1506.02640.
- Alkhatib, R.; Sahwan, W.; Alkhatieb, A.; Schütt, B. A Brief Review of Machine Learning Algorithms in Forest Fires Science. *Appl. Sci.* **2023**, *13*, 8275. [CrossRef]
- Jayasingh, S.K.; Swain, S.; Patra, K.J.; Gountia, D. An Experimental Approach to Detect Forest Fire Using Machine Learning Mathematical Models and IoT. *SN Comput. Sci.* **2024**, *5*, 148. [CrossRef]
- Rehman, A.; Kim, D.; Paul, A. Convolutional neural network model for fire detection in real-time environment. *Comput. Mater. Contin.* **2023**, *77*, 2289–2307. [CrossRef]
- Ghali, R.; Akhloufi, M.A. Deep Learning Approaches for Wildland Fires Using Satellite Remote Sensing Data: Detection, Mapping, and Prediction. *Fire* **2023**, *6*, 192. [CrossRef]
- Keeping, T.; Harrison, S.P.; Prentice, I.C. Modelling the daily probability of wildfire occurrence in the contiguous United States. *Environ. Res. Lett.* **2024**, *19*, 024036. [CrossRef]
- Li, Y.; Xu, S.; Fan, Z.; Zhang, X.; Yang, X.; Wen, S.; Shi, Z. Risk Factors and Prediction of the Probability of Wildfire Occurrence in the China–Mongolia–Russia Cross-Border Area. *Remote Sens.* **2023**, *15*, 42. [CrossRef]
- Villaverde Canosa, I.; Ford, J.; Paavola, J.; Burnasheva, D. Community Risk and Resilience to Wildfires: Rethinking the Complex Human–Climate–Fire Relationship in High-Latitude Regions. *Sustainability* **2024**, *16*, 957. [CrossRef]
- Marey-Perez, M.; Loureiro, X.; Corbelle-Rico, E.J.; Fernández-Filgueira, C. Different Strategies for Resilience to Wildfires: The Experience of Collective Land Ownership in Galicia (Northwest Spain). *Sustainability* **2021**, *13*, 4761. [CrossRef]
- Myagmar-Ochir, Y.; Kim, W. A survey of Video Surveillance Systems in Smart City. *Electronics* **2023**, *12*, 3567. [CrossRef]
- Pan, H.; Badawi, D.; Cetin, A.E. Computationally Efficient Wildfire Detection Method Using a Deep Convolutional Network Pruned via Fourier Analysis. *Sensors* **2020**, *20*, 2891. [CrossRef]

27. Giglio, L.; Boschetti, L.; Roy, D.P.; Humber, M.L.; Justice, C.O. The Collection 6 MODIS burned area mapping algorithm and product. *Remote Sens. Environ.* **2018**, *217*, 72–85. [CrossRef] [PubMed]
28. Ba, R.; Chen, C.; Yuan, J.; Song, W.; Lo, S. SmokeNet: Satellite Smoke Scene Detection Using Convolutional Neural Network with Spatial and Channel-Wise Attention. *Remote Sens.* **2019**, *11*, 1702. [CrossRef]
29. Larsen, A.; Hanigan, I.; Reich, B.J.; Qin, Y.; Cope, M.; Morgan, G.; Rappold, A.G. A deep learning approach to identify smoke plumes in satellite imagery in near-real time for health risk communication. *J. Expo. Sci. Environ. Epidemiol.* **2021**, *31*, 170–176. [CrossRef]
30. Avazov, K.; Mukhiddinov, M.; Makhmudov, F.; Cho, Y.I. Fire Detection Method in Smart City Environments Using a Deep-Learning-Based Approach. *Electronics* **2022**, *11*, 73. [CrossRef]
31. Mukhiddinov, M.; Cho, J. Smart Glass System Using Deep Learning for the Blind and Visually Impaired. *Electronics* **2021**, *10*, 2756. [CrossRef]
32. Toulouse, T.; Rossi, L.; Celik, T.; Akhloufi, M. Automatic fire pixel detection using image processing: A comparative analysis of rule-based and machine learning-based methods. *Signal Image Video Process.* **2016**, *10*, 647–654. [CrossRef]
33. Jiang, Q.; Wang, Q. Large space fire image processing of improving canny edge detector based on adaptive smoothing. In Proceedings of the 2010 International Conference on Innovative Computing and Communication and 2010 Asia-Pacific Conference on Information Technology and Ocean Engineering, Macao, China, 30–31 January 2010; pp. 264–267.
34. Celik, T.; Demirel, H.; Ozkaramanli, H.; Uyguroglu, M. Fire detection using statistical color model in video sequences. *J. Vis. Commun. Image Represent.* **2007**, *18*, 176–185. [CrossRef]
35. Dimitropoulos, K.; Barmopoulos, P.; Grammalidis, N. Spatio temporal flame modeling and dynamic texture analysis for automatic video-based fire detection. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *25*, 339–351. [CrossRef]
36. Geng, X.; Su, Y.; Cao, X.; Li, H.; Liu, L. YOLOFM: An improved fire and smoke object detection algorithm based on YOLOv5n. *Sci. Rep.* **2024**, *14*, 4543. [CrossRef]
37. Li, P.; Zhao, W. Image fire detection algorithms based on convolutional neural networks. *Case Stud. Therm. Eng.* **2020**, *19*, 100625. [CrossRef]
38. Valikhujayev, Y.; Abdusalomov, A.; Cho, Y.I. Automatic Fire and Smoke Detection Method for Surveillance Systems Based on Dilated CNNs. *Atmosphere* **2020**, *11*, 1241. [CrossRef]
39. Li, T.; Zhao, E.; Zhang, J.; Hu, C. Detection of Wildfire Smoke Images Based on a Densely Dilated Convolutional Network. *Electronics* **2019**, *8*, 1131. [CrossRef]
40. Kutlimuratov, A.; Khamzaev, J.; Kuchkorov, T.; Anwar, M.S.; Choi, A. Applying Enhanced Real-Time Monitoring and Counting Method for Effective Traffic Management in Tashkent. *Sensors* **2023**, *23*, 5007. [CrossRef] [PubMed]
41. Wu, S.; Zhang, L. Using popular object detection methods for real time forest fire detection. In Proceedings of the 11th International Symposium on Computational Intelligence and Design (SCID), Hangzhou, China, 8–9 December 2018; pp. 280–284.
42. Kim, B.; Lee, J. A video-based fire detection using deep learning models. *Appl. Sci.* **2019**, *9*, 2862. [CrossRef]
43. Zhao, L.; Liu, J.; Peters, S.; Li, J.; Oliver, S.; Mueller, N. Investigating the Impact of Using IR Bands on Early Fire Smoke Detection from Landsat Imagery with a Lightweight CNN Model. *Remote Sens.* **2022**, *14*, 3047. [CrossRef]
44. Zhao, Y.Y.; Zhu, J.; Xie, Y.K.; Li, W.L.; Guo, Y.K. Improved Yolo-v3 Video Image Flame Real-Time Detection Algorithm. *J. Wuhan Univ. Inf. Sci. Ed.* **2021**, *46*, 326–334.
45. Abdusalomov, A.; Baratov, N.; Kutlimuratov, A.; Whangbo, T.K. An Improvement of the Fire Detection and Classification Method Using YOLOv3 for Surveillance Systems. *Sensors* **2021**, *21*, 6519. [CrossRef]
46. Park, M.; Ko, B.C. Two-Step Real-Time Night-Time Fire Detection in an Urban Environment Using Static ELASTIC-YOLOv3 and Temporal Fire-Tube. *Sensors* **2020**, *20*, 2202. [CrossRef] [PubMed]
47. Mukhiddinov, M.; Abdusalomov, A.B.; Cho, J. Automatic Fire Detection and Notification System Based on Improved YOLOv4 for the Blind and Visually Impaired. *Sensors* **2022**, *22*, 3307. [CrossRef]
48. Talaat, F.M.; ZainEldin, H. An improved fire detection approach based on YOLO-v8 for smart cities. *Neural Comput. Appl.* **2023**, *35*, 20939–20954. [CrossRef]
49. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. *arXiv* **2022**, arXiv:2209.02976.
50. Wang, C.; Bochkovskiy, A.; Liao, H. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.
51. Shi, P.; Lu, J.; Wang, Q.; Zhang, Y.; Kuang, L.; Kan, X. An Efficient Forest Fire Detection Algorithm Using Improved YOLOv5. *Forests* **2023**, *14*, 2440. [CrossRef]
52. Reis, D.; Kupec, J.; Hong, J.; Daoudi, A. Real-Time Flying Object Detection with YOLOv8. *arXiv* **2023**, arXiv:2305.09972.
53. Saydirasulovich, S.N.; Mukhiddinov, M.; Djuraev, O.; Abdusalomov, A.; Cho, Y.-I. An Improved Wildfire Smoke Detection Based on YOLOv8 and UAV Images. *Sensors* **2023**, *23*, 8374. [CrossRef]
54. Girdhar, R.; Carreira, J.; Doersch, C.; Zisserman, A. Video Action Transformer Network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 9–15 June 2019; pp. 244–253.
55. Yang, F.; Yang, H.; Fu, J.; Lu, H.; Guo, B. Learning Texture Transformer Network for Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 5791–5800.

56. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. In *Computer Vision—ECCV*; Springer International Publishing: Cham, Switzerland, 2020; pp. 213–229.
57. Ye, L.; Rochan, M.; Liu, Z.; Wang, Y. Cross-Modal Self-Attention Network for Referring Image Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 9–15 June 2019; pp. 10502–10511.
58. He, X.; Chen, Y.; Lin, Z. Spatial-Spectral Transformer for Hyperspectral Image Classification. *Remote Sens.* **2021**, *13*, 498. [CrossRef]
59. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
60. Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; Jégou, H. Training data-efficient image transformers & distillation through attention. *arXiv* **2020**, arXiv:2012.12877.
61. Valanarasu, J.M.J.; Oza, P.; Hacihaliloglu, I.; Patel, V.M. Medical Transformer: Gated Axial-Attention for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.10662.
62. Abdusalomov, A.B.; Mukhiddinov, M.; Kutlimuratov, A.; Whangbo, T.K. Improved Real-Time Fire Warning System Based on Advanced Technologies for Visually Impaired People. *Sensors* **2022**, *22*, 7305. [CrossRef]
63. Pandey, B.; Pandey, D.K.; Mishra, B.P.; Rhmann, W. A comprehensive survey of deep learning in the field of medical imaging and medical natural language processing: Challenges and research directions. *J. King Saud Univ. Comput. Inf. Sci.* **2022**, *34*, 5083–5099. [CrossRef]
64. Mukhiddinov, M.; Djuraev, O.; Akhmedov, F.; Mukhamadiyev, A.; Cho, J. Masked Face Emotion Recognition Based on Facial Landmarks and Deep Learning Approaches for Visually Impaired People. *Sensors* **2023**, *23*, 1080. [CrossRef] [PubMed]
65. Jocher, G.; Chaurasia, A.; Qiu, J. *Ultralytics YOLO*; Version 8.0.0; Ultralytics: Los Angeles, CA, USA, 2023. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 12 January 2024).
66. Wang, X.; Huang, T.; Gonzalez, J.; Darrell, T.; Yu, F. Frustratingly Simple Few-Shot Object Detection. In Proceedings of the 37th International Conference on Machine Learning, Virtual, 13–18 July 2020; pp. 9919–9928.
67. Xu, X.; Zhang, H.; Ma, Y.; Liu, K.; Bao, H.; Qian, X. TransDet: Toward Effective Transfer Learning for Small-Object Detection. *Remote Sens.* **2023**, *15*, 3525. [CrossRef]
68. Finn, C.; Abbeel, P.; Levine, S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1126–1135.
69. Avazov, K.; Jamil, M.K.; Muminov, B.; Abdusalomov, A.B.; Cho, Y.-I. Fire Detection and Notification Method in Ship Areas Using Deep Learning and Computer Vision Approaches. *Sensors* **2023**, *23*, 7078. [CrossRef] [PubMed]
70. Norkobil Saydirasulovich, S.; Abdusalomov, A.; Jamil, M.K.; Nasimov, R.; Kozhamzharova, D.; Cho, Y.-I. A YOLOv6-Based Improved Fire Detection Approach for Smart City Environments. *Sensors* **2023**, *23*, 3161. [CrossRef] [PubMed]
71. Ergasheva, A.; Akhmedov, F.; Abdusalomov, A.; Kim, W. Advancing Maritime Safety: Early Detection of Ship Fires through Computer Vision, Deep Learning Approaches, and Histogram Equalization Techniques. *Fire* **2024**, *7*, 84. [CrossRef]
72. Sun, C.; Shrivastava, A.; Singh, S.; Gupta, A. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 843–852.
73. Chen, W.-F.; Ou, H.-Y.; Liu, K.-H.; Li, Z.-Y.; Liao, C.-C.; Wang, S.-Y.; Huang, W.; Cheng, Y.-F.; Pan, C.-T. In-Series U-Net Network to 3D Tumor Image Reconstruction for Liver Hepatocellular Carcinoma Recognition. *Diagnostics* **2021**, *11*, 11. [CrossRef] [PubMed]
74. Shah, S.M.; Sun, Z.; Zaman, K.; Hussain, A.; Ullah, I.; Ghadi, Y.Y.; Khan, M.A.; Nasimov, R. Advancements in Neighboring-Based Energy-Efficient Routing Protocol (NBEER) for Underwater Wireless Sensor Networks. *Sensors* **2023**, *23*, 6025. [CrossRef]
75. Aldughayfiq, B.; Ashfaq, F.; Jhanjhi, N.Z.; Humayun, M. YOLO-Based Deep Learning Model for Pressure Ulcer Detection and Classification. *Healthcare* **2023**, *11*, 1222. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

CODAS–Hamming–Mahalanobis Method for Hierarchizing Green Energy Indicators and a Linearity Factor for Relevant Factors’ Prediction through Enterprises’ Opinions

Georgina Elizabeth Riosvelasco-Monroy, Iván Juan Carlos Pérez-Olguín *, Salvador Noriega-Morales, Luis Asunción Pérez-Domínguez, Luis Carlos Méndez-González and Luis Alberto Rodríguez-Picón

Institute of Engineering and Technology, Department of Industrial and Manufacturing Engineering, Universidad Autónoma de Ciudad Juárez, Ciudad Juárez 32310, Mexico; georgina.riosvelasco@uacj.mx (G.E.R.-M.); luis.mendez@uacj.mx (L.C.M.-G.)

* Correspondence: ivan.perez@uacj.mx

Abstract: As enterprises look forward to new market share and supply chain opportunities, innovative strategies and sustainable manufacturing play important roles for micro-, small, and mid-sized enterprises worldwide. Sustainable manufacturing is one of the practices aimed towards deploying green energy initiatives to ease climate change, presenting three main pillars—economic, social, and environmental. The issue of how to reach sustainability goals within the sustainable manufacturing of pillars is a less-researched area. This paper’s main purpose and novelty is two-fold. First, it aims to provide a hierarchy of the green energy indicators and their measurements through a multi-criteria decision-making point of view to implement them as an alliance strategy towards sustainable manufacturing. Moreover, we aim to provide researchers and practitioners with a forecasting method to re-prioritize green energy indicators through a linearity factor model. The CODAS–Hamming–Mahalanobis method is used to obtain preference scores and rankings from a 50-item list. The resulting top 10 list shows that enterprises defined nine items within the economic pillar as more important and one item on the environmental pillar; items from the social pillar were less important. The implication for MSMEs within the manufacturing sector represents an opportunity to work with decision makers to deploy specific initiatives towards sustainable manufacturing, focused on profit and welfare while taking care of natural resources. In addition, we propose a continuous predictive analysis method, the linearity factor model, as a tool for new enterprises to seek a green energy hierarchy according to their individual needs. The resulting hierarchy using the predictive analysis model presented changes in the items’ order, but it remained within the same two sustainable manufacturing pillars: economic and environmental.

Keywords: Mahalanobis distance; green energy supply chain; MCDM; sustainable manufacturing; predictive analysis model; CODAS; Hamming distance

Citation: Riosvelasco-Monroy, G.E.; Pérez-Olguín, I.J.C.; Noriega-Morales, S.; Pérez-Domínguez, L.A.; Méndez-González, L.C.; Rodríguez-Picón, L.A. CODAS–Hamming–Mahalanobis Method for Hierarchizing Green Energy Indicators and a Linearity Factor for Relevant Factors’ Prediction through Enterprises’ Opinions. *Processes* **2024**, *12*, 1070. <https://doi.org/10.3390/pr12061070>

Academic Editors: Sheng Du, Li Jin, Xiongbo Wan and Zixin Huang

Received: 1 May 2024

Revised: 17 May 2024

Accepted: 17 May 2024

Published: 23 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Innovation is seen as an activity that encourages growth, development, and continuous improvement in each department of an enterprise [1]. It is considered a business strategy that enhances the organizational structure for the development of new ideas and the creation or improvement of products, services, and processes for acquiring a competitive advantage [2]. Enterprises must define, develop, and maintain their supply chain to guarantee an innovative and disruptive product.

Disruption presents uncertainty and a ripple effect in the supply chain long after the event has passed [3]. As Ref. [2] describes, innovation and disruption offer great opportunities for enterprises to achieve competitive advantage and growth for the company itself. Decision makers need to focus on strategies for innovative and disruptive events. As described in [3], a disruptive supply chain involves a number of factors such as information

sharing and interconnected physical and cyber infrastructure, or, as suggested in [4], horizontal collaboration strategies to design and develop joint projects in different areas such as logistics, R&D, and knowledge transfer, among others.

The industrial sector is one sector with a rapidly changing challenge and has been identified as the main participant in green energy consumption [5]. For micro-, small, and mid-sized enterprises (MSMEs), the new challenge lies in creating a motivated environment where human capital can increase their abilities and knowledge [6] in order to use their actual resources effectively and efficiently to obtain aggregated value in their offered products and services [7]. The academic literature has identified specific elements to increase the use of innovation and disruption as an enterprise's business strategy, from implementation changes like costs, technology diffusion, intellectual property protection, finance performance, investments, and commercialization through alliances and collaboration ventures where R&D activities, human capital, and knowledge transfer and sources are taken into account to deliver firms' innovation and disruption [1,8–10].

To enable sustainable manufacturing, enterprises must consider environmental and economic implications within the supply chain. As explained by [11], enterprises focus on strategies to increase the efficiency of manufacturing processes, the use of raw materials, or the use of recovered or reused materials, as well as horizontal collaboration business models to engage in innovative projects for sustainable value creation in manufacturing [12]. Key factors include the involvement of decision makers in the considerable number of factors that need to be taken into account, from production planning, manufacturing processes, and innovation to supply chain design and products. Other factors include the selection of raw materials from the perspective of recycling, reusing, or re-furbishing materials, integrating or collaborating with processes that use energy effectively, and working with a logistics strategy that uses low fossil fuel.

Deploying innovation and disruption goes hand-in-hand with green energy initiatives. Faced with the challenge of changing climate damages, countries are working towards several development goals, strategies, and agreements, from the Sustainable Development Goals of the United Nations 2030 Agenda with 17 sustainable goals, to the Paris Agreement, a treaty with the objective of economic and social transformation to be achieved through collaboration in financial assistance, technology development and transfer, and capacity building [13–16]. These strategies aim to incentivize countries towards a green energy change, from production to consumption, and replace the use of fossil fuels [17–19].

To the best of our knowledge, there is no research on existing studies that outline the main green initiatives used to enhance sustainable manufacturing pillars. Furthermore, there is no literature review on the green indicators and measurements that could enhance the sustainable manufacturing pillars—economic, social, and environmental.

Given the importance of achieving sustainable manufacturing considering climate change, the current paper presents an in-depth literature review focused on the sustainable manufacturing pillars, also known as the triple bottom line [13,20]. Overall, recent studies do not provide a comprehensive review of how green energy indicators and their measurements and items can be matched with the three pillars of sustainable manufacturing. The investigation allows for companies in industrial activities to deploy specific actions to enhance sustainable manufacturing, as well as adopt a predictive analytic model for new hierarchization of the items for future research, promoting green energy indicators in sustainable manufacturing. Thus, this review was conducted to fill the gaps in the importance of these indicators and to answer the following questions:

- How can the green energy indicators be implemented within sustainable manufacturing pillars?
- Which are the green energy indicators identified as of higher priority from a comparative correlation from the MCDM methodology?
- How can new enterprises determine the hierarchy of green energy indicators?

This paper's main purpose is to provide a hierarchy of green energy indicators, their measurements, and items aligned with the three pillars of sustainable manufacturing. This is completed from the perspective of the MCDM to implement it as an alliance strategy towards sustainable manufacturing. Moreover, it provides researchers and practitioners with a forecasting method to reprioritize green energy indicators through a linearity factor model. To answer the above research questions, the following objectives are set:

- Present the proposed CODAS–Hamming–Mahalanobis method developed by [12].
- Present the proposed green energy indicators, their measurements, and items identified by [21].
- Develop a green energy matrix with a Likert scale, including the green energy indicators within the three sustainable manufacturing pillars.
- Provide a linearity factor model to forecast the hierarchy of green energy indicators from an additional business' opinion.

This paper is organized as follows: Section 2 presents the results of the literature review of sustainable manufacturing pillars and green energy indicators. Section 3 contains the methodology, followed by the discussion of the empirical findings in Section 4. Section 5 concludes and outlines the theoretical contributions. Finally, Section 6 presents and discusses the practical implications, limitations, and future research.

2. Literature Review

As a result of the above, this paper focuses on green energy indicators that firms can use efficiently and effectively toward sustainable manufacturing. Green energy can have a significant effect on the operational costs and performance achievements in manufacturing processes, among others, but the initiative can fail due to several challenges and barriers [22]. To gain an understanding, the authors retrieved contributions from green energy indicators, specifically those suggested by [21]. Then, a literature analysis of sustainable manufacturing pillars is presented.

2.1. Green Energy Indicators, Measurements and Their Items

Manufacturing businesses are considered responsible for 36% of carbon dioxide emissions and consume approximately a third of global energy production [17]. For this matter, decision makers are aiming for sustainable development because stakeholders and end customers are from developed, developing, and emerging economies. To make better decisions in implementing sustainable factors, managers must pay attention to a more comprehensive approach integrating economic, social, and environmental aspects [23,24]. Riosvelasco et al. [21] present a literature review to identify green energy indicators, measurements, and their items for manufacturing applications towards sustainability, as shown in Table 1. The authors applied the PRISMA 2020 methodology to analyze previous research and identify green energy indicators to ease Mexican companies in the industrial sector to deploy green energy sustainable practices.

Table 1 shows six indicators: (1) technology and innovation, (2) geographical aspects, (3) investment, (4) government regulations, (5) emissions, and (6) sustainable practices. Within these six indicators, each encompasses a total of 18 measurements, and the latter accumulate a total of 50 items. These indicators and their measurements and items are identified to promote actions for Mexican companies to adapt and create a sustainable manufacturing environment. As enterprises look forward to new market shares and supply chain opportunities, innovation plays an important role for MSMEs worldwide. As described by [25], innovative enterprises must create or find a business model that balances everyday activities, innovative strategies, and sustainable manufacturing.

Table 1. Green energy indicators, measurements and items [21].

Green Energy				
Indicators	Measurements	Code	Items	
Technology and Innovation	Technology Capacity	TI01	Improvements in production processes	
		TI02	Development of new production processes towards efficient energy usage	
		TI03	Preventive actions in production processes	
		TI04	Preventive actions in machinery, computer equipment, cooling systems	
		TI05	Have an R&D department	
		TI06	Joining a cluster to participate in trainings, collaborations, and programs for fossil-fuel substitution	
	Collaboration Capacity	TI07	Support in the diffusion of green energy goods and services	
		TI08	Collaborate with research centers or universities in innovative projects toward green energy usage	
		TI09	Participate in industrial symbiosis	
		TI10	Collaboration with enterprises of the same echelon to create new production processes for green energy usage	
		TI11	Protect innovations with intellectual property	
Geographical Aspects	Soil	GA01	Knowing types of green energy that are generated in the area where the enterprise is located	
		GA02	Knowing the destination of residual loads	
	Water	GA03	Have systems for the efficient treatment of water residuals	
		GA04	Using treated water in production processes	
		GA05	Solar cells for generation and usage	
	Electric Energy	GA06	Wind generator for generation and usage	
		GA07	Rely on fossil-fuel usage for machinery, production processes	
Investment	Human Capital	I01	Invest in human capital with knowledge and capabilities in green energy	
		I02	Invest in training for the enterprise's human capital in green energy	
	Energy Efficiency	I03	Identify opportunity areas in the efficient use of energy in machinery, production processes, and transportation	
		I04	Invest in energy efficiency projects in production processes, machinery, and transportation	
Government Regulations	Supply Chain	I05	Integrate raw materials from suppliers that have production processes with an energy efficient usage	
		I06	Buy machinery and equipment that have energy-efficient usage technology	
		GR01	Generate environmental policy regarding green energy usage	
	Policy	GR02	Publish and disseminate about green energy	
		GR03	Facilitate linkage between enterprises and government entities for green energy implementation	
		GR04	Risk assessment in green energy investments	
		GR05	Submit tax offset plans by investing in green energy	
		GR06	Provide training for the usage of green energy in production processes	
		GR07	Offer funds and grants in R&D in production processes with green energy	
		GR08	Promote industrial symbiosis so enterprises implement green energy	
Emissions	Indicators	E01	Reporting of direct emissions from stationary sources	
		E02	Reporting of direct emissions from mobile sources	
		E03	Reporting of indirect emissions	
Sustainable Processes	Residual	SP01	No use of hazardous materials or virgin materials for products in the production process	
		SP02	Production processes designed to avoid waste	
	Reduce	SP03	Reduced material use per unit of production (increased dematerialization)	
		SP04	Develop products with better design, life cycle, durability, and quality in raw materials	
		SP05	Reuse material in production	
	Reuse	SP06	Use of wastes as inputs	
		SP07	Resale of products with minimal defects, unsold products in inventory, unused products	
	Repair	SP08	Use surplus components in inventory and adapt them for another function	
		SP09	Repair and maintenance of products	
		SP10	Collect defective products at centers (branches or points of sale) through the manufacturer or by a third-party company for repairs	
		Refurbish	SP11	Modular product design for ease of disassembly
			SP12	Disassembly of the general structure of the product, checking, cleaning and potentially replacing some components
		Recycle	SP13	Recovering the product at the end of its useful life
			SP14	Ensure the use of recycled raw materials
		Recover	SP15	Capture energy embodied in waste (incineration, use of biomass, among others)

2.2. Sustainable Manufacturing

For enterprises, reaching sustainability goals is an everyday challenge because of the fine decisions entrepreneurs have to make every day. Managers have difficulty selecting and working on strategies for implementing sustainable manufacturing because sustainability presents a variety of factors [17,24]. The path to stabilize the disruptive relationship between industry and the negative impact on the 3Ps (people, planet, and profit) involves a holistic approach relying on economic operational efficiency, environmental compliance, communication, and legal regulations [24].

A literature review was conducted according to the following criteria: (a) year range from 2020 to 2024; (b) article type as review articles; (c) keywords such as “sustainable manufacturing pillars”, “triple bottom line” AND “sustainable manufacturing”, “sustainability” AND “manufacturing”; and (d) access type as open access. For exclusion criteria, the following were taken into account: keywords must be in title, abstract, and keywords. The results obtained ranged from systematic reviews to identify drivers and motives for sustainable manufacturing [26]; a framework for analyzing the product life cycle regarding product design and manufacturing system selection [27] to the identification of several enablers for the adoption of sustainable manufacturing [28].

For manufacturing businesses to work and implement a direct strategy towards sustainable manufacturing, research has been focusing on sustainability measurement that follows activities as the backbone within the supply chain [20], where decision makers have to pay attention to the interaction between society, environment, and companies [29]. Sustainable manufacturing presents three pillars which are: (1) economic, which includes criteria that take into account costs; (2) social, which takes into account the human elements in terms of employees and community; and (3) environmental, criteria that involve reducing emissions, waste, and spike energy, among others [20,24,27,28,30,31], as shown in Table 2.

Table 2. Sustainable manufacturing pillars [20,27,28,30].

Economic Factors	Social Factors	Environmental Factors
Hardware cost	Stakeholder participation	Energy consumption
Applied technology	Employment issues	Holding environmental standards
Software cost	Personnel safety	Environmental planning
Mean time between failure	Personnel training	Inside toxic emissions
Risk level of the system	Acceptance by personnel	Raw material consumption
Research and development cost	Development of management and engineering expertise	Waste cleaning cost
Customer satisfaction	Personnel health	Waste type
Product variety	Personnel wage	Toxic emissions to air, soil, and water
Creativity	Government regulations	Resource availability
Product development stage	Holding related to social standards	Environmental management systems
Competitive enhancement	Poverty	Natural resources efficiency
Current reliability level	Education	Emissions quality
Reliability improvement plan	Culture	
Access difficulty	Lifestyle	
Quality of product/service	Social harmony	
Time efficiency		
Mean time of repair		
Manufacturing system type		
Detailed production scheduling		
System design		
Reparability		
Failures type severity		
Technical feasibility		
Flexibility		
Return on investment		
Demand urgency		

Table 2. Cont.

Economic Factors	Social Factors	Environmental Factors
Spare machine availability		
Spare parts availability		
Tear and wear rate		
Lead time		
Lost production cost		
Personnel training cost		
Wealth		

From this point forward, the authors present a match between the green energy indicators and the three sustainable manufacturing pillars. This decision-making matrix has been designed by considering three sustainable manufacturing pillars, selected for its approach of an innovation perspective from which enterprises can transition towards green energy usage. For the matching process, the three pillars are taken into account. (1) Economic: all firm's costs, repairs, and enhancements, among other factors in different departments; (2) social: factors contributing to personnel, stakeholders, and government; and (3) environmental: factors regarding emissions, consumption and management systems. Table 3 shows the items that involve criteria within each sustainable manufacturing pillar.

Table 3. Design of decision-making matrix.

Indicators	Green Energy		Sustainable Manufacturing Pillars		
	Measurements	Items' Code	Economic	Social	Environmental
Technology and Innovation	Technology Capacity	TI01	X		
		TI02	X		
		TI03			X
		TI04			X
		TI05	X		
		TI06	X		
		TI07			X
	Collaboration Capacity	TI08	X		
		TI09			X
		TI10			X
		TI11	X		
Geographical Aspects	Soil	GA01			X
		GA02			X
	Water	GA03			X
		GA04			X
	Electric Energy	GA05			X
		GA06			X
		GA07			X
Investment	Human Capital	I01		X	
		I02		X	
	Energy Efficiency	I03	X		
		I04	X		
	Supply Chain	I05	X		
		I06	X		
Government Regulations	Policy	GR01			X
		GR02		X	
		GR03		X	
		GR04			X
		GR05			X
	Incentive	GR06		X	
		GR07		X	
		GR08	X		

Table 3. Cont.

Indicators	Green Energy		Sustainable Manufacturing Pillars		
	Measurements	Items' Code	Economic	Social	Environmental
Emissions	Indicators	E01			X
		E02			X
		E03			X
	Residual	SP01			X
		SP02	X		
		SP03			X
	Reduce	SP04	X		
		SP05			X
		SP06			X
	Reuse	SP07			X
		SP08			X
		SP09			X
	Repair	SP10			X
		SP11			X
		SP12	X		
Refurbish	SP13	X			
	SP14			X	
	SP15			X	

3. Methodology

3.1. Research Model

Contributing to the academic literature on sustainable manufacturing towards the use of green energy indicators, measurements, and items, this paper presents a hybrid analysis for hierarchizing green energy indicators using CODAS techniques with distinct distance equations—Hamming and Mahalanobis distances. These center data from an enterprise’s evaluation survey whose experience in different productive and service sectors enhances green energy indicators in multi-criteria decision making. Additionally, we present a linearity factor model for predicting green energy indicators from future enterprise panels. Figure 1 conceptualizes the research model followed to reach this paper’s main objective, taking into account the assigned Likert Scale, the CODAS–Hamming–Mahalanobis method, and the predictive analysis model.

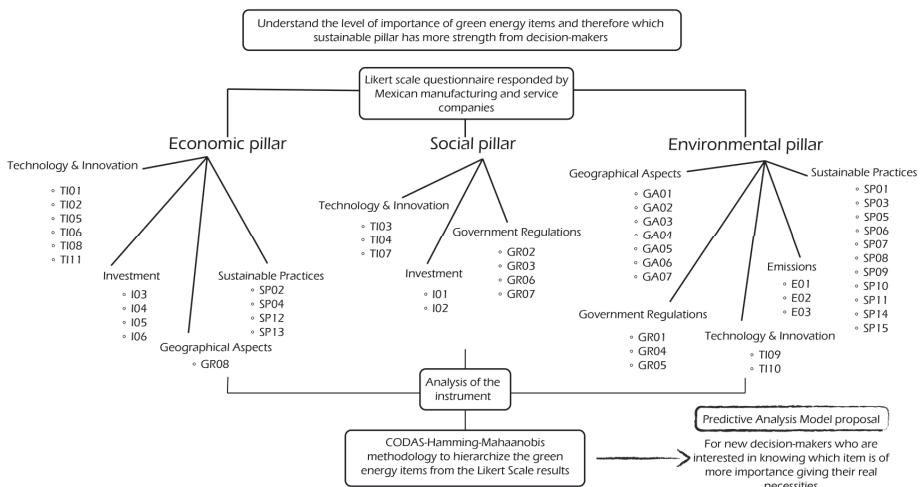


Figure 1. Research model.

3.2. Instrument for Data Collection

The instrument for data collection was then designed from the decision-making matrix, shown in Table 3, that served as a basis for the questionnaire. Table 4 presents the questionnaire, which consisted of 17 questions, divided into 7 sections: (1) demographic data, (2) technology and innovation, (3) geographic aspects, (4) investment, (5) government regulations, (6) emissions, and (7) sustainable practices. The instrument was designed to be answered by a businessperson, manager, supervisor, or person appointed from micro, small, mid-size, or big enterprises.

Table 4. Instrument for data collection.

Green Energy				Likert Scale ¹					
Indicators	Measurements	Items' Code	Question	1	2	3	4	5	
Technology and Innovation	Technology Capacity	TI01	Create improvements in the current production processes for the efficient use of energy in the company.						
		TI02	Develop new production processes for the efficient use of energy.						
		TI03	Maintain preventive actions in the production processes for the efficient use of electric energy.						
	Collaboration Capacity	TI04	Maintain preventive actions in machinery, computer equipment, refrigeration, among others, for the efficient use of electrical energy.						
		TI05	To have a Research and Development department.						
		TI06	Join a cluster to participate in trainings, collaborations and programs for the substitution of fossil fuels for green energies.						
		TI07	Support in the dissemination of products or services for the use of green energies.						
		TI08	Collaborate with Research Centers or Universities for innovation projects with the use of green energies.						
		TI09	Participate in industrial symbiosis (e.g., waste of one company is the raw material of the other company).						
		TI10	Collaborate with companies in the same line of business to create new productive processes for the use of green energies.						
		TI11	Protect innovations with Intellectual Property.						
Geographical Aspects	Soil	GA01	To know the types of green energies generated in the area (city, state) where the company is located.						
	Water	GA02	To know the destination of residual loads (drainage).						
		GA03	To have systems for the efficient use of water (e.g., treatment plant, recycling, reuse systems).						
	Electric Energy	GA04	Use treated water in the company's production processes.						
		GA05	Use solar cells for the generation and use of electrical energy.						
		GA06	Use wind generator for the generation and use of electrical energy.						
		GA07	Dependence on the use of fossil fuels for production processes, machinery, computer equipment, etc.						
Human Capital	I01	Investing in Human Capital with skills and knowledge in green energies.							
	I02	Invest in Human Capital training to increase skills and knowledge in green energy.							
Investment	Energy Efficiency	I03	Identify areas of opportunity in production processes, machinery, computer equipment, transportation for the efficient use of green energy.						
		I04	Invest in energy efficiency projects in production processes, machinery, and transportation, among others.						
	Supply Chain	I05	Integrate raw materials and other materials from companies that use green energy in their processes.						
		I06	Purchase machinery and equipment using green energies.						

Table 4. Cont.

Green Energy		Likert Scale ¹						
Indicators	Measurements	Items' Code	Question	1	2	3	4	5
Government Regulations	Policy	GR01	Create environmental regulation on the use of green energies.					
		GR02	Publish and disseminate green energy.					
		GR03	Linking companies with government agencies to implement green energies.					
	Incentive	GR04	Analyze the risks of green energy investment projects.					
		GR05	Submit tax offset plans when investing in green energies.					
		GR06	Provide training on the use of green energies in production processes.					
		GR07	Offer funds and subsidies for research and development of green energy production processes.					
		GR08	Promote industrial symbiosis so that companies implement actions for the efficient use of green energy in their production processes.					
Emissions	Indicators	E01	Report direct emissions from stationary sources (heavy machinery, boiler, industrial processes, basic furnaces, wastewater treatment plant, emergency plant, etc.).					
		E02	Report direct emissions from mobile sources (goods or personnel transport vehicles, forklifts).					
		E03	Report on indirect emissions (electricity and thermal energy consumption).					
Sustainable Practices	Residual	SP01	No use of hazardous materials or virgin material for products in the production process.					
		SP02	Production processes designed to avoid waste.					
	Reduce	SP03	Reduced material use per unit of production (increased dematerialization).					
		SP04	Develop products with better design, life cycle, durability, quality in raw materials.					
		SP05	Reusing material in production.					
	Reuse	SP06	Use of waste as input material.					
		SP07	Resale of products with minimal defects, unsold products in inventory, unused products.					
	Repair	SP08	Use surplus components in inventory and adapt them for another function.					
		SP09	Product repair and maintenance.					
		SP10	Collect defective products at centers (branches or points of sale) through the manufacturer or by a third-party company for repairs.					
	Refurbish	SP11	Modular product design for easy disassembly.					
		SP12	Disassembly of the general structure of the product, checking, cleaning and potentially replacing some components.					
	Recycle	SP13	Recovering the product at end of life.					
		SP14	Ensuring the use of recycled raw materials.					
	Recover	SP15	Capture energy embodied in waste (incineration, use of biomass, among others).					

¹ Likert Scale: 1—Strongly disagree; 2—Disagree; 3—Indifferent; 4—In agreement; 5—Totally agree.

To give certainty to the questionnaire for the fulfillment of the objectives outlined, a validation process was carried out. Expert judgment was used—a process that requires an accurate, efficient, methodological, and statistical interpretation of the results. The expert panel selection was performed by taking into account the experience and knowledge of sustainability practices, manufacturing processes, and energy efficiency strategies. The expert panel was grouped by three experts, thus obtaining an average Cohen's Kappa equal to 0.95 with perfect agreement between the raters.

Afterward, the questionnaire was made on Limesurvey online software so it could be in a more user-friendly form to be answered by the survey respondents. The questionnaire can be filled out at the following URL: <https://energiasverdes.limesurvey.net/686967?lang=es> (accessed on 21 May 2024) (it is available in Spanish, but will be made available in English to obtain responses from a broader sample).

3.3. Data Collection Process

The instrument for data collection was made available from August 2023 to March 2024, through a digital platform known as Limesurvey. A Likert scale was used to measure the level of interest of each enterprise's decision maker towards each green energy indicator, their measurements and items, with the meanings totally agree (5), in agreement (4), indifferent (3), disagree (2), and strongly disagree (1), as shown in Table 4.

The survey was presented to a Mexican organization and a cluster, Cámara Nacional de la Industria de la Transformación, North Zone (CANACINTRA, for its acronym in Spanish) and Clúster de Energía Chihuahua (CECH, for its acronym in Spanish), to obtain a higher response rate from manufacturing and service companies. Both organizations are in Juarez city, Chihuahua, a city of great importance to the manufacturing sector in the country. Additional responses were received from companies from Chihuahua city because the link was shared to known businessmen. From the elapsed time, from August 2023 to March 2024, a total of 116 survey responses were received. Figure 2 shows the diagram flow for the data collected.

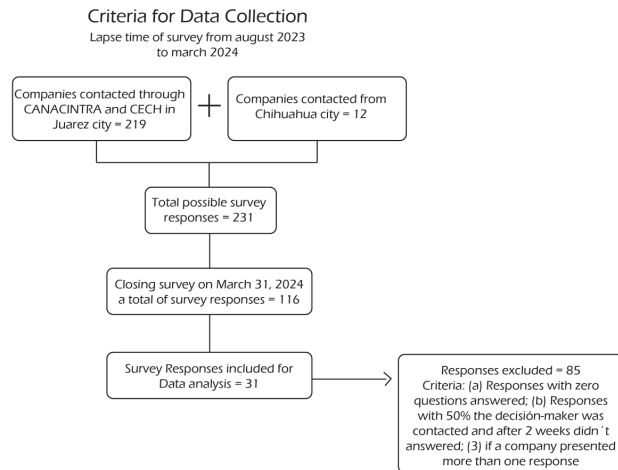


Figure 2. Data collection diagram flow.

As mentioned before, certain criteria were taken into account for excluding companies' responses, such as: (1) only one decision maker from the company could respond to the survey—more than one survey response from the same company were not taken into account; (2) for responses received with 50% of survey responded to, the decision maker was contacted asking them to finish the survey—if they did not respond within a period of two weeks, the response was eliminated; (3) responses with zero items answered and no contact data were discarded. From the demographic section, the following information was obtained from the decision makers: 61.29% were companies located in Juarez city, and 38.70% in Chihuahua City. Table 5 presents demographic data regarding the 31 companies that responded to the survey.

From the decision makers that responded to the survey, their job titles ranged from Assistant Director (1), Quality Coordinator (1), Leader (1), Manager (2), General Manager (6), Director (11), CEO (2), Owner (3), and Teacher (4). Moreover, their level of education was Technician (1), Bachelor's degree (16), Specialization (2), and Master (12).

The next step was to analyze the rest of the six sections through an integrated MCDM method, named CODAS–Hamming–Mahalanobis, presented by [12].

Table 5. Demographic data from companies.

Sector	# of Respondents	Firm Size
Commerce	4	Small = 4
Construction	3	Micro = 1; Small = 1; Big = 1
Government	2	Big = 2
Automotive industry	1	Small = 1
Beverages and Food industry	1	Big = 1
Wood industry	1	Micro = 1
Electric industry	3	Micro = 1; Mid-size = 2
Metal-mechanic industry	1	Micro = 1
Metallurgic industry	1	Big = 1
Residual, urban waste, and special waste management industry	3	Micro = 1; Small = 1; Big = 1
Education services	3	Micro = 1; Small = 1; Big = 1
Other	8	Micro = 2; Small = 1; Mid-size = 3; Big = 2

3.4. Analytic Instruments

The authors of [12] proposed Mahalanobis distance to be used as the second distance measure for the CODAS methodology. The resulting methodology presented better results from the original primary and secondary distance measures (e.g., Euclidean and Taxicab distances), gaining better numbers for Hamming and Mahalanobis distances. From this point on, the present article uses this proposed methodology, named CODAS–Hamming–Mahalanobis to obtain, analyze, and identify green energy items from a sample size defined by the elapsed time. The main distance presented below is the Mahalanobis distance, because Ref. [12] proposed three distance variants, and one of them, the robust variant, showed better results.

Mahalanobis Distance Robust Variant

P.C. Mahalanobis proposed Mahalanobis distance in 1936, where it was defined as a set of parameters of a particular cluster replacing the population via sample statistics [32]. Research presents Mahalanobis distance as a statistical tool that focuses on measurable variables for data-driven decision-making [33,34], as well as a statistical technique measuring a distance point from the center of a multivariate normal distribution [35]. Defined as a distance that measures the distance between variables, different patterns can be analyzed concerning the population parameters [33,36,37], as shown in Equation (1).

Mahalanobis distance:

$$D = \sqrt{(X_i - \bar{X})S^{-1}(X_i - \bar{X})}, \quad (1)$$

where:

1. X_i represents a row vector of a multivariate measurement being observed;
2. S^{-1} is the covariance matrix of the sample;
3. \bar{X} is the mean of the sample.

As globalized manufacturing processes are integrated with cyber-physical features and Industry 4.0, more data are received. Due to this situation, disruptions in statistical themes present an opportunity for researchers to propose Mahalanobis distance measurements with a series of functionalities. The reason for this usage is that Mahalanobis distance can detect normal to abnormal data from a series of multivariate data [37–39]. Alternatively, as described by [33], highlights and analysis outliers are observations that are left out of the known sample and are not consistent with the first sample of data clouding the primary analysis. This presents a wide range of applicability such as multivariate analysis techniques, classification techniques, clustering analysis, discriminant analysis,

and goodness of fit tests, among others [33,37]. In the health system, it is used to classify Chronic Myeloid Leukemia neutrophils [40], studying circulating biomarkers like lipid metabolism, inflammation, and microbiome [41], or for calculating a symmetry metric from segmented NIR spectra [42]. It is also applied in environmental sciences by measuring data and identifying differences to quantify species–environment relationships [35], detecting damage effects on structural health monitoring [43], and detecting and measuring risk levels in the international energy market system [44].

It is differentiated from Euclidean distance by showing correlations and data distribution [33,37]. One of the functionalities presented in the literature is the Mahalanobis–Taguchi System, which is a forecasting method for detecting outliers from the mechanical state of a logistics system for potential failure prediction [45,46]. This can be used to distinguish between criteria selection for strategies within educational institutions [47], and as a forecasting method for consumer satisfaction for vehicle handling [33]. On the other hand, FAST-MCD has been applied in order to achieve a more accurate result with larger datasets [48], and a minimum covariance determinant is used as an approach for multivariate outliers’ detection [49]. Alternatively, Ref. [50] evaluates the resilience system through a TOPSIS model of weighted Mahalanobis distance or an ensemble method for unsupervised learning by applying three options: (1) univariate Mahalanobis distance, (2) Mahalanobis-square distance, and (3) local Mahalanobis-square distance.

As a multi-criteria decision-making (MCDM) tool, Mahalanobis distance facilitates an analysis of a decision matrix of a MCDM problem, such as green energy indicators and their measurements towards sustainable manufacturing in enterprises, as shown in Table 6.

Table 6. Multi-criteria decision matrix.

		Criterion				
Variables	G_i/C_j	C_1	C_2	C_3	...	C_n
	G_1	G_{11}	G_{12}	G_{13}	...	G_{1n}
	G_2	G_{21}	G_{22}	G_{23}	...	G_{2n}
	G_3	G_{31}	G_{32}	G_{33}	...	G_{3n}
	\vdots	\vdots	\vdots	\vdots		\vdots
	G_m	G_{m1}	G_{m2}	G_{m3}	...	G_{mn}

In [12], the authors proposed three variants of the Mahalanobis distance as the secondary measurement distance for hierarchizing horizontal collaboration factors within the CODAS methodology. One of the variants, $MDrv(Cov_rij; ns)$, presented a higher proportion in concordance compared to the Taxicab distance and the other two MD variants proposed. From this point forward, this paper presents the Mahalanobis distance robust variant, $MDrv(Cov_rij; ns)$, for green energy indicators to reach sustainable manufacturing. As presented in [12], this paper follows a specific MCDM technique to determine green energy indicators and their measurements through hierarchical analysis within the sample to successfully carry out decision-making activity.

The Mahalanobis distance robust variant ($Cov_rij; ns$) presented by [12] follows the Mahalanobis distance original equation (Equation 1), modifying the original terms as shown in Equation (2):

$$MDrv = \sqrt{(Cov_Max_{cr} - Cov_Min_{ag})C^{-1}(Cov_Max_{cr} - Cov_Min_{ag})}, \quad (2)$$

where:

4. Cov_Max_{cr} represents the average of the aggregated matrix maximum score of the enterprises’ survey data;
5. C^{-1} is the transposed centered matrix;
6. Cov_Min_{ag} represents the average of the aggregated matrix maximum score of the enterprises’ survey data.

So, the CODAS–Hamming–Mahalanobis methodology follows 8 steps, presented by [12], where Step 5 includes Hamming distance as the primary distance, as shown in Equation (3), and Mahalanobis distance as the secondary distance, as shown in Equation (2).

$$D_4(H_T^1(x_i), H_T^2(x_i)) = \frac{1}{2} \left(\frac{1}{T} \sum_{l=1}^T \frac{|\delta_l^1 - \delta_l^2|}{2\tau + 1} + \max_{l=1,2,\dots,T} \left(\frac{|\delta_l^1 - \delta_l^2|}{2\tau + 1} \right) \right) \tag{3}$$

From the designed instrument to the data collection presented in Tables 4 and 5, respectively, the decision makers from the participating companies responded using a Likert scale. To analyze the resulting scales from the 31 enterprises, an aggregated and normalized matrix was obtained from the enterprises’ evaluation, setting the maximum value criteria as shown in Table 7.

Table 7. Aggregated decision matrix.

Code	EC	SO	EN	Code	EC	SO	EN	Code	EC	SO	EN
TI01	4	0	0	GA07	0	0	2	E03	0	0	3
TI02	4	0	0	I01	0	3	0	SP01	0	0	3
TI03	0	4	0	I02	0	4	0	SP02	4	0	0
TI04	0	3	0	I03	4	0	0	SP03	0	0	4
TI05	4	0	0	I04	3	0	0	SP04	4	0	0
TI06	3	0	0	I05	3	0	0	SP05	0	0	3
TI07	0	3	0	I06	3	0	0	SP06	0	0	4
TI08	4	0	0	GR01	0	0	3	SP07	0	0	3
TI09	0	0	3	GR02	0	3	0	SP08	0	0	4
TI10	0	0	3	GR03	0	4	0	SP09	0	0	4
TI11	3	0	0	GR04	0	0	3	SP10	0	0	3
GA01	0	0	4	GR05	0	0	3	SP11	0	0	4
GA02	0	0	3	GR06	0	3	0	SP12	3	0	0
GA03	0	0	4	GR07	0	4	0	SP13	3	0	0
GA04	0	0	4	GR08	3	0	0	SP14	0	0	3
GA05	0	0	4	E01	0	0	3	SP15	0	0	4
GA06	0	0	4	E02	0	0	3				

EC = economic, SO = social, EN = environmental.

For the criterion’s weight calculation, the ambiguity reduction method was utilized to reduce the ambiguity of the values obtained from the enterprises through the AHP methodology, as presented by [12]. Using the formula $w_j^{AR} = \gamma w_j^{LGE} + \gamma w_j^{ES} + (1 - \gamma)w_j^{AHP}$, where γ represents the impact of the dimensional criterion weighting with respect to the decision makers; w_j^{LGE} is the obtained weighting from the literature review for the critical dimension j ; w_j^{ES} is the obtained weighting from the enterprises’ survey for the critical dimension j ; w_j^{AHP} is the AHP weighting for the critical criterion j ; and w_j^{AR} is the ambiguity reduction weighting for the critical criterion j . Table 8 displays the weighting values obtained.

Table 8. Ambiguity reduction criteria assessment.

Criteria	Economic	Social	Environmental
w_j^{LHC}	0.2734	0.1453	0.5813
w_j^{ES}	0.3045	0.1864	0.5092
w_j^{AHP}	0.6340	0.2600	0.1060
w_j^{AR}	0.4039	0.1972	0.3988

To obtain the w_j^{AHP} , a pair-wise comparison matrix was developed, and then a standardized autovector was generated to obtain w_j , the normalized average value. The consistency index was calculated as $CI = (\lambda_{max} - n_c) / (n_c - 1)$, and the consistency ratio was given by $CR = CI/RI$, which is accepted when it is not greater than 10% of the random index (RI). The resulting CI and RI were obtained as 0.0166 and 5.8%, respectively, demonstrating the consistency index.

The process following each criterion evaluation value was taken from the CODAS–Hamming–Mahalanobis methodology proposed by [12], calculating the normalized decision matrix (Step 2), the weighted normalized decision matrix (Step 3), the negative ideal solution (Step 4), and the main (Hamming distance) and secondary (Mahalanobis distance ($Cov(\bar{r}_{ij}), ns$)) measure (Step 5). This set of distance combinations was used to construct the relative evaluation matrix (Step 6), from which the preference score was obtained for each evaluated alternative (Step 7), and finally, the hierarchy of green energy indicators and their measurements was found (Step 8).

4. Results

Micro-, small, and mid-size enterprises (MSMEs) are well recognized worldwide as each country's economic contribution [51–53], leading local innovation, skill, and distribution of goods and services [54]. Employing around 60% of the labor force and with a 50% average GDP share [55,56], MSMEs tend to become a workforce that complement large enterprises' supply chains.

Overall, MSMEs represent a social and local synergy that drives job creation [57]. Nevertheless, we are in the middle of a globalized, rapidly changing supply chain in which there is an increase in the number of new enterprises, and customers have high expectations of quality, delivery, service, and the uniqueness of each product [58,59]. In addition, there are exceptions where major disruptions can affect and reduce demand from the market, such as the coronavirus pandemic [60] or climate change and environmental emissions caused by fossil-fuel usage [61]; MSMEs face challenging disruptions in enhancing sustainability in manufacturing processes within a global value chain.

By 2019, MSMEs from the OECD countries accounted for one in three people in a micro firm, and two out of three in an SME enterprise [57]. Figure 3 displays the OECD average of persons employed in MSMEs and large enterprises.

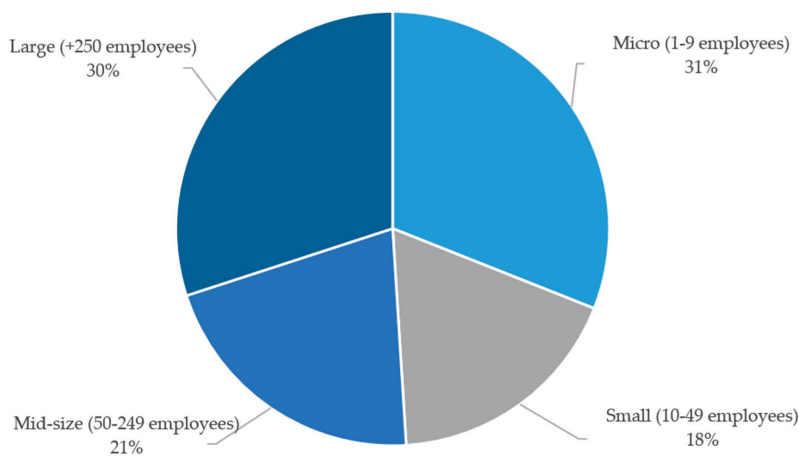


Figure 3. Average of people employed in a MSMEs and large firms [21].

Through the 2008 economic crisis, MSMEs opened up through innovation and the era of digitalization, transforming their business models and supply chain production, facing challenges with technology adoption but gaining market share by adapting digital

platforms as strategic resources and networks [57]. In 2020, facing the coronavirus outbreak, the whole business platform experienced great difficulties. Even though the digital era has aided certain business activities, the majority of economic activities have come to a halt. MSMEs, as the OECD describes, have experienced an impact that has been demonstrated in a few countries, as shown in Table 9, where some of these countries' actual situations are displayed [60].

Table 9. MSMEs situation through COVID-19 outbreak, across OECD countries [60].

Country	SMEs Situation
China	Approximately 60% of its MSMEs have returned to activities, a demand reduction has become the number one challenge.
Germany	Initially, MSMEs did not feel an economic and production dropdown because of their operation in regional markets; by March 9, one-third of firms are expecting a 10% decline in turnover in 2020.
Italy	Approximately 72% of small firms have been affected by the health epidemic; specifically, transport (98.9%), tourism (89.9%), fashion (79.9%), and agro-food (77.7%) due to the demand downfall.
South Korea	SMEs have shown a time reduction in product delivery due to China's factory closure, where 71.8% of SMEs will be affected by the outbreak.
USA	Between the SMEs presenting some kind of damage, 42% are seeing lower sales, and 39% reported supply chain disruption.

Now, in 2024, MSMEs have been subject to more disrupting processes along the value chain, driven partly by climate change. As enterprises represent approximately 90% of each country's economic units, they tend to have a significant footprint.

Overall, in regular times, passing through a pandemic outbreak, or dealing with threats of climate change such as natural disasters and extreme weather, MSMEs are still recognized as a labor force weight for all countries' economies. As described above, MSMEs face everyday challenges from supply and demand sides to financial markets. The literature review displays strategies for enterprises to follow, such as effective supply chain management, technology information for timely handling of data, logistics systems to achieve product delivery, distribution and warehouse storage, and sustainable manufacturing [62–64]. Further strategies include analyzing the working environment, working on alliances, and cooperation among enterprises [10,51,65,66]; additionally, as the Organization of Economic and Cooperation Development describes it, MSMEs should develop and create joint strategies to achieve goals that individually are difficult to reach [67].

Of the 31 enterprises that responded to the survey, 70.96% represent MSMEs and 29.03% large enterprises. Seeing the importance for them to deploy sustainable manufacturing practices, to gain market shares and complement large enterprises' supply chains, the survey analytic instrument shows the preference scores. From these preference scores, the authors obtained a hierarchy list of green energy indicators and their measurements and items, as shown in Table 10.

Table 10. Top 10 green energy indicators and their measurements.

Rank	Indicator	Measurement	Code
1	Technology and Innovation	Collaboration Capacity	TI05
2	Investment	Supply Chain	I05
3	Investment	Energy efficiency	I03
4	Technology and Innovation	Technology Capacity	TI01
5	Sustainable Processes	Residual	SP02
6	Technology and Innovation	Collaboration Capacity	TI08
7	Sustainable Processes	Reduce	SP04
8	Technology and Innovation	Collaboration Capacity	TI11
9	Technology and Innovation	Technology Capacity	TI02
10	Geographical Aspects	Electric energy	GA05

The resulting hierarchy was obtained from an MCDM methodology that took into account 31 decision-making entrepreneurs, within different goods and service sectors, located in the cities Chihuahua and Juarez. It also gives top 10 indicators for enterprises to focus on and start developing ideas and strategies for implementing changes towards sustainable manufacturing. New companies can delve into green energy indicators and their measurements and not know if the hierarchy applies to them; for these new cases, this paper proposes a linearity factor model for predictive analysis. Calculating the sum of the 31 companies from the green energy survey gives a total for each of the 50 items. From this point forward, an index score is obtained, as shown in Equation (4).

$$I_s = \frac{P_s}{s} \quad (4)$$

where P_s is equal to the preference score obtained from the CODAS–Hamming–Mahalanobis method, and s is equal to the sum of the Likert scale evaluation for each of the 50 items made by the panel. With this linearity factor model, each new company that desires to fill in the green energy survey can obtain their green energy indicators, which are hierarchized by this predictive model. This proposed linearity factor model is restricted by the following aspect: if the new responses obtained from several companies are higher than the 10% of the 31 businesses' panel used for the primary evaluation, then the CODAS–Hamming–Mahalanobis method will have to be calculated again to obtain the green energy hierarchy. This restriction exists because if it is higher than 90%, the level of confidence will be affected. Table 11 shows the 32nd business that responded to the green energy survey on 24 April 2024, where the linearity factor model was applied, predicting the new green energy indicator hierarchy from the new enterprise's perspective.

Table 11. Predicting new indicators hierarchy through linearity factor model.

Ps	s	Is	EM-32	New Ps	Code	Items	New Ranking
2.8509	125	0.0228	5	2.9650	I03	Energy Efficiency	1
2.8566	131	0.0218	5	2.9657	SP02	Residual	2
2.8566	130	0.0220	5	2.9665	TI01	Technology Capacity	3
2.8488	118	0.0241	5	2.9695	I05	Supply Chain	4
2.8612	129	0.0222	5	2.9721	SP04	Reduce	5
2.8646	129	0.0222	5	2.9757	TI11	Collaboration Capacity	6
2.8612	122	0.0235	5	2.9785	TI08	Collaboration Capacity	7
2.8453	103	0.0276	5	2.9835	TI05	Collaboration Capacity	8
2.8707	124	0.0232	5	2.9865	TI02	Technology Capacity	9
2.8828	130	0.0222	5	2.9937	GR01	Policy	10

Comparing both top 10 tables, the sustainable manufacturing pillars remained unchanged, i.e., the economic pillar had nine indicators and the environmental pillar had one indicator. On the other hand, the green energy indicators changed in ranking, as the proposed predictive linearity factor model takes into account the enterprise's evaluation within the preference score obtained from the CODAS–Hamming–Mahalanobis method.

5. Discussion

In Mexico, green energy resources represent just 9.74% of enterprises' energy consumption. Policy makers are more focused on raising economic standards and maintaining their global position, so the implications of green energy are given little attention. From a managerial perspective, manufacturing businesses require a lot of change, but it is difficult for decision makers to focus on one or two strategies towards a sustainable manufacturing approach.

The literature presents research from a variety of perspectives regarding green energy definition, usage, implementation, and infrastructure. Riosvelasco et al. [21] propose a literature review using a PRISMA 2020 methodology, focused on identifying green energy

indicators and their measurements. From the obtained final list, previously shown in Table 1, the following are identified as the more frequent indicators and their measurements: six green energy indicators were established, with a total of 18 measurements and 50 items.

From a previous research proposal using an MCDM method, the CODA–Hamming–Mahalanobis methodology was applied to hierarchize the 50-item list from an enterprise’s perspective according to the Likert Scale importance they gave to each green energy indicator and their measurements. From the 50-item list, the top 10 green energy indicators are focused on two of the sustainable manufacturing pillars: nine indicators from the economic pillar and one indicator from the environmental pillar, as shown in Table 12.

Table 12. Green energy hierarchy using the CODAS–Hamming–Mahalanobis method.

Code	Green Energy Indicators		Sustainable Manufacturing Pillars	
	Indicator	Measurement	Economic	Environmental
TI05	Technology and Innovation	Collaboration Capacity	Have an R&D department Integrate to SC suppliers that use green energy in their production process Identify opportunity areas in production processes’ machines, equipment, and transport for efficient use of energy Build improvements in productive processes for the efficient use of energy Production processes design to avoid waste Collaborate with research centers or universities to create and develop innovative green energy projects Develop products with better cycle time, design, durability and raw material quality Protect innovation through IP Develop new product processes for the efficient use of energy	Use of photovoltaic solar cells to generate and use energy
I05	Investment	Supply Chain		
I03	Investment	Energy Efficiency		
TI01	Technology and Innovation	Technology Capacity		
SP02	Sustainable Practices	Residual		
TI08	Technology and Innovation	Collaboration Capacity		
SP04	Sustainable Practices	Reduce		
TI11	Technology and Innovation	Collaboration Capacity		
TI02	Technology and Innovation	Technology Capacity		
GA05	Geographical Aspects	Electric Energy		

It can be observed that 90% of the top 10 items are categorized under the economic pillar for sustainable manufacturing. As described by [68], sustainability takes into account “...the transformation of resources into economically valuable goods...”; that is to say, businesses can gain by rethinking specific actions that will boost economic behavior, such as optimizing material and energy usage and creating production processes that can be maintained by themselves, among other characteristics [69]. The green energy indicators identified are focused on initiatives such as protecting innovation through IP, dynamic collaboration between supply chain members to implement technology, engineering relying

on information, and encouraging human and natural resources towards a sustainable manufacturing environment, among others. Nevertheless, it gives a comprehensive approach to decision makers to organize and manage small projects within actual procedures to create a sustainable culture in the supply chain and in their business, allowing decision makers to take a specific course of action.

Furthermore, the authors propose a predictive analysis model whereby the preference score is taken into account to calculate the index score for new enterprises to evaluate important green energy indicators, and to obtain their ranking of items to deploy specific green energy practices toward sustainable manufacturing. The EM-32's prediction on the green energy hierarchy presented a new ranking in items, such as identifying areas of opportunity to upgrade equipment for efficient use of energy and designing production processes to avoid waste, among other practices. Both hierarchized lists are represented in Figure 4. From the sustainable manufacturing pillars, it can be observed that enterprises are more focused on the economic pillar, which takes into account actions such as product cost, business model, and consumption of resources, among others.

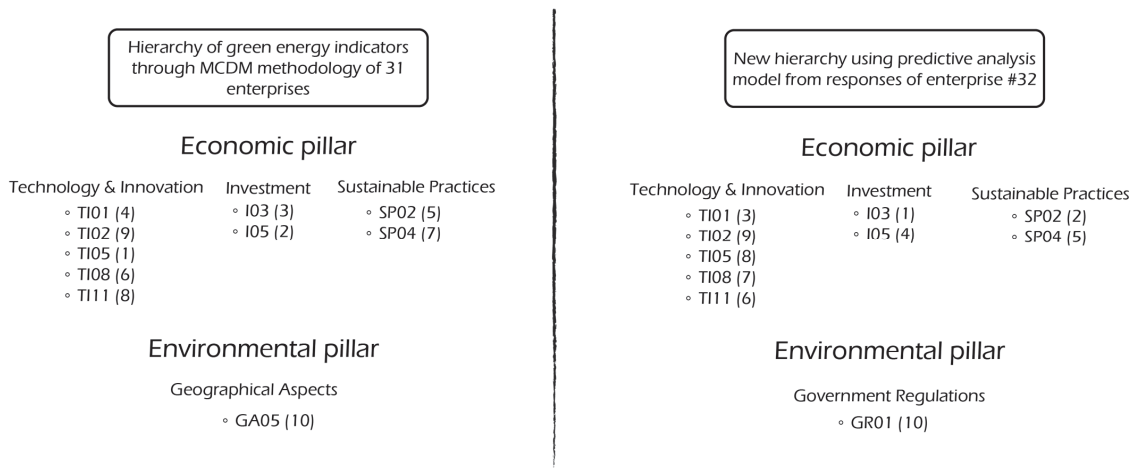


Figure 4. Resulting top 10 hierarchy list from an MCDM methodology and a predictive analysis model.

6. Conclusions

Sustainable manufacturing is happening globally, and nowadays, it affects enterprises in developing or emerging countries due to the lack of systemic procedures and finite strategies for green energy implementation. As mentioned by [68], enterprises know about sustainable manufacturing but do not necessarily know how to deploy specific practices. Our research has contributed a literature review on green energy, in which a hierarchized top 10 list was not found in research, especially from enterprises' perspectives by grading green energy indicators. The following research proposes two models to hierarchize and predict a top 10 list of green energy indicators, measurements, and their items to incentivize decision makers to deploy specific actions and initiatives in their supply chain towards a sustainable manufacturing culture. From the results obtained from the green energy indicators, measurements, and items survey, it is concluded that enterprises search for initiatives from the economic pillar of sustainable manufacturing, giving priority to those focused on deploying actions towards equipment, machinery, and production processes for the efficient use of energy. In addition to this, they seek to collaborate with research centers or universities for the development of practices and designs of new production processes to avoid waste. Not least, they choose to use photovoltaic solar energy as a source of their everyday processes. Besides these results, this article proposes a linearity factor model as

a predictive analysis to allow new enterprises to respond to the survey and obtain their hierarchy in green energy indicators, measurements, and items.

7. Implications and Future Research

Future research on the effects of green energy indicators, measurements, and items is required. Research could extend to empirical studies or case studies, in addition to introducing the development of standards or metrics of performance for each of the top 10 items. This field presents opportunities for future understanding and deployment of specific initiatives for the industrial field. The survey presented in this paper could be broadened to include Mexico's north zone, including the United States, by translating the questionnaire. Moreover, it could include a Gaussian process regression model as the predictive analysis model.

Author Contributions: Conceptualization, G.E.R.-M., I.J.C.P.-O. and S.N.-M.; methodology, G.E.R.-M., I.J.C.P.-O. and L.A.P.-D.; software, G.E.R.-M.; validation, I.J.C.P.-O., L.A.P.-D., L.C.M.-G. and L.A.R.-P.; formal analysis, S.N.-M.; investigation, G.E.R.-M.; resources, G.E.R.-M., I.J.C.P.-O., L.A.P.-D., L.C.M.-G. and L.A.R.-P.; data curation, G.E.R.-M.; writing—original draft preparation, G.E.R.-M.; writing—review and editing, G.E.R.-M., I.J.C.P.-O. and S.N.-M.; visualization, S.N.-M., L.A.P.-D., L.C.M.-G. and L.A.R.-P.; supervision, I.J.C.P.-O. and S.N.-M.; project administration, G.E.R.-M. and I.J.C.P.-O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Survey URL <https://energiasverdes.limesurvey.net/686967?lang=es> (accessed on 8 May 2024).

Acknowledgments: Special thanks for the support from Cluster de Energía Chihuahua (CECH) and Cámara Nacional de la Industria y la Transformación, North Zone (CANACINTRA) for their initiative to share the green energy survey with their members.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sotelo, C.K.C.; Cotón, S.G.H.; Salamanca, P.I.M. La innovación y su interrelación con la competitividad. Sector manufacturero de la zona metropolitana de Guadalajara. *Red Int. Investig. Compet.* **2018**, *8*, 1–21. Available online: <https://riico.net/index.php/riico/article/view/1164/832> (accessed on 8 May 2024).
2. García, M.A.B.; Romero, E.D.; Mena, R.S.A.; López, N.A.S. La innovación y su influencia en el servicio al cliente en la industria del vestido de Puebla, México. *Rev. Int. Adm. Finanz.* **2018**, *11*, 21–37. Available online: <http://www.theibr2.com/RePEc/ibf/riafin/riaf-v11n4-2018/RIAF-V11N4-2018-3.pdf> (accessed on 8 May 2024).
3. Emenike, S.N.; Falcone, G. A review on energy supply chain resilience through optimization. In *Renewable and Sustainable Energy Reviews*; Elsevier Ltd.: Amsterdam, The Netherlands, 2020; Volume 134. [CrossRef]
4. Riosvelasco-Monroy, G.E.; Flores-Amador, J.; Pérez-Olguín, I.J.C. Gestión del Conocimiento a través de la Colaboración Horizontal en el Clúster MACH Knowledge transfer through a Horizontal Collaboration in Cluster MACH. *Rev. Int. Investig. E Innovación Tecnológica* **2019**, *7*, 1–22. Available online: <https://riiit.com.mx/ojs/index.php/riiit/article/view/93> (accessed on 8 May 2024).
5. Yu, S.; Liu, J.; Hu, X.; Tian, P. Does development of renewable energy reduce energy intensity? Evidence from 82 countries. *Technol. Forecast. Soc. Change* **2022**, *174*, 121254. [CrossRef]
6. Choy, C.S.; Yew, W.K.; Lin, B. Criteria for measuring KM performance outcomes in organisations. *Ind. Manag. Data Syst.* **2006**, *106*, 917–936. [CrossRef]
7. Maldonado-Guzmán, G.; Garza-Reyes, J.A.; Pinzón-Castro, S.Y.; Kumar, V. Barriers to innovation in service SMEs: Evidence from Mexico. *Ind. Manag. Data Syst.* **2017**, *117*, 1669–1686. [CrossRef]
8. Hernández, A.S. Factores Relevantes De Competitividad En Pymes. *Eur. Sci. J. ESJ* **2018**, *14*, 193. [CrossRef]
9. Merritt, H. Mexican Technology Based Firms and their Capacity for Innovation: A methodological proposal. *Trayectorias* **2012**, *14*, 33–34.
10. Tobiasen, A.E.; Pettersen, I.B. Exploring open innovation collaboration between SMEs and larger customers. *Balt. J. Manag.* **2018**, *13*, 65–83. [CrossRef]
11. Rashid, S.H.A.; Evans, S.; Longhurst, P. A comparison of four sustainable manufacturing strategies. *Int. J. Sustain. Eng.* **2008**, *1*, 214–229. [CrossRef]

12. Riosvelasco-Monroy, G.E.; Perez-Olguín, I.J.C.; Flores-Amador, J.; Pérez-Domínguez, L.A.; Hernández-Gómez, J.A. Horizontal Collaboration Business Model towards a Sustainable I4.0 Value Creation. In *Sustainable Manufacturing in Industry 4.0: Pathways and Practices*; Gholami, H., Abdul-Nour, G., Sharif, S., Streimikiene, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2023. [CrossRef]
13. Melo, D.I.C.; Queiroz, G.A.; Junior, P.N.A.; de Sousa, T.B.; Yushimito, W.F.; Pereira, J. Sustainable digital transformation in small and medium enterprises (SMEs): A review on performance. In *Heliyon*; Elsevier Ltd.: Amsterdam, The Netherlands, 2023; Volume 9. [CrossRef]
14. Patalas-Maliszewska, J.; Łosyk, H. An approach to assessing sustainability in the development of a manufacturing company. *Sustainability* **2020**, *12*, 8787. [CrossRef]
15. Niewiadomski, P.; Stachowiak, A. Identification of Barriers to Sustainable Manufacturing Implementation—The Perspective of Manufacturers of Parts and Components for Agricultural Transport. *Sustainability* **2024**, *16*, 2244. [CrossRef]
16. Kuś, A.; Grego-Planer, D. A model of innovation activity in small enterprises in the context of selected financial factors: The example of the renewable energy sector. *Energies* **2021**, *14*, 2926. [CrossRef]
17. Shankar, K.M.; Kannan, D.; Kumar, P.U. Analyzing sustainable manufacturing practices—A case study in Indian context. *J. Clean Prod.* **2017**, *164*, 1332–1343. [CrossRef]
18. Polcyn, J.; Us, Y.; Lyulyov, O.; Pimonenko, T.; Kwilinski, A. Factors influencing the renewable energy consumption in selected european countries. *Energies* **2022**, *15*, 108. [CrossRef]
19. Zastempowski, M. Analysis and modeling of innovation factors to replace fossil fuels with renewable energy sources—Evidence from European Union enterprises. *Renew. Sustain. Energy Rev.* **2023**, *178*, 113262. [CrossRef]
20. Neri, A.; Cagno, E.; Lepri, M.; Trianni, A. A triple bottom line balanced set of key performance indicators to measure the sustainability performance of industrial supply chains. *Sustain. Prod. Consum.* **2021**, *26*, 648–691. [CrossRef]
21. Monroy, G.E.R.; Rojas-Ramírez, A.; Noriega-Morales, S. Sustainable practices for the efficient use of green energy in Ciudad Juárez SMEs. *Rev. Cienc. Tecnológicas* **2023**, *6*, e265. [CrossRef]
22. Fatima, Z.; Oksman, V.; Lahdelma, R. Enabling small medium enterprises (Smes) to become leaders in energy efficiency using a continuous maturity matrix. *Sustainability* **2021**, *13*, 10108. [CrossRef]
23. Riosvelasco-Monroy, G.E.; Pérez-Olguín, I.J.C.; Pérez-Domínguez, L.A.; Méndez-González, L.C.; Noriega-Morales, S. Application of the COHRV Conceptual Framework to Enhance Sustainable Manufacturing. *Sustainability* **2022**, *14*, 16804. [CrossRef]
24. Rosen, M.A.; Kishawy, H.A. Sustainable manufacturing and design: Concepts, practices and needs. *Sustainability* **2012**, *4*, 154–174. [CrossRef]
25. Alegre-Vidal, J.; Lapidra-Alcami, R.; Chiva-Gómez, R. Linking operations strategy and product innovation: An empirical study of Spanish ceramic tile producers. *Res. Policy* **2004**, *33*, 829–839. [CrossRef]
26. Hariyani, D.; Mishra, S.; Hariyani, P.; Sharma, M.K. Drivers and motives for sustainable manufacturing system. *Innov. Green Dev.* **2023**, *2*, 100031. [CrossRef]
27. Saxena, P.; Stavropoulos, P.; Kechagias, J.; Salonitis, K. Sustainability assessment for manufacturing operations. *Energies* **2020**, *13*, 2730. [CrossRef]
28. Hariyani, D.; Mishra, S. Organizational enablers for sustainable manufacturing and industrial ecology. *Clean. Eng. Technol.* **2022**, *6*, 100375. [CrossRef]
29. Khan, I.S.; Ahmad, M.O.; Majava, J. Industry 4.0 and sustainable development: A systematic mapping of triple bottom line, Circular Economy and Sustainable Business Models perspectives. *J. Clean. Prod.* **2021**, *297*, 126655. [CrossRef]
30. Nezami, F.G.; Yildirim, M.B. A sustainability approach for selecting maintenance strategy. *Int. J. Sustain. Eng.* **2013**, *6*, 332–343. [CrossRef]
31. Mengistu, A.T.; Dieste, M.; Panizzolo, R.; Biazzo, S. Sustainable product design factors: A comprehensive analysis. *J. Clean. Prod.* **2024**, 142260. [CrossRef]
32. Mahalanobis, P.C. On the Generalized Distance in Statistics. *J. Asiat. Soc. Bengal.* **1930**, *XXVI*, 49–55.
33. Cudney, E.A.; Paryani, K.; Ragsdell, M.K. Applying the Mahalanobis-Taguchi system to vehicle handling. *Concurr. Eng. Res. Appl.* **2006**, *14*, 343–354. [CrossRef]
34. Detwal, P.K.; Soni, G.; Jakhar, S.K.; Shrivastava, D.K.; Madaan, J.; Kayikci, Y. Machine learning-based technique for predicting vendor incoterm (contract) in global omnichannel pharmaceutical supply chain. *J. Bus. Res.* **2023**, *158*, 113688. [CrossRef]
35. Etherington, T.R. Mahalanobis distances for ecological niche modelling and outlier detection: Implications of sample size, error, and bias for selecting and parameterising a multivariate location and scatter method. *PeerJ* **2021**, *9*, e11436. [CrossRef] [PubMed]
36. Pronzato, L.; Wynn, H.P.; Zhigljavsky, A.A. Simplicial variances, potentials and Mahalanobis distances. *J. Multivar. Anal.* **2018**, *168*, 276–289. [CrossRef]
37. Ghorbani, H. Mahalanobis distance and its application for detecting multivariate outliers. *Facta Univ. Ser. Math. Inform.* **2019**, *34*, 583. [CrossRef]
38. Sarmadi, H.; Entezami, A.; Razavi, B.S.; Yuen, K.V. Ensemble learning-based structural health monitoring by Mahalanobis distance metrics. *Struct. Control. Health Monit.* **2021**, *28*, e2663. [CrossRef]
39. Yao, L.; Lin, T.B. Evolutionary mahalanobis distance-based oversampling for multi-class imbalanced data classification. *Sensors* **2021**, *21*, 6616. [CrossRef] [PubMed]

40. Panda, A.; Pachori, R.B.; Sinnappah-Kang, N.D. Classification of chronic myeloid leukemia neutrophils by hyperspectral imaging using Euclidean and Mahalanobis distances. *Biomed. Signal Process. Control.* **2021**, *70*, 103025. [CrossRef]
41. Flores-Guerrero, J.L.; Grzegorzczak, M.A.; Connelly, M.A.; Garcia, E.; Navis, G.; Dullaart, R.P.; Bakker, S.J. Mahalanobis distance, a novel statistical proxy of homeostasis loss is longitudinally associated with risk of type 2 diabetes. *EBioMedicine* **2021**, *71*, 103550. [CrossRef]
42. Huo, J.; Ma, Y.; Lu, C.; Chenggang, L.; Kun, D.; Huaiqi, L. Mahalanobis distance based similarity regression learning of NIRS for quality assurance of tobacco product with different variable selection methods. *Spectrochim. Acta. A Mol. Biomol. Spectrosc. vol.* **2021**, *251*, 119364. [CrossRef] [PubMed]
43. Mousavi, M.; Gandomi, A.H. Structural health monitoring under environmental and operational variations using MCD prediction error. *J. Sound Vib.* **2021**, *512*, 116370. [CrossRef]
44. Xiong, S.; Chen, W. A robust hybrid method using dynamic network analysis and Weighted Mahalanobis distance for modeling systemic risk in the international energy market. *Energy Econ.* **2022**, *109*, 105954. [CrossRef]
45. Asakura, T.; Yashima, W.; Suzuki, K.; Shimotou, M. Anomaly detection in a logistic operating system using the mahalanobis-taguchi method. *Appl. Sci.* **2020**, *10*, 4376. [CrossRef]
46. Chang, Z.P.; Li, Y.W.; Fatima, N. A theoretical survey on Mahalanobis-Taguchi system. *Meas. J. Int. Meas. Confed.* **2019**, *136*, 501–510. [CrossRef]
47. Hadighi, S.A.; Sahebjamnia, N.; Mahdavi, I.; Asadollahpour, H.; Shafieian, H. Mahalanobis-Taguchi System-based criteria selection for strategy formulation: A case in a training institution. *J. Ind. Eng. Int.* **2013**, *9*, 26. [CrossRef]
48. Qiu, Z.; Zhou, B.; Yuan, J. Protein–protein interaction site predictions with minimum covariance determinant and Mahalanobis distance. *J. Theor. Biol.* **2017**, *433*, 57–63. [CrossRef] [PubMed]
49. Leys, C.; Klein, O.; Dominicy, Y.; Ley, C. Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. *J. Exp. Soc. Psychol.* **2018**, *74*, 150–156. [CrossRef]
50. Liu, D.; Qi, X.; Fu, Q.; Li, M.; Zhu, W.; Zhang, L.; Faiz, M.A.; Khan, M.I.; Li, T.; Cui, S. A resilience evaluation method for a combined regional agricultural water and soil resource system based on Weighted Mahalanobis distance and a Gray-TOPSIS model. *J. Clean Prod.* **2019**, *229*, 667–679. [CrossRef]
51. Agostini, L.; Nosella, A. Inter-organizational relationships involving SMEs: A bibliographic investigation into the state of the art. *Long Range Plan.* **2019**, *52*, 1–31. [CrossRef]
52. Farzin, F. Localising the impact of techno-entrepreneurship in Eastern Iran: Birjand’s Science and Technology Park as a local innovation community. *Local Econ. J. Local Econ. Policy Unit* **2017**, *32*, 692–710. [CrossRef]
53. Méndez, L.H.; Villalobos, K.J.O.; Méndez, O.J.M. Perfil operativo de las PYME juarenses. *novaRua* **2014**, *4*, 14–25. Available online: <http://148.210.132.19/ojs/index.php/NovaRua/article/view/304> (accessed on 8 May 2024). [CrossRef]
54. María, L.; Torrejón, P.; Sánchez, L.C. Factores que impactan en la competitividad de la micro y pequeña empresa: Municipio de Huejotzingo Puebla. *Red Int. De Investig. En Compet.* **2016**, 1035–1049. Available online: <https://riico.net/index.php/riico/article/view/1361/1031> (accessed on 8 May 2024).
55. OCDE. *Financiamiento de PyMes y Emprendedores 2018: Un Marcador de la OCDE*; OCDE: Paris, France, 2018.
56. Valdés, J.; Sánchez, G. Las Pymes en el contexto mundial: Sus particularidades en México. *Iberóforum. Revista de Ciencias Sociales de la Universidad Iberoamericana* **2012**, *VII*, 126–156.
57. OECD. *OECD SME and Entrepreneurship Outlook 2019*; OECD: Paris, France, 2019. [CrossRef]
58. Jaharuddin, N.S.; Mansor, Z.D.; Yaakob, S. Assessing the Supply Chain Intelligence Practices of Small Medium Enterprises in Malaysia. *Procedia Econ. Financ.* **2016**, *35*, 515–521. [CrossRef]
59. Sablón-Cossio, N.; Medina-León, A.; Acevedo-Suárez, J.A.; Acevedo-Urquiaga, A.J.; López-Joy, T. Consideraciones sobre la planificación de productos alimenticios en una cadena de suministro comercial. *Ing. Ind.* **2013**, *34*, 353–362.
60. OECD. *Financing SMEs and Entrepreneurs 2020: An OECD Scoreboard*; OCDE: Paris, France, 2020.
61. Ali, W.; Gohar, R.; Chang, B.H.; Wong, W.-K. Revisiting the impacts of globalization, renewable energy consumption, and economic growth on environmental quality in South Asia. *Adv. Decis. Sci.* **2022**, *26*, 1–23.
62. Colin, M.; Galindo, R.; Hernández, O. Information and Communication Technology as a Key Strategy for Efficient Supply Chain Management in Manufacturing SMEs. *Procedia Comput. Sci.* **2015**, *55*, 833–842. [CrossRef]
63. Kumar, M.; Raman, J.; Priya. A Supply Chain Collaboration Model for Product Development with R&D Subsidies. *J. Supply Chain Manag. Syst.* **2015**, *4*. [CrossRef]
64. Sarkis, J.; Gonzalez, E.D.S.; Koh, S.C.L. Effective multi-tier supply chain management for sustainability. *Int. J. Prod. Econ.* **2019**, *217*, 1–10. [CrossRef]
65. Pfeifer, S.; Peterka, S.O.; Stanić, M. Business models of micro businesses: Empirical evidence from creative industries. *J. Contemp. Manag.* **2017**, *22*, 1–19.
66. Xue, X.; Wang, S.; Lu, B. Manufacturing service composition method based on networked collaboration mode. *J. Netw. Comput. Appl.* **2016**, *59*, 28–38. [CrossRef]
67. OCDE. *Economy Raising Productivity in Small Traditional Enterprises*. 2017. Available online: <https://www.oecd.org/mexico/mexico-raising-productivity-in-small-traditional-enterprises.pdf> (accessed on 8 May 2024).

68. Sartal, A.; Bellas, R.; Mejias, A.M.; García-Collado, A. The sustainable manufacturing concept, evolution and opportunities within Industry 4.0: A literature review. *Adv. Mech. Eng.* **2020**, *12*, 1687814020925232. [CrossRef]
69. Athamneh, M.H.A.; Al-Ajlouni, M.M. Unleashing the Power of Green: Sustainable Success through HRM, Supply Chains, and Innovation—A Comprehensive Review of Literature. *Management* **2024**, *1*, 73–105. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Controlling the Friction Coefficient and Adhesive Properties of a Contact by Varying the Indenter Geometry

Iakov A. Lyashenko ^{1,*}, Thao H. Pham ¹ and Valentin L. Popov ^{1,2,*}

¹ Department of System Dynamics and Friction Physics, Institute of Mechanics, Technische Universität Berlin, 10623 Berlin, Germany; pham.19@campus.tu-berlin.de

² Center of Advanced Studies in Mechanics, Tribology, Bio- and Nanotechnologies, Samarkand State University, Samarkand 140104, Uzbekistan

* Correspondence: i.lyashenko@tu-berlin.de (I.A.L.); v.popov@tu-berlin.de (V.L.P.); Tel.: +49-(0)30-31475917 (I.A.L.)

Abstract: In the present paper, we describe a series of laboratory experiments on the friction between rigid indenters with different geometrical forms and an elastic sheet of elastomer as a function of the normal load. We show that the law of friction can be controlled by the shape of the surface profile. Since the formulation of the adhesive theory of friction by Bowden and Tabor, it is widely accepted and confirmed by experimental evidence that the friction force is roughly proportional to the real contact area. This means that producing surfaces with a desired dependence of the real contact area on the normal force will allow to “design the law of friction”. However, the real contact area in question is that during *sliding* and differs from that at the pure normal contact. Our experimental studies show that for indenters having a power law profile $f(r) = c_n r^n$ with an index $n < 1$, the system exhibits a constant friction coefficient, which, however, is different for different values of n . This opens possibilities for creating surfaces with a predefined coefficient of friction.

Keywords: adhesion; controlling friction; surface profile; real contact area

Citation: Lyashenko, I.A.; Pham, T.H.; Popov, V.L. Controlling the Friction Coefficient and Adhesive Properties of a Contact by Varying the Indenter Geometry. *Processes* **2024**, *12*, 1209. <https://doi.org/10.3390/pr12061209>

Academic Editors: Sheng Du, Li Jin, Xiongbo Wan and Zixin Huang

Received: 22 May 2024

Revised: 10 June 2024

Accepted: 11 June 2024

Published: 13 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Friction plays an important role in many industrial processes. Often one seeks to reduce friction in order to lower the energy consumption [1,2]. In other technological applications, the friction coefficient must be high to provide optimal performance, such as in nanostructuring burnishing [3], car braking systems [4], contact of wheels with road surfaces [5], movement transmission [6], etc. Knowledge of the friction behavior in many engineering fields, such as the automotive industry [7], wind turbines [8] or medical devices [9], is of crucial interest.

Adhesion is another important phenomenon that is interesting both as one of the contributing factors to friction and in itself. Adhesion plays an important role in nature and engineering both at the micro- [10,11] and macroscopic scale [12,13]. The pollination process of plants is a biological example [14]. Certain amphibians (tree frogs) and reptiles (geckos) use adhesive forces to climb up or even hang upside down on vertical surfaces. Understanding the underlining mechanisms lead to mimicking their function for engineering purposes [15,16]. An artificial surface with gecko-like adhesive properties was developed by Gorb et al. [17]. In engineering applications, adhesion plays an important role in paints and coatings [18], orthodontics brackets [19] and aeronautical applications [20]. Adhesive contacts have therefore been the subject of intensive research [21–23].

Despite decades of research, it is still impossible to deliberately design surfaces with desired frictional or adhesive properties [24]. The most common way to influence the frictional behavior in dry friction contacts is modifying the surface of the contacting bodies [25,26] either by modifying the surface topography or by applying homogeneous or

heterogeneous thin coatings [27–31]. This approach, however, is still based on trial and error [24].

In macroscopic contacts, the friction force is often roughly proportional to the normal force and, thus, Amontons's law $F_x = \mu F_N$ is valid [32]. On the microscale, on the contrary, many studies show the proportionality of the frictional force to the contact area A , i.e., $F_x = \tau A$ [33–39], where τ is the (approximately constant) tangential stress in the contact area necessary to shear the interface. The widely accepted solution of this controversy is that the contact area in macroscopic contacts is roughly proportional to the normal load [40].

Even if Amontons's law is not valid, one can formally define the coefficient of friction by dividing the tangential force by the normal force, i.e., $\mu = \tau A / F_N$. The ratio A / F_N strongly depends on the contact geometry, which opens up possibilities for creating friction surfaces with predefined friction laws, even with complex specific dependencies $\mu(F_N)$. In normal contacts, the ratio A / F_N can be determined through numerical modeling, for example, by using the boundary element method (BEM) [41] or finite element method (FEM) [42]. However, tangential displacement causes the adhesive contact to lose symmetry. In this case, the contact area usually decreases significantly [21,36,43,44], while the normal force F_N changes only slowly [44–46]. Therefore, the ratio A / F_N also takes on different values. Modeling tangential adhesive contact is a complex task, and to date, the most studied case is that of a spherical indenter, for which there are analytical estimates that allow for the contact area under shear to be evaluated [36,45]. For contact geometries differing from the spherical, the situation becomes much more complicated, and therefore, studying arbitrary contact geometries requires conducting real experiments, which we propose in this work. The main goal of this work is to determine the indenter geometry in which, regardless of the magnitude of the normal load, a constant coefficient of friction $\mu = \tau A / F_N$ is realized. We found such a geometry, which opens up the possibility of manufacturing surfaces with a predetermined coefficient of friction.

A systematic design strategy for producing interfaces with preset frictional properties was proposed by Aymard et al. [24]. In a contact between a rough and a smooth interface, the controllable parameter is the topography of the rough surface. In [24], it was suggested to produce a meta interface consisting of spherical asperities. To target a specific friction law, the number of asperities and their shapes, sizes, height distributions and positions have been varied. The authors illustrate three different friction law types that can be obtained by optimizing the asperity heights.

Enhancing adhesive properties can be achieved by designing the shape of fibrillary adhesive microstructures [17]. A similar approach for the design of a surface meeting preset frictional properties by changing the geometric profile of the asperities is described in the present study.

In our previous work [38], the effect of the indentation depth of spherical indenters on the friction coefficient was studied experimentally. The present work extends this study to the frictional properties of indenters with power law profiles $f(r) = c_n r^n$. The experiments show that for $0 < n < 1$, the friction coefficient μ remains constant over a wide range of normal loads. However, the value of the coefficient of friction depends on the parameters n and c_n . Contrary to [24], where only spherical asperities were used, our design strategy allows us to control the friction coefficient μ by changing the shape of the indenter. This works even for the case of one single indenter. This approach opens opportunities to create new types of surfaces with predefined coefficients of friction. The experimental setup as well as the experimental techniques used in the present study are similar to those described in [38].

2. Materials and Methods

With the knowledge of the tangential force F_x and contact area A , the averaged shear stress $\langle \tau \rangle$ in the contact is given by:

$$\langle \tau \rangle = \frac{F_x}{A}. \quad (1)$$

Often, it is assumed that the shear stress $\langle\tau\rangle$ is equal to some constant value τ_0 , independently of the normal load or contact area [33–37]. If this assumption is valid, the friction force only depends on the *real* contact area A and is calculated as:

$$F_x = \tau_0 A. \quad (2)$$

In particular, the friction law (2) is observed in contacts with strong adhesion, e.g., when a hard indenter is pressed into a soft elastomer. This also applies to friction in contacts that are plastically deformed, where τ_0 represents the yield stress [47].

The friction coefficient μ is defined as:

$$\mu = \frac{F_x}{F_N}, \quad (3)$$

where F_N is the normal force. For most non-adhesive contacts, the friction coefficient μ is approximately constant over some range of normal forces F_N , so that the friction force law takes the classical Amontons form:

$$F_x = \mu F_N. \quad (4)$$

Equations (2) and (4) can be considered as two limiting cases that are valid for strongly adhesive and non-adhesive contacts, respectively. In the classical works [48,49], it is stated, that there exists a transition from the “adhesive” friction mode (Equation (2)) to the “normal” friction mode (Equation (4)) with increasing normal load F_N . But, in our previous experimental work [38], we did not find such a transition in a wide range of external loads. On the contrary, it was shown that the experimental results of [48] can be well described with the “adhesive” friction law (Equation (2)) in the whole range of normal forces used in [48].

The friction coefficient μ can be formally calculated using the standard definition $\mu = F_x/F_N$ (3) independently of whether it is really constant or not. If Equation (2) is valid (adhesive contact case), for a circular contact with radius a , such a formally calculated coefficient of friction is equal to

$$\mu = \frac{\tau_0 A}{F_N} = \frac{\tau_0 \pi a^2}{F_N}. \quad (5)$$

According to this equation, the coefficient of friction depends on the relationship between the contact area A and the force F_N . Consider an axially symmetric indenter with the following power law shape:

$$f(r) = c_n r^n, \quad (6)$$

where r is the radial coordinate. In our previous paper [38], the dependency of the friction coefficient on the normal force F_N was derived for axially symmetric contacts between a hard indenter and an elastic half-space, provided that the friction force is proportional to the contact area, as shown in Equation (2):

$$\mu(F_N) = \pi \tau_0 \left(\frac{(n+1)(1-\nu^2)}{2n E c_n \kappa_n} \right)^{\frac{2}{n+1}} \times \frac{1}{(F_N)^{\frac{n-1}{n+1}}}, \quad (7)$$

where κ_n are constants that can be found in [38], E is the elastic modulus of the half-space and ν is its Poisson’s ratio. Based on this equation, we can conclude the following:

“1” For $0 < n < 1$, the friction coefficient μ increases with an increase in the applied normal force F_N because the normal force F_N increases slower than the contact area A .

“2” For $n = 1$ (conical form), the friction coefficient μ does not depend on the normal force F_N :

$$\mu_{cone} = \frac{2\tau_0(1-\nu^2)}{E c_n}, \quad (8)$$

despite the fact that the friction force is still given by the equation $F_x = \tau_0 A$ (2).

“3” For $n > 1$ (for instance, for a parabolical indenter with $n = 2$), the friction coefficient μ decreases with an increase in the normal force F_N .

“4” For $n \gg 1$, the indenter turns into a flat stamp, where the contact area A does not change and is independent of the normal force F_N . In the limiting case $n \rightarrow \infty$, Equation (7) shows an asymptotic behavior:

$$\mu \propto \frac{1}{F_N}. \quad (9)$$

Note that if the chosen parameter $c_n = a_0^{1-n}$ is inserted into Equation (6), while $n \rightarrow \infty$, the profile $f(r)$ describes a cylindrical stamp with a flat base with radius a_0 .

In our previous work [38], hypothesis “3” was experimentally proven; it was also shown that there was no transition between the “adhesive” friction and “normal” friction modes in the wide range of normal forces. The present work is dedicated to the further experimental verification of all the above formulated hypotheses for adhesive contacts.

3. Results

All the experiments described below were performed in the same way as in [38] but for indenters with various geometrical shapes. Figure 1 shows the scheme of the experiment (left panel) and a real photo of the contact region of the experimental setup (right panel).

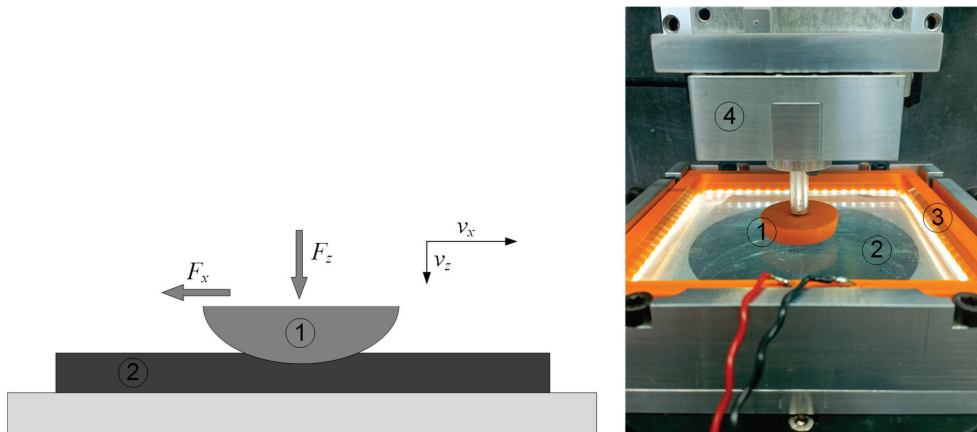


Figure 1. (Left panel) Scheme of the contact between a rigid indenter (1) and an elastic layer (2) located on a rigid glass substrate. Here, F_z and F_x are the normal and tangential contact forces, and v_x and v_z are components of the velocity of the indenter. (Right panel) Photo of the contact region between a hard indenter (1) and a much softer elastomer (2) with surrounding LED illumination (3) and a three-axis force sensor (4). Depicted in the photo is the contact configuration of a 3D-printed indenter with a spherical shape and radius $R = 100$ mm.

In the experiments, the indenters ((1) in Figure 1) were immersed in a CRG N3005 transparent elastomer sheet with a thickness of $h = 5$ mm ((2) in Figure 1). This material is a soft thermoplastic polystyrene-type gel produced by TANAC Co. Ltd., Gifu, Japan [50]. The elastomer sheet was placed on a glass substrate, which allowed for the direct observation of the contact area. To analyze the contact area, it needed to be homogeneously illuminated from all sides, which was provided by a surrounding LED light system ((3) in Figure 1). The contact forces (normal, tangential and lateral, i.e., perpendicular to tangential) were measured with a three-axis force sensor ((4) in Figure 1).

In all the experiments, the indenter was moved simultaneously in the normal and tangential directions with velocities of $v_n = 0.2 \mu\text{m/s}$ and $v_t = 5 \mu\text{m/s}$, respectively. This means that the indenter was immersed in an elastomer sheet under a small angle $\alpha = \arctan(0.2/5) \approx 2.29^\circ$. In such conditions, the contact could be characterized as a tangential contact (friction) but with a slowly increasing indentation depth and, consequently, normal force. In each experiment, the indenter was immersed in the elastomer up to the maximum distance d_{max} and was then lifted until the moment of complete detachment. Supplementary videos show the complete experiments (indentation and pull-off phases), but in this article, only indentation phase is shown. Note that there are also experimental works of other authors with similar elastomers to those we used in this article, for example, [51,52].

3.1. Spherical Profiles

We started with a partial repetition of the results for the spherical indenters with different radii of $R = 50 \text{ mm}$ and $R = 100 \text{ mm}$ from our previous work [38]. The spherical indenters corresponded to a value of $n = 2$ in the profile function $f(r)$ (6). In both experiments, the spheres were indented in the elastomeric sheet to the maximal depth $d_{\text{max}} = 0.6 \text{ mm}$. This means that both the indenters were shifted in the tangential direction up to a distance of $x_{\text{max}} = 15 \text{ mm}$. The results of these experiments can be seen in Figure 2.

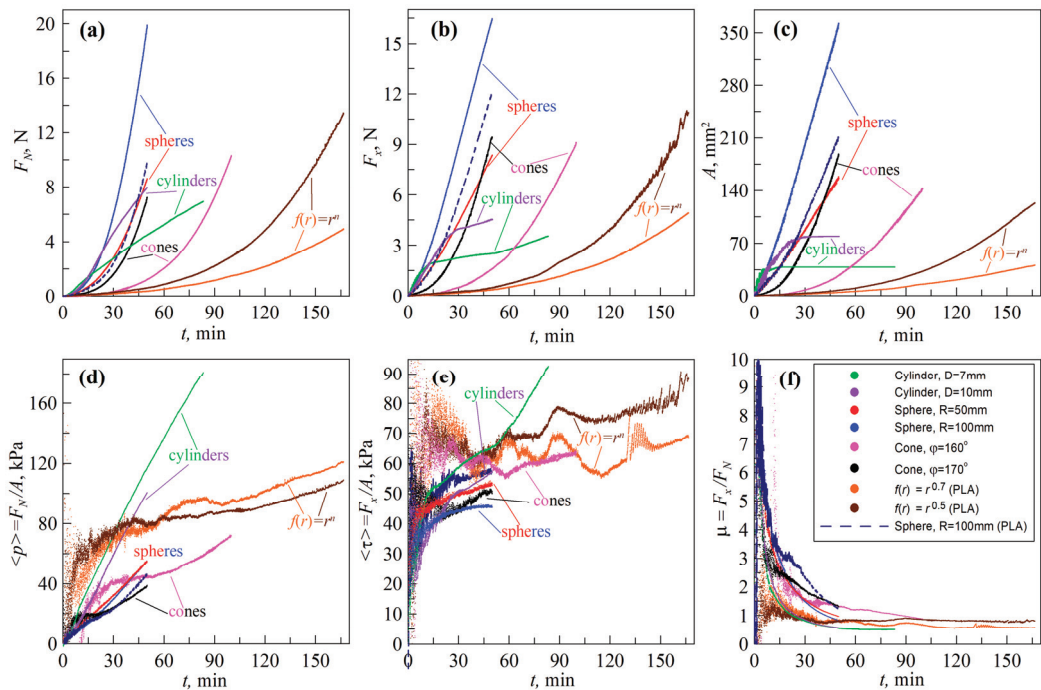


Figure 2. Experimental dependencies of normal F_N (a) and tangential F_x (b) forces, contact area A (c), average contact pressure $\langle p \rangle$ (d), average tangential stress $\langle \tau \rangle$ (e) and friction coefficient μ (f) on time t . Dependencies of indenters with different power profiles $f(r) = c_n r^n$ are presented as follows: cylindrical ($n \rightarrow \infty$), spherical ($n = 2$), conical ($n = 1$) and indenters with values $n < 1$. There were the following two dependencies for each type of indenter: cones with angles $\varphi = 160^\circ$ and 170° , spheres with radii $R = 50 \text{ mm}$ and 100 mm , cylinders with diameters $D = 10 \text{ mm}$ and 7 mm and two indenters with $n = 0.7$ and $n = 0.5$. The dashed lines show the dependencies for a 3D-printed indenter made from PLA material with radius $R = 100 \text{ mm}$; in panels (c,e), these lines appear solid due to fluctuations.

The experimental results for both indenters confirmed hypothesis “3” of Section 2, since the dependences $\mu(t)$ in Figure 2f showed a decrease in the friction coefficient with time. This means that μ decreased with increasing the normal force. Moreover, for a parabolic indenter $n = 2$, Equation (7) leads to the following formula [38,53,54]:

$$\mu(F_N) = \pi\tau_0 \left(\frac{3R}{4E^*} \right)^{\frac{2}{3}} \frac{1}{(F_N)^{1/3}}. \quad (10)$$

Figure 3b shows the dependences of $\mu(F_N)$ in a double-logarithmic scale for all the cases considered in this paper. For convenience, in this figure, several dashed lines are shown, where the upper one shows the dependence $\mu \sim 1/(F_N)^{1/3}$, and it was close to the trend of the experimental dependences for the spherical indenters (red and blue curves).

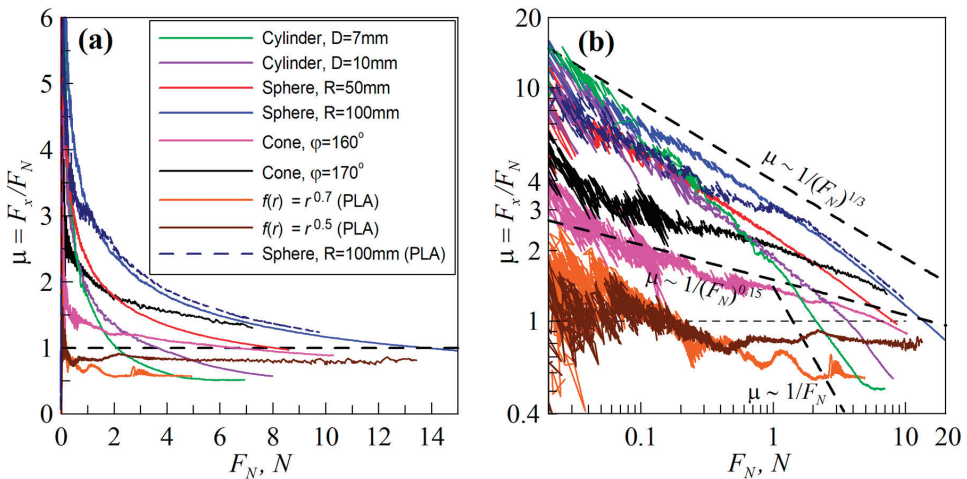


Figure 3. Experimental dependencies of the formally calculated friction coefficient $\mu = F_x/F_N$ on the normal load F_N for indenters with different geometric profiles. (a) Dependences obtained from the data shown in Figure 2a,f; (b) the same dependences in double-logarithmic coordinates.

As the normal force increased, the friction coefficient measured in the experiment began to decrease faster. The reason may be that Equation (10) was derived for the idealized case of the half-space, while the experiment was carried out with an elastomeric sheet with a final thickness of $h = 5$ mm. It was previously shown in [38] that for thinner elastomers, the coefficient of friction μ decreases faster with increasing the normal load. In general, however, experiments with special indenters have shown that Equation (10) (or, in general, Equation (7)) is satisfactorily applicable over a wide range of normal loads.

3.2. Cylindrical Profiles

In order to test the following hypothesis “4” of Section 2, experiments were carried out on indentation of cylindrical stamps with a flat base with diameters of $D = 10$ mm and 7 mm, as shown in Figure 2. A cylinder with a diameter of 10 mm was indented to a depth of $d_{\max} = 0.6$ mm, and at the same time, it was shifted tangentially by $x_{\max} = 15$ mm. A stamp with a smaller diameter of $D = 7$ mm was plunged to a depth of $d_{\max} = 1.0$ mm and shifted by $x_{\max} = 25$ mm. Here and in the following experiments, different indentation depths d_{\max} were selected for the indenters with different geometries to demonstrate the experimental dependencies over a wider range of normal forces, if possible. In the case of a cylinder, for example, the smaller the base diameter, the greater the required indentation depth to achieve a fixed level of normal force.

Equation (2) predicts that in the case of an indenter with a flat base, the contact area A , and hence, the friction force F_x , should not change during tangential motion. However, Figure 2c shows that for both cylindrical indenters, the area A increased to some maximum value at the beginning of the indentation. This was due to the fact that in the experiment, the surfaces of the cylindrical stamp base and the elastomer were not perfectly parallel. Therefore, at the beginning of the indentation, the contact was incomplete, and the contact area built up to the maximal possible value $A_{\max} = \pi D^2/4$. In this regard, during the initial shear phase, the increase in the friction force F_x (see Figure 2b) can be explained by the increase in the contact area, aligning with Equation (2). However, as evident from the experimental relationships, even after the area reached the maximum value of A_{\max} , the friction force F_x continued to increase, although the rate of its increase decreased significantly. This was partly because the value of the tangential stresses $\langle\tau\rangle$ increased with increasing the normal force. However, it should be noted that we calculated the value of $\langle\tau\rangle$ indirectly, as per the ratio $\langle\tau\rangle = F_x/A$, while the directly variable data were just the friction force (force sensor data) and the contact area (observed by the video camera). It should be taken into account that not only can the increase in the tangential stresses in the contact zone lead to an increase in the friction force but also the features observed at the edges of the contact. In the case of a spherical indenter, these effects were not significant because the contact zone grew continuously during the indentation. But, in the case of an indenter with a flat base, the effects at the contact edges could make a significant contribution to the friction force, since the elastomer sheet at the contact edges was deformed in such a way that the indenter was kind of situated in a “pit”, from which it must always “climb out” during tangential shear. This can serve as an additional channel for energy dissipation. In Supplementary Videos S1 and S2, it is possible to trace the change in the contact shape over time for both the experiments with cylindrical indenters. The videos show that during the indentation, the contact area increased to a maximum value, after which a light ring was visible at the edges of the contact, which increased in brightness and size throughout the indentation of the indenter into the elastomer. This suggests that the elastomer was highly deformed at the contact edges, which could resist the tangential movement of the indenter and result in an increase in the frictional force. As a result, the formally calculated tangential stresses $\langle\tau\rangle = F_x/A$ increased, as shown in Figure 2e, where in the case of the cylindrical indenters, the stresses $\langle\tau\rangle$ increased faster than in all the other cases.

It is more reasonable to represent the tangential stresses not as a function of the experiment time (which sets the indentation depth $d = v_z t$) but as a function of the average contact pressure $\langle p \rangle = F_N/A$. Eliminating time t from the dependencies $\langle\tau\rangle(t)$ in Figure 2e and $\langle p \rangle(t)$ in Figure 2d, the tangential stress–pressure relationship shown in Figure 4 was calculated for all the experiments.

It follows from Figure 4 that the dependence of $\tau(p)$ in many cases can be described by a power function of the following form [55,56]:

$$\tau = \tau_0 + \alpha p^\gamma \quad (11)$$

with an exponent of $\gamma \approx 0.2$. However, in the case of a cylindrical indenter with a smaller diameter $D = 7$ mm, as the normal pressure increased, a point was reached where the tangential stresses began to increase more rapidly. This aligns with the power law (11) characterized with a larger exponent of $\gamma \approx 1.0$.

Note that using τ (11) instead of τ_0 in the friction law (5) results in a two-term friction law of the following form:

$$\mu = \tau_0 \left(\frac{A}{F_N} \right) + \alpha \left(\frac{A}{F_N} \right)^{1-\gamma}, \quad (12)$$

where the expression for the average contact pressure $\langle p \rangle = F_N/A$ is used. In our experiments (see Figure 4), the exponent was $\gamma \leq 1$. Note that two-term friction laws have been

used for a long time. As examples, we can cite the classical Amontons–Coulomb friction law [32,57], Derjaguin’s law [23], the law considering friction at the boundary of adhesive contacts [46], the law describing the friction force in the boundary regime [58,59], and so on.

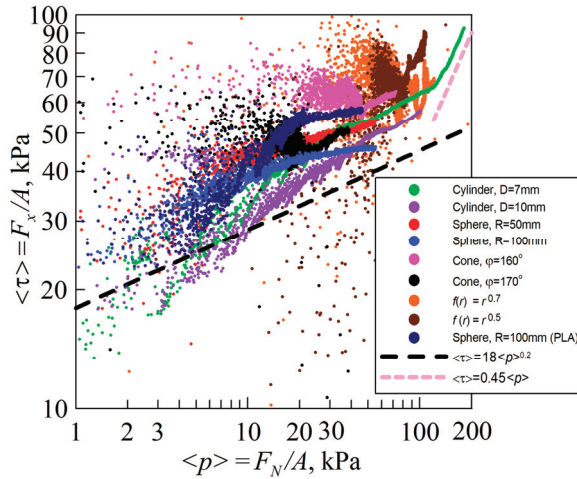


Figure 4. Experimental dependences of the mean shear stress $\langle \tau \rangle = F_x/A$ on the mean contact pressure $\langle p \rangle = F_N/A$ for all experiments, the results of which are shown in Figure 2.

Returning to the experimental results, we can conclude that during the indentation of the cylinder with a flat base, three phases can be distinguished: (1) an increase in the friction force $F_x = \langle \tau \rangle A$, mainly due to the increase in the contact area; (2) a further increase in F_x at the maximum contact area $A = A_{\max} = \pi D^2/4$ due to the growth of tangential stresses according to Equation (11); and (3) an increase in the friction force growth rate due to the even faster growth of the stresses at the subsequent indentation into the elastomer. Note that for the cylinder with diameter $D = 7$ mm, all three phases were present, and the transitions between them can be seen in the relationships shown in Figures 2b and 4. However, for the indenter with a larger diameter $D = 10$ mm, only the first two phases were realized since this experiment was performed in a smaller range of contact pressures $\langle p \rangle$.

Thus, in a certain range of parameters, hypothesis “4” of Section 2 for a cylindrical indenter with a flat base is satisfactorily fulfilled. This follows from Figure 3b, according to which both cylindrical indenters showed a range of forces (approximately at $F_N > 2$ N) in which $\mu \sim 1/F_N$. According to Figure 2, these were the forces at which the contact area was maximized and then remained constant. But, according to the same figure (Figure 2), the friction force F_x in this range of forces consistently increased, which contradicts assumption (2) that was used to derive Equation (9). Therefore, the coefficient of friction shown in Figure 3b decreased slightly slower than $\mu \sim 1/F_N$. However, in the case of the cylindrical indenter, the friction coefficient μ decreased much faster than in the above-described case of a spherical indenter, for which $\mu \sim 1/(F_N)^{1/3}$.

Note that the dependence $\mu \sim 1/F_N$ was violated for the cylindrical indenter with diameter $D = 7$ mm immediately after the change in trend (in the above-mentioned phase (3)), which can be seen in Figure 2b at $t \approx 50$ min. In this phase, the growth rates of the tangential and normal forces were the same, which led to validity of $\mu = \text{const}$ with a further increase in the normal force, as shown in Figure 3a,b for a cylinder with a diameter of $D = 7$ mm at a normal force of $F_N > 5.5$ N. This change in behavior in the $\mu(F_N)$ dependence was not caused by a higher value of the normal force but by a higher contact pressure, since for the second indenter with a diameter of $D = 10$ mm, the $\mu = \text{const}$ region was not observed despite the similar range of normal forces. In Section 2, hypothesis “2” was formulated, stating that the constancy of the friction coefficient, $\mu = \text{const}$, is expected for conical

indenters, as the contact area should increase monotonically with normal indentation, but not for a cylindrical indenter with a constant contact area. Therefore, the situation where $\mu = \text{const}$ for a cylindrical indenter is unexpected and may indicate a transition between the “adhesive” mode of friction, which is given by Equation (2), and “normal” friction (see Equation (4)) with $\mu = \text{const}$. Some classical works, e.g., [48,49], speak about the existence of such a transition, although we did not find such a transition in our recent work [38]. Moreover, it was shown in [38] that the results of [48] can be interpreted by using the single concept of “adhesion” friction over the entire range of experimental parameters. We now find that, for some reason, the cylindrical indenter transitioned to the $\mu = \text{const}$ regime at high contact pressures. This feature was demonstrated by only one of the indenters, as the second indenter was operated in a smaller pressure range. Therefore, unambiguous conclusions about the presence of the transition between friction modes cannot be made, and the study of this issue requires additional experiments. However, this is beyond the scope of the proposed work.

3.3. Conical Profiles

The next step was to test hypothesis “2” of Section 2, stating that for conical indenters, the coefficient of friction μ should be constant at all normal loads. The shape of the conical indenter is given by Equation (6) when $n = 1$, i.e., $f(r) = c_n r$. Therefore, the angle φ at the base of the cone, expressed in radians, is defined by the following formula:

$$\varphi = \pi - 2\arctan(c_n). \quad (13)$$

The case of conical indenters was investigated using two indenters made of steel with angles of $\varphi \approx 160^\circ$ and $\varphi \approx 170^\circ$. The indenter profiles were measured with a 3D laser scanning confocal microscope Keyence VK-X150 (KEYENCE DEUTSCHLAND GmbH, Neu-Isenburg, Germany) using a $10\times$ magnification objective. Figure 5a,b show the profiles of both indenters.

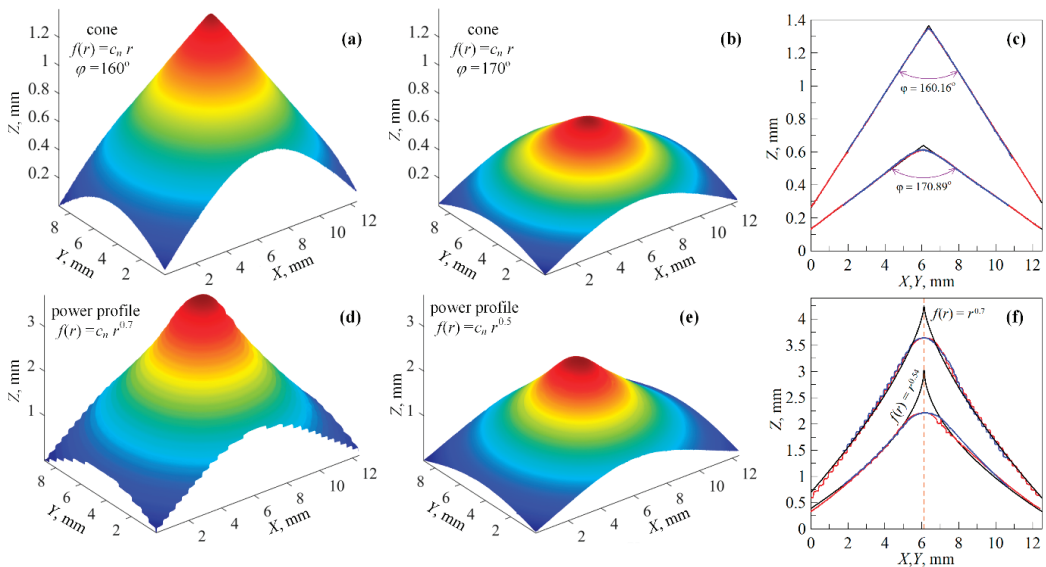


Figure 5. Measured three-dimensional profiles of conical indenters (a,b) and their two-dimensional profiles (c). Panels (d,e,f) show the same as panels (a–c) only for indenters with a profile given by formula $f(r) = c_n r^n$ with two different values of $n < 1$.

Figure 5c shows the two-dimensional profiles of the indenters corresponding to the three-dimensional images. The red and blue curves in Figure 5c show the cuts of the profiles in two perpendicular vertical planes. Their overlapping confirms the axial symmetry of the profiles. The tips of both indenters were slightly rounded; the tips of the regular cones shown in Figure 5c are completed by the straight lines shown in black. The angles between these straight lines represent the exact values of the angles φ (Equation (13)), which were 160.16° and 170.89° . In our case, the rounded corners at the base of the indenters did not interfere with the test of hypothesis “2” of Section 2, since the rounding only affected the results obtained at the beginning of the indentation process. Moreover, the use of rounded corners at the top of the conical punches avoided the occurrence of a region with a singular stress concentration [60,61], which can lead to the destruction of the elastomer surface [62], especially at a tangential indenter shift.

Figure 2 shows the experimental data for both conical indenters. Here, the maximum indentation depth for the indenter with angle $\varphi \approx 160^\circ$ was $d_{\max} = 1.2$ mm at its tangential displacement $x_{\max} = 30$ mm. The indenter with an angle $\varphi \approx 170^\circ$ was immersed into the elastomer to a smaller depth of $d_{\max} = 0.6$ mm, so it was also displaced by a smaller distance $x_{\max} = 15$ mm. As evident from Figure 3b, in both cases, the coefficient of friction decreased with increasing the external load according to the trend $\mu \sim 1/(F_N)^{0.15}$ (black and magenta curves). Thus, the rate of the decrease in the friction coefficient here was lower than in the cases of the cylindrical and spherical indenters, but the coefficient of friction still decreased with the normal force and was not constant, as predicted by Equation (8). This was primarily because the formula is derived on the assumption that the contact is axially symmetric, as in the case of purely normal indentation, but due to tangential shear, the contact quickly lost axial symmetry. At the same time, the main contact zone relative to the indenter center was concentrated at the leading edge of motion, as can be seen in Videos S3 and S4. In the case of the conical indenters, the tangential stresses $\langle \tau \rangle$ increased with the pressure $\langle p \rangle$ according to the same trend as in the cases of the spherical and cylindrical indenters considered before, which follows from Figure 4, namely, $\langle \tau \rangle \sim \langle p \rangle^{0.2}$. However, this build-up came after a decrease in the $\langle \tau \rangle$ value at the beginning of the indentation, which can be seen better in Figure 2e.

Thus, the assumption that the friction coefficient $\mu = F_x/F_N$ does not depend on the applied external load in the case of a conical indenter (hypothesis “2” of Section 2) was not confirmed.

3.4. Power Profiles with Exponent $n < 1$

The last cases considered in this paper were axially symmetric indenters, whose profiles are described by the function $f(r) = c_n r^n$ (6) with an exponent $n < 1$. Hypothesis “1” of Section 2 suggests that in this case, the coefficient of friction μ should increase with increasing the external load rather than decrease as it does for values of the exponent $n > 1$. Since we did not have equipment that would have allowed us to mill indenters from steel with an arbitrary value of the exponent n , we decided to make such indenters from plastic. We printed two indenters with values of $n = 0.7$ and $n = 0.5$ using a “QIDI TECH I Fast FDM 3D Printer” (DI JIA TECHNOLOGY LIMITED, MONGKOK, Kowloon Hong Kong SAR) from Basicfil PLA material in an orange color. In both cases, the coefficient c_n in Equation (6) was equal to 1.

Profiles of the 3D-printed indenters scanned by a 3D microscope are shown in Figure 5d,e. Figure 5f shows the 2D profiles of these indenters, where lines of different colors (red and blue) show profiles corresponding to cross-sections of the 3D surfaces in two mutually perpendicular directions. Here, the black lines show the power functions that approximate these profiles. In the case of $n = 0.7$, the real printed indenter was well described by a power function with a similar exponent. However, in the case of $n = 0.5$, it turned out that in reality, the printed indenter was close to a power function with an exponent of $n = 0.54$, which seemed to be caused by printing inaccuracies. In Figure 5f, the black curves show both profiles completed to the correct power functions of $f(r) = r^n$,

compared to which it can be seen that the indenter peaks were significantly rounded in the center. This was due to the fact that the printer used is not generally capable of printing such narrow areas with a high quality. This feature was similar to that described above with conical indenters. However, as above, the discussed rounding only affected the relationship between the contact forces at the beginning of the indentation, which did not interfere with the objectives of this paper. Figure 5 shows some other features of 3D printing. For example, in the case of $n = 0.7$ in Figure 5d,f we can clearly see the “stepping” of the profile caused by the fact that the printer prints in layers. Moreover, the indenter insignificantly deviated from the axially symmetric shape due to the printing quality.

In the experiment, both indenters were immersed to the same depth of $d_{\max} = 2$ mm, corresponding to their tangential displacement of $x_{\max} = 50$ mm. The indenter with $n = 0.5$ exhibited much higher normal force values, which can be seen in Figure 2a. At the same time, however, both indenters showed similar pressure values over the entire experimental range, as can be seen in Figure 2d. This was due to the different rate of increase in the contact area, which is shown in Figure 2c.

Hypothesis “1” of Section 2 suggests that for an indenter with a power profile of $f(r) = c_n r^n$ with an exponent of $n < 1$, the friction coefficient μ should increase with increasing the normal force. It was assumed that this should occur since the contact area, and therefore the tangential force, will increase faster than the normal force. This, according to Equation (5), should lead to an increase in the coefficient of friction with increasing the normal force. However, Figure 2f shows that, after an initial decrease, the friction coefficient remained constant. Moreover, the plateau $\mu = \text{const}$ was rather long, which was due to the fact that the indenters in the experiments were displaced by the largest distance of $x_{\max} = 50$ mm used in our experiments.

Thus, for indenters with $n < 1$, the behavior assumed for the conical indenter and formulated in hypothesis “2” in Section 2 was realized. As in the cases described above, the deviation from the assumed behavior was caused by a strong violation of the contact symmetry at the tangential shear of the indenter. The dependencies shown in Figure 2e, on average, indicate an increase in the tangential stresses $\langle \tau \rangle$ with increasing the indentation depth (or pressure). However, they were non-monotonic due to the complex processes of contact rearrangement during sliding, which can be clearly seen in Video S5 and Video S6. Note that the value of the stresses $\langle \tau \rangle$ lay in the same range as in all the other cases, despite the fact that both indenters for the case $n < 1$ were made on a 3D printer from plastic (PLA). In all the other experiments, steel indenters were used.

From Figure 3, which shows the dependencies of the friction coefficients on the normal force, it can be seen that in the discussed case $n < 1$, after a rapid decrease, the friction coefficient remained constant. This was especially clearly seen for the case $n = 0.5$, since this experiment was carried out over a larger range of normal forces. Thus, in spite of the fact that, here, the adhesive friction regime was realized, in which the tangential stresses $\langle \tau \rangle$ were close to constant regardless of the load, for this indenter shape, the formally calculated friction coefficient $\mu = F_x / F_N$ showed a constant value of $\mu = \text{const}$ in a wide range of normal forces F_N . Thus the effect of the indenter shape led to the fact that despite the realization of the friction law in the form of $F_x = \tau_0 A$ (2), the classical Amontons law $F_x = \mu F_N$ (4) with a constant friction coefficient μ was also fulfilled.

In this specific case, the coefficient of friction in Equation (4) depended on the geometrical shape of the indenter. For an indenter with an exponent $n = 0.5$, the coefficient of friction is greater than for an indenter with $n = 0.7$.

Note that although we assumed constant stresses τ_0 here, the tangential mean stresses increased slightly with pressure, as in all the previous cases. And, as indicated by Figure 4, for both indenters with exponent $n < 1$, at high pressures $\langle p \rangle$, the stresses $\langle \tau \rangle$ increased according to the power law with an exponent of $\gamma \approx 1.0$, as in the previously considered case of a cylindrical indenter with a smaller diameter of $D = 7$ mm. For a cylinder with a diameter of $D = 7$ mm, the range $\mu = \text{const}$ was also unexpectedly observed just when $\gamma \approx 1.0$, i.e., for higher pressures. Therefore, it remains possible that in the cases of indenters

with exponent $n < 1$, as well as in the case of a cylindrical indenter with a diameter of $D = 7$ mm, we are dealing with a transition between “adhesive” and “normal” friction, the presence of which is indicated by [48,49] but was not detected by us in [38]. The presence of such a transition may be due to the fact that in these three cases, the highest values of the contact pressure $\langle p \rangle$ were realized, as can be seen in Figure 2d. To verify the presence of such a transition between the friction modes, it is necessary to conduct additional experiments in the range of high contact pressures, which is beyond the scope of this paper.

4. Discussion

The effect described in Section 3.4 is that for indenters with power law profiles with an exponent of $n < 1$, the regime of a constant friction coefficient $\mu = \text{const}$ was realized. However, this effect may have also been due to the fact that these indenters were made of plastic (PLA). After all, the properties of the adhesive contact (e.g., the specific work of adhesion) strongly depend on the properties of both contacting surfaces, and all the cases described in this paper for indenters with $n \geq 1$ were carried out with indenters made of steel. Therefore, the cases with indenters $n \geq 1$ (steel) and $n < 1$ (PLA) cannot be compared unambiguously, and in order to understand the differences, it is necessary to establish some kind of “bridge” between them, which would help to make sure that it was the shape’s effect and not the material’s effect. An additional experiment with a spherical indenter with a radius of $R = 100$ mm acted as such a “bridge”. The experiment was completely similar to the one described in Section 3.1 but with the difference that an indenter printed on a 3D printer from the same PLA as the indenters with an exponent of $n < 1$ (see Section 3.4) was used. The results of this additional experiment are shown in the figures above with dashed lines. The detailed course of the experiment can be seen in Supplementary Video S7. Note that the data corresponding to indenters with radii $R = 100$ mm made of steel and PLA were different. This was due to the fact that the used 3D printer prints in layers, so the indenter had a stepped shape, as with the indenters shown in Figure 5d,f. However as can be clearly seen in Figure 3, the $\mu(F_N)$ dependence of the PLA indenter showed exactly the same trend as that of the steel indenter with the same radius $R = 100$ mm. Therefore, the discovered effect, i.e., that for indenters with an exponent of $n < 1$, the friction coefficient does not depend on the external load, was not related to the indenter material but was caused by its specific geometrical shape, in which the contact area increased as fast as the normal force. In this case, according to the expression $\mu = \tau_0 A / F_N$ (5), we have a situation with $\mu = \text{const}$.

In our previous work [38], where adhesive contact was investigated, it was hypothesized that a contact-geometry-dependent and, at the same time, normal-force-independent coefficient of friction should be realized in the case of a conical indenter. In this case, $\mu = \text{const}$ is unique because in it the mutually exclusive laws of friction $F_x = \tau_0 A$ (2) and $F_x = \mu F_N$ (4), where both the tangential stress τ_0 and the coefficient of friction μ are constants, must be simultaneously valid. However, the hypothesis (put forward in [38]) would only be valid if the contact remained axially symmetric during tangential motion. But, in the adhesive contact, the breaking of the axial symmetry at tangential shear violates the hypotheses formulated in [38], which are briefly described in Section 2 of the present paper. However, the general tendency still remains, namely, that with decreasing the exponent n in the profile function $f(r) = c_n r^n$, the friction coefficient μ decreases more and more slowly with increasing the external normal force F_N . Moreover, the form of an indenter for which the laws of friction (2) and (4) are simultaneously valid was found, and it was with an indenter whose profile is given by the power function $f(r) = c_n r^n$ with exponent $n < 1$.

The cases of a constant coefficient of friction, $\mu = \text{const}$, for indenters with $n < 1$, which were studied experimentally in this paper, are of high application value [63]. Here, we studied indenters with two different values of $n = 0.5$ and $n = 0.7$. According to Figure 3, in the case of $n = 0.7$, the coefficient of friction was lower. Thus, there is an indenter shape for which the coefficient of friction on the adhesive contact is constant and does not depend on

the applied load. Moreover, by varying the shape of this indenter, the coefficient of friction can be set. One potential application of the found effect is the creation of surfaces that exhibit constant coefficients of friction in adhesive contacts. On these surfaces, at a certain distance from each other, protrusions should be placed whose profiles would geometrically represent the functions $f(r) = c_n r^n$ with an exponent of $n < 1$. Moreover, these can be, among others, microscopic surfaces with nanoscale protrusions. The protrusions under discussion can be located on the surfaces periodically or randomly, which requires additional study.

Note that in this paper, we often talk about contact symmetry breaking in tangential motion without giving pictures of the contact region. However, the detailed evolution of the contact areas for all the experiments performed can be observed in the supplementary videos to this paper, which are an important part of the paper. Moreover, these videos not only show the phase of the indentation of the indenter at an angle into the elastomer but also the pull-off phase the indenter in the normal direction, which was performed after the indenter reached the maximum indentation depth d_{\max} and tangential displacement x_{\max} .

5. Conclusions

This paper experimentally investigated the influence of the geometric shape of an axially symmetric indenter, which is given by the power law $f(r) = c_n r^n$, on the coefficient of friction in an adhesive contact. Spherical ($n = 2$), conical ($n = 1$) and indenters with an index of $n < 1$ (with $n = 0.5$ and $n = 0.7$) as well as flat-ended stamps were studied. It was shown that the friction coefficient $\mu = F_x/F_N$ decreased with increasing the external force F_N applied to the friction surfaces. This confirms the analytically determined tendency showing that for indenters with smaller indices of n , the friction coefficient decreases slower with increasing the force F_N . It was experimentally shown that in the case when the exponent $n < 1$, the friction coefficient took a constant value independent of the external load. In this case, two laws of friction, which are normally thought of as mutually exclusive, were simultaneously fulfilled. The first of them says that the friction force is proportional to the contact area, since constant-contact tangential stresses independent of the external load are realized. The second law of friction is the classical Amontons law, according to which the coefficient of friction does not depend on the external load. However, this does not mean that the true Amontons law is realized. In the case $n < 1$, the adhesive tangential contact, as before, was characterized by a constant value of the tangential stress τ_0 , which slightly increased with increasing the pressure in the contact interface. In this case, the friction coefficient $\mu = F_x/F_N$, although constant, depends on the exponent n , so it loses its original meaning as a material constant. The found mode can be used for creating surfaces with a given coefficient of friction. These surfaces should have a large number of protrusions with profiles $f(r) = c_n r^n$ with the power $n < 1$. Thus, we proposed a method of creating surfaces with a given friction coefficient, which can be varied by the geometric shape of the indenters (protrusions or roughness on the surfaces). This work is also of fundamental interest because it advances the understanding of the dependence of the coefficient of friction between rough surfaces on the geometric characteristics of their topography.

The question remains open as to what shape an indenter should have so that the coefficient of friction increases with increasing load. To realize such a situation, the contact area must increase faster than the normal force. In normal contact, this condition is realized for $n < 1$. But, in the presence of tangential displacement, as this experiment has shown, due to the symmetry breaking of the contact area, the relationship between the contact area and the normal force is broken, so for indenters with $n < 1$, a constant friction coefficient $\mu = \text{const}$ is observed.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pr12061209/s1>, Video S1: An indenter with a cylindrical base with diameter $D = 7$ mm, which was immersed to a depth of $d_{\max} = 1$ mm with a velocity of $v_n = 0.2$ $\mu\text{m/s}$ into a layer of TANAC CRG N3005 elastomer with a thickness of $h = 5$ mm. At the same time, the indenter was tangentially displaced with a velocity of $v_t = 5$ $\mu\text{m/s}$ until the maximum

distance of $x_{\max} = 25$ mm. After reaching the maximum depth, the indenter was pulled off with the same velocity v_n out of the elastomer layer until the contact was completely broken. Separate panels in the video show the time dependencies of the normal (F_N) and tangential (F_x) forces, the average tangential stress ($\langle \tau \rangle = F_x/A$), the average normal pressure ($\langle p \rangle = F_N/A$), the contact area (A) and the friction coefficient (μ). In addition, the video shows the evolution of the contact zone; it also shows the current values of the indentation depth (d), the tangential shift of the indenter (x) and the time (t) that has passed since the beginning of the indentation. The video relates to the lime-green lines in Figures 2–4 in the article. Video S2: This is similar to Video S1, with the difference being that in this case, the indenter had a base diameter of $D = 10$ mm, and a different indentation depth of $d_{\max} = 0.6$ mm and tangential shift $x_{\max} = 15$ mm were realized. The video relates to the violet lines in Figures 2–4 in the article. Video S3: This is similar to Video S1, with the difference being that in this case, a conical indenter with an angle of $\varphi \approx 160^\circ$ and a different indentation depth of $d_{\max} = 1.2$ mm and tangential shift $x_{\max} = 30$ mm were used. The video relates to the pink lines in Figures 2–4 in the article. Video S4: This is similar to Video S3, with the difference being that in this case, the indenter had an angle of $\varphi \approx 170^\circ$, and a different indentation depth of $d_{\max} = 0.6$ mm and tangential shift $x_{\max} = 15$ mm were realized. The video relates to the black lines in Figures 2–4 in the article. Video S5: This is similar to Video S1, with the difference being that in this case, a 3D-printed indenter of PLA material with a power profile ($f(r) = r^{0.5}$) and a different indentation depth of $d_{\max} = 2$ mm and tangential shift $x_{\max} = 50$ mm were used. The video relates to the brown lines in Figures 2–4 in the article. Video S6: This is similar to Video S5, with the difference being that in this case, the indenter had a profile ($f(r) = r^{0.7}$). The video relates to the orange lines in Figures 2–4 in the article. Video S7: This is similar to Video S1, with the difference being that in this case, a 3D-printed spherical indenter of PLA material with radius $R = 100$ mm and a different indentation depth of $d_{\max} = 0.6$ mm and tangential shift $x_{\max} = 15$ mm were used. The video relates to the dashed dark blue lines in Figures 2–4 in the article.

Author Contributions: Conceptualization, scientific supervision, project administration, writing—review and editing, V.L.P.; methodology, hardware, conducting of experiments, theoretical model and theoretical data analysis, validation, I.A.L.; software, experimental data analysis, visualization, writing—original draft preparation, I.A.L. and T.H.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Deutsche Forschungsgemeinschaft (Project DFG PO 810/55-3).

Data Availability Statement: The datasets generated for this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Holmberg, K.; Kivikytö-Reponen, P.; Härkisaari, P.; Valtonen, K.; Erdemir, A. Global energy consumption due to friction and wear in the mining industry. *Tribol. Int.* **2017**, *115*, 116–139. [CrossRef]
- Holmberg, K.; Erdemir, A. Global impact of friction on energy consumption, economy and environment. *FME Trans.* **2015**, *43*, 181–185. [CrossRef]
- Kuznetsov, V.P.; Tarasov, S.Y.; Dmitriev, A.I. Nanostructuring burnishing and subsurface shear instability. *J. Mater. Process. Technol.* **2015**, *217*, 327–335. [CrossRef]
- Li, W.; Yang, X.; Wang, S.; Xiao, J.; Hou, Q. Research and prospect of ceramics for automotive disc-brakes. *Ceram. Int.* **2021**, *47*, 10442–10463. [CrossRef]
- Oh, Y.; Lee, H. Characteristics of a Tire Friction and Performances of a Braking in a High Speed Driving. *Adv. Mech. Eng.* **2014**, *6*, 260428. [CrossRef]
- Costa, H.L.; Schille, J.; Rosenkranz, A. Tailored surface textures to increase friction—A review. *Friction* **2022**, *10*, 1285–1304. [CrossRef]
- Holmberg, K.; Andersson, P.; Erdemir, A. Global energy consumption due to friction in passenger cars. *Tribol. Int.* **2012**, *47*, 221–234. [CrossRef]
- Pedrero, J.I.; Martínez-López, D.; Calvo-Irisarri, J.; Pleguezuelos, M.; Sanchez, M.B.; Fernandez-Sison, A. Minimum friction losses in wind turbine gearboxes. *Forsch. Ingenieurwes* **2022**, *86*, 321–330. [CrossRef]
- Evangelista, I.; Wencel, D.; Beguin, S.; Zhang, N.; Gilchrist, M.D. Influence of Surface Texturing on the Dry Tribological Properties of Polymers in Medical Devices. *Polymers* **2023**, *15*, 2858. [CrossRef]
- Chernov, S.V.; Makukha, Z.M.; Protsenko, I.Y.; Nepijko, S.A.; Elmers, H.J.; Schönhense, G. Test object for emission electron microscope. *Appl. Phys. A* **2014**, *114*, 1383–1385. [CrossRef]

11. Basu, A.; Adams, G.G.; McGruer, N.E. A review of micro-contact physics, materials, and failure mechanisms in direct-contact RF MEMS switches. *J. Micromech. Microeng.* **2016**, *26*, 104004. [CrossRef]
12. Kesari, H.; Lew, A.J. Effective macroscopic adhesive contact behavior induced by small surface roughness. *J. Mech. Phys. Solids* **2011**, *59*, 2488–2510. [CrossRef]
13. Lyashenko, I.A.; Liashenko, Z.M. Influence of tangential displacement on the adhesion force between gradient materials. *Ukr. J. Phys.* **2020**, *65*, 205–216. [CrossRef]
14. Lin, H.; Gomez, I.; Meredith, J.C. Pollenkitt wetting mechanism enables species-specific tunable pollen adhesion. *Langmuir* **2013**, *29*, 3012–3023. [CrossRef]
15. Takahashi, K.; Berengueres, J.O.L.; Obata, K.J.; Saito, S. Geckos' foot hair structure and their ability to hang from rough surfaces and move quickly. *Int. J. Adhes. Adhes.* **2006**, *26*, 639–643. [CrossRef]
16. Geim, A.K.; Dubonos, S.V.; Grigorieva, I.V.; Novoselov, K.S.; Zhukov, A.A.; Shapoval, S.Y. Microfabricated adhesive mimicking gecko foot-hair. *Nat. Mater.* **2003**, *2*, 461–463. [CrossRef]
17. Gorb, S.; Varenberg, M.; Peressadko, A.; Tuma, J. Biomimetic mushroom-shaped fibrillar adhesive microstructure. *J. R. Soc. Interface* **2007**, *4*, 271–275. [CrossRef]
18. Strivens, T.A. 16—Mechanical properties of paints and coatings. In *Woodhead Publishing Series in Metals and Surface Engineering, Paint and Surface Coatings*, 2nd ed.; Lambourne, R., Strivens, T.A., Eds.; Woodhead Publishing: Cambridge, UK, 1999; pp. 598–620. ISBN 9781855733480. [CrossRef]
19. Jurišić, S.; Jurišić, G.; Jurić, H. Influence of Adhesives and Methods of Enamel Pretreatment on the Shear Bond Strength of Orthodontic Brackets. *Acta Stomatol. Croat.* **2015**, *49*, 269–274. [CrossRef] [PubMed]
20. Scarselli, G.; Corcione, C.; Nicassio, F.; Maffezzoli, A. Adhesive joints with improved mechanical properties for aerospace applications. *Int. J. Adhes. Adhes.* **2017**, *75*, 174–180. [CrossRef]
21. Waters, J.F.; Guduru, P.R. Mode-mixity-dependent adhesive contact of a sphere on a plane surface. *Proc. R. Soc. A* **2010**, *466*, 1303–1325. [CrossRef]
22. Pizzi, A.; Mittal, K.L. *Handbook of Adhesive Technology*, 3rd ed.; CRC Press: Boca Raton, FL, USA, 2017. [CrossRef]
23. Derjaguin, B. Molekulartheorie der äußeren Reibung. *Z. Phys.* **1934**, *88*, 661–675. [CrossRef]
24. Aymard, A.; Delplanque, E.; Dalmas, D.; Scheibert, J. Designing metainterfaces with specified friction laws. *Science* **2024**, *383*, 200–204. [CrossRef] [PubMed]
25. Wos, S.; Koszela, W.; Pawlus, P. The effect of graphite surface texturing on the friction reduction in dry contact. *Tribol. Int.* **2020**, *151*, 106535. [CrossRef]
26. Ibatan, T.; Uddin, M.S.; Chowdhury, M.A.K. Recent development on surface texturing in enhancing tribological performance of bearing sliders. *Surf. Coat. Technol.* **2015**, *272*, 102–120. [CrossRef]
27. Flater, E.E.; Corwin, A.D.; de Boer, M.P.; Carpick, R.W. In situ wear studies of surface micromachined interfaces subject to controlled loading. *Wear* **2006**, *260*, 580–593. [CrossRef]
28. Li, N.; Xu, E.; Liu, Z.; Wang, X.; Liu, L. Tuning apparent friction coefficient by controlled patterning bulk metallic glasses surfaces. *Sci. Rep.* **2016**, *6*, 39388. [CrossRef] [PubMed]
29. Gachot, C.; Rosenkranz, A.; Reinert, L.; Ramos-Moore, E.; Souza, N.; Müser, M.H.; Mücklich, F. Dry Friction Between Laser-Patterned Surfaces: Role of Alignment, Structural Wavelength and Surface Chemistry. *Tribol. Lett.* **2013**, *49*, 193–202. [CrossRef]
30. Rosenkranz, A.; Reinert, L.; Gachot, C.; Mücklich, F. Alignment and wear debris effects between laser-patterned steel surfaces under dry sliding conditions. *Wear* **2014**, *318*, 49–61. [CrossRef]
31. Pakhaliuk, V.I.; Vasilets, V.N.; Poliakov, A.M.; Torkhov, N.A. Reducing the Wear of the UHMWPE Used in the Total Hip Replacement after Low-Pressure Plasma Treatment. *J. Appl. Comput. Mech.* **2022**, *8*, 1035–1042. [CrossRef]
32. Popova, E.; Popov, V.L. The research works of Coulomb and Amontons and generalized laws of friction. *Friction* **2015**, *3*, 183–190. [CrossRef]
33. Carpick, R.W.; Salmeron, M. Scratching the Surface: Fundamental Investigations of Tribology with Atomic Force Microscopy. *Chem. Rev.* **1997**, *97*, 1163–1194. [CrossRef] [PubMed]
34. Degrandi-Contraires, E.; Poulard, C.; Restagno, F.; Léger, L. Sliding friction at soft micropatterned elastomer interfaces. *Faraday Discuss.* **2012**, *156*, 255–265. [CrossRef] [PubMed]
35. Yashima, S.; Romero, V.; Wandersman, E.; Frétigny, C.; Chaudhury, M.K.; Chateauinois, A.; Prevost, A.M. Normal contact and friction of rubber with model randomly rough surfaces. *Soft Matter* **2015**, *11*, 871–881. [CrossRef] [PubMed]
36. Sahli, R.; Pallares, G.; Ducottet, C.; Ben Ali, I.E.; Al Akhrass, S.; Guibert, M.; Scheibert, J. Evolution of real contact area under shear and the value of static friction of soft materials. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 471–476. [CrossRef] [PubMed]
37. Mergel, J.C.; Scheibert, J.; Sauer, R.A. Contact with coupled adhesion and friction: Computational framework, applications, and new insights. *J. Mech. Phys. Solids* **2021**, *146*, 104194. [CrossRef]
38. Lyashenko, I.A.; Pham, T.H.; Popov, V.L. Effect of Indentation Depth on Friction Coefficient in Adhesive Contacts: Experiment and Simulation. *Biomimetics* **2024**, *9*, 52. [CrossRef] [PubMed]
39. Schwarz, U.D.; Zwörner, O.; Köster, P.; Wiesendanger, R. Quantitative analysis of the frictional properties of solid materials at low loads. I. Carbon compounds. *Phys. Rev. B* **1997**, *56*, 6987–6996. [CrossRef]
40. Barber, J.R. Multiscale Surfaces and Amontons' Law of Friction. *Tribol. Lett.* **2013**, *49*, 539–543. [CrossRef]

41. Li, Q.; Pohrt, R.; Lyashenko, I.A.; Popov, V.L. Boundary element method for nonadhesive and adhesive contacts of a coated elastic half-space. *P. I Mech. Eng. J-J. Eng.* **2020**, *234*, 73–83. [CrossRef]
42. Forsbach, F. A Simple Semi-Analytical Method for Solving Axisymmetric Contact Problems Involving Bonded and Unbonded Layers of Arbitrary Thickness. *Machines* **2023**, *11*, 474. [CrossRef]
43. Das, D.; Chasiotis, I. Sliding of adhesive nanoscale polymer contacts. *J. Mech. Phys. Solids* **2020**, *140*, 103931. [CrossRef]
44. Wilhayn, J.; Lyashenko, I.A.; Li, Q.; Popov, V.L. Influence of tangential sliding on the contact area of a macroscopic adhesive contact. *Facta Univ. Ser. Mech. Eng.* **2024**. [CrossRef]
45. Argatov, I.I.; Lyashenko, I.A.; Popov, V.L. Adhesive sliding with a nominal point contact: Postpredictive analysis. *Int. J. Eng. Sci.* **2024**, *200*, 104055. [CrossRef]
46. Popov, V.L.; Li, Q.; Lyashenko, I.A.; Pohrt, R. Adhesion and friction in hard and soft contacts: Theory and experiment. *Friction* **2021**, *9*, 1688–1706. [CrossRef]
47. Shooter, K.V.; Tabor, D. The Frictional Properties of Plastics. *Proc. Phys. Soc. B* **1952**, *65*, 661. [CrossRef]
48. McFarlane, J.S.; Tabor, D. Relation between friction and adhesion. *Proc. R. Soc. Lond. A* **1950**, *202*, 244–253. [CrossRef]
49. Homola, A.M.; Israelachvili, J.N.; McGuiggan, P.M.; Gee, M.L. Fundamental experimental studies in tribology: The transition from “interfacial” friction of undamaged molecularly smooth surfaces to “normal” friction with wear. *Wear* **1990**, *136*, 65–83. [CrossRef]
50. Electronic Resource: Innovation Company TANAC Co., Ltd. Available online: <https://www.k-tanac.co.jp/crystalnone> (accessed on 9 May 2024).
51. Morishita, M.; Kobayashi, M.; Yamaguchi, T.; Doi, M. Observation of spatio-temporal structure in stick–slip motion of an adhesive gel sheet. *J. Phys. Condens. Matter* **2010**, *22*, 365104. [CrossRef]
52. Yamaguchi, T.; Ohmata, S.; Doi, M. Regular to chaotic transition of stick–slip motion in sliding friction of an adhesive gel-sheet. *J. Phys. Condens. Matter* **2009**, *21*, 205105. [CrossRef]
53. Roberts, E.W. Ultralow friction films of MoS₂ for space applications. *Thin Solid. Film.* **1989**, *181*, 461–473. [CrossRef]
54. Martin, J.-M. 13—Superlubricity of Molybdenum Disulfide. In *Superlubricity*; Erdemir, A., Martin, J.-M., Eds.; Elsevier: Amsterdam, The Netherlands, 2007; pp. 207–225. ISBN 978-0-444-52772-1. [CrossRef]
55. Briscoe, B.J.; Tabor, D. The effect of pressure on the frictional properties of polymers. *Wear* **1975**, *34*, 29–38. [CrossRef]
56. Briscoe, B.J. Interfacial friction of polymer composites. general fundamental principles. In *Friction and Wear of Polymer Composites*; Friedrich, K., Ed.; Composite Materials Series; Elsevier: Amsterdam, The Netherlands, 1986; Volume 1, pp. 25–59.
57. Coulomb, C.A. Theorie des Machines Simples. *Mem. Math. Phys. Acad. Sci.* **1785**, *10*, 161–331.
58. Lyashenko, I.A.; Khomenko, A.V.; Metlov, L.S. Nonlinear thermodynamic model of boundary friction. *J. Frict. Wear* **2011**, *32*, 113–123. [CrossRef]
59. Lyashenko, I.A.; Khomenko, A.V.; Zaskoka, A.M. Hysteresis Behavior in the Stick-Slip Mode at the Boundary Friction. *Tribol. Trans.* **2013**, *56*, 1019–1026. [CrossRef]
60. Ciavarella, M. Indentation by nominally flat or conical indenters with rounded corners. *Int. J. Solids Struct.* **1999**, *36*, 4149–4181. [CrossRef]
61. Zhou, W.; Yang, F. Effects of surface stress on the indentation response of an elastic half-space. *Int. J. Mech. Sci.* **2022**, *229*, 107512. [CrossRef]
62. Fooladi Mahani, S.; Liu, C.; Sousa Machado, P.V.; Lin, L.L.; Wen, X.; Jiménez-Piqué, E.; Llanes, L. Contact damage induced by means of conical indentation in hardmetals: Microstructural effects on residual strength. *Int. J. Refract. Hard Met.* **2024**, *118*, 106446. [CrossRef]
63. Slesarenko, V.; Pastewka, L. The bumpy road to friction control. *Science* **2024**, *383*, 150. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Acoustic Effects of Uneven Polymeric Layers on Tunable SAW Oscillators

Ionut Nicolae, Mihaela Bojan and Cristian Viespe *

National Institute for Laser, Plasma and Radiation Physics, 077125 Magurele, Romania;
ionut.nicolae@inflpr.ro (I.N.); mihaela.bojan@inflpr.ro (M.B.)

* Correspondence: cristian.viespe@inflpr.ro

Abstract: Surface acoustic wave (SAW) sensors in tunable oscillator configuration, with a deposited polymeric layer, were used to investigate the layer's impact on the oscillator's resonant frequency. The SAW oscillators were tuned by means of variable loop amplification. Full-range amplification variation led to a resonant frequency increase of ~ 1.7 MHz due to the layer's nonlinear reaction. The layer's morphology and location resulted in a specific resonant frequency–amplitude dependence. Five types of layers were used to test the causal linkage between the layers' morphological parameters or positioning and the SAW oscillator's resonant frequency. The frequency variation trend is almost linear, with a complex minute variation. Small amplitude sigmoids occur at certain attenuation values, due to layer acoustic resonances. Multiple sigmoids were linked with layer resonances of different orders. A good correlation between the layer's thickness and resonance position was found.

Keywords: SAW sensor; tunable oscillator; uneven layer; polymer

1. Introduction

Surface acoustic wave (SAW) delay lines are finding new applications all across the industry, e.g., wireless sensing [1,2], temperature measurement [3], harsh environment monitoring [4–7], lab on chip [8], strain monitoring [9], etc. This interest is well justified, given that SAWs are devices with great potential owing to the complex acousto-electrical interactions governing their functionality. For many years, various techniques to bring this potential to actual use were explored [10–13]. Both frequency variation and amplitude–frequency correlation were evaluated. Furthermore, with a closed-loop oscillator being the most common configuration in which they are deployed, SAW sensors in this configuration inherit all oscillator-specific complexities. It is this specific configuration that enables one of the most interesting applications, which is sensor-based analytics. Additionally, given the extended areas of application outlined above, a technique for extracting complex information from a SAW oscillator would expand its applicability far beyond chemical sensing. As an oscillator at resonance is most sensitive to changes, the idea of a SAW sensor in a multi-frequency oscillator configuration is being diligently pursued by the research community. A recent development that may achieve this goal is the tunable SAW oscillator [14]. In this approach, the oscillator's resonant frequency changes due to amplitude variation. Due to the chemoselective layer's non-linear reaction, changing the signal's amplitude leads to a variation in the oscillator's resonant frequency. The resulting frequency–amplitude (F-A) characteristic is directly affected by the chemoselective layer's non-linear acoustic response [14], making it suitable for usage in a gas detection/identification protocol, as well as complex material characterization. However, reliable data can be achieved only by taking into account all elements impacting the detection process outcome. Among these elements, the chemoselective layer's morphology is of utmost importance. Random thickness variations across the layer's surface that routinely plague the SAW sensors are immediately revealed by F-A characteristics. This renders a detection-identification protocol inoperable, since establishing a baseline is not possible. To circumvent this problem, it is necessary

Citation: Nicolae, I.; Bojan, M.; Viespe, C. Acoustic Effects of Uneven Polymeric Layers on Tunable SAW Oscillators. *Processes* **2024**, *12*, 1217. <https://doi.org/10.3390/pr12061217>

Academic Editors: Sheng Du, Li Jin, Xiongbo Wan and Zixin Huang

Received: 17 April 2024

Revised: 11 June 2024

Accepted: 12 June 2024

Published: 13 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

to identify the frequency–amplitude dependence features that are caused by the layer’s unwanted morphological characteristics.

At this point, we can mention that, besides enabling analytical capabilities, given their dependence on temperature [15,16], an added benefit of such a technique would be eliminating environmental influences, such as temperature variations. Furthermore, it can be applied to other acoustic-based devices with operating modes based on shear oscillation. Quartz crystal micro-balance (QCM) was successfully tested in real-time DNA detection [17] and humidity detection [18]. For applications in molecular biology, the functionality of such a device could be improved by setting up a multi-layer structure with different biochemical functionality in each component sub-layer. Using amplitude variation, the loading degree of each sub-layer could be measured. While such a setup is compatible with an SAW device, a QCM is preferable due to its ability to operate in a liquid environment. Another attractive option is film bulk acoustic resonators (FBARs). This is well justified by their ability to operate at very high frequencies (up to 20 GHz), as well as the notable recent progress in their manufacture [19] and design optimization [20].

This study is intended as a step toward developing an acoustic wave sensor with analytical capabilities, by investigating the effects of layer morphology on F-A dependence.

In our previous research, a large area layer with an average thickness of 0.5 μm was used. While the proposed model in [14] was successfully used to fit the experimental data, it failed to work for smaller area layers of the same material presented in this paper. The aim of this study is to lay the groundwork for the development of this model by outlining the interactions that characterize a layer of non-uniform thickness. In doing so, we will establish in as much detail as possible the effects of the layer’s thickness variation on F-A characteristics. We presumed that while layer thickness and thickness variation amplitude are determinants in the interaction’s outcome, the location with regard to input/output interdigital transducers (IDTs) also plays an important role. Thus, a two-part experiment was envisioned, firstly, to test the location impact by placing identical layers at different locations on the waves’ paths, and secondly, to test the layers of different morphologies. The necessity to isolate the location’s impact required the use of small-diameter layers. Additionally, a small diameter facilitated more accurate layer surface mapping. On the other hand, measurable acoustic effects are the result of thick layers; thus, we opted for a small area and thick layers of polyethyleneimine (PEI). It is expected that the magnitude of unevenness effects is proportional to thickness gradients, so a high gradient of thickness variation due to the small radius of layers is an added benefit.

2. Materials and Methods

Five types of PEI layers were used for investigating the layer morphology and position impact on a tunable SAW closed loop oscillator. In the following description, the layer’s type will be referred to by a capital letter from A to E.

Typically, the morphology of a PEI layer is characterized by thickness variation across its surface. This is both due to random processes during deposition, and, in case of viscous polymers, due to post-deposition fragmentation effects (the so-called “shark skin” effect). To avoid the formation of a layer with complex random relief during deposition was another reason why we opted for small-area layers. Additionally, a small area confers morphological stability, preventing the post-deposition formation of random relief.

To acquire F-A data, we used a setup similar to that presented in [14]. Briefly, the SAW oscillator loop amplification was reduced using a PC-controlled potentiometer. The potentiometer was actuated by a high-precision rotary stage and, for each attenuation value, the corresponding frequency was measured by a CNT-90 frequency counter. A computer program was developed using LabView 2013 IDE (National Instruments, Austin, TX, USA) for measurement process automation and data acquisition, integrating the functionality of both the rotary stage controller and frequency counter. This way, we acquired A-F data at a rigorously identical signal attenuation for each measurement. To identify the relevant morphological parameters of the layers, their thicknesses were measured with an Xi-100

non-contact interferential profilometer made by Ambios Technology Inc. (Santa Cruz, CA, USA). To establish the positioning impact on F-A dependence, it was first necessary to deposit layers with identical F-A characteristics. For this purpose, a FinnTip 10 (Thermo Fisher Scientific Waltham, MA, USA) pipette tip was dipped into PEI solution, allowing a small quantity of liquid to be retained by capillary forces. The retained solution was transferred onto the SAW device by slightly touching the lower tip opening to the piezo-active area. Previous to the deposition process, an extremely thin (<10 kHz) hydrophobic layer was applied on the piezo-active surface. This prevented the uncontrolled expansion of the PEI solution droplet on the surface, resulting in the formation of a reproducible layer after solvent evaporation. To verify the reproducibility of the coating, the deposition technique was tested multiple times. As a reproducibility criterion, the identity of F-A characteristics corresponding to deposited layers was considered. Once the deposition technique for the reproducible layer was established, it was used to deposit one type A layer at distances from the input IDT of 1, 2.5, 5, and 7.5 mm. The type A layers were deposited by the described technique using 5% by weight PEI ethanol solution.

To determine the layer's morphology impact on F-A characteristics, one layer of each type was deposited on the piezo-active area's center, followed by F-A characteristics and layer thickness measurements. The type B layer was deposited by the same technique as type A, but using 1% by weight PEI solution. Type C, D, and E layers were deposited following the same procedure except for the final step, which consisted of drop-casting 0.4, 0.6, and 0.8 μL 1% by weight PEI solution.

The shape of the A, C, D, and E layers was circular, with radii of 2.1, 1.5, 1.6, 2, 2.8 mm. Type B developed a contiguous irregular shape circumscribed inside a circular area of 1.5 mm diameter.

The analysis of the F-A characteristics was based on the model developed in [21]. We adopted the same method as in [21]: x and z in-plane with the piezo-active area, with z along the acoustic wave propagation direction. A layer deposited on the piezo-active area of a SAW device undergoes two shear x - and z -polarized oscillations and one y -polarized compression oscillation. These waves are reflected back to the layer-substrate interface, where they interfere with surface waves on the substrate. The outcome of the interference is given by the phase difference Φ_j between the two oscillations. The oscillation amplitude and velocity of a dissipating layer are given by a complex bulk K modulus and shear G modulus. The wave velocity on a coated SAW sensor is determined by the layer's mass load and the layer-substrate oscillation interference, and is given by [21]:

$$\frac{\Delta v}{v} = -\mathcal{J} \left(\sum_{j=x,y,z} \frac{c_j \beta_j M_j}{\omega} \tanh(i \beta_j h) \right) \quad (1)$$

where

$$\beta_j = \omega \left[\frac{\rho - \frac{E_j}{v_0^2}}{M_j} \right]^{\frac{1}{2}} \quad (2)$$

and

$$E_x = G, E_y \approx 0, E_z = \frac{4G(3K + G)}{3K + 4G} \quad (3)$$

In Equations (1)–(3), Im represents the imaginary part of the expression; index j gives the oscillation's polarization; 'i' is the imaginary unit; c_j is the SAW-film coupling parameter [21]; 'tanh' is the hyperbolic tangent function, where ' ω ' is the wave pulsation $2 \times \pi \times f$ with ' f ' the oscillation frequency of the SAW; ' h ' is the layer thickness; v_0 is surface acoustic wave velocity of un-coated SAW; ρ is the layer density; and $M_x = M_z = G$, $M_y = K$. In this model, phase difference $\Phi_j = R_e(i \beta_j h)$, where R_e represents the real part of the expression in parenthesis.

While the equality $\Delta f/f_0 = \Delta v/v_0$ holds only if the layer covers the entire area between the interdigital transducers [21], which is not our case, the resonant frequency variation

tracks that of the SAW propagation velocity modeled by Equation (1), thus allowing qualitative evaluation.

3. Results and Discussion

For statistically meaningful layer characterization, multiple interferential images were taken, and thickness-related parameters were subsequently obtained by computer image processing. Pixel statistics were made, in which the pixel color values were replaced by their thickness values according to the image color scale.

As in the example in Figure 1, the layer's profile is that of a convex meniscus, with in-layer dips and peaks of various amplitude.

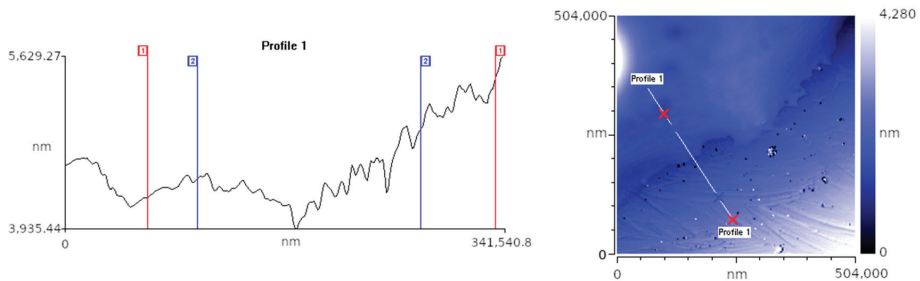


Figure 1. Type A layer interferential image. The image on the left represents the cross section profile along the line in the right image.

Due to the profilometer's viewing area being limited to $504 \mu\text{m} \times 504 \mu\text{m}$ area, only $504 \mu\text{m}$ from the layer's margin could be probed. However, interference profiling provided crucial information for identifying the origin of the observed F-A features.

Histograms for layers thickness and thickness gradients are presented in Figure 2.

In the above-thickness histograms, an abrupt ending toward small values, common to all layers, is visible. This is not surprising, given that the polymer solution droplet assumed a lenticular shape before solvent evaporation. As a result, the layer ends abruptly at the droplet's limit. The thickness lower limit seems to be the result of both solution concentration and droplet volume. In the case of the type B layer, the thickness lower limit is situated at $0.3 \mu\text{m}$, with a maximum around $0.65 \mu\text{m}$. Apart from the type B layer, all other layers present thicknesses no less than $1 \mu\text{m}$. Owing to its different deposition techniques, the type A layer is characterized by a thickness distribution with three local maxima, at 2, 3, and $6 \mu\text{m}$. The type C and D layers are virtually identical in thickness distribution, both starting at $1 \mu\text{m}$, with the most consistent values up to $3.5 \mu\text{m}$, and a smaller fraction that gradually diminishes up to $\sim 5 \mu\text{m}$. The type E layer presents a close resemblance with type C and D, except a $\sim 0.5 \mu\text{m}$ shift toward higher thicknesses. Additionally, it has two maxima centered around $2.7 \mu\text{m}$ and $3.5 \mu\text{m}$.

In the right column of Figure 2, histograms for layer thickness gradients are shown. Consistent thickness differences are present in all five types of layers. The thickness gradient distribution clearly splits the layers into two distinct groups: A and B with thickness gradient histogram maxima at less than $0.03 \mu\text{m}$; and C, D, and E with thickness gradient distribution maxima at more than $0.03 \mu\text{m}$. Additionally, types A and B present insignificant thickness variations above $0.15 \mu\text{m}$, while types C, D, and E exhibit important areas with gradients higher than $0.2 \mu\text{m}$. While most of it is due to the increasing thickness toward the layer's center, it is clear that in-layer thickness irregularities also account for a significant proportion of gradients, as can be seen in Figure 1.

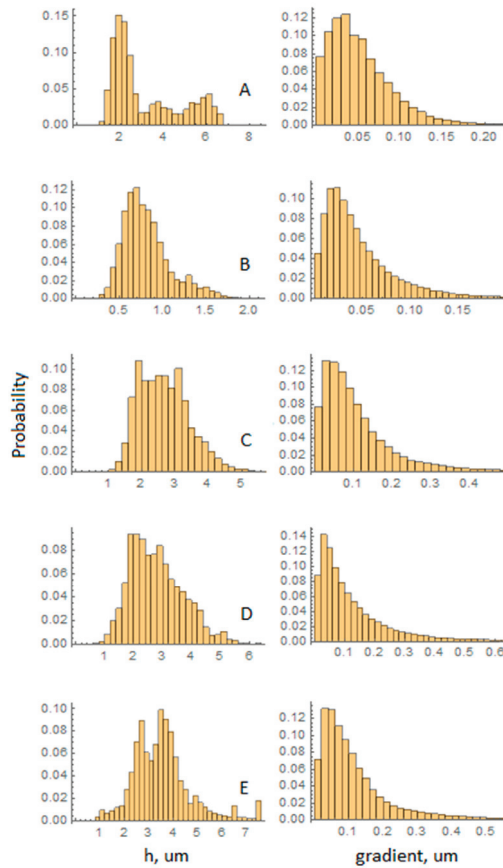


Figure 2. Thickness (left) and thickness gradients (right) histograms A to E letters indicate corresponding layer.

Figure 3 depicts an F-A characteristic corresponding to a B-type layer placed on the piezo active area center. The oscillator's resonant frequency shows a steady increase with signal amplitude attenuation, with a shallow sigmoid occurrence. While common to all layer types, sigmoid amplitude and position varies. The sigmoid positioning is the result of layer resonances, which are dependent on thickness.

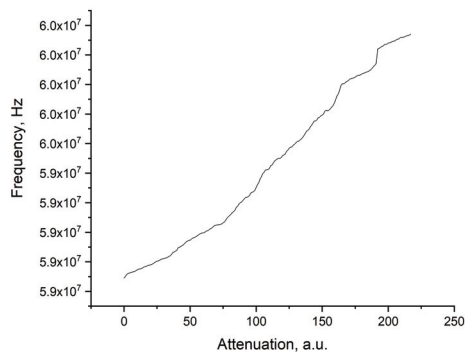


Figure 3. F-A characteristics of type B layer.

Multiple mini plateaus occur, and they could indicate spontaneous signal amplitude modulation. This is because zones of increased frequency stability were connected with carrier envelope resonances of a delay line oscillator [22].

The overall profile of the F-A is that of an increase in resonant frequency with attenuation, with full-range attenuation variation leading to an increase in the resonant frequency of ~1.7 MHz. This can be explained by the non-Newtonian behavior of the polymers, consisting of decreasing viscosity η and shear loss modulus G'' with shear rate [23,24]:

$$\eta = \frac{G''}{\omega} \sim \frac{1}{\dot{\gamma}(t)} \quad (4)$$

where $\dot{\gamma}(t)$ is the shear rate.

As the signal's amplitude decreases due to attenuation, the shear rate $\dot{\gamma}(t)$ decreases, also leading to increased viscosity and shear loss module G'' . Considering the dependence of β_z on G'' , this leads to the observed increase in resonant frequency.

Additionally, the layer's acoustic reaction contributes to shaping the F-A characteristic. As the layer's acoustic impedance changes due to the amplitude scan, so does the phase shift Φ_z of the reflected oscillations. While $\Phi_z = R_e(\beta_z h)$ [21], the loss modulus variation will cause a change in phase difference Φ_z due to relation (3), which mingles the pure complex loss modulus G'' in the real part of β_z . Thus, loss modulus variation nevertheless exerts a measurable influence on the Φ_z , consequently changing F-A characteristics.

Based on the above, the resonant frequency variation should be attributed to amplitude-driven changes in PEI shear loss module. This can be direct, due to β_z dependence on G'' , or indirect, via non-gravimetric effects, namely interference between the layer's shear oscillation and acoustic waves at the layer-SAW interface. The existence of acoustical effects is further confirmed by the presence of the sigmoid shapes at 110, 165, and 185 attenuations. Additionally, the orientation of the sigmoid is consistent with an increase in phase difference [21] implied by relations (3) and (4).

Type A layer was used to determine the correlation between the layer's position and F-A characteristic features. Isolating the positioning influence from other factors imposed the use of layers of identical F-A characteristics. As mentioned before, the deposition process reproducibility was tested by comparing the F-A characteristics of four type A layers. Figure 4 depicts the F-A characteristic of four type A layers placed in the center of the piezo-active area. Multiple common elements can be noted, the most significant being the sigmoidal region in the 100–120 a.u. attenuation range. The sigmoid consists of two regions of similar inclination separated by a mini plateau. Mini plateaus are present at both sigmoid ends. Many other plateaus occur with F-A characteristics, and most of them are similar in size and, aside from a small vertical shift, position.

The F-A characteristics in Figure 4 can be divided into three regions. First, the 0–65 a.u. attenuation is characterized by the A1 layer departure from the profile of the other three layers. From 65 to 160 a.u. attenuation, the F-A is characterized by a high degree of coincidence, both in features and position. In the last portion, the coincidence departure is most notable, both in features and position. This could indicate higher system instability at lower amplification, which exacerbated the differences in layer morphology. In the 0–100 a.u. attenuation range, the A2, A3, and A4 layers are virtually indistinguishable, their superposition being almost perfect. The A1 layer presents a vertical shift of ~30 kHz in the 0–25 and 45–65 a.u. attenuation ranges. Overall, Figure 1 proves that the deposition technique produces layers with similar F-A characteristics, supporting the validity of positioning impact investigation based on type A layers.

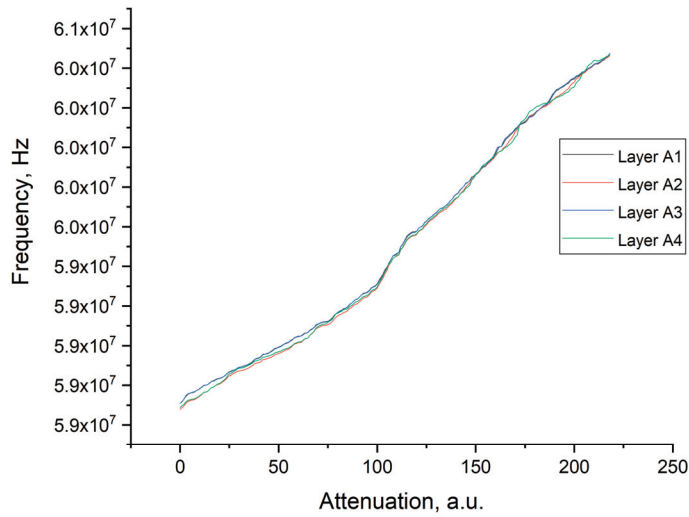


Figure 4. A-F dependence reproducibility for type A layers.

Figure 5 depicts the F-A characteristics corresponding to the four distances from input IDT at which a type A layer was placed. It is worth noting that, despite clear differences, the F-A characteristics still retain significant common elements, confirming the layers' quasi-identical morphology.

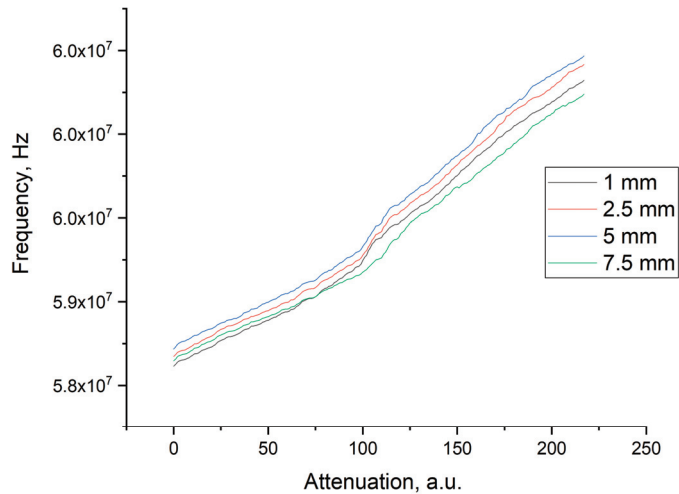


Figure 5. A-F for type A layer at different distances from input IDT.

While consistent efforts to preserve the experiment's integrity were made, there are some differences arising from occurrences out of our control, as is visible in both Figures 4 and 5. The presence of these random features will render our evaluation more difficult. To avoid drawing unfounded conclusions, we will limit ourselves to considering only major elements and parameters such as frequency variation rate, sigmoids, etc.

Aside from a relatively small vertical shift, the first three characteristics are parallel. Surprisingly, and in opposition to the resonant frequency increasing with attenuation, as the signal amplitude presumably increases with proximity to the IDT input, the characteristics are shifted vertically toward higher frequencies. At this point, the most certain information

we possess is the increase in frequency at lower signal amplitudes. This is supported by the values in Table 1, which indicate a lower frequency variation range for the 1, 2.5, and 7.5 mm layers. Thus, we are compelled to assume that this vertical shift is also caused by a decrease in the signal's amplitude, which implies the existence of a wave energy loss mechanism. In order to explain this behavior, the energy loss must be highly dependent on the layer's location. At this stage, it seems that the layer exerts a lens-like effect on SAW's wavefront, increasing its divergence. Also at this stage, it seems that the layer convex contour exerts a lens-like effect on SAW's wavefront, strongly increasing its convergence. After passing the focus point, the wavefront expands, diminishing its energy in a distance-dependent manner, thus lowering the signal's amplitude at receiving IDT. In turn, this causes a general decrease in signal amplitude all across the system, including a decrease in the oscillation amplitude of layers. This hypothesis can also explain the anomalous 7.5 mm F-A characteristics. The anomaly consists of the vertical shift toward lower frequencies (equivalent to layer oscillation with higher amplitudes) while being placed further from the emitting IDT (lower SAW amplitude due to attenuation). That means that the amplitude attenuation with distance is outweighed by the reduction in divergence-induced energy loss due to the shorter distance to the reception IDT, with the net result being an increase in SAW amplitude. Nevertheless, this does not explain the lower frequency variation range for the 7.5 mm layer and the diminished sigmoid amplitude.

Table 1. Relation between layer position and total frequency variation.

Distance from input IDT (mm)	1	2.5	5	7.5
Frequency variation (Hz)	1,705,990.5	1,743,723.5	1,750,369.5	1,590,019.1

In conclusion, displacement from the central position results in both a frequency shift and decrease in the frequency variation range due to reduced oscillation amplitude. It can be seen in Table 1 that the trend is well-defined, and it is the result of two distance-dependent wave energy loss mechanisms: layer-induced wavefront divergence, and wave attenuation with distance. Maximum frequency variation corresponds to the center position, 1.75 MHz, which corresponds to a minimum in wave energy loss.

F-A for all five layers can be seen in Figure 6. There are two elements that differentiate the F-A dependence: frequency increase rates and layer resonance position.

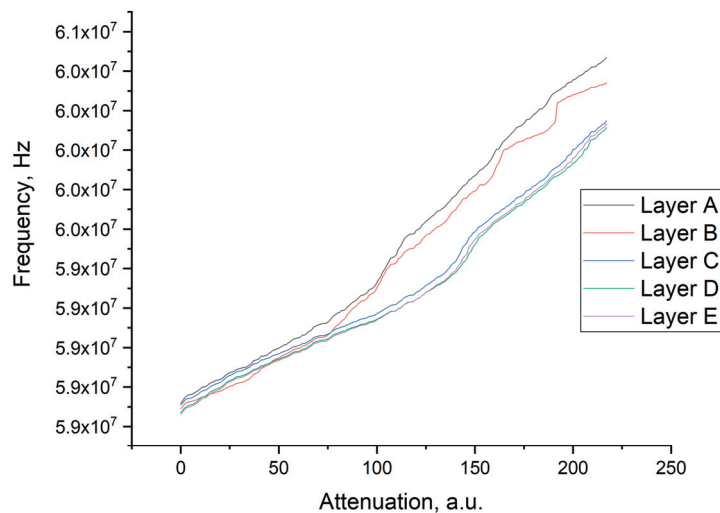


Figure 6. F-A for different layer types.

A good correlation between layer thicknesses and the position of layer resonances can be observed in Figure 6. In the cases of type C, D, and E layers, the layer resonance is situated virtually at the same frequency, which is consistent with their measured thickness.

Comparing the resonances of type B and any of the C, D, or E layers, the frequency spacing is not large enough to accommodate two different harmonics. Thus, it is probable that the lower resonance (in layers C, D, and E) is a second-order resonance. Given that the phase difference between reflected and substrate acoustic waves marks is $n \times \pi/2$ ($n = 1, 3, 5, \dots$) at layer resonance for a second-order layer resonance, the corresponding thickness ratio between the two layers should be 3. It should be noted that the sigmoids for the type C, D, and E layers are well defined and virtually identical. This suggests that the associated layer thickness causing this resonance is situated in an overlapping region of the type C, D, and E thickness histograms. The most probable thickness value is $\sim 2.4 \mu\text{m}$, leading to a very plausible (based on the thickness histogram in Figure 2 for the type B layer) thickness resonance value of $0.8 \mu\text{m}$ for sigmoids on type B F-A characteristic at 59.9 and 60.2 MHz. The $2.4 \mu\text{m}$ value for layer thickness is compatible with type A layer thickness distribution, and its departure from the thickness distribution maximum at $2 \mu\text{m}$ could account for the not-so-well-defined sigmoid shape.

Considering that the type B layer is the thinnest layer, the sigmoid around 59.3 MHz can be regarded as anomalous. This could be explained by the fact that the inner areas of the type B layer might be thick enough to accommodate the corresponding higher-order resonance. This explanation is thus far consistent with our experimental data interpretation.

Keeping in mind that the thickness distributions, as shown in Figure 2, are kept from probing the layers outer areas, and thickness distribution could be shifted toward higher values, this fits very well with type C, D, and E morphologies.

A striking quasi-total F-A similitude can be noted for the type C, D, and E layers, in spite of their different area and masses. While this correlates well with their similar morphology, as revealed by thickness-related parameters in Figure 2, it raises questions about the impact of size and mass on F-A characteristics. In this case, the two parameters have antagonistic effects on signal amplitude. On the one hand, a larger quantity of material will dissipate proportionally more energy. On the other hand, a higher radius layer will focus the wavefront less sharply, followed by a less pronounced divergence. Thus, the size impact for type C, D, and E layers may have been diminished due to the two effects canceling each other. As a result, the oscillation amplitude is almost identical for the C, D, and E layers, which is manifested in their F-A quasi-identical shape.

Aside from layer resonance position, the F-A differ in frequency variation rate, which is visibly smaller for C, D, and E-type layers. This signifies that the layers undergo overall lower amplitude oscillations, i.e., the amplitude variation range is smaller. While the amplitude attenuation caused by the added mass accounts for the different slopes, this does not explain the lower sigmoidal amplitudes for C, D, and E characteristics. Since the lower sigmoidal frequency excursion is the hallmark of higher loss parameter [21], this could signify the presence of an additional wave energy loss mechanism. Such a mechanism could be due to shear wave interference among adjacent areas of different thicknesses. Thus, the shear oscillatory energy returned by a portion of a given thickness will undergo attenuation by interference with neighboring portions of different thicknesses, proportional to the border length and phase difference between the two shear oscillations. This interpretation is very well supported by the thickness gradient histogram, as shown in Figure 2, which is highest for the C, D, and E layers.

4. Conclusions

PEI layers were deposited on SAW devices and used to determine the causal linkage between layer morphology and positioning, as well as F-A features. Identifying the morphology features responsible for F-A parameters was enabled by the deposition of PEI layers with reduced area. Thus, a causal connection between morphological features and F-A parameters was possible. As the reproduction of F-A for distinct layers of the same

type was possible, this enabled the effects of the layer's positioning on F-A characteristics to be isolated. The results strongly suggest the existence of a position-dependent energy loss mechanism, presumably due to layer-induced wavefront divergence.

F-A features specific to each type of layer were explained by resonances of layer areas of different thicknesses and by the presence of energy loss mechanisms via layer-induced wave divergence, and inter-layer shear wave interference, due to differences in layer thickness.

Notably, while a liability for gas detection, this method dependence on layer characteristics could constitute the basis for an advanced material testing method. Additionally, combined with the wireless capabilities of SAW devices, this could provide an accessible early warning method for an advanced process monitoring method.

The future development of this method is conditioned by a process to circumvent its dependence on layer morphology. Developing better layer deposition techniques with higher morphology reproducibility might not be always possible, nor economically viable. Thus, the development of computational methods could prove instrumental in implementing a detection/identification protocol based on this method. In this respect, the most important result is the dependence on the attenuation of the terms outside the hyperbolic tangent in Equation 2, which is nonexistent in the model used in [14], this is a very important consideration for future model development.

Author Contributions: Conceptualization, I.N.; methodology, I.N. and C.V.; software, I.N.; validation, I.N. and C.V.; investigation, I.N. and M.B.; resources, C.V.; writing—original draft preparation, I.N.; writing—review and editing, I.N., M.B. and C.V.; funding acquisition, C.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Romanian Ministry of Research, Innovation and Digitalization under Romanian National Program LAPLAS VII—contract no. 30N/2023.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Shu, L.; Peng, B.; Yang, Z.; Wang, R.; Deng, S.; Liu, X. High-Temperature SAW Wireless Strain Sensor with Langasite. *Sensors* **2015**, *15*, 28531–28542. [CrossRef] [PubMed]
- Pan, Y.; Molin, Q.; Guo, T.; Zhang, L.; Cao, B.; Yang, J.; Wang, W.; Xue, X. Wireless passive surface acoustic wave (SAW) technology in gas sensing. *Sens. Rev.* **2021**, *41*, 135–143. [CrossRef]
- Ruan, Y.; Chen, Y.; Wu, Y.; Shi, M.; Du, Y.; Song, Z.; Chen, Y.; Dong, H.; Zhang, C.; Teng, J. Novel SAW Temperature Sensor with Pt/Ti/AlN/Mo/AlN/Si Structure for High Temperature Application. *Chemosensors* **2023**, *11*, 225. [CrossRef]
- Zhang, J.; Jin, H.; Dong, S.; Ding, R.; Chen, J.; Xuan, W.; Gao, F.; Luo, J. Surface Acoustic Wave Strain Sensor with Ultra-Thin Langasite. *IEEE Sens. J.* **2022**, *22*, 11509–11516. [CrossRef]
- Pei, G.; Ma, B.; Zhang, Z.; Wang, C.; Deng, J.; Luo, J. High-temperature Pt-Al₂O₃ composite nano-thick interdigital electrodes for surface acoustic wave sensors. *Mater. Chem. Phys.* **2022**, *291*, 126697. [CrossRef]
- Yan, X.; Tan, Q.; Li, X.; Xue, T.; Li, M. Test and Analysis of SAW High Temperature Strain Sensor Based on Langasite. *IEEE Sens. J.* **2022**, *22*, 12622–12628. [CrossRef]
- Zhou, X.; Tan, Q.; Liang, X.; Lin, B.; Guo, T.; Gan, Y. Novel Multilayer SAW Temperature Sensor for Ultra-High Temperature Environments. *Micromachines* **2021**, *12*, 643. [CrossRef] [PubMed]
- Nair, M.P.; Teo, A.J.T.; Li, K.H.H. Acoustic Biosensors and Microfluidic Devices in the Decennium: Principles and Applications. *Micromachines* **2022**, *13*, 24. [CrossRef] [PubMed]
- Feng, B.; Jin, H.; Fang, Z.; Yu, Z.; Dong, S.; Luo, J. Flexible Strain Sensor Based on Ultra-Thin Quartz Plate. *IEEE Sens. J.* **2021**, *21*, 18571–18577. [CrossRef]
- Ricco, A.J.; Martin, S.J. Multiple-frequency SAW devices for chemical sensing and materials characterization. *Sens. Actuators B Chem.* **1993**, *10*, 123–131. [CrossRef]
- Cular, S.; Bhethanabotla, V.R.; Branch, D.W. Hexagonal surface acoustic wave devices for enhanced sensing and materials characterization. *Proc. IEEE Ultrason. Symp.* **2005**, *4*, 2309–2312.
- Yadava, R.D.S.; Kshetrimayum, R.; Khaneja, M. Multifrequency characterization of viscoelastic polymers and vapor sensing based on SAW oscillators. *Ultrasonics* **2009**, *49*, 638–645. [CrossRef] [PubMed]
- Seidel, W.; Hesjedal, T. Multimode and multifrequency gigahertz surface acoustic wave sensors. *Appl. Phys. Lett.* **2004**, *84*, 1407–1409. [CrossRef]

14. Nicolae, I.; Viespe, C.; Miu, D.; Marcu, A. Analyte discrimination by SAW sensor variable loop amplification probing. *Sens. Actuators B Chem.* **2022**, *358*, 131480. [CrossRef]
15. Wang, H.; Zhang, L.; Zhou, Z.; Lou, L. Temperature Performance Study of SAW Sensors Based on AlN and AlScN. *Micromachines* **2023**, *14*, 1065. [CrossRef] [PubMed]
16. Zhang, J.; Jin, H.; Chen, J.; Xuan, W.; Ding, R.; Dong, S.; Luo, J. High temperature effects on surface acoustic wave strain sensor. *Sens. Actuators A Phys.* **2022**, *338*, 113464. [CrossRef]
17. Kumagai, H.; Furusawa, H. Real-Time Monitoring of a Nucleic Acid Amplification Reaction Using a Mass Sensor Based on a Quartz-Crystal Microbalance. *Biosensors* **2024**, *14*, 155. [CrossRef] [PubMed]
18. Li, Y.; Huang, X.; Chen, Q.; Yao, Y.; Pan, W. Nanochitin/Mxene Composite Coated on Quartz Crystal Microbalance for Humidity Sensing. *Nanomaterials* **2023**, *13*, 3135. [CrossRef] [PubMed]
19. Patel, R.; Adhikari, M.S.; Bansal, D.; Johar, A. Hybrid Film Bulk Acoustic Resonator Fabricated Using a Single Bragg Reflector. *J. Electron. Mater.* **2024**, *53*, 2910–2920. [CrossRef]
20. Patel, R.; Adhikari, M.S.; Tripathi, S.K.; Sahu, S. Design, Optimization and Performance Assessment of Single Port Film Bulk Acoustic Resonator through Finite Element Simulation. *Sensors* **2023**, *23*, 8920. [CrossRef]
21. Martin, S.J.; Frye, G.C.; Senturia, S.D. Dynamics and Response of Polymer-Coated Surface Acoustic Wave Devices: Effect of Viscoelastic Properties and Film Resonance. *Anal. Chem.* **1994**, *66*, 2201–2219. [CrossRef]
22. Planat, M.; Lardet-Vieudrin, F.; Martin, G.; Dos Santos, S.; Marianneau, G. Frequency stability enhancement from carrier-envelope resonance in a surface acoustic wave delay line oscillator. *J. Appl. Phys.* **1996**, *80*, 2509. [CrossRef]
23. Shaw, M.T.; MacKnight, W.J. *Introduction to Polymer Viscoelasticity*, 3rd ed.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2005; ISBN 13: 978-0-471-74045-2.
24. Ferry, J.D. *Viscoelastic Properties of Polymers*, 3rd ed.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 1980; p. 43, ISBN 9780471048947.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Image Analysis Techniques Applied in the Drilling of a Carbon Fibre Reinforced Polymer and Aluminium Multi-Material to Assess the Delamination Damage

Rúben D. F. Sousa Costa^{1,2}, Marta L. S. Barbosa², Filipe G. A. Silva^{1,3,4}, Tiago E. F. Silva¹, Abílio M. P. de Jesus^{1,2}, Francisco J. G. Silva^{1,5}, Luís M. P. Durão^{1,5} and João Manuel R. S. Tavares^{1,2,*}

¹ Laboratório Associado em Energia, Transportes e Aeronáutica, Rua Dr. Roberto Frias 400, 4200-465 Porto, Portugal; rdcosta@inegi.up.pt (R.D.F.S.C.); fgsilva@ispgaya.pt (F.G.A.S.); tesilva@inegi.up.pt (T.E.F.S.); ajesus@fe.up.pt (A.M.P.d.J.); fgs@isep.ipp.pt (F.J.G.S.); lmd@isep.ipp.pt (L.M.P.D.)

² Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal; up201808522@edu.fe.up.pt

³ Instituto Politécnico Gaya (ISPGaya), Av. dos Descobrimentos 333, 4400-103 Vila Nova de Gaia, Portugal

⁴ proMethuus, Instituto Politécnico de Viana do Castelo, Rua Escola Industrial e Comercial Nun'Álvares, 4900-347 Viana do Castelo, Portugal

⁵ CIDEM, ISEP, Politécnico do Porto, Rua Dr. Bernardino de Almeida 431, 4200-072 Porto, Portugal

* Correspondence: tavares@fe.up.pt

Abstract: Due to the high abrasiveness and anisotropic nature of composites, along with the need to machine different materials at the same time, drilling multi-materials is a difficult task, and usually results in material damage, such as uncut fibres and delamination, hindering hole functionality and reliability. Image processing and analysis algorithms can be developed to effectively assess such damage, allowing for the calculation of delamination factors essential to the quality control of hole inspection in composite materials. In this study, a digital image processing and analysis algorithm was developed in Python to perform the delamination evaluation of drilled holes on a carbon fibre reinforced polymer (CFRP) and aluminium (Al) multi-material. This algorithm was designed to overcome several limitations often found in other algorithms developed with similar purposes, which frequently lead to user mistakes and incorrect results. The new algorithm is easy to use and, without requiring manual pre-editing of the input images, is fully automatic, provides more complete and reliable results (such as the delamination factor), and is a free-of-charge software. For example, the delamination factors of two drilled holes were calculated using the new algorithm and one previously developed in Matlab. Using the previous Matlab algorithm, the delamination factors of the two holes were 1.380 and 2.563, respectively, and using the new Python algorithm, the results were equal to 3.957 and 3.383, respectively. The Python results were more trustworthy, as the first hole had a higher delamination area, so its factor should be higher than that of the second one.

Keywords: image processing; image analysis; Python; delamination evaluation; multi-material

Citation: Sousa Costa, R.D.F.; Barbosa, M.L.S.; Silva, F.G.A.; Silva, T.E.F.; de Jesus, A.M.P.; Silva, F.J.G.; Durão, L.M.P.; Tavares, J.M.R.S. Image Analysis Techniques Applied in the Drilling of a Carbon Fibre Reinforced Polymer and Aluminium Multi-Material to Assess the Delamination Damage. *Processes* **2024**, *12*, 1258. <https://doi.org/10.3390/pr12061258>

Academic Editors: Jiaqiang E, Sheng Du, Xiongbo Wan, Zixin Huang and Li Jin

Received: 11 April 2024

Revised: 5 June 2024

Accepted: 11 June 2024

Published: 19 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Image processing and analysis techniques are based on mathematical functions to achieve the intended outcomes for the desired task [1]. Nowadays, many of the advanced engineering developing software tools, such as Matlab (MathWorks, Natick, MA, USA), version 2019 B, or Python (Python Software Foundation, Wilmington, DE, USA), version 5.1.5., already include many of these functions, and the algorithms developed almost only depend on the correct combination of these available functions. Four primary focus areas are usually considered in image processing and analysis: image improvement or enhancement, image restoration, image compression, and image analysis. These focus areas are typically associated with artificial intelligence and are usually designated as computer vision [2]. The first and last areas are the most important in multi-material

drilling processes. In the image improvement or enhancement process, the characteristics of the input image are improved, for example, by accentuating the contrast and reducing the noise. On the other hand, image analysis aims to interpret a given image by measuring or classifying its properties. These techniques can be applied in various fields, such as engineering, science, and healthcare [3].

Usually, the first step in image processing is to convert the input colour image into a grayscale image to simplify the following steps. Afterward, a noise-smoothing technique may be applied to remove unwanted elements in the image surrounding the object(s) of interest, which can be achieved using, for example, a bilateral filter [4] or low-pass filter [5]. Another helpful technique to improve the input image's quality is contrast stretching [6], which increases the contrast between the darkest and brightest regions in the image while preserving the relative differences among the intermediate values. The intensity of the pixels of the object of interest can also be enhanced using sharpening or histogram equalization techniques [7] by modifying the image's intensity distribution based on its histogram or using a high-pass filter.

Another essential function in image processing is the erosion operation, commonly used to erode the boundaries of the object of interest in an image. It can be used, for example, to remove noise, to isolate individual elements, or to reduce the object's size. In the same area, the dilation operation is often used to expand the object's boundaries in a binary image. It can also be used to fill gaps, to connect broken objects, or to increase the object's size. Both have an "iterations" parameter, specifying how often the techniques are applied to the image and how they can be increased to intensify their result [8].

In many applications, after image pre-processing, it becomes necessary to apply the thresholding technique [9], which is the simplest method for performing image segmentation. This technique turns the image into a binary form so that a specific region of interest (ROI) of the image can be emphasised and its subregions can be uniformised. Some other important techniques usually used for this task include the Hough transform technique [10], to detect geometric shapes in images, such as circles; the snake algorithm [11], to find the contour that best approximates the border of an object; and the level set model for image segmentation [12], which implements a geometric active segmentation contour model. From a boundary detection perspective, the Canny, Prewitt, Roberts, and Sobel edge detectors [13,14] are some of the most used operators that perform contour detection in the image under study, using an approximation of its intensity gradient through finite differences. For image analysis, the region-growing algorithm [15] is commonly used to segment image regions by growing a region from a seed point, and then measurements can be obtained from a particular region, allowing for further analyses.

Due to the many possibilities that the usual image processing and analysis techniques offer in solving real-world problems, different applications exist to put them into practice. These range from the automatic evaluation of metallographic microstructure phases [16] and surface characterisations [17] to damage detection in parts resulting from mechanical tests [18] or manufacturing processes in engineering. In the scope of drilling composites and multi-materials, mainly by a combination of composite and metal, several damages usually arise, such as burrs and fibre delamination. This happens mainly due to composite materials' anisotropic nature, making it harder to perform a clean cut. This property, along with the high abrasiveness of the composite fibres and the drill's need to cut through different materials simultaneously, causes severe tool wear, leading to an imperfect surface on the hole's border [19]. This kind of damage may sometimes go unnoticed in high batch production, which could have catastrophic consequences, especially for this process' most frequent applications, the aeronautical and automotive industries [20].

To quantify the hole delamination, some criteria have been proposed in the literature relating the key properties of the damaged regions and those of the original hole. These can comprise just the diameter (F_d), introduced by Chen et al. [21] in 1997, or just the area (F_a), presented for the first time by Mohan et al. [22] in 2007. Later, in the same year, Davim et al. [23] suggested an improved version of the delamination measurement, the

adjusted delamination factor (F_{da}), which is considered a better approximation of reality by combining the diameter and the area at the same time. This can be expressed as:

$$F_d = \frac{D_{max}}{D_{nom}}, \quad (1)$$

$$F_a = \frac{A_d}{A_{nom}}, \quad (2)$$

$$F_{da} = F_d + \frac{A_d}{(A_{max} - A_{nom})} (F_d^2 - F_d), \quad (3)$$

where D_{nom} and A_{nom} correspond to the nominal hole's diameter and area (in mm and mm², respectively), D_{max} and A_{max} to the maximum delamination diameter and area considering the maximum diameter (in mm and mm², respectively), and A_d to the actual delamination area (in mm²) (Figure 1). If the damage around the hole is circular and has a low amplitude, it is not completely considered in the F_d approach, and the same happens for the F_a if there is just a thin crack or a damaged peak but without a great area. That is why the adjusted delamination factor, F_{da} , provides a more complete and more reliable result.

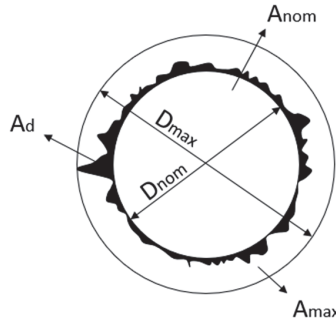


Figure 1. Diameters and areas used to calculate the delamination factors F_d , F_a and F_{da} .

Several studies assess delamination damage in composite materials using image processing and analysis techniques to identify if the non-conformities are within the acceptable limits imposed by the industry quality patterns. Durão et al. [24] used damage assessment image-based methods through computational processing of data extracted from radiographic images and compared them with mechanical test results, bearing and delamination assessment tests, and analytical models, obtaining valuable information regarding the delamination extension and mechanical strength of the drilled parts. In more recent studies, Durão et al. [25] and Devesa et al. [26] used similar methods to calculate the fractal dimension of the damaged region's boundary. Geier et al. [27] predicted drilling-induced burr occurrence in CFRP composites using image processing and analysis by studying the fibre orientation in CFRP sections and their local properties, achieving an accuracy of 64–97% compared to real drilling experiments. Alejandro et al. [28] also autonomously inspected drilled holes in CFRP panels using image processing and analysis techniques to determine the geometric errors and delamination factors for several holes and compared them with experimental results. Despite most delamination assessment approaches through image processing and analysis techniques being applied in just plain composites, machined multi-material processing has been gaining relevance in recent years, even if only a few studies exist on this topic. Isbilir and Ghassemieh [29] studied the hole's quality and respective delamination factors of drilled CFRP/titanium stacks using image processing and analysis techniques, and Marques et al. [30] successfully determined the delamination factors of holes in drilled composite and aluminium (CFRP/Al) multi-materials via computational analysis of X-ray images. Several non-destructive tests could acquire images of the hole

and surrounding areas for the damage extension analysis and evaluation, such as radiography, ultrasound, acoustic emission, C-scan, and computer tomography (CT) [24]. Despite this, the delamination damage was typically assessed using X-ray radiography, as it is the cheapest and most easily accessible imaging modality.

The problem is that the common software used to assess drilled multi-material hole damage had some gaps. For example, in many of them, the original hole images required manual pre-editing to obtain consistent results, which were sometimes incorrect, nonetheless. They also required several user inputs of key functions that, if wrongly selected, led to problems in their functioning. Additionally, the main limitation of solutions developed in Matlab is that it is a paid software, whereas those developed in Python have no such limitation and can be improved easily with new functions associated with the Python libraries that appear every day. Therefore, the objective of this work consisted in the development of an algorithm in Python with an intuitive interface, i.e., an enhanced graphical user interface (GUI), which uses image processing and analysis techniques to obtain comprehensive information about the quality of drilled holes in CFRP/Al multi-material, namely the extent of delamination, from X-ray images. The algorithm was then compared to a previous algorithm developed in Matlab to confirm its superior performance, mainly in terms of image improvement and accuracy of delamination factors. For this, it was necessary to perform the division of the problem under study into two main steps:

- Image processing, where the input image was firstly improved by enhancing its most relevant features and via contrast refinement and noise smoothing to retain only the regions of interest with enhanced quality.
- Image segmentation and analysis, where the region of interest was selected through proper thresholding and finally assessed, resulting in the evaluation of the hole's damage based on delamination factors derived from its perimeter and area.

2. Materials and Methods

2.1. Image Acquisition

This work was based on a drilling campaign using a DMG Mori DMU 60 eVo (DMG Mori, Bielefeld, Germany) CNC machine on a three-layered multi-material (CFRP/Al/CFRP) with a 4.5 mm thickness, using a 6 mm diameter conventional twist drill with a point angle of 120° and helix angle of 30°. The multi-material was evenly distributed, with each layer having a thickness of 1.5 mm. The metal used was Al2024-T3, and the composite was a cross-ply (0/90°) with a thermosetting matrix (epoxy) composed of 7 plies.

For the acquisition of the intended images from the drilled holes, a radiography system associated with a Kodak RVG 5100 (Eastman Kodak Company, Rochester, NY, USA) digital sensor was used, Figure 2a, since the X-ray could effectively pass through both the CFRP and aluminium, revealing the damage present in them, as has been considered in several studies in the literature. Two foam supports were used on both sides of the sensor for the stacks' stability and to ensure that all the images were obtained with the same distances between the sensor and the plate and the sensor and the emissary. Firstly, the samples were immersed in a contrasting liquid, Analar Normapur Diiodomethane, for 30 min. Thus, when the X-ray system was employed, it revealed the zones where the liquid penetrated the delamination areas, which were clearly distinguished from the unaffected material because the liquid was opaque to radiography. As already mentioned, a Satelec X-mind radiography system was used, Figure 2b, with a 0.16 s exposition time, alongside the image acquisition software in Matlab and Python languages.

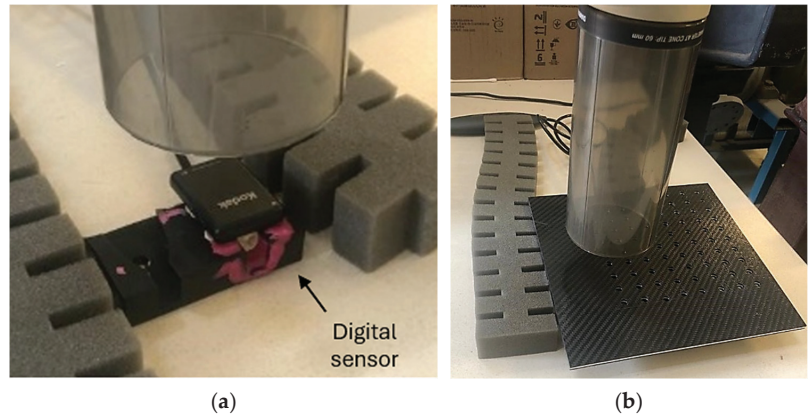


Figure 2. (a) The Kodak RVG 5100 digital sensor and (b) a sample of the multi-material under X-ray analysis.

2.2. Algorithm Pipeline

Python was used to develop the image processing and analysis algorithm, mainly to identify the ROIs, i.e., the hole and damaged areas, and assess them. Its main steps consisted of image processing to determine the damaged and hole regions, image segmentation to determine the ROIs, and image analysis to calculate the delamination factors (Equations (1)–(3)). Finally, the results were shown to the user. The scheme in Figure 3 depicts the tasks performed in the novel algorithm. The development was based on the defined objectives and a literature search for Python functions that could be used to address the tasks involved. Combining different functions to obtain the desired outcome was possible with a trial-and-error method. The libraries in Python used to develop the new algorithm are shown in numpy v1.20.3, OpenCV v4.5.2, matplotlib v3.4.3 and scipy v1.7.3.

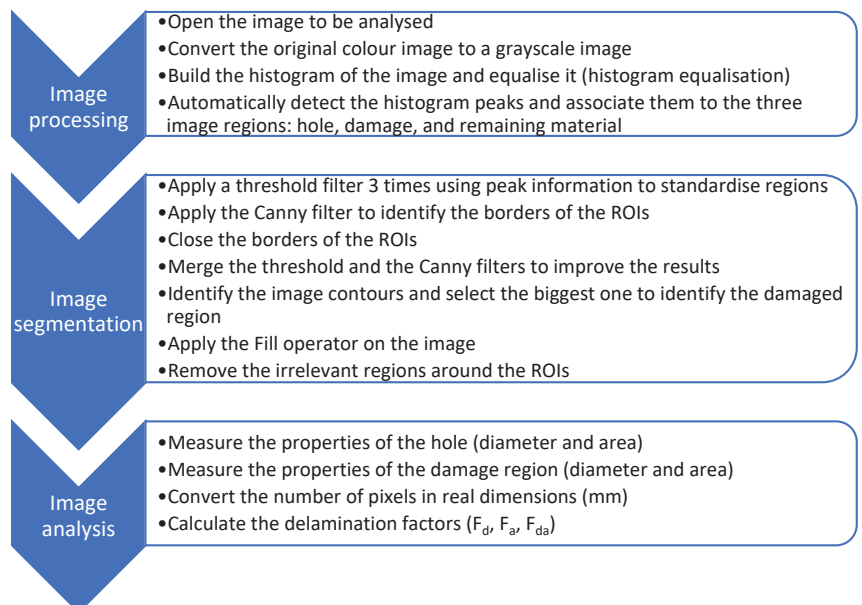


Figure 3. Pipeline of the novel image processing and analysis algorithm developed in Python.

3. Algorithm Comparison

3.1. Previous Algorithm

With the previous image processing and analysis algorithm developed in Matlab, when a new image is given as the algorithm's input, the first step is to manually choose the ROI's borders using a two-step selection grid, Figure 4, so that the algorithm may restrict the analysis to the specified region. From this selection and based on a conversion factor from pixels to millimetres (mm) that the user needs to update in the algorithm code before running it, the hole's radius is calculated first in pixels and then in mm.

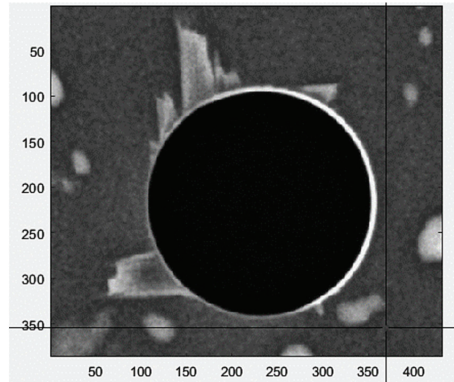


Figure 4. Grid selection of the ROI's borders of the hole image under study in the previous Matlab algorithm.

Afterward, the algorithm suggests a threshold that usually never corresponds to the intended one. It is either too high, as in Figure 5a, meaning that almost no delamination is detected and the hole circle is not entirely detected, or too low, as in Figure 5b, with unwanted noise surrounding the hole's region and disrupting the real delamination analysis. Therefore, the algorithm asks the user if the threshold is acceptable, and if the user declines it, the algorithm then asks the user to choose another value in an iterative method until a satisfying result is obtained, as in Figure 5c. Thus, an adequate threshold needs to be selected by the user based on visual analysis, otherwise, the delamination factors are miscalculated.

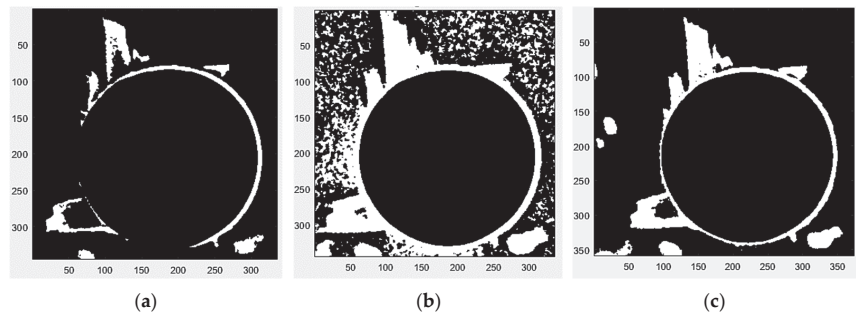


Figure 5. The effect of the selected ROIs threshold value in the previous Matlab algorithm: (a) too high, (b) too low, and (c) adequate threshold value.

Then, if the user accepts the threshold, it is necessary to zoom in on the image and press a random keyboard key. The problem is that the user can only determine that a pixel on the hole's border needs to be selected after pressing the key, and the user cannot zoom the image out at this step. Hence, an error occurs if a bad selection is made, and the process

must be reinitiated to continue. A similar process starts for the user to select a pixel on the delamination contour. Figure 6a and Figure 6b represent the hole and delamination contour detection, respectively, on the top left corner region of the image in Figure 5c.

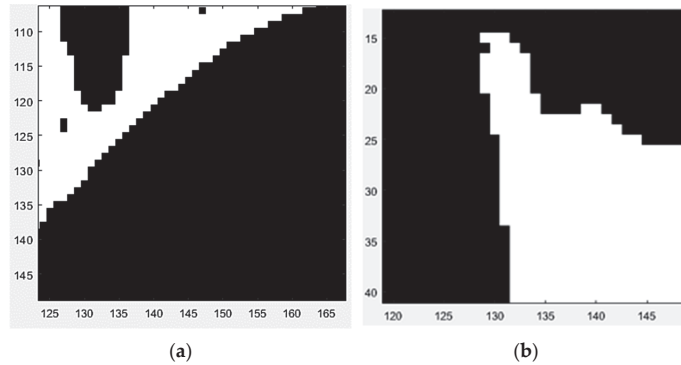


Figure 6. Results obtained by the previous Matlab algorithm: the (a) hole and (b) delamination contour detection on the top left corner region of the image in Figure 5c.

At the final step of running the previous algorithm, it returns three images, one highlighting the hole region, Figure 7a, another representing the hole and damaged regions merged, Figure 7b, and finally, the binary image to be assessed, Figure 7c. Alongside these visual representations, the delamination factor values are presented to quantify the damage from the drill effect on the hole's periphery.

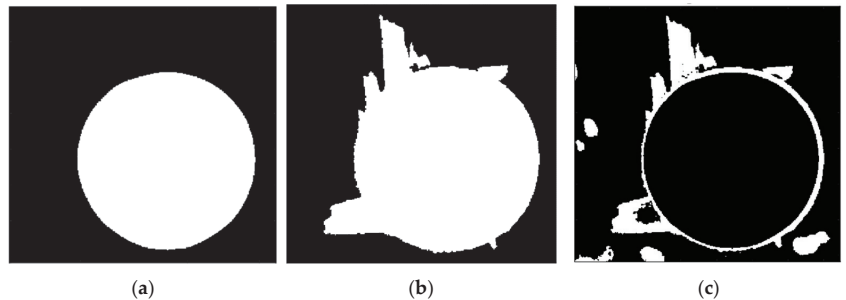


Figure 7. Results obtained by the previous Matlab algorithm: (a) the hole region, (b) the hole + damaged regions, and (c) the final binary image to be assessed.

3.2. Novel Algorithm

Considering the operational concerns mentioned earlier, a new algorithm needed to be developed, to enable easier operation and better result accuracy. Thus, a new set of steps was designed as follows:

- Image processing to determine the damaged and hole regions, i.e., image segmentation or determination of the ROIs.
- Calculating the delamination parameters: F_d , F_a , F_{da} .
- Showing the results to the user.

Figure 8 shows an example of a grayscale image imported to the new Python algorithm.

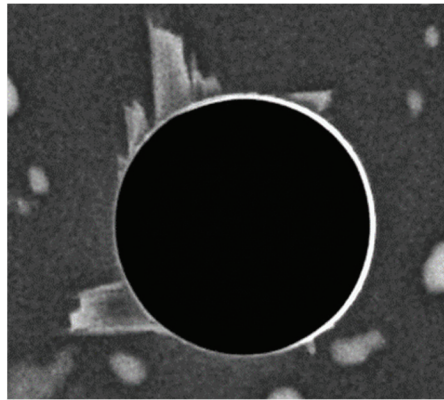


Figure 8. An example of a hole image imported to the new Python algorithm.

3.2.1. Image Segmentation—Determination of the ROI

The best way to analyse an image is to look at its histogram, like the one in Figure 9a. The histogram shows the number of pixels along each bright intensity. The original image essentially contained three groups of pixels, black, grey, and white, which could be analysed in the peaks of the respective histogram (Figure 9a) although the last one, with nearly 250 bins, was not very pronounced. To automatically determine the threshold indicated for each image, the value of each peak was calculated, Figure 9b, with only a small range given to define the values of that peak. The purpose of these two graphics was to show the approach used in the algorithm to identify the hole and damaged regions.

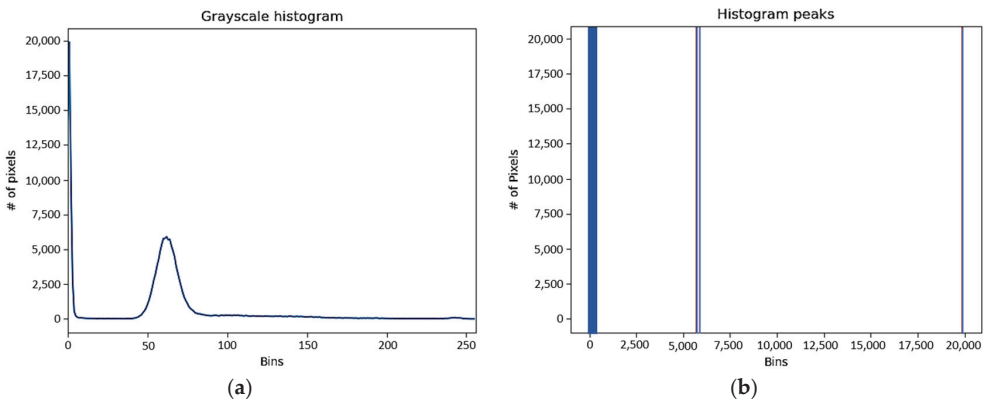


Figure 9. (a) The grayscale histogram of the image in Figure 8 and (b) the respective histogram peaks.

The images in Figure 10 were obtained from the histogram data. Figure 10a,b resulted from the values of the peaks identified in Figure 9, showing the hole and damaged regions, respectively, by simply applying a threshold to those values given the respective interval. Figure 10c combines Figure 10a,b, showing the hole and damaged regions.

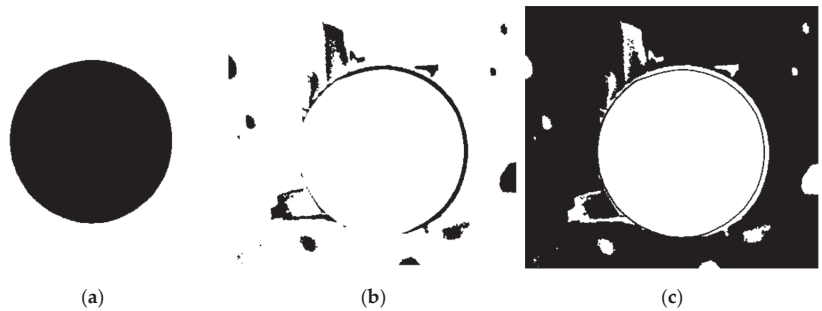


Figure 10. The output obtained by the new Python algorithm from the image in Figure 8: (a) the hole region, (b) the damaged region, and (c) the hole + damaged regions.

In the case of Figure 11, two different filters were applied to the original image. The Canny filter is commonly used to find objects' edges in images. This filter was tested, and examples of its results are shown in Figure 11a. Since this filter only shows the hole's edges, and there can be cases where the edges are not all closed, and there are gaps in the objects' borders, basic morphological filters were tested to improve the results. The combination of dilation and erosion operators after applying the Canny filter was the method that improved the results the most, Figure 11b.

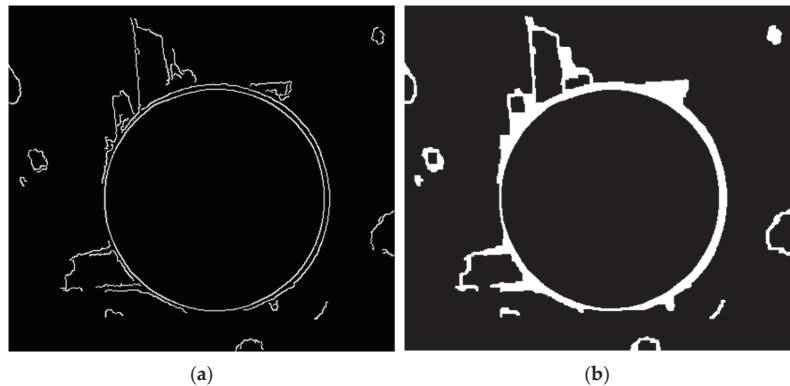


Figure 11. Images after applying the (a) Canny edges detector and (b) erosion filter to the image in Figure 10c.

Figure 12a depicts the result of combining the previously applied filters with the threshold technique. Figure 12b shows the damaged and hole regions merged together, derived from Figure 12a by closing its borders, filling in the insides, and selecting only the region with the biggest area. At this step, the unwanted surrounding noise disappeared due to the combined action of the previously used filters. Figure 12c was obtained by replacing the central area (white pixels) of Figure 12b with the value of the hole region (black pixels), showing only the damaged region, so a proper comparison with the original image could be possible.

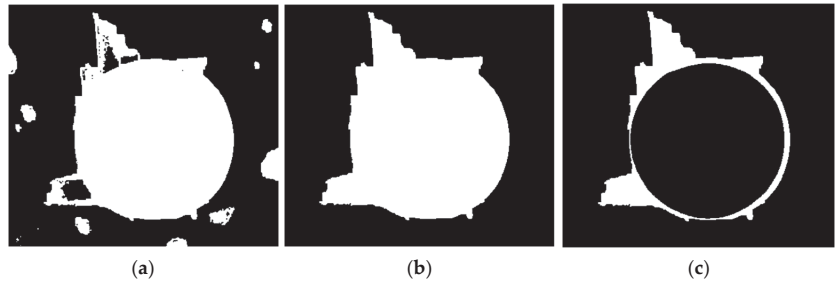


Figure 12. The output obtained by the new Python algorithm from the image in Figure 8: after the (a) filter combination, (b) filled damaged + hole region, and (c) damaged region.

3.2.2. Calculating the Delamination

This step calculates the delamination parameters by automatically counting the pixels corresponding to each region. Then, by knowing the base radius of the drilled hole, the conversion factor is calculated so that the delamination factors can be determined in mm. The X-ray images processed by the Python algorithm permitted a specific delamination evaluation to be quantified through delamination factor criteria (Equations (1)–(3)).

To determine the diameters for calculating the delamination factors, first, the contours of the hole and the damaged regions were identified by thin internal lines. Secondly, the smallest circle that included those contours was calculated on the thicker lines, corresponding to the hole's inner diameter, D_{nom} , Figure 13a, and the outer diameter of the circle better adjusted to the complete delaminated region, D_{max} , Figure 13b. In the first image, a slight difference between the two lines in the top left-hand corner is seen, which means that the hole was not perfectly circular, possibly caused by more loads being applied on that side during the drilling process.

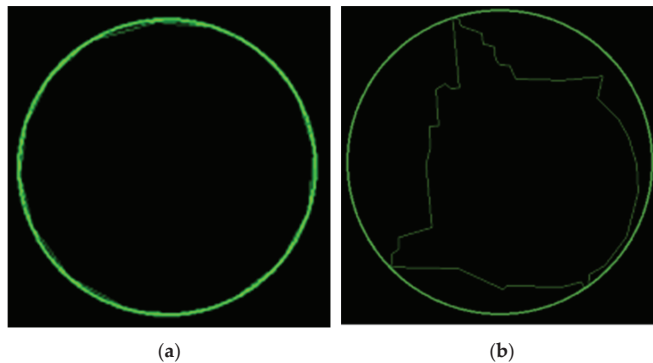


Figure 13. The identification process of the (a) inner diameter, D_{nom} , and (b) delaminated outer diameter, D_{max} , by the new algorithm.

4. Novel Algorithm's Results

In order not to simply write the new algorithm's results in the user console, as the user may not be proficient in programming with Python or may use different code compilers, the results were presented under the image form for clearer output. For comparison purposes, the two algorithms were put side by side to highlight their differences, and the improvements brought by the new one in the analysis of two holes performed with different sets of parameters with distinct delamination levels.

Hole 1 was drilled with a feed of 0.3 mm/rev. and a cutting speed of 150 m/min. Comparing its original image, Figure 14a, with the corresponding binary image obtained

by the previously developed Matlab algorithm, Figure 14b, and the image obtained by the new Python algorithm, Figure 14c, a clear difference was observed. Particularly noticeable was the identification of the delaminated zones around the drilled hole, which in this new solution was highlighted more, with the delamination regions fully filled, contrary to the result obtained by the previous algorithm. Moreover, the presence of contrasting liquid drops near the hole, which could not be completely removed despite thoroughly cleaning the sample, leaving only the delaminated zones impregnated with it, can sometimes affect the damage assessment. The new algorithm successfully considered these as image noise and removed them so there was no interference with the delaminated region around the hole. This ensured a higher accuracy of the results, making the evaluation more precise and effective.

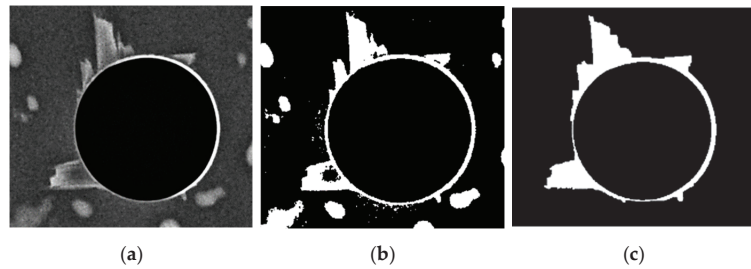


Figure 14. (a) The original image of Hole 1, (b) the output from the Matlab algorithm of the Hole 1 image, and (c) the corresponding output from the Python algorithm.

An example of a less damaged hole, namely of the Hole 2 image, drilled with a lower feed, 0.15 mm/rev., and the same cutting speed, 150 m/min., is depicted in Figure 15a. The problem with the previous algorithm, Figure 15b, was that if the selected threshold was too high, the hole's boundaries were not properly identified, but if it was too low, as shown in the image, there was too much noise around the hole. As clearly seen, the new algorithm, Figure 15c, was designed to identify better the less prominent delamination zones around the hole, i.e., the ones where the contrast liquid did not penetrate as effectively or were simply less deep and not so easily distinguishable with the naked eye. In the new algorithm, combining the Canny, dilation and erosion operators, and threshold filter removed the noise from the input image without compromising the hole boundaries, thus overcoming this problem.

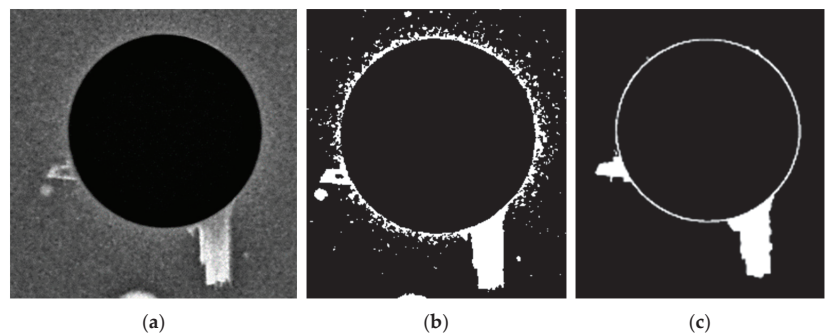


Figure 15. (a) The original image of Hole 2, (b) the output from the Matlab algorithm of the Hole 2 image, and (c) the corresponding output from the Python algorithm.

With the application of the new algorithm, obtaining more accurately adjusted delamination factors from images of holes with great delamination was possible. It also resulted in logical values regarding the tool used, and the chosen machining parameters, namely

the feed and cutting speed. Table 1 indicates the delamination factors calculated using Equations (1)–(3) with the data obtained from the images of Holes 1 and 2 by the previous and the new algorithms, along with the respective drilling parameters. Noticeably, the increase in the delamination factor obtained by the new algorithm was directly related to the feed increment.

Table 1. The comparison of the values of the studied delamination damage factors obtained from the images of Holes 1 and 2 using the two algorithms.

Hole Image	Feed (mm/rev)	Cutting Speed (m/min)	Algorithm	F_d	F_a	F_{da}
1	0.30	150	Previous	2.014	−0.239	1.380
			New	1.496	4.106	3.957
2	0.15	150	Previous	2.332	0.121	2.563
			New	1.366	3.492	3.383

The previous algorithm had problems with the image processing step and with calculating the delamination factors. Although the values of the adjusted delamination factor (F_{da}) seemed more reliable in the previous algorithm, as they were lower, it quickly became apparent that they were incorrect. The value in the Hole 2 image was almost twice as high as that in the Hole 1 image, and it can be perceived visually that the Hole 1 image contained more delamination than the second image. Using the new algorithm, the Hole 1 image had a higher adjusted delamination factor (F_{da}), but the two values were closer to each other, which is more realistic. Additionally, the delamination factor considering the area (F_a) obtained by the previous algorithm was too low, which does not make sense. Since it is calculated via the area, it should be higher than one, which only happens with the new algorithm.

All the calculated delamination factors are considerably above the acceptable limit in the aeronautical industry, which is 1.200 [31], which means that, as confirmed visually, the holes exhibit severe damage and would not be accepted in a quality inspection. Nevertheless, with the aid of the new Python algorithm, a correct analysis can be performed, ensuring better hole assessment and higher quality in obtaining good results in future experiments.

5. Discussion

The literature contemplates some works in this field, although the most common ones focus on drilling composites, such as CFRP. For example, Krishnaraj et al. [32] performed drilling tests in a 4.2-mm-thick unidirectional CFRP, obtaining an F_d of 1.33 with a feed of 0.2 mm/rev. and a cutting speed of 15 mm/min. In the work of Durão et al. [24], a 4-mm-thick unidirectional epoxy CFRP was drilled with feeds of 0.03, 0.1, and 0.2 mm/rev. and a constant cutting speed of 52 m/min, which resulted in F_d , F_{da} values of 1.29, 1.47, 1.61, 1.81, 2.04, and 2.33, respectively. In both studies, the F_d values were close to the multi-material ones, but in the second, the F_{da} values were relatively lower. Studies regarding the delamination quantification and analysis of drilling multi-materials, specifically CFRP/Al, are also being conducted, although not very prominently. Nevertheless, few articles can be found with the delamination values clearly stated and, from the existing ones, the majority only address the F_d factor, the simplest delamination factor. This is the case in Mahdi et al. [33], where they drilled into a 7.6-mm-thick woven CFRP/Al stack with an uncoated classical twist drill, obtaining F_d values of around 1.2 using a feed of 0.0023 mm/min. and a cutting speed of 178 m/min. Furthermore, Ekici et al. [34] drilled a 5-mm-thick woven CFRP/Al multi-material composed of seven alternate material layers with a twist drill containing a TiN–TiAlN PVD coating, achieving F_{da} results of 1.8 using a feed of 0.2 mm/rev. and a cutting speed of 85 m/min. This value is lower than the one obtained

in this present study, which means that the tool coating and stack configuration greatly influence the final hole quality when drilling multi-materials.

6. Conclusions

The main objective of this work was to develop an algorithm in Python that could solve the problems felt using an algorithm previously developed in Matlab in the correct processing and analysis of X-ray images of drilled holes in a multi-material, specifically, a CFRP/Al material. The established goals were reached with the new algorithm, in addition to the fact that it is free, meaning that it is available for any user to benefit from, in contrast to the previous one, which required paid software. With the implementation of the novel algorithm, it is now possible to select an image with the hole delamination as an input, and automatically identify the different zones without manual intervention. This avoids errors usually made by inexperienced users, such as selecting the wrong image threshold, which may lead to a failure and the need to start the procedure from the beginning, whereas in the Python algorithm, the threshold is defined automatically and, thus, is always the preferred option. Besides this, the developed algorithm correctly segments the input image into the hole, delamination, remaining material, and noise, giving more reliable results when calculating the delamination factors. To corroborate this, the quality of two drilled holes was assessed using both algorithms. Using the Matlab algorithm, the delamination factors obtained were equal to 1.380 and 2.563 for the two holes, respectively, while the results obtained using the Python algorithm were equal to 3.957 and 3.383, respectively. However, the first results were incorrect, as could easily be confirmed visually, the first hole had a higher delamination area, which was not suggested by the results obtained by the Matlab algorithm. In contrast, the results obtained by the Python algorithm were correct. In addition, the user of the new Python algorithm did not need any previous experience in image processing and analysis, and therefore, the analysis was quicker, and the results were more accurate. This algorithm can be applied to drilling images and any procedure involving delamination and fibre pull-out in composite materials. In the future, the new algorithm can be tested in other hole-making processes to quantify the damage around the hole, and possible improvements can be implemented to achieve results even more accurately for those processes.

Author Contributions: Conceptualization, R.D.F.S.C., M.L.S.B. and J.M.R.S.T.; methodology, R.D.F.S.C., F.G.A.S. and J.M.R.S.T.; software, R.D.F.S.C. and M.L.S.B.; validation, L.M.P.D.; formal analysis, T.E.F.S.; investigation, R.D.F.S.C. and M.L.S.B.; resources, T.E.F.S. and A.M.P.d.J.; data curation, R.D.F.S.C.; writing—original draft preparation, R.D.F.S.C.; writing—review and editing, T.E.F.S., F.G.A.S. and J.M.R.S.T.; visualization, M.L.S.B.; supervision, A.M.P.d.J. and F.J.G.S.; project administration, T.E.F.S. and A.M.P.d.J.; funding acquisition, T.E.F.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Project “Hi-rEV—Recuperação do Setor de Componentes Automóveis” (C644864375-00000002), cofinanced by “Plano de Recuperação e Resiliência” (PRR), “República Portuguesa”, through NextGeneration EU program.

Data Availability Statement: The used data can be provided upon request.

Acknowledgments: The authors would like to thank Rui Silva and Miguel Figueiredo from DEMecFEUP for supporting the X-ray analysis.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mohan, A.; Poobal, S. Crack detection using image processing: A critical review and analysis. *Alex. Eng. J.* **2018**, *57*, 787–798. [CrossRef]
2. Gonzalez, R.E.; Rafael, C. *Woods, Digital Image Processing*, 3rd ed.; Taylor & Francis: Upper Saddle River, NJ, USA, 2008.
3. Ahad, M.A.R.; Kobashi, S.; Tavares, J.M.R.S. Advancements of Image Processing and Vision in Healthcare. *J. Healthc. Eng.* **2018**, *2018*, 3–6. [CrossRef] [PubMed]

4. Gavaskar, R.G.; Chaudhury, K.N. Fast Adaptive Bilateral Filtering. *IEEE Trans. Image Process.* **2019**, *28*, 779–790. [CrossRef] [PubMed]
5. Yin, D.; Lopes, R.G.; Shlens, J.; Cubuk, E.D.; Gilmer, J. A fourier perspective on model robustness in computer vision. *Adv. Neural Inf. Process. Syst.* **2019**, *32*. [CrossRef]
6. Huang, S.; Yan, W.; Jean, S.; Mavromatis, S. Shallow-Water image enhancement using relative global histogram stretching based on adaptive parameter acquisition. *Multimed. Model.* **2018**, *71*, 453–465. [CrossRef]
7. Ulutas, G.; Ustubioglu, B. Underwater image enhancement using contrast limited adaptive histogram equalization and layered difference representation. *Multimed. Tools Appl.* **2021**, *80*, 15067–15091. [CrossRef]
8. Said, K.A.M.; Jambek, A.B. Analysis of Image Processing Using Morphological Erosion and Dilation. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2021. [CrossRef]
9. Machado, J.M.; Tavares, J.M.R.S.; Camanho, P.P.; Correia, N. Automatic void content assessment of composite laminates using a machine-learning approach. *Compos. Struct.* **2022**, *288*, 115383. [CrossRef]
10. Zhang, Y.; Yuan, L.; Liang, W.; Member, S.; Zeng, Z. Carbon Fiber-Reinforced Polymer Composites Texture Angle Regression Based on the Improved Deep Hough Network. In *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–14. [CrossRef]
11. Cai, W.; Ma, Q.; Wang, Y.; Gan, X. Crashworthiness analysis and multi-objective optimization of Al/CFRP hybrid tube with initial damage under transverse impact. *Polym. Compos.* **2023**, *44*, 7953–7971. [CrossRef]
12. Wang, Z.; Wan, L.; Xiong, N.; Zhu, J.; Ciampa, F. Variational level set and fuzzy clustering for enhanced thermal image segmentation and damage assessment. *NDT E Int.* **2021**, *118*, 102396. [CrossRef]
13. Vinista, P.; Milton, M. A Novel Modified Sobel Algorithm for Better Edge Detection of Various Images. *Int. J. Emerg. Technol. Eng. Res.* **2019**, *7*, 25–31.
14. Barbosa, M.; Renna, F.; Dourado, N.; Costa, R. Construction of an Algorithm for Three-Dimensional Bone Segmentation from Images Obtained by Computational Tomography. *Stud. Comput. Intell.* **2023**, *1112*, 47–71. [CrossRef]
15. Feng, Q.; Gao, B.; Lu, P.; Woo, W.L.; Yang, Y.; Fan, Y.; Gu, L. Automatic seeded region growing for thermography debonding detection of CFRP. *NDT E Int.* **2018**, *99*, 36–49. [CrossRef]
16. De Albuquerque, V.H.C.; Silva, C.C.; de S, T.I.; Farias, J.P.; Tavares, J.M.R.S. Automatic evaluation of nickel alloy secondary phases from SEM images. *Microsc. Res. Tech.* **2011**, *74*, 36–46. [CrossRef] [PubMed]
17. Basivi, P.K.; Hamieh, T.; Kakani, V.; Pasupuleti, V.R.; Sasikala, G.; Heo, S.M.; Kim, C.W. Exploring advanced materials: Harnessing the synergy of inverse gas chromatography and artificial vision intelligence. In *TrAC-Trends in Analytical Chemistry 173*; Elsevier B.V.: Amsterdam, The Netherlands, 2024. [CrossRef]
18. Cruz, D.J.; Amaral, R.L.; Santos, A.D.; Tavares, J.M.R.S. Application of Digital Image Processing Techniques to Detect Through-Thickness Crack in Hole Expansion Test. *Metals* **2023**, *13*, 1197. [CrossRef]
19. Zhang, L.; Liu, Z.; Tian, W.; Liao, W. Experimental studies on the performance of different structure tools in drilling CFRP/Al alloy stacks. *Int. J. Adv. Manuf. Technol.* **2015**, *81*, 241–251. [CrossRef]
20. Blanco, D.; Rubio, E.M.; Marín, M.M.; Davim, J.P. Advanced materials and multi-materials applied in aeronautical and automotive fields: A systematic review approach. *Procedia CIRP* **2021**, *99*, 196–201. [CrossRef]
21. Chen, W.C. Some experimental investigations in the drilling of carbon fiber-reinforced plastic (CFRP) composite laminates. *Int. J. Mach. Tools Manuf.* **1997**, *37*, 1097–1108. [CrossRef]
22. Mohan, N.S.; Kulkarni, S.M.; Ramachandra, A. Delamination analysis in drilling process of glass fiber reinforced plastic (GFRP) composite materials. *J. Mater. Process. Technol.* **2007**, *186*, 265–271. [CrossRef]
23. Davim, J.P.; Rubio, J.C.; Abrao, A.M. A novel approach based on digital image analysis to evaluate the delamination factor after drilling composite laminates. *Compos. Sci. Technol.* **2007**, *67*, 1939–1945. [CrossRef]
24. Durão, L.M.P.; Tavares, J.M.R.S.; de Albuquerque, V.H.C.; Gonçalves, D.J.S. Damage evaluation of drilled carbon/epoxy laminates based on area assessment methods. *Compos. Struct.* **2013**, *96*, 576–583. [CrossRef]
25. Silva, J.M.; Ferreira, F.; Abreu, S.M.; Matos, J.E.; Durão, L.M.P. Correlation of drilling damage with mechanical strength: A geometrical approach. *Compos. Struct.* **2017**, *181*, 306–314. [CrossRef]
26. Devesa, L.F.S.; Matos, J.E.; Durão, L.M.P. Composites Part C: Open Access Experimental assessment of delamination extension on carbon/epoxy drilled plates. *Compos. Part C Open Access* **2021**, *5*, 100144. [CrossRef]
27. Geier, N.; Póka, G.; Jacsó, Á.; Pereszlai, C. A method to predict drilling-induced burr occurrence in chopped carbon fibre reinforced polymer (CFRP) composites based on digital image processing. *Compos. Part B Eng.* **2022**, *242*, 110054. [CrossRef]
28. Hernandez, A.; Maghami, A.; Khoshdarregi, M. A Machine Vision Framework for Autonomous Inspection of Drilled Holes in CFRP Panels. In Proceedings of the 2020 6th International Conference on Control, Automation and Robotics (ICCAR), Singapore, 20–23 April 2020; Volume 2020, pp. 669–675. [CrossRef]
29. Isbilir, O.; Ghassemieh, E. Comparative study of tool life and hole quality in drilling of CFRP/titanium stack using coated carbide drill. *Mach. Sci. Technol.* **2013**, *17*, 380–409. [CrossRef]
30. Marques, F.; Silva, F.G.A.; Silva, T.E.F.; Rosa, P.A.R.; Marques, A.T.; de Jesus, A.M.P. Delamination of Fibre Metal Laminates Due to Drilling: Experimental Study and Fracture Mechanics-Based Modelling. *Metals* **2022**, *12*, 1262. [CrossRef]

31. Ahmad Sobri, S.; Whitehead, D.; Mohamed, M.; Mohamed, J.J.; Mohamad Amini, M.H.; Hermawan, A.; Norizan, M.N. Augmentation of the delamination factor in drilling of carbon fibre-reinforced polymer composites (CFRP). *Polymers* **2020**, *12*, 2461. [CrossRef]
32. Krishnaraj, V.; Zitoune, R.; Collombet, F.; Davim, J.P. Challenges in drilling of multi-materials. *Mater. Sci. Forum* **2013**, *763*, 145–168. [CrossRef]
33. Mahdi, A.; Turki, Y.; Habak, M.; Salem, M.; Bouaziz, Z. Experimental study of thrust force and surface quality when drilling hybrid stacks. *Int. J. Adv. Manuf. Technol.* **2020**, *107*, 3981–3994. [CrossRef]
34. Ekici, E.; Motorcu, A.R.; Yıldırım, E. An Experimental Study on Hole Quality and Different Delamination Approaches in the Drilling of CARALL, a New FML Composite. *FME Trans.* **2021**, *49*, 950–961. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

A Novel Data Mining Framework to Investigate Causes of Boiler Failures in Waste-to-Energy Plants

Dong Wang ¹, Lili Jiang ², Måns Kjellander ³, Eva Weidemann ^{3,4}, Johan Trygg ⁴ and Mats Tysklind ^{4,*}

¹ Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, 2628 CN Delft, The Netherlands; guanerguan@gmail.com

² Department of Computing Science, Umeå University, SE-901 87 Umeå, Sweden; lili.jiang@umu.se

³ Umeå Energi, SE-901 05 Umeå, Sweden; mans.kjellander@umeaenergi.se (M.K.);
eva.weidemann@umu.se (E.W.)

⁴ Department of Chemistry, Umeå University, SE-907 36 Umeå, Sweden; johan.trygg@umu.se

* Correspondence: mats.tysklind@umu.se

Abstract: Examining boiler failure causes is crucial for thermal power plant safety and profitability. However, traditional approaches are complex and expensive, lacking precise operational insights. Although data-driven approaches hold substantial potential in addressing these challenges, there is a gap in systematic approaches for investigating failure root causes with unlabeled data. Therefore, we proffered a novel framework rooted in data mining methodologies to probe the accountable operational variables for boiler failures. The primary objective was to furnish precise guidance for future operations to proactively prevent similar failures. The framework was centered on two data mining approaches, Principal Component Analysis (PCA) + K-means and Deep Embedded Clustering (DEC), with PCA + K-means serving as the baseline against which the performance of DEC was evaluated. To demonstrate the framework's specifics, a case study was performed using datasets obtained from a waste-to-energy plant in Sweden. The results showed the following: (1) The clustering outcomes of DEC consistently surpass those of PCA + K-means across nearly every dimension. (2) The operational temperature variables T-BSH3rm, T-BSH2l, T-BSH3r, T-BSH1l, T-SbSH3, and T-BSH1r emerged as the most significant contributors to the failures. It is advisable to maintain the operational levels of T-BSH3rm, T-BSH2l, T-BSH3r, T-BSH1l, T-SbSH3, and T-BSH1r around 527 °C, 432 °C, 482 °C, 338 °C, 313 °C, and 343 °C respectively. Moreover, it is crucial to prevent these values from reaching or exceeding 594 °C, 471 °C, 537 °C, 355 °C, 340 °C, and 359 °C for prolonged durations. The findings offer the opportunity to improve future operational conditions, thereby extending the overall service life of the boiler. Consequently, operators can address faulty tubes during scheduled annual maintenance without encountering failures and disrupting production.

Keywords: power plants; failure analysis; data mining; deep embedded clustering

Citation: Wang, D.; Jiang, L.; Kjellander, M.; Weidemann, E.; Trygg, J.; Tysklind, M. A Novel Data Mining Framework to Investigate Causes of Boiler Failures in Waste-to-Energy Plants. *Processes* **2024**, *12*, 1346. <https://doi.org/10.3390/pr12071346>

Academic Editor: Maoyin Chen

Received: 15 May 2024

Revised: 18 June 2024

Accepted: 21 June 2024

Published: 28 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A boiler is an essential component in thermal power plants that utilize various fuels, including coal, oil, nuclear, or waste. Functioning as heat exchangers, boilers transform purified water into high-pressure steam through heat radiation from hot flue gas. This steam subsequently drives turbine blades for electricity generation. Typically, a boiler comprises economizers, evaporators, superheaters, and a steam drum, although the specific configuration may vary depending on the design and function of the power plant [1–4]. Given the harsh operating conditions of elevated temperature, pressure, corrosive substances, and mechanical stress, boilers are prone to frequent failures. Boiler tube failures account for the majority of unplanned shutdowns in power plants [5]. These failures commonly manifest as tube ruptures, significantly compromising both the safety and revenue of a power plant. In the event of tube rupture, the steam generation process can be halted, or worse, it

might lead to more serious accidents, compelling a complete plant shutdown for necessary repairs [6–9]. Such unplanned downtime leads to substantial economic ramifications for the plant. Research indicates that the average cost of a single day of unscheduled power plant downtime in Europe is approximately EUR 100,000 [10].

Investigating the causes of boiler failures holds significant importance for the safety and profitability of power plants. Extensive research has been dedicated to probing the origins of boiler failures, with a predominant focus on chemical and physical mechanisms. These culprits can be generally classified into several categories, including short-term overheating, long-term overheating (high-temperature creep), caustic corrosion from the water/steam side, hydrogen attack from the water/steam side, high-temperature corrosion from the fireside, and dew point corrosion from the fireside [11–16]. These phenomena often occur concurrently and can be intricately interconnected. For example, caustic corrosion can set the stage for hydrogen attack. When substantial quantities of alkaline compounds deposit on the inner surface of a tube, they initiate a reaction with the oxide layer, resulting in the depletion of it. Consequently, the hydroxide ions continue to interact with the inner material of the tube, leading to caustic corrosion. Simultaneously, atomic hydrogen is generated. The atomic hydrogen diffuses into the tube wall, where it reacts with metal carbide, forming methane. The accumulation of methane can result in the formation of cracks in the tube wall, a phenomenon known as hydrogen attack [5]. However, if the oxide layer remains intact and accumulates gradually over time, it can diminish heat exchange between the water/steam and flue gas. This reduction in heat exchange fosters localized overheating, which can significantly contribute to tube creep or fatigue [17,18].

Inspecting failed tubes typically demands complex chemical treatments and expensive equipment, such as Scanning Electron Microscopes [12,19]. Furthermore, findings from one part of the boiler may not be relevant to another due to variations in design and operating conditions among different sections of the boiler. Even for the same boiler component, conclusions may not apply consistently over time, given the dynamic nature of the surrounding environment. For example, variations in fuel mixtures can introduce fluctuations in the environment around the boiler, a common occurrence in waste-to-energy (WtE) plants where the quality of municipal solid waste is uncontrollable [20,21]. Furthermore, some studies indicate that prior corrosion experiences can influence the current rate of corrosion [22].

The ultimate objective of uncovering the root causes of failures is to leverage these insights to inform future operations and proactively prevent similar incidents. Unfortunately, conventional examination methods struggle to pinpoint the exact parameters and their specific values that contributed to the failure. The conventional examination results typically yield general recommendations on adjusting operating conditions, but these fall short of offering precise guidance to operators. Regarding operational guidance, an efficient approach to failure investigation should prioritize the connection between a failure and precise operating parameters without delving extensively into the intricacies of the failure mechanism, especially considering the intricate and variable nature of the aforementioned boiler failure mechanisms. Therefore, it is advisable to harness historical operational monitoring data and apply suitable data science methodologies for failure analysis.

Only very few data science applications related to boilers in power plants have been documented in the literature. For instance, one study demonstrated the high effectiveness of a data-driven approach comprising Wavelet Packet Transform analysis and Deep Neural Network in detecting boiler tube leakages [23]. Another developed two short-term forecasting models (Convolutional Neural Network (CNN) and Long Short-Term Memory Network) for predicting three safety indicators of a supercharged boiler. Both models yielded excellent results, but CNN was preferred due to its lower computational cost [24]. Additionally, an Extreme Gradient Boosting model, fine-tuned with a Particle Swarm Optimization algorithm, accurately predicted the metal temperature time series, enabling the early detection of metal temperature anomalies in a coal-fired boiler [25]. Furthermore, the Extra-Tree classifier and Minimum Redundancy Maximum Relevance model were found to

be highly effective in selecting the most relevant sensors for detecting faults in turbines and boilers, respectively. The results indicated a substantial reduction in the number of sensors needed for fault detection and a significant increase in detection accuracy [26]. Moreover, three individual machine learning algorithms, Random Forest, Lasso, and Support Vector Regression, along with the ensemble model based on them, were employed to forecast boiler faults in a thermal power plant by predicting the key performance indicators of the boiler. The findings indicated that the ensemble model outperformed all three individual models, delivering a highly satisfactory outcome [27].

However, the literature presents two gaps. Firstly, there is a lack of data science applications specifically focused on analyzing root causes of boiler failures. Secondly, all the prior studies are based on supervised learning, which is not suitable for scenarios where operational data lacks clear labels, a common occurrence in engineering settings, including the case study addressed in this research. Motivated by these gaps, this study introduces a novel and methodical framework that integrates engineering expertise with data science methods to investigate the causes of boiler failures and improve future operational practices. Beginning with formulating the boiler failure investigation problem into a data mining problem, the framework encompasses data preprocessing, model building and selection, and result evaluation and analysis, culminating in the provision of precise operational recommendations to prevent future boiler failures. The data science techniques employed predominantly include Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA), K-means clustering, and Deep Embedded Clustering (DEC). This framework is designed and leveraged to achieve the aforementioned objective, which is pinpointing the exact operational parameters and their specific values that contributed to boiler failures so that similar failures in the future can be proactively prevented by adjusting process operations.

This paper is structured as follows. Following this introduction, the subsequent section introduces the case study subject and the datasets used in this research. The case study was conducted on a WtE facility situated in Umeå, Sweden. Its purpose was to demonstrate the details of the framework and validate the framework's applicability in a real engineering context. The Section 3 that follows presents the framework, the chosen data science techniques, and the rationale behind their adoption. The results derived from the case study and the ensuing discussion are then presented in the subsequent section. Finally, the Section 5 summarizes the key findings of this research.

2. Overview of Umeå Waste-to-Energy Plant and Data Origin

The subject of the case study is the WtE plant located in Umeå, Sweden, operated by Umeå Energi. Umeå WtE plant is a 65 MW Combined Heat and Power (CHP) plant fueled by approximately 50% municipal solid waste and 50% industrial waste. Boasting a waste processing capacity of around 20 t/h, the plant operates roughly 8000 h per year and undergoes an annual scheduled maintenance shutdown.

Illustrated in Figure 1 is the boiler-related layout of the Umeå WtE plant. Waste is introduced through the hopper for incineration on the grate, and the resulting flue gas traverses four flue gas passages until it reaches the flue gas treatment modules. The initial three passages are vertically oriented, primarily relying on radiation for heat transfer, while the fourth passage is horizontal and characterized by convective heat exchange. In the initial three passages, numerous tubes containing water/steam are positioned along the inner walls. These tubes serve dual purposes: functioning as evaporators within the boiler system and acting as safeguards against overheating for the walls. In the fourth passage lies the central segment of the boiler arrangement, consisting of one evaporator unit, three superheater units, another evaporator unit, and three economizer units, arranged from left to right. Within this segment, water/steam typically flows counter to the flue gas to facilitate convective processes. Within the economizers, boiler feed water is raised to a temperature below boiling point under certain water pressure. Concurrently, the flue gas surrounding the economizers achieves the desired (lower) temperature for subsequent flue

gas treatment. Following the economizers, the heated water ascends to the uppermost steam drum situated atop the flue gas passages. Subsequently, the water in the steam drum flows through the downcomers to reach the evaporators, where it undergoes a phase transition into wet steam before ascending back to the steam drum. Within the steam drum, a separator works to transform the wet steam into saturated steam. This saturated steam is extracted from the upper section of the steam drum and subsequently undergoes additional heating in the superheaters to attain the status of superheated steam. The superheating process is crucial for optimizing turbine efficiency and ensuring its continued optimal performance.

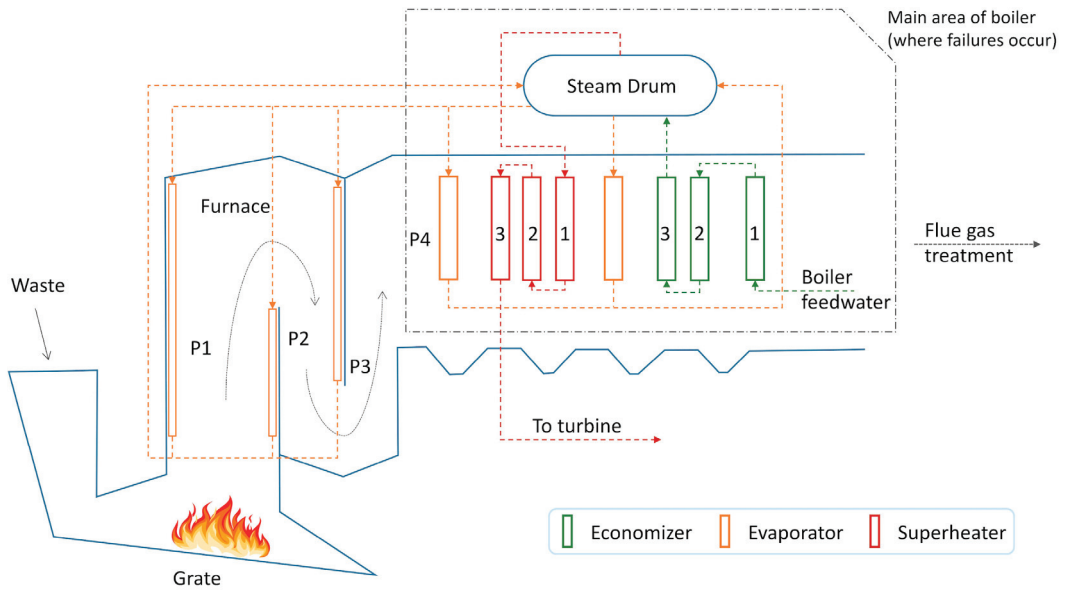


Figure 1. Boiler-related layout of Umeå WtE plant. P1, P2, P3, and P4 represent the first, second, third, and fourth flue gas passages, respectively.

The entirety of the plant is monitored by numerous online sensors. With the assistance of the engineers at the Umeå WtE plant, 66 of them (presented in Table S1 in Supplementary Material) were identified to possess potential associations with boiler failure occurrences. Consequently, there were 66 variables in the case study datasets. Throughout the case study, a total of three boiler failures were examined, each corresponding to a specific repair stoppage. The timeframes for these stoppages were derived from the log. Closely proximate failures were analyzed collectively, resulting in the investigation of two datasets (as outlined in Table 1). The time spans of the datasets were decided by setting the starting points three to five months (depending on the availability of data) before the initial stoppage. This approach ensured an adequate number of observations for evaluating distinctions between normal and abnormal operational conditions (further elaborated on in Section 3.1). The datasets were obtained at a 30 min resolution through averaging, despite the original data being of a higher resolution. Averaging was employed for two main purposes: noise reduction and, notably, mitigation of the time-lag impact caused by the movement of water, steam, and flue gas.

Table 1. Summary information of datasets.

Dataset ID	No. of Failures/Stoppages	Data Resolution	Dataset Size (Row × Column)
A	2	30 min	5808 × 66
B	1	30 min	7856 × 66

3. Methodology

3.1. The Framework

The investigation into the causes of boiler failure in this study is primarily grounded in the inference that there are certain abnormal conditions giving rise to the failure, and these abnormal conditions persist until the operators detect the failures and halt the process line. Hence, the primary phase of abnormal conditions ceases around the time of the commencement of stoppage/repair. Preceding the occurrence of abnormal conditions there exists a period characterized by normal operational conditions. Through a comparison of variable values under normal and abnormal conditions, we can identify which variables deviate from the expected behavior and consequently lead to failure.

However, identifying the normal and abnormal periods presents a two-fold challenge. First, the monitored data lack labels, aside from the logging of boiler repair events. Second, the criteria for classifying operational conditions as abnormal may differ from one tube to another and across various time periods, owing to variations in the functions of different tubes and the potential degradation of their properties over time. Thus, to the authors' best knowledge, case-based unsupervised clustering stands as the sole fitting approach for identifying normal and abnormal periods in this study. The specific method of unsupervised clustering applied in this study is K-means [28].

Figure 2 shows the flowchart of the failure analysis framework in this study. Following the initial data cleansing process, the application of Discrete Wavelet Transform (DWT) served to effectively eliminate any noise stemming from the sensors. Next, embedding techniques were implemented to mitigate noise that may exist among different variables. Importantly, the utilization of embedding also aids in averting the curse of dimensionality [29], as it effectively reduces the dimensionality of the data. This study employed two distinct embedding techniques. The first approach was Principal Component Analysis (PCA), whereas the second approach was a Deep Neural Network (DNN) integrated within the structure of Deep Embedded Clustering (DEC). Following the embedding process, the transformed data were input into K-means, producing the final clustering results. PCA + K-means served as the baseline against which the performance of DEC was evaluated. For PCA + K-means, the developments of PCA and K-means are loosely combined as the information flow is unidirectional from PCA to K-means. Conversely, in DEC, the DNN and K-means are seamlessly connected and trained simultaneously and iteratively. The information flow in DEC is bidirectional: from DNN to K-means, further extending to KL divergence, and reciprocally from KL divergence back to K-means and DNN. Having obtained the initial clustering results that categorized all observations into three distinct clusters, the subsequent task was to determine the identity of each cluster. Initially, the repair cluster (period of stoppage) can be discerned by referencing the operational log, as the log indicates when the boiler underwent repair and subsequently resumed operation. Following this, the contiguous timeframe directly preceding the repair event can be recognized as the cluster indicative of abnormal operating conditions. Finally, the continuous timeframe preceding the cluster of abnormal conditions can be designated as the cluster of normal operating conditions. Once the clusters were identified, an assessment and comparison of the clustering outcomes between PCA + K-means and DEC were conducted from an operational perspective. This evaluation aimed to determine the optimal clustering result. Based on this optimal clustering result, histograms were constructed for each individual variable. The purpose was to scrutinize potential disparity in distribution patterns between clusters under normal and abnormal conditions. To quantify this distribution disparity, the Normalized Peak Shift (NPS) metric was employed. It assesses the normalized shift in peak values (the most frequent values) within two distinct distributions. Variables that displayed a noticeable shift, characterized by NPS values surpassing the threshold of 30%, were identified as contributors to failure occurrences. These identified variables require vigilant monitoring to proactively prevent the recurrence of similar failures. Furthermore, recommendations concerning their values during production were formulated based on observations of their distributions under both normal and abnormal conditions.

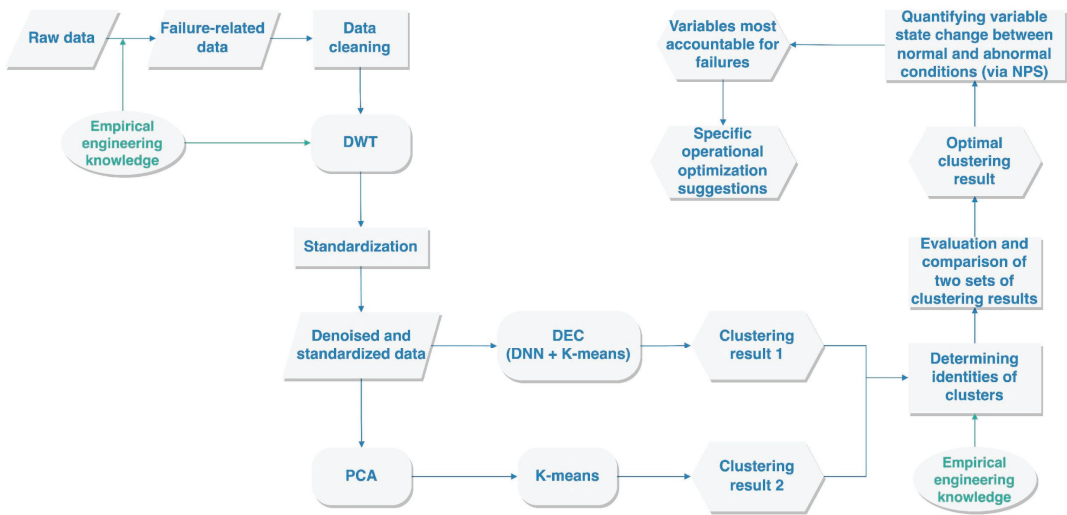


Figure 2. Flowchart of the boiler failure investigation framework. Empirical engineering knowledge refers to the knowledge engineers have gained from their operating or maintenance experience, and it is grounded in physical or chemical mechanisms in the engineering context.

In addition to the utilization of data science techniques, the experiential insights contributed by WtE plant engineers held a substantial influence within the framework. The term ‘empirical engineering knowledge’ within the framework pertains to the experiential knowledge garnered by the engineers through their operational and maintenance experiences. This encompasses their specialized engineering expertise in the realms of chemistry and physics. This form of knowledge served as an important complement within this framework, ensuring the data science methodologies were effectively employed to align seamlessly with the study’s objectives. For example, as described in Section 2, the engineers helped to narrow down relevant variables significantly. Moreover, empirical engineering knowledge was sought when setting the noise threshold in the DWT process. More importantly, it was employed to evaluate and compare different clustering results, and, finally, to select the optimal one.

3.2. Discrete Wavelet Transform

Discrete Wavelet Transform (DWT) is a powerful tool for denoising signal data [30]. DWT-based denoising typically comprises three steps: Decomposition, Thresholding, and Reconstruction.

Decomposition: Solve the DWT coefficients from the decomposition expansion of the signal with noise. Given a signal $s(t)$, decompose it using DWT to obtain the approximation coefficients $c_{j_0}(k)$ and detail coefficients $d_j(k)$. The decomposition expansion can be expressed as Equation (1):

$$s(t) = \sum_k c_{j_0}(k) \varphi_{j_0,k}(t) + \sum_k \sum_{j=j_0}^{\infty} d_j(k) \psi_{j,k}(t) \quad (1)$$

Here, $\psi_{j,k}(t) = 2^{\frac{j}{2}} \psi(2^j t - k)$ is the wavelet function, and $\varphi_{j_0,k}(t) = 2^{\frac{j_0}{2}} \varphi(2^{j_0} t - k)$ is the scaling function associated with the wavelet function. j is the level parameter, and k is the translation parameter.

Thresholding: Keep the detail coefficients associated with the signal as they are, and replace the ones related to noise with zeros. Given the detail coefficients $d_j(k)$, apply a

threshold T to it to suppress the noise. This study adopted the hard thresholding approach that is presented in Equation (2):

$$\tilde{d}_j(k) = \begin{cases} d_j(k) & \text{if } |d_j(k)| \geq T \\ 0 & \text{if } |d_j(k)| < T \end{cases} \quad (2)$$

Reconstruction: Reconstruct the signal with the modified coefficients. Equation (3) demonstrates the reconstructed and denoised signal $\tilde{s}(t)$ using the original approximation coefficients $c_{j_0}(k)$ and the modified detail coefficients $\tilde{d}_j(k)$:

$$\tilde{s}(t) = \sum_k c_{j_0}(k) \varphi_{j_0,k}(t) + \sum_k \sum_{j=j_0}^{\infty} \tilde{d}_j(k) \psi_{j,k}(t) \quad (3)$$

For the DWT work in this study, we used the Python package *PyWavelets* (version: 1.1.1) [31]. Specifically, we use the *wavedec* application programming interface (API) for multilevel decomposition with the arguments 'db6' for *wavelet* and 5 for *level*.

3.3. Principal Component Analysis

Principal Component Analysis (PCA) yields several principal components (PCs), which are the result of mapping the raw data's variation space to a new space of lower dimensionality. All the PCs are linear combinations of the original variables, but the PCs are orthogonal to each other. The number of PCs is determined by maximizing the total variation explained by the PCs, while minimizing the noise remaining. Typically, PCA is conducted by calculating the covariance matrix of the original data, which is followed by eigenvalue decomposition of the covariance matrix. The eigenvectors from the decomposition define the directions of the PCs. The eigenvectors are sorted according to their corresponding eigenvalues, and larger eigenvalues represent greater capability of explaining variation by the corresponding PCs [32]. For the PCA work in this study, we used the API *sklearn.decomposition.PCA* in the Python package *sickit-learn* (version: 0.24.0) [33].

3.4. K-Means

The idea of K-means clustering is quite straightforward: all the observations in the dataset are grouped into k clusters based on their distances to each other, minimizing the distances among observations within each cluster, while maximizing the distances among different clusters [28]. To be specific, the objective of K-means is to minimize E , as presented in Equation (4):

$$E = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (4)$$

Here, k is the set number of clusters, C_i is the i th cluster, and μ_i is the mean vector (centroid) of C_i . Since the total variance is constant, minimizing E is equivalent to maximizing the variance among different clusters.

However, minimizing E is an NP-hard problem. Thus, the following heuristic algorithm is used:

- (1) Randomly generate k initial centroids within the dataset.
- (2) Generate new clusters by assigning every observation to its nearest centroid.
- (3) Calculate the centroids of the new clusters.
- (4) Repeat Steps 2 and 3 until convergence is reached.

For the K-means work (for both PCA + K-means and DEC) in this study, we used the Python API *sklearn.cluster.KMeans* in the Python package *sickit-learn* (version: 0.24.0) [33] with the parameters *n_clusters* = 3, *tol* = 0.001, and *random_state* = 5. The number of clusters for K-means was set to 3, because for every case, there are three categories of operating conditions for the boiler—normal conditions, abnormal conditions, and repair/stoppage.

3.5. Deep Embedded Clustering

Deep Embedded Clustering (DEC) is a method that learns variable embedding and observation clustering simultaneously using deep neural networks (DNN) and K-means [34]. Instead of clustering the original data X into k clusters, DEC first maps X nonlinearly onto a new space Z with much lower dimensionality. The mapping is conducted through a DNN with the parameters θ . Subsequently, DEC learns the centroids set $\{\mu_i \in Z\}_{i=1}^k$ and the parameter θ simultaneously. DEC consists of two stages:

- (1) Using a stacked autoencoder (SAE) to initialize the parameters θ .
- (2) Iterating the process of generating an auxiliary target distribution and minimizing the Kullback–Leibler (KL) divergence between the soft assignment q_{ij} and the auxiliary target distribution p_{ij} . By doing this, the parameters θ are optimized.

SAE is applied because much research has demonstrated its capability of consistently yielding good representations (results of mapping) for real-world datasets [35–37]. As shown by Figure S1 in Supplementary Material, SAE consists of an encoder and a decoder, and their structures are symmetric with respect to one another. The low-dimension layer in the middle is the embedded space. The activation function applied for the SAE (except for the embedded layer and the reconstruction layer) in this study is ReLU [38]. The training is performed by minimizing the least-square loss between the input layer and the reconstruction layer. Once initialization is carried out, the encoder part is selected to concatenate with K-means in Stage (2) for further training.

In Stage (2), the loss function KL divergence is expressed in Equation (5):

$$KL(P \parallel Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (5)$$

The term q_{ij} mentioned above is defined in Equation (6):

$$q_{ij} = \frac{\left(1 + \frac{\|z_i - \mu_j\|^2}{\alpha}\right)^{-\frac{\alpha+1}{2}}}{\sum_{j'} \left(1 + \frac{\|z_i - \mu_{j'}\|^2}{\alpha}\right)^{-\frac{\alpha+1}{2}}} \quad (6)$$

Here, $z_i \in Z$ corresponds to $x_i \in X$, and α are the degrees of freedom of the Student's t distribution. q_{ij} indicates the probability of assigning sample i to cluster j .

The term p_{ij} mentioned above is defined in Equation (7):

$$p_{ij} = \frac{\frac{q_{ij}^2}{f_j}}{\sum_{j'} \frac{q_{ij'}^2}{f_{j'}}} \quad (7)$$

Here, $f_j = \sum_i q_{ij}$ are soft cluster frequencies.

The DEC work in this study was carried out based on the Keras script written by Xifeng Guo [39]. The original script was designed to perform image clustering, but we customized it to fit this study.

3.6. Key Hyperparameters of Models

Since the primary focus of this study is the optimization of the WtE process, the description and discussion of the data science methodology are presented concisely. Therefore, only the core hyperparameters of the models are discussed in this paper, while any API arguments not explicitly mentioned are retained at their default settings. We adopted the Grad Student Descent approach [40] for tuning all the model hyperparameters. In our analyses, two key hyperparameters took center stage: the number of PCs (*npc*) for PCA + K-means, and the number of neurons in the embedded layer (*nn_el*) for DEC. They both dictate the dimensionality of the embedded spaces. To facilitate optimization, we

defined an identical range, specifically {2, 3, 4, 5, 6, 7, 8}, for both of these hyperparameters' tuning. In cases where the optimal outcome for either approach emerged at 2 or 8, an exploration of 1 or 9 would be initiated to assess the potential for yielding a new optimal result. This iterative process would continue until the superior outcome was no longer derived from the boundary values of the specified range. For DEC, additional significant hyperparameters included the number of hidden layers within the encoder (*nhl*) and the number of neurons within these layers (*nn_hl*). Given the datasets' moderate scales, the optimization range for *nhl* was designated as {1, 2, 3}, with 2 consistently identified as the optimal selection across all datasets. To enhance tuning efficiency, we maintained uniformity in *nn_hl* across all hidden layers for a specific dataset. Nevertheless, the optimization ranges and optimal values for *nn_hl* differed among datasets.

3.7. Normalized Peak Shift

Normalized Peak Shift (NPS) was introduced as the metric to evaluate the state difference of each variable between normal conditions and abnormal conditions. It is based on the notion that the most frequently observed value (peak value of a distribution) under certain conditions can effectively encapsulate the variable's state under those conditions. Thus, by estimating the peak values' shift between two distributions, the state change of the variable of interest can be quantified. To enhance the clarity and utility of this metric, the range of variable values under normal conditions (excluding extreme values) is utilized to normalize the shift, resulting in NPS values presented as percentages. NPS can be calculated by Equation (8):

$$NPS = \frac{|f(x_n) - h(x_a)|}{\max_{1 \leq i < j \leq k} |x_{ni} - x_{nj}|} \quad (8)$$

Here, $f(x_n)$ is the Probability Mass Function (PMF) of the variable values under normal conditions (x_n), while $h(x_a)$ is the PMF of the variable values under abnormal conditions (x_a). k is the number of observations under normal conditions after excluding the observations with extreme variable values.

4. Results and Discussion

4.1. Results for Dataset A

There are two boiler failures in Dataset A. Within the DEC approach, we defined the optimization range for *nn_hl* as {70, 80, 90, 100}, determining that 80 emerged as the optimal value. By employing this optimal *nn_hl*, we obtained the DEC's optimal clustering outcome when *nn_el* was set to 2. Within the PCA+K-means approach, we observed that none of the values within the set {1, 2, 3, 4, 5, 6, 7, 8} proved to be an effective *mpc* value that was capable of delineating a distinct separation between the normal conditions cluster and the abnormal conditions cluster. Accordingly, as summarized in Table 2, the optimal clustering result for Dataset A was achieved using DEC with *nhl*, *nn_hl*, and *nn_el* set to 2, 80, and 2, respectively. The optimal clustering result is shown in Figure 3. The complete compilation of results obtained from both DEC and PCA + K-means under various hyperparameter settings can be accessed in Section S3 in the Supplementary Material.

Applying the method expounded in Section 3.1, the three clusters in Figure 3 can be readily identified. Cluster 2 is the repair periods/stoppages caused by the failures since it matches the timelines of repair according to the log information. Consequently, Cluster 1 can be identified as the abnormal conditions directly contributing to the failures, while Cluster 0 is the normal conditions. Based on this clustering result, the NPS values of each variable were calculated. The histograms of the variables displaying substantial disparities between normal and abnormal conditions ($NPS > 30\%$) are demonstrated in Figure 4.

As illustrated in Figure 4, 17 variables exhibit NPS values surpassing 30%. Notably, T-FaG2 demonstrates the highest value at 94.0%, while T-BSH1r records the lowest at 33.8%. The substantial number of involved variables coupled with elevated average NPS values underscores an extensive and noteworthy shift in the state between these two operational conditions. It is noteworthy that the variables T-FaG1 and T-FaG2 correspond to the

two temperature sensors in closest proximity to the incineration area, thus their values are expected to be significantly higher compared to others. However, as depicted in Figure 4, a substantial portion of T-FaG2 values is remarkably low (nearing 0) under both normal and abnormal conditions, in contrast to the regular patterns observed in other variables' values. This observation leads to the inference that the T-FaG2 sensor experienced prolonged malfunction while the data were recorded. Consequently, despite having the highest NPS value among all variables, T-FaG2 is excluded from further consideration and analysis.

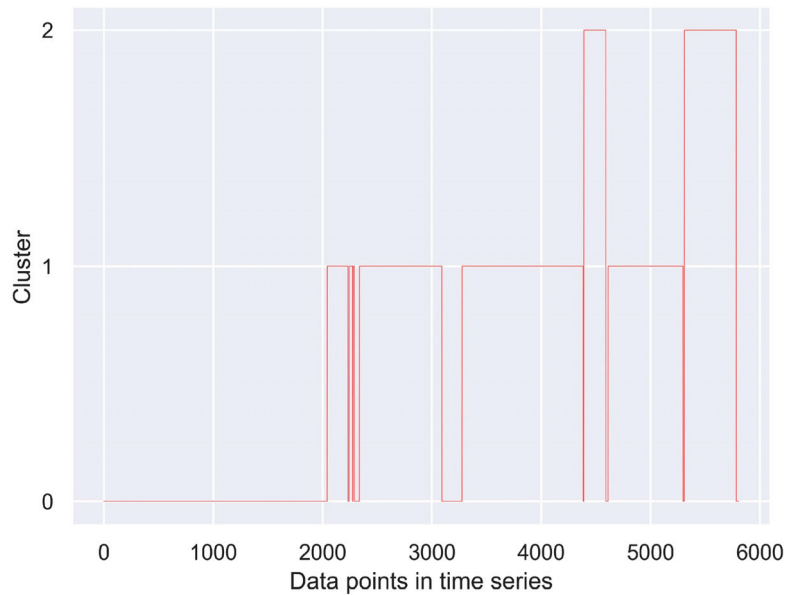


Figure 3. Optimal clustering result of Dataset A. Cluster 2 represents repair periods/stoppages, Cluster 1 abnormal conditions, and Cluster 0 normal conditions.

Among the variables displayed in Figure 4, all except T-BEM3r1 pertain to flue gas temperatures within the superheater area (red modules in Figure 1). This collective observation implies a significant overheating issue across the entirety of the superheater area, which emerges as the most likely culprit behind the failures encountered in Dataset A. The top three variables ranked by their NPS values are T-BSH3rm, T-BSH2l, and T-BSH3r, exhibiting NPS values of 57.2%, 56.7%, and 53.4%, respectively. This signifies that the temperatures of flue gas at “Superheater 3 roof middle”, “Superheater 2 left”, and “Superheater 3 right” deviated considerably from the normal operational temperatures. Therefore, these variables stand out as the primary contributors to the failures observed in Dataset A. Given the adjacency of all three superheaters, the temperatures within the superheater area are naturally linked and interdependent. Thus, prioritizing the management of the top three influential variables has the potential to effectively address the overarching overheating concern throughout the entire superheater area. According to Figure 4, the peak operational values of T-BSH3rm, T-BSH2l, and T-BSH3r during normal conditions were recorded as 527 °C, 432 °C, and 482 °C, respectively. Under abnormal conditions, these values escalated to 594 °C, 471 °C, and 537 °C, correspondingly. To ensure future production safety, it is advisable to maintain the operational levels of T-BSH3rm, T-BSH2l, and T-BSH3r around 527 °C, 432 °C, and 482 °C respectively. Additionally, it is crucial to prevent these values from reaching or exceeding 594 °C, 471 °C, and 537 °C for prolonged durations. Please refer to Section 4.3 for the discussion on the results and underlying mechanisms.

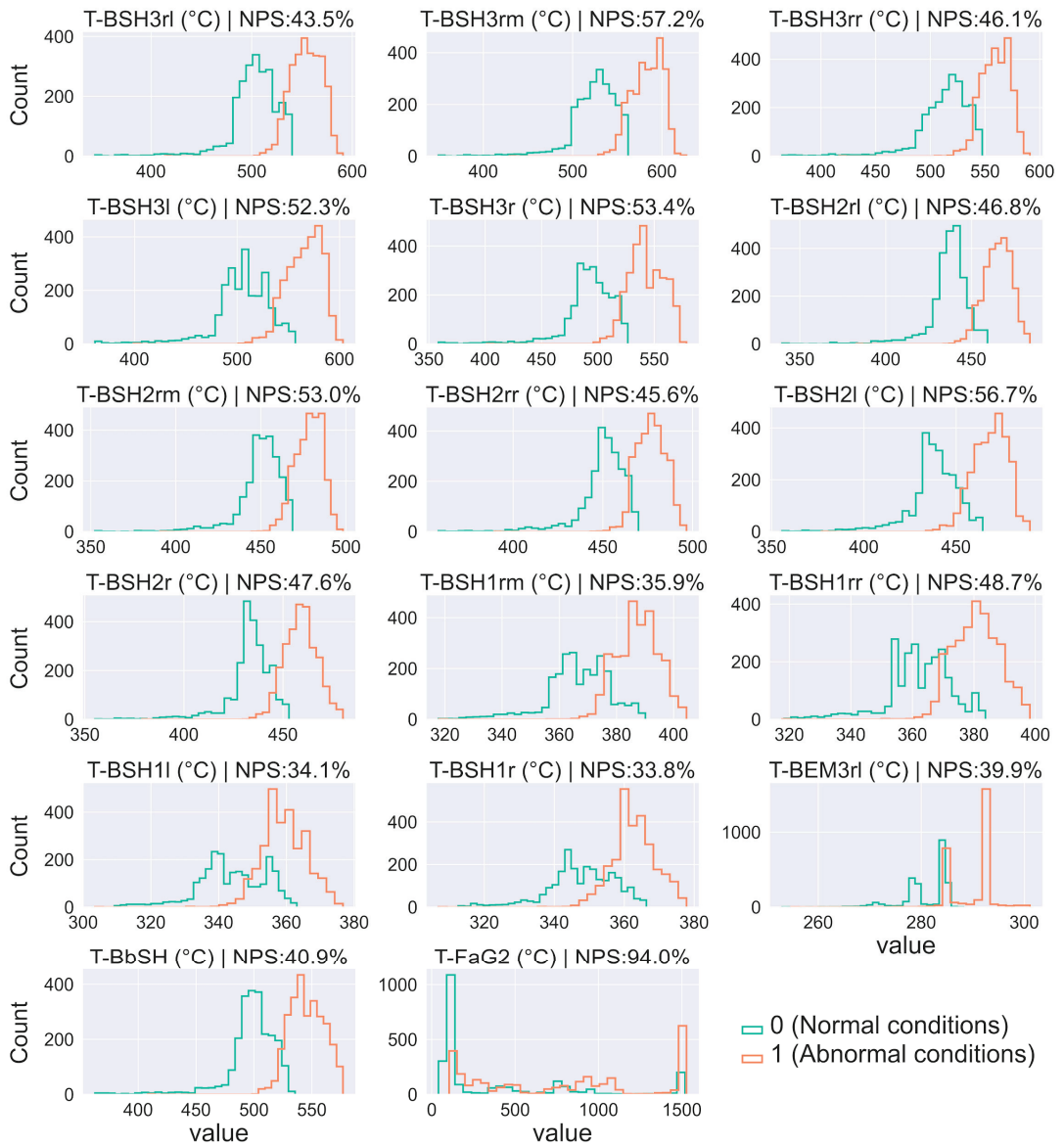


Figure 4. Histograms of Dataset A variables with significant shifts between normal and abnormal conditions. The histograms are independent of each other. For each of them, the x-axis indicates the variable's name, unit, and NPS value.

4.2. Results for Dataset B

There is one boiler failure in Dataset B. Within the DEC approach, we defined the optimization range for nm_hl as {80, 100, 128, 156}, determining that 128 emerged as the optimal value. By employing this optimal nm_hl , we obtained the DEC's optimal clustering outcome when nm_el was set to 8. Within the PCA+K-means approach, we observed that none of the values within the set {2, 3, 4, 5, 6, 7, 8, 9} proved to be an effective npc value that was capable of delineating a distinct separation between the normal conditions

cluster and the abnormal conditions cluster. Accordingly, as summarized in Table 2, the optimal clustering result for Dataset A was achieved using DEC with *nhl*, *nm_hl*, and *nm_el* set to 2, 128, and 8, respectively. The optimal clustering result is shown in Figure 5. The complete compilation of results obtained from both DEC and PCA + K-means under various hyperparameter settings can be accessed in Section S4 in the Supplementary Material.

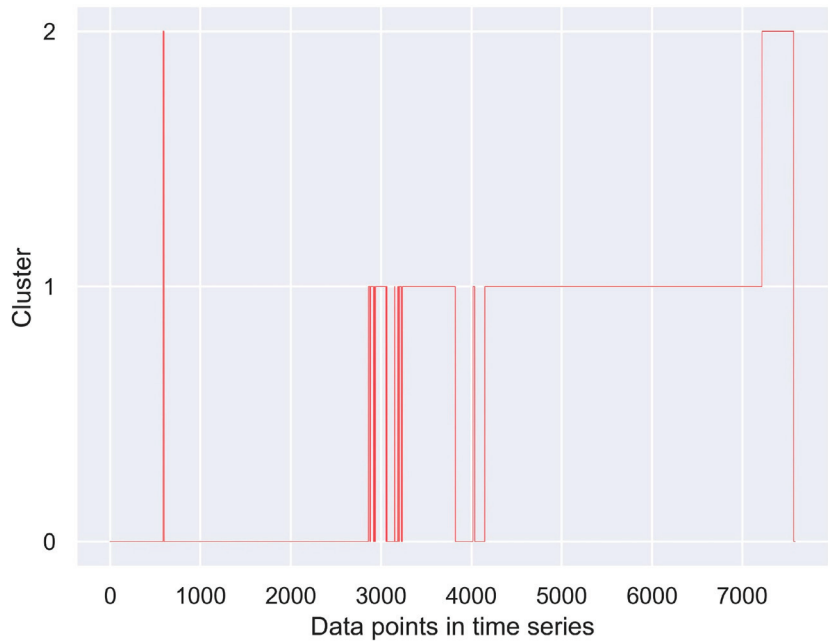


Figure 5. Optimal clustering result of Dataset B. Cluster 2 represents repair periods/stoppages, Cluster 1 abnormal conditions, and Cluster 0 normal conditions. The extremely brief appearance of Cluster 2 at the beginning reflects a transient fault in the monitoring system.

Table 2. Optimal hyperparameter values for the DEC models on Datasets A and B.

Dataset ID	<i>nhl</i>	<i>nm_hl</i>	<i>nm_el</i>
A	2	80	2
B	2	128	8

Applying the method expounded in Section 3.1, three distinct clusters can be discerned. Cluster 2 signifies stoppages, Cluster 1 denotes abnormal conditions, and Cluster 0 corresponds to normal conditions. Nevertheless, a brief segment of Cluster 2 is evident within the initial phase of Cluster 0. Engineers at Dáva 1 believed that this occurrence is likely unrelated to the boiler failure, suggesting it may reflect a transient glitch within the monitoring system. Based on this clustering result, the NPS values of each variable were calculated. The histograms of the variables displaying substantial disparities between normal and abnormal conditions ($NPS > 30\%$) are demonstrated in Figure 6.

As illustrated in Figure 6, 10 variables exhibit NPS values surpassing 30%. Notably, T-BSH11 demonstrates the highest value at 89.5%, while T-BEM2rl records the lowest at 32.5%. It is worth noticing that T-FaG2 holds the second-highest NPS value (74.0%) among all the variables. However, the histogram pattern of T-FaG2 closely resembles its counterpart in Dataset A. Consequently, based on the analysis of Dataset A, T-FaG2 is disregarded for subsequent consideration and analysis, despite its elevated NPS value.

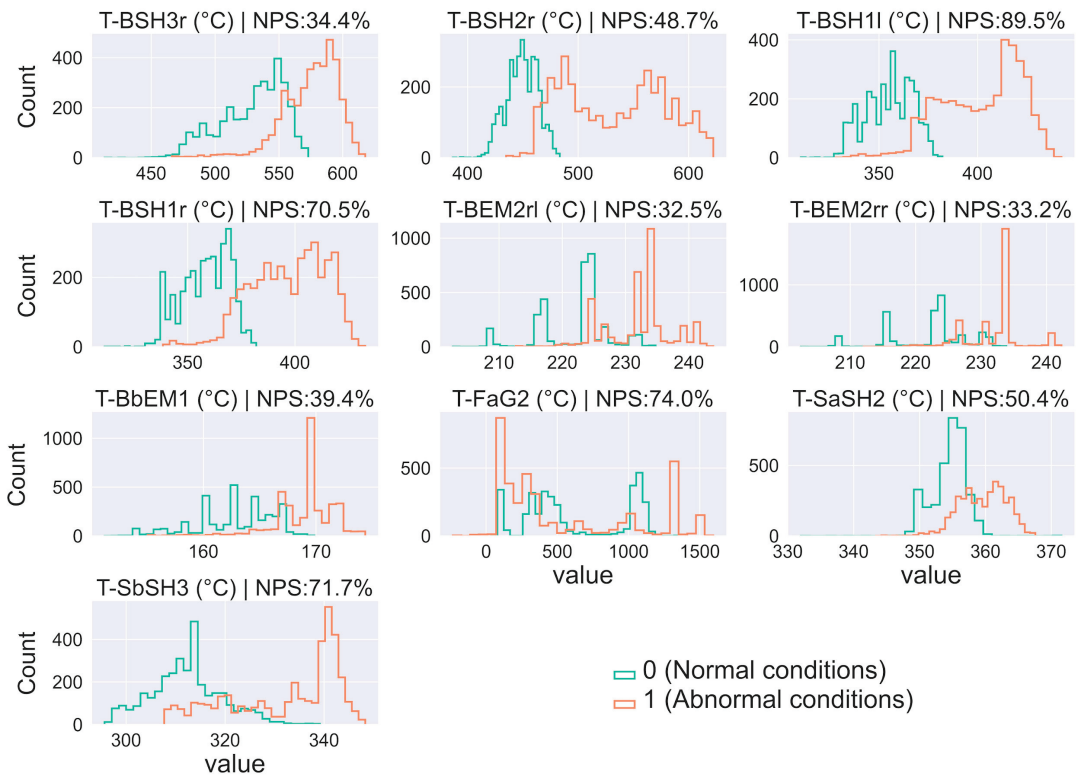


Figure 6. Histograms of Dataset *B* variables with significant shifts between normal and abnormal conditions. The histograms are independent of each other. For each of them, the *x*-axis indicates the variable value, the *y*-axis indicates the count/frequency of values, and the title includes the variable's name, unit, and NPS value.

In contrast to Dataset *A*, the variables in Figure 6 comprise a more balanced combination of temperatures from both economizers (green modules in Figure 1) and superheaters. Among them, T-BbEM1, T-BEM2rr, and T-BEM2rl are temperatures within the economizer sector, while the remaining variables pertain to temperatures in the superheater sector. In addition, this implies that the temperature of the evaporator situated between the first superheater and the third economizer likely underwent a comparable shift pattern, despite the absence of a designated sensor for the evaporator. The presence of these variables in Figure 6 suggests an overall overheating of the entire fourth flue gas pass during the abnormal conditions. What resembles Dataset *A* is the consistent prominence of variables associated with the superheaters. The top three variables for Dataset *B* ranked by their NPS values are T-BSH1l, T-SbSH3, and T-BSH1r, recording NPS values of 89.5%, 71.7%, and 70.5% respectively. They are the temperatures of flue gas at “Superheater 1 left”, steam prior to “Superheater 3”, and flue gas at “Superheater 1 right”. Additionally, T-SaSH2 and T-BSH2r exhibit noteworthy NPS values of 50.4% and 48.7%, respectively, ranking as the fourth and fifth most influential variables. Importantly, all five variables correspond to either flue gas or steam temperatures within the superheater area, signifying an intensified overheating specifically concentrated in the superheater region within the already overheated fourth flue gas pass. Thus, these variables emerge as the primary causative factors behind the failure observed in Dataset *B*. Given the linkage and interdependence between the temperatures in the fourth flue gas pass, prioritizing the management of the top three influential variables has the potential to effectively address the overheating con-

cern throughout the entire fourth flue gas pass. According to Figure 6, the peak operational values of T-BSH11, T-SbSH3, and T-BSH1r during normal conditions were recorded as 356 °C, 313 °C, and 368 °C, respectively. Under abnormal conditions, these values escalated to 412 °C, 340 °C, and 408 °C, correspondingly. It is worth noting that T-BSH11 and T-BSH1r were also recognized as significant contributors to the failures in Dataset A (see Figure 4), despite not being part of the top three ranked variables. In Dataset A, their respective peak values during normal conditions were 338 °C and 343 °C, while during abnormal conditions, these values increased to 355 °C and 359 °C. Consequently, to uphold the highest safety standards, it is advisable to maintain the operational levels of T-BSH11, T-SbSH3, and T-BSH1r around 338 °C, 313 °C, and 343 °C, respectively. Additionally, it is crucial to prevent these values from reaching or exceeding 355 °C, 340 °C, and 359 °C for prolonged durations. Please refer to Section 4.3 for the discussion on the results and underlying mechanisms.

4.3. Discussion on the Results and Underlying Mechanisms

Four types of operational variables—temperature, pressure, chemical concentration, and fluid flow rate—were investigated in this study. Among these, the results illustrated that elevated temperatures, particularly those in close proximity to both inside and outside of superheaters, emerge as the predominant cause for boiler failures. This observation aligns with numerous prior investigations into boiler failure, which employed conventional chemical and physical methodologies. Engineers and researchers widely acknowledge that elevated temperatures can directly result in the rupture of boiler pipes or expedite the occurrence of such ruptures.

For example, thermal fatigue is prevalent with boilers. Thermal fatigue arises when metal components undergo substantial fluctuations in temperature, particularly during repetitive cycles of heating and cooling. These fluctuations can lead to substantial variations in thermal expansion among the structural elements. Depending on the magnitude of the thermal shock experienced, failure may manifest within a few cycles. This process induces multiaxial stresses on the affected surfaces, giving rise to microcracks along the pipe's surface. Once initiated, these cracks continue to propagate with each subsequent cycle [41–43]. Hence, it is imperative to avert situations that introduce significant temperature fluctuations, such as frequent adjustments to burner settings, inconsistent fuel supply, or excessive on/off cycles.

Overheating is another significant temperature-related cause of boiler failures, which encompasses both short-term and long-term overheating. The short-term overheating problem occurs when pipes experience elevated temperatures and insufficient cooling, often causing the pipe temperature to exceed the eutectoid transformation temperature of the pipe materials. Moreover, the rise in material temperature can induce a significant escalation of stress within the pipe, potentially surpassing the pipe's yield point. The short-term overheating problem can be triggered by factors like water/steam deprivation, flow stagnation due to blockages, uneven flame temperature, etc. [44–46].

Conversely, long-term overheating transpires over an extended duration, as the term implies. Long-term overheating problems have been reported as the most common cause of boiler failures [47–49]. The continuous exposure to elevated temperatures, often surpassing intended or recommended operational thresholds, can result in the deterioration of the pipes' microstructure, characterized by phenomena such as graphitization and spheroidization. Along with the constant stress on the pipes, this gradually leads to a slow, time-dependent deformation (creep) and eventually the rupture of the pipes [50–52]. Long-term overheating may manifest as a result of various underlying factors, including inadequate circulation, scaling, and flame impingement [53,54]. For instance, the presence of scales on the interior surface of pipes can contribute to the occurrence or exacerbation of long-term overheating issues. The low thermal conductivity of scales results in the potential evaporation of water beneath them when exposed to excessive heat. This process of evaporation can progressively elevate the pH of the water to a critical point, thereby

fostering conditions conducive to localized caustic corrosion or embrittlement. Ultimately, this sequence of events culminates in the eventual failure of the boiler [55–57].

4.4. Factors Contributing to DEC's Superior Performance over PCA + K-Means

As can be seen in Sections S3 and S4 of the Supplementary Material, the clustering outcomes achieved by DEC consistently surpass those of PCA + K-means across nearly every dimension. This discrepancy in performance can be attributed to the mechanisms inherent in these two methods. PCA + K-means comprises transforming the original variable space through linear PCA embedding, followed by K-means clustering in the transformed space. Conversely, DEC utilizes an encoder module (DNN) extracted from a pre-trained SAE for embedding, enabling simultaneous, iterative training of DNN and K-means bidirectionally. The DNN forwards the embedded information to K-means clustering, further extending to KL divergence. Reciprocally, DNN and K-means receive feedback from KL divergence for optimization, creating a seamless and iterative process. To be more specific, the DEC system involves iterative refinement of the non-linearly embedded space and cluster centroids based on KL divergence feedback. In contrast, the PCA + K-means approach utilizes linear embedding and lacks iterative feedback optimization. Furthermore, DEC benefits from a wider array of tunable core hyperparameters (such as *nhl* and *nm_hl*), whereas PCA + K-means is limited to *npc* alone.

4.5. Significance of Study and Limitations

As mentioned in the Section 1, this study focuses on an operational-parameter-oriented investigation of failure causes, rather than an exhaustive examination of intricate physical or chemical mechanisms. The objective is not to pinpoint precise physical or chemical reactions leading to failures, but rather to optimize future operational conditions for the purpose of prolonging the boiler's overall service life. Through this approach, operators can address faulty tubes during scheduled annual maintenance without encountering failures and disrupting production. The notable benefit of this study is its accessibility to operators, as the outcomes are straightforward, encompassing solely operational parameters and their recommended values. Informed by the findings, operators can adjust the production process as needed to ensure that the operating parameters remain within a secure range. This failure investigation framework is applicable not only to WtE plants but also potentially to any production line characterized by numerous operating parameters, even when lacking operating-condition labels in the data.

As aforementioned in Section 4.3, elevated temperature-induced mechanisms are the primary causes of failures studied in this case study. However, precisely identifying the specific mechanisms responsible for the failures is beyond the scope of this research, as outlined in the Section 1. Various mechanisms can intertwine and interact, contributing to the occurrence of failures. Given this complexity, extensive additional research and traditional examination approaches are necessary, rather than relying solely on data-driven methods.

The variables identified as culprits in the investigation of failures—specifically, elevated flue gas and steam temperatures—directly triggered cracking, bulging, or bursting. However, various other factors might also have played a role in the eventual failures, including long-term corrosion, physical stress or impact, and oxidation, which are beyond the scope of this study's findings. Monitoring some of these parameters might not be feasible, resulting in unavailable data. Alternatively, for other parameters, relevant factors were monitored, such as the concentration of acidic compounds (SO₂, HCl, etc.) known to contribute to corrosion. The impact of these acidic compounds is usually not immediately significant but rather accumulates gradually over time, implying that data collected within a specific period may not accurately capture their true influence. This deduction has been substantiated by the fact that none of the variables relating to acidic compounds was identified as influential for the failures.

5. Conclusions

A novel and methodical data mining framework was introduced for conducting operational-level (focused on operating parameters) investigations into the attribution of boiler failures. The framework centered on two data mining approaches, PCA + K-means and DEC, with PCA + K-means serving as the baseline against which the performance of DEC was evaluated. To demonstrate the framework's specifics, a case study was performed using datasets obtained from a WtE plant in Sweden. Within the case study, different operational conditions were clustered and identified, followed by the quantification of shifts in variable states between normal and abnormal conditions. Based on this quantification, we pinpointed the variables that played a substantial role in causing failures and recommended their safe operational values to forestall similar incidents in the future. The major findings of the case study are as follows:

- (1) The clustering outcomes of DEC consistently surpass those of PCA + K-means across nearly every dimension. This is attributed to DEC's iterative refinement of the non-linearly embedded space and cluster centroids based on KL divergence feedback.
- (2) T-BSH3rm, T-BSH2l, T-BSH3r, T-BSH1l, T-SbSH3, and T-BSH1r emerged as the most significant contributors to the three failures recorded in the two datasets. This underscores the critical importance of vigilant monitoring and precise temperature control of the superheaters to ensure safe production.
- (3) It is advisable to maintain the operational levels of T-BSH3rm, T-BSH2l, T-BSH3r, T-BSH1l, T-SbSH3, and T-BSH1r around 527 °C, 432 °C, 482 °C, 338 °C, 313 °C, and 343 °C, respectively. Additionally, it is crucial to prevent these values from reaching or exceeding 594 °C, 471 °C, 537 °C, 355 °C, 340 °C, and 359 °C for prolonged durations.

The findings offer the opportunity to improve future operational conditions, thereby extending the overall service life of the boiler. Consequently, operators can address faulty tubes during scheduled annual maintenance without encountering failures and disrupting production. In future research, by examining a broader range of failures, we can develop a repository of diverse influential variables and their recommended operational values. This resource can facilitate more comprehensive, precise, and reliable production operation and management.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pr12071346/s1>, Table S1: List of failure-related operational variables with full name, short name, and unit, of each; Figure S1: DEC structure; Figure S2: Complete clustering results for PCA+K-means for dataset A; Figure S3: Complete clustering results for DEC for dataset A; Figure S4: Complete clustering results for PCA+K-means for dataset B; Figure S5: Complete clustering results for DEC for dataset B.

Author Contributions: Conceptualization, D.W., M.K., E.W. and M.T.; Methodology, D.W.; Validation, D.W., L.J., M.K., E.W. and J.T.; Formal analysis, D.W.; Investigation, D.W., M.K., E.W. and M.T.; Resources, J.T. and M.T.; Data curation, D.W., M.K. and E.W.; Writing—original draft, D.W.; Writing—review & editing, D.W., L.J., J.T. and M.T.; Visualization, D.W.; Supervision, L.J., J.T. and M.T.; Project administration, J.T. and M.T.; Funding acquisition, M.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are not publicly available due to commercial restrictions.

Acknowledgments: The present work was performed as part of the Green Technology and Environmental Economics Research and Collaboration Platform (Green TEE) at Umeå University. Green TEE is a collaborative interface between municipal companies and academic researchers directed toward improving the sustainable performance of cities. The authors acknowledge support from the Green TEE platform for performing this project. The authors would also like to acknowledge support from Umeå Energi, Umeå, Sweden, in organizing study visits and providing the data required in order to perform this study.

Conflicts of Interest: Authors Måns Kjellander and Eva Weidemann were employed by the company Umeå Energi. All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Nomenclature

API	Application Programming Interface
DEC	Deep Embedded Clustering
DM	Data Mining
DNN	Deep Neural Network
DWT	Discrete Wavelet Transform
HF	High-pass Filter
ID	Induced Draft
KL	Kullback–Leibler
LF	Low-pass Filter
n _{hl}	number of hidden layers of the encoder
nn _{el}	number of neurons in the embedded layer of DEC
nn _{hl}	number of neurons in the hidden layer of DEC
npc	number of PCs
PC	Principal Component
PCA	Principal Component Analysis
PMF	Probability Mass Function
SAE	Stacked Autoencoder
WtE	Waste-to-Energy

References

1. Agarwal, S.; Suhane, A. Study of boiler maintenance for enhanced reliability of system A review. *Mater. Today Proc.* **2017**, *4*, 1542–1549. [CrossRef]
2. Barma, M.; Saidur, R.; Rahman, S.; Allouhi, A.; Akash, B.; Sait, S.M. A review on boilers energy use, energy savings, and emissions reductions. *Renew. Sustain. Energy Rev.* **2017**, *79*, 970–983. [CrossRef]
3. Liu, J.; Zhao, J.; Zhu, Q.; Huo, D.; Li, Y.; Li, W. Methanol-based fuel boiler: Design, process, emission, energy consumption, and techno-economic analysis. *Case Stud. Therm. Eng.* **2024**, *54*, 103885. [CrossRef]
4. Elwardany, M. Enhancing Steam Boiler Efficiency through Comprehensive Energy and Exergy Analysis: A Review. *Process Saf. Environ. Prot.* **2024**, *184*, 1222–1250. [CrossRef]
5. Saha, A. Boiler tube failures: Some case studies. In *Handbook of Materials Failure Analysis with Case Studies from the Chemicals, Concrete and Power Industries*; Elsevier: Amsterdam, The Netherlands, 2016; pp. 49–68.
6. Kumar, S.; Kumar, M.; Handa, A. Combating hot corrosion of boiler tubes—A study. *Eng. Fail. Anal.* **2018**, *94*, 379–395. [CrossRef]
7. Shokouhmand, H.; Ghadimi, B.; Espanani, R. Failure analysis and retrofitting of superheater tubes in utility boiler. *Eng. Fail. Anal.* **2015**, *50*, 20–28. [CrossRef]
8. Xue, S.; Guo, R.; Hu, F.; Ding, K.; Liu, L.; Zheng, L.; Yang, T. Analysis of the causes of leakages and preventive strategies of boiler water-wall tubes in a thermal power plant. *Eng. Fail. Anal.* **2020**, *110*, 104381. [CrossRef]
9. Hu, W.; Xue, S.; Gao, H.; He, Q.; Deng, R.; He, S.; Xu, M.; Li, Z. Leakage failure analysis on water wall pipes of an ultra-supercritical boiler. *Eng. Fail. Anal.* **2023**, *154*, 107670. [CrossRef]
10. Baglee, D.; Gorostegui, U.; Jantunen, E.; Sharma, P.; Campos, J. How can SMEs adopt a new method to advanced maintenance strategies? A Case study approach. In Proceedings of the COMADEM 2017 30th International Congress & Exhibition on Condition Monitoring and Diagnostic Engineering Management, Lancashire, UK, 10–13 July 2017.
11. Ichihara, T.; Koike, R.; Watanabe, Y.; Amano, Y.; Machida, M. Hydrogen damage in a power boiler: Correlations between damage distribution and thermal-hydraulic properties. *Eng. Fail. Anal.* **2023**, *146*, 107120. [CrossRef]
12. Haghghat-Shishavan, B.; Firouzi-Nerbin, H.; Nazarian-Samani, M.; Ashtari, P.; Nasirpour, F. Failure analysis of a superheater tube ruptured in a power plant boiler: Main causes and preventive strategies. *Eng. Fail. Anal.* **2019**, *98*, 131–140. [CrossRef]
13. Ding, Q.; Tang, X.-F.; Yang, Z.-G. Failure analysis on abnormal corrosion of economizer tubes in a waste heat boiler. *Eng. Fail. Anal.* **2017**, *73*, 129–138. [CrossRef]
14. Mudgal, D.; Ahuja, L.; Bhatia, D.; Singh, S.; Prakash, S. High temperature corrosion behaviour of superalloys under actual waste incinerator environment. *Eng. Fail. Anal.* **2016**, *63*, 160–171. [CrossRef]
15. Pal, U.; Kishore, K.; Mukhopadhyay, S.; Mukhopadhyay, G.; Bhattacharya, S. Failure analysis of boiler economizer tubes at power house. *Eng. Fail. Anal.* **2019**, *104*, 1203–1210. [CrossRef]
16. Pramanick, A.; Das, G.; Das, S.K.; Ghosh, M. Failure investigation of super heater tubes of coal fired power plant. *Case Stud. Eng. Fail. Anal.* **2017**, *9*, 17–26. [CrossRef]

17. Jones, D. Creep failures of overheated boiler, superheater and reformer tubes. *Eng. Fail. Anal.* **2004**, *11*, 873–893. [CrossRef]
18. Kain, V.; Chandra, K.; Sharma, B. Failure of carbon steel tubes in a fluidized bed combustor. *Eng. Fail. Anal.* **2008**, *15*, 182–187. [CrossRef]
19. Guo, H.; Fan, W.; Liu, Y.; Long, J. Experimental investigation on the high-temperature corrosion of 12Cr1MoVG boiler steel in waste-to-energy plants: Effects of superheater operating temperature and moisture. *Process Saf. Environ. Prot.* **2024**. [CrossRef]
20. Gu, B.; Jiang, S.; Wang, H.; Wang, Z.; Jia, R.; Yang, J.; He, S.; Cheng, R. Characterization, quantification and management of China’s municipal solid waste in spatiotemporal distributions: A review. *Waste Manag.* **2017**, *61*, 67–77. [CrossRef] [PubMed]
21. Tsiliyannis, C.A. Enhanced waste to energy operability under feedstock uncertainty by synergistic flue gas recirculation and heat recuperation. *Renew. Sustain. Energy Rev.* **2015**, *50*, 1320–1337. [CrossRef]
22. Paz, M.; Zhao, D.; Karlsson, S.; Liske, J.; Jonsson, T. Investigating corrosion memory: The influence of previous boiler operation on current corrosion rate. *Fuel Process. Technol.* **2017**, *156*, 348–356. [CrossRef]
23. Sohaib, M.; Kim, J.-M. Data driven leakage detection and classification of a boiler tube. *Appl. Sci.* **2019**, *9*, 2450. [CrossRef]
24. Jia, X.; Sang, Y.; Li, Y.; Du, W.; Zhang, G. Short-term forecasting for supercharged boiler safety performance based on advanced data-driven modelling framework. *Energy* **2022**, *239*, 122449. [CrossRef]
25. Cui, Z.; Xu, J.; Liu, W.; Zhao, G.; Ma, S. Data-driven modeling-based digital twin of supercritical coal-fired boiler for metal temperature anomaly detection. *Energy* **2023**, *278*, 127959. [CrossRef]
26. Khalid, S.; Hwang, H.; Kim, H.S. Real-world data-driven machine-learning-based optimal sensor selection approach for equipment fault detection in a thermal power plant. *Mathematics* **2021**, *9*, 2814. [CrossRef]
27. Qin, H.; Yin, S.; Gao, T.; Luo, H. A data-driven fault prediction integrated design scheme based on ensemble learning for thermal boiler process. In Proceedings of the 2020 IEEE International Conference on Industrial Technology (ICIT), Buenos Aires, Argentina, 26–28 February 2020; pp. 639–644.
28. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [CrossRef]
29. Verleysen, M.; François, D. The curse of dimensionality in data mining and time series prediction. In Proceedings of the International work-conference on artificial neural networks, Barcelona, Spain, 8–10 June 2005; pp. 758–770.
30. Burrus, S.; Burrus, C.S.; Gopinath, R.A.; Guo, H.; Odegard, J.A.N.E.A.; Selesnick, I.W.A. *Introduction to Wavelets and Wavelet Transforms: A Primer*; Prentice Hall: Hoboken, NJ, USA, 1998.
31. Lee, G.R.; Gommers, R.; Waselewski, F.; Wohlfahrt, K.; O’Leary, A. PyWavelets: A Python package for wavelet analysis. *J. Open Source Softw.* **2019**, *4*, 1237. [CrossRef]
32. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]
33. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
34. Xie, J.; Girshick, R.; Farhadi, A. Unsupervised deep embedding for clustering analysis. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 20–22 June 2016; pp. 478–487.
35. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [CrossRef]
36. Le, Q.V. Building high-level features using large scale unsupervised learning. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 8595–8598.
37. Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.-A. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.
38. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 807–814.
39. Guo, X. DEC.py. Available online: <https://github.com/XifengGuo/DEC-keras/blob/master/DEC.py> (accessed on 25 February 2020).
40. Gencoglu, O.; van Gils, M.; Guldogan, E.; Morikawa, C.; Sützen, M.; Gruber, M.; Leinonen, J.; Huttunen, H. HARK Side of Deep Learning—From Grad Student Descent to Automated Machine Learning. *arXiv* **2019**, arXiv:1904.07633.
41. Magnus, C.; Pardeshi, A. Investigation into the failure of a superheater tube in a power generation plant utilizing waste material combustion in a furnace. *Eng. Fail. Anal.* **2024**, *156*, 107838. [CrossRef]
42. Ahmad, J.; Purbolaksono, J.; Beng, L. Thermal fatigue and corrosion fatigue in heat recovery area wall side tubes. *Eng. Fail. Anal.* **2010**, *17*, 334–343. [CrossRef]
43. Lee, N.-H.; Kim, S.; Choe, B.-H.; Yoon, K.-B.; Kwon, D.-i. Failure analysis of a boiler tube in USC coal power plant. *Eng. Fail. Anal.* **2009**, *16*, 2031–2035. [CrossRef]
44. Ahmad, J.; Rahman, M.M.; Zuhairi, M.; Ramesh, S.; Hassan, M.; Purbolaksono, J. High operating steam pressure and localized overheating of a primary superheater tube. *Eng. Fail. Anal.* **2012**, *26*, 344–348. [CrossRef]
45. Hosseini, R.K.; Yareiee, S. Failure analysis of boiler tube at a petrochemical plant. *Eng. Fail. Anal.* **2019**, *106*, 104146. [CrossRef]
46. Munda, P.; Husain, M.M.; Rajinikanth, V.; Metya, A. Evolution of microstructure during short-term overheating failure of a boiler water wall tube made of carbon steel. *J. Fail. Anal. Prev.* **2018**, *18*, 199–211. [CrossRef]
47. Deshmukh, S.; Dhamangaonkar, P. A review paper on factors that causes the bulging failure of the metal tube. *Mater. Today Proc.* **2022**, *62*, 7610–7617. [CrossRef]
48. Lobleby, G.R.; Al-Otaibi, W.L. Diagnosing boiler tube failures related to overheating. *Adv. Mater. Res.* **2008**, *41*, 175–181. [CrossRef]

49. Rahman, M.; Purbolaksono, J.; Ahmad, J. Root cause failure analysis of a division wall superheater tube of a coal-fired power station. *Eng. Fail. Anal.* **2010**, *17*, 1490–1494. [CrossRef]
50. Hayazi, N.F.; Shamsudin, S.R.; Wardan, R.; Sanusi, M.S.M.; Zainal, F.F. Graphitization damage on seamless steel tube of pressurized closed-loop of steam boiler. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *701*, 012042. [CrossRef]
51. Nutal, N.; Gommès, C.J.; Blacher, S.; Pouteau, P.; Pirard, J.-P.; Boschini, F.; Traina, K.; Cloots, R. Image analysis of pearlite spheroidization based on the morphological characterization of cementite particles. *Image Anal. Stereol.* **2010**, *29*, 91–98. [CrossRef]
52. Pérez, I.U.; Da Silveira, T.L.; Da Silveira, T.F.; Furtado, H.C. Graphitization in low alloy steel pressure vessels and piping. *J. Fail. Anal. Prev.* **2011**, *11*, 3–9. [CrossRef]
53. da Silveira, R.M.S.; Guimarães, A.V.; Oliveira, G.; dos Santos Queiroz, F.; Guzela, L.R.; Cardoso, B.R.; Araujo, L.S.; de Almeida, L.H. Failure of an ASTM A213 T12 steel tube of a circulating fluidized bed boiler. *Eng. Fail. Anal.* **2023**, *148*, 107188. [CrossRef]
54. McIntyre, K.B. A review of the common causes of boiler failure in the sugar industry. *Proc. S. Afr. Sug. Technol. Ass.* **2002**, *75*, 355–364.
55. Dooley, R.B.; Bursik, A. Hydrogen damage. *PowerPlant Chem.* **2010**, *12*, 122.
56. Dooley, R.B.; Bursik, A. Caustic gouging. *PowerPlant Chem.* **2010**, *12*, 188–192.
57. Kim, Y.-S.; Kim, W.-C.; Kim, J.-G. Bulging rupture and caustic corrosion of a boiler tube in a thermal power plant. *Eng. Fail. Anal.* **2019**, *104*, 560–567. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Review

Toward Enhanced Efficiency: Soft Sensing and Intelligent Modeling in Industrial Electrical Systems

Paul Arévalo ^{1,2,*} and Danny Ochoa-Correa ¹

¹ Department of Electrical Engineering, Electronics and Telecommunications (DEET), University of Cuenca, Balzay Campus, Cuenca 010107, Ecuador; danny.ochoac@ucuenca.edu.ec

² Department of Electrical Engineering, University of Jaén, 23700 Linares, Spain

* Correspondence: warevalo@ujaen.es

Abstract: This review article focuses on applying operation state detection and performance optimization techniques in industrial electrical systems. A comprehensive literature review was conducted using the preferred reporting items for systematic reviews and meta-analyses (PRISMA) methodology to ensure a rigorous and transparent selection of high-quality studies. The review examines in detail how soft sensing technologies, such as state estimation and Kalman filtering, along with hybrid intelligent modeling techniques, are being used to enhance efficiency and reliability in the electrical industry. Specific case studies are analyzed in areas such as electrical network monitoring, fault detection in high-voltage equipment, and energy consumption optimization in industrial plants. The PRISMA methodology facilitated the identification and synthesis of the most relevant studies, providing a robust foundation for this review. Additionally, the article explores the challenges and research opportunities in applying these techniques in specific industrial contexts, such as steel metallurgy and chemical engineering. By incorporating findings from meticulously selected studies, this work offers a detailed, engineering-oriented insight into how advanced technologies are transforming industrial processes to achieve greater efficiency and operational safety.

Keywords: intelligent modeling; soft sensing; industrial electrical systems; state estimation; machine-learning

Citation: Arévalo, P.; Ochoa-Correa, D. Toward Enhanced Efficiency: Soft Sensing and Intelligent Modeling in Industrial Electrical Systems. *Processes* **2024**, *12*, 1365. <https://doi.org/10.3390/pr12071365>

Academic Editors: Sheng Du, Xiongbo Wan, Zixin Huang and Li Jin

Received: 2 June 2024

Revised: 28 June 2024

Accepted: 28 June 2024

Published: 30 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The electrical industry is continuously evolving, driven by technological advancements to enhance the efficiency and reliability of industrial electrical systems. In this context, soft sensing and intelligent modeling have emerged as promising research areas. Soft sensing involves using virtual sensors and advanced algorithms to estimate electrical systems' state and operating conditions [1–3]. Meanwhile, intelligent modeling entails the development of mathematical models and optimization techniques to improve the performance of these systems [2]. The motivation behind this work stems from the growing demand for more efficient and reliable electrical systems in the industry. Soft sensing and intelligent modeling offer an innovative approach to addressing these challenges by enabling better equipment condition monitoring, early fault detection, and operational performance optimization.

Moreover, with the rise of industry 4.0 and the digitization of industrial processes, there is an urgent need to adopt advanced technologies that enhance the efficiency and competitiveness of companies. By understanding the context and importance of soft sensing and intelligent modeling in the electrical industry, this study aims to analyze how these technologies can improve industrial electrical systems' efficiency and reliability. Through a comprehensive analysis of case studies and current trends, we seek to provide a detailed, engineering-oriented perspective on how these innovations transform industrial processes and pave the way for greater operational efficiency and safety.

The electrical industry is undergoing a revolution driven by soft sensing, incorporating virtual sensors and advanced algorithms. For instance, in the biomedical field, algorithms

such as nonlinear weighted total variation image reconstruction enhance the accuracy of electrical capacitance tomography [1,2]. Additionally, dynamic latent structures with time-varying parameters are being used to predict hard-to-measure variables in virtual sensing applications [3]. Soft tissue-based sensors are also being effectively employed in practical applications, such as assisted gloves [3]. In state estimation, various approaches are being applied to improve the efficiency and reliability of electrical systems. For example, the combination of electromagnetic flow tomography and electrical tomography through Bayesian estimation enables precise image reconstruction [4–6].

Furthermore, advanced models are being developed to estimate the state of charge of energy storage systems like supercapacitors [7] and the state of health (SOH) of lithium-ion batteries [8]. Operational state detection is crucial for the efficiency and safety of electrical systems. Methods such as the real-time monitoring of current consumption in industrial electrical equipment allow for proactive machine management and an early detection of deviations and faults [9–11]. Other innovative approaches include impedance measurement in superconducting circuits during the operation and remote detection of defects in insulating materials using optical thermography [12–14].

In the realm of industrial electrical system efficiency, performance optimization techniques play a crucial role. For example, electrical impedance tomography is used in medical and industrial applications to estimate the internal electrical properties of biological tissues through voltage and current measurements on their surface [15]. In the field of solar energy, artificial neural networks (ANN) are applied for maximum power point tracking and fault detection in partially shaded photovoltaic systems [16]. Electrical capacitance tomography visualizes multiphase systems in industrial applications [17]. For the control of brushless direct current motors, a system based on an adaptive neuro-fuzzy inference system (ANFIS) optimized by the bacterial foraging optimization algorithm has been developed [18]. In rotary ultrasonic machining, a dual compensation approach for frequency and impedance using a multi-objective genetic algorithm has significantly improved transmission efficiency and ultrasonic vibration stability [19]. In thermoelectric generation systems with non-uniform temperature distribution, an optimal control technique based on an equilibrium optimization algorithm has been proposed [20]. For frequency and voltage stabilization in hybrid energy systems, a second-order active disturbance rejection control strategy optimized with ANN has been developed [21]. In cogeneration plants integrated with photovoltaic energy, a modified firefly algorithm combined with machine learning techniques has been used to predict plant efficiency and optimize its operation [22].

In the field of industrial electrical systems, efficiency and reliability are critical for optimizing performance and reducing operational costs. Intelligent techniques like state estimation and Kalman filtering are pivotal in fault detection and condition monitoring [23]. Condition monitoring and fault detection and diagnosis are essential for preventing severe damage to rotating machinery, such as induction motors [24]. Electrical machines, which are widely used in industry, also require advanced techniques for condition monitoring and predictive fault diagnosis [25]. In the context of industry 4.0, disruptive technologies like the internet of things (IoT) and artificial intelligence (AI) are revolutionizing communication and control in smart industries [26]. Battery management in energy storage and electric propulsion applications is another area where efficiency and safety are paramount [27]. Battery management systems incorporate functional safety techniques to ensure battery cells' safe and efficient operation [28]. Estimating batteries' SOH is crucial for ensuring their safe operation and optimizing their lifecycle [29]. Despite significant advances in soft sensing and intelligent modeling in industrial electrical systems, several research gaps still need detailed attention. For instance, while condition monitoring and fault diagnosis have been active areas of interest, there remains a need for more robust and accurate techniques for early anomaly detection and real-time fault prediction.

Furthermore, although the application of disruptive technologies like IoT and AI has shown promise, further research is necessary to fully understand their effective integration into industrial environments and their impact on the efficiency and reliability of electrical

systems. Another area requiring greater attention is battery management in energy storage applications, where optimizing the SOH of batteries continues to be a significant challenge. In summary, while considerable progress has been made, important research gaps must be addressed to advance efficiency and operational safety in the electrical industry.

This paper proposes a comprehensive literature review on using soft sensing and intelligent modeling in industrial electrical systems. The aim is to synthesize the latest advancements in this field, critically analyze the selected studies, identify emerging trends, and explore practical applications. The PRISMA method will be employed to ensure the rigor and transparency of the review, facilitating the effective identification of relevant studies and a coherent synthesis of their findings. This systematic and transparent approach will allow us to evaluate the quality of the research, identify potential knowledge gaps, and highlight areas needing further investigation. Consequently, this review will provide a detailed, engineering-oriented perspective on how these technologies transform industrial processes, leading to greater efficiency and operational safety.

The remainder of this paper is organized as follows: Section 2 presents a literature review and methodology, outlining the theoretical framework and research methods used. Section 3 provides a descriptive analysis of the literature, summarizing key findings from the reviewed studies. Section 4 details the results of our analysis. Section 5 discusses the implications of these findings, and finally, Section 6 concludes the paper, highlighting the main contributions and suggesting avenues for future research.

2. Literature Review Methodology

2.1. Study Selection Criteria

The bibliographic resources for this literature review were sourced from three prestigious databases relevant to the research area. Scopus, IEEE Xplore, and MDPI. These databases were chosen due to their extensive coverage of high-quality research articles in electrical and industrial systems, mainly focusing on innovative technologies such as soft sensing and intelligent modeling. The search terms were carefully selected to capture the most relevant literature. These terms include “soft sensing”, “intelligent modeling”, “industrial”, and “electrical”, among others. Combining these terms ensures that the search encompasses a broad yet targeted spectrum of research articles that align with the objectives of this review.

The years 2019–2024 were selected as the most appropriate period for mapping knowledge in this study’s thematic area. This period marks a significant phase of technological advancements and increased research activity in soft sensing and intelligent modeling. This timeframe captures the latest developments and emerging trends, providing a comprehensive overview of the current state of research and its practical applications in industrial electrical systems. Table 1 summarizes the search terms and queries defined for the literature review.

Table 1. Search terms and queries utilized for the literature review.

Database	Search Terms	Query String
Scopus	Journal articles published between 2019 and 2024, written in English, that include either “Industrial” or “Electrical” along with either “Soft Sensing” or “Intelligent Modeling” in the title, abstract, or keywords.	TITLE-ABS-KEY (“Industrial” OR “Electrical”) AND (“Soft Sensing” OR “Intelligent Modeling”) AND PUBYEAR > 2018 AND PUBYEAR < 2025 AND (LIMIT-TO (DOCTYPE, “ar”)) AND (LIMIT-TO (LANGUAGE, “English”))
IEEE Xplore	Journal articles published between 2019 and 2024, written in English, that include either “Industrial” or “Electrical” along with both “Soft Sensing” and “Intelligent Modeling” in all metadata.	(“All Metadata”:Industrial) OR (“All Metadata”:Electrical) AND (“All Metadata”:Soft Sensing) AND (“All Metadata”:Intelligent Modeling)
MDPI	Journal articles published between 2019 and 2024, written in English, that include either “Industrial” or “Electrical” along with both “Soft Sensing” and “Intelligent Modeling”.	Industrial OR Electrical AND Soft Sensing AND Intelligent Modeling

The search terms were derived from this study's preliminary background and introduction, highlighting the electrical industry's continuous evolution and the emergence of promising research areas such as soft sensing and intelligent modeling. These technologies leverage virtual sensors and advanced algorithms to estimate electrical systems' state and operational conditions, thus enhancing monitoring, fault detection, and performance optimization. The growing demand for more efficient and reliable electrical systems in the industry, coupled with the advancement of industry 4.0 and digitalization, underscores the necessity of adopting these advanced technologies.

2.2. Search Process and Selection of Studies

The PRISMA approach ensures a rigorous and transparent review process, enabling the identification and synthesis of relevant studies while minimizing bias [30]. This methodology is well-suited for our research as it provides a structured framework for selecting high-quality studies, which is critical for understanding the advancements in soft sensing and intelligent modeling within industrial electrical systems. Figure 1 illustrates the flowchart of the literature review process. As depicted in the figure, the review begins by applying the search terms and queries in Table 1. This initial search yielded 279 items: 207 from Scopus, 62 from IEEE Xplore, and 10 from MDPI. With these raw results, the authors have assigned a coding system to the items to facilitate subsequent bibliometric processing. Articles from Scopus are coded as S-XX, those from IEEE Xplore as IEEE-XX, and those from MDPI as MDPI-XX.

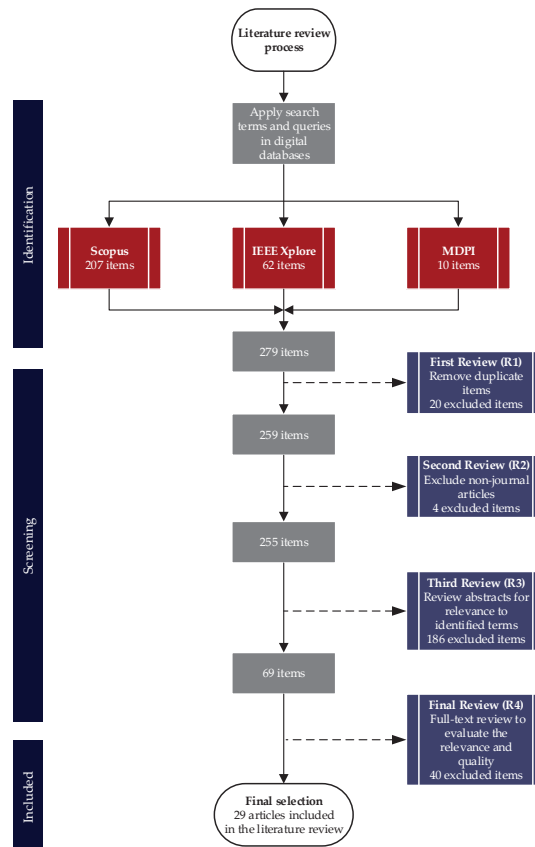


Figure 1. Flowchart of the literature review process.

The first review stage (R1) involves the removal of duplicate items. During this stage, 20 duplicate documents were identified and withdrawn. Next, the second review stage (R2) excludes items that are not journal articles, such as review papers, conference papers, book chapters, and books. While these documents are undoubtedly valuable and contribute significantly to the field, primary research articles published in peer-reviewed journals provide more direct and less biased information, essential for constructing a robust and unbiased literature review. This stage revealed that four items had bypassed the initial database filters and were subsequently rejected. The critical mass comprises 255 articles, which will undergo a preliminary bibliometric analysis.

Figure 2 illustrates the distribution of the preselected works across the different digital databases, revealing a predominance in Scopus, which accounts for 80.39% of the total. IEEE Xplore follows this with 15.69%, and MDPI with 3.92%. This distribution was expected, as Scopus encompasses a broader range of scientific publishers and journals than the other two databases under consideration. Scopus is known for its extensive indexing of diverse and high-impact journals, naturally resulting in more relevant articles being retrieved from this database.

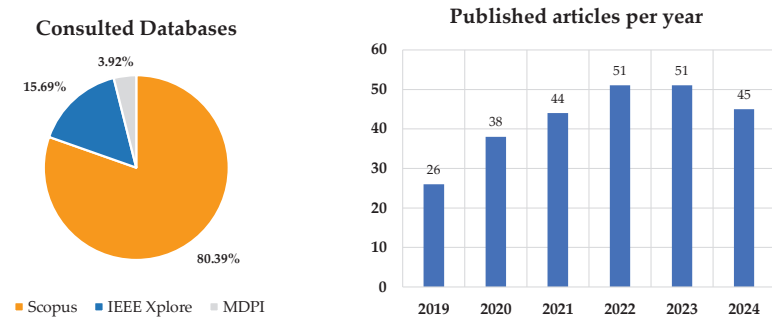


Figure 2. Distribution of articles by digital database and year of publication.

The smaller proportion of MDPI's works highlights the need to generate more scientific material related to the topic within this publishing house. This represents a niche the present work aims to address, contributing to expanding research in soft sensing and intelligent modeling within the MDPI database. Figure 2 also depicts the publication trend over the past five years, showing a clear and sustained increase in publications. The statistics indicate a notably high number of publications in 2024 despite the year not being halfway through. This trend suggests a continuous and growing scientific interest in the topic addressed by this research. The upward trajectory in publication numbers over the past five years indicates the increasing relevance and importance of research in soft sensing and intelligent modeling in industrial electrical systems. This sustained growth trend underscores the ongoing advancements and the expanding body of knowledge in this field, reflecting a robust and dynamic area of scientific inquiry that this literature review aims to cover comprehensively.

Moreover, Figure 3 shows the distribution of articles according to the journals hosted in the consulted databases. This distribution provides insights into which journals are most actively publishing research on soft sensing and intelligent modeling, helping to identify key publications and potential outlets for future research in this area.

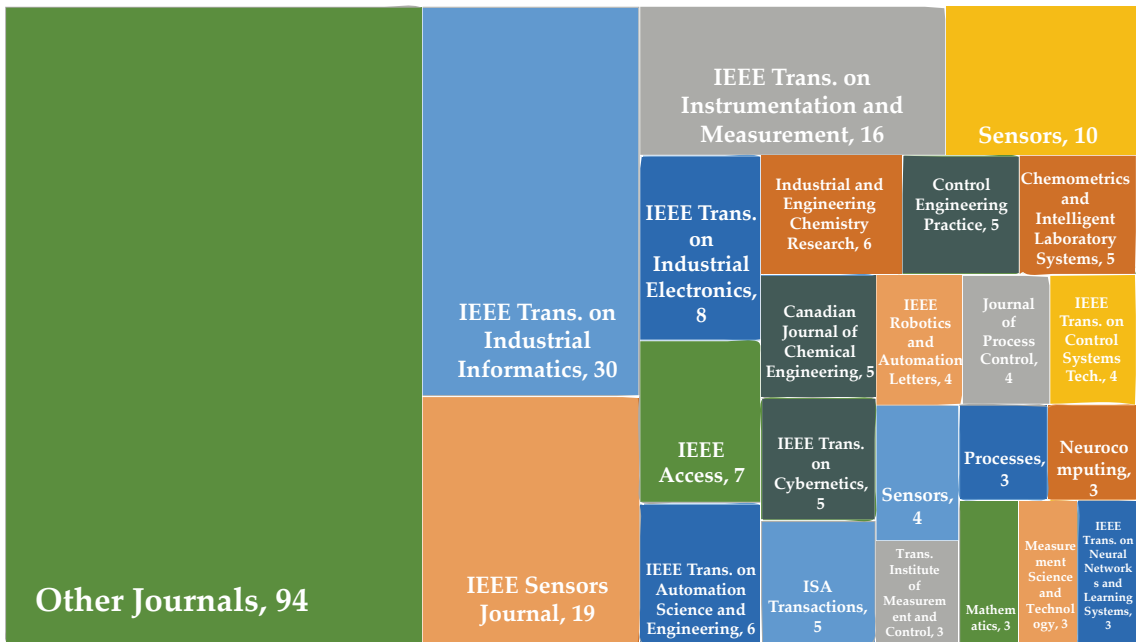


Figure 3. Distribution of articles by journal.

Now, with the metadata of these 255 articles, the keywords from each article are extracted to cluster them and generate a word cloud map. This action helps to identify the frequency of terms used. Based on the most prevalent keywords in Figure 4, the following combinations of terms can be defined: soft sensing, state estimation, Kalman filtering, intelligent modeling, industrial electrical systems, fault detection, energy optimization, condition monitoring, predictive maintenance, machine learning in electrical systems, IoT in industry, electrical network monitoring, high-voltage equipment, battery health estimation, and energy consumption optimization.

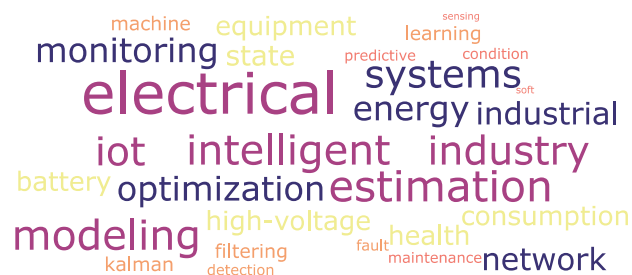


Figure 4. Word cloud map of the keywords in the preselected articles.

In the following review step (R3), each item's abstract was thoroughly reviewed for its relevance to the terms identified in the word cloud map. This exhaustive process resulted in the withdrawal of 186 articles, leaving only 69 whose abstracts are closely aligned with the identified terms to proceed to the next stage. Then, in the review step (R4), a full-text review of each work was conducted to determine the relevance of the topics concerning the identified terms and the main focus of this research. For this purpose, the research team defined a series of criteria to evaluate each of the 69 items using a five-level Likert scale

for the evaluation metrics. The evaluation criteria and metrics devised for the conducted literature review are provided in Table 2.

Table 2. Evaluation criteria and metrics for full-text review.

Criterion	Description	Evaluation Metrics
Relevance to Research	How closely does the paper align with the review's focus on soft sensing, intelligent modeling, etc.	1 = Not relevant, 2 = Slightly relevant, 3 = Moderately relevant, 4 = Highly relevant, 5 = Essential
Quality of Research	The rigor and reliability of the research methodology used.	1 = Poor, 2 = Fair, 3 = Good, 4 = Very Good, 5 = Excellent
Innovation and Originality	The originality and contribution of the paper to the field.	1 = Not original, 2 = Slightly original, 3 = Moderately original, 4 = Very original, 5 = Groundbreaking
Clarity and Presentation	The clarity of writing, structure, and presentation of the paper.	1 = Poor, 2 = Fair, 3 = Good, 4 = Very Good, 5 = Excellent
Depth of Analysis	The depth and thoroughness of the analysis provided in the paper.	1 = Superficial, 2 = Basic, 3 = Adequate, 4 = In-depth, 5 = Comprehensive
Applicability	The practical applicability of the research findings to industrial electrical systems.	1 = Not applicable, 2 = Slightly applicable, 3 = Moderately applicable, 4 = Very applicable, 5 = Highly applicable
References and Citations	The number and quality of references and citations used in the paper.	1 = Poor, 2 = Fair, 3 = Good, 4 = Very Good, 5 = Excellent
Impact and Influence	The impact and influence of the paper within the research community.	1 = Low, 2 = Fair, 3 = Good, 4 = Very Good, 5 = High
Scope and Coverage	The scope of the paper and the extent to which it covers relevant topics.	1 = Narrow, 2 = Limited, 3 = Adequate, 4 = Broad, 5 = Comprehensive
Case Studies and Examples	The paper provides the inclusion and quality of case studies or practical examples.	1 = None, 2 = Few/poor quality, 3 = Adequate, 4 = Several/good quality, 5 = Many/high quality
Technical Accuracy	The accuracy and reliability of the technical content presented.	1 = Poor, 2 = Fair, 3 = Good, 4 = Very Good, 5 = Excellent
Future Research Directions	The paper discusses future research directions and potential advancements.	1 = None, 2 = Limited, 3 = Adequate, 4 = Extensive, 5 = Comprehensive

A minimum threshold score of 36 out of 60 (60%) has been established to ensure the inclusion of relevant and high-quality studies in this literature review. This threshold ensures that selected articles closely align with the review's focus on soft sensing, intelligent modeling, and related topics in industrial electrical systems. Each article is evaluated on criteria such as relevance, the quality of research, innovation, clarity, the depth of analysis, applicability, references, author expertise, impact, scope, case studies, technical accuracy, and future research directions. Setting the threshold at 60% guarantees that the included studies meet a sufficient methodological rigor and practical relevance standard while allowing for a comprehensive and inclusive review. This balanced approach ensures that the literature review incorporates valuable contributions without being overly restrictive, thus providing a robust and insightful analysis of the current state of research in the field.

Figure 5 shows the final scores achieved by each item at this stage. Based on the results, 29 articles meet the predefined minimum threshold; therefore, the remaining articles have been discarded.



Figure 5. Final scores achieved by each item at the final stage of the review process.

Appendix A, Table A1, provides a comprehensive summary of the articles that successfully passed all stages of the literature review process and are the subject of analysis in the following section.

3. Descriptive Analysis of the Literature

Following the systematic selection process in the literature review, 29 items were identified as highly relevant to the research focus. These items were selected based on a thorough full-text review, ensuring their alignment with this study’s core themes and objectives. The comprehensive evaluation of these works is summarized in Table A1, which provides detailed information about each selected article, including their titles and core proposals.

To gain a deeper understanding of the current state of research and to facilitate a structured analysis, the selected literature was subsequently clustered into four main topics: intelligent modeling and optimization (IMO), soft sensing techniques, machine learning and neural networks, and process monitoring and optimization. This categorization was

derived from carefully examining the articles' titles and abstracts, identifying each work's primary focus and contributions.

- **Intelligent Modeling and Optimization** encompasses articles focusing on advanced modeling techniques and optimization strategies to enhance industrial processes' efficiency and effectiveness. This cluster includes works that utilize innovative approaches such as Bayesian learning, just-in-time learning (JITL), and intelligent optimization frameworks to address various challenges in industrial systems.
- **Soft Sensing Techniques** includes research that develops and applies soft sensing methods for industrial processes. These techniques involve indirect measurements and data-driven models to estimate process variables that are difficult or impossible to measure directly. The articles in this cluster explore various statistical and machine-learning methods to improve the accuracy and reliability of soft sensors.
- **Machine Learning and Neural Networks** covers studies that leverage machine learning algorithms and neural network models to solve complex problems in industrial settings. This cluster includes works implementing deep learning, auto-encoders, and other neural network-based methods to enhance predictive accuracy and process control.
- **Process Monitoring and Optimization** focuses on articles that aim to monitor and optimize industrial processes through data-driven and model-based approaches. This cluster includes research on state estimation, process monitoring, and dynamic optimization, utilizing techniques such as variational Bayesian learning and dynamic latent variable models.

The distribution of selected items across these four clusters is depicted in the Venn diagram in Figure 6. This figure illustrates the categorization of the selected literature, visually representing the research landscape and highlighting the areas of concentration within the field. By organizing the literature into these distinct topics, we can better understand the current trends and gaps in research, guiding future investigations and innovations in soft sensing and intelligent modeling in industrial electrical systems.

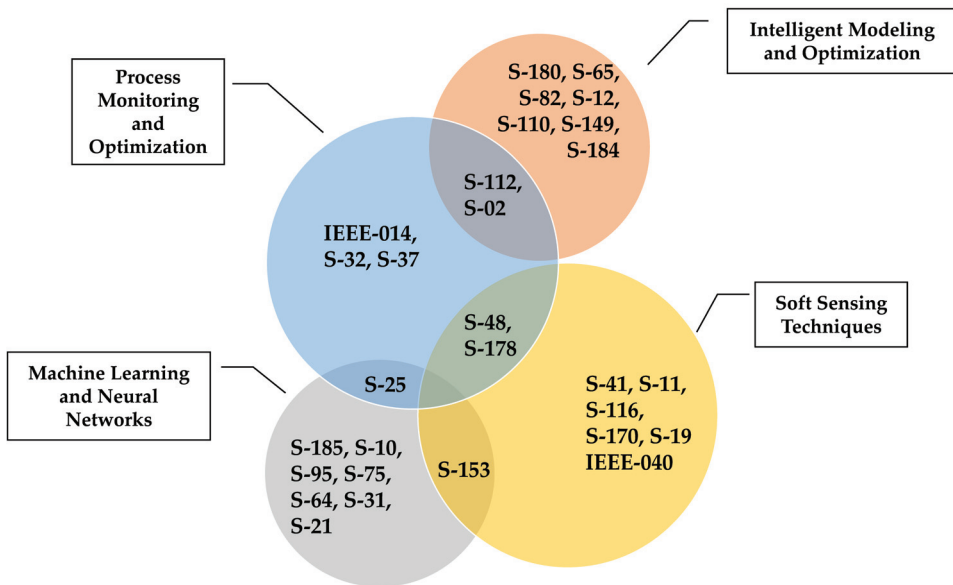


Figure 6. Venn diagram showing the clustering of selected articles into four main topics.

4. Results

After applying the PRISMA method, four main themes have been identified, representing significant research areas in IMO. These themes encompass a variety of approaches and applications within the field, including advanced modeling techniques, soft sensing, machine learning, and process optimization. In the following sections, we will detail each of these themes, examining key contributions from the literature and discussing the challenges and future trends in each area.

4.1. Intelligent Modeling and Optimization

4.1.1. Definition

IMO is an advanced approach that employs techniques in industrial electrical systems to enhance processes' efficiency, productivity, and reliability [31]. It utilizes intelligent algorithms and machine learning methods to analyze complex data and make optimal real-time decisions. This approach finds application across various industrial sectors, including energy, manufacturing, chemistry, and automation, aiming to optimize the operation of electrical networks, improve product quality, and design safer and more efficient production processes [32,33]. Despite advancements, challenges persist regarding model interpretation and integration into real-time systems. However, the evolution of technologies such as cloud computing and the IoT is expected to continue driving the development and application of IMO in the future [33,34].

4.1.2. Applications

IMO has experienced significant growth in recent decades, with a wide range of applications across various industrial sectors. For instance, in the metallurgical sector, the study presented in article [31] proposes a semi-supervised online soft sensor model to predict silicon content in industrial blast furnaces. This approach leverages semi-supervised learning to absorb valuable information from unlabeled data, resulting in improved predictive performance compared to traditional soft sensors. Moreover, in the industrial process sector, article [35] introduces a β -variational autoencoder regression model for observing and measuring multimode industrial processes. This approach has been successfully applied in predicting variable quality for multimode industrial processes, demonstrating its efficacy and superiority in soft sensor modeling in complex industrial environments. In the field of chemical engineering, the study presented in [34] proposes an efficient (JITL) framework for soft sensing in industrial processes. This approach has been applied in ebullated bed hydrogenation and physical separation units, demonstrating higher predictive accuracy and operational efficiency compared to other methods. These current applications illustrate the versatility and effectiveness of IMO techniques in a variety of industrial contexts. From metallurgy to chemical engineering, IMO has shown its ability to enhance processes, optimize performance, and reduce operating costs across diverse industries.

4.1.3. Current Challenges

Despite significant advances achieved in the field of IMO, several important challenges persist that require attention and innovative solutions. For example, in the study on the soft sensing of liquefied petroleum gas (LPG) processes using deep learning [33], the challenge of the interpretability of AI and deep learning-based models is highlighted. While these models can offer high levels of predictive accuracy, their opacity regarding how they arrive at those predictions can be a barrier to adoption in industrial environments where a clear understanding of the model's decision-making process is required. Additionally, in the IMO for a smart energy hub [31] article, the need to address the computational complexity associated with optimizing energy hub models is noted. As energy systems become more interconnected and diversified, there is a need to develop more efficient optimization methods to handle the vast data and complexity of modern energy systems.

Another significant challenge is highlighted in the study on Bayesian (JITL) [36], where the efficient selection of relevant samples and accurate base model construction are

addressed. While JITL techniques offer the advantage of updating localized models in real-time, identifying relevant samples and constructing accurate base models remain areas of active research and development.

4.1.4. Future Trends

Future trends in the field of IMO point towards greater integration of advanced technologies and innovative approaches to address emerging challenges and leverage new opportunities in a variety of industrial applications. A significant trend is the increasing use of machine learning techniques and real-time predictive modeling in industrial environments, as highlighted in the study on the soft sensing of LPG processes using deep learning [33,37]. This trend is expected to continue, with more sophisticated AI models enabling more precise and efficient real-time decision-making across a variety of industrial processes.

Additionally, the optimization of energy systems, as discussed in the article on IMO for a smart energy hub [31], is expected to become increasingly important as companies seek to maximize the efficiency and profitability of their energy operations. This could involve the development of more advanced optimization algorithms and the implementation of more sophisticated automation technologies to manage and control complex energy systems. Another significant future trend is using Bayesian approaches and JITL methods in industrial applications, as mentioned in [36]. These approaches are expected to continue evolving and improving, focusing more on an efficient selection of relevant samples, accurate base model construction, and uncertainty management in industrial processes.

4.2. Soft Sensing Techniques

4.2.1. Definition

Soft sensing techniques use computational models to estimate difficult-to-measure variables in industrial processes, leveraging statistical methods and machine learning algorithms [38,39]. These methods predict quality parameters, enhancing process monitoring without expensive direct measurement instruments. By utilizing labeled and unlabeled data, they aim for accurate real-time predictions, which are crucial for operational efficiency and product quality [1]. For instance, the principal component-based semi-supervised extreme learning machine (PCSELM) model combines labeled and unlabeled data to enhance model accuracy, while a neural network approach optimizes input layers in dynamic batch processes, improving predictive performance [40,41].

4.2.2. Applications

Soft sensing techniques are crucial across various industries, enhancing process monitoring and product quality prediction. In metallurgy, they predict silicon content in blast furnaces, a hard-to-measure variable. The PCSELM model in [38] showcases superior predictive accuracy through semi-supervised learning, leveraging labeled and unlabeled data. Dynamic batch processes like penicillin fermentation and injection molding benefit from a neural network-based approach outlined in [40], optimizing information relevance layer by layer and outperforming existing methods.

JITL combines mutual information and partial least squares to improve soft sensor accuracy [42]. This method addresses non-linearity and collinearity, yielding more precise similarity measures and predictions. Additionally, a two-step statistical learning approach for batch process soft sensing [43] reduces data dimensionality. It constructs reliable soft sensors, enhancing accuracy in industrial applications such as personal care product manufacturing.

4.2.3. Current Challenges

Despite advancements in soft sensing techniques, challenges persist. Interpreting complex models, especially those based on deep learning, poses a significant hurdle due to their opaque nature, hindering trust and acceptance in industrial settings [38]. Compu-

tational complexity is another issue, as real-time data integration and processing, such as in JITL, demand efficient measures and strain resources [42]. Integrating domain-specific knowledge into data-driven models remains challenging despite promising results from methods like neural network-based representation learning [40]. Handling variability and complexity in industrial processes, such as non-linearity and multimodality, present further difficulties, exemplified by the challenge of managing multiple modes without prior information [44].

4.2.4. Future Trends

Future trends in soft sensing techniques involve integrating advanced machine learning and statistical methods to tackle current challenges and exploit new opportunities in industrial applications. One key trend is the growing adoption of sophisticated machine learning algorithms like deep learning and JITL to improve real-time predictive accuracy and decision-making. For instance, models like PCSELM demonstrate the potential for incorporating more advanced semi-supervised and unsupervised learning techniques to enhance model robustness and flexibility [38]. These models also focus on enhancing interpretability and transparency to facilitate industrial adoption. Techniques like the weighted autoregressive dynamic latent variable model aim to make predictions more understandable and trustworthy [45]. For most modern industrial processes with strong nonlinear and multimodal characteristics, the traditional linear PLS-based soft sensor may not work well. Meanwhile, the traditional global modeling approach has a high demand for data representation capability in the face of complex data distribution, which poses a challenge to soft sensing [46]. In addition, the unbalanced nature of data distribution exacerbates the model's neglect of local information to some extent, which enhances the overall prediction difficulty of the model. To this end, based on the PLS, a novel quality-relevant feature clustering (QRFC) model is proposed for the first time in this article from the view of the local modeling of probabilistic fusion. In the QRFC, the PLS can give reasonable and explanatory guidance on the initial feature space for the modeling.

Another significant trend is optimizing energy systems using soft sensing techniques to maximize efficiency and profitability. Advanced optimization algorithms and automation technologies will improve the management and control of complex energy systems [38]. Bayesian approaches and JITL strategies will expand, providing more efficient real-time data processing and decision-making methods. Integrating multimodal data and cluster-based multiple models will further drive the evolution of soft sensing techniques [38,40]. Lastly, incorporating soft sensing models into real-time process control frameworks will be crucial for the adaptive and automated control of industrial processes. Advancements in model training, optimization, and classifier development will be necessary for optimal application timing [43]. These advancements will refine soft sensing techniques, making them more robust, interpretable, and capable of handling modern industrial complexities, enhancing process efficiency, product quality, and operational decision-making across various sectors.

4.3. Machine Learning and Neural Networks

4.3.1. Definition

Machine learning and neural networks represent significant advancements in AI, enabling the creation of models that can learn from data, make predictions, and support decision-making processes. Machine learning encompasses a broad range of algorithms that allow computers to learn patterns from data, improving their performance over time without being explicitly programmed. Neural networks, a specialized subset of machine learning, mimic the human brain's structure and function, using interconnected layers of nodes (neurons) to process information and recognize patterns. These networks are particularly adept at handling large and complex datasets, making them invaluable for tasks that require high levels of pattern recognition, such as image and speech recognition, natural language processing, and predictive analytics [47].

4.3.2. Applications

Machine learning and neural networks find broad applications across industries, including industrial automation. ANFISs integrated with computer-aided design (CAD) data facilitate the precise control of robotic systems, enabling them to navigate obstacles and achieve desired positions with high accuracy [47]. In soft sensing, machine learning models, such as Echo State Networks (ESNs), estimate key variables in industrial processes. ESNs handle high-dimensional, nonlinear data and improve prediction accuracy by integrating auto-encoders to address dimension disaster problems [48]. The JITL strategy enhances the adaptability and accuracy of real-time process control models [42].

Data augmentation techniques, like DAWI-VSG shown in [49], expand datasets with virtual samples, which is crucial where real-world data is limited. Using methods like singular value decomposition to generate and refine virtual samples significantly boosts the predictive power of soft sensors [49]. Causal discovery algorithms, such as weight comparison causal mining (WCCM), uncover underlying relationships between process variables, enhancing feature selection for improved prediction accuracy and model interpretability [50]. Neural networks with upper and lower bound constraints handle missing values, ensuring prediction reliability even with incomplete or anomalous training data [51].

4.3.3. Current Challenges

Despite their widespread adoption, machine learning and neural networks encounter significant challenges. Handling high-dimensional data, leading to the “curse of dimensionality,” poses a significant hurdle for developing accurate and efficient models. Solutions like distributed ESNs with auto-encoders reduce input dimensionality and cluster input attributes using algorithms [48]. Modeling multimode processes, common in industrial settings, is another challenge. Traditional models struggle with their dynamic and multimode characteristics. Gaussian mixture models integrated with variational autoencoders (GM-GVAER) effectively capture dynamic features and handle multimode properties, improving soft sensor model accuracy [52].

Dealing with missing data persists as a challenge. Incomplete or abnormal data can lead to inaccurate predictions. Approaches like neural networks with upper and lower bound constraints for estimated missing values ensure robust predictions, preventing the overestimation of prediction errors and maintaining model reliability [51]. Utilizing unlabeled data efficiently is a limitation of traditional soft sensors. Semi-supervised models like the quality regularization-based semi-supervised adversarial transfer model (QR-SATM) leverage adversarial transfer learning to incorporate unlabeled data effectively. They enhance model performance by pretraining with unlabeled data and fine-tuning with labeled data [53].

4.3.4. Future Trends

The future of machine learning and neural networks in industrial applications is promising, with several key trends expected to drive advancements. Enhanced feature extraction capabilities will be crucial, with more sophisticated neural network structures, such as convolutional neural networks (CNNs) and recurrent neural networks, being integrated to extract complex features from industrial processes. This will likely improve model performance and prediction accuracy in various applications. Adaptive models that can dynamically adjust to changing process conditions and varying data sequences are also on the horizon. Such models will enhance the flexibility and applicability of neural networks in real-time industrial environments, allowing for more responsive and efficient process control [51].

Improving the interpretability of neural networks is another important trend. Mechanisms like attention layers can be added to models to highlight the importance of different variables and clarify causal relationships. This will improve the understanding of how models make predictions and enhance their utility in anomaly detection and process optimization. Combining domain-specific knowledge with data-driven approaches is expected

to further enhance the robustness and reliability of machine learning models. Researchers can create more comprehensive models that better understand complex industrial processes by integrating limited prior knowledge into causal discovery algorithms and other data-driven methods.

4.4. Process Monitoring and Optimization

4.4.1. Definition

Soft sensing, a pivotal aspect of contemporary industrial operations, deviates from conventional sensor-based monitoring approaches. While traditional methods rely heavily on physical sensors for data acquisition, soft sensing leverages computational models to estimate crucial process variables, providing a cost-effective and scalable solution [54]. In industrial processes, soft sensing involves developing and deploying algorithms to infer process variables based on available data, often integrating machine learning and statistical techniques [33]. By employing these algorithms, industries can reduce reliance on physical sensors, which may be costly to install and maintain, while still obtaining accurate insights into process behavior.

4.4.2. Applications

Soft sensing applications span various industries, reflecting their versatility and utility in diverse operational contexts. For example, soft sensors integrated with advanced deep learning techniques have revolutionized monitoring capabilities in the oil refining sector, optimizing processes such as LPG purification with unprecedented accuracy and efficiency [33]. Similarly, in chemical manufacturing, data-driven modeling frameworks have empowered engineers to predict product yields and process dynamics with remarkable precision, driving improvements in operational efficiency and profitability [55].

Moreover, soft sensing is applicable in pharmaceutical sectors, where precise control over manufacturing processes is essential to ensure product quality and regulatory compliance [36]. By harnessing soft sensing techniques, pharmaceutical companies can monitor critical process parameters in real-time, facilitating timely interventions and quality assurance measures.

4.4.3. Current Challenges

Despite its transformative potential, soft sensing confronts several challenges that require ongoing research and innovation to address effectively. One significant challenge revolves around the dynamic nature of industrial processes, which often exhibit non-stationary behaviors and transient phenomena [56]. Adapting conventional modeling approaches to accommodate such dynamic environments necessitates the development of novel methodologies capable of capturing and responding to evolving process dynamics.

Additionally, data scarcity poses a significant obstacle, particularly in industries with limited access to large volumes of high-quality data [49]. Overcoming this challenge requires developing data augmentation and synthesis techniques, enabling the generation of representative datasets for training soft sensing models. Furthermore, ensuring the robustness and reliability of soft sensing algorithms in the face of noisy or incomplete data remains a critical area of research and development.

4.4.4. Future Trends

Several emerging trends are poised to shape the future of soft sensing in industrial applications. One such trend is the increasing integration of soft sensing with advanced analytics and AI techniques, including deep learning and reinforcement learning [44]. By leveraging these cutting-edge technologies, industries can unlock new capabilities in process monitoring, optimization, and control, paving the way for autonomous and adaptive manufacturing systems. Moreover, the advent of edge computing and IoT technologies holds promise for decentralized soft sensing solutions, enabling real-time monitoring and control at the device level [41]. This trend towards edge-based soft sensing architectures has

the potential to enhance scalability, resilience, and responsiveness in industrial operations, particularly in sectors characterized by distributed and interconnected production systems.

5. Discussions

Figure 7 illustrates the key findings and future directions identified in the review of IMO techniques applied in industrial electrical systems. Four main themes emerge from the analysis, each represented by a distinct set of blocks in the diagram. The first block highlights the importance of IMO techniques, showcasing their growing applications across various industrial sectors. These techniques leverage advanced algorithms and machine learning methods to enhance efficiency, productivity, and process reliability. The second set of blocks focuses on soft sensing techniques, emphasizing their role in estimating difficult-to-measure variables in industrial processes. These techniques use statistical methods and machine learning algorithms to predict key quality parameters, enabling the improved monitoring and control of processes. Moving to the third set of blocks, machine learning and neural networks take center stage, demonstrating their versatility and utility in various industrial domains. These technologies enable the development of intelligent models for process control, predictive analysis, and anomaly detection, driving improvements in operational efficiency and product quality. Finally, the fourth set of blocks highlights process monitoring and optimization strategies, underscoring the shift towards autonomous and adaptive manufacturing systems. Integrating edge computing and IoT technologies promises real-time device-level monitoring and control, enhancing industrial operations’ scalability, resilience, and responsiveness.

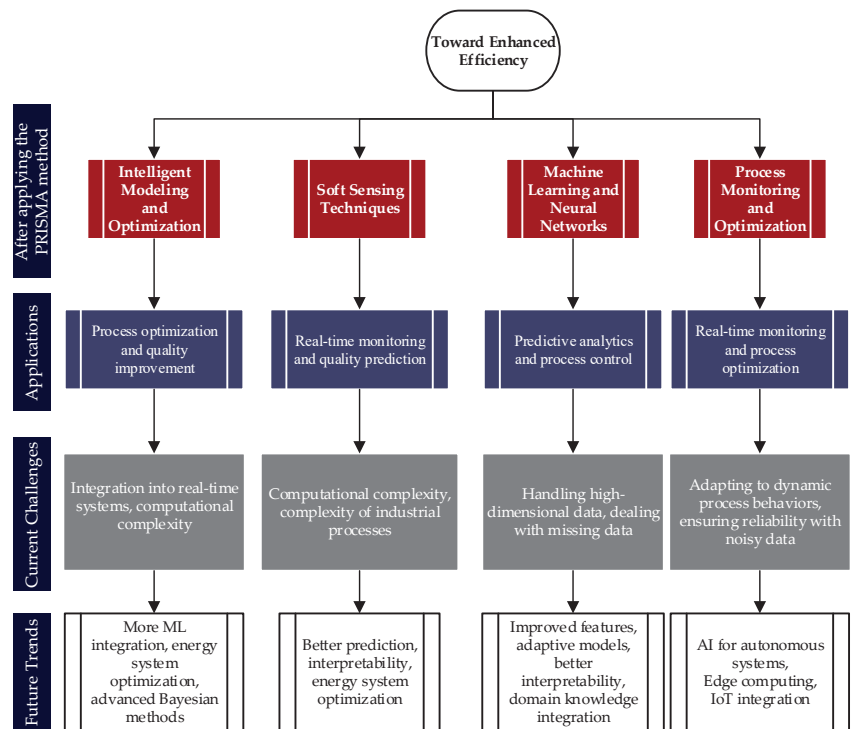


Figure 7. Trends and innovations in industrial electrical systems optimization.

In the optimization of intelligent manufacturing, challenges such as the interpretability of complex models and the management of incomplete data persist. Interpretability is crucial since complex models, although accurate, often function as “black boxes.” To

enhance transparency, explanatory models like decision trees, post-hoc techniques such as local interpretable model-agnostic explanations (LIME) and shapley additive explanations (SHAP), and advanced visualization tools can be used. On the other hand, the management of incomplete data is a significant issue due to sensor failures and communication errors. To address this, data imputation techniques, the development of robust models, and semi-supervised learning algorithms can be employed. Additionally, error detection and correction can improve data quality in real-time. These combined strategies can increase the reliability and applicability of intelligent manufacturing technologies, allowing for more effective advancements in the field.

The review underscores the critical necessity of integrating domain-specific knowledge into learning models for industrial electrical systems. This integration is pivotal for enhancing the effectiveness and applicability of advanced algorithms in practical industrial environments. By infusing industry-specific insights into algorithmic frameworks, these models can achieve heightened predictive accuracy and operational efficiency. Achieving effective integration involves embedding practical industrial insights directly into algorithmic processes. For example, in fields like metallurgy and chemical engineering, understanding the intricacies of process dynamics and material behavior significantly bolsters the precision of predictive models. This alignment ensures that algorithmic outputs resonate more closely with real-world industrial conditions, thereby enhancing interpretability and fostering greater acceptance among industry stakeholders.

Furthermore, the integration of domain-specific knowledge facilitates the development of resilient models capable of adapting to dynamic operational challenges. While challenges such as model interpretability and adaptation to complex industrial environments persist, hybrid models that blend data-driven approaches with expert knowledge frameworks offer promising solutions. These methodologies not only optimize process efficiency and reduce operational costs but also pave the way for adaptive and sustainable industrial practices. Looking ahead, future research should focus on refining methodologies that seamlessly integrate domain-specific knowledge into learning models. This includes exploring advanced hybrid models and developing frameworks that effectively capture and formalize industry-specific knowledge into algorithmic solutions. By addressing these challenges, the potential of intelligent modeling and optimization in advancing industrial electrical systems towards greater efficiency, reliability, and sustainability can be fully realized.

6. Conclusions

This study used the PRISMA method to review literature using three prominent databases: Scopus, IEEE Xplore, and MDPI. It focused on soft sensing and intelligent modeling research in electrical and industrial systems. After applying selection criteria, 255 relevant articles were identified, with a significant predominance of publications obtained from Scopus, followed by IEEE Xplore and MDPI. The publication trend during the study period shows a sustained increase in scientific interest in these areas, reflecting the growing importance of research in soft sensing and intelligent modeling in industrial electrical systems. The method was filtered down to a total of 29 articles, which met the systematic literature review criteria defined by the authors.

Based on this, a review of the main contributions of these articles was carried out. In conclusion, this comprehensive review of IMO techniques in industrial electrical systems has identified key areas of application, current challenges, and future trends. Four main themes have been explored: IMO, soft sensing techniques, machine learning and neural networks, and process monitoring and optimization strategies. The applications of these techniques are diverse and encompass industrial sectors such as metallurgy, chemistry, and manufacturing. Examples have been highlighted, such as quality prediction in industrial furnaces, the monitoring of dynamic processes, and the control of industrial robots. However, significant challenges persist, such as the interpretability of complex models, managing incomplete data, and adapting to dynamic environments. Furthermore, greater

integration of domain-specific knowledge into automatic learning models is required to improve their robustness and reliability.

Looking to the future, the integration of advanced technologies like deep learning and the IoT is poised to enhance real-time monitoring and control in industrial environments. Deep learning can process large datasets to improve predictive maintenance, fault detection, and quality control. For example, CNNs and RNNs can analyze sensor data to predict equipment failures, reducing downtime and maintenance costs. The IoT enables the interconnection of industrial devices for seamless data exchange and real-time monitoring. IoT sensors can continuously collect data from industrial processes, which deep learning algorithms can analyze to optimize operations and detect anomalies. This leads to smart factories where automated systems make real-time decisions, improving efficiency, reducing energy consumption, and enhancing safety. The continuous optimization of algorithms and interdisciplinary collaboration will be crucial to fully leveraging these technologies.

Author Contributions: Conceptualization, D.O.-C. and P.A.; methodology, D.O.-C.; software, P.A.; validation, D.O.-C. and P.A.; formal analysis, D.O.-C.; investigation, P.A.; resources, D.O.-C.; data curation, P.A.; writing—original draft preparation, D.O.-C.; writing—review and editing, P.A.; visualization, D.O.-C.; supervision, P.A.; project administration, D.O.-C.; funding acquisition, D.O.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors thank the Universidad de Cuenca (UCUENCA), Ecuador, for easing access to the facilities of the Micro-Grid Laboratory of the Faculty of Engineering, for allowing the use of its equipment, and for authorizing members of its staff to provide the technical support necessary to carry out the experiments described in this article. The results of this research will serve as input for developing the project titled «Planeamiento conjunto de la expansión optima de los sistemas eléctricos de generación y transmisión», Proj. code: VIUC_XX_2024_3_TORRES_SANTIAGO, winner of the XX Concurso Universitario de Proyectos de Investigacion promoted by the Vicerrectorado de Investigacion of UCUENCA, a department to which the authors also wish to express their gratitude.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Overall information on the selected studies for the literature review.

N°	ID	Ref.	Complete Title	Article's Proposal	Authors, Year
1	S-180	[57]	Soft sensing of silicon content via bagging local semi-supervised models	Bagging local semi-supervised models (BLSM) improve the online prediction of silicon content in industrial blast furnaces, leveraging hidden information in process variables.	He et al., 2019
2	S-65	[35]	Mode Information Separated β -VAE Regression for Multimode Industrial Process Soft Sensing	An advanced regression model, SW- β -VAE and MA-SW- β -VAER, analyzes and measures multimode industrial processes using soft sensors.	Shen et al., 2023
3	S-41	[38]	Principal Component-Based Semi-Supervised Extreme Learning Machine for Soft Sensing	The PCSELM model enhances soft sensor performance by utilizing both labeled and unlabeled data.	Shi et al., 2023
4	S-11	[40]	Data-Driven Soft Sensing for Batch Processes Using Neural Network-Based Deep Quality-Relevant Representation Learning	A neural network-based deep quality-relevant representation learning approach improves soft sensing in dynamic batch processes by optimizing quality-relevant information.	Jiang et al., 2023

Table A1. Cont.

N°	ID	Ref.	Complete Title	Article's Proposal	Authors, Year
5	S-153	[42]	A novel (JITL) strategy for soft sensing with improved similarity measure based on mutual information and PLS	A new similarity measure method combining mutual information and partial least squares improves JITL-based soft sensor modeling for industrial processes.	Song et al., 2020
6	S-82	[32]	Intelligent modeling and detailed analysis of drying hydration thermal and spectral characteristics for convective drying of chicken breast slices	Convective drying characteristics of chicken breast slices at various temperatures are analyzed using ANN and semi-empirical models to optimize drying processes.	Kumar et al., 2019
7	S-32	[56]	Probabilistic stationary subspace regression model for soft sensing of nonstationary industrial processes	A probabilistic stationary subspace regression (PSSR) model enhances soft sensing for nonstationary industrial processes by capturing mathematical correlations between variables.	Zhao et al., 2024
8	S-112	[31]	IMO for smart energy hub	An IMO method for a smart energy hub model improves operation efficiency and reduces computational burden for multiple energy systems.	Liu et al., 2019
9	S-184	[36]	Bayesian (JITL) and its Application to Industrial Soft Sensing	Bayesian JITL (BJITL) improves soft sensor performance in industrial processes by enhancing relevant sample selection and base learner construction.	Shao et al., 2020
10	S-185	[47]	Intelligent Modeling and Simulation of the Inverse Kinematics Redundant 3-DOD Cooperative Using Solidworks and MATLAB/Simmechanics	Intelligent modeling with an ANFIS controller and CAD data enhances the inverse kinematics of redundant industrial manipulator robots.	Bahani et al., 2022
11	IEEE-014	[54]	Intelligent State Estimation for Continuous Fermenters Using Variational Bayesian Learning	Variational Bayesian learning algorithms accurately estimate states in continuous fermenters, focusing on improving estimation with a random transition probability matrix.	S. Gao and S. Zhao and X. Luan and F. Liu et al., 2021
12	S-12	[37]	Robust Sparse Gaussian Process Regression for Soft Sensing in Industrial Big Data Under the Outlier Condition	A robust sparse Gaussian process regression method addresses outliers in large datasets to improve model construction for industrial processes.	Huang et al., 2024
13	S-116	[43]	A two-step multivariate statistical learning approach for batch process soft sensing	A two-step approach using partial least squares and multiway partial least squares designs soft sensors for product quality prediction in industrial processes.	Hicks et al., 2021
14	S-170	[45]	Soft Sensing Applications for Non-Stable Processes Based on a Weighted High-Order Dynamic Information Structure	A novel weighted autoregressive dynamic latent variable (WARDLV) model addresses challenges of autocorrelation and non-stable features in industrial processes.	Zhang et al., 2020
15	S-110	[36]	Powder composition monitoring in continuous pharmaceutical solid-dosage form manufacturing using state estimation—Proof of concept	A model-based approach using a moving-horizon state estimator improves the monitoring of powder composition in continuous solid-dosage form manufacturing.	Destro et al., 2021

Table A1. Cont.

N°	ID	Ref.	Complete Title	Article's Proposal	Authors, Year
16	S-149	[34]	Adaptive ensemble learning strategy for semi-supervised soft sensing	An adaptive ensemble learning strategy for soft sensors enhances regression performance with limited labeled samples using semi-supervised learning.	Shi et al., 2020
17	S-21	[34]	Efficient JITL framework for nonlinear industrial chemical engineering soft sensing based on adaptive multi-branch variable scale integrated CNNs.	An efficient JITL framework (EJITL-AMVs-ICNN) enhances real-time updating of local models in chemical processes, improving prediction accuracy and reducing elapsed time.	Chen et al., 2023
18	S-02	[33]	Soft Sensing of LPG Processes Using Deep Learning	The integration of soft sensors and deep learning in oil-refinery processes enhances monitoring efficiency and predictive accuracy for de-ethanization and debutanization.	Sifakis et al., 2023
19	S-19	[46]	Probabilistic Fusion Model for Industrial Soft Sensing Based on QRFC.	A novel QRFC model based on PLS improves soft sensor performance in nonlinear and multimodal industrial processes.	Yang et al., 2023
20	S-25	[49]	Novel virtual sample generation method based on data augmentation and weighted interpolation for soft sensing with small data	A virtual sample generation method (DAWI-VSG) enhances soft sensing datasets with high-quality samples to improve prediction accuracy in industrial processes.	Song et al., 2023
21	S-31	[53]	Quality Regularization-Based Semisupervised Adversarial Transfer Model With Unlabeled Data for Industrial Soft Sensing	QR-SATM leverages unlabeled data for more accurate soft sensor predictions.	He et al., 2024
22	S-64	[50]	Neural Network Weight Comparison for Industrial Causality Discovering and Its Soft Sensing Application	WCCM algorithm discovers variable relationships to guide process modeling and control optimization in industrial processes.	He et al., 2023
23	S-10	[48]	A New Distributed Echo State Network Integrated With an Auto-Encoder for Dynamic Soft Sensing	A distributed ESN model integrated with an auto-encoder (AE-DESNm) handles high-dimensional data and improves dynamic soft sensor performance.	He et al., 2023
24	S-95	[52]	Gaussian mixture deep dynamic latent variable model with application to soft sensing for multimode industrial processes	A deep dynamic latent variable regression model (GM-GVAER) enhances soft sensing in multimode industrial processes by capturing dynamic features.	Xu et al., 2022
25	S-75	[51]	Neural networks with upper and lower bound constraints and its application on industrial soft sensing modeling with missing values	A neural network method with upper and lower bound constraints handles missing values in soft sensor data, improving prediction reliability.	Lu et al., 2022
26	IEEE-040	[39]	A Data-Driven Soft Sensing Approach Using Modified Subspace Identification With Limited Iterative Expectation-Maximization	An EM-SID algorithm improves predictive ability in data-driven soft sensor modeling by addressing biased system matrices estimation.	W. Guo and T. Pan and Z. Li and S. Chen et al., 2020

Table A1. Cont.

N°	ID	Ref.	Complete Title	Article's Proposal	Authors, Year
27	S-37	[55]	Data-driven intelligent modeling framework for the steam cracking process	A data-driven intelligent modeling framework (LARD-MARS) balances prediction accuracy and computational speed in the steam cracking process.	Zhao et al., 2023
28	S-178	[44]	Soft Sensing of a Nonlinear Multimode Process Using a Self Organizing Model and Conditional Probability Density Analysis	A self-organizing model and conditional probability density analysis handle nonlinear multimode processes for improved soft sensor performance.	Wang et al., 2019
29	S-48	[41]	Weighted target feature regression neural networks based soft sensing for industrial process	A weighted target feature regression neural network (WTFNER) was proposed to address the difficulties in measuring product quality online in industrial processes, demonstrating superior performance and generalization in simulations compared to other methods.	Guo, X. and Wang, Q. and Li, Y., 2024.

References

- Li, K.Z.; Chandrasekera, T.C.; Li, Y.; Holland, D.J. A Non-Linear Reweighted Total Variation Image Reconstruction Algorithm for Electrical Capacitance Tomography. *IEEE Sens. J.* **2018**, *18*, 5049–5057. [CrossRef]
- Ying, Z.; Wang, Y.; Wang, J.; He, Y.C. A Dynamic Latent Structure With Time-Varying Parameters for Virtual Sensing of Industrial Process With Irregular Missing Data. *IEEE Access* **2021**, *9*, 77017–77027. [CrossRef]
- Cheng, N.; Low, J.H.; Ang, B.W.K.; Goh, A.J.Y.; Yeow, C.H. Soft Fabric-Based Pneumatic Sensor for Bending Angles and Contact Force Detection. *IEEE Sens. J.* **2019**, *19*, 1269–1279. [CrossRef]
- Wang, F.H.; Cao, H.J.; Wang, P.P.; Guo, X.K.; Han, J.L.; Dong, H.F.; Zhang, X.P.; Wang, X.D. A signal analysis and identification scheme for an online multiphase micron-sized particle analyzer system. *Meas. Sci. Technol.* **2021**, *32*, 085111. [CrossRef]
- Miao, L.W.; Ma, Y.X.; Wang, J.P. ROI-Based Image Reconstruction of Electrical Impedance Tomography Used to Detect Regional Conductivity Variation. *IEEE Trans. Instrum. Meas.* **2014**, *63*, 2903–2910. [CrossRef]
- Arif, M.Z.; Seppänen, A.; Vauhkonen, M. State estimation approach to dual-modal imaging of two-phase flow based on electromagnetic flow tomography and electrical tomography. *Inverse Probl.* **2023**, *39*, 084003. [CrossRef]
- Maity, S.; Saha, M.; Saha, P.; Khanra, M. Fractional calculus-based modeling and state-of-charge estimation of supercapacitor. *J. Energy Storage* **2024**, *81*, 110317. [CrossRef]
- Li, G.Z.; Li, B.; Li, C.; Wang, S. State-of-health rapid estimation for lithium-ion battery based on an interpretable stacking ensemble model with short-term voltage profiles. *Energy* **2023**, *263*, 126064. [CrossRef]
- Muthuramalingam, L.; Chandrasekaran, K.; Xavier, F.J. Electrical parameter computation of various photovoltaic models using an enhanced jumping spider optimization with chaotic drifts. *J. Comput. Electron.* **2022**, *21*, 905–941. [CrossRef]
- Singla, M.K.; Gupta, J.; Nijhawan, P. Solid oxide fuel cell parameter estimation using enhanced LSHADE algorithm and Newton Raphson method. *Int. J. Energy Res.* **2022**, *46*, 23341–23352. [CrossRef]
- Fabrizio, M.A.; Behrens, F.H.; Bianchini, D. Monitoring of Industrial Electrical Equipment using IoT. *IEEE Lat. Am. Trans.* **2020**, *18*, 1425–1432. [CrossRef]
- Christensen, M.B.B.; Bednarek, M.J.; Denz, R.; Koch, P.; Ludwin, J.; Rodriguez-Mateos, F.; Podzorny, T.; Ravaioli, E.; Steckert, J.; Verweij, A.; et al. Enabling Real-Time Impedance Measurements of Operational Superconducting Circuits of Accelerator Magnets. *IEEE Trans. Appl. Supercond.* **2024**, *34*, 4004705. [CrossRef]
- Liu, L.S.; Mei, H.W.; Guo, C.J.; Tu, Y.X.; Wang, L.M.; Liu, J.B. Remote Optical Thermography Detection Method and System for Silicone Polymer Insulating Materials Used in Power Industry. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 5782–5790. [CrossRef]
- Xu, H.; Zhao, X.Y.; Yin, Q.Y.; Dou, J.T.; Liu, R.P.; Wang, W.A. Isolating switch state detection system based on depth information guidance. *Electron. Res. Arch.* **2024**, *32*, 836–856. [CrossRef]
- Kumar, R.; Kumar, S.; Sengupta, A. Optimization of Bio-Impedance Techniques-Based Monitoring System for Medical & Industrial Applications. *IETE J. Res.* **2022**, *68*, 3843–3854. [CrossRef]
- Olabi, A.G.; Abdelkareem, M.A.; Semeraro, C.; Al Radi, M.; Rezk, H.; Muhaisen, O.; Al-Isawi, O.A.; Sayed, E.T. Artificial neural networks applications in partially shaded PV systems. *Therm. Sci. Eng. Prog.* **2023**, *37*, 101612. [CrossRef]
- Lei, J.; Liu, Q.B.; Wang, X.Y.; Yan, R.S. Distributed electrical capacitance tomography reconstruction with data and sparsity priors. *Signal Process.* **2021**, *181*, 107922. [CrossRef]
- Soto-Rivera, O.O. Optimization of a portable nanoparticle detection device. *Tecnol. En Marcha* **2020**, *33*, 187–202. [CrossRef]

19. Sivarani, T.S.; Jawhar, S.J.; Kumar, C.A.; Kumar, K.P. Novel bacterial foraging-based ANFIS for speed control of matrix converter-fed industrial BLDC motors operated under low speed and high torque. *Neural Comput. Appl.* **2018**, *29*, 1411–1434. [CrossRef]
20. Long, Z.L.; Zhang, J.G.; Gao, Q.B.; Zhao, H.; Li, Y.M. Novel Double Compensation for Impedance-Frequency Characteristics of Rotary Ultrasonic Machining via Multiobjective Genetic Algorithm. *IEEE Trans. Autom. Sci. Eng.* **2021**, *18*, 1928–1938. [CrossRef]
21. Mansoor, M.; Mirza, A.F.; Duan, S.H.; Zhu, J.; Yin, B.Q.; Ling, Q. Maximum energy harvesting of centralized thermoelectric power generation systems with non-uniform temperature distribution based on novel equilibrium optimizer. *Energy Convers. Manag.* **2021**, *246*, 114694. [CrossRef]
22. Safiullah, S.; Rahman, A.; Lone, S.A.; Hussain, S.M.S.; Ustun, T.S. Robust frequency-voltage stabilization scheme for multi-area power systems incorporated with EVs and renewable generations using AI based modified disturbance rejection controller. *Energy Rep.* **2022**, *8*, 12186–12202. [CrossRef]
23. Reddy, B.K.; Singh, A.K. Optimal Operation of a Photovoltaic Integrated Captive Cogeneration Plant with a Utility Grid Using Optimization and Machine Learning Prediction Methods. *Energies* **2021**, *14*, 4935. [CrossRef]
24. AlShorman, O.; Irfan, M.; Saad, N.; Zhen, D.; Haider, N.; Glowacz, A.; AlShorman, A. A Review of Artificial Intelligence Methods for Condition Monitoring and Fault Diagnosis of Rolling Element Bearings for Induction Motor. *Shock Vib.* **2020**, *2020*, 1–20. [CrossRef]
25. Kudelina, K.; Asad, B.; Vaimann, T.; Rassolkin, A.; Kallaste, A.; Khang, H.V. Methods of Condition Monitoring and Fault Detection for Electrical Machines. *Energies* **2021**, *14*, 7459. [CrossRef]
26. Jabbar, A.; Jamshed, M.A.; Abbasi, Q.; Imran, M.A.; Ur-Rehman, M. Leveraging the Role of Dynamic Reconfigurable Antennas in Viewpoint of Industry 4.0 and Beyond. *Research* **2023**, *6*, 0110. [CrossRef]
27. See, K.W.; Wang, G.F.; Zhang, Y.; Wang, Y.P.; Meng, L.Y.; Gu, X.Y.; Zhang, N.; Lim, K.C.; Zhao, L.; Xie, B. Critical review and functional safety of a battery management system for large-scale lithium-ion battery pack technologies. *Int. J. Coal Sci. Technol.* **2022**, *9*, 36. [CrossRef]
28. Yang, K.Y.; Zhang, L.S.; Zhang, Z.J.; Yu, H.Q.; Wang, W.T.; Ouyang, M.Z.; Zhang, C.; Sun, Q.; Yan, X.Y.; Yang, S.C.; et al. Battery State of Health Estimate Strategies: From Data Analysis to End-Cloud Collaborative Framework. *Batter. -Basel* **2023**, *9*, 351. [CrossRef]
29. Rahman, M.A.; Saleh, T.; Jahan, M.P.; McGarry, C.; Chaudhari, A.; Huang, R.; Tauhiduzzaman, M.; Ahmed, A.; Al Mahmud, A.; Bhuiyan, M.S.; et al. Review of Intelligence for Additive and Subtractive Manufacturing: Current Status and Future Prospects. *Micromachines* **2023**, *14*, 508. [CrossRef] [PubMed]
30. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Int. J. Surg.* **2021**, *88*, 105906. [CrossRef] [PubMed]
31. Liu, T.; Zhang, D.; Dai, H.; Wu, T. Intelligent modeling and optimization for smart energy hub. *IEEE Trans. Ind. Electron.* **2019**, *66*, 9898–9908. [CrossRef]
32. Kumar, D.; Tarafdar, A.; Kumar, Y.; Badgujar, P.C. Intelligent modeling and detailed analysis of drying, hydration, thermal, and spectral characteristics for convective drying of chicken breast slices. *J. Food Process Eng.* **2019**, *42*, e13087. [CrossRef]
33. Sifakis, N.; Sarantinoudis, N.; Tsinarakis, G.; Politis, C.; Arampatzis, G. Soft Sensing of LPG Processes Using Deep Learning. *Sensors* **2023**, *23*, 7858. [CrossRef] [PubMed]
34. Chen, Y.; Li, A.; Li, X.; Xue, D.; Long, J. Efficient JITL framework for nonlinear industrial chemical engineering soft sensing based on adaptive multi-branch variable scale integrated convolutional neural networks. *Adv. Eng. Inform.* **2023**, *58*, 102199. [CrossRef]
35. Shen, B.; Yao, L.; Yang, Z.; Ge, Z. Mode Information Separated β -VAE Regression for Multimode Industrial Process Soft Sensing. *IEEE Sens. J.* **2023**, *23*, 10231–10240. [CrossRef]
36. Shao, W.; Ge, Z.; Song, Z. Bayesian Just-in-Time Learning and Its Application to Industrial Soft Sensing. *IEEE Trans. Ind. Inform.* **2020**, *16*, 2787–2798. [CrossRef]
37. Huang, H.; Peng, X.; Du, W.; Zhong, W. Robust Sparse Gaussian Process Regression for Soft Sensing in Industrial Big Data Under the Outlier Condition. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 3512811. [CrossRef]
38. Shi, X.; Kang, Q.; Bao, H.; Huang, W.; An, J. Principal Component-Based Semi-Supervised Extreme Learning Machine for Soft Sensing. *IEEE Trans. Autom. Sci. Eng.* **2023**, 1–11. [CrossRef]
39. Guo, W.; Pan, T.; Li, Z.; Chen, S. A Data-Driven Soft Sensing Approach Using Modified Subspace Identification With Limited Iterative Expectation-Maximization. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 9272–9283. [CrossRef]
40. Jiang, Q.; Wang, Z.; Yan, S.; Cao, Z. Data-Driven Soft Sensing for Batch Processes Using Neural Network-Based Deep Quality-Relevant Representation Learning. *IEEE Trans. Artif. Intell.* **2023**, *4*, 602–611. [CrossRef]
41. Guo, X.; Wang, Q.; Li, Y. Weighted target feature regression neural networks based soft sensing for industrial process. *Can. J. Chem. Eng.* **2024**, *102*, 840–852. [CrossRef]
42. Song, Y.; Ren, M. A novel just-in-time learning strategy for soft sensing with improved similarity measure based on mutual information and PLS. *Sensors* **2020**, *20*, 3804. [CrossRef] [PubMed]
43. Hicks, A.; Johnston, M.; Mowbray, M.; Barton, M.; Lane, A.; Mendoza, C.; Martin, P.; Zhang, D. A two-step multivariate statistical learning approach for batch process soft sensing. *Digit. Chem. Eng.* **2021**, *1*, 100003. [CrossRef]
44. Wang, L.; Zeng, J.; Liang, X.; He, Y.; Luo, S.; Cai, J. Soft Sensing of a Nonlinear Multimode Process Using a Self Organizing Model and Conditional Probability Density Analysis. *Ind. Eng. Chem. Res.* **2019**, *58*, 14267–14274. [CrossRef]

45. Zhang, L.; He, Y.; Yan, T.; Wang, Y.; Gu, D.; Du, H.; Li, W. Soft Sensing Applications for Non-Stable Processes Based on a Weighted High-Order Dynamic Information Structure. *IEEE Access* **2020**, *8*, 212055–212065. [CrossRef]
46. Yang, Z.; Yao, L.; Shen, B.; Wang, P. Probabilistic Fusion Model for Industrial Soft Sensing Based on Quality-Relevant Feature Clustering. *IEEE Trans. Ind. Inform.* **2023**, *19*, 9037–9047. [CrossRef]
47. Bahani, A.; Chhibat, E.H.E.; Samri, H.; Attar, H.A.E. Intelligent Modeling and Simulation of the Inverse Kinematics Redundant 3-Dof Cooperative Using Solidworks and Matlab/Simmechanics. *Int. J. Tech. Phys. Probl. Eng.* **2022**, *14*, 78–88.
48. He, Y.L.; Chen, L.; Xu, Y.; Zhu, Q.X.; Lu, S. A New Distributed Echo State Network Integrated With an Auto-Encoder for Dynamic Soft Sensing. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 2500308. [CrossRef]
49. Song, X.L.; He, Y.L.; Li, X.Y.; Zhu, Q.X.; Xu, Y. Novel virtual sample generation method based on data augmentation and weighted interpolation for soft sensing with small data. *Expert Syst. Appl.* **2023**, *225*, 120085. [CrossRef]
50. He, Y.; Kong, X.; Yao, L.; Ge, Z. Neural Network Weight Comparison for Industrial Causality Discovering and Its Soft Sensing Application. *IEEE Trans. Ind. Inform.* **2023**, *19*, 8817–8828. [CrossRef]
51. Lu, Y.; Yang, D.; Li, Z.; Peng, X.; Zhong, W. Neural networks with upper and lower bound constraints and its application on industrial soft sensing modeling with missing values. *Knowl. -Based Syst.* **2022**, *243*, 108510. [CrossRef]
52. Xu, J.; Cai, Z. Gaussian mixture deep dynamic latent variable model with application to soft sensing for multimode industrial processes. *Appl. Soft Comput.* **2022**, *114*, 108092. [CrossRef]
53. He, Y.L.; Chen, L.; Zhu, Q.X. Quality Regularization-Based Semisupervised Adversarial Transfer Model With Unlabeled Data for Industrial Soft Sensing. *IEEE Trans. Ind. Inform.* **2024**, *20*, 1190–1197. [CrossRef]
54. Gao, S.; Zhao, S.; Luan, X.; Liu, F. Intelligent State Estimation for Continuous Fermenters Using Variational Bayesian Learning. *IEEE Trans. Ind. Inform.* **2021**, *17*, 8429–8437. [CrossRef]
55. Zhao, Q.; Bi, K.; Qiu, T. Data-driven intelligent modeling framework for the steam cracking process. *Chin. J. Chem. Eng.* **2023**, *61*, 237–247. [CrossRef]
56. Zhao, H.; Xu, J.; Wang, P. Probabilistic stationary subspace regression model for soft sensing of nonstationary industrial processes. *Can. J. Chem. Eng.* **2024**, *102*, 1558–1568. [CrossRef]
57. He, X.; Ji, J.; Liu, K.; Gao, Z.; Liu, Y. Soft sensing of silicon content via bagging local semi-supervised models. *Sensors* **2019**, *19*, 3814. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Leakage Quantification in Metallic Pipes under Different Corrosion Exposure Times

Alaa Agala, Muhammad A. Khan *, Feiyang He and Abdulaaziz Alnuman

School of Aerospace, Transport and Manufacturing, Cranfield University, Bedford MK43 0AL, UK; alaa.a.agala@cranfield.ac.uk (A.A.); feiyang.he@cranfield.ac.uk (F.H.); engr.a.alnoaman@gmail.com (A.A.)

* Correspondence: muhammad.a.khan@cranfield.ac.uk

Abstract: The combined effects of aqueous corrosion, stress factors, and seeded cracks on leakage in cast iron pipes have not been thoroughly examined due to the complexity and difficulty in predicting their interactions. This study seeks to address this gap by investigating the interdependencies between corrosion, stress, and cracks in cast iron pipes to optimise the material selection and design in corrosive environments. Leakage experiments were conducted under simulated localised corrosive conditions and internal pressure, revealing that leakage increased from 0 to 25 mL with crack sizes of 0.5 mm, 0.8 mm, 1 mm, and 1.2 mm, along with corrosion times of 0, 120, 160, and 200 h, and varying stress levels. An empirical model was developed using a curve-fitting approach to map the relationships among corrosion time, crack propagation, and leakage amount. The results demonstrate that the interaction between corrosion, stress, and crack propagation was complex and nonlinear, and the leakage amount increased from 0.7 to 0.10 mm every 15 min, as evidenced by SEM microstructure images and empirical data.

Keywords: leakage; crack; degradation mechanism; crack propagation; pressure; stress; corrosion time; pipe failure

1. Introduction

Cast iron pipes have been widely used in water distribution systems due to their robustness and cost-effectiveness. However, despite these advantages, the long-term reliability of cast iron pipes is frequently compromised by various factors, including corrosion and mechanical stress. These issues often lead to significant maintenance challenges and potential failures [1,2]. The complex interaction between corrosion and stress, particularly in the presence of cracks, remains inadequately understood.

Corrosion in cast iron pipes, typically increased in aqueous environments, gradually deteriorates the material, increasing the possibility of cracking and eventual failure [3]. Stress arising from either internal pressures or external loads can further deteriorate these conditions by promoting the initiation and propagation of cracks [4–12]. The coupled effects of corrosion and stress complicate the prediction and management of pipe leakage due to their nonlinear interaction [12].

Existing research has explored the individual effects of corrosion and stress on cast iron pipes [4,13–23]. Researchers have documented localised corrosion mechanisms and their impact on material integrity [24]. Similarly, Dzioba et al. have investigated how mechanical stress influences crack growth and pipe failure [25,26]. However, comprehensive studies examining the combined effects of these factors on leakage rates are sparse. The complex interdependencies between corrosion, stress, and cracks necessitate a comprehensive understanding of their impact on pipe leakage.

This study aims to address this research gap by investigating the combined effects of corrosion, stress, and cracks on the leakage behaviour of cast iron pipes. By simulating localised corrosive conditions and applying varying stress levels, the aim is to demonstrate

Citation: Agala, A.; Khan, M.A.; He, F.; Alnuman, A. Leakage Quantification in Metallic Pipes under Different Corrosion Exposure Times. *Processes* **2024**, *12*, 1529. <https://doi.org/10.3390/pr12071529>

Academic Editor: Carlos Sierra Fernández

Received: 17 June 2024
Revised: 15 July 2024
Accepted: 18 July 2024
Published: 20 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

how these factors influence leakage rates. The experimental variables involve crack sizes from 0.5 mm, 0.8 mm, 1 mm, and 1.2 mm and corrosion times from 0 to 160 h under different stress levels. The empirical model developed through a curve-fitting approach will map the relationships among these variables and leakage amount. They will provide a deeper understanding of the mechanisms driving leakage in cast iron pipes, with the optimal selection and design of materials in corrosive environments. Ultimately, this study seeks to improve the maintenance and design of water distribution systems [26].

2. Methodology

The experimental methodology was designed to investigate the combined effects of aqueous corrosion, stress, and cracks on cast iron pipes. The preparation of the samples for the corrosion tests was conducted to ensure the reliability and relevance of the findings. In this study, CI pipes were tested under controlled internal pressures while monitoring the resulting stress. To simulate different levels of structural integrity, these pipes were pre-damaged with cracks of various sizes. Additionally, the duration of exposure to corrosive environments was altered to emulate long-term degradation effects.

2.1. Material

The major chemical constituents of cast iron, which are critical to its corrosion resistance, were analysed and are presented in Table 1.

Table 1. Major chemical constituents of cast iron.

Material	C	S	P	Mn	Si	Fe
HT200 (EN-GJL200) Cast iron	3.2	0.12	0.015	0.9	1.6	93–97%

2.2. Sample Preparation

The experiment used a solution containing 600 parts per million (ppm) of 3.5% NaCl dissolved in distilled water. This concentration was chosen based on the significant influence of chloride ions on cast iron's corrosion behaviour. The specimens, detailed in Table 2, were prepared with varying crack sizes from 0.5 to 1.2 mm to simulate the realistic damages with different severity. Crack length and depth were constant at 50 and 3 mm, respectively, for convenient machining. All specimens were subjected to different reasonable corrosion durations. In this study, 120 h was determined as the time of appearance of corrosion with a 40 h step to observe the different effects of corrosion over time [26].

Table 2. The details of sample preparation.

Crack Size	Pipe Length	Crack Length	Wall Thickness	Crack Depth	Number of Tubes Needed	Corrosion Time
0.5 mm	200 mm	50 mm	4 mm	3 mm	3	0, 120, 160 and 200 h
0.8 mm	200 mm	50 mm	4 mm	3 mm	3	0, 120, 160 and 200 h
1.0 mm	200 mm	50 mm	4 mm	3 mm	3	0, 120, 160 and 200 h
1.2 mm	200 mm	50 mm	4 mm	3 mm	3	0, 120, 160 and 200 h

The specimens were immersed in a NaCl solution for varying durations (0, 120, 160, and 200 h) to allow sufficient interaction time for the NaCl to initiate the corrosion process.

2.3. Data Collection and Analysis

The physical effects of corrosion, such as changes in crack propagation and leakage amount, were measured and analysed using a multi-linear regression approach. This

empirical model mapped the relationships among corrosion time, stress, crack propagation, and leakage amount.

By following this methodology, this study aimed to produce reliable and transferable results that could enhance the understanding of corrosion behaviour in cast iron pipes and inform the material selection and design in corrosive environments.

2.4. Experimental Setup

Figure 1 illustrates the experimental setup used in this study. The test arrangement comprised two removable end sections attached to a corroded pipe section using couplings. One end section was connected to the water supply network via a combination turbine flow meter from GPI® (Enfield, UK), and the downstream section was equipped with a calibrated pressure transducer from Baker Hughes (Houston, TX, USA). Threaded cast iron rods and end supports were used to maintain the system intact. Data collection was facilitated through a data logger, which recorded readings from the flow meter and pressure transducer. Pressure readings were recorded every second, while the flow rate was measured by the pulse generated by the flow meter for every minute of flow.

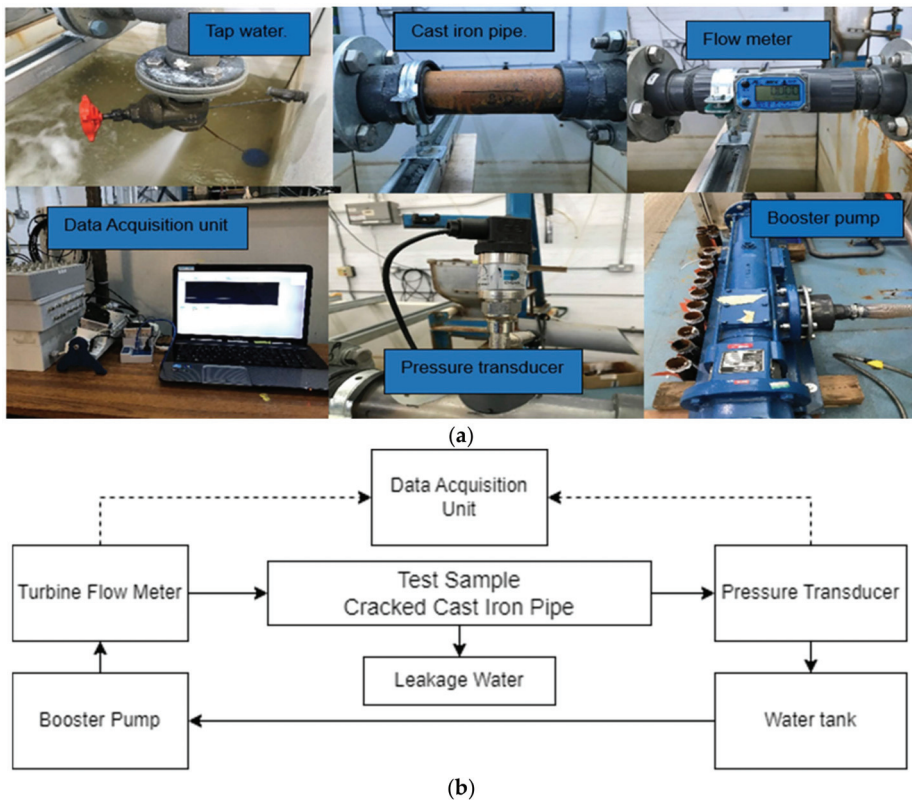


Figure 1. (a) Experimental setup; (b) conceptual diagram showing the whole equipment system connection.

The system was pressurised using a booster pump operating at 6 bars, regulated by a pressure regulator set at 4 bars. Additionally, a booster pump was equipped with an expansion vessel to reduce the frequency of pump starts. Any leakage from the system was discharged into the atmosphere, and aside from the leakage flow rate, there was no other flow present in the system. The pressure transducer and leakage observation point were at the same horizontal level. The experimental procedure involved incrementally

increasing and decreasing flow and pressure. By doing so, the stress factor of the pipeline was calculated under varying durations of corrosion when subjected to internal pressure. The stress factor, denoted as K , refers to the stress distribution around a crack. It represents the measurement of stress and deformation at the crack tip [27]. Griffith's strength and energy theory postulates that the stress intensity factor, K , can be linked to the energy released during crack growth, expressed by Equation (1) [28].

$$K = Y\sigma\sqrt{\pi a} \quad (1)$$

where K = stress intensity factor for a particular mode of cracking;

a = initial crack length;

σ = remote stress;

Y = a geometric factor.

In the experiment, seeded cracks of specific sizes were introduced into each specimen pipeline. These cracks, consistently sized at 0.5, 0.8, 1.0, and 1.2 mm, were placed at similar positions within each pipe, after which the water source was connected, and the pump was used to pressurise the pipes. The experiment continued with the gradual increase in water pressure, which increased the stresses at the tips of the pre-existing cracks, leading to their propagation. This development led to leakage, which was quantitatively measured by collecting the escaped water in a container. This process was repeated at different pressure intensities, enabling the analysis of the impact of varying pressures on crack growth and associated leakage rates. Initially focusing on specific crack dimensions, the experiment was methodically repeated with progressively larger cracks to amass extensive data on leakage rates across diverse crack sizes and pressure magnitudes. These compiled data were then carefully evaluated, and the insights are methodically organised and summarised in the next section, with the objective of deriving significant conclusions from the experimental observations.

2.5. Data Collection and Analysis

The physical effects of corrosion, such as changes in crack propagation and leakage amount, were measured and analysed using a curve-fitting approach, namely multi-linear regression (MLR). This method is used to analyse the relationships between a dependent variable and multiple independent variables. It minimises the differences between the observed and predicted outcomes based on the least-square principle.

In this study, an MLR model was applied to investigate how parameters like corrosion time, crack size, and content affect the mechanical properties of cast iron pipes [29].

3. Results and Discussion

3.1. Calculation of the Stresses at the Tips of the Cracks

The results from stress calculations using Equation (2) are presented in Tables 3–6. These tables display the stress intensity factors for various crack lengths under different applied stresses.

$$\begin{aligned} Kc &= Y\sigma_s\sqrt{\pi a} \\ \sigma_s &= Kc/(Y\sqrt{\pi a}) \end{aligned} \quad (2)$$

Table 3. Results of stress at the crack tip for corrosion time 0 h.

Corrosion Time 0 h		
Initial Crack Size (a) (mm)	Load (N)	Stress σ (MPa)
0.5	10,200	510
0.8	9900	495
1.0	9800	490
1.2	6000	300

Table 4. Results for stress at the crack tip for corrosion time 120 h.

Corrosion Time 120 h		
Initial Crack Size (<i>a</i>) (mm)	Load (N)	Hoop Stress σ (MPa)
0.5	9100	455
0.8	9000	450
1.0	8000	400
1.2	7900	395

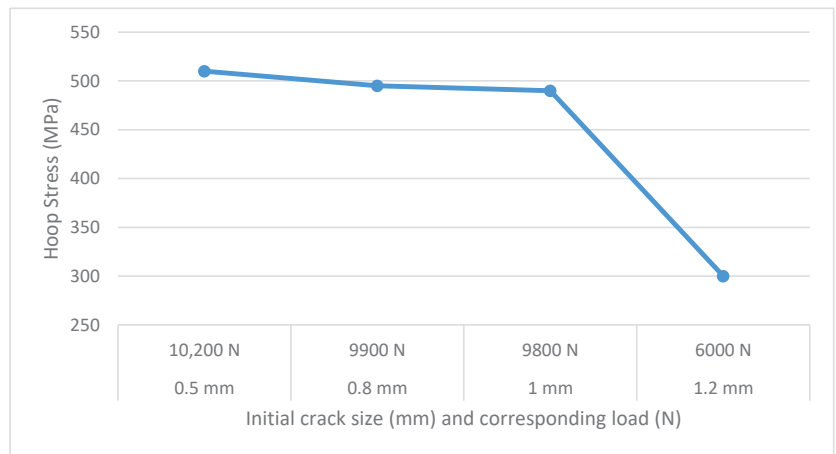
Table 5. Results of stress at the crack tip for corrosion time 160 h.

Corrosion Time 160 h		
Initial Crack Size (<i>a</i>) (mm)	Load (N)	Stress σ (MPa)
0.5	7500	375
0.8	7300	365
1.0	7000	350
1.2	6800	340

Table 6. Results of stress at the crack tip for corrosion time 200 h.

Corrosion Time 200 h		
Initial Crack Size (<i>a</i>) (mm)	Load (N)	Stress σ (MPa)
0.5	5000	250
0.8	4900	245
1.0	4500	225
1.2	3800	190

Figures 2–5 indicate the crack sizes and the corresponding pressures at which the pipes were tested. Each data point on these graphs represents a stress measurement correlating to a specific state of crack and corrosion severity against the applied internal pressure. It is worth noting that we observed some drastically decreased stress due to the rapidly reduced loading measurements for some experimental conditions.

**Figure 2.** Relationship between initial crack and stress value at corrosion time 0 h.

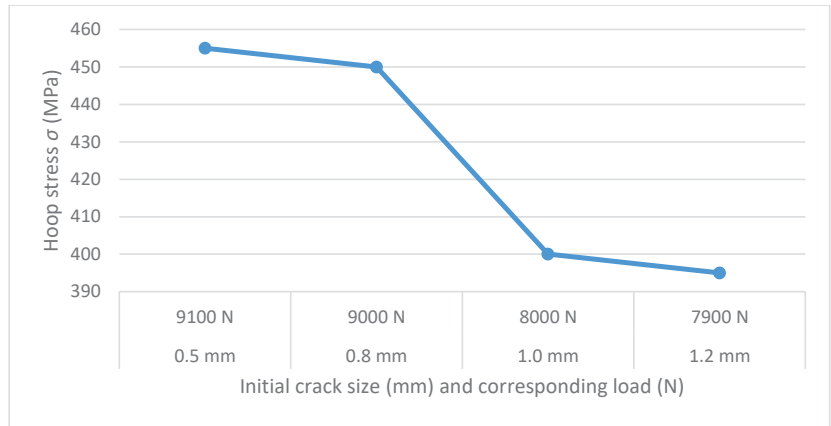


Figure 3. Relationship between initial crack and stress value for corrosion time 120.

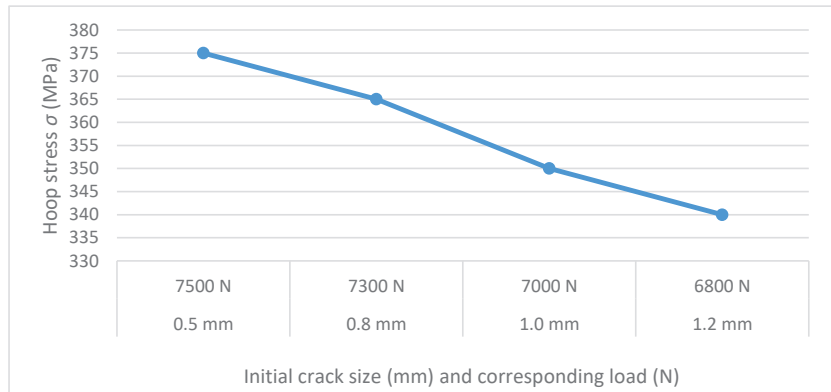


Figure 4. Relationship between initial crack and stress value for corrosion time 160 h.

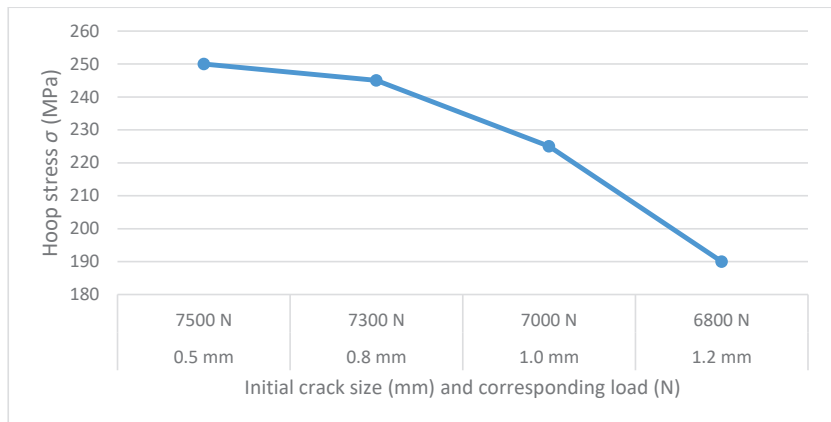


Figure 5. Relationship between initial crack and stress value for corrosion time of 200 h.

Contrary to initial assumptions, we found reduced hoop stress with increased crack size and corrosion exposure, as shown in Tables 3–6. Typically, one would expect larger

cracks and more significant corrosion to compromise the material's strength, potentially increasing stress. However, this anticipated increase in stress was not observed. The reason for this anomaly is linked to the experimental methodology. In this study, the internal pressure within the pipes was deliberately reduced from 6 to 4 bar for specimens with more extensive crack growth. This approach is a standard safety protocol in testing, implemented to prevent the risk of pipe failure. However, this decrease in internal pressure in more damaged pipes influenced the observed stress outcomes.

The findings suggest that the experimental procedures, adjusted based on the severity of pipe damage, produced corresponding modifications to the pressure levels. As a result, the hoop stress, indicative of the circumferential stress within the pipe walls, decreased as the internal pressure was intentionally lowered. This reduction served as a countermeasure to the compromised structural integrity due to increased crack size and extent of corrosion. From a research standpoint, these results are crucial for defining CI pipes' operational limits and failure thresholds. They can be used to develop predictive models to estimate the service life of such pipes and determine the necessary operational pressure adjustments for ongoing safety.

In summary, the experimental outcomes highlight the effects of damage progression on hoop stress in CI pipes and demonstrate the efficacy of pressure modulation as a failure prevention strategy. This information is invaluable for maintaining and safely operating systems that depend on a CI piping infrastructure. By including the variable of crack propagation, the experiment provides a comprehensive understanding of the failure mechanisms in CI pipes. It also assesses how pressure modulation can act as a preventative measure against risks associated with material degradation. The results indicate that adjusting the internal pressure in response to crack propagation can extend pipe life and decrease the leakage of sudden rupture.

3.2. Crack Propagation

In this study, the behaviour of cracks in brittle materials like cast iron was analysed using principles such as the stress intensity factor (SIF), which was employed to quantify the stress conditions near the tips of cracks and predict the onset of crack propagation. The influence of crack propagation in CI pipes was multifaceted, affected by internal pressure variations, exposure to corrosive elements, and different initial crack sizes. The corrosive environmental influences are clearly illustrated in Figure 6.

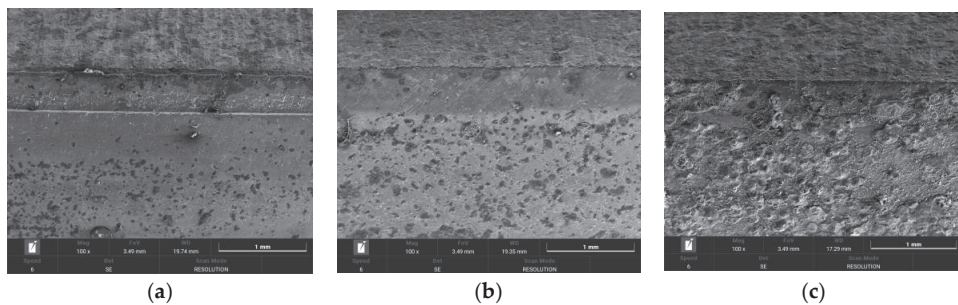


Figure 6. Corrosion conditions of the pipe surfaces after (a) 0 h, (b) 120 h, and (c) 200 h.

These images were captured through scanning electron microscopy (SEM) with $\times 100$ magnification. They show the surface morphology near the cracks on the CI pipe surface. The initial seeded cracks propagate throughout the captured area, as shown in the upper part of the images. Compared with Figure 6a, the pipe surface in Figure 6b exhibits more areas of pitting and roughness, with cracks irregular in shape and varying in depth, indicating the corrosion processes after 120 h. The morphology in Figure 6c shows further

degradation. The surface is covered with a heterogeneous layer of corrosion products. Corrosives penetrate deeper into the substrate.

The incremental changes in crack length can be visualised and measured under various conditions, as demonstrated in Figure 7. The analysis revealed that crack propagation increased from 0.55 mm to 2.43 mm. It was observed that larger cracks tended to grow more rapidly, as shown in Tables 7–10.

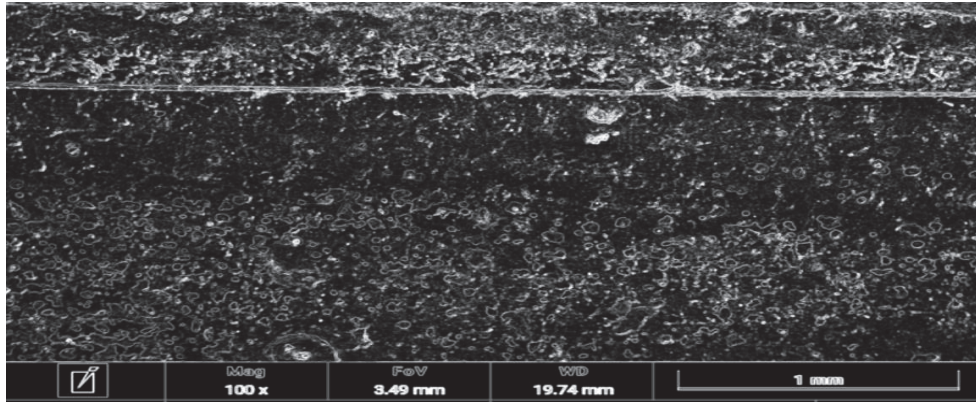


Figure 7. Visualised crack in the image processing software imageJ 1.54i.

Table 7. Crack propagation for the initial crack size of 0.5 mm and corrosion time.

Initial Crack Size 0.5 mm			
Corrosion Time 0 h	Corrosion Time 120 h	Corrosion Time 160 h	Corrosion Time 200 h
Crack Propagation mm/15 min	Crack Propagation mm/15 min	Crack Propagation mm/15 min	Crack Propagation mm/15 min
0.55	0.80	1.20	1.65
0.60	0.95	1.25	1.70
0.63	1.02	1.26	1.78
0.70	1.1	1.30	1.80

Table 8. Crack propagation for the initial crack size of 0.8 mm and corrosion time.

Initial Crack Size 0.8 mm			
Corrosion Time 0 h	Corrosion Time 120 h	Corrosion Time 160 h	Corrosion Time 200 h
Crack Propagation mm/15 min	Crack Propagation mm/15 min	Crack Propagation mm/15 min	Crack Propagation mm/15 min
0.58	1.20	1.35	1.83
0.62	1.23	1.38	1.88
0.64	1.30	1.40	1.93
0.74	1.33	1.44	1.97

Table 9. Crack propagation for the initial crack size of 1 mm and corrosion time.

Initial Crack Size 1 mm			
Corrosion Time 0 h	Corrosion Time 120 h	Corrosion Time 160 h	Corrosion Time 200 h
Crack Propagation mm/15 min	Crack Propagation mm/15 min	Crack Propagation mm/15 min	Crack Propagation mm/15 min
0.75	1.35	1.47	2.10
0.77	1.39	1.50	2.12
0.80	1.42	1.55	2.15
0.84	1.44	1.59	2.25

Table 10. Crack propagation for the initial crack size of 1.2 mm and corrosion time.

Initial Crack Size 1.2 mm			
Corrosion Time 0 h	Corrosion Time 120 h	Corrosion Time 160 h	Corrosion Time 200 h
Crack Propagation mm/15 min	Crack Propagation mm/15 min	Crack Propagation mm/15 min	Crack Propagation mm/15 min
0.77	1.47	1.62	2.30
0.80	1.52	1.70	2.35
0.82	1.55	1.90	2.40
0.85	1.59	2.09	2.43

This study comprehensively details the dynamics of crack propagation in CI pipes, presenting the results across various initial crack sizes (0.5, 0.8, 1.0, and 1.2 mm) as follows:

3.2.1. Initial Crack Size 0.5 mm

The data on corrosion time and crack propagation reveal a gradual increase in crack growth over time, as shown in Table 7. Initially, at 0 h, the crack growth ranged between 0.55 mm and 0.70 mm within 15 min. After 120 h of corrosion, the crack size increased to a range between 0.80 mm and 1.1 mm. This growth continued, reaching a range between 1.2 mm and 1.30 mm at 160 h. By 200 h, the crack size further increased, ranging between 1.65 mm and 1.80 mm. This significant increase in crack size due to prolonged corrosion time notably impacts stress resistance and elevates the risk associated with material failure.

3.2.2. Initial Crack Size 0.8 mm

The pattern resembles that of the 0.5 mm scenario but with greater increases in crack size over time, as shown in Table 8. Larger initial cracks show faster growth rates, emphasising the importance of monitoring and intervention, especially for larger cracks prone to rapid deterioration.

3.2.3. Initial Crack Size 1 mm

The data on corrosion time and crack propagation reveal a consistent pattern of increasing crack growth over time, consistent with previously observed trends but showing even larger increases in crack size, as shown in Table 9. This substantial growth intensifies durability and safety concerns, as the increasing crack size significantly impacts the material's stress resistance and elevates the risk of failure.

3.2.4. Initial Crack Size 1.2 mm

The data on corrosion time and crack propagation show a marked increase in crack growth rate from 0.77 to 2.43 mm, indicating significant material weakening due to 200 h of corrosion, as shown in Table 10. The rapid growth from the initial crack size underscores the material's high susceptibility to corrosion, intensifying durability and safety concerns.

This study reveals a consistent trend: As the duration of corrosion extended across all specimens (0, 120, 160, and 200 h) with an initial crack size of 0.5 mm, crack propagation increased from 0.55 to 1.80 mm per hour. Similarly, with an initial crack size of 0.8 mm, the crack extended from 0.58 to 1.97 mm, indicating a 4% increase. Likewise, with an initial crack size of 1 mm, the crack size propagated from 0.75 to 2.25 mm, marking a 3% increase. For an initial crack size of 1.2 mm, the crack propagated from 0.77 to 2.43 mm, representing a 5% increase. These findings underscore the significance of factoring in both the initial crack size and environmental conditions in material integrity assessments and lifespan predictions.

4. Empirical Modelling of Leakage, Crack Propagation, and Corrosion

4.1. The Use of Multi-Linear Regression to Identify the Relationships between Corrosion Time, Crack Size, and Leakage Amount

The results in Table 2 are illustrated in a 3D scatter plot, as shown in Figure 8, with a fitted polynomial surface to visualise the relationship between the variables. The independent variables, namely the corrosion time (CT, hours) and crack size (CZ, mm), are plotted along the x- and y-axes, respectively, while the dependent variable, leakage amount (LA), is represented on the z-axis. Each blue dot on the plot signifies an observed data point, correlating specific CT and CZ values to a measured LA value.

$$LA(x,y) = 8.069 + (-30.45)x + 0.01402y + 41.51x^2 + (-0.176)xy + 0.0002779y^2 + (-17.33)x^3 + 0.07944x^2y + 0.0005892xy^2 \quad (3)$$

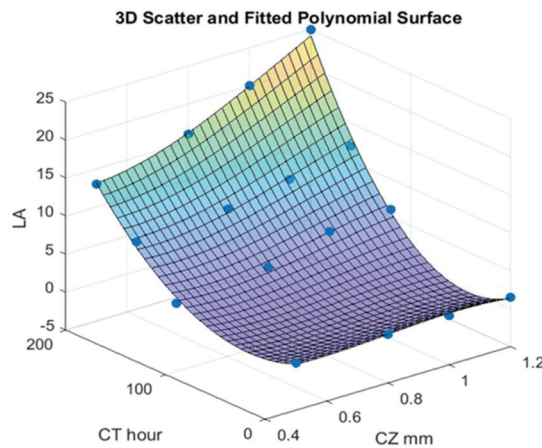


Figure 8. The relationship between crack size (CZ), corrosion time (CT), and leakage amount (LA). Blue dots represent the raw test data.

At lower CT hours and CZ mm, both near zero, LA starts at a negative value, around -0.5 , as indicated by the blue colour at the corner of the surface. As CT hours and CZ mm increase, LA increases, exemplified by a value around 8 mL at CT = 100 h and CZ = 0.6 mm. When CT and CZ approach their maxima, near 200 h and 1.2 mm, respectively, LA becomes positive, potentially reaching 20 mL, showing above the zero plane on the surface. A polynomial surface fitted through these data points visually represents the predicted LA values, colour-coded from 0 to 20, reflecting a relationship between CT, CZ, and LA. This model predicts LA values within the observed ranges, enhancing the understanding of the dynamics between these variables. The analysis examines how crack size and corrosion duration influence leakage amount. It demonstrates that larger cracks and longer corrosion times significantly contribute to the enhancement of leakage.

4.2. Methods for Evaluating the Accuracy of the Prediction Model

In this study, the effectiveness of the empirical model was assessed by applying evaluation metrics to gauge the model's performance. Specifically, we used the mean-squared error to quantify the accuracy of a model's predictions, as demonstrated in Equation (4). This parameter assesses the variance between the predicted and actual values [30].

$$MSE = (1/n) \times \sum (\text{actual forecast})^2 \quad (4)$$

where \sum is the sum, n is the sample size, actual is the actual data value, and forecast is the data value forecast.

4.3. Empirical Analysis and Results

The empirical analysis demonstrates significant degradation in cast iron pipes with larger crack sizes and extended corrosion periods of 0, 120, 160, and 200 h. This degradation is attributed to the increased SIF and the adverse effects of corrosion processes, such as pitting and stress corrosion cracking, causing material degradation. The model's findings are visually represented through a polynomial surface, where transitions in colour hues illustrate the detrimental impact of prolonged corrosion and increased crack dimensions on the pipe, as well as increased leakage amount.

4.4. Validation of the Empirical Model

The validation of the empirical model is presented in this section. Equation (3) models the relationship among crack characteristics, corrosion duration, and leakage quantity. The model was validated using the same material types but with varying crack sizes and corrosion durations. The crack sizes selected for this study were 0.9, 1.5, and 1.7 mm, while the corrosion durations were set at 100, 130, and 180 h, respectively, as presented in Table 11.

Table 11. Crack size and corrosion time values.

Crack Size (mm)	Corrosion Time (h)
0.9	100, 130, and 180
1.5	100, 130, and 180
1.7	100, 130, and 180

The predicted values for corrosion time, crack size, and leakage amount obtained from the model were compared to numerical estimations in the existing papers, as shown in Table 12. This model, formulated from empirical data, is derived from Equation (4).

Table 12. Comparison of leakage amounts.

Corrosion Time 100 h			
Crack Size (mm)	Leakage Quantity (mL/h)		
	Experimental	Empirical	Error %
0.9	3	2.8	0.4
1.5	5	4.7	0.6
1.7	6	6.2	0.2
Corrosion Time 130 h			
Crack Size (mm)	Leakage Quantity (mL/h)		
	Experimental	Empirical	Error %
0.9	8	7.9	4.45
1.5	8.3	9	0.33
1.7	10	10.3	0.5
Corrosion Time 180 h			
Crack Size (mm)	Leakage Quantity (mL/h)		
	Experimental	Empirical	Error %
0.9	17.5	18	0.28
1.5	20	20.24	0.4
1.7	26	25.4	0.21

Good agreement was observed between this model and the numerical models in [31–33]. Furthermore, as documented in Table 12, the MSE was employed to validate the results for leakage amount. A comparison of the MSE results between the empirical and numerical

models indicated an error range of 0.40% to 8% across all parameters. The error in estimating the leakage amount was below 10%, demonstrating the accuracy of the proposed model's predictions. Despite increased error margins in cases of larger cracks or prolonged corrosion, these results emphasise the necessity of establishing validation criteria that are tailored to the application's criticality.

The validation results confirm the empirical model's efficacy in predicting leakage. The low error percentages observed in most scenarios reinforce the model's precision and reliability. However, the noted variations in error rates, especially under conditions of increased crack sizes and prolonged corrosion durations, highlight areas where the model might require further improvement. These variations could be attributed to the complexities inherent in the corrosion process and the behaviour of larger cracks under prolonged exposure to corrosion.

5. Conclusions

Model development plays a crucial role in engineering, especially in assessing structural integrity under various conditions. This study focused on the impact of corrosion and mechanical loads on structural dynamics and crack propagation. An empirical model was formulated to include variables such as crack size, duration of corrosion exposure, and leakage quantity. The validation of the model was achieved through both experimental and numerical methods. The investigation revealed that crack propagation behaviour varied with changes in corrosion duration and crack size, particularly in scenarios involving large cracks and prolonged corrosion exposure.

This study also revealed the crack propagation characteristics in cast iron (CI) pipes exposed to corrosive environments. Our findings indicated a linear progression of crack growth regardless of the crack size and corrosion duration. Multiple cracks were observed on surfaces with pre-existing rectangular-shaped cracks, suggesting the usefulness of analysing crack propagation using pre-seeded cracks of specific geometries. However, future research should consider using naturally occurring cracks for more representative insights.

Experimental observations showed that corrosion progressively undermines material integrity, leading to crack initiation and expansion. As exposure to corrosive conditions extends, cracks tend to grow larger, which accelerates fluid leakage by providing more efficient channels for flow. This results in a significant increase in water leakage. A key finding was the correlation between model parameters and the reduction in material integrity, which decreased with increasing corrosion time and crack size. These results confirm that material strength reduction is an important factor to consider in such studies. The experimental data support this, indicating that longer corrosion duration and larger crack size lead to a decrease in material integrity.

Author Contributions: Conceptualization, A.A. (Alaa Agala) and M.A.K.; methodology, A.A. (Alaa Agala) and M.A.K.; validation, A.A. (Alaa Agala); formal analysis and A.A. (Abdulaaziz Alnuman) and A.A. (Alaa Agala); investigation, A.A. (Alaa Agala) and A.A. (Abdulaaziz Alnuman); resources, M.A.K.; writing—original draft preparation, A.A. (Alaa Agala); writing—review and editing, A.A. (Alaa Agala), F.H. and M.A.K.; visualization, A.A. (Alaa Agala) and F.H.; supervision, M.A.K.; project administration, M.A.K.; funding acquisition, M.A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Makar, J.M.; Desnoyers, R.; McDonald, S.E. *Failure Modes and Mechanisms in Cast Grey Iron Pipe*; CRC Press: Boca Raton, FL, USA, 2001.
- Rajani, B.; Kleiner, Y. Comprehensive Review of Structural Deterioration of Water Mains: Physically Based Models. *Urban Water* **2001**, *3*, 151–164. [CrossRef]
- Melchers, R.; Jeffrey, R. The Critical Involvement of Anaerobic Bacterial Activity in Modelling the Corrosion Behaviour of Mild Steel in Marine Environments. *Electrochim. Acta* **2008**, *54*, 80–85. [CrossRef]
- Zagórski, A.; Matysiak, H.; Tsyrylnyk, O.; Zvirko, O.; Nykyforchyn, H.; Kurzydłowski, K. Corrosion and Stress-Corrosion Cracking of Exploited Storage Tank Steel. *Mater. Sci.* **2004**, *40*, 113–117. [CrossRef]
- Olisa, S.C.; Khan, M.A.; Starr, A. Elastic Wave Mechanics in Damaged Metallic Plates. *Symmetry* **2023**, *15*, 1989. [CrossRef]
- Alqahtani, I.; Starr, A.; Khan, M. Fracture Behaviour of Aluminium Alloys under Coastal Environmental Conditions: A Review. *Metals* **2024**, *14*, 336. [CrossRef]
- Alqahtani, I.; Starr, A.; Khan, M. Coupled Effects of Temperature and Humidity on Fracture Toughness of Al–Mg–Si–Mn Alloy. *Materials* **2023**, *16*, 4066. [CrossRef] [PubMed]
- Alqahtani, I.; Starr, A.; Khan, M. Investigation of the Combined Influence of Temperature and Humidity on Fatigue Crack Growth Rate in Al6082 Alloy in a Coastal Environment. *Materials* **2023**, *16*, 6833. [CrossRef] [PubMed]
- Samuel, O.C.; Starr, A.; Muhammad, K.A. Monitoring Evolution of Debris-Filled Damage Using Pre-Modulated Wave and Guided Wave Ultrasonic Testing (GWUT). *Measurement* **2022**, *199*, 111558. [CrossRef]
- Kamei, K.; Khan, M.A. Current Challenges in Modelling Vibrational Fatigue and Fracture of Structures: A Review. *J. Braz. Soc. Mech. Sci. Eng.* **2021**, *43*, 77. [CrossRef]
- Niazi, H.; Eadie, R.; Chen, W.; Zhang, H. High PH Stress Corrosion Cracking Initiation and Crack Evolution in Buried Steel Pipelines: A Review. *Eng. Fail. Anal.* **2021**, *120*, 105013. [CrossRef]
- Ryakhovskikh, I.V.; Bogdanov, R.I. Model of Stress Corrosion Cracking and Practical Guidelines for Pipelines Operation. *Eng. Fail. Anal.* **2021**, *121*, 105134. [CrossRef]
- Agala, A.; Khan, M.; Starr, A. Degradation Mechanisms Associated with Metal Pipes and the Effective Impact of LDMs and LLMs in Water Transport and Distribution. *Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci.* **2022**, *237*, 1855–1876. [CrossRef]
- Akhi, A.H.; Dhar, A.S. Stress Intensity Factors for External Corrosions and Cracking of Buried Cast Iron Pipes. *Eng. Fract. Mech.* **2021**, *250*, 107778. [CrossRef]
- Al-mashhadani, M.H.; Ahmed, A.A.; Hussain, Z.; Mohammed, S.A.; Yusop, R.M.; Yousef, E. Inhibition of Corrosion: Mechanisms and Classifications an Overview. *J. Pure Sci.* **2020**, *25*, 1–9. [CrossRef]
- Sancy, M.; Gourbeyre, Y.; Sutter, E.M.M.; Tribollet, B. Mechanism of Corrosion of Cast Iron Covered by Aged Corrosion Products: Application of Electrochemical Impedance Spectrometry. *Corros. Sci.* **2010**, *52*, 1222–1227. [CrossRef]
- Biezma, M.V. The Role of Hydrogen in Microbiologically Influenced Corrosion and Stress Corrosion Cracking. *Int. J. Hydrogen Energy* **2001**, *26*, 515–520. [CrossRef]
- Liyanage, K.T.H.; Dhar, A.S. Effects of Corrosion Pits on Wall Stresses in Cast-Iron Water Mains. *J. Pipeline Syst. Eng. Pract.* **2017**, *8*. [CrossRef]
- Cronin, D.S. Assessment of Corrosion Defects in Pipelines. Ph.D. Thesis, University of Waterloo, Waterloo, ON, Canada, 2000.
- Mohebbi, H.; Li, C.Q. Experimental Investigation on Corrosion of Cast Iron Pipes. *Int. J. Corros.* **2011**, *2011*, 17. [CrossRef]
- Yamamoto, K.; Mizoguti, S.; Yoshimitsu, K.; Kawasaki, J. Relation between Graphitic Corrosion and Strengthdegradation of Cast Iron Pipe. *Corros. Eng.* **1983**, *32*, 157–162. [CrossRef]
- Cerit, M.; Genel, K.; Eksi, S. Numerical Investigation on Stress Concentration of Corrosion Pit. *Eng. Fail. Anal.* **2009**, *16*, 2467–2472. [CrossRef]
- Dean, S.W.; Grab, G.D. Corrosion of Carbon Steel by Concentrated Sulfuric Acid. *Mater. Perform.* **1985**, *24*, 21–25.
- Stumm, W. Investigation of the Corrosive Behavior of Waters. *J. Sanit. Eng. Div.* **1960**, *86*. [CrossRef]
- Dzioba, I.; Zvirko, O.; Lipiec, S. *Assessment of Operational Degradation of Pipeline Steel Based on True Stress–Strain Diagrams*; Springer International Publishing: Berlin/Heidelberg, Germany, 2021; Volume 102, ISBN 9783030580735.
- Rajani, B.; Makar, J. A Methodology to Estimate Remaining Service Life of Grey Cast Iron Water Mains. *Can. J. Civ. Eng.* **2000**, *27*, 1259–1272. [CrossRef]
- Kim, J.; Bae, C.; Woo, H.; Kim, J.; Hong, S. Assessment of Residual Tensile Strength on Cast Iron Pipes. In *Pipelines 2007: Advances and Experiences with Trenchless Pipeline Projects, Proceedings of the ASCE International Conference on Pipeline Engineering and Construction, Boston, MA, USA, 8–11 July 2007*; American Society of Civil Engineers (ASCE): Reston, VA, USA, 2007; Volume 62. [CrossRef]
- Wang, W.; Zhou, A.; Fu, G.; Li, C.Q.; Robert, D.; Mahmoodian, M. Evaluation of Stress Intensity Factor for Cast Iron Pipes with Sharp Corrosion Pits. *Eng. Fail. Anal.* **2017**, *81*, 254–269. [CrossRef]
- Khademi, F.; Behfarnia, K. Evaluation of Concrete Compressive Strength Using Artificial Neural Network and Multiple Linear Regression Models. *Int. J. Optimization Civ. Eng.* **2016**, *6*, 423–432.
- Efron, B. How Biased Is the Apparent Error Rate of a Prediction Rule? *J. Am. Stat. Assoc.* **1986**, *81*. [CrossRef]
- Raja, V.S. Grand Challenges in Metal Corrosion and Protection Research. *Front. Met. Alloys* **2022**, *1*, 894181. [CrossRef]

32. Ren, L.; Jiang, T.; Jia, Z.-G.; Li, D.-S.; Yuan, C.-L.; Li, H.-N. Pipeline Corrosion and Leakage Monitoring Based on the Distributed Optical Fiber Sensing Technology. *Measurement* **2018**, *122*, 57–65. [CrossRef]
33. Ilie, M.C.; Chiş, T.V.; Maior, I.; Răducanu, C.E.; Deleanu, I.M.; Dobre, T.; Pârvulescu, O.C. Experimental Investigation and Modeling: Considerations of Simultaneous Surface Steel Droplets' Evaporation and Corrosion. *Metals* **2023**, *13*, 1733. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Zone Model Predictive Control with Ellipsoid Softening Target in Chemical Processes

Xin Wan

Department of Automation, China University of Petroleum Beijing, Beijing 102249, China; xinwanauto@gmail.com

Abstract: A zone control algorithm is proposed that considers both economic performance indicators and control performance indicators. Unlike classic set point control, zone control expands the control target into a convex set. In this study, an ellipsoid is used as the control target, and the advantages of the ellipsoid target are explained in terms of overall stability and computational load. After defining the distance measurement function and appropriate terminal constraints, an objective function that considers both control performance and optimization performance is constructed. A theoretical analysis shows that the proposed control algorithm satisfies the Lyapunov stability criterion. The superiority of the ellipsoid control target in handling complex multivariable control tasks is also demonstrated. This method has significant potential value in practical industrial applications, helping to unleash the potential control performance and economic benefits of zone control systems. Finally, the feasibility and stability of the algorithm are verified through a typical chemical process simulation.

Keywords: zone control; model predictive control; ellipsoid control target; dual-stage control; stability; time limited control

1. Introduction

In process control, the system usually has complex features such as multivariable, strong coupling, large dead time and nonlinearity. These complex features bring difficulties to the design of classic controllers. Therefore, many modern control theories have been established to overcome the complex characteristics of the system. Model predictive control (MPC), as a model-based multi-variable control algorithm, has been widely studied and applied [1–4]. In the MPC algorithm, the physical constraints of the system are used as the corresponding hard constraints of the controller, and the control objective of the system is reflected by the objective function, which is the soft constraint. By continuously solving the optimization problem at each time, the best control law at each sampling time can be obtained. Due to the limitation of computers' computing power, the early MPC could only satisfy a simple paradigm. With the development of computer hardware and software, great progress has been made in the field of linear programming and nonlinear programming, and corresponding solutions for complex optimization problems have been provided. Therefore, MPC control algorithms suitable for various scenarios can be developed. In industrial applications, MPC control strategies have been used to cope with various complex multi-variable systems. At the same time, the MPC control strategy is also developing towards high flexibility. In the theoretical part, researchers have proposed a variety of design paradigms based on the Lyapunov stability criterion.

The realization of an MPC controller depends on solving optimization problems with constraints. These optimization problems use the observations of the system state variables at the current sampling time to predict the state trajectory in the future finite horizon, and they output the first element of the optimal control sequence under a given optimization goal. At the next sampling time, the above solution process is repeated to obtain a new optimal control law. Rawlings and Muske [5] developed an infinite horizon predictive

Citation: Wan, X. Zone Model Predictive Control with Ellipsoid Softening Target in Chemical Processes. *Processes* **2024**, *12*, 1611. <https://doi.org/10.3390/pr12081611>

Academic Editor: Francisco Vazquez

Received: 1 July 2024

Revised: 28 July 2024

Accepted: 29 July 2024

Published: 31 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

control that includes input and state constraints. Research shows that state constraints can be removed under certain conditions, and the solution of the optimization problem can be obtained by finite-dimensional quadratic programming. Grimm et al. [6] developed a finite horizon MPC controller for unconstrained nonlinear systems and proved that the closed-loop system is still stable without terminal constraints. Primbs and Nevistić [7] discussed the relationship between the prediction horizon and the stability of the closed-loop system. For compact initial conditions, there is a prediction horizon such that any horizon greater than this value can satisfy stability. Lee et al. [8] developed a constrained nonlinear predictive controller. The algorithm includes terminal constraints and relies on an online linear programming problem. At the same time, it is pointed out that suitable terminal constraints can ensure the asymptotic stability.

With the development of MPC technology, its application scenarios have become more diverse. Since the MPC control strategy is based on a mathematical model, a model with a complex mathematical form [9] or a model with a complex operating rule [10] can obtain better simulation results under the MPC framework. There are many ways to obtain the reference model in MPC. Wibowo and Saad [11] analyzed and compared multiple identification methods and designed an identification strategy. The identification model can reproduce the main dynamic characteristics of the real system. Similarly, the data-driven MPC strategy [12] and the MPC controller based on encrypted Lyapunov technology [13] have been successfully applied. Dubay [14] developed a self-optimized MPC controller for the highly nonlinear injection moulding process to ensure product quality. Oravec et al. [15] developed a robust MPC control strategy with soft constraints for the nonlinear process with asymmetric dynamic characteristics and improved the performance of the system in the response process through soft constraints. In the building temperature control system, Ławryńczuk and Ocioń [16] used a double-layer structure of an optimizer and MPC controller. The optimizer can calculate the optimal operating point online to minimize energy consumption. The simulation results show the effectiveness of this strategy. In addition, Zhao and Go [17] developed a two-layer MPC control strategy. The collision avoidance reference trajectory is calculated by the upper MPC controller, and the trajectory tracking is realized by the robust feedback controller. Similarly, the distributed MPC strategy developed by Dai et al. [18] also implements functions such as collision avoidance and obstacle avoidance for multiple agents. Rahman et al. [19] achieved precise control of the blow-line Kappa number by feedforwarding lignin content measurements using near-infrared spectroscopy and MPC controller. Zhao et al. [20] used fractional order MPC technology to achieve improved control performance of the steam/water loop. He et al. [21] employed MPC technology to achieve optimized tool paths for forming parts with varying wall angles in incremental sheet forming.

Based on MPC, researchers have developed economic model predictive control (EMPC) that can directly use economic index functions as objective functions. This control strategy has been fully developed in theory and application as MPC strategy [22]. In fact, the implementation paradigm of EMPC is similar to that of MPC, but there are differences between the two control strategies for the proof of stability. Adeodu et al. [23] developed an infinite horizon EMPC controller, which represented the objective function of the infinite horizon by an approximate term, and verified the effectiveness of the strategy through simulation experiments. Liu and Liu [24] discussed the impact of the finite horizon in the terminal loss function on the performance of the closed-loop system. Grüne [25] considered an EMPC control strategy without terminal constraints and gave an upper bound for the loss of this strategy. Similarly, based on MPC, integrating economic optimization factors in its solution process can still construct a control strategy with economic optimization behavior [26,27]. In addition, there are control performances corresponding to economic performance indicators, such as the form of control targets. When the control target expands from a set point to a convex set, the problem can be summarized as zone control. González and Odloak [28] developed a zone MPC control strategy, which can ensure that the closed-loop stability is satisfied both inside and outside the zone control

target. Graciano et al. [29], Capron and Odloak [30] used multi-layer MPC, where the top layer is an optimizer, and the bottom layer is a zone MPC controller. The optimizer can calculate the optimal operating point, and the MPC controller can move the system to the preset operating point under constraint conditions. Ferramosca et al. [31] developed a zone tracking MPC control strategy, using a distance metric to ensure recursive feasibility and local optimality, and verified the performance of the proposed strategy through simulation experiments. Guan et al. [32] developed a zone control algorithm based on soft constraints to reduce the frequency of zone constraint violations and increase the stability of the closed-loop system. It is worth mentioning that under the framework of zone control, the system still has additional degrees of freedom after entering the zone target. According to this degree of freedom, an optimization strategy can be designed to improve the economic benefits of the zone control system [33].

The main contributions of this paper are as follows:

- Propose a zone predictive control algorithm for ellipsoid control targets.
- Analyze typical switching control strategies and geometric forms in multi-stage control systems.
- Theoretically analyze the finite-time occurrence of switching in switching strategies.
- Verify the effectiveness of the control strategy through numerical simulations.

In this paper, a typical switching control strategy is introduced in a multi-stage control system. The differences between the geometric forms of the zone control targets are then explained. Subsequently, a control algorithm corresponding to the ellipsoid control target is designed, and the stability of the closed-loop system is theoretically ensured. Common problems in multi-stage control are discussed sequentially from Section 2, and reasonable solutions are provided. The geometric form of the control target is analyzed, leading to the selection of the ellipsoid control target as being more suitable for this strategy. In Section 3, a zone predictive control algorithm is developed, and the stability of its closed-loop system is analyzed. Section 4 presents numerical simulation experiments used to verify the effectiveness of the proposed control strategy.

2. Motivation

2.1. The Two Stages Control Strategy

In zone control, the trajectory of the system state can be divided into two stages, as shown in Figure 1. In the first stage, the system state is outside the control target. At this stage, the controller should take control performance as the main factor and weaken optimization operations such as economic optimization or quality optimization. In the second stage, the system state is inside the control target, and optimization operations such as economic optimization should be considered at this time. In addition, the controller of the second stage needs to ensure the asymptotic stability of the closed-loop system and prevent the system state from deviating from the control target due to optimized operation. In other words, the controller needs to ensure the convergence of the zone control targets during the optimization stage. Therefore, zone control can be regarded as a multi-objective control problem, and its control objective in the first stage is a set of points. The switching points of these two stages are strictly defined in some scenarios. For example, in process control, when the system state enters the zone control target, it can be considered that the first stage of the control process has ended. The next step is to start the economic optimization operation, which is to start the second stage to improve the economic performance of the entire closed-loop system. From the perspective of actual production, when the product has not yet reached the quality score that meets the customer's requirements, there is no need to consider improving economic efficiency during the production. It is meaningful to consider the economic optimization in the production process only when the various indicators of the product meet the requirements. Therefore, the two-stage control strategy is more reasonable in process control.

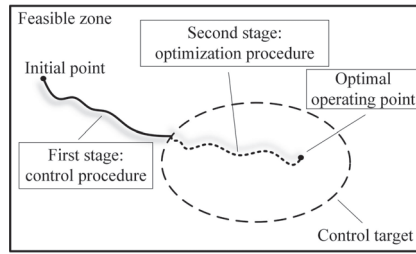


Figure 1. The schematic of two stages of zone control.

From the point of view of control theory, there are two possible reasons that explain why the system does not need to perform economic optimization when the system state is outside the control target.

Within the framework of EMPC or optimal control, the optimal trajectory depends on the mathematical model of the system. Usually, the controller designer chooses the local linearized model near the steady-state operating point as the reference model. This model can show the characteristics of the system within a given accuracy error near the steady-state point. Once the operating range is not within this neighborhood, accuracy errors cannot be guaranteed. Therefore, there is a varying model accuracy error between the current state of the system and the steady-state operating point, and the model error becomes larger as the distance increases, as shown in Figure 2.

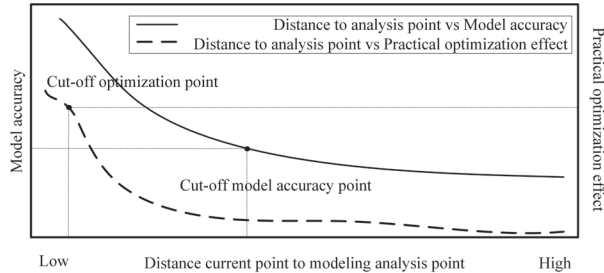


Figure 2. The schematic of optimization effect and model accuracy vs. distance to operating point.

Consider a discrete system as follows:

$$x(k + 1) = f(x(k), u(k)) \tag{1}$$

Assume that the steady-state point (x_s, u_s) satisfies $x_s = f(x_s, u_s)$. Define $f_e(x(k), u(k)) = f(x(k), u(k)) - Ax(k) - Bu(k)$ as the model error of the linear model; then, the linearization model at the steady-state point (x_s, u_s) is shown in Equation (2).

$$x(k + 1) = Ax(k) + Bu(k) + f_e(x(k), u(k)) \tag{2}$$

where $A = \frac{\partial f(x, u)}{\partial x} \Big|_{x=x_s, u=u_s}$, $B = \frac{\partial f(x, u)}{\partial u} \Big|_{x=x_s, u=u_s}$.

Assumption 1. Linearization model error $\|f_e(x(k), u(k))\| \leq \alpha_f(\|x(k) - x_s\|)$, where the monotonically increasing and zero-starting \mathcal{K} -class function $\alpha_f(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ is strictly increasing and satisfies $\alpha_f(0) = 0$.

Lemma 1. For any $\varepsilon > 0$, there exists an ellipsoid $\mathbb{B}(x_s, \varepsilon)$, so that for $\forall x \in \mathbb{B}(x_s, \varepsilon)$, the linearized system $x(k+1) = Ax(k) + Bu(k) + f_e(x(k), u(k))$ of $x(k+1) = f(x(k), u(k))$ satisfies $\|f_e(x(k), u(k))\| \leq \varepsilon$.

Proof. According to assumption 1, $\|f_e(x(k), u(k))\| \leq \alpha_f(\|x(k) - x_s\|)$. Select $\delta \in (0, \varepsilon]$; let $\alpha_f(\|x(k) - x_s\|) = \delta$; then, $\|x(k) - x_s\| = \alpha_f^{-1}(\delta)$ is established. So, for $\forall x(k) \in \{y \mid \|y - x_s\| \leq \alpha_f^{-1}(\delta)\}$, both $\|f_e(x(k), u(k))\| \leq \varepsilon$ and $\mathbb{B}(x_s, \varepsilon) = \{y \mid \|y - x_s\| \leq \alpha_f^{-1}(\delta)\}$ are established. \square

From Remark 1, the necessity of multi-stage control can be derived. That is, the model error within the specified range can meet the given requirements.

Therefore, in practical applications, the control target or the cut-off optimization point must be determined by the characteristics of the process and operating conditions, including safety, availability, reliability, accuracy, weight, and size.

On the other hand, the control strategy of optimizing and controlling separation can reduce the risk of failure in the optimization process. Conventional control may result in the optimization operation not being as expected, as shown in the left half of Figure 3. The controller drive system advances along the predicted optimal trajectory, and the system's feasible domain limiting factors prevent the system from advancing along the set trajectory. Finally, the system is limited to the same contour as the best working point, but it does not enter the setting zone. From the experimental phenomenon, the system is locked at the boundary of the feasible domain, resulting in a boundary effect. Therefore, the control strategy should be designed to avoid boundary effects.

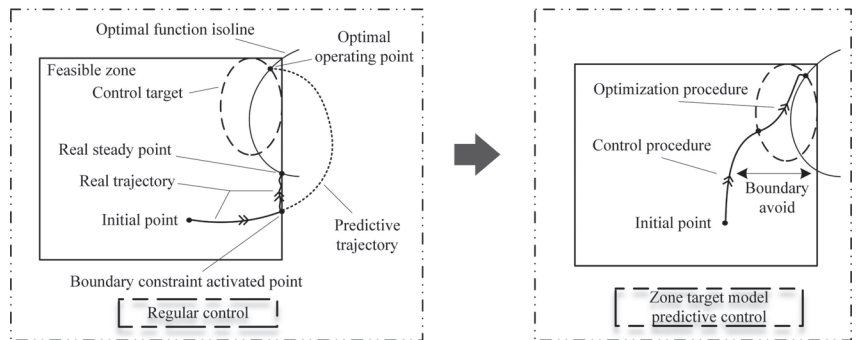


Figure 3. The schematic of regular control vs. zone target model predictive control.

In the design of the control strategy, additional parts should be added to avoid boundary effects, as shown in Figure 4. Feasible domain preprocessing is performed at the beginning of the algorithm, while two asynchronous monitoring events are designed to analyze the behavior of the system online. Within this framework, there are two independent processing strategies that ensure that the system can optimally control the system during the control process and optimization process. In a two-stage control strategy, boundary effects can be effectively avoided. The trajectory of the system under this control strategy can be referred to on the right half of Figure 3. The state of the system is driven to the set zone first and then gradually stabilized to the optimal operating point.

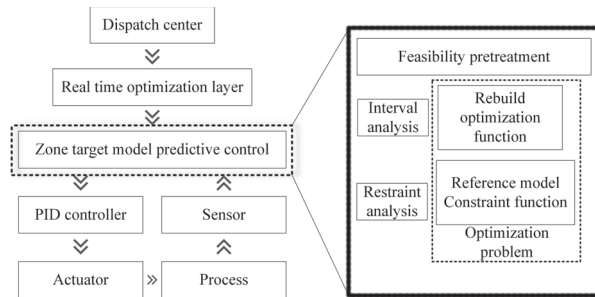


Figure 4. The framework of the zone control system.

2.2. The Ellipsoid Control Target

In this section, the spatial geometry of the zone control target is the focus. This problem is analyzed from two different perspectives: algorithm complexity and control performance. First, let us take the cube and ellipsoid in space as examples to find out the computer resources that the algorithm needs to consume in these two cases. When the controller needs to determine the distance between the current state and the set target, a spatial search algorithm is initiated to obtain a more accurate distance. In the spatial cube control target, the n -dimensional setting target divides the space into $3^n - 1$ sub-parts. If it is possible to determine which sub-portion the current state belongs to, it is easier to determine the desired distance, as shown on the left side of Figure 5. Once the sub-portion can be determined, the distance can be determined by the projection of the point to the surface. However, the judgment of the sub-parts needs to be compared with all the vertices of the control target, and the number increases as the dimension increases. And the more complicated spatial structure form will bring more difficulties to the calculation of the distance.

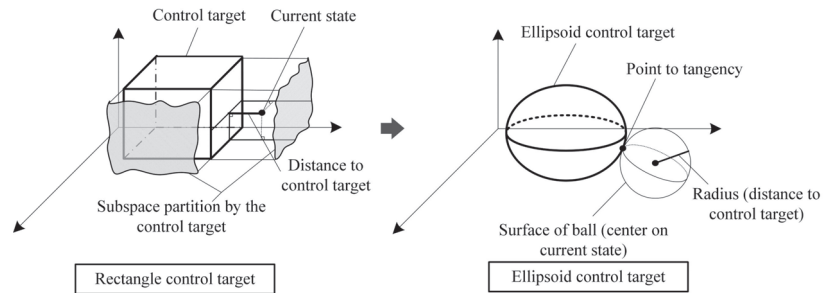


Figure 5. The schematic of the distance to the rectangle target vs. the distance to the ellipsoid target.

The situation changes when the control target is an ellipsoid in space. Construct a sphere in the space with the current state as the center of the sphere and gradually tangential to the space ellipsoid by gradually increasing the radius of the sphere. At this time, the radius of the sphere is the distance from the current state to the spatial ellipsoid. Since the tangent points are all on the two spherical surfaces, it can be solved by establishing a parametric equation system. This method can be referred to on the right half of Figure 5. In addition, from the perspective of asymptotic stability, the ellipsoid target can be more reasonable. When the system enters the zone target through the first phase. In order to ensure the progressive stability of the system, it is necessary to estimate a terminal attraction domain. Within this terminal attraction domain, the system can use an explicit controller to stabilize the closed-loop system along a trajectory at the optimal operating point. In Figure 6, regions I and III are safety zones, and the trajectory of the system within

these zones is still inside the control target. However, Area II cannot guarantee the above requirements. The system's trajectory may be driven out of the control target by the explicit controller, at which point the controller will switch. In this case, the system may rebuild the optimal trajectory, which represents the failure of this optimization operation. Sometimes, a high-frequency switching of the controller is also triggered, which is what controller designers do not want. To define the control target, an ellipsoid is used as shown on the right side of Figure 6. Because an inscribed ball can always be found inside the ellipsoid, and an explicit controller can be designed inside it, the running track of the system is always bounded. Therefore, a spatial ellipsoid is used as a target to avoid optimization failure. In summary, in the framework of zone control, the ellipsoid control target is more suitable than the spatial cube control target.

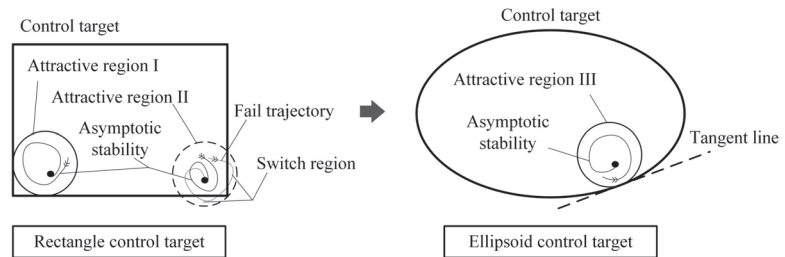


Figure 6. The schematic of control performance of the rectangle target vs. the ellipsoid target.

In some application scenarios, a given control target is often not an ellipsoid. Therefore, an approximation is proposed to ensure that the control target can be transformed into an ellipsoid. In the above description, the general spatial cube has vertices, and the critical region that cannot satisfy the local feedback control law overlaps with these vertices. Therefore, these key areas should be replaced by smooth curves. In order to obtain a smoother boundary, a space ellipsoid can be used for replacement. The replacement target can guarantee the calculation performance and ensure the stability of the closed-loop system. In addition, the ellipsoid control target has more research results than the spatial cube. At the same time, it can also meet the needs of real industrial control, so research in this area has certain significance.

The optimization algorithm inside the model predictive control will generate large fluctuations at the boundary of the derivative discontinuity. In other words, analytic expressions are not available at the cusp, which makes the design and performance tuning of the controller difficult. From the perspective of algorithms, a quadratic programming problem with smooth boundaries has a higher efficiency. Therefore, smooth boundaries can improve the reliability of the algorithm. In a specific application, in order to make the ellipsoid target directly applicable, some approximations need to be made. Take the second-order system as an example, mainly to give a more specific description, as shown in Figure 7.

In Figure 7, method (a) employs an intrinsic approximation to reduce the actual control target. This method is suitable for control systems with stricter control targets because the set target in the controller is a subset of the original target, and the steady state of the system does not deviate from the original set target. At the same time, it can also reduce the impact of overshoot to a certain extent.

Method (b) selects a control target that intersects with the original set target. Using this method will expand or crop the original target, so one needs to weigh the advantages and disadvantages of these two control objectives in order to obtain the best results.

Method (c) is an external approximation, in which case the original control target is magnified. The resulting goal includes the original target and a portion of the additional area. This approach is applicable to non-strict control objectives, and new targets need to be accepted by process requirements.

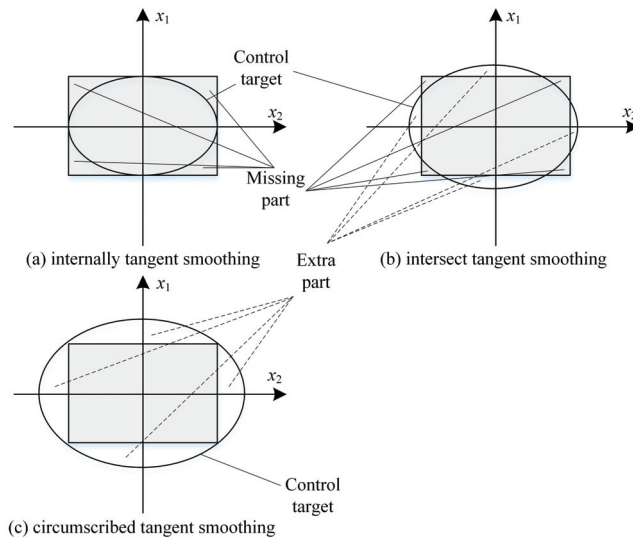


Figure 7. Typical cases on smoothing method.

2.3. Related Work

In zone control, a control strategy based on switching is designed. Compared with the conventional control strategy, the switching strategy can achieve higher economic benefits. Field theory is used to analyze the difference between the optimal trajectories of MPC and EMPC, concluding that the two control strategies are searching in different gradient fields. Based on these differences, a switching strategy for zone control is designed [34]. When the system state is outside the control target, MPC is employed to improve control performance. When the system state is within the control target, EMPC is utilized to enhance economic performance. Additionally, a boundary effect is present in zone control, where the input variable is manipulated by the controller to the boundary of the input constraint. To avoid boundary effects in real devices, a method is designed to project the optimal operating point at the boundary of the control target to a suboptimal operating point within the target. To improve control performance and avoid boundary effects of the system in transient or steady states, the target function of the controller is also modified. Zone control can also abstract a special class of application scenarios, such as a control system that includes margins. Margins often exist in the original design of the control system to extend the effective use of equipment. Therefore, margins are often ignored in the design process of control strategies, which can prevent the maximization of equipment performance. For example, in a heat exchange network, margins are used to optimize the control strategy, leading to improved system performance [35]. Considering margins as virtual state variables can transform conventional set point control into zone control, allowing for both economic performance and control performance to be considered in different zones to achieve overall optimal performance. Consequently, zone control, as a design tool for multivariable system controllers, can coordinate economic performance and control performance while avoiding negative effects such as boundary effects. Within the framework of zone control, margins can also be defined as redundancy [33]. The specific function of redundancy is to ensure that adjustments to a certain input variable within a small range do not affect control performance. Thus, a non-switching control strategy is developed to achieve coordination of control and optimization in the zone control system, addressing the challenge of switching on the switching surface.

3. Zone Model Predictive Control

The zone predictive control proposed in this paper is divided into two parts. When the system state is outside the ellipsoid control target, the distance measurement function is used as the performance index, which can indicate the distance between the current state of the system and the ellipsoid target. When the system state is inside the ellipsoid, the distance measurement function can be removed. However, only this distance measurement function cannot guarantee the stability of the system, so our objective function includes a finite horizon terminal constraint similar to EMPC [24]. In the framework of this paper, in order to ensure the validity of the constraints, appropriate assumptions are needed.

Assumption 2. *The applicable range of the linear model is within the ellipsoid $\mathbb{B}(x_s, \varepsilon)$ defined by Remark 1, so the model error of the linearized model can be considered to meet the requirements.*

Suppose the discrete equation of the system is shown in Equation (1) and satisfies $x(k) \in \mathbb{X} \subset \mathbb{R}^n$, $u(k) \in \mathbb{U} \subset \mathbb{R}^m$, where \mathbb{U} is the domain and \mathbb{X} is the range. In addition, $l(x(k), u(k)) : \mathbb{X} \times \mathbb{U} \rightarrow \mathbb{R}$ is defined, and the steady-state operating point (x_s, u_s) is defined by the following optimization problem in Equation (3).

$$\begin{aligned} (x_s, u_s) = \arg \min & l(x, u) \\ \text{s.t. } & x = f(x, u) \\ & x \in \mathbb{X}, u \in \mathbb{U} \end{aligned} \tag{3}$$

The optimal solution of the above optimization problem can be used as the closed-loop steady-state operating point of EMPC, but the optimality cannot be guaranteed for some process control with zone control targets. The optimization problem corresponding to the classic EMPC [36] is shown in Equation (4):

$$\begin{aligned} \min J(k) = & \sum_{i=0}^{N-1} l(x(k+i|k), u(k+i|k)) \\ & + V_N(x(k+N|k)) \\ \text{s.t. } & x(k+i+1|k) = f(x(k+i|k), u(k+i|k)) \\ & x(k+i|k) \in \mathbb{X} \\ & u(k+i|k) \in \mathbb{U} \\ & x(k+N|k) \in \mathbb{X}_F \end{aligned} \tag{4}$$

where $V_N(x(k+N|k))$ belongs to the terminal loss function. Since $x(k+N|k) \in \mathbb{X}_F$, let $\mathbb{X}_F = \mathbb{B}(x_s, \varepsilon)$; then, in the open-loop predicted trajectory $x(k+j|k), j \geq N$, the open-loop trajectory can be calculated using the linearization equation. The above problems can ensure that the closed-loop trajectory of the system can obtain the optimal value under the condition of the economic loss function $l(x(k), u(k))$. However, in some process control systems, there are zone control targets \mathbb{X}_{sp} , and these targets can be written as $\mathbb{X}_{sp} = \{x \in \mathbb{R}^n | x_i^{\min} \leq x_i \leq x_i^{\max}, i \in I_{[1, \dots, m]}\}$. EMPC cannot reflect the optimality of the zone target, that is, the rapidity of the system state entering the control target. To this end, a new metric is defined to indicate the distance $d(x(k), \mathbb{X}_{sp}) \in \mathbb{R}$ between the current state and the set target, with the calculation method provided in Equation (5).

$$\begin{aligned} d(x(k), \mathbb{X}_{sp}) = \min & d \\ \text{s.t. } & x_e \in \mathbb{X}_{sp} \\ & \|x(k) - x_e\| \leq d \end{aligned} \tag{5}$$

Therefore, the optimality or suboptimality of the distance $d(x(k), \mathbb{X}_{sp})$ cannot be guaranteed in a closed-loop system with an EMPC controller.

3.1. Approach to Zone Model Predictive Control

The optimization problem corresponding to the zone MPC controller can be obtained by appropriately modifying the EMPC controller. According to the proposed distance

measurement index, combined with the standardized definition, the mathematical form of the zone MPC controller can be obtained.

Definition 1. Ellipsoid control target $\mathbb{X}_{sp}(\gamma, \rho) = \{x \in \mathbb{R}^n | (x - \gamma)^T \rho (x - \gamma) \leq 1\}$, where $\gamma \in \mathbb{R}^n, \rho \in \mathbb{R}^{n \times n}, \rho > 0, \rho^T = \rho$.

Definition 2. The distance $d(x(k), \mathbb{X}_{sp}(\gamma, \rho))$ represents the distance between the current state $x(k)$ and the set target $\mathbb{X}_{sp}(\gamma, \rho)$.

$$d(x, \mathbb{X}_{sp}) = \begin{cases} \inf\{r | r = \|x_{\text{ff}} - x\|, x_{\text{ff}} \in \mathbb{X}_{sp}\} & x \notin \mathbb{X}_{sp} \\ 0 & x \in \mathbb{X}_{sp} \end{cases} \quad (6)$$

Definition 3. Terminal loss function $\bar{I}(x, u) = \|x - x_s\| + \|u - u_s\|$.

Based on the above definition, the optimization problem corresponding to the ZMPC controller is given as Equation (7).

$$\begin{aligned} \min J(x(k)) &= \sum_{i=0}^{N-1} d(x(k+i|k), \mathbb{X}_{sp}) \\ &\quad + V_h(x(k+N|k)) \\ \text{s.t. } &x(k+i|k) \in \mathbb{X} \\ &u(k+i|k) \in \mathbb{U} \\ &x(k+i+1|k) = f(x(k+i|k), u(k+i|k)) \\ &x(k+N|k) \in \mathbb{X}_F \end{aligned} \quad (7)$$

where $\mathbb{X}_F = \mathbb{B}(x_s, \epsilon)$, $V_h(x(k+N|k)) = \sum_{i=0}^{h-1} \bar{I}(x(k+N+i|k), u(k+N+i|k))$.

3.2. Stability Analysis

In this section, the known optimal sequence is used to construct a new trajectory, which is then made to serve as a Lyapunov candidate function. The construction method is different from Grüne [25], and the method of expanding in the prediction time domain of the terminal constraint conditions is adopted. First, the behavior of the system is limited within the terminal constraints.

Lemma 2. If there is a local controller $u = g(x) : \mathbb{X} \rightarrow \mathbb{U}$ in $\mathbb{B}(x_s, \epsilon)$ such that $\|x - x_s\| \geq \|f(x, g(x)) - x_s\|$ for any $x \in \mathbb{B}(x_s, \epsilon)$, then $\bar{I}(x, g(x)) - \bar{I}(f(x, g(x)), u') \geq 0$ is satisfied, where $u' \in \mathbb{U}$.

Proof.

$$\begin{aligned} \bar{I}(x, g(x)) &= \|x - x_s\| + \|g(x) - u_s\| \\ &\geq \|f(x, g(x)) - x_s\| + \|g(x) - u_s\| \\ &= \bar{I}(f(x, g(x)), u') - \|u' - u_s\| + \|g(x) - u_s\| \end{aligned} \quad (8)$$

Therefore, $u' \in \mathbb{U}$ can be selected so that $\|g(x) - u_s\| - \|u' - u_s\| \geq 0$ holds; then, $\bar{I}(x, g(x)) - \bar{I}(f(x, g(x)), u') \geq 0$.

□

Theorem 1. If there exists (x', u') such that $\bar{I}(x(k+N), u(k+N)) - \bar{I}(x', u') \geq 0$, then $V(x(k)) \geq \sum_{i=1}^{N-1} d(x(k+i), \mathbb{X}_{sp}) + \sum_{i=1}^{h-1} \bar{I}(x(k+N+i), u(k+N+i)) + \bar{I}(x', u')$ holds. $V(x(k)) = \min J(x(k))$ is the objective function value corresponding to the optimal solution of the optimization problem in Equation (7).

Proof. If $x(k+i), i = 0, \dots, N+h-1$ is the optimal solution, it must satisfy $x(k+N) \in \mathbb{B}(x_s, \epsilon)$ and $\mathbb{B}(x_s, \epsilon) \subset \mathbb{X}_{sp}$. So, $d(x(k+N), \mathbb{X}_{sp}) = 0$; then, $\sum_{i=0}^{N-1} d(x(k+i), \mathbb{X}_{sp}) - \sum_{i=1}^N d(x(k+i), \mathbb{X}_{sp}) = d(x(k), \mathbb{X}_{sp}) - d(x(k+N), \mathbb{X}_{sp}) \geq 0$. Assuming that there is a

local controller $u = g(x)$, it can be seen from Remark 2 that $\bar{I}(x(k+N), u(k+N)) - \bar{I}(x(k+N+1), u(k+N+1)) \geq 0$. So, choose $x' = f(x(k+N+h-1), u(k+N+h-1))$ and $u' = u_s$; then, $\bar{I}(x(k+N+h-1), u(k+N+h-1)) - \bar{I}(x', u') \geq 0$. Therefore:

$$\begin{aligned}
 & \sum_{i=0}^{h-1} \bar{I}(x(k+N+i), u(k+N+i)) \\
 & - \sum_{i=1}^{h-1} \bar{I}(x(k+N+i), u(k+N+i)) - \bar{I}(x', u') \\
 & = \bar{I}(x(k+N), u(k+N)) - \bar{I}(x', u') \\
 & = \bar{I}(x(k+N), u(k+N)) \\
 & \quad - \bar{I}(x(k+N+1), u(k+N+1)) \\
 & \quad + \bar{I}(x(k+N+1), u(k+N+1)) \\
 & \quad - \bar{I}(x(k+N+2), u(k+N+2)) \\
 & \quad \vdots \\
 & \quad + \bar{I}(x(k+N+h-1), u(k+N+h-1)) \\
 & \quad - \bar{I}(x', u') \\
 & \geq 0
 \end{aligned} \tag{9}$$

□

It can be seen from Remark 1 that a new trajectory sequence $x(k+1+i|k+1)$, $i = 0, \dots, N+h-1$ can be constructed through the optimal trajectory $x(k+i|k)$, $i = 0, \dots, N+h-1$ at time k and satisfy all constraints, where $x(k+1+i|k+1) = x(k+1+i|k)$, $i = 0, \dots, N+h-2$, $x(k+1+N+h-1|k+1) = x'$. Therefore, $x(k+1+i|k+1)$, $i = 0, \dots, N+h-1$ can be used as a set of feasible solutions for the optimization problem in Equation (7), that is, $V(x(k)) \geq J(x(k+1|k+1))$. And because $V(x(k+1|k+1)) \leq J(x(k+1|k+1))$, then $V(x(k)) \geq V(x(k+1))$. Since $V(x(k)) \geq 0$ and, according to definitions 2 and 3, it is bounded; $V(x(k))$ can be used as the Lyapunov function of the closed-loop system [37].

The performance loss function mentioned above can only guarantee asymptotic stability. Therefore, the terminal loss function is considered to be changed to another form to ensure that the system achieves strong convergence, meaning that it can enter the zone control target within a limited time.

Definition 4. Terminal loss function $V_N(x) = (x - x_s)^T P(x - x_s)[1]$, where P is a symmetric positive definite matrix, and x_s belongs to the interior point of the stable feasible range, that is, $x_s \in \{x | x = f(x, u), u \in U\} \cap X$ is satisfied.

Lemma 3. If the linearization system of $x(k+1) = f(x(k), u(k))$ at a steady-state operating point x_s and its neighborhood $\mathbb{B}(x_s, \varepsilon)$ is completely controllable, then for any $x(0) \in \mathbb{B}(x_s, \varepsilon)$, there must be a local controller that makes $x(k)$ enter an area inside $\mathbb{B}(x_s, \varepsilon)$ within a finite time.

Proof. Let $x(k+1) = f(x(k), u(k))$ the linearized system at the operating point x_s in its neighborhood $\mathbb{B}(x_s, \varepsilon)$ as $x(k+1) = Ax(k) + Bu(k)$, where $A = \frac{\partial f(x, u)}{\partial x} |_{x=x_s, u=u_s}$, $B = \frac{\partial f(x, u)}{\partial u} |_{x=x_s, u=u_s}$. If there is a steady-state feasible region X_s and $\mathbb{B}(x_s, \varepsilon) \subset X_s$ is satisfied, the linearized system (A,B) can be considered to be completely controllable [38,39] in $\mathbb{B}(x_s, \varepsilon)$. Let $e(k) = x(k) - x_s$; then, $e(k+1) = x(k+1) - x_s$. which is

$$\begin{aligned}
 e(k+1) &= Ax(k) + Bu(k) - x_s \\
 &= A(e(k) + x_s) + Bu(k) - x_s \\
 &= Ae(k) + Ax_s + Bu(k) - x_s \\
 &= Ae(k) + (A - I)x_s + Bu(k)
 \end{aligned} \tag{10}$$

From $x_s = Ax_s + Bu_s$, it follows that $x_s = (I - A)^{-1}Bu_s$, so:

$$\begin{aligned} e(k+1) &= Ae(k) + (A - I)(I - A)^{-1}Bu_s + Bu(k) \\ &= Ae(k) + B(u(k) - u_s) \end{aligned} \tag{11}$$

Let $u(k) = v(k) + u_s$, and finally $e(k+1) = Ae(k) + Bv(k)$. The pole configuration can make the closed-loop system $e(k+1) = A_F e(k)$ under the action of the state feedback control law $v(k) = -Fe(k)$ satisfy all the poles of A_F . It is inside the unit circle, where $A_F = A - FB$. According to the Lyapunov equation [39], if you choose a symmetric positive definite matrix Q , there must be only one symmetric positive definite matrix P so that the following equation holds.

$$A_F^T P A_F + P = -Q \tag{12}$$

Then select $V(e(k)) = e^T(k)Pe(k)$; there is

$$\begin{aligned} V(e(k+1)) - V(e(k)) &= e^T(k+1)Pe(k+1) - e^T(k)Pe(k) \\ &= e^T(k)A_F^T P A_F e(k) - e^T(k)Pe(k) \\ &= e^T(k)(A_F^T P A_F + P)e(k) \\ &= -e^T(k)Qe(k) \end{aligned} \tag{13}$$

Consider choosing $V_\alpha > 0$ so that a closed ball satisfies $\{x|V_\alpha \geq e^T Q e\} \subset \mathbb{B}(x_{sp}, \epsilon)$; then, when $x(k) \in \mathbb{B}(x_{sp}, \epsilon)$, $x(k) \notin \{x|V_\alpha \geq e^T Q e\}$, $V(e(k+1)) - V(e(k)) \leq -V_\alpha$. Then, there is

$$\begin{aligned} V(e(k)) - V(e(0)) &\leq -kV_\alpha \\ V(e(k)) &\leq V(e(0)) - kV_\alpha \\ V_\alpha &\leq V(e(k)) \leq V(e(0)) - kV_\alpha \\ k &\leq \frac{V(e(0))}{V_\alpha} - 1 \end{aligned} \tag{14}$$

So, $x(k)$ enters the area $\{x|V_\alpha \geq e^T Q e\}$ within a finite time, and the local controller is $u(k) = -Fx(k) + Fx_s + u_s$. \square

Lemma 4. *If the terminal loss function in Equation (7) is replaced by $V_N(x(k+N|k)) = (x(k+N|k) - x_s)^T P(x(k+N|k) - x_s)$, and it is assumed that the open-loop prediction trajectory corresponding to the optimal solution $u(k+i|k)$, $i = 0, \dots, N-1$ of the optimization problem at time k is $x(k+i|k)$, $i = 0, \dots, N$, then another optimal solution sequence $u(k+1+i|k+1)$, $i = 0, \dots, N-1$ can be constructed, and it is a feasible solution sequence of Equation (7) with $x(k+1|k+1) = x(k+1|k)$ as the initial value and satisfies $J(x(k+1|k+1)) - V(x(k|k)) < -\sigma$, where $\sigma > 0$, $V(x)$ represents the value of the objective function corresponding to the optimal solution of the optimization problem in Equation (7) with x as the initial condition.*

Proof. According to Lemma 3, construct a solution sequence to satisfy $u(k+1+i|k+1) = u(k+1+i|k)$, $i = 0, \dots, N-2$, $u(k+N|k+1) = -Fx(k) + Fx_s + u_s$. Then, there is $x(k+1+i|k+1) = x(k+1+i|k)$, $i = 0, 1, \dots, N-1$; so,

$$\begin{aligned} &J(x(k+1|k+1)) - V(x(k|k)) \\ &= \text{dist}(x(k+N|k+1)) - \text{dist}(x(k|k)) \\ &+ V_N(x(k+N+1|k+1)) - V_N(x(k+N|k)) \\ &\leq -\text{dist}(x(k|k)) - V_\alpha \\ &\leq -V_\alpha \end{aligned} \tag{15}$$

In Equation (7), the terminal constraint set $x(k+N|k) \in \mathbb{X}_F$ can be defined as the linearized neighborhood $\mathbb{B}(x_s, \epsilon)$ in Lemma 3. Then, the above conclusion holds, where $\sigma = V_\alpha$ is given by Lemma 3. \square

Theorem 2. *If the optimization problem after replacing the terminal loss function in Lemma 3 is used as the controller, the state variables of the closed-loop system can enter the zone control target in a limited time.*

Proof. From Lemma 4, it is known that $J(x(k+1)) - V(x(k)) < -V_\alpha$, and from the optimality of the optimization problem, it follows that $V(x(k+1)) \leq J(x(k+1))$. Thus, $V(x(k+1)) - V(x(k)) < -V_\alpha$. Therefore, $V(x(k)) - V(x(0)) < -kV_\alpha$, that is, $V(x(k)) < V(x(0)) - kV_\alpha$. And because $V(x(k)) \geq 0$, there is $V(x(0)) - kV_\alpha \geq 0$, that is, $k \leq \frac{V(x(0))}{V_\alpha}$. Combining Lemma 3, it is known that when $x(0) \notin \mathbb{X}_{sp}$, the system state can enter an area $\{x \mid V_\alpha \geq e^T Q e\}$ inside $\mathbb{B}(x_{sp}, \varepsilon)$ within a limited time. Since $\{x \mid V_\alpha \geq e^T Q e\} \subset \mathbb{B}(x_{sp}, \varepsilon) \subset \mathbb{X}_{sp}$, the system state must enter the zone control target within a limited time. Let $T_0 = \frac{V(x(0))}{V_\alpha}$. Then, for any $k \geq T_0$, $x(k) \in \mathbb{X}_{sp}$ holds. \square

From the above conclusion of strong convergence, it can be seen that the state of the system must enter the zone control target within a limited time. Therefore, the control algorithm proposed in this paper can ensure the stability of the zone control system.

4. Case Study

In this section, a simulation example is used to demonstrate the application of the MPC control strategy for the ellipsoid control target. To provide a more intuitive representation of the ellipsoid control target and to verify the effectiveness and stability of the proposed control algorithm, a continuous stirred tank reactor (CSTR) system is utilized to validate the ideas. Consider a lumped parameter model in Equation (16).

$$\begin{aligned}
 V \frac{dC}{dt} &= -k_0 e^{-\frac{E}{RT}} CV + Q_0(C_0 - C) \\
 V\rho C_P \frac{dT}{dt} &= D_h k_0 e^{-\frac{E}{RT}} CV + Q_0 \rho C_P (T_0 - T) - q_c \\
 q_c &= \frac{U_a Q_{cw} C_{pc}}{U_a + Q_{cw} C_{pc}} (T - T_c)
 \end{aligned} \tag{16}$$

where the k_0 is the reaction rate constant, E is the reaction activation energy (J/mol), R is the ideal gas constant (J/mol·K), V is the reactor volume (m^3), C is the concentration of A in reactor (mol/L), T is the temperature of reactor (K), C_p is the heat capacity of the mixture (J/mol·K), ρ is the density of the mixture (kg/m^3), D_h is the reaction heat (J/mol), Q_0 is the inlet flow rate (L/min), C_0 is the inlet concentration of A (mol/L), T_0 is the inlet temperature (K), q_c is the heat transfer rate of cooling water (J/min), Q_{cw} is the cooling water flow rate (L/min), T_c is the temperature of the cooling water (K), C_{pc} is the heat capacity of the cooling water (J/mol·K), U_a is the heat transfer coefficient ($W/m^2 \cdot K$), Q_0 is the inlet flow rate (L/min), and Q_{cw} is the cooling water flow rate (L/min). The system's structure is as follows in Figure 8. In this model, the variables are defined in Table 1.

Table 1. The description of each variable.

Variable Name	Description
u_1	Feed concentration
u_2	Coolant flow
x_1	Reactor concentration
x_2	Reactor temperature

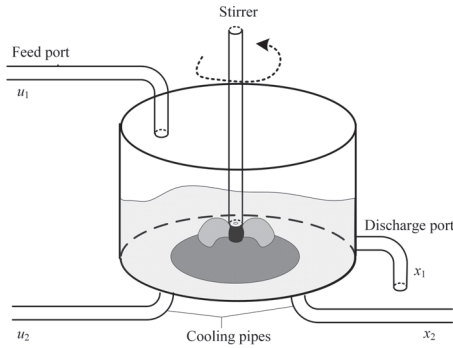


Figure 8. Model of continuous stirred tank reactor (CSTR) system.

According to the invariant set theory [40] and Remark 1, a feedback matrix H exists that can ensure the satisfaction of Equation (17).

$$(A + BH)^T P (A + BH) \leq P \tag{17}$$

To meet the constraint condition, the linear matrix inequality is given by Equation (18).

$$\begin{bmatrix} u_{\max}^2 I & Y \\ Y^T & Q \end{bmatrix} \geq 0 \tag{18}$$

where $Q = P^{-1}, H = YQ^{-1}$.

Proof. The input constraints of the system can be expressed.

$$\|u(k)\|_2 \leq u_{\max} \tag{19}$$

It can be rewritten as follows with state feedback matrix H .

$$\begin{aligned} \|u(k)\|_2^2 &= \|Hx(k)\|_2^2 \\ &= \left\| YQ^{-1/2}P^{1/2}x(k) \right\|_2^2 \\ &= x^T(k)P^{1/2}(Q^{-1/2}Y^TYQ^{-1/2})P^{1/2}x(k) \\ &\leq u_{\max}^2 \end{aligned} \tag{20}$$

The equivalent condition of the above inequality is as follows [41]:

$$\begin{aligned} Q^{-1/2}Y^TYQ^{-1/2} &\leq u_{\max}^2 I \\ Q - Y^T \frac{1}{u_{\max}^2} Y &\geq 0 \end{aligned} \tag{21}$$

With the Schur complement lemma, the linear matrix inequality can be obtained.

$$\begin{bmatrix} u_{\max}^2 I & Y \\ Y^T & Q \end{bmatrix} \geq 0 \tag{22}$$

□

A centralized parameter model was constructed and discretized with a period of 1 min near the operating point. Without loss of generality, an unmeasurable disturbance u_d was

considered in this model. The discrete state-space model with dimensionless transformation and equalization can be described as follows:

$$x(k+1) = \begin{bmatrix} 0.7415 & -0.0015 \\ 1.1935 & 0.0657 \end{bmatrix} x(k) + \begin{bmatrix} 0.15 & 0 \\ 0 & -0.91 \end{bmatrix} u(k) + u_d(k) \quad (23)$$

The system constraints are as follows in Table 2.

Table 2. The variables and constraints of the model.

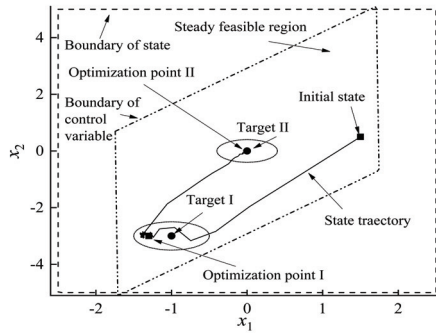
Variable Name	Constraint
u_1	$[-3, 3]$
u_2	$[-3, 3]$
x_1	$[-2.5, 2.5]$
x_2	$[-5, 5]$

Giving the initial state $x(0) = [2.5 \ 3.5]^T$, control target $P = \text{diag}([2 \ 2])$. The requirement drives the system state into the set zone primarily and then stabilizes the system. Combined with Equation (17) and Equation (18), the appropriate state feedback matrix H can be computed as

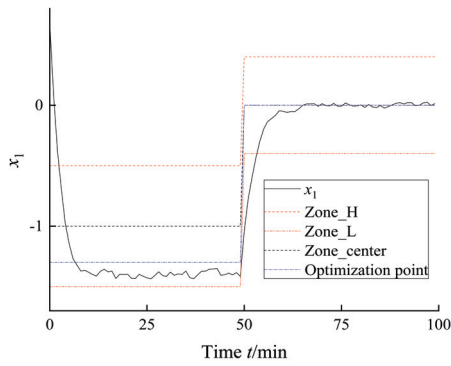
$$H = \begin{bmatrix} -0.1667 & 0.0022 \\ 0.7170 & 0.0278 \end{bmatrix} \quad (24)$$

In scenario I, the predictive horizon was 10. The coefficient of zone target penalty was 10 and optimization penalty 50. The first target was an ellipsoid target whose center was $[-1 \ -3]$ and radius $[0.5 \ 0.5]$. The optimization point was set to $[-1.3 \ -3]$. The second target for the center was $[0 \ 0]$ and the radius $[0.4 \ 0.4]$. The optimization point was set as the center $[0 \ 0]$. The system trajectory is shown in Figure 9a. The system states can be steadied in or near the optimization point on disturbance. The system has two stages in each target. First of all, the system should be governed into the target rapidly. Then, the states should be steadied at the optimization point. As shown in Figure 9b,c, the states can be attracted into the target and then stabilized at the economic optimization point. In scenario II, the controller parameters were the same as in scenario I; however, the first target was changed in the center to $[-1.7 \ -3]$ and radius $[0.5 \ 0.5]$. In addition, the system had the same second target with scenario I. The first optimization was set as $[-1.7 \ -3]$. The system states can also be tracked by the optimization points and be governed in the set target as shown in Figure 10.

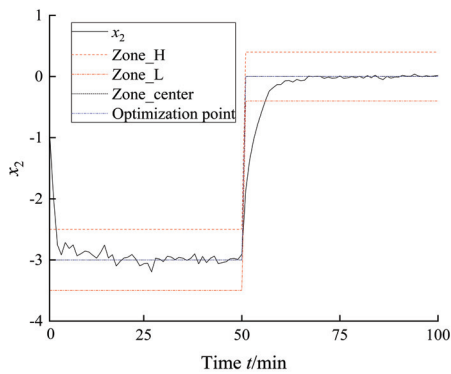
In scenario III, each parameter and target were set to be the same as those of scenario II. In addition, the first optimization point was $[-1.7 \ -2]$. This point was in the first target but out of the steady feasible region as shown in Figure 11a. Therefore, the system should be steady in the boundary of the first target as shown in Figure 11c, called boundary effect. In this extreme situation, this control algorithm can also govern the system into the set target and maintain the asymptotic stability. The algorithm's stability and control effect meet expected requirements in these three scenarios.



(a) Scenario I: Trajectory of the system states.

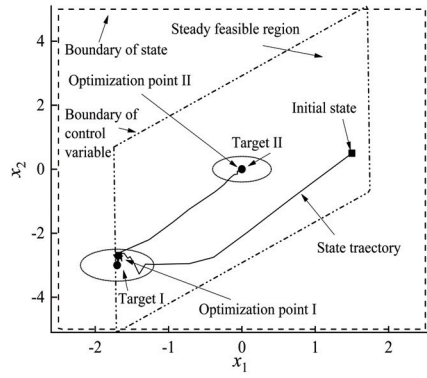


(b) Scenario I: Simulation result of x_1 .

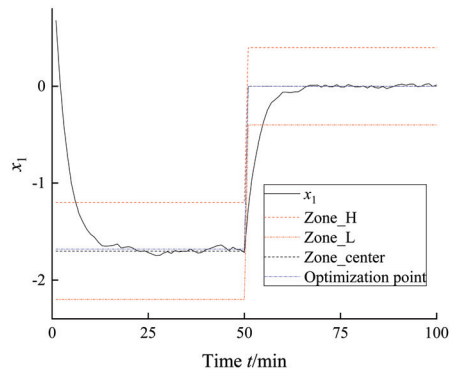


(c) Scenario I: Simulation result of x_2 .

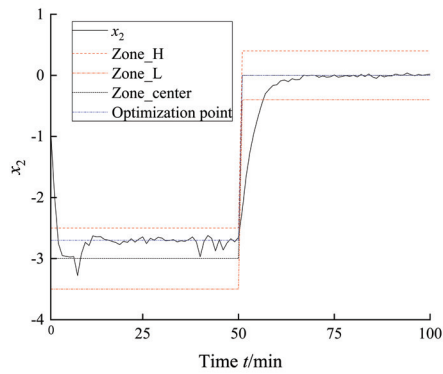
Figure 9. Simulation results of scenario I.



(a) Scenario II: Trajectory of the system states.

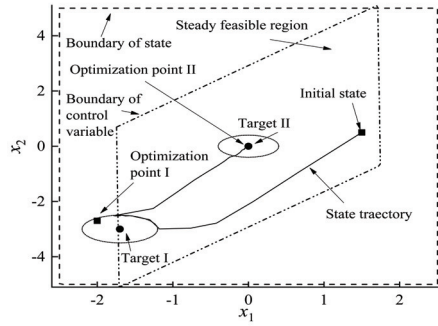


(b) Scenario II: Simulation result of x_1 .

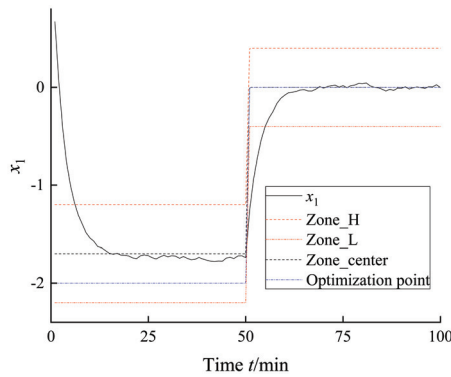


(c) Scenario II: Simulation result of x_2 .

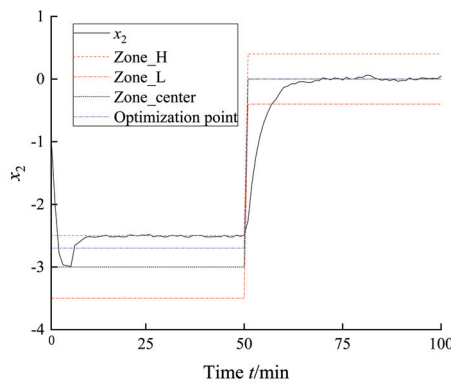
Figure 10. Simulation results of scenario II.



(a) Scenario III: Trajectory of the system states.



(b) Scenario III: Simulation result of x_1 .



(c) Scenario III: Simulation result of x_2 .

Figure 11. Simulation results of scenario III.

5. Conclusions

In this paper, the selection of targets in the zone control task is first discussed. From the perspective of computational complexity and closed-loop control performance, it is de-

terminated that the control target should be in the form of an ellipsoid. For the design of ellipsoids, three typical methods of boundary smoothing are proposed, which can shape the space cube into an ellipsoid. A predictive control algorithm based on the ellipsoid control target is then constructed. The generalized Euclidean distance is used as the objective function of the controller and is segmented from the prediction time domain. The algorithm is subsequently deduced and analyzed, and the closed-loop asymptotic stability of the algorithm is verified. Finally, the feasibility and effectiveness of the ellipsoid target predictive control algorithm are verified using a typical CSTR model.

Future research should focus on the following three aspects: limitations of model assumptions, computational complexity, and applicability. Complex models may affect real-time computational performance, while overly simplified models could introduce significant dynamic errors. Therefore, it is necessary to address these issues comprehensively and design appropriate adjustment mechanisms to enable timely switching between different linear operating points, thereby further enhancing the control performance of the system.

Funding: This research was funded by Science Foundation of China University of Petroleum, Beijing (No. 2462021YJRC022).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that may appear to influence the work reported in this paper.

References

1. Mayne, D.Q.; Rawlings, J.B.; Rao, C.V.; Scokaert, P.O. Constrained model predictive control: Stability and optimality. *Automatica* **2000**, *36*, 789–814. [CrossRef]
2. Mayne, D.Q. Model predictive control: Recent developments and future promise. *Automatica* **2014**, *50*, 2967–2986. [CrossRef]
3. Sen, M.; Singh, R.; Ramachandran, R. A hybrid MPC-PID control system design for the continuous purification and processing of active pharmaceutical ingredients. *Processes* **2014**, *2*, 392–418. [CrossRef]
4. Huang, Y.S.; Sheriff, M.Z.; Bachawala, S.; Gonzalez, M.; Nagy, Z.K.; Reklaitis, G.V. Evaluation of a combined MHE-NMPC approach to handle plant-model mismatch in a rotary tablet press. *Processes* **2021**, *9*, 1612. [CrossRef] [PubMed]
5. Rawlings, J.B.; Muske, K.R. The Stability of Constrained Receding Horizon Control. *IEEE Trans. Autom. Control* **1993**, *38*, 1512–1516. [CrossRef]
6. Grimm, G.; Messina, M.J.; Tuna, S.E.; Teel, A.R. Model predictive control: For want of a local control Lyapunov function, all is not lost. *IEEE Trans. Autom. Control* **2005**, *50*, 546–558. [CrossRef]
7. Primbs, J.A.; Nevistić, V. Feasibility and stability of constrained finite receding horizon control. *Automatica* **2000**, *36*, 965–971. [CrossRef]
8. Lee, Y.I.; Kouvaritakis, B.; Cannon, M. Constrained receding horizon predictive control for nonlinear systems. *Automatica* **2002**, *38*, 2093–2102. [CrossRef]
9. Yu, Y.; Luo, X.; Liu, Q. Model predictive control of a dynamic nonlinear PDE system with application to continuous casting. *J. Process Control* **2018**, *65*, 41–55. [CrossRef]
10. Pourdehi, S.; Karimaghvae, P. Stability analysis and design of model predictive reset control for nonlinear time-delay systems with application to a two-stage chemical reactor system. *J. Process Control* **2018**, *71*, 103–115. [CrossRef]
11. Wibowo, T.C.S.; Saad, N. MIMO model of an interacting series process for Robust MPC via System Identification. *ISA Trans.* **2010**, *49*, 335–347. [CrossRef] [PubMed]
12. Thombre, M.; Mdoe, Z.; Jäschke, J. Data-driven robust optimal operation of thermal energy storage in industrial clusters. *Processes* **2020**, *8*, 194. [CrossRef]
13. Kadakia, Y.A.; Suryavanshi, A.; Alnajdi, A.; Abdullah, F.; Christofides, P.D. Encrypted model predictive control of a nonlinear chemical process network. *Processes* **2023**, *11*, 2501. [CrossRef]
14. Dubay, R. Self-optimizing MPC of melt temperature in injection moulding. *ISA Trans.* **2002**, *41*, 81–94. [CrossRef] [PubMed]
15. Oravec, J.; Bakošová, M.; Galčíková, L.; Slávik, M.; Horváthová, M.; Mészáros, A. Soft-constrained robust model predictive control of a plate heat exchanger: Experimental analysis. *Energy* **2019**, *180*, 303–314. [CrossRef]
16. Ławryńczuk, M.; Ocloń, P. Model Predictive Control and energy optimisation in residential building with electric underfloor heating system. *Energy* **2019**, *182*, 1028–1044. [CrossRef]
17. Zhao, W.; Go, T.H. Quadcopter formation flight control combining MPC and robust feedback linearization. *J. Frankl. Inst.* **2014**, *351*, 1335–1355. [CrossRef]

18. Dai, L.; Cao, Q.; Xia, Y.; Gao, Y. Distributed MPC for formation of multi-agent systems with collision avoidance and obstacle avoidance. *J. Frankl. Inst.* **2017**, *354*, 2068–2085. [CrossRef]
19. Rahman, M.; Avelin, A.; Kyprianidis, K. An approach for feedforward model predictive control of continuous pulp digesters. *Processes* **2019**, *7*, 602. [CrossRef]
20. Zhao, S.; Cajo, R.; De Keyser, R.; Ionescu, C.M. The potential of fractional order distributed MPC applied to steam/water loop in large scale ships. *Processes* **2020**, *8*, 451. [CrossRef]
21. He, A.; Wang, C.; Liu, S.; Meehan, P.A. Switched model predictive path control of incremental sheet forming for parts with varying wall angles. *J. Manuf. Process.* **2020**, *53*, 342–355. [CrossRef]
22. Ellis, M.; Durand, H.; Christofides, P.D. A tutorial review of economic model predictive control methods. *J. Process Control* **2014**, *24*, 1156–1178. [CrossRef]
23. Adeodu, O.; Omell, B.; Chmielewski, D.J. On the theory of economic MPC: ELOC and approximate infinite horizon EMPC. *J. Process Control* **2019**, *73*, 19–32. [CrossRef]
24. Liu, S.; Liu, J. Economic model predictive control with extended horizon. *Automatica* **2016**, *73*, 180–192. [CrossRef]
25. Grüne, L. Economic receding horizon control without terminal constraints. *Automatica* **2013**, *49*, 725–734. [CrossRef]
26. Vaccari, M.; Pannocchia, G. A modifier-adaptation strategy towards offset-free economic MPC. *Processes* **2016**, *5*, 2. [CrossRef]
27. Suwartadi, E.; Kungurtsev, V.; Jäschke, J. Sensitivity-based economic NMPC with a path-following approach. *Processes* **2017**, *5*, 8. [CrossRef]
28. González, A.H.; Odloak, D. A stable MPC with zone control. *J. Process Control* **2009**, *19*, 110–122. [CrossRef]
29. Graciano, J.E.A.; Jäschke, J.; Le Roux, G.A.; Biegler, L.T. Integrating self-optimizing control and real-time optimization using zone control MPC. *J. Process Control* **2015**, *34*, 35–48. [CrossRef]
30. Capron, B.D.O.; Odloak, D. An extended Linear Quadratic Regulator with zone control and input targets. *J. Process Control* **2015**, *29*, 33–44. [CrossRef]
31. Ferramosca, A.; Limon, D.; González, A.H.; Odloak, D.; Camacho, E.F. MPC for tracking zone regions. *J. Process Control* **2010**, *20*, 506–516. [CrossRef]
32. Guan, S.; Wu, X.; Wu, Z. Model predictive zone control with soft constrained appending margin. *Asian J. Control* **2021**, *23*, 2776–2785. [CrossRef]
33. Wan, X.; Luo, X.L. Economic optimization of chemical processes based on zone predictive control with redundancy variables. *Energy* **2020**, *212*, 118586. [CrossRef]
34. Wan, X.; Liu, B.J.; Liang, Z.S.; Luo, X.L. Switch approach of control zone and optimization zone for economic model predictive control. *Chem. Eng. Trans.* **2017**, *61*, 181–186.
35. Sun, L.; Zha, X.; Luo, X. Coordination between bypass control and economic optimization for heat exchanger network. *Energy* **2018**, *160*, 318–329. [CrossRef]
36. Rawlings, J.B.; Angeli, D.; Bates, C.N. Fundamentals of economic model predictive control. In Proceedings of the 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), IEEE, Maui, HI, USA, 10–13 December 2012; pp. 3851–3861.
37. Lyapunov, A.M. The general problem of the stability of motion. *Int. J. Control* **1992**, *55*, 531–534. [CrossRef]
38. Gilbert, E.G. Controllability and observability in multivariable control systems. *J. Soc. Ind. Appl. Math. Ser. A Control* **1963**, *1*, 128–151. [CrossRef]
39. Kalman, R.E. Lectures on controllability and observability. In Proceedings of the Centro Internazionale Matematico Estivo Seminar on Controllability and Observability, Bologna, Italy, 1–9 July 1968; number NASA-CR-113869.
40. Kerrigan, E.C.; Maciejowski, J.M. Invariant sets for constrained nonlinear discrete-time systems with application to feasibility in model predictive control. In Proceedings of the IEEE Conference on Decision and Control, IEEE, Sydney, Australia, 12–15 December 2000; Volume 5, pp. 4951–4956.
41. Boyd, S.; El Ghaoui, L.; Feron, E.; Balakrishnan, V. *Linear Matrix Inequalities in System and Control Theory*; SIAM: Philadelphia, PA, USA, 1994.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

An Integrated Approach of Fuzzy AHP-TOPSIS for Multi-Criteria Decision-Making in Industrial Robot Selection

Ngoc-Tien Tran ¹, Van-Long Trinh ¹ and Chen-Kuei Chung ^{2,*}

¹ School of Mechanical and Automotive Engineering, Hanoi University of Industry, No. 298 Cau Dien Street, Bac Tu Liem District, Hanoi 100000, Vietnam; tientn@hau.edu.vn (N.-T.T.); longtv@hau.edu.vn (V.-L.T.)

² Department of Mechanical Engineering, National Cheng Kung University, Tainan 701, Taiwan

* Correspondence: ckchung@mail.ncku.edu.tw

Abstract: In recent times, industrial robots have gained immense significance and popularity in various industries. They not only enhance labor safety and reduce costs but also greatly improve productivity and efficiency in the production process. However, selecting the most suitable robot for a specific production process is a complex task. There are numerous criteria to consider, often conflicting with each other, making decision-making challenging. In order to tackle this problem, the multi-criteria decision-making (MCDM) method is employed, which aids in ranking decisions based on criteria weights. However, traditional MCDM methods are now considered outdated, and researchers are concentrating on hybrid models that include multiple MCDM techniques to tackle decision-making problems effectively. This study presents an effective MCDM model that integrates Fuzzy-AHP-TOPSIS to evaluate and choose the best robot. The Fuzzy-AHP is utilized to establish a set of weights for the evaluation criteria. Subsequently, the proposed technique analyzes, prioritizes, and chooses the best robot option from the ranking list for the factory. The experimental results demonstrate that by employing the integrated fuzzy analytical hierarchy process, taking into account parameter weights and expert judgment, the robots are identified in order of best to worst alternatives to factories. The outcomes of this research possess significant implications for robot selection and can be applied in various fields to cater to production requirements.

Keywords: industrial robots; MCDM; fuzzy-AHP; fuzzy-TOPSIS

Citation: Tran, N.-T.; Trinh, V.-L.; Chung, C.-K. An Integrated Approach of Fuzzy AHP-TOPSIS for Multi-Criteria Decision-Making in Industrial Robot Selection. *Processes* **2024**, *12*, 1723. <https://doi.org/10.3390/pr12081723>

Academic Editors: Sheng Du, Li Jin, Xiongbo Wan and Zixin Huang

Received: 15 July 2024

Revised: 10 August 2024

Accepted: 14 August 2024

Published: 16 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The advancement of technology in the 4.0 industrial revolution has unlocked remarkable opportunities for heightened automation in production [1,2]. It is crucial to address the imperative of researching and implementing industrial robots to substitute human labor in production processes [3–5]. The utilization of robots in diverse industries such as automobile manufacturing, electronics, food and beverages, healthcare, and services has demonstrated notable impacts [6–8]. Especially in the manufacturing industry, the introduction of various robot types has resulted in a wide range of attributes, including functions, technical specifications, load capacity, speed, and price [9,10]. This diversity has presented challenges in choosing the best robot for a factory. The decision-making process for choosing the ideal robot aims not only to achieve cost effectiveness and efficient production but also to optimize other aspects of the production process, such as labor safety, productivity, product quality, space and resource optimization, flexibility, and reduced production time [11,12]. Making errors in decision-making regarding the selection of industrial robots can impact a factory's ability to compete in the market for both productivity and product quality [13]. Consequently, choosing the appropriate robot for a certain industry application and production environment has become a complex challenge, particularly given the multitude of robot types available in the market. Decision makers must take into account subjective and objective factors, as well as the benefits and costs associated with

each option [14,15]. In practice, the criteria for selecting industrial robots often conflict with each other, have different units of measurement, and require trade-offs, making it challenging to compare and make decisions [16–18]. To address this issue, researchers have put forth various methods and models for robot selection [19,20]. These models encompass computer-aided approaches, statistical methods, and optimization models for enhancing production system performance [21–23]. In 1991, the authors proposed a regression model to identify robots that outperformed others in terms of manufacturer specifications at a given cost [24]. Diagram and matrix methods have been utilized since 2006 to effectively compare and identify robots, as well as to store and retrieve robot data for various applications [25]. While these methods address the issue of robot selection, they may lack flexibility when dealing with complex variables, and the models can become intricate and confusing, particularly in large-scale factories. Conversely, determining the optimal robot involves a decision-making process that takes into account a number of competing subjective and objective factors, resembling an MCDM problem. MCDM techniques have proven to be valuable in handling such complicated problems, and researchers have employed various MCDM methods to tackle robot selection challenges [26–29]. The authors of one study employed the MCDM technique combined with the weighting technique for decision-making in the powder-mixed electrical discharge machining process [30]. The evaluation based on distance from average solution (EDAS) technique was effectively used to handle the robot selection problem [31]. Another MCDM technique, known as Analytic Hierarchy Process (AHP) secondary analysis, incorporates both subjective and objective criteria to make robot selections [32]. However, the traditional MCDM methods may not align with reality as decision-making for each option depends on the evaluator's subjective opinion and related assessments, which are often vague and imprecise [33,34]. In many cases, accurately determining ratings and weights for performance is challenging. To address this issue, fuzzy set theory was created to represent uncertainties in predictions, human perception, and other factors. This led to the creation of Fuzzy MCDM (FMCDM) techniques [35]. Researchers have utilized FMCDM methods to tackle this problem [36–38]. The first people to introduce decision-making procedures in fuzzy contexts were Bellman and Zadeh (1970) [39]. Generally, a fuzzy function defines a fuzzy number, where each value in the set is assigned a membership degree ranging from 0 to 1 [40]. Octagonal fuzzy numbers are often considered the optimal solution for addressing load transmission problems in fuzzy environments [41]. The triangular fuzzy number (TFN), which represents the decision maker's status in complex problems, can be an effective means of conveying information [42]. In MCDM models, fuzzy numbers are employed to manage the evaluator's subjectivity and uncertainty. The novel hexagonal fuzzy approximation's characteristics are examined, and a group MCDM issue using index matrices is used to show the practicality of the proposed method [43]. Multi-criteria selections are made using fuzzy numbers in MCDM approaches as Fuzzy-TOPSIS (Fuzzy Technique for Order Preference by Similarity to Ideal Solution), Fuzzy-AHP, Fuzzy-MOORA, etc. [44–47]. In order to prioritize the order for multi-criteria assessment of industrial robot systems, Cengiz Kahraman developed a fuzzy hierarchy approach based on the TOPSIS model [48]. In [49], the authors utilized the Fuzzy-AHP technique to achieve optimal robot selection. The utilization of fuzzy numbers in MCDM multi-criteria techniques offers a more objective, multi-perspective, and realistic assessment when considering criteria for selecting the optimal solution.

Several new MCDM models have been developed by researchers, which improve making choices, accuracy and strategy [50–53]. MCDM approaches are becoming more and more popular because of their capacity to evaluate and contrast many options. Meanwhile, traditional MCDM methods are progressively going out of style. For complicated decision-making scenarios, a single MCDM tool is often insufficient [54,55]. In order to achieve more effective decision-making, it is necessary to integrate two or more MCDM models together, in addition to combining MCDM with other methods [56,57]. The primary objective of merging these techniques is to leverage the advantages of each tool and overcome the drawbacks of individual models. To predict the ideal replacement robot, Goswani

et al. [58] used a unique hybrid MCDM model that incorporates COPRAS and ARAS. For robot choosing, the authors integrated FAQT-2 and concluded that the suggested hybrid MCDM approach is more dependable and consistent compared to the traditional MCDM method [59]. Table 1 highlights a selection of notable research studies in the current body of literature pertaining to the evaluation of robots.

Table 1. Reference list for MCDM studies for robotics evaluation.

MCDM Method	Criteria for Evaluating Robots	Results	Reference
Entropy, TOPSIS	Mechanical Weight, Repeatability, Payload, Maximum Reach, Average Power Consumption.	The study determined that Robot-7 is the optimal selection for arc welding tasks. This robot has a mechanical weight of 501 kg, a repeatability of 0.15 mm, a load capacity of 6 kg, a maximum reach of 4368 mm, and a power consumption of 2.5 kW.	[60]
BW, EDAS	Load Capacity, Repeatability, Velocity Ratio, Degree of Freedom.	The proposed method offers several advantages, including increased consistency and reduced computational requirements.	[61]
EDAS	Purchase Cost, Load Capacity, Repeatability, Man–Machine Interface, Man–Machine Interface, Vendor’s Service Contract.	Compared with other MCDM methods (such as AHP, TOPSIS, VIKOR, ELECTRE, PROMETHEE, MOORA, WASPAS, GRA, ROV, and OCRA), the EDAS method is simpler and easier to apply in selecting industrial robots.	[62]
TOPSIS-ARAS, COPRAS-ARAS	Load Capacity, Repeatability Error, Handling Coefficient, Velocity, Cost.	Based on the evaluation, Robot-12 achieved the highest rating and was identified as the optimal choice. This study validates the effectiveness of the hybrid models TOPSIS-ARAS and COPRAS-ARAS in enhancing the accuracy of rankings and delivering consistent and dependable results in the selection of industrial robots.	[63]
SAW, TOPSIS, LINMAP, VIKOR, ELECTRE-III and NFM	Load Capacity, Repeatability Error, Cost, Vendor’s Service Quality, Programming Flexibility.	In case study 1, Robot 2 and Robot 3 emerged as the top choices for pick-and-place tasks. Case study 2 revealed that Robot 1 and Robot 3 received the highest ratings among the considered robots.	[64]
CRITIC, MABAC	Load Capacity, Memory Capacity, Manipulator Reach, Maximum Tip Speed, Repeatability.	In case study 3, Robot 2 was consistently identified as the optimal selection among the four robots considered by most MCDM methods. Robot R3 attained the highest ranking, signifying its suitability for pick-and-place operations in flexible manufacturing systems. Among the evaluated robots, Robot R1 received the lowest ranking. The study also conducted a comparison of the ranking results with other MCDM methods to validate the accuracy and reliability of the proposed method.	[65]
QFD, MPR	Payload Capacity, Workspace, Accuracy, Repeatability, Life Expectancy, Programmable Flexibility, Safety and Security, Purchase Cost, Maintenance Cost, Operation Cost.	The key criterion in the selection of an industrial robot is load capacity, and the most critical technical requirement is the drive system.	[66]
COCOSO, TOPSIS, VIKOR, MOORA	Load Capacity, Repeatability, Maximum Tip Speed, Memory Capacity, Manipulator Reach.	According to the COCOSO method, R3 emerges as the best robot based on the MW, SD, and CRITIC weight distribution methods. However, R1 is considered the best robot according to the EM method, and R3 is favored according to the AHP method.	[67]

Table 1. Cont.

MCDM Method	Criteria for Evaluating Robots	Results	Reference
SWARA, CoCoSo	Payload, Mechanical weight, Repeatability, Reach, Cost, Power Consumption.	The Fanuc P-350iA/45 robot has been selected as the most suitable robot for painting applications. These results have also been compared and cross-referenced with other popular MCDM methods such as TOPSIS, VIKOR, COPRAS, PROMETHEE, and MOORA, demonstrating a high degree of similarity in the ranking patterns among these methods, affirming the effectiveness of the SWARA-CoCoSo method.	[68]
Rough-MABAC	Payload, Horizontal Reach, Vertical Reach, Repeatability, Weight, Power Rating, Cost, Flexibility, Safety, Welding Performance, Maintainability, Ease of Programming.	The research findings indicate that Robot A6 is the most suitable choice, ranking at the top of the list, followed by Robots A3, A13, A10, A5, A9, A4, A11, A1, A14, A7, A12, A8, and finally Robot A2. The robots are categorized into two main groups, efficient and inefficient, based on their positions in the approximate boundary regions.	[69]
PIPRECIA-TOPSIS	Payload, Weight of Robot, Repeatability, Reach.	The PIPRECIA technique identifies payload as the most crucial criterion based on a predefined priority order, and the TOPSIS method recommends the FANUC 100iD/10L model as the best arc welding robot.	[70]
BWM, G-BWM	Velocity, Repeatability, Load Capacity, Cost, Quality, Memory Capacity, Manipulator Reach.	The results indicate that Robot 2 is the best robot. The G-BWM (group best–worst method) demonstrates greater effectiveness compared to the G-AHP (Group Analytic Hierarchy Process) method due to its lower overall violation and deviation, as well as requiring fewer comparisons, resulting in reduced computational requirements.	[71]
MCGDM-IP	Cost, Handling Coefficient, Load Capacity, Repeatability, Velocity.	Robot R11 achieved the highest ranking among the evaluated robots, while Robot R4 received the lowest ranking. The MCGDM-IP method improved the satisfaction level of the group by 2.12% compared to the simple additive weighting (SAW) method.	[72]
CODAS, COPRAS, COCOSO, MABAC, VIKOR	Payload, Speed, Reach, Mechanical Weight, Repeatability, Cost, Power Consumption.	The results indicate that the HY1010A-143 robot is evaluated as the most suitable for painting applications according to four out of the five methods used. The KF121 robot is evaluated as the least suitable for painting applications by all of the MCDM methods.	[73]
AHP	General Criteria, Structure/Architecture Criteria, Reliability Criteria, Application Criteria, Performance Criteria, Safety Criteria.	The AHP method is applied to evaluate the cobots based on the predefined criteria. The cobot with the highest overall priority weight (A1) is considered the most suitable based on the given criteria and AHP evaluation.	[74]
WSM, WPM, WASPAS, MOORA, MULTIMOORA	Load Capacity, Maximum Tip Speed, Repeatability, Memory Capacity, Manipulator Reach.	The results indicate that among the applied MCDM methods, the MULTIMOORA (MOORA with Complete Multiplicative Form) method is the most robust and less affected by changes in the criteria weights. The robot ranking results show that the Cybotech V15 Electric Robot (R3) is often the best choice in most of the methods.	[75]
COPRAS	Repeatability Error, Load Capacity, Maximum Tip Speed, Memory Capacity, Manipulator Reach.	The Cincinnati Milacrone T3-726 Robot (A2) achieved the highest ranking with a Qi value of 0.1946 and a Ui value of 100.00, securing first position. The COPRAS method has been demonstrated to be effective in the evaluation and selection of industrial robots, aligning well with the results from previous studies.	[76]

Table 1. Cont.

MCDM Method	Criteria for Evaluating Robots	Results	Reference
AHP	Load Capacity, Reach, Weight, Repeatability, Power Consumption, Dexterity, Service	Based on the AHP method, the robot structure R2 is selected as the most optimal choice.	[77]
GRA	Load Capacity, Repeatability Error, Velocity Ratio, Degrees of Freedom.	Robot R3 achieved the highest score with a grey relational grade of 0.9434 and was ranked first.	[78]
AHP	Technical Criteria: Movement, Shaft Speed, Reach, Repeatability, Allowable Moment, Load: Robot Mass, Robot Reach, Vertical Reach, Horizontal reach Other Criteria: Capacity, Cost, Flexibility, Mounting Type, Welding Type.	Among the analyzed 15 industrial robots, the robot with code A4 achieved the highest weight of approximately 16%, followed by A5 with approximately 15%, and A2 and A9 both scoring $\approx 10\%$. Robot A4 excelled in criteria such as repeatability (C1.2), robot weight (C2.2), and power (C3.1), obtaining the highest score in these aspects.	[79]

We introduced the hybrid approach of Fuzzy TOPSIS and Fuzzy AHP, which has not been previously explored in research on robot selection. Our results demonstrated the effectiveness of this evaluation method and the article not only brings practical value to robot selection in manufacturing but also contributes to the knowledge base of MCDM methods, particularly the combination of fuzzy AHP and fuzzy TOPSIS in modern production environments. In this study, the criterion weights are determined using the Fuzzy-AHP, while the robot alternatives are evaluated and ranked using the Fuzzy-TOPSIS. Specifically, we employ the Fuzzy-AHP-TOPSIS combination model to rank eight different robots based on their attributes, as evaluated by experts. The alternatives are organized in order of increasing closeness to both positive and negative ideal solutions. To address the uncertainty and subjectivity of the evaluators, we incorporate fuzzy numbers into the model. By integrating the Fuzzy-AHP-TOPSIS model, we successfully selected the optimal robot for our factory with a high degree of reliability. Moving forward, we plan to enhance this hybrid MCDM model by incorporating additional techniques and methods to ensure the best possible decision-making outcomes.

2. Materials and Methods

2.1. Fuzzy Numbers

Fuzzy set theory is employed to address uncertainty stemming from imprecision or ambiguous information. According to this theory, an ordered pair collection (X being a subset of the real numbers \mathbb{R}) is referred to as a fuzzy set ($F = \{(\psi, \mu_F(\psi)) | \psi \in X\}$). A membership function called $\mu_F(\psi)$ and $\mu_F(\psi)$ gives each element a number between 0 and 1. The pairwise comparison matrices of the AHP integrate fuzzy set theory. The triangular fuzzy number (TFN) is commonly utilized to represent the judgments of experts and is denoted by $F(\underline{f}, f, \bar{f})$. The parameters $F(\underline{f}, f, \bar{f})$ represent the minimum, intermediate (i.e., most favorable), and maximum values used to quantify uncertain judgments. The following determines the TFN's membership function:

$$\mu_F(\psi) = \begin{cases} 0, & \psi < \underline{f} \\ (\psi - \underline{f})(\bar{f} - \underline{f})^{-1}, & \underline{f} \leq \psi < f \\ (\bar{f} - \psi)(\bar{f} - f)^{-1}, & f \leq \psi \leq \bar{f} \\ 0, & \psi > \bar{f} \end{cases} \quad (1)$$

A collection of criteria, represented as $C = \{c_1, c_2, c_3, \dots, c_i\}$, and a set of alternatives, represented as $A = \{A_1, A_2, A_3, \dots, A_j\}$, are taken into consideration while making a multi-criteria choice. Every alternative is analyzed using a specified set of criteria. The selection of each criterion is followed by the analysis of its utilization level for each alternative

A_j . Thus, the value range analysis for each criterion can be expressed as $F_{A_i}^1, F_{A_i}^2, \dots, F_{A_i}^m$ ($i = 1, 2, 3, \dots, n$), where $F_{A_i}^j$ ($j = 1, 2, 3, \dots, m$) represents the TFNs.

Utilizing fuzzy numbers facilitates the management of ambiguity and uncertainty within expert evaluations. By employing fuzzy numbers, experts can articulate their judgments using qualitative terms like “low”, “medium”, and “high”, which can then be translated into fuzzy numbers. This approach offers greater flexibility compared to traditional methods that demand precise score values, particularly beneficial when experts are uncertain about their assessments. Additionally, fuzzy numbers enable a more nuanced aggregation of diverse judgments, capturing the breadth of expert opinions and evaluations, a capability that conventional methods may lack. For instance, in a traditional scenario, an expert might assign a score of 3 out of 10 for the importance of the criterion “Cost”. In contrast, employing fuzzy numbers, the expert could rate this criterion as “Medium High”, resulting in a conversion to a triangular fuzzy number (2, 3, or 4). This method enhances the accuracy of representing the uncertainties and ambiguities inherent in the expert’s evaluation.

2.2. Fuzzy AHP

The AHP is a MCDM technique used in order to establish priorities among various criteria. It involves making pairwise comparisons between the criteria and alternatives, which are utilized to calculate the weights used to rank the alternatives (Figure 1).

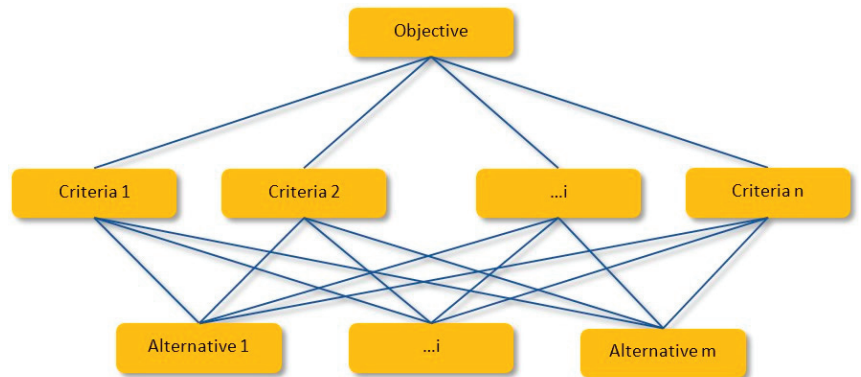


Figure 1. Hierarchical structure diagram.

However, in real-life decision-making scenarios, making clear and accurate comparisons can be challenging due to the presence of imprecision and subjectivity. Moreover, traditional AHP may not fully capture human reasoning and accurately represent expert opinions when comparing alternatives. The Fuzzy-AHP is an extension of the traditional AHP method that incorporates fuzzy number theory into its framework. This approach addresses the limitations of the conventional AHP by allowing decision makers to communicate their assessments utilizing linguistic variables or fuzzy numbers. By considering uncertainty in decision criteria and alternatives, the Fuzzy-AHP method facilitates a more flexible and diverse decision-making process. It is a widely employed method in the field of MCDM [80–83]. The Fuzzy-AHP approach makes use of a fuzzy pairwise comparison matrix. The priority weights are obtained by solving a fuzzy linear equation system. These resulting weights are then used to rank the alternatives based on their overall scores. In the Fuzzy-AHP method, the weight vector is determined by following these steps. $F = (\tilde{f})_{n \times m} = (\underline{f}_{ij}, f_{ij}, \bar{f}_{ij})_{n \times m}$ is a fuzzy pairwise comparison matrix:

Step 1: Calculate the fuzzy aggregation range. For each object, the fuzzy synthetic extent value is computed as follows:

$$U_i = \sum_{j=1}^m F_{A_i}^j \otimes \left(\sum_{i=1}^n \sum_{j=1}^m F_{A_i}^j \right)^{-1} = \left(\sum_{j=1}^m \underline{f}_j, \sum_{j=1}^m f_j, \sum_{j=1}^m \bar{f}_j \right) \otimes \left(\sum_{i=1}^n \left(\sum_{j=1}^m \underline{f}_j, \sum_{j=1}^m f_j, \sum_{j=1}^m \bar{f}_j \right) \right)^{-1} \tag{2}$$

We calculate the fuzzy set value as follows:

$$\sum_{i=1}^n \left(\sum_{j=1}^m \underline{f}_j, \sum_{j=1}^m f_j, \sum_{j=1}^m \bar{f}_j \right) = \left(\sum_{i=1}^n \underline{f}_i, \sum_{i=1}^n f_i, \sum_{i=1}^n \bar{f}_i \right) \tag{3}$$

Thus, Equation (2) becomes

$$U_i = \sum_{j=1}^m F_{A_i}^j \otimes \left(\sum_{i=1}^n \sum_{j=1}^m F_{A_i}^j \right)^{-1} = \left(\sum_{j=1}^m \underline{f}_j, \sum_{j=1}^m f_j, \sum_{j=1}^m \bar{f}_j \right) \otimes \left(\frac{1}{\sum_{i=1}^n \underline{f}_i}, \frac{1}{\sum_{i=1}^n f_i}, \frac{1}{\sum_{i=1}^n \bar{f}_i} \right) \tag{4}$$

Step 2: The degree of possibility of $F_2 = (\underline{f}_2, a_2, \bar{f}_2) \geq F_1 = (\underline{f}_1, a_1, \bar{f}_1)$ is defined as follows:

$$V(F_2 > F_1) = height(A_1 \cap A_2) = \begin{cases} 1 & f_2 \geq \bar{f}_1 \\ 0 & \underline{f}_1 \leq \bar{f}_2 \\ \frac{f_1 - \bar{f}_2}{(f_2 - \bar{f}_2) - (f_1 - \bar{f}_1)} & \text{otherwise} \end{cases} \tag{5}$$

To compare two fuzzy numbers F_1 and F_2 , regarding the values of $V(F_1 \geq F_2)$ or $V(F_2 \geq F_1)$, we consider the highest intersection point G between their corresponding membership functions μ_{F_1} and μ_{F_2} , with corresponding value g (Figure 2). The values of $V(F_1 \geq F_2)$ and $V(F_2 \geq F_1)$ can be calculated to compare between two fuzzy numbers of F_1 and F_2 .

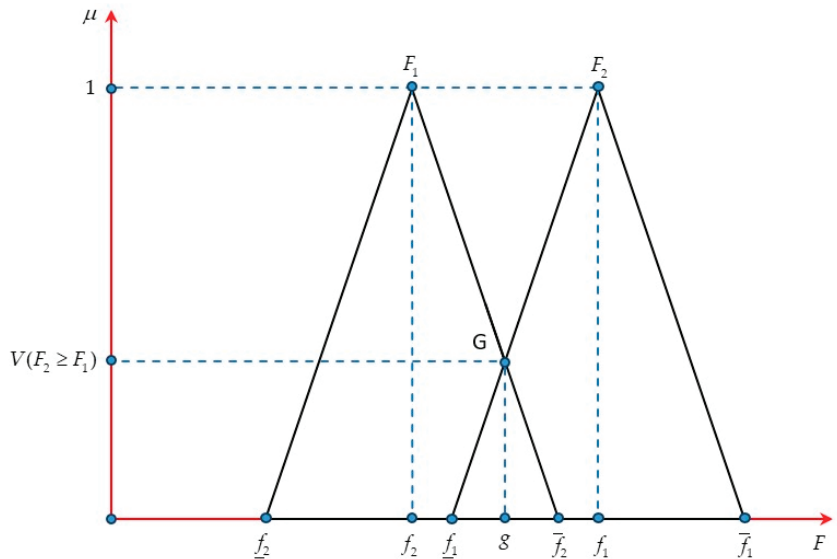


Figure 2. Value of fuzzy numbers F_1 and F_2 .

Step 3: We calculate the minimum level at which fuzzy number F is greater than fuzzy number $F_i (i = 1, 2, 3, \dots, k)$ as follows:

$$V(F \geq F_1, F_2, \dots, F_k) = \min V(F \geq F_1) \quad (6)$$

The weight vector is given by

$$W = (\min V(F \geq F_1), \min V(F \geq F_2), \dots, \min V(F \geq F_n))^T \quad (7)$$

Step 4: We calculate the normalized weight vectors as follows:

$$W = (W_1, W_2, \dots, W_3)^T \quad (8)$$

where W is a non-fuzzy number.

2.3. Fuzzy TOPSIS

The classic TOPSIS method operates under the assumption that if each local criterion increases or decreases monotonically, determining the ideal solution becomes straightforward. The ideal solution includes the highest achievable values for each local criterion, whereas the negative ideal solution includes the lowest obtainable values. To account for uncertainty and imprecision, the traditional TOPSIS approach was expanded to include a fuzzy variant [84–88]. The idea behind the Fuzzy-TOPSIS approach is that the chosen option ought to be closest to the positive ideal solution (PIS), which reduces the cost criteria and maximizes the benefit criterion, while being the furthest from the negative ideal solution (NIS). The implementation procedure for Fuzzy-TOPSIS is as follows (Figure 3):

Step 1: Determine the evaluation of criteria and alternative options. Suppose we have a decision group consisting of K individuals. The fuzzy evaluation of the C_j criterion for the A_i alternative by the k th decision maker is represented by $\tilde{x}_{ij}^k = (a_{1ij}^k, a_{2ij}^k, a_{3ij}^k)$. The weight of the C_j criterion is represented as $\tilde{w}_j^k = (w_{j1}^k, w_{j2}^k, w_{j3}^k)$.

Step 2: Determine the combined fuzzy weight for the criterion as well as the overall fuzzy ranking for the possibilities. The following approach may be used to obtain the aggregate fuzzy evaluation (abbreviated as $\tilde{x}_{ij}^k = (a_{1ij}^k, a_{2ij}^k, a_{3ij}^k)$) of the i th criteria that replaces the j th criterion:

$$\begin{cases} a_{1ij} = \min_k \{a_{1ij}^k\} \\ a_{2ij} = \frac{1}{K} \sum_{k=1}^K a_{2ij}^k \\ a_{3ij} = \max_k \{a_{3ij}^k\} \end{cases} \quad (9)$$

The aggregated fuzzy weight, $\tilde{w}_j = (w_{j1}, w_{j2}, w_{j3})$, for the C_j criterion is given by

$$\begin{cases} w_{j1} = \min_k \{w_{j1}^k\} \\ w_{j2} = \frac{1}{K} \sum_{k=1}^K w_{j2}^k \\ w_{j3} = \max_k \{w_{j3}^k\} \end{cases} \quad (10)$$

Step 3: Compute the normalized fuzzy decision matrix $\tilde{R} = [\tilde{r}_{ij}]$, in which

$$\tilde{\psi}_{ij} = \left(\frac{a_{ij}}{\hat{c}_{ij}}, \frac{b_{ij}}{\hat{c}_{ij}}, \frac{c_{ij}}{\hat{c}_{ij}} \right) \text{ and } \hat{c}_{ij} = \max_i \{c_{ij}\} \text{ (for the benefit criteria)} \quad (11)$$

$$\tilde{\psi}_{ij} = \left(\frac{\hat{a}_{ij}}{c_{ij}}, \frac{\hat{a}_{ij}}{b_{ij}}, \frac{\hat{a}_{ij}}{a_{ij}} \right) \text{ and } \hat{a}_{ij} = \min_i \{a_{ij}\} \text{ (for the cost criteria)} \quad (12)$$

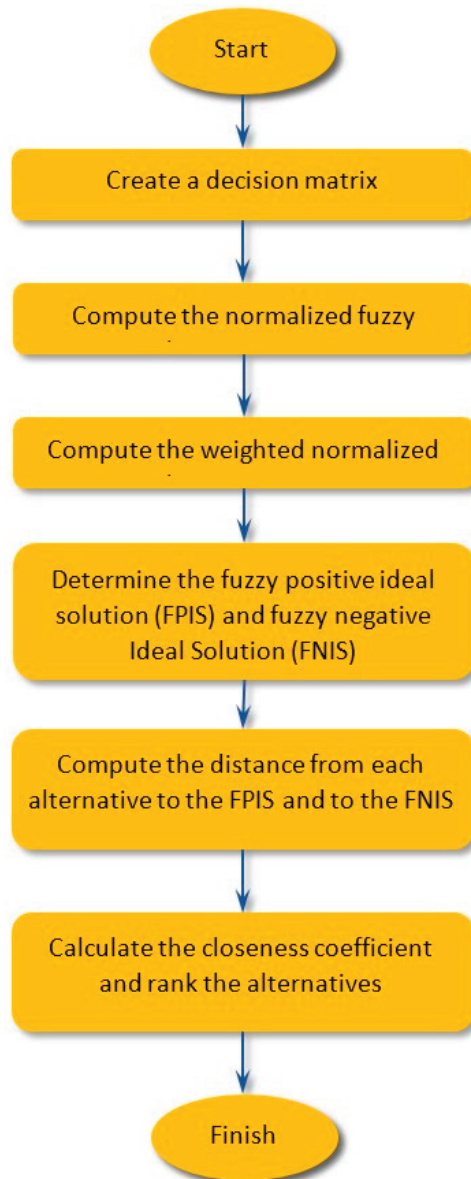


Figure 3. Flowchart of the fuzzy TOPSIS process.

Step 4: Compute the matrix \tilde{V} using the equation below:

$$\tilde{V} = \tilde{v}_{ij} = \tilde{\psi}_{ij} \times w_j \quad (13)$$

Step 5: Determine the fuzzy positive ideal solution (FPIS) and fuzzy negative ideal solution (FNIS) as described below:

$$\hat{T}_{\max} = (\hat{v}_{\max 1}, \hat{v}_{\max 2}, \hat{v}_{\max 3}, \dots, \hat{v}_{\max n}), \hat{v}_{\max j} = \max_i \{\tilde{v}_{ij}\} \quad (14)$$

$$\hat{T}_{\min} = (\hat{v}_{\min 1}, \hat{v}_{\min 2}, \hat{v}_{\min 3}, \dots, \hat{v}_{\min n}), \hat{v}_{\min j} = \min_i \{\tilde{v}_{ij}\} \tag{15}$$

Step 6: Compute the distance from each alternative to the FPIS and to the FNIS.

$$\begin{cases} d_i^+ = \sum_{j=1}^n d(\tilde{v}_{ij}, \hat{v}_{\max j}) \\ d_i^- = \sum_{j=1}^n d(\tilde{v}_{ij}, \hat{v}_{\min j}) \end{cases} \tag{16}$$

Step 7: Calculate the closeness coefficient FT_i using the following formula:

$$FT_i = \frac{d_i^-}{d_i^- + d_i^+} (i = 1, 2, 3, \dots, n) \tag{17}$$

Step 8: Rank the alternatives based on FT_i . The best choice is shown by the alternative with the highest FT_i .

3. Results and Discussion

In order to enhance the reliability of the MCDM model, we have incorporated two techniques, namely Fuzzy-AHP and Fuzzy-TOPSIS, to select the optimal robot for our factory based on specific criteria. Firstly, we utilize the Fuzzy-AHP to establish a set of weights for the evaluation criteria. Subsequently, we employ the Fuzzy-TOPSIS to assess and rank the available robot options. By employing the Fuzzy-AHP method to establish the weights for the evaluation criteria, we have achieved greater objectivity and accuracy compared to the conventional weight set determined by experts in Fuzzy-TOPSIS. The research process diagram is illustrated in Figure 4.

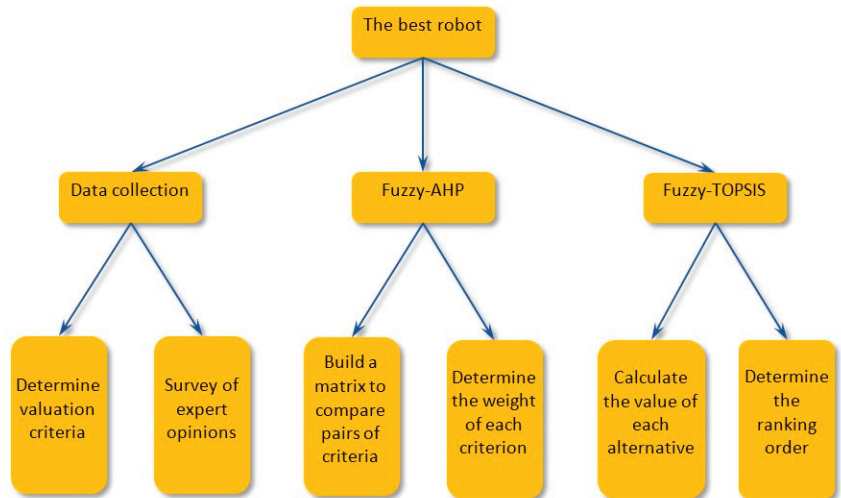


Figure 4. Components for calculating the optimal robots.

The Fuzzy-AHP-TOPSIS hybrid model is employed to assess and rank objects by evaluating criteria through a pairwise comparison matrix and ranking objects based on their proximity to positive and negative ideal solutions. Fuzzy numbers are utilized in the model to mitigate uncertainty and subjectivity associated with the evaluator. The model employs Fuzzy-AHP-TOPSIS to evaluate criteria, aiding in the selection of the optimal robot. The process involves the following steps:

Step 1: Determine evaluation criteria.

In order to make a decision in selecting the optimal robot, it is necessary to establish criteria that align with the factory's requirements. These criteria represent the specific attributes that robot manufacturers have incorporated into their products. After careful consideration of the factory's operations, we identified and defined the criteria, which are outlined in Table 2.

Table 2. List of criteria used in robot selection.

No.	Criteria	Units	Symbol
1	Mechanical Weight	Kg	MW
2	Velocity	m/s	VL
3	Payload	Kg	PL
4	Maximum Reach	Mm	MR
5	Average Power Consumption	Kw	APC
6	Cost	\$	CO

The selection of criteria for evaluating and choosing the optimal robot in a manufacturing environment is crucial for the effectiveness and accuracy of the decision-making process. The six criteria chosen in this study are Mechanical Weight, Velocity, Payload, Maximum Reach, Average Power Consumption, and Cost. These criteria were selected based on their comprehensive relevance and critical importance in determining the suitability of a robot for industrial applications. Descriptions of these criteria are as follows:

Mechanical weight pertains to the weight of the robot itself, impacting aspects such as mobility, installation prerequisites, and the structural support essential in the production setting. This weight factor influences the ease of integration into existing systems and the overall adaptability for deployment across various sectors of the factory. Lighter robots may offer easier installation and relocation, whereas their heavier counterparts might necessitate sturdier infrastructure.

Velocity denotes the rate at which a robot can execute its assigned tasks, directly affecting production efficiency and cycle durations. Enhanced speeds can enhance productivity, diminishing the time needed for each operation and augmenting the factory's output rate. This aspect holds particular significance in high-speed manufacturing scenarios where time optimization is paramount.

Payload signifies the maximum weight a robot can manage, a critical factor in ensuring its capability to execute tasks without mechanical strain. This capacity significantly influences the robot's applicability for specific tasks, especially in handling hefty materials or products during manufacturing processes. A higher load capacity empowers the robot to handle larger or heavier components, thereby amplifying its versatility.

Maximum reach denotes the farthest distance a robot can extend to execute its tasks, influencing its access to different parts of the work area. This metric determines the robot's operational range and its ability to function effectively in larger or more intricate setups. A greater reach enables the robot to cover more ground without requiring frequent repositioning, thereby enhancing its efficiency and adaptability across various applications.

Average power consumption quantifies the energy utilized by the robot during its operation, impacting operational expenses. Lower power consumption leads to reduced operational costs and contributes to a more sustainable production process. This criterion holds significance for factories striving to minimize energy usage and lower overall expenditures.

Cost encompasses the initial purchase price, installation expenses, and ongoing maintenance costs, all pivotal factors in any investment evaluation. It ensures that the investment aligns with the financial constraints of the factory and aids in comparing the economic viability of different robot options. Effectively managing costs is crucial for sustaining profitability and attaining a favorable return on investment.

Once the essential criteria were established, we conducted a thorough search to identify robots that possess the desired attributes. The robots, along with their corresponding attribute parameters, are listed in Table 3.

Table 3. Numerical data for robot selection [60].

Alternative	MW	VL	PL	MR	APC	CO
Robot 1	145	1.33	12	1441	1.0	722
Robot 2	27	1.11	8	911	0.5	485
Robot 3	170	1.26	4	1500	0.6	965
Robot 4	272	0.65	20	1650	3.4	671
Robot 5	250	0.04	25	2409	2	690
Robot 6	230	0.25	10	1925	5.6	325
Robot 7	501	1.01	6	4368	2.5	400
Robot 8	215	1.21	8	1801	5.05	690

Step 2: Survey experts' opinions.

Initially, the study constructs a fuzzy evaluation table for the weight vector. The values corresponding to the semantic level, ranging from low to high, are presented in Table 4.

Table 4. Fuzzy evaluation scores for the weight vector.

Linguistic Terms	Scale of Fuzzy Number	Units
Absolutely strong (AS)	(2, 2.5, 3)	$\tilde{9}$
Very strong (VS)	(1.5, 2, 2.5)	$\tilde{8}$
Fairly strong (FS)	(1, 1.5, 2)	$\tilde{7}$
Slightly strong (SS)	(1, 1, 1.5)	$\tilde{6}$
Equal (E)	(1, 1, 1)	$\tilde{5}$
Slightly weak (SW)	(2/3, 1, 1)	$\tilde{4}$
Fairly weak (FW)	(0.5, 2/3, 1)	$\tilde{3}$
Very weak (VW)	(0.4, 0.5, 2/3)	$\tilde{2}$
Absolutely weak (AW)	(1/3, 0.4, 0.5)	$\tilde{1}$

To ensure practical applicability, we sought expert opinions by consulting individuals who are recognized as experts in the field of industrial robots. These experts possess extensive knowledge and experience in the domain, as listed in Table 5.

Table 5. List of the experts.

Experts	Age	Education	Experience in the Field (Years)
Decision maker 1 (DM 1)	58	Associate Professor of Mechanical Engineering	>15
Decision maker 2 (DM 2)	62	Associate Professor of Robotics Engineering	>20
Decision maker 3 (DM 3)	58	Associate Professor of Manufacturing Processes	>25
Decision maker 4 (DM 4)	65	Professor of Management Science and Engineering Management	>20
Decision maker 5 (DM 5)	66	Professor of Mechatronics Engineering	>30

The aforementioned experts assessed the criteria using a fuzzy evaluation table for the weight vector, as outlined in Table 6.

Table 6. Evaluation of criteria by experts.

Criteria	DM 1	DM 2	DM 3	DM 4	DM 5
MW	FS	FS	FS	VS	FS
VL	AS	VS	VS	AS	VS
PL	AS	AS	AS	AS	AS
MR	VS	VS	VS	VS	AS
APC	VW	VW	VW	FW	VW
CO	FW	FW	FW	FW	FW

Step 3: Construct a pairwise comparison matrix for the criteria.

For each pair of criteria A and B, we establish their relationship using the value scale determined by fuzzy numbers. Experts then assess the levels of superiority and inferiority between the criteria. The semantic relationship between the evaluation criteria is depicted in Table 7.

Table 7. Relationship between two criteria according to linguistic terms.

Criteria	High Priority			Equal		Low Priority			Criteria	
A	$\tilde{9}$	$\tilde{8}$	$\tilde{7}$	$\tilde{6}$	$\tilde{5}$	$\tilde{4}$	$\tilde{3}$	$\tilde{2}$	$\tilde{1}$	B

Based on the established Fuzzy-AHP method, we have a pairwise comparison matrix between the criteria, as shown in Table 8.

Table 8. Pairwise comparison matrix between criteria.

Criteria	MW	VL	PL	MR	APC	CO
MW	(1, 1, 1)	(1, 15/14, 30/19)	(1, 15/13, 5/3)	(15/14, 30/19, 25/12)	(1, 8/5, 21/10)	(1, 11/10, 8/5)
VL	(19/30, 14/15, 1)	(1, 1, 1)	(1, 11/10, 8/5)	(1, 6/5, 17/10)	(8/5, 21/10, 13/5)	(1, 8/5, 21/10)
PL	(3/5, 13/15, 1)	(5/8, 10/11, 1)	(1, 1, 1)	(1, 15/14, 30/19)	(13/10, 9/5, 13/10)	(13/10, 9/5, 23/10)
MR	(12/25, 19/30, 14/15)	(10/17, 5/6, 1)	(19/30, 14/15, 1)	(1, 1, 1)	(1, 7/5, 19/10)	(1, 13/10, 9/5)
APC	(10/21, 5/8, 1)	(5/13, 10/21, 5/8)	(10/13, 5/9, 10/13)	(10/19, 5/7, 1)	(1, 1, 1)	(1, 7/5, 19/10)
CO	(5/8, 10/11, 1)	(10/21, 5/8, 1)	(10/23, 5/9, 10/13)	(5/9, 10/13, 1)	(10/19, 5/7, 1)	(1, 1, 1)

Step 4: Determine the fuzzy weight of each criterion.

Based on the evaluation of the criteria, we have a table of fuzzy weight values of each criterion, as seen in Table 9 below.

Table 9. Fuzzy weight value of each criterion.

Fuzzy Weight \tilde{w}_j	Value
\tilde{w}_1	(0.132, 0.2, 0.33)
\tilde{w}_2	(0.131, 0.206, 0.3190)
\tilde{w}_3	(0.121, 0.192, 0.263)
\tilde{w}_4	(0.098, 0.16, 0.247)
\tilde{w}_5	(0.085, 0.121, 0.2)
\tilde{w}_6	(0.076, 0.121, 0.195)

Step 5: Expert evaluation of alternative options based on criteria.

Initially, we have a fuzzy evaluation score table for the alternatives, which captures the expert assessments for each criterion listed in Table 10.

Table 10. Fuzzy evaluation scores for alternatives.

Linguistic Terms	Fuzzy Core
Very poor (VP)	(0, 0, 1)
Poor (P)	(0, 1, 3)
Medium poor (MP)	(1, 3, 5)
Fair (F)	(3, 5, 7)
Medium good (MG)	(5, 7, 9)
Good (G)	(7, 9, 10)
Very good (VG)	(9, 10, 10)

We consulted experts to gather their perspectives on alternative options for each criterion. Appendix A contains Tables A1–A6, which present the various issues discussed.

Step 6: Construct a decision matrix.

We establish a fuzzy decision matrix for the criteria, which summarizes the evaluations and scores for each criterion. The decision matrix for all criteria is depicted in Table 11.

Table 11. Decision matrix.

Alternative	MW	VL	PL	MR	APC	CO
Robot 1	(0, 0.6, 2.2)	(8.6, 9.8, 10)	(5.4, 7.4, 9.2)	(0, 0.8, 2.6)	(5.8, 7.8, 9.4)	(0, 0.6, 2.2)
Robot 2	(0, 0, 1)	(2.2, 4.2, 6.2)	(0, 0.8, 2.6)	(0, 0, 1)	(9, 10, 10)	(5.8, 7.8, 9.4)
Robot 3	(0, 1, 3)	(7.4, 9.2, 10)	(0, 0.6, 2.2)	(0, 1, 3)	(7.4, 9.2, 10)	(0, 0.2, 1.4)
Robot 4	(6.6, 8.6, 9.8)	(0, 0.4, 1.8)	(7.8, 9.4, 10)	(1.4, 3.4, 5.4)	(0, 1, 3)	(2.2, 4.2, 6.2)
Robot 5	(5.4, 7.4, 9.2)	(0, 0.8, 2.6)	(8.6, 9.8, 10)	(7.8, 9.4, 10)	(2.6, 4.6, 6.6)	(0.8, 2.6, 4.6)
Robot 6	(2.2, 4.2, 6.2)	(0, 0, 1)	(2.6, 4.6, 6.6)	(5, 7, 9)	(0, 0.2, 1.4)	(9, 10, 10)
Robot 7	(8.6, 9.8, 10)	(0.6, 2.2, 4.2)	(0.2, 1.4, 3.4)	(9, 10, 10)	(1.4, 3.4, 5.4)	(7.8, 9.4, 10)
Robot 8	(0.6, 2.2, 4.2)	(5.4, 7.4, 9.2)	(1, 3, 5)	(3.4, 5.4, 7.4)	(0, 0.4, 1.8)	(3.8, 5.8, 7.8)

Step 7: Compute the distance from each alternative to the FPIS and FNIS.

The values of the FPIS and FNIS are determined using Equation (16). The distances from the options to the FPIS and FNIS are presented in Table 12.

Table 12. Distance from alternatives to FPIS and FNIS.

Alternative	Robot 1	Robot 2	Robot 3	Robot 4	Robot 5	Robot 6	Robot 7	Robot 8
d_i^+	0.2656	0.4608	0.3525	0.2307	0.1653	0.3168	0.2131	0.2673
d_i^-	0.2803	0.0819	0.1953	0.3183	0.3835	0.2319	0.3301	0.2843

Step 8: Utilize the Fuzzy-TOPSIS method to evaluate the robots.

For each option A_i , we calculate a closeness coefficient FT_i , which is presented in Table 13, indicating the relative closeness of each option to the ideal solution.

Table 13. Ranking of alternatives.

Alternative	Robot 1	Robot 2	Robot 3	Robot 4	Robot 5	Robot 6	Robot 7	Robot 8
FT_i	0.5134	0.1508	0.3565	0.5798	0.6987	0.4227	0.6077	0.5154
Rank	5	8	7	3	1	6	2	4

Within the proposed model, a higher coefficient value FT_i signifies a greater preference or optimality as per the decision maker's inclinations. As per the closeness coefficients FT_i detailed in Table 13, Robot 5 emerges as the most optimal choice to meet the factory's needs, while Robot 2 is identified as the least favorable option. This ranking highlights the model's efficacy in accurately distinguishing among alternatives. The key innovation

of this method lies in its fusion of the strengths of both methodologies within a fuzzy logic framework, effectively addressing the inherent uncertainty and subjectivity in expert evaluations. Traditional methods often falter in handling the ambiguity and imprecision intrinsic to expert judgments, as they necessitate precise numerical inputs. In contrast, the fuzzy logic framework empowers experts to convey their preferences using linguistic terms that are subsequently transformed into triangular fuzzy numbers. This approach diminishes the reliance on exact figures and better encapsulates the uncertainty in expert evaluations. By employing this dual approach, the chosen robot optimizes benefits while minimizing costs and other adverse factors.

4. Conclusions

The solution to the problem of selecting optimal robots in production brought significant benefits to the factory. It not only alleviates the challenges associated with choosing the most suitable robot from a multitude of conflicting criteria but also delivers numerous advantages to the factory as a whole. Appropriately designed and selected robots can operate with enhanced efficiency, thereby increasing production output and labor productivity. In this study, the integrated Fuzzy-AHP-TOPSIS model is employed, leveraging the AHP technique to establish criteria weights and employing the TOPSIS method to evaluate and rank robot options. The proposed fuzzy solution enhances objectivity in evaluating criteria by utilizing nine fuzzy numbers for pairwise comparisons. Triangular fuzzy numbers are employed to expand the evaluation possibilities. The expert system is coordinated to construct a criteria comparison matrix, which is then used to determine the weights for the set of objective criteria. Based on the established fuzzy weight set, the TOPSIS fuzzification strategy is employed to select the option that is closest to the positive ideal solution (PIS), optimizing the benefit criterion while minimizing the cost criteria, and farthest from the negative ideal solution (NIS). The incorporation of fuzzy numbers mitigates uncertainty and subjectivity in the evaluation process, resulting in a more accurate ranking of alternatives compared to traditional MCDM methods. Based on this ranking, the optimal robot option can be selected for the factory, facilitating effective decision-making in manufacturing environments. However, it is important to note that this study has certain limitations. The number of robots included in the evaluation is relatively small. Nevertheless, this serves as a foundation for testing the Fuzzy-AHP-TOPSIS-integrated model and paves the way for its application to broader problems with a more diverse range of options. In the near future, we plan to implement this model with a larger dataset and enhance it by integrating additional MCDM methods to achieve the highest level of accuracy. Nevertheless, it is essential to acknowledge that this study has certain constraints. The evaluation includes a relatively small number of robots, potentially limiting the ability to comprehensively encompass the diversity and intricacies of available market options. Additionally, the model's current applicability is restricted to a specific set of criteria and may necessitate adjustments for various industrial contexts or additional criteria. Despite these limitations, this study establishes the groundwork for testing the Fuzzy-AHP-TOPSIS-integrated model and sets the stage for its application to broader issues with a wider array of options, such as expanding the dataset, integrating additional MCDM methods, and incorporating real-time data.

Author Contributions: Conceptualization, N.-T.T., V.-L.T. and C.-K.C.; methodology, N.-T.T., V.-L.T. and C.-K.C.; validation, N.-T.T. and V.-L.T.; formal analysis, N.-T.T. and V.-L.T.; investigation, N.-T.T., V.-L.T. and C.-K.C.; resources, N.-T.T., V.-L.T. and C.-K.C.; data curation, N.-T.T. and V.-L.T.; writing—original draft preparation, N.-T.T. and V.-L.T.; writing—review and editing, N.-T.T., V.-L.T. and C.-K.C.; visualization, N.-T.T. and V.-L.T.; supervision, C.-K.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially sponsored by the National Science and Technology Council (NSTC), Taiwan, under no. NSTC113-222-1-E006-099MY3.

Data Availability Statement: Data are presented in the coauthors' research results and schematic drawings are available on request.

Acknowledgments: The work was supported in part by School of Mechanical and Automotive Engineering, Hanoi University of Industry, Hanoi 100000, Vietnam.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

For the mechanical weight criterion, we have an expert evaluation table for the alternative options, as illustrated in Table A1.

Table A1. Expert's assessment of alternatives according to the MW.

Criteria	Alternative	DM 1	DM 2	DM 3	DM 4	DM 5
MW	Robot 1	P	P	P	VP	VP
	Robot 2	VP	VP	VP	VP	VP
	Robot 3	P	P	P	P	P
	Robot 4	G	G	G	G	MG
	Robot 5	MG	MG	MG	G	MG
	Robot 6	F	F	F	MP	MP
	Robot 7	VG	VG	VG	G	VG
	Robot 8	MP	MP	P	MP	P

For the velocity criterion (VL), we have an expert evaluation table for the alternative options, as illustrated in Table A2.

Table A2. Expert's assessment of alternatives according to the VL.

Criteria	Alternative	DM 1	DM 2	DM 3	DM 4	DM 5
VL	Robot 1	VG	VG	G	VG	VG
	Robot 2	F	F	MP	F	MP
	Robot 3	G	G	VG	G	G
	Robot 4	P	VP	VP	VP	P
	Robot 5	P	P	P	P	VP
	Robot 6	VP	VP	VP	VP	VP
	Robot 7	MP	MP	MP	P	P
	Robot 8	MG	MG	MG	G	MG

For the payload criterion, we have an expert evaluation table for the alternative options, as illustrated in Table A3.

Table A3. Expert's assessment of alternatives according to the PL.

Criteria	Alternative	DM 1	DM 2	DM 3	DM 4	DM 5
PL	Robot 1	MG	G	MG	MG	MG
	Robot 2	P	P	P	P	VP
	Robot 3	VP	VP	P	P	P
	Robot 4	G	VG	G	VG	G
	Robot 5	VG	VG	VG	VG	G
	Robot 6	F	F	F	F	MP
	Robot 7	P	P	P	MP	P
	Robot 8	MP	MP	MP	MP	MP

For the maximum reach criterion, we have an expert evaluation table for the alternative options, as illustrated in Table A4.

Table A4. Expert’s assessment of alternatives according to the MR.

Criteria	Alternative	DM 1	DM 2	DM 3	DM 4	DM 5
MR	Robot 1	P	P	P	P	VP
	Robot 2	VP	VP	VP	VP	VP
	Robot 3	P	P	P	P	P
	Robot 4	MP	MP	MP	MP	F
	Robot 5	G	G	G	MG	MG
	Robot 6	MG	MG	MG	MG	MG
	Robot 7	VG	VG	VG	VG	VG
	Robot 8	F	F	F	MG	F

For the average power consumption criterion, we have an expert evaluation table for the alternative options, as illustrated in Table A5.

Table A5. Expert’s assessment of alternatives according to the APC.

Criteria	Alternative	DM 1	DM 2	DM 3	DM 4	DM 5
APC	Robot 1	MG	G	G	MG	MG
	Robot 2	VG	VG	VG	VG	VG
	Robot 3	G	VG	G	G	G
	Robot 4	P	P	P	P	P
	Robot 5	F	F	F	F	MP
	Robot 6	VP	VP	VP	VP	P
	Robot 7	MP	MP	MP	MP	F
	Robot 8	VP	VP	VP	P	P

For the cost criterion, we have an expert evaluation table for the alternative options, as illustrated in Table A6.

Table A6. Expert’s assessment of alternatives according to the CO.

Criteria	Alternative	DM 1	DM 2	DM 3	DM 4	DM 5
CO	Robot 1	P	P	P	MP	MP
	Robot 2	MG	G	G	MG	MG
	Robot 3	VP	VP	VP	VP	P
	Robot 4	F	F	MP	F	MP
	Robot 5	MP	MP	MP	MP	P
	Robot 6	VG	VG	VG	VG	VG
	Robot 7	G	VG	G	G	VG
	Robot 8	F	F	MG	MG	F

References

- Brecher, C.; Müller, A.; Dassen, Y.; Storms, S. Automation technology as a key component of the industry 4.0 production development path. *Int. J. Adv. Manuf. Technol.* **2021**, *117*, 2287–2295. [CrossRef]
- Papulová, Z.; Gažová, A.; Šufliarský, L. Implementation of automation technologies of industry 4.0 in automotive manufacturing companies. *Procedia Comput. Sci.* **2022**, *200*, 1488–1497. [CrossRef]
- Pedersen, M.R.; Nalpanitidis, L.; Andersen, R.S.; Schou, C.; Bøgh, S.; Krüger, V.; Madsen, O. Robot skills for manufacturing: From concept to industrial deployment. *Robot. Comput.-Integr. Manuf.* **2016**, *37*, 282–291. [CrossRef]
- Dean-Leon, E.; Ramirez-Amaro, K.; Bergner, F.; Dianov, I.; Lanillos, P.; Cheng, G. Robotic technologies for fast deployment of industrial robot systems. In Proceedings of the IECON 2016—42nd Annual Conference of the IEEE Industrial Electronics Society 2016, Florence, Italy, 23–26 October 2016.
- De Backer, K.; DeStefano, T.; Menon, C.; Suh, J.R. Industrial Robotics and the Global Organisation of Production. 2018. Available online: <https://www.oecd-ilibrary.org/docserver/dd98ff58-en.pdf?expires=1723607367&id=id&accname=guest&checksum=AEE2669ACDD759EE9C1C92406FE64B4D> (accessed on 15 July 2024).
- Ogbemhe, J.; Mpofu, K.; Tlale, N.; Ramatsetse, B. Application of robotics in rail car manufacturing learning factory: A case of welding complex joints. *Procedia Manuf.* **2019**, *31*, 316–322. [CrossRef]
- Iqbal, J.; Khan, Z.H.; Khalid, A. Prospects of robotics in food industry. *Food Sci. Technol.* **2017**, *37*, 159–165. [CrossRef]

8. Tran, N.T.; Ngo, T.D.; Nguyen, D.K.; Son, P.X.; Thai, N.H. Mapping and path planning for the differential drive wheeled mobile robot in unknown indoor environments using the rapidly exploring random tree method. In *Lecture Notes in Mechanical Engineering 2021, Proceedings of the RCTEMME 2021, Hanoi, Vietnam, 10–12 December 2021*; Springer Nature: Singapore, 2021; pp. 516–527.
9. Fosch-Villaronga, E.; Drukarch, H. Accounting for diversity in robot design, testbeds, and safety standardization. *Int. J. Soc. Robot.* **2023**, *15*, 1871–1889. [CrossRef]
10. Campilho, R.D.S.G.; Silva, F.J.G. Industrial process improvement by automation and robotics. *Machines* **2023**, *11*, 1011. [CrossRef]
11. Barosz, P.; Gołda, G.; Kampa, A. Efficiency analysis of manufacturing line with industrial robots and human operators. *Appl. Sci.* **2020**, *10*, 2862. [CrossRef]
12. Amici, C.; Pellegrini, N.; Tiboni, M. The robot selection problem for mini-parallel kinematic machines: A task-driven approach to the selection attributes identification. *Micromachines* **2020**, *11*, 711. [CrossRef]
13. Rao, R.V.; Patel, B.K.; Parnichkun, M. Industrial robot selection using a novel decision making method considering objective and subjective preferences. *Robot. Auton. Syst.* **2011**, *59*, 367–375. [CrossRef]
14. Chatterjee, P.; Mondal, S.; Chakraborty, S. A comparative study of preference dominance-based approaches for selection of industrial robots. *Adv. Prod. Eng. Manag.* **2014**, *9*, 5. [CrossRef]
15. Villa Silva, A.J.; Pérez-Domínguez, L.; Martínez Gómez, E.; Luviano-Cruz, D.; Valles-Rosales, D. Dimensional analysis under linguistic pythagorean fuzzy set. *Symmetry* **2021**, *13*, 440. [CrossRef]
16. İc, Y.T.; Yurdakul, M.; Dengiz, B. Development of a decision support system for robot selection. *Robot. Comput.-Integr. Manuf.* **2013**, *29*, 142–157.
17. Choe, M.S.; Yang, W.C. A reasonable method for industrial robot selection combined with several multi-attribute decision making methods. *Int. J. Interact. Des. Manuf.* **2024**, *18*, 2481–2489. [CrossRef]
18. Khan, N.A.; Kumar, A.; Rao, N. A hybrid robot selection model for efficient decisive support system using fuzzy logic and genetic algorithm. *Int. J. Syst. Assur. Eng. Manag.* **2024**, *15*, 2120–2129. [CrossRef]
19. Garg, C.P.; Görçün, Ö.F.; Kundu, P.; Küçükönder, H. An integrated fuzzy MCDM approach based on Bonferroni functions for selection and evaluation of industrial robots for the automobile manufacturing industry. *Expert Syst. Appl.* **2023**, *213*, 118863. [CrossRef]
20. Ketipi, M.; Koulouriotis, D.; Karakasis, E. Robot evaluation and selection part b: A comparative analysis. *Int. J. Adv. Manuf. Technol.* **2014**, *71*, 1395–1417. [CrossRef]
21. Bhangale, P.P.; Saha, S.; Agrawal, V. A dynamic model based robot arm selection criterion. *Multibody Syst. Dyn.* **2004**, *12*, 95–115. [CrossRef]
22. Bader, F.; Rahimifard, S. A methodology for the selection of industrial robots in food handling. *Innov. Food Sci. Emerg. Technol.* **2020**, *64*, 102379. [CrossRef]
23. Zhang, Y.; Abdullah, S.; Ullah, I.; Ghani, F. A new approach to neural network via double hierarchy linguistic information: Application in robot selection. *Eng. Appl. Artif. Intell.* **2024**, *129*, 107581. [CrossRef]
24. Khouja, M.; Booth, D.E. A decision model for the robot selection problem using robust regression. *Decis. Sci.* **1991**, *22*, 656–662. [CrossRef]
25. Rao, R.V.; Padmanabhan, K. Selection, identification and comparison of industrial robots using digraph and matrix methods. *Robot. Comput.-Integr. Manuf.* **2006**, *22*, 373–383. [CrossRef]
26. Sen, D.K.; Datta, S.; Patel, S.K.; Mahapatra, S.S. Multi-criteria decision making towards selection of industrial robot: Exploration of promethee ii method. *Benchmarking Int. J.* **2015**, *22*, 465–487. [CrossRef]
27. Sampathkumar, S.; Augustin, F.; Narayanamoorthy, S.; Ahmadian, A.; Ferrara, M.; Kang, D. Centroid and graded mean ranking methods for intuitionistic trapezoidal dense fuzzy set to solve mcdm problems of robot selection. *Int. J. Fuzzy Syst.* **2024**, 1–24. [CrossRef]
28. Athawale, V.M.; Chakraborty, S. A comparative study on the ranking performance of some multi-criteria decision-making methods for industrial robot selection. *Int. J. Ind. Eng. Comput.* **2011**, *2*, 831–850. [CrossRef]
29. Fu, Y.; Li, M.; Luo, H.; Huang, G.Q. Industrial robot selection using stochastic multicriteria acceptability analysis for group decision making. *Robot. Auton. Syst.* **2019**, *122*, 103304. [CrossRef]
30. Bui, H.A.; Tran, N.T.; Nguyen, D.L. Multi-criteria decision making in the powder-mixed electrical discharge machining process based on the cocoso, spotis algorithms and the weighting methods. *Int. J. Mod. Manuf. Technol.* **2023**, *15*, 69–79. [CrossRef]
31. Garg, H.; Sharaf, I.M. A new spherical aggregation function with the concept of spherical fuzzy difference for spherical fuzzy EDAS and its application to industrial robot selection. *Comput. Appl. Math.* **2022**, *41*, 212. [CrossRef]
32. Goh, C.-H. Analytic hierarchy process for robot selection. *J. Manuf. Syst.* **1997**, *16*, 381–386. [CrossRef]
33. Bairagi, B.; Dey, B.; Sarkar, B.; Sanyal, S. Selection of robotic systems in fuzzy multi criteria decision-making environment. *Int. J. Comput. Syst. Eng.* **2015**, *2*, 32–42. [CrossRef]
34. Hung, C.-Y.; Chang, P.Y.; Huang, Y.-H. Comparison of fuzzy-based mcdm and non-fuzzy mcdm in setting a new fee schedule for orthopedic procedures in taiwan’s national health insurance program. In Proceedings of the 8th WSEAS International Conference on APPLIED MATHEMATICS, Tenerife, Spain, 16–18 December 2005; Volume 3, pp. 321–332.
35. Dubois, D.; Prade, H. Fuzzy numbers: An overview. In *Readings in Fuzzy Sets for Intelligent Systems*; Morgan Kaufmann: Cambridge, MA, USA, 1993; pp. 112–148. [CrossRef]

36. Petrović, G.; Mihajlović, J.; Čojbašić, Ž.; Madić, M.; Marinković, D. Comparison of three fuzzy mcdm methods for solving the supplier selection problem. *Facta Univ. Ser. Mech. Eng.* **2019**, *17*, 455–469. [CrossRef]
37. Kahar, N. Comparative study of smart and fmcmm methods in smartphone selection decision support system. *Int. J. Image Graph. Signal Process.* **2021**, *13*, 1. [CrossRef]
38. Chen, L.; Pan, W. Review fuzzy multi-criteria decision-making in construction management using a network approach. *Appl. Soft Comput.* **2021**, *102*, 107103. [CrossRef]
39. Bellman, R.E.; Zadeh, L.A. Decision-making in a fuzzy environment. *Manag. Sci.* **1970**, *17*, B-141–B-164. [CrossRef]
40. Kumar, R.; Khepar, J.; Yadav, K.; Kareri, E.; Alotaibi, S.D.; Viriyasitavat, W.; Gulati, K.; Kotecha, K.; Dhiman, G. A systematic review on generalized fuzzy numbers and its applications: Past, present and future. *Arch. Comput. Methods Eng.* **2022**, *29*, 5213–5236. [CrossRef]
41. Gurukumaresan, D.; Duraisamy, C.; Srinivasan, R. Optimal solution of fuzzy transportation problem using octagonal fuzzy numbers. *Comput. Syst. Sci. Eng.* **2021**, *37*, 415–421. [CrossRef]
42. Wang, F. Preference degree of triangular fuzzy numbers and its application to multi-attribute group decision making. *Expert Syst. Appl.* **2021**, *178*, 114982. [CrossRef]
43. Nayagam, V.L.G.; Murugan, J. Hexagonal fuzzy approximation of fuzzy numbers and its applications in mcdm. *Complex Intell. Syst.* **2021**, *7*, 1459–1487. [CrossRef]
44. Liu, Y.; Eckert, C.M.; Earl, C. A review of fuzzy ahp methods for decision-making with subjective judgements. *Expert Syst. Appl.* **2020**, *161*, 113738. [CrossRef]
45. Stanković, M.; Stević, Ž.; Das, D.K.; Subotić, M.; Pamučar, D. A new fuzzy marcos method for road traffic risk analysis. *Mathematics* **2020**, *8*, 457. [CrossRef]
46. Makkı, A.A.; Abdulaal, R.M.S. A hybrid MCDM approach based on Fuzzy MEREC-G and Fuzzy RATMI. *Mathematics* **2023**, *11*, 3773. [CrossRef]
47. Gürbüz, F.; Erdinç, G. Selecting the best hotel using the Fuzzy-Moora method with a new combined weight approach. In Proceedings of the International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey, 19–21 October 2018; pp. 1–8.
48. Kahraman, C.; Kaya, I.; Evik, S.; Ates, N.Y.; Gülbay, M. Fuzzy multi-criteria evaluation of industrial robotic systems using topsis. In *Fuzzy Multi-Criteria Decision Making: Theory and Applications with Recent Developments*; Springer: Boston, MA, USA, 2008; pp. 159–186.
49. Kutlu Gündoğdu, F.; Kahraman, C. Spherical fuzzy analytic hierarchy process (AHP) and its application to industrial robot selection. In *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making, Proceedings of the INFUIS 2019 Conference, Istanbul, Turkey, 23–25 July 2019*; Springer International Publishing: Cham, Switzerland, 2020; pp. 988–996.
50. Tan, T.; Mills, G.; Papadonikolaki, E.; Liu, Z. Combining multi-criteria decision making (mcdm) methods with building information modelling (bim): A review. *Autom. Constr.* **2021**, *121*, 103451. [CrossRef]
51. Shyr, H.-J.; Shih, H.-S. A hybrid mcdm model for strategic vendor selection. *Math. Comput. Model.* **2006**, *44*, 749–761. [CrossRef]
52. Hu, S.-K.; Lu, M.-T.; Tzeng, G.-H. Exploring smart phone improvements based on a hybrid mcdm model. *Expert Syst. Appl.* **2014**, *41*, 4401–4413. [CrossRef]
53. Wu, H.-Y.; Chen, J.-K.; Chen, I.-S.; Zhuo, H.-H. Ranking universities based on performance evaluation by a hybrid mcdm model. *Measurement* **2012**, *45*, 856–880. [CrossRef]
54. Zarei, E.; Ramavandi, B.; Darabi, A.H.; Omidvar, M. A framework for resilience assessment in process systems using a fuzzy hybrid mcdm model. *J. Loss Prev. Process Ind.* **2021**, *69*, 104375. [CrossRef]
55. Roshanaei, V.; Vahdani, B.; Mousavi, S.M.; Mousakhani, M.; Zhang, G. Cad/cam system selection: A multi-component hybrid fuzzy mcdm model. *Arab. J. Sci. Eng.* **2013**, *38*, 2579–2594. [CrossRef]
56. Li, H.; Wang, W.; Fan, L.; Li, Q.; Chen, X. A novel hybrid mcdm model for machine tool selection using fuzzy dematel, entropy weighting and later defuzzification vikor. *Appl. Soft Comput.* **2020**, *91*, 106207. [CrossRef]
57. Alizadeh, R.; Soltanisehat, L.; Lund, P.D.; Zamanisabzi, H. Improving renewable energy policy planning and decision-making through a hybrid mcdm method. *Energy Policy* **2020**, *137*, 111174. [CrossRef]
58. Goswami, S.S.; Behera, D.K. Developing fuzzy-AHP-integrated hybrid mcdm system of copras-aras for solving an industrial robot selection problem. *Int. J. Decis. Support Syst. Technol. (IJDSST)* **2023**, *15*, 1–38. [CrossRef]
59. Soltan, H.; Janada, K.; Omar, M. FAQT-2: A customer-oriented method for MCDM with statistical verification applied to industrial robot selection. *Expert Syst. Appl.* **2023**, *226*, 120106. [CrossRef]
60. Chodha, V.; Dubey, R.; Kumar, R.; Singh, S.; Kaur, S. Selection of industrial arc welding robot with TOPSIS and Entropy MCDM techniques. *Mater. Today: Proc.* **2022**, *50*, 709–715. [CrossRef]
61. Rashid, T.; Ali, A.; Chu, Y.-M. Hybrid BW-EDAS MCDM methodology for optimal industrial robot selection. *PLoS ONE* **2021**, *16*, e0246738. [CrossRef]
62. Yalçın, N.; Uncu, N. Applying EDAS as an applicable MCDM method for industrial robot selection. *Sigma J. Eng. Nat. Sci.* **2019**, *37*, 779–796.
63. Goswami, S.S.; Behera, D.K.; Afzal, A.; Razak Kaladgi, A.; Khan, S.A.; Rajendran, P.; Subbiah, R.; Asif, M. Analysis of a robot selection problem using two newly developed hybrid MCDM models of TOPSIS-ARAS and COPRAS-ARAS. *Symmetry* **2021**, *13*, 1331. [CrossRef]

64. Shanmugasundar, G.; Kalita, K.; Ćep, R.; Chohan, J.S. Decision models for selection of industrial Robots—A comprehensive comparison of multi-criteria decision making. *Processes* **2023**, *11*, 1681. [CrossRef]
65. Ahmad, S.; BİNGÖL, S.; Wakeel, S. A hybrid multi-criteria decision making method for robot selection in flexible manufacturing system. *Middle East J. Sci.* **2020**, *6*, 68–77.
66. Büyükoçkan, G.; Ilıcak, Ö.; Feyzioğlu, O. An integrated QFD approach for industrial robot selection. In *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems, Proceedings of the IFIP WG 5.7 International Conference, APMS 2021, Nantes, France, 5–9 September 2021*; Proceedings, Part III, 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 561–570.
67. Eisa, A.; Fattouh, M. The significance of weighting in multicriteria decision-making methods: A case study on robot selection. *ERJ. Eng. Res. J.* **2023**, *46*, 339–352. [CrossRef]
68. Kumar, V.; Kalita, K.; Chatterjee, P.; Zavadskas, E.K.; Chakraborty, S. A SWARA-CoCoSo-based approach for spray painting robot selection. *Informatica* **2022**, *33*, 35–54. [CrossRef]
69. Agarwal, S.; Chakraborty, S.; Prasad, K.; Chakraborty, S. A rough multi-attributive border approximation area comparison approach for arc welding robot selection. *Jordan J. Mech. Ind. Eng.* **2021**, *15*, 169–180.
70. Kumar, V.; Gupta, P.; Mohata, A.; Das, P.P. Selection of industrial arc welding robot using integrated PIPRECIA-TOPSIS model. In *AIP Conference Proceedings 2023*; AIP Publishing: Melville, NY, USA, 2023; Volume 2786.
71. Ali, A.; Rashid, T. Best–worst method for robot selection. *Soft Comput.* **2021**, *25*, 563–583. [CrossRef]
72. Zhao, J.; Sui, Y.; Xu, Y.; Lai, K. Industrial robot selection using a multiple criteria group decision making method with individual preferences. *PLoS ONE* **2021**, *16*, e0259354. [CrossRef]
73. Shanmugasundar, G.; Sapkota, G.; Ćep, R.; Kalita, K. Application of MEREC in multi-criteria selection of optimal spray-painting robot. *Processes* **2022**, *10*, 1172. [CrossRef]
74. Mecheri, S.S.; Greene, C.M. Collaborative robot selection using analytical hierarchical process. *Int. J. Rapid Manuf.* **2019**, *8*, 326–344. [CrossRef]
75. Karande, P.; Zavadskas, E.; Chakraborty, S. A study on the ranking performance of some MCDM methods for industrial robot selection problems. *Int. J. Ind. Eng. Comput.* **2016**, *7*, 399–422. [CrossRef]
76. Mondal, S.; Kuila, S.; Singh, A.K.; Chatterjee, P. A complex proportional assessment method-based framework for industrial robot selection problem. *Int. J. Res. Sci. Eng.* **2017**, *3*, 368–378.
77. Breaz, R.E.; Bologa, O.; Racz, S.G. Selecting industrial robots for milling applications using AHP. *Procedia Comput. Sci.* **2017**, *122*, 346–353. [CrossRef]
78. Muduli, K.; Pumwa, J.; Yadav, D.K.; Kumar, R.; Tripathy, S. A grey relation approach for selection of industrial robot. In *Proceedings of the 2018 International Conference on Information Technology (ICIT) 2018*, Bhubaneswar, India, 19–21 December 2018; pp. 141–144.
79. Simion, M.; Socaciu, L.; Giurgiu, O.; PETRIȘOR, S.M. The selection of industrial robots for military industry using AHP method: A case study. *Acta Tech. Napoc.-Ser. Appl. Math. Mech. Eng.* **2018**, *61*. Available online: <https://atna-mam.utcluj.ro/index.php/Acta/article/view/987> (accessed on 9 August 2024).
80. Fedai, Y. Optimization of drilling parameters in drilling of mwcnt-reinforced gfrp nanocomposites using fuzzy AHP-weighted Taguchi-based MCDM methods. *Processes* **2023**, *11*, 2872. [CrossRef]
81. Luyen, L.A.; Thanh, N.V. Logistics service provider evaluation and selection: Hybrid SERVQUAL-FAHP-TOPSIS model. *Processes* **2022**, *10*, 1024. [CrossRef]
82. Hien, D.N.; Thanh, N.V. Optimization of cold chain logistics with fuzzy MCDM model. *Processes* **2022**, *10*, 947. [CrossRef]
83. Zhu, X.; Xiao, Y.; Xiao, G.; Deng, X. Research on driving factors of collaborative integration implementation of lean-green manufacturing system with industry 4.0 based on Fuzzy AHP-DEMATEL-ISM: From the Perspective of Enterprise Stakeholders. *Processes* **2022**, *10*, 2714. [CrossRef]
84. Alhakami, W. Computational study of security risk evaluation in energy management and control systems based on a fuzzy MCDM method. *Processes* **2023**, *11*, 1366. [CrossRef]
85. Alojaiman, B. A multi-criteria decision-making process for the selection of an efficient and reliable IOT application. *Processes* **2023**, *11*, 1313. [CrossRef]
86. Liu, R.; Solangi, Y.A. An analysis of renewable energy sources for developing a sustainable and low-carbon hydrogen economy in China. *Processes* **2023**, *11*, 1225. [CrossRef]
87. Abdullah, F.M.; Al-Ahmari, A.M.; Anwar, S. An integrated fuzzy dematel and fuzzy TOPSIS method for analyzing smart manufacturing technologies. *Processes* **2023**, *11*, 906. [CrossRef]
88. Thanh, N.V. A dynamic decision support system for sustainable supplier selection under fuzzy environment. *Processes* **2022**, *10*, 1576. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

AI-Based Integrated Smart Process Sensor for Emulsion Control in Industrial Application

Inga Burke ^{1,*}, Sven Salzer ², Sebastian Stein ³, Tom Olatomiwa Olakunle Olusanya ¹, Ole Fabian Thiel ¹ and Norbert Kockmann ^{1,*}

¹ Laboratory of Equipment Design, Department of Biochemical and Chemical Engineering, TU Dortmund University, Emil-Figge-Straße 68, 44227 Dortmund, Germany

² Ark Vision Systems GmbH & Co. KG, Limburger Straße 51, 35799 Merenberg, Germany; sven.salzer@ark-vision-systems.com

³ SystemKosmetik Produktionsgesellschaft für Kosmetische Erzeugnisse mbH, Raiffeisenstraße 2, 86692 Münster, Germany; sebastian.stein@systemkosmetik.de

* Correspondence: inga.burke@tu-dortmund.de (I.B.); norbert.kockmann@tu-dortmund.de (N.K.)

Abstract: In industry, reliable process supervision is essential to ensure efficient, safe, and high-quality production. The droplet size distribution represents a critical quality attribute for emulsification processes and should be monitored. For emulsion characterization, image-based analysis methods are well-known but are often performed offline, leading to a time-delayed and error-prone process evaluation. The use of an integrated smart process sensor to characterize the emulsification process over time enables the real-time evaluation of the entire system. The presented integrated smart process sensor consists of an optical measurement flow cell built into a camera system. The overall system is placed in a bypass system of a production plant for emulsification processes. AI-based image evaluation is used in combination with a feature extraction method (You Only Look Once version 4 (YOLOv4) and Hough circle (HC)) to characterize the process over time. The sensor system is installed in the plant and tested with different cosmetic products. Various iteration, prototyping, and test steps for the final sensor design are performed prior to this in a laboratory test setup. The results indicate robust and accurate detection and determination of the droplet size in real time to improve product control and save time. For benchmarking the integrated smart process sensor, the results are compared with common analysis methods using offline samples.

Keywords: emulsification process; camera system; flow measurement cell; AI-based image analysis; process monitoring; integrated process sensor

Citation: Burke, I.; Salzer, S.; Stein, S.; Olusanya, T.O.O.; Thiel, O.F.;

Kockmann, N. AI-Based Integrated Smart Process Sensor for Emulsion Control in Industrial Application.

Processes **2024**, *12*, 1821. <https://doi.org/10.3390/pr12091821>

Academic Editors: Sheng Du, Li Jin, Xiongbo Wan and Zixin Huang

Received: 31 July 2024

Revised: 19 August 2024

Accepted: 23 August 2024

Published: 27 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In industrial production, process monitoring is crucial to ensure efficient, safe, and high-quality processes. Continuous monitoring ensures early detection and correction of deviations, thus securing high product quality as well as high efficiency and economic production [1,2]. In the food, chemical, and cosmetic industries, emulsions are often involved, and the accurate estimation of the droplet concentration and size distribution is crucial for the stability of the emulsion. The morphology and droplet size distribution significantly affect this stability. Consequently, a thorough understanding of the emulsification conditions and the resulting droplet characteristics is crucial for product assessment and process control. In particular, for the emulsification process, the droplet size distribution emerges as a critical quality attribute that requires continuous monitoring [2–4]. Established methods for determining the droplet size distribution, such as manual image analysis or light diffraction, are often time-consuming or labor-intensive, which is not practical for many applications. Specifically, the quality evaluation of highly dispersed products can be challenging and the implementation of real-time analysis is difficult [5–9]. Nevertheless, optical image analysis has proven to be a cost-effective and easy-to-use method that

can extract features with high spatial and temporal resolution. The detection of such features enables the determination of droplet sizes and droplet size distributions [4,10,12–14]. Image characteristics such as contrast, edges, and color (grey-scale) are used for this detection, resulting in distinctive features. Offline analysis of process images has some limitations in particular. These include the lack of representativeness due to sampling, resulting in a temporal offset, as well as a location dependence of the measurement due to the sampling point. In the worst-case scenario, this leads to a poor representation of the overall system and possibly reduced product quality and lower process control due to a delay in time [5,6,15]. Another challenge in the optical image analysis of emulsification processes is the segmentation of overlapping droplets [6,9,16,17]. Real-time online analysis offers many advantages over sampling, including reduced risk of contamination and the ability to integrate into the production process by using real-time data to adjust production parameters [5–7]. In addition, automation and standardization of optical analysis reduce the number of sources of error, resulting in a more robust and representative analysis while reducing costs. Since sampling and analysis are performed simultaneously, time-dependent effects such as coalescence and droplet agglomeration do not occur [8,18–20].

Automated online analysis methods are increasingly popular in the process industry, and their potential can also be used to determine characteristic emulsion parameters [4,8,15,21]. New potentials arise due to more efficient methods and progress in artificial intelligence (AI) research and development [4,22]. The combination of traditional measurement instruments with AI algorithms that feature increasingly powerful network architectures enables the processing of large amounts of data with high precision in a short time [14,15,23–29]. New network architectures offer opportunities for real-time analysis and, consequently, process analysis, especially in the field of image-based process evaluation [22]. The use of smart camera systems enables advanced, precise, and fast evaluation and characterization of processes. These integrated sensors enable users to influence process control and result in potentially more efficient process management. These so-called intelligent or smart sensors are attracting growing attention in pharmaceutical and chemical process engineering [18,30]. The development of an innovative, noninvasive optical smart sensor is inspired by the design thinking method [31], wherein a final classification of the sensor system is completed based on the Technology Readiness Level (TRL) [32]. Design thinking is a creative approach to problem solving that aims to develop innovative solutions that meet user needs. This approach is user-centered and iterative, involving interdisciplinary teams to understand and solve complex problems [31]. These are defined in advance and are focused on in this publication as the real-time characterization of industrial emulsification processes. To evaluate the sensor design and development level, the Technology Readiness Level is used. This framework ensures systematic advancement and validation and helps to refine the sensor design at each stage, ensuring it meets industry standards and user requirements. The TRL is a scale (1–9) to assess the level of development of new technologies based on a systematic analysis. It was developed in 1988 by NASA for the evaluation of space technologies and has become an evaluation standard for other areas of technology [32].

Abidin et al. [8] presents a comprehensive overview of the existing measurement techniques for emulsion characterization. These are divided into in situ and external methods, making a distinction between indirect and direct methods. Since the focus of this work is on real-time monitoring and process control, the in situ image analysis methods such as particle vision microscope (PVM), stereo-microscope, and endoscope are of special interest [5,8]. A PVM is an invasive measurement technique that is used to characterize particulate systems. The PVM probe uses six near-IR lasers focused through a hexagonal lens array to illuminate particles, with back-scattered light collected and relayed to a charge-coupled device (CCD) array, capturing digital images. This illumination enhances the grey-scale structure, aiding with fully automatic image recognition [8]. In stereo-microscopy, the process images are captured with an external camera. This is in sync with the flashing in the vessel. Measurements are limited to droplets near the wall of the vessel and can be affected by the curvature

of the wall. Depending on the camera used, blurring and low contrast may occur, so the image evaluation may be subject to errors [8]. For in situ image analysis, an endoscope with a camera is inserted into the dispersion. The camera captures images of droplets near the glass window of the endoscope, allowing real-time measurement of droplet sizes, especially in areas with high breakage rates, such as the impeller zone. Sharp images are ensured by integrating a light source into the endoscope with a fiber-optic-guided stroboscopic flash. A cover tube at the tip of the endoscope prevents droplet interference and allows precise analysis of droplet size [8]. SOPAT, for example, designs and develops photo-optical image-based measurement systems to characterize multi-phase systems. Depending on the probe system, which is based on the endoscope technique, particles with size ranges of 1.5 to 7700 μm can be analyzed [33]. In addition to the usage of probe systems, measurement flow cells provide the opportunity for the characterization of multi-phase systems [16,34–36]. Schmalenberg et al. [36] presented a temperature-controlled minichannel flow-cell for non-invasive particle measurements in solid–liquid flow. Burke et al. [16] designed a 3D-printed modular optical flow cell for image-based droplet size measurements in emulsification processes.

In order to combine the advantages of image-based methods and automated evaluation methods based on AI, in this contribution, the measurement cell presented in [16] is further developed and integrated into a smart camera system. In general, the optical sensor system should provide optical accessibility to the industrial emulsification process and combine it with a direct evaluation system that allows real-time process monitoring. Therefore, AI-based image analysis is combined with a new methodology for non-invasive droplet size measurement and integrated into an emulsification process. In addition to that, the developed camera setup, including its processor board, image sensor board, and optics, is presented and combined with the optical measurement flow cell. To enable process evaluation, integration of the final system into the process plant is necessary. Together, these steps enable the evaluation of emulsions in a bypass system and provide a measurement system for droplet size characterization within an industrial process plant. For the integration of a sensor system as well as for its development, different key aspects are considered and discussed in this work. In particular, this work aims to develop, validate, and integrate an optical smart sensor to automate the analysis and control of an industrial emulsification process and enable real-time analysis of industrial emulsions. For this, an AI-based object detection approach is transferred to an edge device and its usage in industrial processes is evaluated. Embedding the smart sensor in the system and evaluating the process on an edge device constitutes an innovative approach that guarantees straightforward installation and replacement.

2. Materials and Methods

The development of a smart optical sensor involves several design steps, including optical accessibility to the emulsion, the development of a camera system that provides suitable resolution, an automated online method for the determination of the emulsion droplet size, and the integration of the measurement system into an industrial production plant. This section deals with the sensor development strategy and the industrial plant in which the smart optical sensor is integrated. The working principle and the evaluation strategy of AI-based optical detection are explained here, and the development and application of an analytical strategy is described.

2.1. Sensor Development Strategy

The development of an applicable smart measurement system in the form of an optical sensor to monitor emulsification processes is based on several requirements. These include:

- Optical access to the product, allowing a large number of different formulations to be optically resolved and evaluated;
- No influence on process performance due to the integrated intelligent sensor;
- Simple integration into the industrial plant and user-friendliness;

- Robust design.

The sensor development strategy requires an iterative approach based on the design thinking process. Parts of this process use a rapid prototyping method using additive manufacturing to enable rapid adaptation. This iterative workflow is visualized in Figure 1.

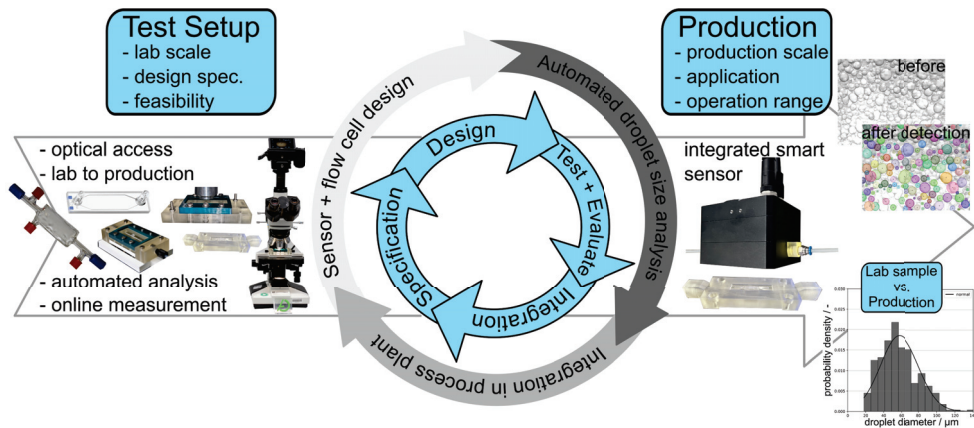


Figure 1. Schematic overview of the iterative sensor development strategy, including the optimization steps for sensor development and plant integration. The strategy includes the progress from the idea and requirements definition over the design of a test setup to the integration of the sensor system into the production plant. This requires continuous validation and evaluation of the state of the system. In particular, the measurement flow cell design, the components and composition of the camera parts, the integration of the measurement flow cell into the camera, the evaluation option of the resulting droplets, and the integration of the system were decisive [37,38].

Firstly, critical process conditions and quality attributes were identified, and specifications for the smart optical sensor were defined. Based on the available process information, a laboratory-scale test setup was constructed to represent the industrial process. A description of the test setup is given in [16]. The design of an optical flow measurement cell was iteratively developed and evaluated in the test setup to allow optical accessibility to the emulsion [16]. The focus is on optical accessibility and the use of optical methods to evaluate the images of the resulting emulsion. Here, a general proof of the concept is shown, and it builds the foundation for the next step of the camera integration. Based on these results and the sensor requirements, the key aspects were defined:

- Optical sensor and flow cell design;
- Automated droplet size analysis;
- Integration into the production plant.

For the transition of the optical evaluation of the emulsification process from the laboratory to the production scale, optimization of the individual key aspects is performed. For this purpose, we adapt the measurement flow cell presented in [16], and the development and integration into the camera system are iteratively advanced. In this optimization framework, the periphery and the requirements are defined, and the design of the prototype is adapted, tested, and integrated. The optimization of the other main aspects is done simultaneously.

The automated analysis method for process characterization includes AI-based droplet detection. The first optimization steps are presented in [17]. The implementation of AI-based evaluation requires some adaptations to execute this method on the embedded sensor board of the camera system. The integration of the complete measurement system into the production plant is the last step to transfer the measurement principle from the laboratory

to the production plant. The prototype solution is evaluated by comparing its performance with the originally used evaluation method.

2.2. Industrial Emulsification Process

The smart optical process sensor is integrated into an existing industrial production plant for emulsions that focus on cosmetic and medical products, as illustrated in Figure 2. This plant has a total volume of 4 tons and is equipped with a temperature control system in the form of a double jacket. A bypass pipe is integrated at the lowest point of the stirred vessel. The bypass consists of an DN 80 pipe section, to which the disperser (Unimix S-Jet, EKATO HOLDING GmbH, Freiburg, Germany) is connected. At this point, a toothed-ring disperser performs both the conveying through the bypass and the dispersion of the emulsion ingredients. After the disperser, a steel recirculation pipe is transferred to a hose section then reverts to a steel pipe and is fed back into the vessel. This reduces the vibrations caused by the emulsification process.

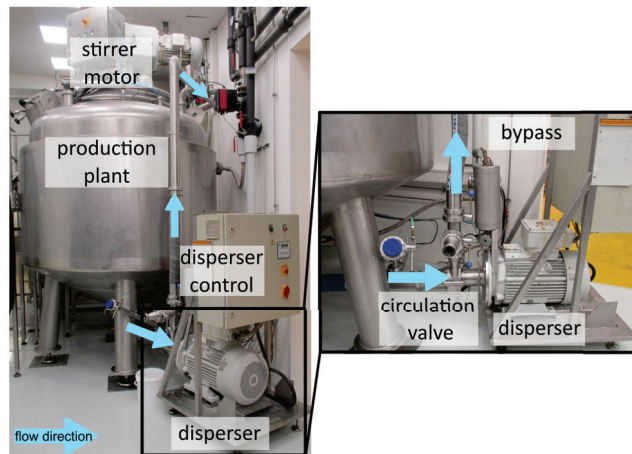


Figure 2. Image of the industrial plant for emulsification processes, including a closer view of the disperser unit and the flow direction of the emulsion, at SystemKosmetik Produktionsgesellschaft für kosmetische Erzeugnisse mbH, Münster, Bavaria, Germany.

Depending on the different formulations, the emulsion process proceeds as follows. First, the water and oil phases are heated externally and fed to the heated vessel after reaching the process temperature. Then, the emulsification process starts by starting the disperser. Once a pre-emulsion has been achieved, additional components such as emulsifiers, additives, etc., are added, and the emulsification process continues. As soon as the emulsion has the desired properties, the process temperature is decreased, the emulsion is cooled down, and the filling process is initiated. The entire system is then cleaned.

2.3. AI-Based Droplet Size Determination

The AI-based droplet size determination uses the object detection algorithm You Only Look Once (YOLO) version v4 (v4). This is a real-time single-stage detector that localizes and classifies objects in a single step. Based on previous contributions, this method ensures a precise and reliable determination of droplet sizes [17,37]. This approach is examined for its application in industrial emulsions. The used model is converted to enable its application to the used embedded smart sensor. The detection accuracy and detection time of the configuration for the present smart sensor is performed based on statistical evaluation of the detected droplet sizes (see Section 3.1). For detection, a confidence score (CS) of 0.6 and image input sizes of (640 × 540 pixels) are investigated to evaluate their influence on detection performance.

2.4. Validation of Droplet Size Determination

A comparative independent image analysis methodology is used to validate the AI-based droplet size determination. For this, manual image evaluation is performed using ImageJ (version 1.53k; Java 1.8.0_172) [39]. The results of this statistical evaluation are compared with the results of the AI-based image analysis to confirm its trustworthiness and accuracy. Further, a comparison of the different observed process states during an emulsification process is performed. Here, the resulting DSD captured within the optical measurement flow cell and the DSD from analyzing a corresponding sample under a microscope are compared. This analytical strategy offers the possibility of evaluating the process at different points in time and determining the progress of the process over time. The accuracy and applicability of the sensor are evaluated.

3. Results

The next section deals with the results of the design process of the integrated smart process sensor and its characterization in terms of the first test runs.

3.1. Optical Smart Process Sensor

The integrated optical smart process sensor is briefly discussed. The resulting individual components along with the design-thinking process and the final integration of the entire measurement system into the plant are evaluated. The resulting total measurement system is illustrated in Figure 3.

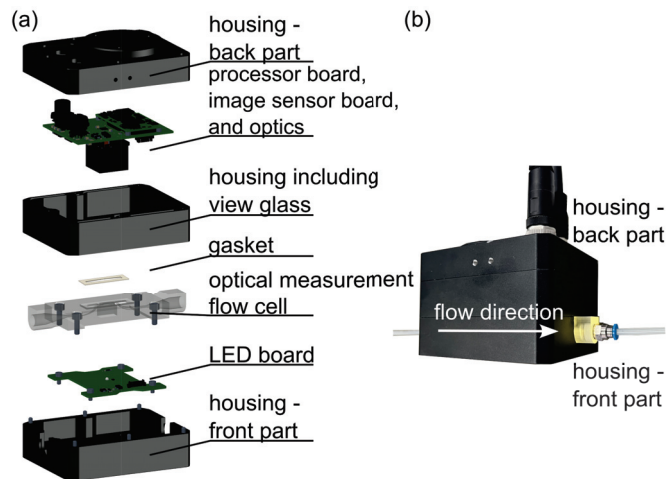


Figure 3. Explosion view of the optical sensor design, including the measurement flow cell as well as the camera setup in (a) and an image of the assembled sensor in (b).

The measurement system is organized into three key functional units:

- Camera unit, including the processor board, optical sensor, and focusable lens;
- Optical measurement flow cell;
- LED illumination unit.

The individual components of the measurement system are independent and interchangeable. This enables rapid adaptation to changes in processes such as color, turbidity, and formulation of emulsions containing different additives. The individual components and their development are described in more detail in the following.

3.1.1. Optical Measurement Flow Cell

The design of an optical flow cell for droplet size measurement in laboratory emulsification processes is described in [16]. Transferring this measurement system from laboratory to production scale and integrating the measurement cell into a camera system requires further adaptation and iteration steps, which are illustrated in Figure 4.

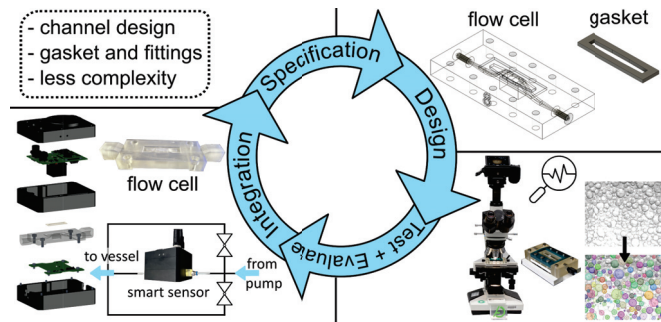


Figure 4. Schematic overview of the iterative measurement flow cell design, including the optimization steps for camera integration. The strategy includes the progress from the idea [16] and the defining of the requirements for the design of a test setup to the integration of the measurement flow cell into the production plant. This requires continuous validation and evaluation of the design and possible system integration strategies.

For this adaptation and optimization of the measurement system, the following requirements were formulated:

- Reduction of complexity;
- Reduction of components to simplify handling and minimize leakages;
- Minimization of the distance between the emulsion and the camera lens;
- Applicability for the evaluation of industrial emulsions;
- Integration into industrial system—focus on fluid connections.

The test system measurement flow cell is designed with a modular structure, allowing the adjustment of channel depth within the observation window. The developed optical flow cell showed a first feasibility in optical process accessibility and provides the opportunity for optical process evaluation. The modular structure of the flow cell results in different channel depths, as well as high complexity. The new design intends to solve these challenges by the flow cell frame being manufactured as a single part, thus reducing the complexity and the number of components. The integration of the flow cell into the developed camera setup requires a minimum distance between the measurement flow cell channel and the objective lens for focusing. The front part of the camera housing seals the measurement flow cell against the view glass with a flat gasket when fitted. The detailed housing concept of the camera is described in Section 3.1.2.

The iterative development process requires the application and evaluation of the adapted measurement cell design for industrial emulsions, which was examined in the laboratory test setup. This led to modifications to the channel width and exposure strategy. In addition, the fluidic connections for integrating the sensor system into the production plant were defined and adapted. In particular, the position and nominal width of the fluidic connections are crucial for achieving a tight, pressure-resistant system and simple handling during operation. For a deeper understanding, the key iterations of the measurement flow cell are summarized in Table A1.

The final measurement flow cell design has the following specifications. The dimension of the measurement flow cell is $85 \times 40 \times 10$ mm (length \times width \times height) with a minimal channel diameter of $d = 1.6$ mm. The fluid connections are set to $1/8''$ threads at both sides, providing the connection of the bypass hose. An expansion of the cross

section is slowly performed within the measurement cell frame. The circular cross section broadens and is changed to a rectangular cross section to allow the emulsion to flow optimally in relation to the observation window. Here, the channel depth slowly tapers from 0.25 mm to 0.025 mm, resulting in a channel depth of 0.025 mm, while the width and length of the channel are 5 mm and 3 mm, respectively, at the observation window. The optical measurement flow cell features a symmetric design, allowing the emulsion to flow in either direction. To ensure a watertight seal, a fitting gasket is printed with a dimension of $40 \times 12 \times 1.5$ mm. This gasket is designed to perfectly fit into the measurement flow cell frame. The flow cell design boasts a cutout at the bottom of the flow cell that improves translucency and provides a defined place for the LED. A sketch of the optical measurement flow cell and the gasket is shown in Figure 5. CAD sketches are provided here: <https://github.com/TUDoAD/AI-based-integrated-smart-process-sensor-for-emulsion-control-in-industrial-application> (accessed on 30 July 2024).

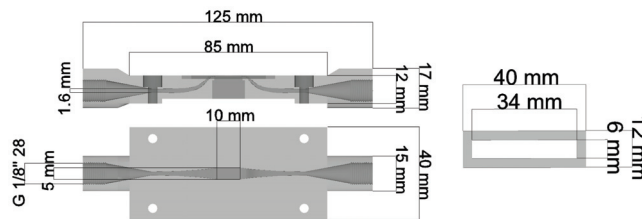


Figure 5. Technical drawing of the final design of the measurement cell frame and gasket, including important dimensions.

3.1.2. Camera System and Illumination Strategy

The camera system was developed on the basis of the results of the laboratory tests. During development, special focus was paid to the image sensor, the optics, the exposure, and the processor board. The following requirements for the integration and application of the optical measurement system for industrial emulsification processes were defined:

- Adequate magnification and visualization of the emulsion droplets;
- Definable and reproducible focusing;
- Exposure time, exposure mode, and light temperature;
- Hardware of the embedded processor for real-time evaluation;
- Housing concept and gasket.

The final setup consists of a Sony IMX327 (FRAMOS GmbH, Taufkirchen, Germany) image sensor, which is combined with a TB4M-220-4 (Lensation GmbH, Karlsruhe, Germany) lens. This combination demonstrates the basic visualization capabilities of an emulsion. The choice of lens produces a distortion-free and perspective-correct image. The selected lens has an optical magnification factor of 2.22 and focuses at a working distance of 1 to 4 mm. The image sensor was selected on the basis of resolution, light sensitivity, availability, and sensor size and is from the machine vision and automotive sectors. The specifications of the image sensor can be found in Table 1 below.

Table 1. Properties of the used image sensor for the camera system.

Properties	Sony IMX327
Resolution/pixel	1920 × 1080 (2 MP)
Sensor size/"	1/2.8
Pixel size/ μm	2.9
Max. image diagonal/mm	6.46
Shutter/-	Rolling

Initial lighting tests with a ring-shaped arrangement of eight LEDs were followed by the use of a single high-power LED with a color temperature of 5000 Kelvin (daylight) and a Color Rendering Index (CRI) CRI = 70 for the demonstrator. This was confirmed as a result of the test setup. Thus, a light capacity of a maximum of 135 lm is achieved. The LED is mounted centrally in the channel in front of the measurement flow cell for transmitted light illumination and can also be placed outside the image center for testing purposes. This also allows comparative measurements between direct and indirect illumination. A piezo-electrically adjustable focus module is used for precise electronic focusing of the lens to the desired image plane. The hardware concept of the camera is based on an NXP iMX8M Plus processor core, which is particularly suitable for an intelligent camera platform due to its MIPI/CSI2 interface to a high-resolution image sensor and its hardware blocks for graphics processing and machine learning. The selected embedded processor i.MX8M Plus already contains blocks for machine learning hardware acceleration (NPU with 2.3 TOPS) and a graphics processing unit (GPU).

The newly developed Ark Vision hardware consists of a base board, system-on-module (SOM) board, and sensor board. The baseboard contains the following components:

- Power supply unit for input voltage of 9–60 V DC;
- Interfaces to sensor board and electronic focus module;
- Ethernet PHY;
- USB interface;
- Digital inputs and outputs for triggering and signaling;
- LED driver for controlling the lighting.

The SOM board is equipped with the NXP embedded processor iMX8M Plus, while the sensor board has the image sensor and a mounting option for lens mounts or electronic focus modules.

The components of the demonstrator board are installed in a housing. The housing components are manufactured using a 3D-printing rapid prototyping process for verification purposes, allowing quick changes and adjustments after construction or testing. The housing concept is adapted constructively to ensure that the components used fit the modular concept of the demonstrator. The back part (see Figure 3), which serves as a base for the main board, including the SOM board and the sensor board, is used for this purpose. The focus module is placed on the sensor board. The middle section is placed on top, which closes off the electronics with a sealed glass pane. The front part with the integrated measurement flow cell and LED lighting forms the final part. The measurement flow cell presented in Section 3.1.1 is inserted into the demonstrator case. The challenge is to seal all components, which is achieved with two circumferential tongue-and-groove seals between the three housing parts, and to seal the measurement flow cell against the camera electronics. For this purpose, the middle and back sections form a self-contained unit with a sight glass. The front part with the measurement flow cell seals against the window with a flat gasket and can be removed as a separate block for cleaning. Finally, the sealed electrical connection between the LED assembly (front part) and the main assembly (back part) through the housing walls must be considered.

3.1.3. AI-Based Evaluation for Edge Device

Based on the YOLOv4 model presented in [17], the AI-based image evaluation is examined with respect to its use on the selected i.MX8M embedded processor mentioned in Section 3.1.2. The YOLOv4 architecture is divided into a backbone, neck, and head. The backbone is in charge of feature extraction, the neck collects feature maps from the previous stage, and the head predicts different object classes and bounding boxes. Here, for the backbone, CSPDarknet53 is chosen, the neck comprises a spatial pyramid pooling (SPP) block and a path aggregation network (PAN) block, and the head uses three YOLO blocks. This is based on the originally presented YOLO model [40]. The customized and optimized model for droplet detection is presented in [17]. For this contribution, this model is transferred for use on an edge device to build a smart sensor system. The cam-

era system (see Section 3.1.2), including its processor, is used for this since the process monitoring and characterization are implemented on it. A conversion of the openCV-based model into the TensorFlow framework is necessary for the use of the hardware acceleration blocks for ML (NPU with 2.3 TOPS) and the graphics processing unit (GPU) installed on the processor [41]. The script described in [42] is used for this purpose. Besides the framework conversion—and, thus, the transfer from an openCV-based model to a TensorFlow lite (tflite) model—no further changes to the AI-model were performed. In order to ensure simple and good handling for the operator, it is necessary to create application software (App), which simply starts the camera to enable the evaluation of the droplet images.

An environment containing libraries for Python 3.10 and the Linux system must first be set up on the camera to run the image analysis on the edge device. Due to a limited main memory (2 GB RAM), the dependencies for numpy, pandas, openCV, and especially tflite libraries were reduced in order to save main memory when creating the app version for the AI-based image analysis.

Image evaluation using a tflite model also requires an image input size of around 640×640 pixels. Larger input images are scaled down to this size and vice versa, which results in a reduction in resolution and consequently a degradation in droplet detection and size determination. However, in order not to lose any image information, the input images (1920×1080 pixels) are cut into smaller image sections, resulting in a total of six sub-images for the AI-based image evaluation. The app version of the image analysis enables for simple operation. The image analysis process is started by the operator by starting the app. The required memory to load the model is around 10 MB, while the buffer memory (250 MB) for the image data in the tensor format occupies the most memory and depends on the amount of data. Also, the memory for prediction requires 10 MB. These values refer to the evaluation of five emulsion images. The location of the processing videos is defined so that the ML pipeline uses these videos for processing. The following steps are performed in a loop:

1. Reading the video and extracting one frame per second;
2. Cutting the individual images into six sub-images;
3. Optional: adjustment of contrast;
4. Starting AI-based droplet detection using YOLO and size determination using HC;
5. Saving the determined droplet diameters and statistical parameters in a .csv file;
6. Requesting if a new video is available in the folder to restart at Step 1.

In addition to the detection images, the output of the detection is the statistical evaluation of the resulting droplet sizes. This statistical evaluation enables process characterization and process control for different process time steps.

3.2. Sensor Integration

The integration of the optical sensor (see Figure 3) into the system requires several requirements that have been taken into account. These are defined below:

- Installation of a bypass system with flow control;
- Representative sampling;
- GMP-based hygiene requirements with regard to cleaning/maintenance;
- Connection to the system control unit;
- Reduction of vibrations on the measurement system as well as simple maintenance and servicing.

Based on the knowledge of the laboratory test setup, a bypass system is installed to integrate the optical sensor in the production plant. This involves a pipe system for the flow and reflux of the optical sensor system connected to the circulation pipe of the test system. The design ensures a high flow rate in the cleaning phase and has little impact on the flow of the emulsion. An influence on the resulting droplet size was not measured. The flow to the measurement system is integrated into the bypass after the flexible hose pipe to ensure that

the vibrations caused by the disperser are removed from the system and not transferred by the bypass to the optical smart process sensor. The reflux is located just before the stirred vessel. A ball valve is installed on the flow pipe to enable maintenance and intensive cleaning work and to regulate the flow during the test phase. A flexible hose system is used to connect the bypass to the smart process sensor to reduce additional vibrations.

A diaphragm pump (Midgetbox, Debem Deutschland GmbH, Holzkirchen, Germany) is installed in the feed to ensure a constant and controllable flow rate. To regulate the flows of different formulations with different viscosities through the optical sensor, a multiple bypass system was developed in which several hose connections that are closable by valves are installed in series. This multiple bypass system is shown in Figure 6. The bypass significantly reduces the flow velocity, which is necessary to enable the optical analysis of the emulsion using the optical sensor. The general setup is located close to the processing plant.

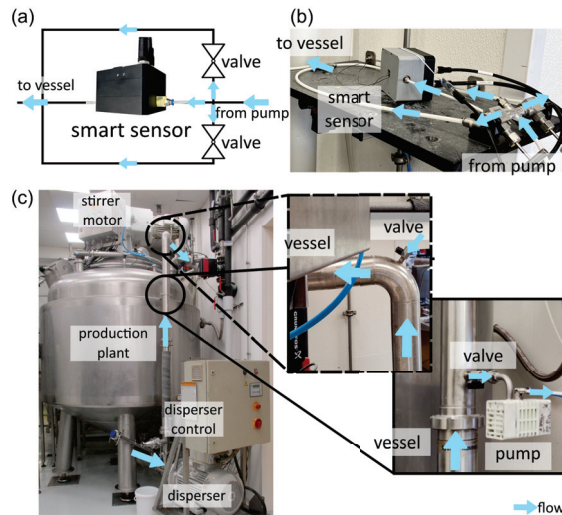


Figure 6. (a) Schematic sketch of the multiple bypass system and the integrated optical sensor and (b) an image of the bypass setup, including the optical sensor, at SystemKosmetik Produktionsgesellschaft für kosmetische Erzeugnisse mbH, Münster, Bavaria, Germany. (c) Image of the industrial plant for emulsification processes, including the position of the bypass with the integrated optical sensor.

The control unit of the system is used to supply the measurement system with power and to ensure an Ethernet connection. This is required for data transmission. In addition, the camera is connected to the bypass diaphragm pump, provides the signal to the bypass, and initiates pumping through the bypass. This signal is triggered electrically via a General Purpose Output (GPO) of the camera system. In the future, the data on the quality of the processed emulsion will be utilized to manage the entire process. This will be achieved through direct communication between the camera system and the control unit using an Ethernet connection.

3.3. Validation of Measurement System

Diverse industrial emulsions, Probio body lotion (batch number: 2329005), and collagen cream (batch number: 2335213) were measured at different process times to validate the smart process sensor described in Section 3.1. For benchmarking, the results are compared with common analysis methods using offline samples or reference values. In addition to a qualitative evaluation of the resulting emulsion images, the reproducibility of the AI-based measurement method is assessed. Droplet sizes are determined for an example emulsion at three different time steps. These time steps represent the same state during

the emulsification process, as they are recorded with a Δt of 1 s, with no direct flow in the measurement cell at the time of recording.

Figure 7a illustrates the ML-pipeline used for emulsion image processing and analysis. Here, the first step is the cutting of the 1920×1080 pixel input image into six sub-images. An additional contrast adjustment can also be performed here. Next, the system performs AI-based droplet detection using YOLOv4, size determination using HC, as well as generation of the .csv file, which contains the droplet diameters and statistical parameters. In Figure 7b, the detection of the three example images is shown, while Figure 7c illustrates the corresponding statistical evaluation. The results are presented in boxplots with additional information about the number of droplets. The images show a high-contrast representation of the emulsion, which enables an automated method for droplet size analysis. In addition, the total image is in focus, and the used telecentric objective ensures that no perspective distortion is present.

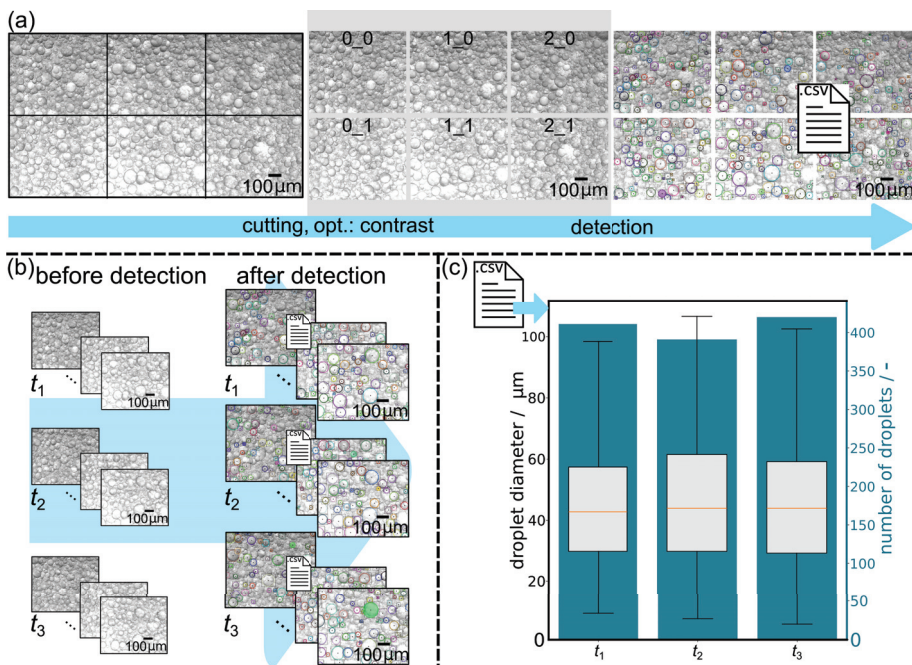


Figure 7. (a) The ML-pipeline for an example emulsion image. (b) Three examples of images, including their detection at different time steps t during an industrial emulsification process. The time steps presented show the same state during the emulsification step since they are recorded with a Δt of 1 s with no flow in the measurement cell at the time of recording. (c) The corresponding boxplots for the three different time steps. The droplet size is on the primary y-axis, while the total number of detected droplets is illustrated as bars that correspond to the secondary y-axis.

The detection results (Figure 7c) were evaluated with a confidence score of $CS = 0.6$, which was identified in previous studies [16]. Given that the three selected images show a very similar state of the emulsification process, it is expected that the number of droplets detected and the statistical evaluation will be consistent. This evaluates and validates the reproducibility of the analysis method. The number of droplets detected fluctuates between 427, 403, and 430 droplets per input image for the states t_1 , t_2 , and t_3 , respectively. Statistical analysis demonstrates an approximation of the results with a CS of 0.6. Both the median and the mean deviate by a maximum of $\Delta d_{50} = 1.95 \mu\text{m}$ and $\Delta \bar{d} = 1.09 \mu\text{m}$, respectively. These deviations are observed between the states t_1 and t_2 . When considering

the interquartile range (IQR), a maximum deviation of $3.3\ \mu\text{m}$ is again observed between the first two states. These minor fluctuations are not solely attributable to detection and size determination errors. Due to the very low flow rate and movement within the measurement cell, the three recorded states (see Figure A1) are very similar but not identical.

Additional time steps and industrial emulsions were measured in order to validate the measurement system and to evaluate the application possibilities for process monitoring and control. In addition to the validation of the resulting droplet size, the evaluation time was assessed. Figure 8 presents three distinct emulsion images captured within the measurement system and analyzed using the AI-based evaluation method implemented in the camera's processor. The resulting droplet sizes are compared with reference values from the produced batches from SystemKosmetik Produktionsgesellschaft für kosmetische Erzeugnisse mbH, Münster, Bavaria, Germany. For this comparison, samples from the same time states examined for this analysis were observed under a microscope and evaluated. In addition, a comparative analysis using ImageJ was performed, assessing at least 100 droplets per image.

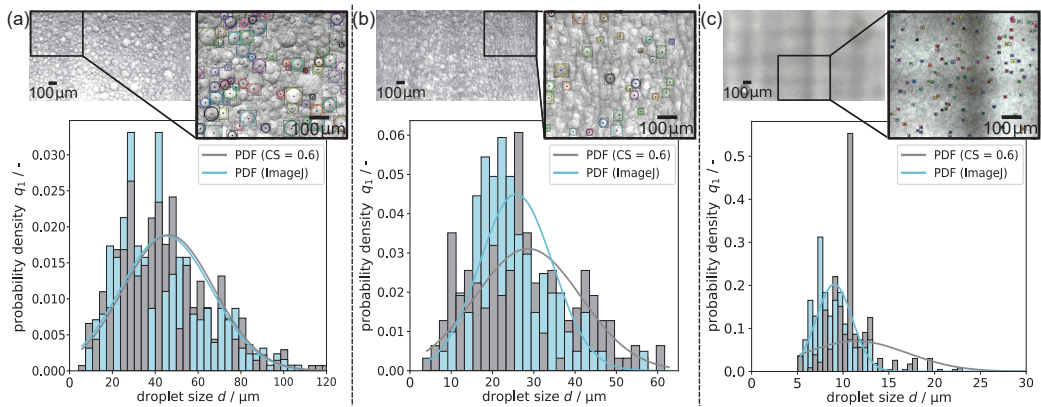


Figure 8. The statistical evaluation of various industrial emulsions. Here, the input image and an example of the sub-image after the detection are illustrated. The statistical evaluation is presented as a histogram, which shows the results of the AI-based evaluation using YOLOv4 and HC as well as the comparison results of a manual evaluation method. (a–c) Three different examples. (a,b) The same formulation (collagen cream) for different time steps and (c) a different formulation (Probio bodylotion) for final products. Dark grey: evaluation using YOLOv4 and HC with a CS of 0.6, and light blue: manual evaluation using ImageJ.

Figure 8a shows the comparison of the manual and AI-based statistical evaluation in the form of a histogram for Figure 7 time step t_3 to evaluate the trustworthiness of the AI-based droplet size evaluation. Comparative analysis of the distributions reveals a consistent trend in the DSDs. In total, 400 droplets were labeled for this comparison. The median of the distribution as determined using the AI-based method is $d_{50} = 43.55\ \mu\text{m}$, whereas the median from manual evaluation is $d_{50} = 42.41\ \mu\text{m}$, resulting in a deviation of $\Delta d_{50} = 1.14\ \mu\text{m}$. Furthermore, the mean values of both distributions were analyzed, with YOLOv4 + HC producing $\bar{d}_{\text{YOLOv4}} = 46.31\ \mu\text{m}$ and ImageJ $\bar{d}_{\text{ImageJ}} = 45.01\ \mu\text{m}$. The IQR shows a difference of $\Delta d = 0.52\ \mu\text{m}$. Overall, the measured droplet sizes are credible, with minor deviations attributed to the droplet labeling process. Given the camera resolution, where one pixel corresponds to $1.2\ \mu\text{m}$, the assigned label length for each droplet significantly impacts the measurement, especially for smaller droplets.

Figure 8b presents again the validation of the AI-based method for a later time step than that presented in Figure 8a. This leads to smaller droplet sizes and probably to a narrower distribution. The median of the distribution, which was determined using the

AI method, is $d_{50} = 28.71 \mu\text{m}$. The median of the manual evaluation is $d_{50} = 24.07 \mu\text{m}$. The deviation is therefore $\Delta d_{50} = 4.74 \mu\text{m}$. In addition, the mean values of both distributions were calculated ($\bar{d}_{\text{YOLOv4}} = 29.65 \mu\text{m}$ and $\bar{d}_{\text{ImageJ}} = 25.69 \mu\text{m}$), as were the IQRs, which differ by a delta of $\Delta d = 7.74 \mu\text{m}$. The number of evaluated droplets using YOLOv4 + HC was 134, while 100 droplets were labeled manually. At this point, the gap between manual and AI-based evaluation is more pronounced than before, mainly due to the influence of image resolution. Motion blur further complicates the AI-based evaluation, as the edges of the droplets are less prominent. In addition, the AI-based method shows lower sensitivity for smaller droplets than the manual method, as can be seen from histograms and IQRs. This lower sensitivity is particularly noticeable if the droplet classes are widely scattered in one image. Nevertheless, this lower sensitivity is a consequence of the pixel-to-micrometer ratio as a result of the small magnification and sensor resolution. The last comparison focuses on the final product of an emulsion that has already been cooled. Thus, a very narrow and small droplet size distribution is expected. The number of droplets detected was 427, while 100 droplets were labeled for the ImageJ evaluation. The median between the distributions differs by $\Delta d_{50} = 1.59 \mu\text{m}$, the mean differs by $\Delta \bar{d} = 1.75 \mu\text{m}$, and the IQR differs by $\Delta d = 0.06 \mu\text{m}$. The histogram shown in Figure 8c has a class width of $0.54 \mu\text{m}$, indicating the importance of accurate labeling. Consequently, small differences in the defined pixel length of a droplet have a large influence on the resulting droplet size. Therefore, only those droplets whose edges were 100% recognized were evaluated manually. The reference value, which was evaluated for this end product in the laboratory of SystemKosmetik Produktionsgesellschaft, is a value of $d_{50} = 9.02 \mu\text{m}$, indicating a deviation in the AI-based results ($\Delta d_{50} = 2.11 \mu\text{m}$). However, the overall deviations in the median and mean values are again explained by the pixel-to-micrometer ratio, which has a large influence on the resulting droplet size. If the flow speed is too high for the chosen shutter speed, motion blur occurs, amplifying the effect of the pixel-to-micrometer ratio as the droplet edges become noticeably blurred. The measurement system itself shows no influence on the droplet size and does not affect droplet measurement. Increasing the image resolution and the use of a greater magnification are recommended for further studies.

Considering the processing time of the individual images, the single droplet detection and evaluation of a single image using the ML pipeline, which is running on the camera's processor, is completed within approximately four seconds, plus approximately 30 s to initialize the detection's interpreter. However, this processing time depends on the number of droplets detected per image and may therefore differ. In comparison, the manual evaluation of approximately 100 droplets takes at least 10 min. The acceleration of the process evaluation in this way enables real-time evaluation and characterization of an emulsification process and provides the possibility of process control. For this, further optimization of AI-based image evaluation can be performed on the edge device. Here, the accuracy of the model and the detection time can be considered. In general, performance optimization can be achieved by hardware and software changes. Improvement in detection performance, particularly for small droplets, can be achieved with further training, higher image resolution, and greater magnification. The conversion of YOLOv4 weights for implementation in a TensorFlow framework may have resulted in a reduction in accuracy. Direct training based on a TensorFlow model is a conceivable approach. Additionally, adapting the methodology to a YOLOv7 model is planned, which is expected to not only enhance accuracy but also reduce detection time. This finding has been supported by initial studies and is well-documented in the literature [43,44]. The current investigations utilize only the camera's CPU for the evaluation methodology, indicating that leveraging the GPU or the processor's NPU would accelerate the evaluation process.

Overall, the integrated smart process sensor is suitable for process monitoring and potential process control. However, optimizing detection performance is desirable, as it can substantially expand the sensor's range of applications further. Based on the TRL level, a level-7 sensor is presented since it is a system prototype demonstration in an operational environment showing first validated results.

4. Conclusions and Outlook

This contribution presents the development, testing, and validation of an integrated smart process sensor. The design strategy represents an iterative approach that considers a measurement flow cell for optical accessibility to the emulsion, a camera system as the optical sensor, the coupling of these two components, and the integration of the total measurement system into an industrial emulsion production plant. In addition to the development of the sensor, AI-based droplet size determination for application on the camera processor was modified and validated with respect to the accuracy and applicability for the characterization of the emulsification process. The integration of the sensor system into a production plant for emulsions—more precisely cosmetic products—shows the possibility of evaluation of the product and process during operation without delay in time. The validation of the measurement system shows that the system produces consistent and reproducible results. In addition, the applicability of the measurement system was validated using a benchmark method. The developed bypass system and the sensor system do not influence the droplet size. The resulting DSDs differed only slightly, and the system is sufficiently accurate to be used on an industrial scale. The first application of the measurement system shows its usability for different types of industrial emulsions focusing on cosmetic products. Here, its use to determine different droplet size ranges is shown. In particular, the detection of smaller droplet sizes and the investigation in flow, resulting in motion blur, show an increasing deviation from the manual method and require critical consideration. However, future steps are intended to optimize this deviation and include both hardware and software adaptations. These optimization steps increase the accuracy of the detection itself and the image resolution. The general measurement concept as well as the combination of the optical measurement cell and the camera system, shows a reliable evaluation of the emulsification process, allows faster process evaluation, and provides a basis for process control. Furthermore, the optical measurement cell must be tested for long-term use. Consequently, the durability and pressure resistance of fluidic connections in particular should be investigated. However, the measurement system is sealed against dust and water by the camera housing, ensuring that the measurement is not affected. The camera housing is usable for total operating temperatures up to 85 °C and a humidity up to + 95 % r.h.

For future investigations, the focus will be on adapting AI-based evaluation methods for explicit use on an edge device. A later version of the YOLO evaluation algorithm will be considered for this purpose. In addition, the camera system is to be equipped with a larger sensor (5 MP instead of 2 MP) to achieve a higher resolution of the emulsion images. The evaluation using the TRL level includes a further qualification of the methodology, which is achieved by comparing a variety of emulsions. Further, a material and production study needs to be performed to test and possibly increase the lifetime of the optical measurement cell. Further validation of the presented approach considering the longevity allows for a higher classification of the measurement system based on the TRL level. Finally, the real-time data obtained by the sensor can be used for process control. The evaluation of intermediate and end products is relevant here. Current knowledge indicates that an increase in efficiency is expected for process execution, allowing an increase in plant availability.

Author Contributions: Conceptualization, I.B., S.S. (Sven Salzer), S.S. (Sebastian Stein) and N.K.; methodology, I.B. and N.K.; software, I.B., S.S. (Sven Salzer) and T.O.O.O.; validation, I.B., S.S. (Sven Salzer), S.S. (Sebastian Stein) and N.K.; formal analysis, I.B., S.S. (Sven Salzer) and S.S. (Sebastian Stein); investigation, I.B., S.S. (Sven Salzer), S.S. (Sebastian Stein), T.O.O.O. and O.F.T.; resources, given in the text; data curation, I.B., S.S. (Sven Salzer), S.S. (Sebastian Stein), T.O.O.O. and O.F.T.; writing—original draft preparation, I.B.; writing—review and editing, I.B., S.S. (Sven Salzer), S.S. (Sebastian Stein) and N.K.; visualization, I.B.; supervision, I.B. and N.K.; project administration, S.S. (Sven Salzer), S.S. (Sebastian Stein) and N.K.; funding acquisition, N.K. All authors have read and agreed to the published version of the manuscript.

Funding: The authors thank the German Federal Ministry for Economic Affairs and Climate Action (BMWK) for funding this research as part of AiF (support codes: KK5168501 PR0, KK5173701 PR0, and KK5057302 PR0).

Data Availability Statement: The associated CAD sketches for this article are available on GitHub at <https://github.com/TUDoAD/AI-based-integrated-smart-process-sensor-for-emulsion-control-in-industrial-application> (accessed on 30 July 2024). The repository may be subject to changes due to further contributions. Thus, the state of the repository as described in this paper is available on Zenodo: <https://zenodo.org/doi/10.5281/zenodo.12819799> (accessed on 30 July 2024). The adaptations for the AI-based droplet detection are based on: https://github.com/TUDoAD/DropletDetection_YOLOv4; <https://zenodo.org/doi/10.5281/zenodo.10938289> (accessed on 30 July 2024).

Acknowledgments: The authors would like to thank Carsten Schrömges (TU Dortmund University, BCI Laboratory of Equipment Design), Michael Wagner and Thomas Pelz (SystemKosmetik Produktionsgesellschaft für kosmetische Erzeugnisse mbH), as well as Jens Clees (Ark Vision Systems GmbH & Co. KG for their technical support, and Patrick Becker (Ark Vision Systems GmbH & Co. KG for his support during the software integration on the Ark Vision SmartCam. Additionally, we would like to thank Robin Fortmann (TU Dortmund University, BCI Laboratory of Equipment Design) for the further adaption and preparation of the 3D-printed measurement flow cell.

Conflicts of Interest: Author Sven Salzer is employed by the company Ark Vision Systems GmbH & Co. KG. Author Sebastian Stein is employed by the SystemKosmetik Produktionsgesellschaft für Kosmetische Erzeugnisse mbH. The authors declare that the research was conducted in the frame of the iSPS project in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funding agency had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

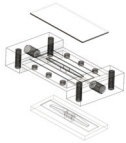
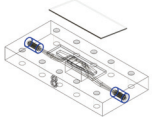
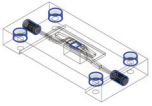
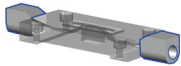
The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
App	Application Software
CCD	Charge-Coupled Device
CPU	Central Processing Unit
CS	Confidence Score
CRI	Color Rendering Index
DC	Direct Current
DSD	Droplet Size Distribution
GMP	Good Manufacturing Practice
GPO	General Purpose Output
GPU	Graphics Processing Unit
H	Height
HC	Hough Circle
IQR	Interquartile Range
L	Length
LED	Light-Emitting Diode
ML	Machine Learning
NPU	Neural Processing Unit
PAN	Path Aggregation Network
PDF	Probability Density Function
PHY	Physical Layer
PVM	Particle Vision Microscope
RAM	Random-Access Memory
SOM	System-on-Module
SPP	Spatial Pyramid Pooling
TFLITE	TensorFlow Lite
TOPS	Trillions or Tera Operations per Second

TRL	Technical Readiness Level
USB	Universal Serial Bus
W	Width
v4	Version 4
YOLO	You Only Look Once

Appendix A

Table A1. History of the measurement flow cell designs that were iteratively generated. Changes compared to the preceding prototype are marked in blue in the images. The first prototype is related to [16], while the last is used for the presented work. For the experimental setup, consider the test setup [16].

CAD Sketch	No. of Components (Total)	Channel Dimension/ mm × mm × mm *1	Fluid Connection	Change *2
	7	3 × 3 × 0.025–0.250	1/8"	-
	4	3 × 3 × 0.025	hose nozzle	Reduction in complexity; Changes fluid connection for faster testing in test setup
	3	3 × 5 × 0.025	5 mm/M5 inside camera housing	Wider channel; Changes fluid connection + screwing for camera/plant integration
	3	3 × 5 × 0.025	8 mm/M8 outside camera housing	Fluid connection outside camera housing; Multiple threads for longer lifetime

*1 At the observation window (W × L × H); *2 Changes to previous prototype.

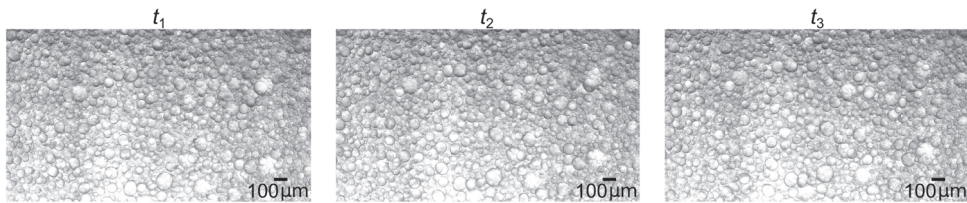


Figure A1. Input images for reproducibility test of the AI-based measurement using the smart process sensor. From left to right, the different time steps t_1 , t_2 , and t_3 are presented.

References

- Jiang, Z. Online Monitoring and Robust, Reliable Fault Detection of Chemical Process Systems. In Proceedings of the 33rd European Symposium on Computer Aided Process Engineering (ESCAPE33), Athens, Greece, 18–21 June 2023; pp. 1623–1628. [CrossRef]
- Tadros, T. *Emulsions—Formation, Stability, Industrial Applications*; De Gruyter: Berlin, Germany; Boston, MA, USA, 2016. [CrossRef]
- Hunter, T.; Pugh, R.; Franks, G.; Jameson, G. The Role of Particles in Stabilising Foams and Emulsions. *Adv. Colloid Interface Sci.* **2008**, *137*, 57–81. [CrossRef] [PubMed]
- Khosravi, H.; Thaker, A.; Donovan, J.; Ranade, V.; Unnikrishnan, S. Artificial Intelligence and Classic Methods to Segment and Characterize Spherical Objects in Micrographs of Industrial Emulsions. *Int. J. Pharm.* **2024**, *649*, 123633. [CrossRef]

5. Panckow, R.; Reinecke, L.; Cuellar, M.; Maaß, S. Photo-Optical In-Situ Measurement of Drop Size Distributions: Applications in Research and Industry. *Oil Gas Sci. Technol.* **2017**, *72*, 14. [CrossRef]
6. Maaß, S.; Rojahn, J.; Hänsch, R.; Kraume, M. Automated Drop Detection Using Image Analysis for Online Particle Size Monitoring in Multiphase Systems. *Comput. Chem. Eng.* **2012**, *45*, 27–37. [CrossRef]
7. Emmerich, J.; Tang, Q.; Wang, Y.; Neubauer, P.; Junne, S.; Maaß, S. Optical Inline Analysis and Monitoring of Particle Size and Shape Distributions for Multiple Applications: Scientific and Industrial Relevance. *Chin. J. Chem. Eng.* **2019**, *27*, 257–277. [CrossRef]
8. Abidin, M.; Raman, A.; Nor, M. Review on Measurement Techniques for Drop Size Distribution in a Stirred Vessel. *Ind. Eng. Chem. Res.* **2013**, *52*, 16085–16094. [CrossRef]
9. Brás, L.; Gomes, E.; Ribeiro, M.; Guimarães, M. Drop Distribution Determination in a Liquid-Liquid Dispersion by Image Processing. *Int. J. Chem. Eng.* **2009**, *2009*, 746439. [CrossRef]
10. Bowler, A.; Bakalis, S.; Watson, N.A. Review of In-line and On-line Measurement Techniques to Monitor Industrial Mixing Processes. *Chem. Eng. Res. Des.* **2020**, *153*, 463–495. [CrossRef]
11. Neuendorf, L.; Müller, P.; Lammers, K.; Kockmann, N. Convolutional Neural Network (CNN)-Based Measurement of Properties in Liquid-Liquid Systems. *Processes* **2023**, *11*, 1521. [CrossRef]
12. Wu, Y.; Gao, Z.; Rohani, S. Deep Learning-based Oriented Object Detection for In situ Image Monitoring and Analysis: A Process Analytical Technology (PAT) Application for Taurine Crystallization. *Chem. Eng. Res. Des.* **2021**, *170*, 444–455. [CrossRef]
13. Huo, Y.; Zhang, F. In-situ Detection of Micro Crystals During Cooling Crystallization Based on Deep Image Super-Resolution Reconstruction. *IEEE Access* **2021**, *9*, 31618–31626. [CrossRef]
14. Lins, J.; Harweg, T.; Weichert, F.; Wohlgemuth, K. Potential of Deep Learning Methods for Deep Level Particle Characterization in Crystallization. *Appl. Sci.* **2022**, *12*, 2465. [CrossRef]
15. Unnikrishnan, S.; Donovan, J.; MacPherson, R.; Tormey, D. An Integrated Histogram-Based Vision and Machine-Learning Classification Model for Industrial Emulsion Processing. *IEEE Trans. Ind. Inform.* **2020**, *16*, 5948–5955. [CrossRef]
16. Burke, I.; Assies, C.; Kockmann, N. Rapid Prototyping of a Modular Optical Flow Cell for Image-Based Droplet Size Measurements in Emulsification Processes. *J. Flow Chem.* **2024**. [CrossRef]
17. Burke, I.; Dhayaparan, T.; Youssef, A.S.; Schmidt, K.; Kockmann, N. Two Deep Learning Methods in Comparison to Characterize Droplet Sizes in Emulsification Flow Processes. *J. Flow Chem.* **2024**. [CrossRef]
18. Kockmann, N. Digital Methods and Tools for Chemical Equipment and Plants. *React. Chem. Eng.* **2019**, *4*, 1522–1529. [CrossRef]
19. Kadlec, P.; Gabrys, B.; Strandt, S. Data-driven Soft Sensors in the Process Industry. *Comput. Chem. Eng.* **2009**, *33*, 795–814. [CrossRef]
20. Neto, J.; Mota, A.; Lopes, G.; Coelho, B.; Frazão, J.; Moura, A.; Oliveira, B.; Sieira, B.; Fernandes, J.; Fortunato, E.; et al. Open-source Tool for Real-Time and Automated Analysis of Droplet-Based Microfluidic. *Lab Chip* **2023**, *23*, 3238–3244. [CrossRef] [PubMed]
21. Unnikrishnan, S.; Donovan, J.; Tormey, D.; Macpherson, R. Emulsion Quality Evaluation Using Automated Image Analysis. *EasyChair Prepr.* **2022**, 8762. [CrossRef]
22. Habib, G.; Qureshi, S. Optimization and Acceleration of Convolutional Neural Networks: A Survey. *J. King Saud Univ.—Comput. Inf. Sci.* **2022**, *34*, 4244–4268. [CrossRef]
23. Neuendorf, L.; Khaydarov, V.; Schler, C.; Kock, T.; Fischer, J.; Urbas, L.; Kockmann, N. Artificial Intelligence-based Module Type Package-compatible Smart Sensors in the Process Industry. *Chemie-Ingenieur-Technik* **2023**, *95*, 1546–1554. [CrossRef]
24. Sibirtsev, S.; Zhai, S.; Neufang, M.; Seiler, J.; Jupke, A. Mask R-CNN Based Droplet Detection in Liquid-Liquid Systems, Part 2: Methodology for Determining Training and Image Processing Parameter Values Improving Droplet Detection Accuracy. *Chem. Eng. J.* **2023**, *473*, 144826. [CrossRef]
25. Sibirtsev, S.; Zhai, S.; Jupke, A. Mask R-CNN Based Droplet Detection in Liquid-Liquid Systems, Part 3: Model Generalization for Accurate Processing Performance Independent of Image Quality. *Chem. Eng. Res. Des.* **2024**, *202*, 161–168. [CrossRef]
26. Schäfer, J.; Schmitt, P.; Hlawitschka, M.; Bart, H. Measuring Particle Size Distributions in Multiphase Flows Using a Convolutional Neural Network. *Chemie-Ingenieur-Technik* **2019**, *91*, 1688–1695. [CrossRef]
27. Liu, J.; Kuang, W.; Liu, J.; Gao, Z.; Rohani, S.; Gong, J. In-situ Multiphase Flow Imaging for Particle Dynamic Tracking and Characterization: Advances and Applications. *Chem. Eng. J.* **2022**, *438*, 135554. [CrossRef]
28. Manee, V.; Zhu, W.; Romagnoli, J. A Deep Learning Image-Based Sensor for Real-Time Crystal Size Distribution Characterization. *Ind. Eng. Chem. Res.* **2019**, *58*, 23175–23186. [CrossRef]
29. Gao, Z.; Wu, Y.; Bao, Y.; Gong, J.; Wang, J.; Rohani, S. Image Analysis for In-line Measurement of Multidimensional Size, Shape, and Polymorphic Transformation of L-Glutamic Acid Using Deep Learning-Based Image Segmentation and Classification. *Cryst. Growth Des.* **2018**, *18*, 4275–4281. [CrossRef]
30. Kockmann, N.; Bittorf, L.; Krieger, W.; Reichmann, F.; Schmalenberg, M.; Soboll, S. Smart Equipment—A Perspective Paper. *Chem. Ing. Tech.* **2018**, *90*, 1806–1822. [CrossRef]
31. Bundesministerium des Innern und für Heimat. Design Thinking. Available online: https://www.orghandbuch.de/Webs/OHB/DE/OrganisationshandbuchNEU/4_MethodenUndTechniken/Methoden_A_bis_Z/Design_Thinking/Design%20Thinking_node.html (accessed on 12 July 2024).
32. Manning, C. Technology Readiness Levels. Available online: <https://www.nasa.gov/directorates/somd/space-communications-navigation-program/technology-readiness-levels/> (accessed on 12 July 2024).

33. Sopat. Available online: <https://www.sopat.de/de/> (accessed on 18 July 2024).
34. Dinter, R.; Helwes, L.; Vries, S.; Jegatheeswaran, K.; Jibben, H.; Kockmann, N. 3D-Printed Open-Source Sensor Flow Cells for Microfluidic Temperature, Electrical Conductivity, and pH Value Determination. *J. Flow Chem.* **2024**, *14*, 469–479. [CrossRef]
35. Glotz, G.; Kappe, C. Design and Construction of an Open Source-based Photometer and its Applications in Flow Chemistry. *React. Chem. Eng.* **2018**, *3*, 478–486. [CrossRef]
36. Schmalenberg, M.; Sallamon, F.; Haas, C.; Kockmann, N. Temperature-Controlled Minichannel Flow-Cell for Non-Invasive Particle Measurements in Solid-Liquid Flow. In Proceedings of the ASME 2020 18th International Conference on Nanochannels, Microchannels, and Minichannels (ICNMM2020), Orlando, FL, USA, 12–15 July 2020. [CrossRef]
37. Burke, I.; Youssef, A.S.; Kockmann, N. Design of an AI-supported Sensor for Process Relevant Parameters in Emulsification Processes. In Proceedings of the Dresdner Sensor-Symposium, Dresden, Germany, 5–7 December 2022; pp. 218–223. [CrossRef]
38. Analytics, H. Durchfluss-Küvette 137-QS, Quarzglas High Performance, 1 mm Schichtdicke. Available online: <https://www.analytics-shop.com/de/hl137-1-40> (accessed on 18 July 2024).
39. Schneider, C.; Rasband, W.; Eliceiri, K. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675. [CrossRef] [PubMed]
40. Bochkovskiy, A.; Wang, C.; Liao, H. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. [CrossRef]
41. IMXMLUG_6.6.23_2.0.0—i.MX Machine Learning User’s Guide. Available online: <https://www.nxp.com/docs/en/user-guide/IMX-MACHINE-LEARNING-UG.pdf> (accessed on 16 July 2024).
42. Laskowski, P. GitHub Repository—Convert_Darknet_YOLO_to_TensorFlow (patryklaskowski). Available online: https://github.com/patryklaskowski/Convert_Darknet_YOLO_to_TensorFlow (accessed on 16 July 2024).
43. Wang, C.; Bochkovskiy, A.; Liao, H. YOLOv7: Trainable Bag-of-freebies Sets New State-of-the-art for Real-time Object Detectors. *arXiv* **2022**, arXiv:2207.02696. [CrossRef]
44. Durve, M.; Orsini, S.; Tiribocchi, A.; Montessori, A.; Tucny, J.; Lauricella, M.; Camposeo, A.; Pisignano, D.; Succi, S. Benchmarking YOLOv5 and YOLOv7 Models with DeepSORT for Droplet Tracking Applications. *Eur. Phys. J.* **2023**, *46*, 32. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Review

Advanced-Functional-Material-Modified Electrodes for the Monitoring of Nitrobenzene: Progress in Nitrobenzene Electrochemical Sensing

Khursheed Ahmad and Tae Hwan Oh *

School of Chemical Engineering, Yeungnam University, 280 Daehak-Ro, Gyeongsan 38541, Republic of Korea; khursheed@yu.ac.kr

* Correspondence: taehwanoh@ynu.ac.kr

Abstract: Nitrobenzene (NB) is one of the nitro-aromatic compounds that is extensively used in various chemical industries. Despite its potential applications, NB is considered to be a toxic compound that has significant hazardous effects on human health and the environment. Thus, it can be said that the NB level should be monitored to avoid its negative impacts on human health. In this vein, the electrochemical method has emerged as one of the most efficient sensing techniques for the determination of NB. The sensing performance of the electrochemical techniques depends on the electro-catalytic properties and conductivity of the electrode materials. In the past few years, various electrode materials, such as conductive metal ions, semiconducting metal oxides, metal-organic frameworks, and two-dimensional (2D) materials, have been used as the electrode material for the construction of the NB sensor. Thus, it is worth summarizing previous studies on the design and synthesis of electrode materials for the construction of the NB sensor. In this mini-review article, we summarize the previous reports on the synthesis of various advanced electrode materials, such as platinum (Pt) nanoparticles (NPs), silver (Ag) NPs, carbon dots (CDs), graphene, graphitic carbon nitride (g-C₃N₄), zinc stannate (ZnSnO₃), cerium oxide (CeO₂), zinc oxide (ZnO), and so on. Furthermore, the impacts of different electrode materials are systematically discussed for the sensing of NB. The advantages of, limitations of, and future perspectives on the construction of NB sensors are discussed. The aim of the present mini-review article is to enhance the knowledge and overall literature, working towards the construction of NB sensors.

Keywords: nitrobenzene; metal oxides; electrochemical sensor; MOF

Citation: Ahmad, K.; Oh, T.H. Advanced-Functional-Material-Modified Electrodes for the Monitoring of Nitrobenzene: Progress in Nitrobenzene Electrochemical Sensing. *Processes* **2024**, *12*, 1884. <https://doi.org/10.3390/pr12091884>

Academic Editors: Pier Gianni Medaglia, Chi He, Li Jin, Sheng Du, Xiongbo Wan and Zixin Huang

Received: 11 July 2024

Revised: 6 August 2024

Accepted: 27 August 2024

Published: 2 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the present scenario, environmental pollution is considered to be one of the major challenges globally [1,2]. Various organic and inorganic compounds, such as the nitro (-NO₂) group containing aromatic compounds (nitrobenzene (NB; C₆H₅NO₂), nitrophenol (NP), and nitrotoluene (NT)), hydrazine, catechol, hydrogen peroxide, etc., are widely used in the chemical and dye industries [3–8]. In particular, NB is widely used in various applications, such as rubber, chemicals, dyes, explosives, pesticides, urethane polymers, the pharmaceutical industry, and analgesic drugs [9–12]. Despite its various advantages and uses, NB has a significant negative impact on human health and the environment due to its hazardous nature. The United States Environmental Protection Agency classifies it as a group 2B carcinogen [13]. NB presents severe health risks as it can be easily inhaled, ingested, and absorbed through the skin [14,15]. In addition, NB may enter the environment through various routes, such as industrial discharge, land leaching, accidental spills, and chemical waste. Thus, it can be understood that NB has serious negative effects on human health and the environment. Therefore, it is of great significance to monitor the level of NB to avoid its hazardous effects. In this regard, conventional methods, such as gas chromatography (GC), surface plasmonic resonance (SPR), Raman spectroscopy (RS), and

high-performance liquid chromatography (HPLC), are widely used for the determination of NB [13–20]. Unfortunately, these conventional methods are complex, expensive, and challenging to use for the determination of NB [20]. Thus, it is necessary to find an alternative sensing approach for the sensing of NB with high sensitivity and selectivity.

Recently, the electrochemical method has received a tremendous amount of attention because of its excellent sensitivity, simplicity, fast response, selectivity, storage stability, and repeatability [21]. These types of electrochemical sensors have a low environmental impact due to their excellent reusability and stability and the high stability of the catalyst. Thus, researchers are highly motivated towards the design and fabrication of highly efficient electrochemical sensors [22]. Electrochemical methods involve the use of electro-catalysts, which may significantly influence the performance of the fabricated sensors towards the detection of NB. In the past few years, various electro-catalytic materials, such as transition metal oxides, polymers, metal–organic frameworks (MOFs), carbon dots, metal carbides/nitrides, and composites, have been explored for electrochemical sensing applications [23–26].

In this mini-review article, we compiled previously published reports on the fabrication of the NB sensor using various nanostructured metal oxides, polymers, carbon-based materials, MOFs, and hybrid composites as electrode materials. To the best of our knowledge, in our literature survey, no review report is available which summarizes the different electrode materials towards the determination of NB. This is the first mini-review article to compile the advancements in the electrochemical fabrication of NB sensors. This review provides insights into the electrode materials for the construction of the NB sensor. It is believed that the present mini-review article will be useful to researchers working to improve the selectivity and sensitivity of NB sensors.

2. Progress in the Fabrication of NB Sensors

Previously, various advanced electrode materials based on transition metal oxides, polymers, carbon-based materials, MOFs, and hybrid composite materials were extensively used in the construction of NB sensors. In this section, we summarize previously reported articles on the fabrication of the NB sensor using metal oxides, carbon-based materials (carbon nanotubes, graphene, etc.), metal dichalcogenides, and other materials.

2.1. Metal-Oxide-Based Electrode Materials

In previous years, metal oxides and their composites were extensively used for various electrochemical sensing applications due to their decent electro-catalytic properties, conductivity, eco-friendliness, and cost-effective properties. Cerium dioxide (CeO_2) has good electro-catalytic properties and can be used in the fabrication of electrochemical sensors. It is worth exploring such materials for the fabrication of the NB sensor. In this context, Sangili et al. [27] synthesized CeO_2 nanoparticles (NPs) by employing the hydrothermal method, and the morphological properties of the synthesized CeO_2 NPs were characterized by transmission electron microscopy (TEM); the presence of Ce and O elements was confirmed via energy-dispersive X-ray spectroscopy (EDX). The authors reported that CeO_2 NPs have a uniform spherical structure, and the size of the prepared particles lies in a range of 2–5 nm. The prepared CeO_2 NPs were further proposed as an NB-sensing material, and the sensing capability of the CeO_2 -NP-modified electrode was studied by using the cyclic voltammetry (CV) and differential pulse voltammetry (DPV) techniques. The authors found that the fabricated electrode demonstrated an excellent LOD of $0.092 \mu\text{M}$ with a decent sensitivity of $1.1166 \mu\text{A} \mu\text{M}^{-1} \text{cm}^{-1}$. The proposed sensor also demonstrated a good linear dynamic range of 0.1 to $50 \mu\text{M}$. This work also reported the real sample investigations of the fabricated electrode in tap water and river water with acceptable recoveries. Arul et al. [28] synthesized iron oxide ($\alpha\text{-Fe}_2\text{O}_3$) by using the co-precipitation method at room temperature (RT). SEM studies demonstrated that $\alpha\text{-Fe}_2\text{O}_3$ consists of micro/nanorods that may be beneficial for better electron transport. $\alpha\text{-Fe}_2\text{O}_3$ micro/nanorods have a high specific surface area of $67 \text{m}^2/\text{g}$, and this may

enhance electron transportation during redox reactions. The α - Fe_2O_3 micro/nanorods were deposited onto the surface of a GC electrode, and its electrochemical performance was checked by using the CV method. The α - Fe_2O_3 /GC electrode exhibited an LOD of 30.4 ppb with a sensitivity of 446 nA/ μM for the detection of NB. The reported articles showed that manganese ferrite (MnFe_2O_4 ; spinel ferrite) has excellent properties and can be used in various applications such as sensing and energy storage devices. In this vein, Sang et al. [29] reported the hydrothermal preparation of MnFe_2O_4 for the fabrication of an NB sensor. The synthesized MnFe_2O_4 colloid nanocrystals exhibited good physicochemical properties, and the authors used the synthesized material as an electro-catalyst towards the fabrication of an NB sensor. The MnFe_2O_4 colloid nanocrystal-modified electrode shows improved performance towards the sensing of NB, and an LOD of 4 mM was achieved. Zinc stannate (ZnSnO_3) is a perovskite material, which has robust stability and decent electrical properties, suggesting its potential for electrochemical sensing applications. In this regard, Vinoth et al. [30] studied the role of ZnSnO_3 -incorporated graphitic carbon nitride ($\text{g-C}_3\text{N}_4$) as a sensing material for the detection of NB. The authors prepared a $\text{ZnSnO}_3/\text{g-C}_3\text{N}_4$ composite by employing novel strategies, as shown in Figure 1a. The authors authenticated the formation of $\text{ZnSnO}_3/\text{g-C}_3\text{N}_4$ via the XRD technique (Figure 1b). The high and sharp peak intensity suggest the good crystalline nature of the prepared $\text{ZnSnO}_3/\text{g-C}_3\text{N}_4$. The GC electrode was modified with $\text{ZnSnO}_3/\text{g-C}_3\text{N}_4$ for the sensing of NB. The $\text{ZnSnO}_3/\text{g-C}_3\text{N}_4$ -modified GC electrode exhibited a good LOD of 2.2 μM with a sensitivity of 0.05857 $\mu\text{A } \mu\text{M}^{-1} \text{cm}^{-2}$. The $\text{ZnSnO}_3/\text{g-C}_3\text{N}_4$ -modified GC electrode also demonstrated a linear range of 30–100 μM towards the determination of NB using the LSV technique. The $\text{ZnSnO}_3/\text{g-C}_3\text{N}_4$ -modified GC electrode also exhibited excellent selectivity in the presence of various interfering substances.

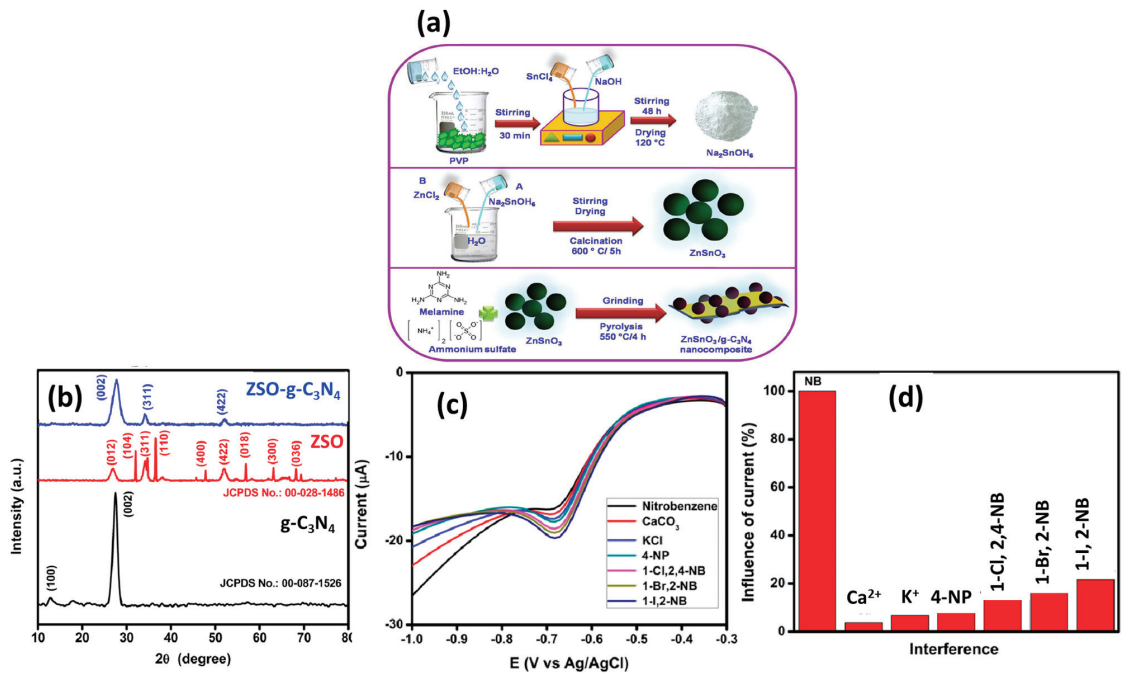


Figure 1. (a) Schematic picture for the synthesis of $\text{ZnSnO}_3/\text{g-C}_3\text{N}_4$ composite. (b) XRD pattern of the ZnSnO_3 , $\text{g-C}_3\text{N}_4$, and $\text{ZnSnO}_3/\text{g-C}_3\text{N}_4$ composite. (c) Selectivity test (DPV curves) of the NB sensor in presence of various interfering substances. (d) Selectivity results for NB sensing. Reprinted with permission [30].

The DPV graphs are shown in Figure 1c, which indicate that the presence of interfering substances does not influence the performance of the ZnSnO₃/g-C₃N₄-modified GC electrode. The ZnSnO₃/g-C₃N₄-modified GC electrode demonstrated excellent selectivity for the detection of NB in the presence of interfering materials (K⁺, Ca²⁺, 4-nitrophenol, 1-bromo,2-nitrobenzene, 1-chloro,2,4-dinitrobenzene, 1-iodo, and 2-nitrobenzene), as shown in Figure 1d. The ZnSnO₃/g-C₃N₄-modified GC electrode also exhibited good recovery in real samples, which suggested its application for practical purposes.

Magnesium oxide (MgO) is a fascinating alkaline metal oxide, which possesses excellent catalytic, adoption, chemical, and mechanical properties, which makes MgO an efficient electrode material for the construction of cost-effective electrochemical sensors. The catalytic properties of MgO can be further improved by combing it with carbon-based materials. Kokulnathan et al. [31] proposed the synthesis of a GO/MgO composite using a benign approach, as shown in Figure 2a. The authors modified the GC electrodes with GO, MgO, and GO/MgO as catalysts and evaluated their performance towards a reduction in NB using the CV and DPV methods (Figure 2b).

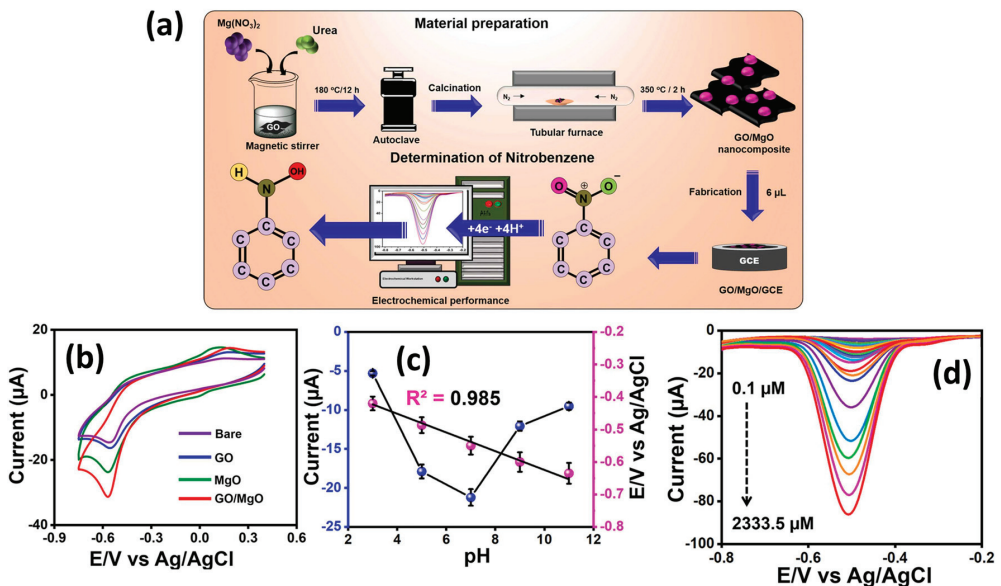


Figure 2. (a) Schematic representation of the synthesis of GO/MgO composite, fabrication of GO/MgO/GC electrode and its working for NB sensing. (b) CV graphs of the different electrodes for NB sensing. (c) Current value versus pH graph for the sensing of NB using GO/MgO/GC electrode. (d) DPV graphs of the GO/MgO/GC electrode in presence of various concentrations of NB. Reprinted with permission [31].

It was clearly seen that GO/MgO-modified GC electrode has higher catalytic activity for the improved NB reduction, as shown in Figure 2b. The pH of the solution was also optimized, and higher activity was observed at a pH of 7.0 (Figure 2c). The authors also adopted the DPV technique for further electrochemical investigations and found that DPV is more sensitive compared to the CV method. The effect of various concentrations was studied using a GO/MgO-modified GC electrode, as shown in Figure 2d. It is clear that synergistic interactions between GO and MgO enhanced the charge transfer properties of the modified electrode and improved the reduction in NB using the DPV method. The GO/MgO-modified GC electrode demonstrated a low LOD of 0.01 μM with two linear dynamic ranges of 0.1 to 38.9 μM and 58.5 to 2333.5 μM. The reported NB sensor exhibited excellent repeatability, stability, and selectivity. Acceptable recoveries of 99.35% to 99.80%

were also observed for real sample investigations for the GO/MgO-modified GC electrode towards the detection of NB in real water samples. Manganese dioxide (MnO_2) has good electro-catalytic properties and may be a good electrode for the fabrication of NB sensors. In this regard, Chellappa et al. [32] reported a simple hydrothermal method for the synthesis of MnO_2 nanorods (NRs), as depicted in Figure 3a. The MnO_2 NRs were coated onto the bare surface of the GC electrode, and its sensing activity for the sensing of NB was evaluated by using the CV and DPV methods.

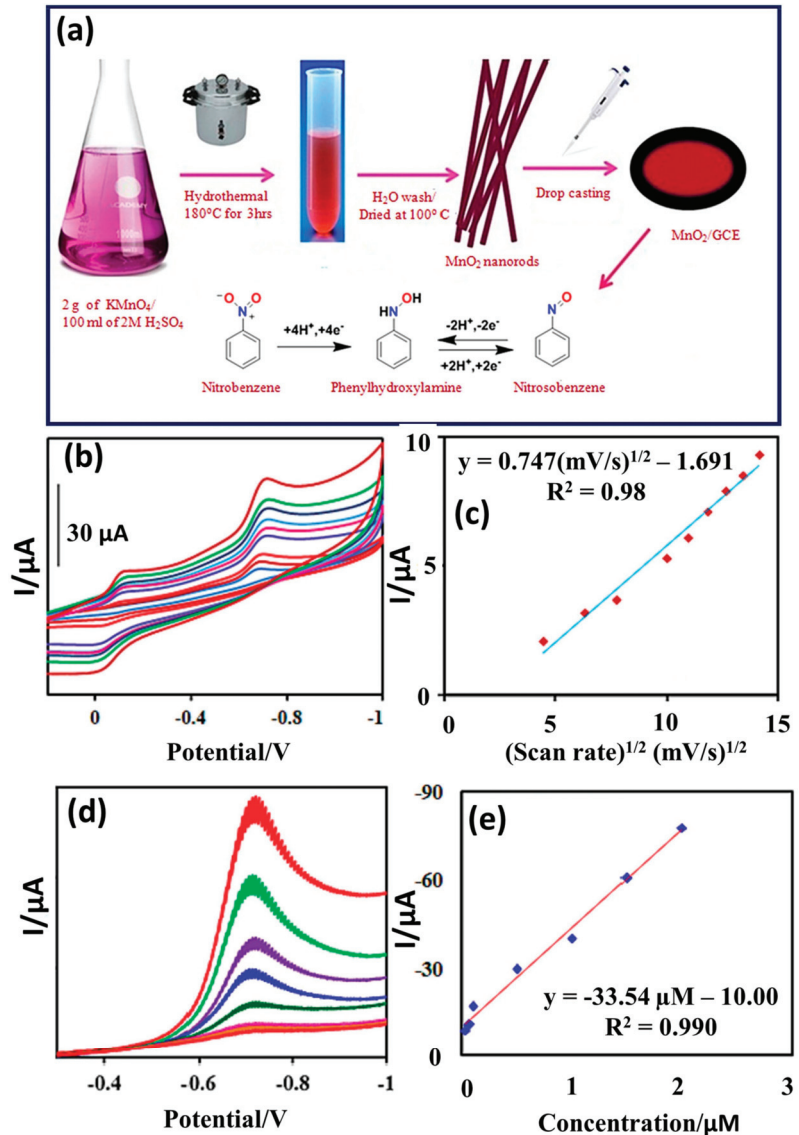


Figure 3. (a) Schematic diagram presents the synthetic procedure for MnO_2 NRs. (b) CV of the MnO_2 NRs/GCE in presence of NB at different scan rates. (c) Calibration plot between current response and square root of scan rate. (d) DPV curves of the MnO_2 NRs/GCE in presence of different concentrations of NB. (e) Calibration plot between current response and concentration of NB. Reprinted with permission [32].

CV graphs of the MnO₂ NRs/GCE in the presence of NB at different applied scan rates are shown in Figure 3b. It is seen that the current response increases with respect to the applied scan rate. The calibration plot shows a regression coefficient (R^2) value of 0.98 (Figure 3c). Thus, this suggests that the current response for redox reactions linearly increases. It is considered that DPV may be a more sensitive technique for the determination of NB. Thus, the authors studied the electro-catalytic activities of MnO₂ NRs/GCE in the presence of different concentrations of NB. The DPV results for MnO₂ NRs/GCE indicated that current response for the redox reactions also increases with increasing concentrations of NB (Figure 3d), and this response was found to be linear (Figure 3e), as suggested by the R^2 value of 0.99. The authors proposed that DPV is relatively more sensitive for the detection for NB in comparison to CV. The MnO₂ NRs/GCE demonstrated excellent recovery in real sample studies, with a reasonably good LOD of 0.025 μM and a linear range of 0.03 to 2 μM for the detection of NB, with high stability. Kokulnathan et al. [33] demonstrated that zinc oxide NRs/copper tin sulfide nanoflowers (ZnO-NRs@CTS-NFs) can be synthesized by using the hydrothermal route and used it as an NB sensing material. The proposed ZnO-NRs@CTS-NF-based NB sensor exhibited an LOD of 0.002 μM and sensitivity of 1.30 $\mu\text{A}/\mu\text{M cm}^2$, with two linear ranges of 0.01–17.2 and 17.2–203 μM using the DPV method. The ZnO-NRs@CTS-NF-based NB sensor also exhibited acceptable recovery in real samples, such as rivers, ponds, and industrial water.

2.2. Carbon-Family-Based Electrode Materials

Carbon-based materials, such as graphene, carbon nanotubes, and graphitic carbon nitride, have decent catalytic properties and surface areas. Sang et al. [34] modified the surface area of the bare GCE using multi-walled carbon nanotubes (MWCNTs) as an electrode material for the determination of NB. The performance of the MWCNTs/GCE was evaluated in acid electrolytes using the CV and DPV methods. The authors found that hydroxyl-containing MWCNT-modified GCE (MWCNTs-OH/GCE) has excellent performance for the determination of NB compared to the pristine MWCNTs/GCE. This showed that the functionalization of MWCNTs with an OH group significantly improved the sensing properties of the pristine MWCNTs. Thus, it would be worth using the OH-functionalized MWCNT-modified electrodes as working electrodes for the detection of NB. The functionalized (f)-MWCNTs have an excellent conductive nature and improved catalytic properties due to the presence of functional groups on the surface of MWCNTs. Therefore, it is of great significance to utilize the f-MWCNTs as sensing materials for the construction of an NB sensor. Govindasamy et al. [35] proposed the use of f-MWCNTs as electrode materials for NB sensing applications.

The authors also used nafion as a binder to improve the adhesiveness/stability of the f-MWCNTs on the surface of a screen-printed carbon electrode (SPCE). The electrochemical performance of the f-MWCNT-modified SPCE (f-MWCNTs/SPCE) was determined in the presence of NB, and the obtained results suggested that f-MWCNTs/SPCE has good electro-catalytic properties. The f-MWCNTs/SPCE exhibited a wide dynamic linear range of 50 nm to 1170 μM . A sensitivity of 0.6685 $\mu\text{A}/\mu\text{M cm}^2$ was obtained using f-MWCNTs/SPCE for the determination of NB. A decent LOD value of 45 nM was also reported for the sensing of NB using f-MWCNTs/SPCE. Furthermore, f-MWCNTs/SPCE also demonstrates satisfactory stability, repeatability, and reproducibility. The authors successfully recovered the NB from a human urine sample.

Thirumalraj et al. [36] demonstrated the role of electrochemically activated graphite (EAG) as catalysts for the sensing of NB. The authors prepared an EAG-modified screen-printed electrode (EAG-SPE) and explored this as the working electrode for the determination of NB. The sensing performance of EAG-SPE was determined by employing the CV method under different pH conditions of phosphate-buffered saline (PBS) solutions. The EAG-SPE exhibited higher catalytic properties under a pH of 7.0, and the authors reported a pH of 7.0 as the optimal condition. Furthermore, the authors checked the effects of different applied scan rates and found that current responses for the determination

of NB increase with increasing applied scan rates. Chronoamperometry (CA) is one of the most sensitive methods for the sensing of toxic compounds and biomolecules. Thus, the authors also adopted the CA method for further sensing experiments. The effect of various concentrations of NB on the current response of EAG-SPE has been investigated using the CA method. The authors found that current responses/signals rapidly increase with the addition/spike of NB at different concentrations. EAG-SPE demonstrated an excellent LOD of 0.06 μM and sensitivity of 1.445 $\mu\text{A } \mu\text{M}^{-1} \text{ cm}^{-2}$ [36]. The EAG-SPE also demonstrated its excellent selective nature for the determination of NB in the presence of various interfering compounds. Ma et al. [37] also reported the fabrication of an NB sensor by employing macro-/meso-porous carbon material (MMPCM) catalysts. The MMPCMs were prepared via a sonication-assisted pyrolysis method using a silica template. The synthesized MMPCMs were characterized by the SEM method, which suggested the presence of a honeycomb-like surface morphology with lots of macro-pores (diameter range = ~ 330 nm). These morphological features improve the electron transport and improve the sensing activity of modified GC electrodes. Thus, the authors modified the surface of the GC electrode using the MMPCM catalyst and employed linear sweep voltammetry (LSV) for the determination of NB. The authors studied the effects of various concentrations of NB (0.2, 0.4, 1, 2, 4, 6, 10, 15, 20, 30, 40 μM) on the performance of the MMPCM catalyst-modified GC electrode. They found that the current response increases with respect to the concentration of the NB. An LOD of 8 nM was obtained with a sensitivity of 2.36 $\mu\text{A } \mu\text{M}^{-1}$. They also reported that the MMPCM catalyst-modified GC electrode has good selectivity for the detection of NB in the presence of various interfering molecules, such as NO_3^- , Mn^{2+} , CO_3^{2-} , SO_4^{2-} , PO_4^{3-} , CH_3COO^- , and Mg^{2+} . The MMPCM catalyst-modified GC electrode also demonstrated good stability and reproducibility.

It is well known that nitrogen (N)-doped materials may have significantly improved catalytic properties due to the generation of active sites on N-doped materials. It is of great significance to prepare a N-doped catalyst for electrochemical sensing applications. In this context, Liu et al. [38] synthesized N-doped hollow carbon nanospheres (NHCPs) by employing novel strategies. The mono-dispersed SiO_2 microspheres were used as a template with dopamine as the N source. The synthetic process for the preparation of NHCPs is described in Figure 4a.

The authors authenticated the formation of NHCPs through XRD and Fourier-transform infrared (FT-IR) spectroscopy. The NHCPs were treated at a temperature of 750 $^\circ\text{C}$ and denoted as NHCPs-750. The NHCPs-750 was coated on the GC surface, and the performance of the NHCPs-750-GC electrode was determined by the CV, electrochemical impedance spectroscopy (EIS), and DPV methods. The NHCPs-750-GC has a higher current response for the sensing of NB compared to the GC electrode. The DPV curves of the NHCPs-750-GC for different concentrations of NB are shown in Figure 4b. It is clear that the current linearly increases with increasing concentrations of NB (inset of Figure 4b). The NHCPs-750-GC demonstrated an LOD of 2.29 μM with high sensitivity of 436 $\mu\text{A mM}^{-1}$. The NHCPs-750-GC also has an excellent linear range of 5–2610 μM for the sensing of NB. The presence of pyridinic N in the NHCPs-750-GC electrode improved the catalytic behavior for the sensing of NB. The authors also evaluated the selectivity of the NHCPs-750-GC electrode for the sensing of NB in the presence of various interferences. Figure 4c shows the excellent selectivity of NHCPs-750-GC for the determination of NB in the presence of various interferences. In other work, Sakthivel et al. [39] prepared a novel catalyst material for the sensing of NB. The authors synthesized a chitin hydrogel-stabilized graphite (GR-CHI) composite using simple strategies and fabricated the surface of the GC electrode with the prepared GR-CHI as a catalyst material. The physiochemical investigations showed that synthesized GR-CHI is formed by strong interaction between GR and CHI. GR-CHI/GCE was further employed as a working electrode, and CV was adopted as the voltammetric sensing approach for the detection of NB. The CV results showed that the current response for the redox peaks increases with increasing scan rates, and sensing of NB is an adsorption-controlled process. The performance of the

GR-CHI/GCE was further studied by employing the CA method. The current response increases with increasing concentrations of NB, and this increment was found to be linear. The GR-CHI/GCE exhibited an LOD of 37 nM and linear range of 0.1 to 594.6 μM using the amperometric (*i-t*) method. The sensor (GR-CHI/GCE) demonstrates good selectivity, excellent practicality, and consistent repeatability in detecting NB in laboratory water samples. β -cyclodextrin (β -CD) possesses notable physicochemical properties, which make it valuable in various applications, such as sensing, drug delivery, and environmental remediation. β -CD is a chemically stable molecular structure with biocompatibility and less toxicity, which make it a suitable candidate for the fabrication of electrochemical sensors to monitor toxic substances. In this vein, Velmurugan et al. [40] designed and prepared β -CD on the graphene oxide (GO) surface using a benign synthetic procedure (ultra-sonication method). The authors optimized the loading level of β -CD on the GC surface to improve the performance of the NB sensor. The β -CD/GO/SPCE was used as the NB sensor, and its performance was checked using the CV and LSV techniques. The authors reported that the presence of synergistic interactions between β -CD and GO improved the performance of β -CD/GO/SPCE towards the determination of NB. The β -CD_{1.2 mg}/GO/SPCE was found to be an optimized electrode for the sensing of NB. The β -CD_{1.2 mg}/GO/SPCE exhibited a linear range of 0.5 to 1000 μM , with an LOD of 0.184 μM . β -CD_{1.2 mg}/GO/SPCE also has acceptable reproducibility, good sensitivity, and decent stability for the sensing of NB. The authors also proposed the potential of β -CD_{1.2 mg}/GO/SPCE for practical purposes due to good recoveries in real samples.

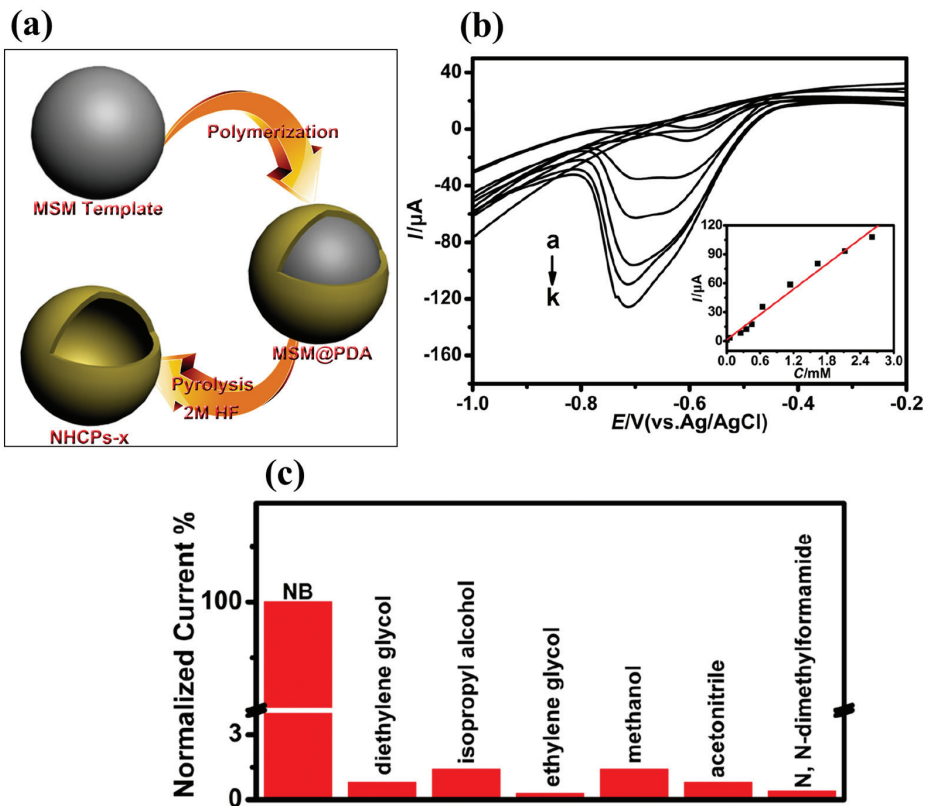


Figure 4. (a) Schematic picture for the preparation of NHCPs_{s-x}. (b) DPV graphs of the NHCPs-750-GC electrode in different concentration of NB. (c) Selectivity nature of NHCPs-750-GC electrode for NB sensing. Reprinted with permission [38].

Kubendhiran et al. [41] explored the potential of a reduced graphene oxide/nickel tetraphenyl porphyrin (GRGO/Ni-TPP) nanocomposite as a sensing material for the fabrication of an NB sensor. The authors synthesized GRGO via an eco-friendly method, utilizing caffeic acid as the reducing agent. Furthermore, the authors reported that the GRGO/Ni-TPP composite material was formed by π - π stacking interactions between RGO and Ni-TPP. The interaction and vibrational band in the prepared GRGO/Ni-TPP were confirmed by FTIR analysis. The synthesized GRGO/Ni-TPP was drop casted on the active surface of the bare GC electrode. The GRGO/Ni-TPP-modified GC electrode was further used for the sensing of NB. EIS studies also revealed that the GRGO/Ni-TPP/GC electrode has higher conductivity with a low charge transfer resistance value compared to the GC electrode. The CV graph of the GRGO/Ni-TPP/GC electrode was recorded in the absence and presence of NB.

Figure 5a reveals that no redox peaks appeared for GRGO/Ni-TPP/GCE in the absence of NB. However, sharp redox peaks were observed in the presence of NB. The GRGO/Ni-TPP/GCE demonstrated a higher current response compared to the bare GCE (b), GRGO/GCE (c), NiTPP/GCE (d), and GO/NiTPP/GCE (e), as shown in Figure 5a. The CV of the GRGO/Ni-TPP/GC was also obtained under different pH conditions, and the authors found that GRGO/Ni-TPP/GCE is highly active under pH conditions of 7.0 (Figure 5b). The GRGO/Ni-TPP/GCE exhibited an LOD of 0.14 μM and sensitivity of 1.277 $\mu\text{A } \mu\text{M}^{-1} \text{ cm}^{-2}$. A wide linear dynamic range of 0.5 to 878 μM was also observed for GRGO/Ni-TPP/GCE towards the sensing of NB. Figure 5c shows the DPV results for GRGO/Ni-TPP/GCE in the presence of different concentrations of NB. It is seen that the current response linearly increases with increasing concentrations of NB (Figure 5d). The DPV results demonstrated better performance compared to the CV results. The GRGO/Ni-TPP/GCE also demonstrated excellent selectivity towards NB detection. The real sample investigations in water samples suggested that the GRGO/Ni-TPP/GCE can be used for practical applications. The improved performance of GRGO/Ni-TPP/GCE may be ascribed to the presence of the synergism between the GRGO and Ni-TPP.

In recent years, a new form of carbon material such as carbon dots (CDs) has received extensive attention from the scientific community because of its extraordinary optoelectronic properties, biocompatibility, ease of surface modification, and low toxicity. The CDs can be synthesized by hydrothermal methods using various C-sources. The utilization of waste materials to form CDs is of great significance. Thus, Bressi et al. [42] synthesized CDs using orange peel waste as a C source via an eco-friendly hydrothermal carbonization/electrochemical bottom-up synthetic process. The transformation of orange peel waste to CDs was confirmed by employing various sophisticated techniques. The authors reported that synthesized CDs have excellent electrochemical properties and deposited it onto the surface of an SPC electrode. The CD-modified SPC electrode was used as an NB sensor using DPV technology. The CD-modified SPC electrode exhibited an LOD, sensitivity, and linear range of 13 nM, 9.36 $\mu\text{A}/\mu\text{M cm}^2$, and 0.1–200 μM , respectively. The excellent long-term stability, selectivity, and repeatability of the CD-modified SPC electrode suggested its potential for commercialization. Pandiyarajan et al. [43] also proposed the fabrication of a novel NB sensor using innovative strategies. In this vein, the authors proposed the synthesis of a Ag NP-decorated N-[3-(trimethoxysilyl) propyl]ethylenediamine (EDAS)-modified g-C₃N₄ composite for the construction of an NB sensor. The authors optimized the concentration of Ag NPs to improve the catalytic properties of the EDAS/(g-C₃N₄-Ag). The EDAS/(g-C₃N₄-Ag)-modified GC electrode demonstrated an LOD of 2 μM , sensitivity of 0.594 $\text{A M}^{-1} \text{ cm}^{-2}$, and linear range of 5 to 50 μM for the sensing of NB. Zhang et al. [44] reported the synthesis of 2D mesoporous carbon nitride (OMCN) using SBA-15 mesoporous silica as the template and melamine as the precursor. The synthesized OMCN was used as an electrode material for the modification of the GC electrode. The OMCN-modified GC electrode demonstrated an LOD of 1.52 μM for the sensing of NB.

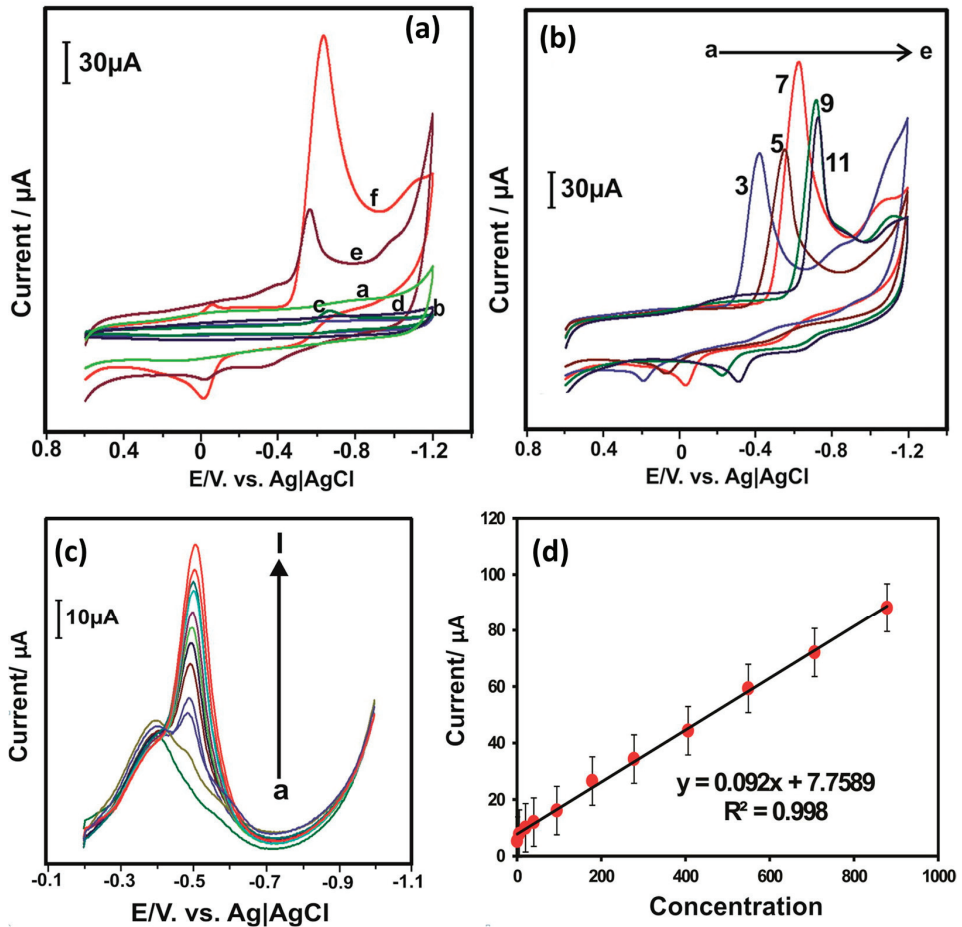


Figure 5. (a) CV curve of the (a) GRGO/Ni-TPP/GCE in the absence of NB. CV of (b) bare GCE, (c) GRGO/GCE, (d) NiTPP/GCE, (e) GO/NiTPP/GCE and (f) GRGO/Ni-TPP/GCE in the presence of NB. (b) CV responses of the GRGO/Ni-TPP/GCE electrode in the presence of NB under different pH conditions. (c) DPV curves of the GRGO/Ni-TPP/GCE in the presence of different concentration of NB. (d) Corresponding calibration curve between current response and concentration of NB. Reprinted with permission [41].

2.3. Bimetallic Materials and Metal Nanoparticle-Based Electrode Materials

Bimetallic materials with nanostructures have received great interest because of their excellent and unique chemical and physical properties and have been explored in various applications including sensors. Nickel–copper (Ni–Cu) is an environmental-friendly bimetallic material and has various advantages, such as low cost, high catalytic activity, and less-toxic nature. Previously, Yan et al. [45] adopted the Ni–Cu alloy electrode for the determination of NB using simple strategies. The authors used the electro-deposition method for the preparation of a Ni–Cu alloy electrode. The prepared Ni–Cu alloy electrode was characterized by X-ray diffraction (XRD) and scanning electron microscopy (SEM)-based analysis. The XRD results demonstrated the presence of (111) and (220) diffraction planes for the presence of the Ni–Cu alloy. The decent intensity of the diffraction peaks suggested a moderate crystalline nature of the prepared Ni–Cu alloy electrode. The SEM results also indicated the presence of small particles, which agglomerated with a rough

surface. The authors stated that these surface properties may be useful for obtaining better electro-catalytic behavior of the Ni-Cu alloy electrode for the sensing of NB. A polarization method was used for the determination of NB, and the authors reported a limit of detection (LOD) of 4×10^{-5} M with a sensitivity of 298 $\mu\text{A}/\text{mM}$. The linear dynamic range of 0.1–20 mM has been reported for the sensing of NB with a correlation coefficient (R^2) of 0.995 using the Ni-Cu alloy electrode. Platinum (Pt) NPs are a significant and highly active catalyst due to their excellent conductivity and catalytic properties. Zhang et al. [46] assembled Pt NPs on macroporous carbon (MPC) by using a simple and benign one-step microwave-assisted heating method. The synthesized Pt NPs/MPC was characterized by TEM, which revealed that the synthesized material has a uniform surface morphology. Further, the authors modified the GC electrode with Pt NPs/MPC (nafion was used as a binder) and investigated its performance for the determination of NB using the LSV method. The Pt NPs/MPC-modified GC electrode exhibited an LOD of 50 nM with a wide linear dynamic range of 1–200 μM . The Pt NPs/MPC-modified GC electrode also exhibited decent recovery in real samples. Rameshkumar et al. [47] prepared silver nanoparticles (Ag NPs) on an amine-functionalized SiO_2 sphere-modified electrode. The Ag NPs/ SiO_2 -modified electrode was then used as an NB sensor, and the electrochemical properties of this Ag NPs/ SiO_2 -modified electrode was determined by CV and square wave voltammetry (SWV). The CV results indicated that the Ag NPs/ SiO_2 -modified electrode has higher catalytic activity compared to the GC electrode. The effect of various concentrations of NB was studied by employing SWV, and the authors found that the increase in the current response of the Ag NPs/ SiO_2 -modified electrode is directly proportional to the concentration of NB. The improved performance of the Ag NPs/ SiO_2 -modified electrode may be ascribed to the presence of a large number of Ag NPs on the SiO_2 surface. The Ag NPs/ SiO_2 -modified electrode exhibited an LOD of 500 nM for the sensing of NB derivatives. Thirumalraj et al. [48] explored the potential of an alumina-polished GC electrode for the sensing of NB. The CV curves of the alumina/GC electrode are shown in Figure 6a.

The authors optimized the pH of the solution to obtain higher electro-catalytic activity in the alumina/GC electrode. According to Figure 6a, it is clear that the alumina/GC electrode has higher catalytic properties under a pH of 5. The authors also used the DPV method for further investigations and found that DPV is a more efficient method for the sensing of NB. The current responses increase with increasing an concentration of NB (Figure 6b). Furthermore, the authors also investigated the selectivity of the alumina/GC electrode for the sensing of NB in the presence of various interfering materials. The selectivity studies are summarized in Figure 6c. It is suggested that alumina/GC electrodes have good selective nature for the sensing of NB in the presence of different interfering molecules. The alumina/GC electrode demonstrated an LOD of 0.15 μM . The alumina/GC electrode also showed good real sample studies in lake, tap, waste water, and drinking samples.

It is also reported that gold nanoparticles (Au NPs) may further improve the electrochemical performance of NB sensors. In this vein, Gupta et al. [49] reported the fabrication of an electrochemical sensing scaffold (ESS) for the determination of NB. The authors grew Au NPs on mesoporous SiO_2 microspheres and used them as a catalyst for the determination of NB. The Au NPs/ SiO_2 /GCE was employed as an NB sensor, and its electrochemical activities were checked by using the CV, DPV, and CA methods. The EIS also suggested the presence of a relatively high conductive nature and catalytic properties of the Au NPs/ SiO_2 -modified GC electrode. The DPV technique suggested that the current response increases with increasing concentrations of NB. A wide linear dynamic range of 0.1 μM to 2.5 mM has been observed. An excellent LOD of 15 nM was achieved for the sensing of NB using Au NPs/ SiO_2 /GCE. The authors also reported excellent selectivity of the Au NPs/ SiO_2 /GCE towards the determination of NB in various interfering compounds. Rameshkumar et al. [50] proposed the fabrication of an NB sensor by employing simple protocols. The authors prepared silicate sol–gel-stabilized (SSG) Ag NPs using a benign one-pot synthetic process, and a mixture of hydrazine, nitric acid, and ammonium chloride was used as the reduction solution system. Initially, the authors used colorimetric methods for the determination of Hg (II) ions and

reported an LOD of 5 μM . Furthermore, the authors explored the synthesized Ag NPs for the sensing of NB. The authors modified the active area of the bare GC electrode using Ag NPs as the catalyst and SWV as the sensing technique. The lowest LOD of 1 μM was reported by the authors for the sensing of NB using the square wave voltammetry (SWV) method. The enhanced performance of the NB sensor may be attributed to the presence of highly catalytic Ag NPs on the surface of the GC electrode.

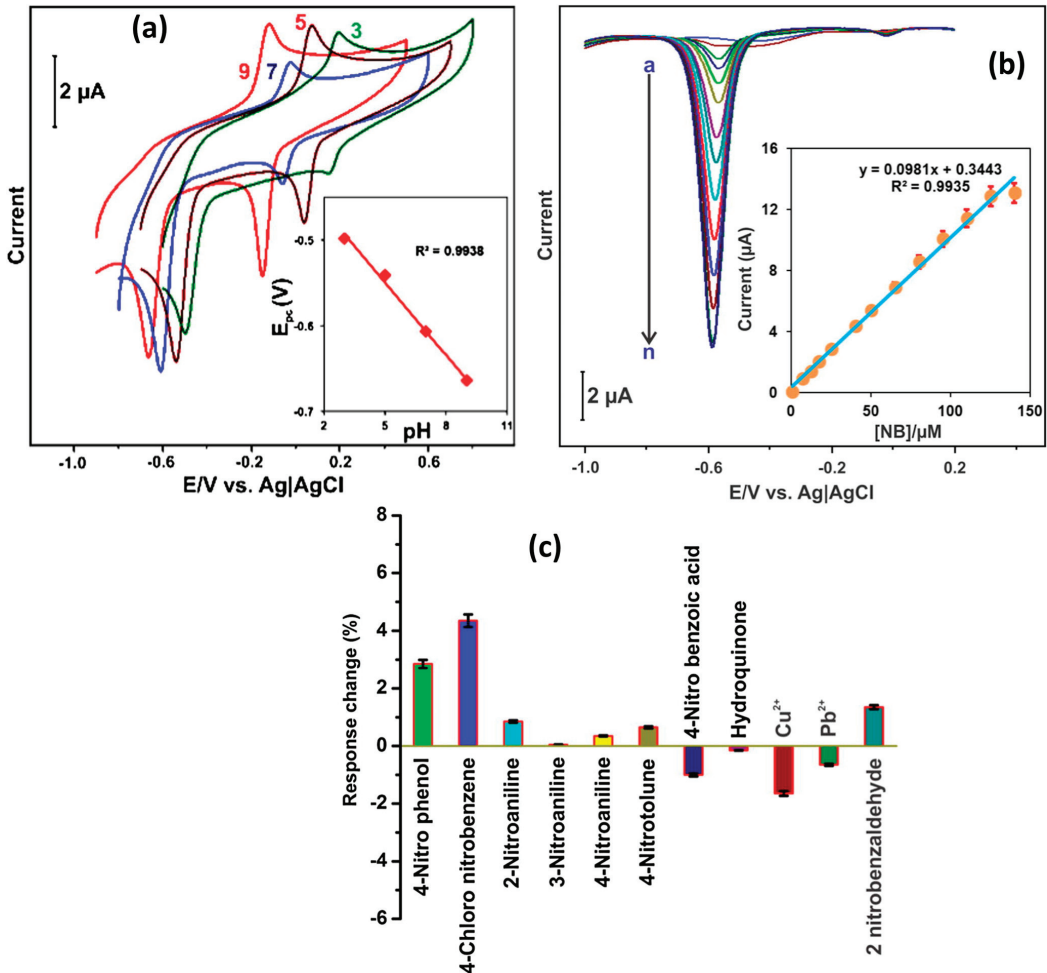


Figure 6. (a) CV patterns of the alumina/GCE for the presence of 50 μM NB at different pHs. (b) DPV curves of the alumina/GCE in the presence of different concentrations of NB. (c) Selectivity of alumina/GCE for NB sensing. Reprinted with permission [48].

2.4. Metal Dichalcogenides, Polymers, MOF, Metal Sulfides, and Other Electrode Materials

It is considered that nanoscale electrode materials play a crucial role in the design and fabrication of highly sensitive electrochemical sensors. Molybdenum disulfide (MoS_2) is a widely used dichalcogenide in the construction of electrochemical sensors and energy storage applications. The catalytic properties of MoS_2 can be further improved by incorporating MoS_2 with rGO for electrochemical reactions. Nehru et al. [51] designed and prepared flower-like MoS_2 nanosheet arrays (MoS_2 NSA)/rGO hybrid composite material using the hydrothermal method. Furthermore, a GC electrode was modified with the

prepared MoS₂ NSAs/rGO as the sensing catalyst material, and CV/DPV techniques were adopted for the determination of NB. Figure 7a presents the EIS results for the differently modified electrodes under similar conditions. It is seen that MoS₂ NSAs/rGO/GCE has low charge transfer resistance (R_{ct}) and high conductivity. The obtained CV results for the different modified GCEs are shown in Figure 7b. The MoS₂ NSAs/rGO/GCE exhibited a higher current response, which may be ascribed to the improved conductivity and better electro-catalytic properties. Figure 7c shows the CVs of the MoS₂ NSAs/rGO/GCE at different scan rates, indicating that the current response linearly increases with increasing scan rates, and it is a diffusion-controlled process (Figure 7d). In terms of the effect of mass loading, as also studied, the authors reported 6 μ L catalysts as the optimum amount for the sensing of NB (Figure 7e,f). The authors also recorded the CVs of the bare GCE, MoS₂ NSAs/GCE, and MoS₂ NSAs/rGO/GCE in the presence of NB. MoS₂ NSAs/rGO/GCE exhibited a higher current response for the detection of NB compared to bare GCE and MoS₂ NSAs/GCE (Figure 7g,h). Figure 7i shows DPV curves of the MoS₂ NSAs/rGO/GCE at various concentrations of NB, and calibration curves between the current response and concentration of NB are depicted in Figure 7j. The authors found that the current value of the MoS₂ NSAs/rGO/GCE increases linearly with increasing concentrations of NB. This proposed sensor demonstrated a linear range of 0.005 to 849.505 μ M with a sensitivity of 1.8985 μ A/ μ M cm² for the determination of NB. The MoS₂ NSAs/rGO/GCE also demonstrated an LOD of 0.0072 μ M, with excellent cyclic stability, reproducibility, and repeatability. The authors investigated the selectivity test for the constructed MoS₂ NSAs/rGO/GCE towards the determination of NB and found that MoS₂ NSAs/rGO/GCE has excellent selectivity (Figure 7k,l). The MoS₂ NSAs/rGO/GCE exhibits acceptable recoveries in real sample investigations, which suggests that MoS₂ NSAs/rGO/GCE is a promising candidate for the sensing of NB in real life applications.

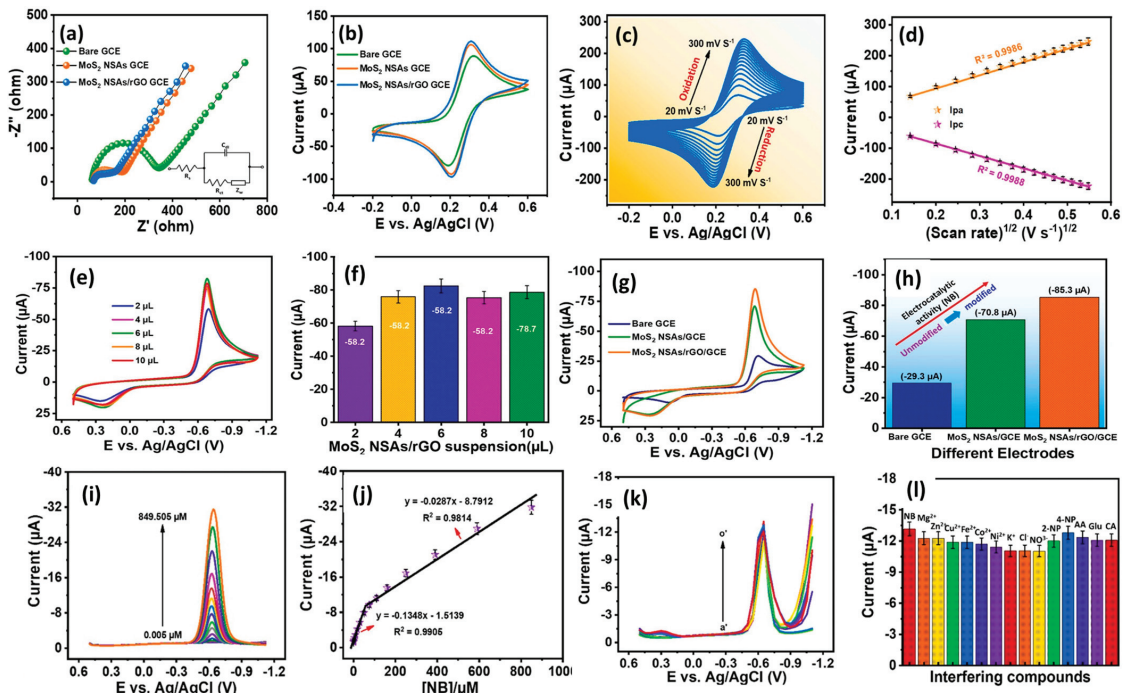


Figure 7. (a) Nyquist curves and (b) CV graphs of bare GCE, MoS₂ NSAs/GCE, and MoS₂ NSAs/rGO/GCE in [Fe(CN)₆]^{3−/4−} redox probe (0.1 M KCl). (c) CV curves of MoS₂ NSAs/rGO/GCE at different scan rates

in the redox system and (d) corresponding calibration plot between peak current response and square root of scan rate. (e) CV curves and (f) cathodic current response of different volume of MoS₂ NSAs/rGO/GCE in the presence of 200 μM NB in 0.05 M PBS. (g) CV curves and (h) cathodic current response of bare GCE, MoS₂ NSAs/GCE, and MoS₂ NSAs/rGO/GCE in the presence of 200 μM NB. (i) DPV curves of the MoS₂/NSA/rGO/GCE in different concentrations of NB. (j) Calibration curve between current responses and concentration of NB. (k) DPV curves and (l) corresponding relative peak of the MoS₂/NSA/rGO/GCE for selectivity test towards the sensing of NB in the presence of interfering substances. Reprinted with permission [51].

Recently, Papavasileiou et al. [52] reported the synthesis of 2D vanadium diselenide nanoflakes (VSe₂ NFs) for the construction of an NB sensor. The VSe₂ was coated on a GC electrode, and CV/DPV was used for the determination of NB. An LOD of 0.03 μM was reported with a linear range of 0.1 to 4 μM. The proposed sensor also showed acceptable recovery of 96% in real sample investigations. Karthik et al. [53] also reported strontium molybdate (SrMoO₄) microflower/three-dimensional (3D) nitrogen-doped reduced graphene oxide aerogels (N-rGO) for the sensitive detection of NB in water samples. The SrMoO₄/N-rGO-modified GC electrode demonstrated a remarkable LOD of 2.1 nM and linear range of 7.1 nM to 1.0 mM for the detection of NB. The performance of the SrMoO₄/N-rGO-modified GC electrode also suggested that it can be used for real sample investigations with acceptable recoveries in a range of 96.1–99.6%. Polymers are well known for their excellent conductive nature, and their composite may be a promising electrode material for various electrochemical applications. In this context, Ramirez et al. [54] prepared a polymer nanocomposite (PNC) on flowers like the hierarchical rutile phase of the titanium dioxide (TiO₂) nanorod microsphere. The authors optimized various conditions and reported the synthesis of unique surface morphologies (flower-like/cauliflower). The surface morphology of the synthesized PNC on TiO₂/GO was characterized by employing SEM and TEM analysis. The authors modified the graphite electrode with the synthesized TiO₂/GO and adopted this modified electrode as an NB sensor. The authors found that the sensing performance of the TiO₂/GO-modified graphite electrode increases with increasing concentrations of NB. An improved LOD of 2.64 ppb was reported for the detection of NB using a TiO₂/GO-modified graphite electrode. The TiO₂/GO-modified graphite electrode also demonstrated acceptable repeatability, stability, and reproducibility towards the sensing of NB. The TiO₂/GO-modified graphite electrode also exhibited excellent recovery in real sample applications. Yadav et al. [55] reported the synthesis of a Au NP-incorporated zinc-based metal–organic framework (MOF-5) and characterized by various sophisticated physicochemical techniques. The synthesized Au-MOF-5 was deposited onto the surface of a GC electrode and used as an electrochemical sensor towards the sensing of NB. The Au-MOF-5-modified GC electrode demonstrated improved electro-catalytic properties compared to the bare GC electrode, and this improved performance may be ascribed to the presence of the electrode material on the surface of the GC electrode. An interesting LOD of 1 μM with a sensitivity of 0.23 μA/μM cm² was observed for the sensing of NB by using a Au-MOF-5-modified GC electrode. The excellent selectivity of the Au-MOF-5-modified GC electrode may be attributed to synergistic interactions. Zeolitic imidazole framework (ZIF) materials are well-known highly porous materials with a high surface area. The ZIF materials may be a promising material for various electrochemical applications. It would be of great significance to propose and fabricate ZIF-based hybrid or metal-doped materials for the construction of electrochemical sensors. An et al. [56] synthesized ZIF-67 material using simple protocols, and a synthetic process is shown in Figure 8a. The Co-NC was prepared by using ZIF-67 and PS nanospheres as the template and carbonizing process.

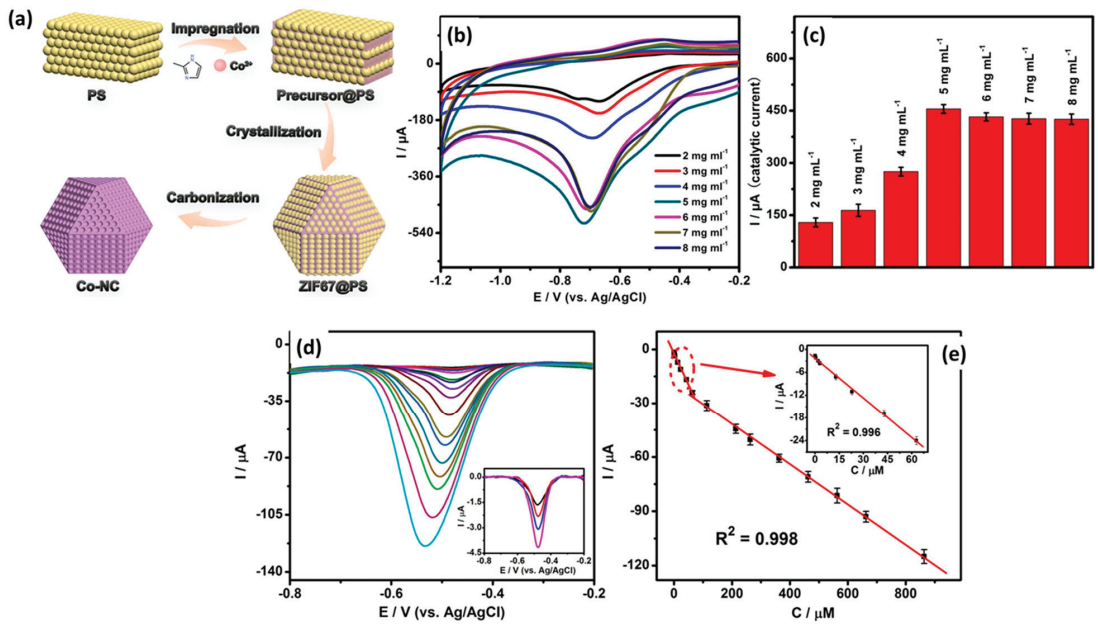


Figure 8. (a) Schematic representation for the preparation of ZIF67@PS-derived Co-NC. (b) CV curves and (c) corresponding current values of the Co-NC-800-GCE (different mass loadings) in the presence of NB. (d) DPV curves of Co-NC-800-GCE in different concentrations of NB and (e) corresponding calibration plot between peak current and concentration of NB. Reprinted with permission [56].

The authors also optimized the temperature of Co-NC synthesis and found that 800 °C is the suitable temperature for the preparation of a porous structure. The presence of N-atoms in the carbon framework improves the catalytic properties of the prepared materials. The Co-NC was coated on a GC electrode and used as a working electrode towards the sensing of NB. The loading mass of the catalyst was also optimized as 5 mg/mL using CV (Figure 8b,c). The effects of different concentrations were also studied using Co-NC-800-GCE. The DPV graphs demonstrated that the current linearly increases with increasing concentrations of NB (Figure 8d,e). The Co-NC-800-GCE demonstrated good selectivity for the determination of NB in various interfering substances. The Co-NC-800-GCE showed an LOD of 0.086 μM and linear range of 0.1 μM to 0.863 mM. The stability and better selectivity of the Co-NC-800-GCE was attributed to the presence of the robust electrode material on the GC surface. This work proposed the construction of a simple and highly selective NB sensor using robust electrode material. Li et al. [57] reported a MOF-conductive polymer composite film-modified electrode as an NB sensor, which demonstrated an LOD of 0.047 μM and linear range of 0.05 to 1 μM and 1 to 100 μM . Copper sulfide (CuS) is a semiconducting metal sulfide and has an optical band gap in the range of 1.23 to 2.0 eV. CuS has been widely used in various catalytic applications and energy-related applications such as lithium sulfur batteries and energy storage devices. It is also well reported that combining CuS may improve the properties of the hybrid composite materials. Carbon-based materials are desirable materials to form the composite materials. It is also known that doping with heteroatoms may further create defects and improve the catalytic properties of carbon-based materials. In this vein, Yuan et al. [58] designed and reported the facile synthesis of a CuS-loaded boron, nitrogen co-doped carbon composite (CuS-BCN) material using a simple synthetic method (Figure 9a). The XRD results suggested the formation of CuS-BCN, whereas SEM analysis revealed the presence of the surface structural morphology of the prepared CuS-BCN composite. The bare GC electrode was modified using

CuS-BCN as a catalyst material, and its electrochemical performance was checked using the CV, EIS, and SWV methods. The EIS results showed that the CuS-BCN-coated GC electrode has a low resistance value compared to the bare GC electrode and revealed that the CuS-BCN-coated GC electrode has high conductivity. The CV results also demonstrated the presence of the relatively higher electro-catalytic behavior of the CuS-BCN-coated GC electrode (Figure 9b). The SWV demonstrated an excellent LOD of 0.12 μM using a CuS-BCN-coated GC electrode as an NB sensor. Two linear ranges of 0.5 to 150 μM and 150 to 1000 μM were reported for the sensing of NB using a CuS-BCN-coated GC electrode (Figure 9c,d). It was observed that SWV is more sensitive compared to the CV. Improved selectivity was observed for the detection of NB in the presence of various interfering substances by using a CuS-BCN coated GC electrode. The authors also reported excellent NB recovery in real water samples using a CuS-BCN-coated GC electrode. Thus, the authors proposed that the CuS-BCN-coated GC electrode is a promising sensing candidate for the determination of NB.

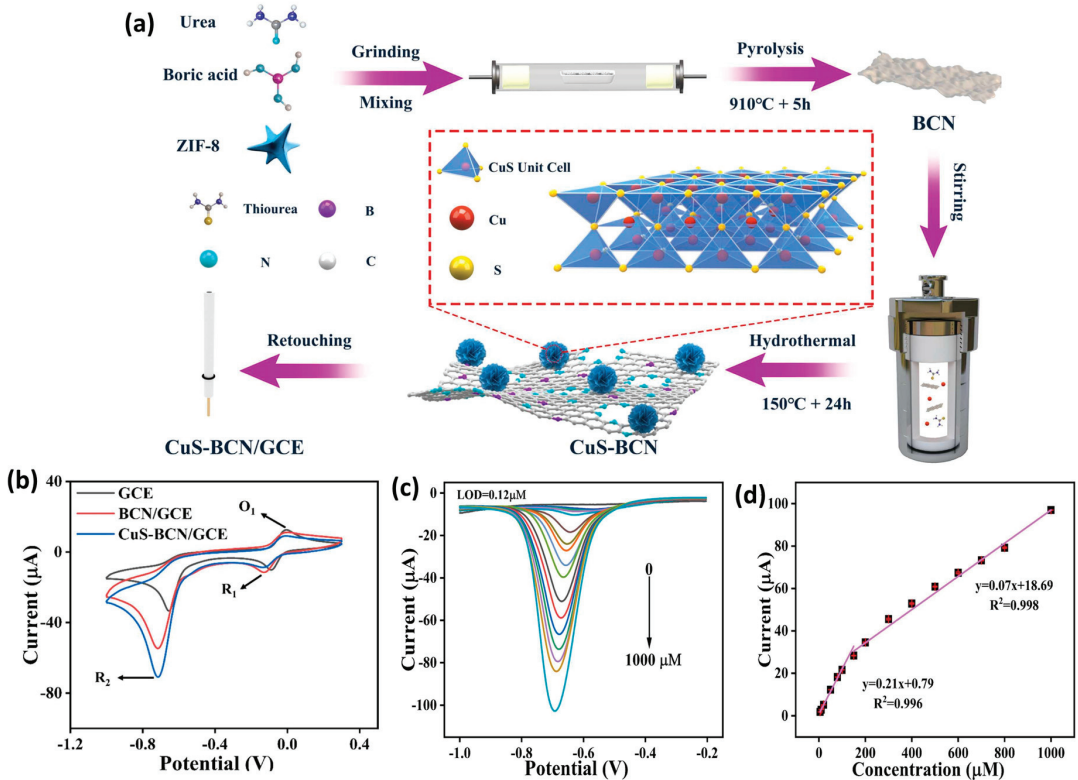


Figure 9. (a) Schematic representation for the synthesis of CuS-BCN and surface modification of GC electrode for the sensing of NB. (b) CV curves of the GCE, BCN/GCE, and CuS-BCN/GCE in the presence of NB. (c) SWV curves of the CuS-BCN/GCE at different concentrations of NB and (d) corresponding calibration plot between peak current and concentration of NB. Reprinted with permission [58].

Recent years have witnessed a rapid surge in the synthesis and use of layered double hydroxides (LDHs) due to their excellent intrinsic properties for catalytic activity and interesting adsorption properties. LDH materials can be prepared via a hydrothermal method, and their properties can be tuned by incorporating with other materials. Li et al. [59] prepared a novel electrode material (Ni/Fe LDH) functionalized with sodium

dodecyl sulfate (Ni/Fe(SDS)-LDH) for electrochemical sensing applications. This proposed 2D Ni/Fe(SDS)-LDH material was deposited on a GC electrode for electrochemical sensing studies. The electro-catalytic activities of the Ni/Fe(SDS)-LDH-modified GC electrode was evaluated by employing CV and DPV methods. The authors clearly found that the Ni/Fe(SDS)-LDH-modified GC electrode has higher catalytic properties compared to the bare GC electrode, suggesting that the presence of Ni/Fe(SDS)-LDH on the active surface of the GC electrode enhanced its catalytic behavior for the detection of NB. A low LOD of 0.093 μM was achieved using a Ni/Fe(SDS)-LDH-modified GC electrode for the determination of NB. The DPV results revealed that the Ni/Fe(SDS)-LDH-modified GC electrode has excellent selectivity for the sensing of NB, and this may be ascribed to the interactions/bonding between the NB and Ni/Fe(SDS)-LDH-modified GC electrode (Figure 10a). The CV curves of the different electrodes in the absence and presence of NB are shown in Figure 10b. It is seen that Ni/Fe(150 mg SDS)-LDH/GCE has high catalytic properties for the sensing of NB. The DPV graphs for Ni/Fe(0–200 mg SDS)-LDH/GCE are shown in Figure 10c. It is observed that Ni/Fe(150 mg SDS)-LDH/GCE has high electro-catalytic activities towards NB. The Ni/Fe(0–200 mg SDS)-LDH/GCE also demonstrated good recoveries for real sample investigations in tap water and underground water samples.

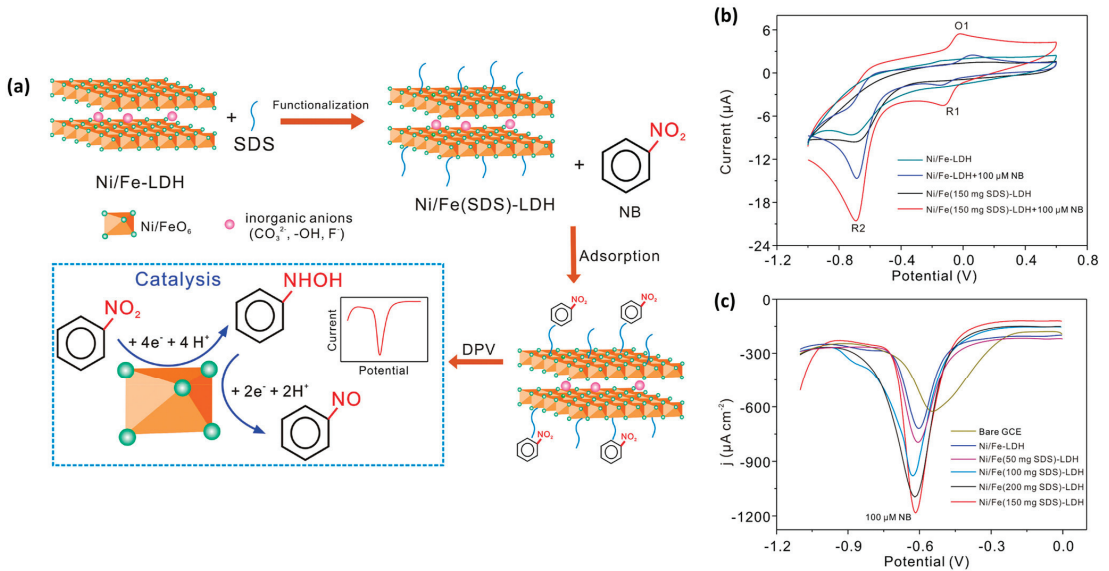


Figure 10. (a) Schematic representation for the working mechanism of Ni/Fe(SDS)-LDH-modified GC electrode towards the determination of NB. (b) CV curves of the different electrodes in absence and presence of NB. (c) DPV curves of the different electrodes in the presence of NB. Reprinted with permission [59].

Perovskite materials have received extensive attention because of their excellent optoelectronic properties. Zhang et al. [60] synthesized an inorganic perovskite-based composite with high hydrophilicity. The synthesized CsPbBr₃/TDPA exhibited high water stability due to the presence of oleylamine molecules. The ECL technique was used for the sensing of NB using CsPbBr₃/TDPA as a sensing material. The authors reported an LOD of 0.05 μM with a linear range 1 mM to 0.1 μM towards the determination of NB. Rastogi et al. [61] reported a novel sensing system consisting of palladium nanoparticle-decorated polymer-silica (Pd-GG-g-PAM-silica). The authors modified the GC electrode using Pd-GG-g-PAM-silica as a sensing catalyst, and its performance was checked by employing the LSV, DPV, and CA techniques. The Pd-GG-g-PAM-silica-modified GC electrode demonstrated good reproducibility, acceptable repeatability, and high stability for the reduction/sensing of NB. An LOD of 0.06 mM with two linear ranges of 1 to 1900 mM and 1900 to 3900 mM

were reported for the sensing of NB using a Pd-GG-g-PAM-silica-modified electrode. The authors reported that real sample recoveries are acceptable using the spike method. In the above sections, we described numerous electrode materials and their applications in the fabrication of NB sensors. The performance of the above-discussed NB sensors is summarized in Table 1 [27–61].

Table 1. Electrochemical parameters of the published NB sensors [27–61].

Catalyst	LOD (μM)	Sensitivity ($\mu\text{A}/\mu\text{M cm}^2$)	Linear Range	Technique	References
CeO ₂ NPs	0.092	1.1166	0.1–50	DPV	[27]
$\alpha\text{-Fe}_2\text{O}_3$ micro/nanorods	30.4 ppb	446 nA/ μM	-	CV	[28]
MnFe-Na/GCE	4 mM	-	-	CV	[29]
ZnSnO ₃ -gC ₃ N ₄ /GCE	2.2	0.05857	-	LSV	[30]
GO/MgO-NC/GCE	0.01	-	58.5–2333.5	DPV	[31]
MnO ₂ /GCE	0.025	-	0.03–2	DPV	[32]
ZnO-NRs@CTS-NFs	0.002	3.13	0.01–17.2; 17.2–203	DPV	[33]
MWNT-OH/GCE	0.08	-	-	DPV	[34]
f-MWCNTs/SPCE	45 nM	0.6685	50 nm to 1170 μM	Amperometry	[35]
EAG-SPE	0.06	1.445	-	CA	[36]
MMPCMs	8 nM	2.36	0.2–40	LSV	[37]
NHCPS-750-GCE	2.296	436 $\mu\text{A mM}^{-1}$	2–2610	DPV	[38]
GR-CHI	37 nM	-	0.1–594.6	Amperometry	[39]
$\beta\text{-CD}_{1.2\text{ mg}}$ /GO	0.184	-	0.5–1000	LSV	[40]
GRGO/Ni-TPP	0.14	1.277	0.5–878	DPV	[41]
CDs/SPCE	13 nM	9.36	0.1–200	DPV	[42]
EDAS/(g-C ₃ N ₄ -Ag)NC	2	0.594 $\text{A M}^{-1} \text{cm}^{-2}$	5–50	SWV	[43]
OMCN	1.52	-	-	Amperometry	[44]
Ni-Cu	40	298 $\mu\text{A}/\text{mM}$	0.1–20 mM	Polarization	[45]
Pt NPs/MPC	50 nM	-	1–200	LSV	[46]
Ag NPs/SiO ₂	0.5	-	-	SWV	[47]
Alumina/GC	0.15	-	0.5–145.5	DPV	[48]
GC/Au-MSM	15 nM	-	0.1 μM to 2.5 mM	DPV	[49]
GC/TPDT-Ag NPs	1	-	-	SWV	[50]
MoS ₂ NSAs/rGO	0.0072	1.8985	-	DPV	[51]
VSe ₂ /GCE	0.03	-	0.1–4	DPV	[52]
SMO/N-rGO	2.1	-	1.1 mM to 2.5 mM	CV	[53]
TiO ₂ /GO	2.64 ppb	40.6 $\mu\text{A}/\text{ppb}$	2–8 ppb	CV	[54]
GC/Au-MOF-5	15.3	0.43	-	CV	[55]
Co-NC	0.086	-	0.1 μM –0.863 mM	DPV	[56]
PCN-222(Fe)/p ₃ HT-p ₃ TPA/GCE	0.05	-	-	DPV	[57]
CuS-BCN/GCE	0.120	-	0.5–150, 150–1000	SWV	[58]
Ni/Fe(SDS)-LDH	0.093	15.79	10–100	DPV	[59]
CsPbBr ₃ /TDPA	0.05	-	1 mM to 0.1 μM	CV/ECL	[60]
Pd-GG-g-PAM-silica	0.06	-	1–1900, 1900–3900	DPV	[61]

3. Conclusions and Future Perspectives

We conclude that various electrode materials based on metal oxides (e.g., magnesium oxide, zinc oxide), graphene, carbon nitride, metal nanoparticles (e.g., silver or gold), metal-organic frameworks, or their composites have been reviewed for the determination of NB. The published articles demonstrated that the sensitivity and detection limit of NB sensors are the

two major key parameters, which should be improved. In addition, electrochemical sensors should have excellent anti-interfering properties. Each material has its own advantages and disadvantages. From the published reports, it has been observed that metal oxide such as the cerium oxide-modified electrode demonstrated an excellent detection limit using the DPV method. On the other hand, the Ni-Cu-modified electrode exhibits a poor detection limit using the polarization method. Thus, the polarization method is not a highly efficient approach for the electrochemical sensing of NB. The CV-method-based sensing results demonstrated lower performance compared to the amperometry. This showed that CV is also a less sensitive method for the detection of NB compared to DPV or amperometry. It is clear that the DPV method is a more efficient and sensitive technique for the determination of NB. In addition, metal nanoparticles (gold, silver, platinum, etc.) have a highly conductive nature, but they are not suitable materials for NB sensing applications due to their high cost and poor performance. It is seen that metal oxides with hybrid composite materials are more efficient materials for the construction of NB sensors. However, some challenges still exist, such as poor adhesiveness of metal oxide particles on the surface of electrodes, the use of binders which reduce the catalytic activity and conductivity of the modified electrodes, and the low conductive nature of the metal oxides. It is required to combine metal oxides with highly conducting MXenes to form efficient electrode materials for the construction of NB sensors. Thus, it is believed that the performance of NB sensors can be further enhanced by incorporating metal oxides with MXenes.

Author Contributions: Conceptualization, K.A. and T.H.O.; writing—original draft preparation, K.A.; writing—review and editing, T.H.O.; supervision, T.H.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wu, Y.; Qi, H.; Li, B.; Zhanhua, H.; Li, W.; Liu, S. Novel hydrophobic cotton fibers adsorbent for the removal of nitrobenzene in aqueous solution. *Carbohydr. Polym.* **2017**, *155*, 294–302. [CrossRef]
2. Raza, W.; Ahmad, K.; Kim, H. Fabrication of defective graphene oxide for efficient hydrogen production and enhanced 4-nitrophenol reduction. *Nanotechnology* **2021**, *32*, 495404. [CrossRef]
3. Nemakal, M.; Aralekallu, S.; Mohammed, I.; Pari, M.; Reddy, K.R.V.; Sannegowda, L.K. Nanomolar detection of 4-aminophenol using amperometric sensor based on a novel phthalocyanine. *Electrochim. Acta* **2019**, *318*, 342–353. [CrossRef]
4. Sajjan, V.A.; Aralekallu, S.; Nemakal, M.; Palanna, M.; Prabhu, C.P.K.; Sannegowda, L.K. Nanomolar detection of 4-nitrophenol using Schiff-base phthalocyanine. *Microchem. J.* **2021**, *164*, 105980. [CrossRef]
5. Ahmad, K.; Mohammad, A.; Ansari, S.N.; Mobin, S.M. Construction of graphene oxide sheets based modified glassy carbon electrode (GO/GCE) for the highly sensitive detection of nitrobenzene. *Mater. Res. Express* **2018**, *5*, 075601. [CrossRef]
6. Mohammad, A.; Ahmad, K.; Rajak, R.; Mobin, S.M. Binder Free Modification of Glassy Carbon Electrode by Employing Reduced Graphene Oxide/ZnO Composite for Voltammetric Determination of Certain Nitroaromatics. *Electroanalysis* **2018**, *30*, 274–282. [CrossRef]
7. Ahmad, K.; Mohammad, A.; Mathur, P.; Mobin, S.M. Preparation of SrTiO₃ perovskite decorated rGO and electrochemical detection of nitroaromatics. *Electrochim. Acta* **2016**, *215*, 435–446. [CrossRef]
8. Kumar, P.; Khan, M.Q.; Khan, R.A.; Ahmad, K.; Kim, H. Hydrothermal Synthesis of MnO₂/Reduced Graphene Oxide Composite for 4-Nitrophenol Sensing Applications. *Inorganics* **2022**, *10*, 219. [CrossRef]
9. Beauchamp, R.O.; Irons, R.D.; Rickert, D.E.; Couch, D.B.; Hamm, T.E.; Lyon, J.P. A critical review of the literature on nitrobenzene toxicity. *Crit. Rev. Toxicol.* **1982**, *11*, 33–84. [CrossRef]
10. Hanif, S.; Junaid, H.M.; Munir, F.; Waseem, M.T.; Majeed, S.; Shahzad, S.A. AIEE active new fluorescent and colorimetric probes for solution and vapor phase detection of Nitrobenzene: A reversible mechanochromism and application of logic gate. *Microchem. J.* **2022**, *175*, 107227. [CrossRef]
11. Huang, X.L.; Liu, L.; Gao, M.L.; Han, Z.B. A luminescent metal-organic framework for highly selective sensing of nitrobenzene and aniline. *RSC Adv.* **2016**, *6*, 87945–87949. [CrossRef]
12. Riskin, M.; Tel-Vered, R.; Lioubashevski, O.; Willner, I. Ultrasensitive surface plasmon resonance detection of trinitrotoluene by a bis-aniline-cross-linked Au nanoparticles composite. *J. Am. Chem. Soc.* **2009**, *131*, 7368–7378. [CrossRef]
13. Rajeevan, G.; Ramesh, A.; Madanan, A.S.; Varghese, S.; Abraham, M.K.; Shkhaier, A.I.; Indongo, G.; Arathy, B.K.; George, S. Efficient nanostructured Cs₂CuBr₂Cl₂ perovskite as a fluorescent sensor for the selective “Switch Off” detection of nitrobenzene. *Spectrochim. Acta Part A Mol. Biomol. Spectr.* **2024**, *318*, 124481. [CrossRef]

14. Babae, S.; Beiraghi, A. Micellar extraction and high performance liquid chromatography-ultra violet determination of some explosives in water samples. *Anal. Chim. Acta* **2010**, *662*, 9–13. [CrossRef]
15. Berg, M.; Bolotin, J.; Hofstetter, T.B. Compound-specific nitrogen and carbon isotope analysis of nitroaromatic compounds in aqueous samples using solid-phase microextraction coupled to GC/IRMS. *Anal. Chem.* **2007**, *79*, 2386–2393. [CrossRef]
16. Kumar, V.; Maiti, B.; Chini, M.K.; De, P.; Satapathi, S. Multimodal Fluorescent Polymer Sensor for Highly Sensitive Detection of Nitroaromatics. *Sci. Rep.* **2019**, *9*, 7269. [CrossRef]
17. Furstenberg, R.; Kendziora, C.A.; Stepnowski, J.; Stepnowski, S.V.; Rake, M.; Papantonakis, M.R.; Nguyen, V.; Hubler, G.K.; McGill, R.A. Stand-off detection of trace explosives via resonant infrared photothermal imaging. *Appl. Phys. Lett.* **2008**, *93*, 224103. [CrossRef]
18. Ko, H.; Chang, S.; Tsukruk, V.V. Porous substrates for label-free molecular level detection of nonresonant organic molecules. *ACS Nano* **2009**, *3*, 181–188. [CrossRef]
19. Sun, X.; Wang, Y.; Lei, Y. Fluorescence based explosive detection: From mechanisms to sensory materials. *Chem. Soc. Rev.* **2015**, *44*, 8019–8061. [CrossRef]
20. Sikka, R.; Kumar, P. Optical Sensing Capability Evaluation for Methylammonium Based Perovskites for Explosive. *J. Fluoresc.* **2023**, *33*, 1677–1682. [CrossRef]
21. Ahmad, K.; Khan, R.A. Design and fabrication of nitrogen-doped tungsten trioxide-based 4-nitrobenzene sensor. *J. Mater. Sci. Mater. Electron.* **2024**, *35*, 1061. [CrossRef]
22. Moro, G.; Bottari, F.; Loon, J.V.; Bois, E.D.; Wael, K.D.; Moretto, L.M. Disposable electrodes from waste materials and renewable sources for (bio)electroanalytical applications. *Biosens. Bioelectron.* **2019**, *146*, 111758. [CrossRef]
23. Wang, L.; Zhang, M.; Li, Y.; Chen, X.; Qin, H.; Yang, J.; Fan, S.; Wu, H. Construction of porphyrinic manganese-organic frameworks based on structural regulation for electrochemical determination of nitrobenzene in water and vegetable samples. *Front Chem.* **2024**, *12*, 1380551. [CrossRef]
24. Niyitanga, T.; Chaudhary, A.; Ahmad, K.; Kim, H. Titanium Carbide ($Ti_3C_2T_x$) MXene as Efficient Electron/Hole Transport Material for Perovskite Solar Cells and Electrode Material for Electrochemical Biosensors/Non-Biosensors Applications. *Micromachines* **2023**, *14*, 1907. [CrossRef]
25. Niyitanga, T.; Ahmad, K.; Chaudhary, A.; Kim, H. Carbon dots as efficient electrode material for hydrogen peroxide sensing applications: A mini review. *Inorg. Chem. Commun.* **2023**, *156*, 111249. [CrossRef]
26. Rhouati, A.; Berkani, M.; Vasseghian, Y.; Golzadeh, N. MXene-based electrochemical sensors for detection of environmental pollutants: A comprehensive review. *Chemosphere* **2022**, *291*, 132921. [CrossRef]
27. Sangili, A.; Annalakshmi, M.; Chen, S.M.; Chen, T.W.; Kumaravel, S.; Govindasamy, M. A Facile synthesis of ultra-small cerium oxide nanoparticles for enhanced Electrochemical Detection of Nitrobenzene in water samples. *Int. J. Electrochem. Sci.* **2018**, *13*, 6135–6143. [CrossRef]
28. Arul, N.S.; Mangalaraj, D.; Kumar, P.N.; Kim, E.; Devi, P.; Han, J.I. Synthesis and characterization of α - Fe_2O_3 Micro-/Nanorods-modified glassy carbon electrode for electrochemical sensing of nitrobenzene. *Ceram. Int.* **2015**, *41*, 5568–5573. [CrossRef]
29. Sang, Y.; Cui, Y.; Li, Z.; Ye, W.; Li, H.; Zhao, X.S.; Guo, P. Electrochemical reaction of nitrobenzene and its derivatives on glassy carbon electrode modified with $MnFe_2O_4$ colloid nanocrystal assemblies. *Sens. Actuators B Chem.* **2016**, *234*, 46–52. [CrossRef]
30. Vinoth, S.; Rajaiitha, P.M.; Pandikumar, A. In-situ pyrolytic processed zinc stannate incorporated graphitic carbon nitride nanocomposite for selective and sensitive electrochemical determination of nitrobenzene. *Compos. Sci. Technol.* **2020**, *195*, 108192. [CrossRef]
31. Kokulnathan, T.; Jothi, A.I.; Chen, S.M.; Almutairi, G.; Ahmed, F.; Arshi, N.; AlOtaibi, B. Integrating graphene oxide with magnesium oxide nanoparticles for electrochemical detection of nitrobenzene. *J. Environ. Chem. Eng.* **2021**, *9*, 106310. [CrossRef]
32. Chellappa, V.; Meenakshisundaram, N.; Annaraj, J.; Sagadevan, S. Hydrothermal synthesis of MnO_2 nanorods for efficient electrochemical detection of environmental anthropogenic pollutants and nitrobenzene. *Inorg. Chem. Commun.* **2024**, *160*, 112015. [CrossRef]
33. Kokulnathan, T.; Vishnuraj, R.; Wang, T.J.; Pullithadathil, B.; Rangarajan, M.; Ahmed, F.; Alshahrani, T. Strongly coupled design of zinc oxide-nanorods/copper tin sulfide-nanoflowers nanostructures: An electrochemical study in 4-nitrochlorobenzene detection. *Chem. Eng. J.* **2024**, *479*, 147747. [CrossRef]
34. Sang, Y.; Wang, B.; Wang, Q.; Zhao, G.; Guo, P. Insights into the electrocatalysis of nitrobenzene using chemically-modified carbon nanotube electrodes. *Sci. Rep.* **2014**, *4*, 6321. [CrossRef]
35. Govindasamy, M.; Mani, V.; Chen, S.M.; Subramani, B.; Devasenathipathy, R.; Tamilarasan, S. Highly Sensitive Amperometric Sensor for Nitrobenzene Detection Using Functionalized Multiwalled-Carbon Nanotubes Modified Screen Printed Carbon Electrode. *Int. J. Electrochem. Sci.* **2016**, *11*, 10837–10846. [CrossRef]
36. Thirumalraj, B.; Palanisamy, S.; Chen, S.M. An Amperometric Nitrobenzene Electrochemical Sensor Based on Electrochemically Activated Graphite Modified Screen Printed Carbon Electrode. *Int. J. Electrochem. Sci.* **2015**, *10*, 4173–4182. [CrossRef]
37. Ma, J.; Zhang, Y.; Zhang, X.; Zhu, G.; Liu, B.; Chen, J. Sensitive electrochemical detection of nitrobenzene based on macro-/mesoporous carbon materials modified glassy carbon electrode. *Talanta* **2012**, *88*, 696–700. [CrossRef]
38. Liu, M.; Zhang, T.; Ren, H.; Wang, L.; Meng, T.; Zhao, J.; Wang, H.; Zhang, Y. Nitrogen-doped hollow carbon nanospheres for highly sensitive electrochemical sensing of nitrobenzene. *Mater. Res. Bull.* **2018**, *104*, 15–19. [CrossRef]
39. Sakthivel, R.; Palanisamy, S.; Chen, S.M.; Ramaraj, S.; Velusamy, V.; Yi-Fan, P.; Hall, J.M.; Ramaraj, S.K. A robust nitrobenzene electrochemical sensor based on chitin hydrogel entrapped graphite composite. *J. Taiwan Inst. Chem. Eng.* **2017**, *80*, 663–668. [CrossRef]

40. Velmurugan, M.; Karikalan, N.; Chen, S.M.; Dai, Z.C. Studies on the influence of β -cyclodextrin on graphene oxide and its synergistic activity to the electrochemical detection of nitrobenzene. *J. Colloid Interface Sci.* **2017**, *490*, 365–371. [CrossRef]
41. Kubendhiran, S.; Sakthinathan, S.; Chen, S.M.; Tamizhdurai, P.; Shanthi, K.; Karuppiyah, C. Green reduction of reduced graphene oxide with nickel tetraphenyl porphyrin nanocomposite modified electrode for enhanced electrochemical determination of environmentally pollutant nitrobenzene. *J. Colloid Interface Sci.* **2017**, *497*, 207–216. [CrossRef] [PubMed]
42. Bressi, V.; Chiarotto, I.; Ferlazzo, A.; Celesti, C.; Michenzi, C.; Len, T.; Iannazzo, D.; Neri, G.; Espro, C. Voltammetric Sensor Based on Waste-Derived Carbon Nanodots for Enhanced Detection of Nitrobenzene. *ChemElectroChem* **2023**, *10*, e202300004. [CrossRef]
43. Pandiyarajan, C.; Rameshkumar, P.; Murugesan, S.; Selvaraj, M. Silver nanoparticles-supported graphitic-like carbon nitride for the electrochemical sensing of nitrobenzene and its derivatives. *J. Mater. Sci. Mater. Electron.* **2021**, *32*, 19912–19924. [CrossRef]
44. Zhang, Y.; Bo, X.; Nsabimana, A.; Luhana, C.; Wang, G.; Wang, H.; Li, M.; Guo, L. Fabrication of 2D ordered mesoporous carbon nitride and its use as electrochemical sensing platform for H_2O_2 , nitrobenzene, and NADH detection. *Biosens. Bioelectron.* **2014**, *53*, 250–256. [CrossRef]
45. Yan, Z.; Xu, Z.; Zhang, W.; Zhao, S.; Xu, Y. A Novel Electrochemical Nitrobenzene Sensor Based on NiCu Alloy Electrode. *Int. J. Electrochem. Sci.* **2012**, *7*, 2938–2946. [CrossRef]
46. Zhang, Y.; Zeng, L.; Bo, X.; Wang, H.; Guo, L. Electrochemical study of nitrobenzene reduction using novel Pt nanoparticles/macroporous carbon hybrid nanocomposites. *Anal. Chim. Acta* **2012**, *752*, 45–52. [CrossRef]
47. Rameshkumar, P.; Ramaraj, R. Electroanalysis of nitrobenzene derivatives and nitrite ions using silver nanoparticles deposited silica spheres modified electrode. *J. Electroanal. Chem.* **2014**, *731*, 72–77. [CrossRef]
48. Thirumalraj, B.; Palanisamy, S.; Chen, S.M.; Thangavelu, K.; Periakaruppan, P.; Liu, X.H. A simple electrochemical platform for detection of nitrobenzene in water samples using an alumina polished glassy carbon electrode. *J. Colloid Interface Sci.* **2016**, *475*, 154–160. [CrossRef]
49. Gupta, R.; Rastogi, P.K.; Ganesan, V.; Yadav, D.K.; Sonkar, P.K. Gold nanoparticles decorated mesoporous silica microspheres: A proficient electrochemical sensing scaffold for hydrazine and nitrobenzene. *Sens. Actuators B Chem.* **2017**, *239*, 970–978. [CrossRef]
50. Rameshkumar, P.; Viswanathan, P.; Ramaraj, R. Silicate sol–gel stabilized silver nanoparticles for sensor applications toward mercuric ions, hydrogen peroxide and nitrobenzene. *Sens. Actuators B Chem.* **2014**, *202*, 1070–1077. [CrossRef]
51. Nehru, R.; Kumar, B.S.; Chen, C.W.; Dong, C.W. Sphere-like MoS_2 nanosheet arrays/reduced graphene oxide hybrid electrocatalyst for accurate electrochemical monitoring of toxic pollutant. *J. Environ. Chem. Eng.* **2022**, *10*, 108687. [CrossRef]
52. Papavasileiou, A.V.; Antonatos, N.; Luxa, J.; Děkanovský, L.; Ashtiani, S.; Fomekong, R.L.; Sofer, Z. Two-dimensional VSe_2 nanoflakes as a promising sensing electrocatalyst for nitrobenzene determination in water samples. *Electrochim. Acta* **2024**, *475*, 143653. [CrossRef]
53. Karthik, R.; Chavan, P.R.; Sukanya, R.; Dhakal, G.; Shim, J.-J.; Breslin, C.B. Flower-like strontium molybdate anchored on 3D N-rich reduced graphene oxide aerogel composite: An efficient catalyst for the detection of lethal pollutant nitrobenzene in water samples. *Compos. Part B Eng.* **2023**, *256*, 110649. [CrossRef]
54. Ruiz-Ramirez, M.M.; Silva-Carrillo, C.; Hinostroza-Mojarro, J.J.; Rivera-Lugo, Y.Y.; Valle-Trujillo, P.; Trujillo-Navarrete, B. Electrochemical sensor for determination of nitrobenzene in aqueous solution based on nanostructures of TiO_2/GO . *Fuel* **2021**, *283*, 119326. [CrossRef]
55. Yadav, D.K.; Ganesan, V.; Sonkar, P.K.; Gupta, R.; Rastogi, P.K. Electrochemical investigation of gold nanoparticles incorporated zinc based metal-organic framework for selective recognition of nitrite and nitrobenzene. *Electrochim. Acta* **2016**, *200*, 276–282. [CrossRef]
56. An, S.; Shang, N.; Zhang, J.; Nsabimana, A.; Su, M.; Zhang, S.; Zhang, Y. Fabrication of electrocatalytically active, cobalt-embedded nitrogen-doped ordered macroporous carbon for sensitive detection of nitrobenzene. *Colloids Surf. A Physicochem. Eng. Asp.* **2022**, *653*, 130078. [CrossRef]
57. Li, Y.P.; Zhuge, R.X.; Zhang, T. MOF-Conductive polymer composite electrode as electrochemical sensor of nitrobenzene. *Inorg. Chem. Commun.* **2023**, *154*, 110904. [CrossRef]
58. Yuan, C.; Li, N.; Zhang, X.; Wang, Y.; Zhou, Z.; Zhang, L.; Zhou, M.; Hu, G. Flower-like copper sulfide-decorated boron-nitrogen co-doped carbon-modified glassy carbon electrode for selective and sensitive electrochemical detection of nitrobenzene in natural water. *Colloids Surf. A Physicochem. Eng. Asp.* **2023**, *675*, 132011. [CrossRef]
59. Li, S.S.; Fang, J.H.; Li, L.; Zhu, M.; Zhang, F.; Zhang, B.Y.; Jiang, T.J.; Zhang, Y.X. An ultra-sensitive electrochemical sensor of Ni/Fe-LDH toward nitrobenzene with the assistance of surface functionalization engineering. *Talanta* **2021**, *225*, 122087. [CrossRef]
60. Zhang, W.X.; Chen, J.S.; Liu, X.P.; Mao, C.J.; Jin, B.-K. An electrochemiluminescent sensor based on hydrophilic $CsPbBr_3/TDPA$ nanocrystals for sensitive detection of nitrobenzene. *Sens. Diagn.* **2023**, *2*, 445–456. [CrossRef]
61. Rastogi, P.K.; Ganesan, V.; Krishnamoorthi, S. Palladium nanoparticles incorporated polymer-silica nanocomposite based electrochemical sensing platform for nitrobenzene detection. *Electrochim. Acta* **2014**, *147*, 442–450. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

The Physical Properties and Crystal Structure Changes of Stabilized Ice Cream Affected by Ultrasound-Assisted Freezing

Anna Kamińska-Dwórznička^{1,*} and Anna Kot^{2,*}

¹ Department of Food Engineering and Process Management, Faculty of Food Sciences, Warsaw University of Life Sciences (WULS-SGGW), Nowoursynowska 159C, 02-776 Warsaw, Poland

² Department of Functional and Organic Food, Institute of Human Nutrition Sciences, Warsaw University of Life Sciences (WULS-SGGW), Nowoursynowska 159C, 02-776 Warsaw, Poland

* Correspondence: anna_kaminska1@sggw.edu.pl (A.K.-D.); anna_kot1@sggw.edu.pl (A.K.)

Abstract: In this study, the effect of ultrasound-assisted freezing with frequencies of 21.5 and 40 kHz, and a power of 2.4 kW in the chopped mode of milk ice cream in comparison to a standard freezer on the freezing course and formed crystal structure was examined. The first part of the research included the preparation of an ice cream mixture on the basis of skimmed milk with the addition of an emulsifier, locust bean gum, xanthan gum, ι -carrageenan and a reference mixture without stabilizer addition. Ultrasound-assisted freezing shortened the processing time of both stabilized and non-stabilized ice cream. Stabilized samples of milk ice cream exposed to ultrasound (US) at a frequency of 21.5 kHz were characterized by the most homogeneous structure, consisting of crystals with the smallest diameters among all of the tested samples, the size of which, after 3 months of storage at $-18\text{ }^{\circ}\text{C}$, was $7.8\text{ }\mu\text{m}$ (for the reference sample, it was $14.9\text{ }\mu\text{m}$). The ice recrystallization inhibition (IRI effect) in the samples after US treatment with a frequency of 40 kHz was also observed, regardless of the addition of stabilizers, which may suggest that sonication with these parameters could replace or limit the addition of these substances.

Keywords: ultrasound-assisted freezing; milk ice cream; stabilizers; crystal structure

Citation: Kamińska-Dwórznička, A.; Kot, A. The Physical Properties and Crystal Structure Changes of Stabilized Ice Cream Affected by Ultrasound-Assisted Freezing. *Processes* **2024**, *12*, 1957. <https://doi.org/10.3390/pr12091957>

Academic Editor: Dariusz Dziki

Received: 31 July 2024

Revised: 10 September 2024

Accepted: 10 September 2024

Published: 12 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The most popular types of ice cream include milk ice cream, obtained by freezing a liquid pasteurized ice cream mixture, consisting of whole or skimmed milk (in liquid, concentrated or powdered form), cream, sugar and various flavor additives. The consistency and melting of ice cream are strongly related to the coalescence of destabilized fat globules that occurs during freezing, and the ratio of free and bound water content [1]. When freezing an ice cream mixture, a three-dimensional network of partially aggregated fat globules is formed, which surrounds air bubbles, stabilizing the air phase and, as a result, improving the ice cream's resistance to melting [2]. Emulsifiers and stabilizers are usually added to improve consistency and limit the unfavorable phenomenon of ice recrystallization. It occurs due to temperature fluctuations during storage and transportation, causing unfavorable ice crystals to grow. This is the main problem in the production of frozen food, especially ice cream, but with the appropriate use and combination of stabilizers, which are water-binding polysaccharides, as well as the relevant selection of other ingredients, it is possible to obtain a product that is more resistant to temperature changes [3].

By improving the conditions of the heat transfer process during ultrasound-assisted freezing in the production of milk ice cream, the rate of ice crystal nucleation is increased and freezing is faster, while controlling the size of ice crystals. Acoustic cavitation initiated by ultrasonic waves causes micro-streaming, which promotes the formation of ice nuclei, accelerates heat and mass transfer, and regulates the shape and size distribution of ice crystals during freezing. Moreover, the formation of high pressure by the breakdown of cavitation bubbles contributes to an increase in the equilibrium freezing temperature of

the liquid and, as a result, to the nucleation of crystals, which are smaller in size in the finished product. This may also shorten the freezing process time. As a result of these micro turbulences, secondary nucleation occurs, which involves cracking and breaking the structure of crystals that have stopped forming [4].

Sonication is a promising method in ice cream production that can shorten processing time, improve quality and efficiency, and ensure safety while extending the shelf life. It was discovered that the use of ultrasound activates phase changes in biological cells thanks to cavitation. As a result of the emission of acoustic waves, gas bubbles are formed, which increase in volume and disappear rapidly. The second effect is the alternating dynamic compression and expansion of the material sample by an acoustic wave acting in the material, with additional solid–liquid cavitation [5,6]. During the crystallization process, it controls the growth of ice crystals and limits sensory and nutrient changes as a result of the non-thermal process. When used during the homogenization of the mixture, a more effectively dispersed emulsion is created, and the resulting ice cream has a smoother and softer texture due to limited crystal growth [7]. It was already observed that ultrasound homogenization contributed to smaller ice crystals and had a positive influence on the ice crystal structure in milk ice cream. Ice crystal diameters lower than even 10 μm were observed for the sample after US homogenization and without any stabilizer addition [8].

Currently, the majority of research on ultrasonic-assisted freezing focuses on plant and meat materials, while few publications concern food ice cream. Their example mainly examined the influence of ultrasound in the pasteurization or homogenization process, not the freezing process itself.

The ultrasonic pasteurization and homogenization of an ice cream mixture with the addition of tomato seed oleo-gel were tested [9]. These treatments allowed for reducing the size of the ice crystals formed during freezing and accelerated the formation of a three-dimensional structure. Moreover, during pasteurization, ultrasonic waves reduced the fat and carbohydrate content, lowering the overall energy value of the ice cream. In turn, using the example of mango sorbet showed the effect of ultrasonic-assisted freezing on the crystal structure, obtaining a product with smaller crystals, and a more even distribution and regular shape. This way, the quality of the sorbet was improved, while the freezing time was reduced, which is important from an economic point of view [10]; however, it is also pivotal to check this effect on the classic milk ice cream as a product with a higher fat and protein content. It is important to answer the question about the role of US not during the homogenization part, but during the freezing of the emulsion.

2. Materials and Methods

2.1. Materials

For milk ice cream production, the following ingredients were used: 45.5% addition of 0.5% fat UHT milk (Mlekpól, Grajewo, Poland), 7% milk powder, low-fat content, 40% or 30% fat cream (Piątnica, Poland), 7% sucrose (Diamant, Miejska Górka, Poland), 0.4% emulsifier (Fooding, Shanghai, China) and 0.1% stabilizers (for the control sample, 0.1% more of powdered milk instead). The mixture recipe was modified by the addition of stabilizers: 0.01% ι -carrageenan (Sigma for Merck, St. Louis, MO, USA), 0.02% xanthan gum (Fooding, Shanghai, China) and 0.07% LBG (Locust Bean Gum) (Fooding, Shanghai, China).

2.2. Preparation of Ice Cream Mixes

First, the dry and liquid ingredients were weighed and mixed separately, then combined and homogenized using a BOSH MaxoMixx blender (Gerlingen, Germany) with a power of 750 W for about 2 min to obtain a uniform mixture.

Then, the ice cream mixes were pasteurized at a temperature of 75 °C for 1.5 min in Thermomix TM31-1 and poured into a glass beaker, which was placed in ice water and left to lower the temperature to 25 °C. After that, 100 mL was isolated from the cooled mixture and prepared for the other analysis. The remaining volume was stored at 4 °C for 24 h for the maturation step of the ice cream mixture.

2.3. Physicochemical Property Evaluation

2.3.1. Density

The density of the prepared ice mixture was determined using the pycnometric method using a 25 cm³ vacuum pycnometer; the method has already been shown [1,11]. The density value of the mixture was obtained as the ratio of mass to the known volume of the sample. The measurement was performed in triplicate.

2.3.2. Overrun

To determine the overrun (O) of the finished ice cream, a glass cylinder with a volume of 25 cm³ was weighed, then filled with the sample and weighed again. This way, the ice mixture was treated before and after freezing, obtaining the masses of samples w_1 and w_2 , respectively, to the formula [11,12]:

$$O [\%] = \frac{w_1 - w_2}{w_2} \times 100\%$$

2.3.3. pH

The pH value was measured with the Electrode Elmetron EPP3t (Zabrze, Poland) with the temperature sensor Pt-1000B. The measurement was always taken under the same conditions of the temperature and humidity in the same laboratory. The electrode was immersed in the sample of the tested ice cream, and the result was displayed on a small screen of the device. The test was performed in triplicate, according to the device producer's recommendation.

2.3.4. Freezing Point Analysis

The freezing point of the stabilized and non-stabilized ice cream mixtures was determined using a Marcel OS 3000 osmometer (Marcel, Zielonka, Poland), as also shown in the previous study [10]. The device measures the freezing temperature for 10 µL of the sample with an accuracy of 0.002 °C. Measurements were performed in triplicate.

2.4. Freezer Ice Cream Production

The milk ice cream mixture preparation samples were then frozen in two small freezers, G3 Ferrari G20035 Cremosa (Modena, Italy). The freezing conditions for stabilized and non-stabilized samples were exactly 40 min of the process. The change of temperature was recorded in 60 s intervals by the thermocouples connected to the MPI-LAB temperature recorder (Metronic Instruments, North Shields, UK) connected to the PC. After freezing time, the temperature of both samples was at the level of −6 °C. Frozen samples of milk ice cream were packed into 2 boxes for 300 mL (for each variant) and stored for further microscopy analysis after 24 h, 1 week, 1 and 3 months.

2.5. Ultrasound-Assisted Freezing

The cubic plastic boxes containing 300 mL of ice creams were placed in a bath of a two-chamber cryostat-type freezer made of stainless steel, with an aggregate for freezing closed and open samples, filled with the cooling liquid (standard coolant—propylene glycol, Borygo, Boryszew ERG, Poland). The cryostat allows for the study of ultrasound-assisted freezing processes on the principle of immersion, i.e., by immersion in a non-boiling liquid. The ultrasound frequency of the right chamber was at the level of 21.5 kHz ± 10% and the left chamber at 40 kHz ± 10%, power 2.4 kW (chopped operation), and it has already been shown in the previous study [10]. The process was carried out in two chambers at the temperature of the bath −12 °C for the proper time to adjust −6 °C in the center of each sample. The following variants were carried out:

- Cryostat without the US for samples with (CS) and without stabilizers (CNS);
- Cryostat 21.5 kHz for samples with (CS21) and without stabilizers (CNS21), 10 min of US treatment, chopped operation;

- Cryostat 40 kHz for samples with (CS40) and without stabilizers (CNS40), 10 min of US treatment, chopped operation.

The temperature changes were monitored using two thermocouples placed in the thermal centers of the samples in both chambers (records were taken every 120 s by the MPI-LAB temperature recorder connected to the PC). The freezing process was carried out to achieve $-6\text{ }^{\circ}\text{C}$ in the thermal center of the tested samples (the temperature was chosen according to the temperature of the samples frozen in a conventional freezer).

In order to analyze the course of freezing for both methods (US cryostat and conventional freezer), the recorded data were used to prepare freezing curves for each method and each sample (Figure 1). The method has already been described by the authors [13]. Freezing processes and measurements were performed in triplicate.

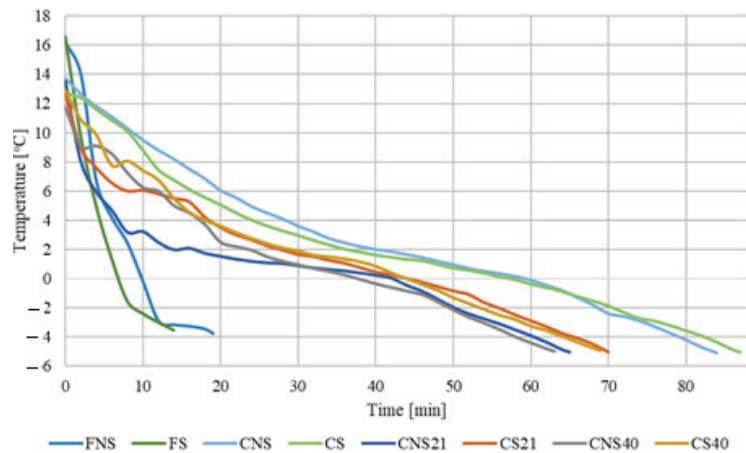


Figure 1. Characteristics of averaged freezing curves for stabilized and non-stabilized ice cream frozen used conventional and immersion methods without and with ultrasound support at frequencies of 21.5 and 40 kHz.

2.6. Microscopy Structure Analysis

To prepare the sample for image analysis (after the production cycle), a small piece of ice cream was taken from the center of the plastic box from at least 2 different locations, a minimum of 3 cm away from the ice cream surface, and placed on an object slide by using a spatula and then covered by the cover slip placed on the top of the sample (which gives at least two slides for each sample). The samples were prepared in a freezing chamber and transferred into a microscope occupied with the cooling system (Linkam LTS420). This system eliminated the influence of the ambient temperature.

The recrystallization process was then analyzed based on the images of ice crystals taken after 24 h, 1 week, 1 and 3 months at a temperature of $-18\text{ }^{\circ}\text{C}$. A microscope (Olympus BX53) with the cooling system Linkam LTS420, with $10\times$ and $50\times$ lenses and a camera (Olympus SC50), was used. The microscope was occupied with an LED light source with a power equivalent to 30 W for a halogen lamp. Camera settings: exposure time $31\text{ }\mu\text{s}$ – 2.74 s , pixel size $2.2 \times 2.2\text{ }\mu\text{m}$, refresh 15 frames per second.

The obtained images were then analyzed using NIS Elements D software. From 300 to 500 crystals were manually marked for a particular sample, and then the area, equivalent diameter and standard deviation were calculated using the NIS Elements D Imaging software (version 5.30.00, Nikon), based on the method used before [10].

2.7. Statistical Analysis

For the statistical analysis of the obtained results (in order to determine the effect of the addition of stabilizers), a one-way analysis of variance was used in a completely

random design, and homogeneous groups were determined using the Tukey test at the significance level of $\alpha = 0.05$. The R programming language was used in the R Commander environment (version 4.3.0).

The freezing curves were prepared based on the temperature changes, using Microsoft Excel 365 Office (version 2208, 2022). Then, regression curves were plotted on their basis, obtaining the corresponding equations, and regression coefficients were compared using a one-way analysis of variance and the Tukey test.

The frequency distribution of the crystal size was computed using Microsoft Excel (version 2208, 2022) macro data analysis. The relative frequency of any class interval was calculated as the number of crystals in that class divided by the total number of crystals and expressed as a percentage. The parameter X_{50} was analyzed as an average diameter (D_A) for 50% of crystals in the sample. The average diameter (D_A) and standard deviations (SD) of each class were also calculated. The method has already been shown [10,13].

3. Results and Discussion

3.1. Physicochemical Properties of Ice Cream

The density of the ice cream mixture affects the overrun value of the finished ice cream and depends mainly on the composition of the mixture, i.e., the content of fat-free dry matter, sugars and stabilizers. Generally, its value is estimated in the range of 1.0544–1.1232 g/cm³ and the addition of stabilizers could significantly increase the density of ice cream mixtures [14]. Statistical analysis for the obtained ice cream mixes did not show a significant difference (Table 1); therefore, it was concluded that the stabilizing additives did not affect the density of the examined samples.

Table 1. Density, pH, overrun and freezing point.

Sample	Density	pH	Overrun	Freezing Point
NS	1.161 ± 0.030 ^a	6.35 ± 0.01 ^a	16.4 ± 1.6 ^b	−1.8 ± 0.1 ^a
S	1.157 ± 0.221 ^a	6.31 ± 0.00 ^a	25.2 ± 0.7 ^a	−2.4 ± 0.2 ^b

Explanatory notes: NS—reference sample (without any stabilizers), S—LBG, xanthan with the addition of ι -carrageenan, ($\bar{x} \pm sd$); ^{a,b} mean values denoted by different superscripts differ statistically at $\alpha = 0.05$.

It was proved that a stabilizing mix of LBG and ι -carrageenan with the combination of guar gum significantly increased the density of ice cream based on whey [1]; however, this effect could be related to the whey protein properties. Therefore, the negative effect of the high density of the ice cream mix on the overrun might have been observed [8].

During mixing and freezing, reactions of proteins and surfactants occur, resulting in the formation and stabilization of the ice cream mixture as a foam [15]. In this particular research, ice cream with added stabilizers contained in its structure 25.2% air, thus having a higher overrun value compared to ice cream without the addition of stabilizers (overrun at the level of 16.4%) (Table 1). Hydrocolloids affect the degree of aeration of the ice cream with varying effectiveness, and depend on the type, presence and concentration of other ingredients, including another hydrocolloid and its synergistic effect [3,16]. It was also proved [17] that the overrun of milk ice cream ranged from 65.04% to 72.54%; nonetheless, it is strictly dependent on the ice cream recipes [3]. According to research by Kot et al. (2023) and Romulo et al. (2021) [8,15], the use of locust bean gum increases the aeration level of ice cream more effectively than the addition of guar gum or a combination of these hydrocolloids. In turn, other studies observed that the addition of xanthan gum to rice ice cream resulted in better aeration than the addition of guar gum [18]. Moreover, it has already been observed that ι -carrageenan addition improves milky ice cream aeration (overrun higher than 30%) and elongated the melting time of the tested samples [8].

The pH of the tested samples did not show a significant difference (Table 1), which has already been presented for stabilized and non-stabilized whey and milky ice cream [1,8].

An important ingredient of any type of ice cream is carbohydrates, which lower the freezing point of the ice cream mixture [8,19,20]. Ice cream without stabilizers had a lower freezing point than the reference sample without any addition (Table 1). Góral et al. (2018) examined the freezing point of ice cream based on a coconut drink with different levels of LBG addition, and they discovered that the higher concentration of LBG lowered the freezing point of the examined samples [12]. On the other hand, it was also proved that the addition of ι -carrageenan did not significantly influence the freezing point of milky ice cream in comparison to the addition of its hydrolysates with a lower molecular mass [8].

3.2. Freezing Course Analysis

The time and course of freezing are related to the arrangement and size of ice crystals. By shortening the process time, it is possible to create smaller and more evenly distributed crystals in the sample, while improving the quality of frozen food [3,13].

Based on the graph of freezing curves (Figure 1) and the values of regression coefficients (statistical analysis—Supplementary Materials, Figure S1), it can be concluded that the course of freezing in the freezer of both examined samples was similar. The lack of significant statistical differences indicates the lack of influence of stabilizers on the course of freezing and the linearity of temperature changes during it. Therefore, the use of a combination of locust bean gum, xanthan gum and ι -carrageenan did not affect the nature of freezing. In research on the freezing process of strawberry sorbet, it was found that the addition of a combination of κ - and ι -carrageenan slowed down the freezing rate.

In the immersion method under the influence of sonication, the freezing time was effectively shortened, regardless of the set frequency compared with the freezing time of the reference sample (Figure 1, Table 2).

Table 2. Freezing time.

Sample	Time [min]	Sample	Time [min]
FNS	19 ± 4 ^c	FS	14 ± 0 ^c
CNS	84 ± 0 ^a	CS	87 ± 4 ^c
CNS21	65 ± 1 ^b	CS21	70 ± 3 ^b
CNS40	63 ± 4 ^b	CS40	69 ± 4 ^b

Explanatory notes: NS—reference sample (without any stabilizers), S—LBG, xanthan with the addition of ι -carrageenan, F—samples from the freezer, C—samples from cryostat, C21—samples from cryostat US 21.5 kHz, C40—samples from cryostat US 40 kHz. ($\bar{x} \pm sd$); ^{a-c} mean values denoted by different superscripts differ statistically at $\alpha = 0.05$.

The regression coefficient curves obtained from the equations (statistical analysis—Supplementary Materials, Figures S2 and S3) show that the freezing curves are parallel to each other, and the course of freezing for each of the cryostat samples did not differ significantly. Therefore, ultrasound did not relevantly affect temperature changes during the process, which is surprising because it shortened its time. It is possible that during the emission of waves with a frequency of 40 kHz, simultaneous cavitation and heat generation were more intense, and therefore, more temperature fluctuations were observed than during freezing at a frequency of 21.5 kHz (Figure 1). This would also explain the lack of influence of the frequency of the given waves on the freezing time (Table 2), because during sonication at a higher frequency, more heat was released, which required a longer cooling time for the product. In theory, freezing should be accelerated as a result of intensified cavitation [6].

The reason for this result could be the even distribution of heat throughout the product, caused by microturbulence in the coolant that create a mixing effect. Moreover, cavitation effects contribute to the disintegration of air bubbles, which hinder heat conduction and effective freezing [21]. Their removal by ultrasonic waves allows for this process to be accelerated. The factors explaining this effect may be the same as in the case of ice cream

without stabilizers, i.e., more uniform cooling of the product and reduced air insulation. As in the case of ice cream without stabilizers, the statistical analysis excluded the influence of stabilizers on the effect of ultrasound; therefore, in the production of dairy ice cream with the recipe given in this study, sonication can be considered as a treatment beneficial in terms of freezing time, regardless of the presence of added stabilizers. The same effect was also noticed for sorbet frozen in similar conditions [10]. In this sorbet study, despite similar conditions, the total time for sample freezing was much shorter (not longer than 16 min in cryostat) than for milky ice cream in this particular study (Table 2). This effect can be explained by the different water content of the ice cream mixes and totally different basal ingredients with a higher fat content. The presence of fat globules in milky ice cream and air bubbles pressed into the mixture during preparation could influence the heat transfer of the freezing process that is US-assisted, which was also already confirmed in the previous study [21–23].

In previous research [24], the ultrasound effect was used in the process of freezing chicken breast samples in a cooling tunnel with forced air flow, which shortened the total freezing time by approximately 11%. A similar conclusion was reached by [6], who found that thanks to the use of ultrasonic waves, the process of freezing potato samples was significantly shortened. In turn, in carp and pork samples, a reduction in the size and more uniform distribution of ice crystals was observed after ultrasonic-assisted immersion freezing during 180 days of storage compared to the samples air-frozen and immersion-frozen without sonication.

3.3. Ice Crystal Structure Analysis

The sizes and shapes of ice crystals in the ice cream structure are the main factors determining the fine texture of this product. Ice cream is considered a good-quality product only when the ice crystals' diameters are smaller than the detection threshold, which means, for some of the researchers, lower than 25 μm , and the rest, even more than 50 μm [3,25].

Comparing the ice crystal diameters from the classic freezer and cryostat without the US (Tables 3 and 4), we can conclude that stabilizers effectively stopped the growth of ice crystals after all analyzed periods of storage.

Table 3. Comparison of ice crystal size distribution in ice cream samples frozen in a conventional freezer and cryostat with and without US.

Variant		Minimal Diameter [μm]	Maximal Diameter [μm]	D_A in Class with the Highest Frequency [μm]
FNS	24 h	2.38	14.72	7.39 ± 2.23 ^{d,e,f}
	1 week	3.03	16.10	7.53 ± 2.84 ^d
	1 month	4.57	24.49	7.45 ± 2.21 ^{d,e}
	3 months	8.02	29.75	17.99 ± 4.05 ^a
CNS	24 h	1.98	18.38	6.60 ± 2.59 ^{e,f}
	1 week	3.51	21.58	6.41 ± 2.20 ^f
	1 month	3.00	20.55	8.91 ± 3.39 ^c
	3 months	4.05	27.08	16.90 ± 4.27 ^a
CNS21	24 h	2.74	16.87	6.54 ± 3.28 ^{e,f}
	1 week	2.03	15.25	6.78 ± 4.45 ^{d,e,f}
	1 month	4.12	22.32	6.86 ± 1.95 ^{d,e,f}
	3 months	3.77	24.42	11.84 ± 4.44 ^b
CNS40	24 h	1.93	8.34	4.44 ± 1.54 ^g
	1 week	1.96	10.96	4.32 ± 1.68 ^g
	1 month	1.80	19.35	9.31 ± 4.40 ^b
	3 months	3.29	29.85	11.07 ± 4.64 ^b

Explanatory notes: NS—reference sample (without any stabilizers), F—samples from the freezer, C—samples from cryostat, C21—samples from cryostat US 21.5 kHz, C40—samples from cryostat US 40 kHz. ^{a–g} means in the same row indicated by different letters were significantly different ($\alpha = 0.05$).

Table 4. Comparison of ice crystal size distribution in ice cream samples frozen in a conventional freezer and cryostat, with and without US.

Variant		Minimal Diameter [μm]	Maximal Diameter [μm]	D_A in Class with the Highest Frequency [μm]
FS	24 h	1.43	17.08	5.45 ± 3.08 ^{h,i}
	1 week	2.21	17.27	5.82 ± 2.43 ^{g,h}
	1 month	3.17	12.65	6.03 ± 1.83 ^{f,g,h}
	3 months	2.70	28.58	12.90 ± 4.51 ^b
CS	24 h	3.71	8.56	6.72 ± 1.07 ^{e,f,g}
	1 week	2.35	14.54	6.49 ± 2.85 ^{e,f,g}
	1 month	3.98	10.70	6.77 ± 1.38 ^{e,f}
	3 months	3.73	27.32	14.89 ± 4.55 ^a
CS21	24 h	2.06	7.98	5.29 ± 1.45 ^{h,i,j}
	1 week	2.81	10.26	5.11 ± 1.96 ^{h,i,j}
	1 month	4.12	15.25	6.86 ± 3.08 ^e
	3 months	3.77	24.42	7.84 ± 3.35 ^d
CS40	24 h	1.91	7.12	4.36 ± 1.26 ^{ij}
	1 week	1.81	12.51	4.58 ± 1.80 ^j
	1 month	1.96	20.63	9.33 ± 2.92 ^c
	3 months	2.15	20.75	9.90 ± 2.47 ^c

Explanatory notes: S—LBG, xanthan with the addition of ι -carrageenan, F—samples from the freezer, C—samples from cryostat, C21—samples from cryostat US 21.5 kHz, C40—samples from cryostat US 40 kHz. ^{a–j} means in the same row indicated by different letters were significantly different ($\alpha = 0.05$).

After 3 months of storage, the biggest ice crystals were noticed for a stabilized sample of ice cream prepared with cryostat use, at the level of $15 \mu\text{m}$. The crystals in the stabilized sample from a cryostat were larger, which may be the result of the fact that the freezer, thanks to the stirrer, injects air, limiting the growth of the crystals and breaking them into smaller ones. Kamińska-Dwórznicza et al. (2020) found that in a conventional freezer, due to aeration and mixing during freezing, the heat transfer conditions are different than in a cryostat, and therefore smaller ice crystals are formed [13].

The formation of smaller crystals was observed in the ice cream frozen with ultrasound assistance, and the smallest and more evenly distributed crystals were achieved at a wave frequency of 21.5 kHz and for the stabilized samples (Figures 2 and 3). The average crystal diameter after 3 months of storage was at the level of $8 \mu\text{m}$, and at a frequency of 40 kHz, it was nearly $10 \mu\text{m}$ (Table 4). A similar result was obtained for mango sorbet, where it was noticed that the crystals reached the lowest size of $12.11 \mu\text{m}$ after the application of ultrasound at a frequency of 21.5 kHz [10]. Sonication with a wave frequency of 20 kHz resulted in a reduction in the X_{50} parameter (average diameter for 50% of crystals in the sample) and also in gelatin gel samples from 28 to $47 \mu\text{m}$ [26].

It is possible that waves with the lower of the two given frequencies (21.5 kHz) may be more effective in inhibiting recrystallization in milk ice cream and sorbets (Figure 3); this was already proved for liquid water or a solution containing more air [27]. Ice mixtures are a multiphase system which, due to the content of air bubbles, is also indeed a solution containing more air and in a sense a foam system.

It can be assumed that the ultrasonic waves caused more uniform cooling throughout the entire volume of the product and, as a result, the nucleation of crystals occurred in different parts of the product at a similar time, thanks to which more crystal nuclei were formed and they mutually limited their growth.

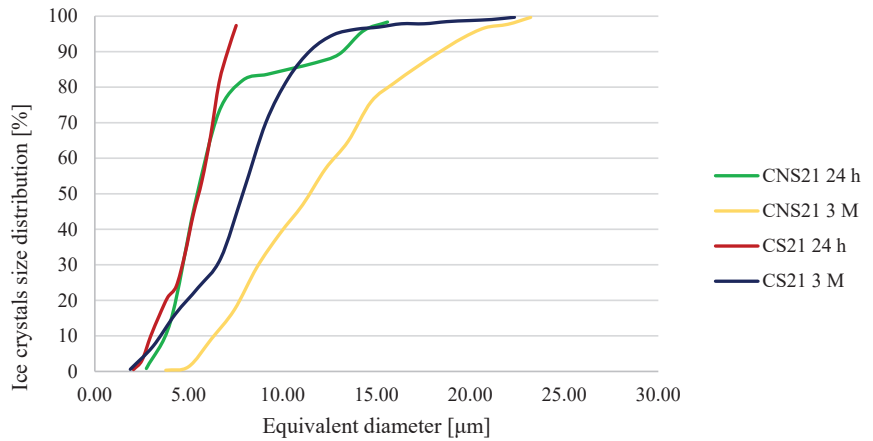


Figure 2. Ice crystal size distribution in stabilized and non-stabilized ice cream frozen using immersion methods, without and with ultrasound support at frequencies of 21.5 kHz.

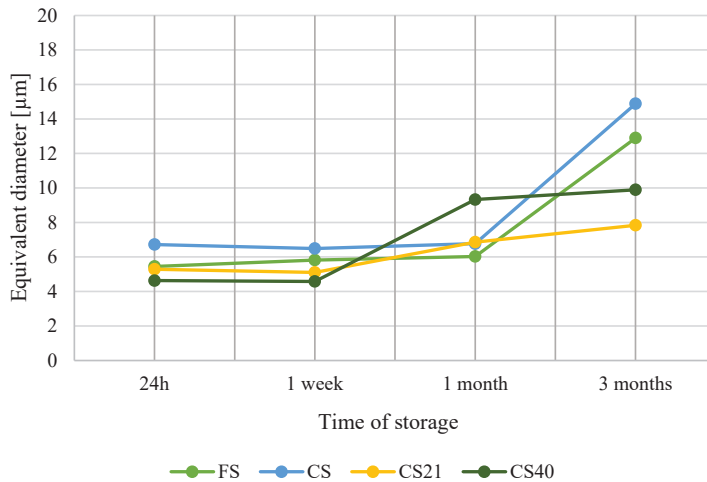


Figure 3. Recrystallization curves for each variant of freezing for stabilized ice cream samples.

Consequently, the application of ultrasonic-assisted freezing at a frequency of 40 kHz, a change in the size of ice crystals occurred in the non-stabilized ice cream sample after 1 month. Then, their growth was inhibited, because the values of the X_{50} parameters after one month and three months were similar (Figure 4). The same tendency was observed in the data regarding the stabilized samples. This might be the result of ultrasonic treatment, which slowed down the recrystallization rate, regardless of the influence of stabilizers. In ice mixtures with the addition of stabilizers, during ripening, a coalescence process may occur for fat globules that increase in size. This destabilization process may be beneficial in the traditional freezing method, but when using ultrasound, it may adversely affect the formation of larger ice crystals [28].

Using waves of this frequency, it was possible to reduce the size of the crystals of frozen grass carp fish (*Ctenopharyngodon idella*) and preserve the original structure of muscle fibers to a greater extent [29]. It was also examined [26] that the differences in structure depend on the intensity of emitting ultrasonic waves. That is why it is so important to select the best parameters, because too little or too much ultrasound action may result in the formation of crystals with larger diameters.

Microscopic analysis gives information not only about ice crystals' sizes, but also about their shapes [3]. After 3 months of storage, for the stabilized and non-stabilized samples from the freezer, the differences in the crystal structure were clearly visible. For the non-stabilized samples, more round crystals with free space between were observed. For the stabilized sample, a more angular and spatial arrangement was noticed. We can assume that a lack of free spaces was observed as a result of the recrystallization mechanisms (Figure 5b).

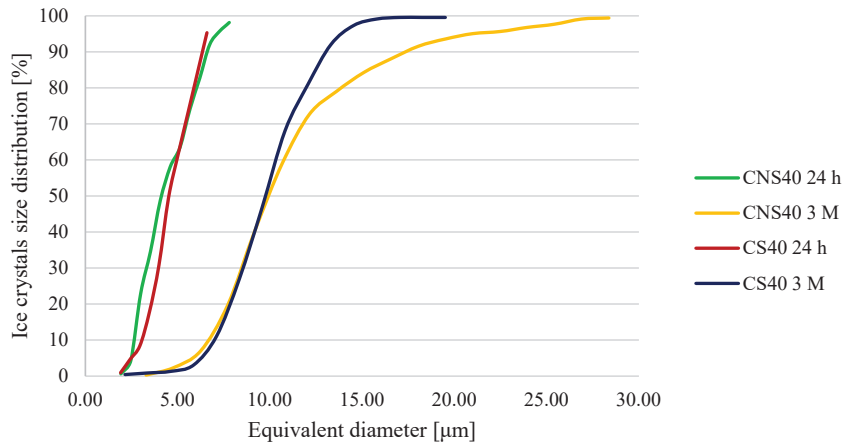


Figure 4. Ice crystal size distribution in stabilized and non-stabilized ice cream frozen using immersion methods, without and with ultrasound support at frequencies of 40 kHz.

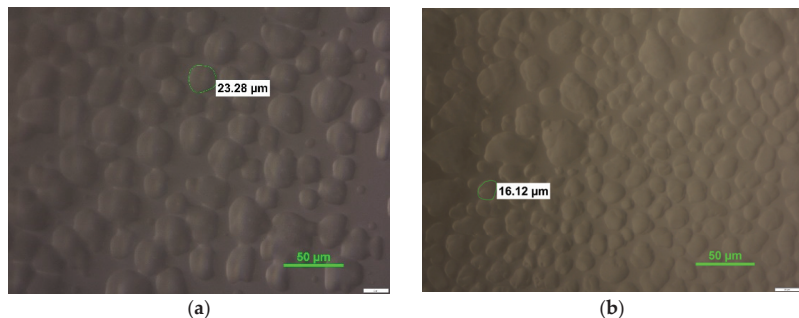


Figure 5. Comparison of the images of ice cream frozen in the freezer, not stabilized (a) and stabilized (b), after 3 months of storage at the temperature of -18°C (magnification $50\times$).

Kiran-Yildirim et al. (2020) examined the influence of various carrageenan fractions (κ -, ι -, λ -) on the recrystallization in model sucrose solutions. They concluded that in the sample with the addition of ι -carrageenan, crystals were more angular and more closely packed just after 50 h of storage [30].

Looking at the crystal morphology of an ice cream sample with and without stabilization, frozen using the immersion method without ultrasound (Figure 6), one can notice a large number of oval crystals, situated close to each other. It is difficult to observe any differences between stabilized and non-stabilized samples. As it was already written [10,13] with a lack of mixing and aerating during freezing in a cryostat gave a more dense structure with larger ice crystals.

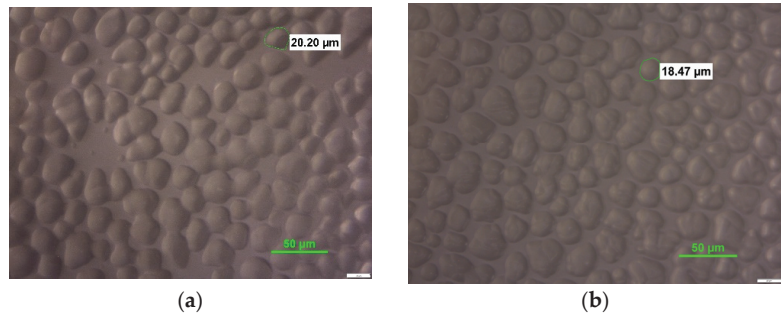


Figure 6. Comparison of the images of ice cream frozen with a cryostat (without the US), not stabilized (a) and stabilized (b), after 3 months of storage at the temperature of $-18\text{ }^{\circ}\text{C}$ (magnification $50\times$).

After 3 months of storage, in samples without stabilizers and after US treatment (21.5 kHz), the spaces in the crystal lattice increased (Figure 7a). Presumably, as a result of coalescence, the crystals grew together, leaving behind free spaces filled with ice masses.

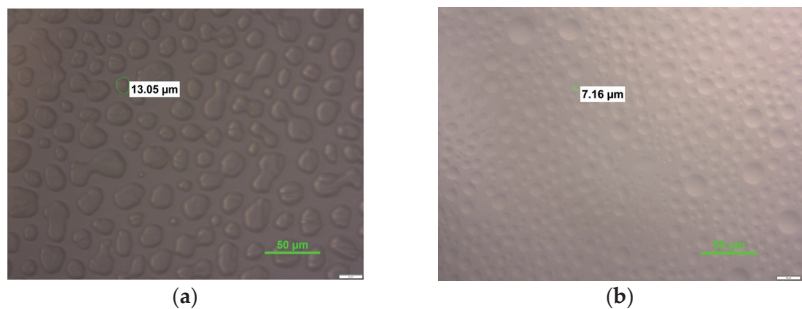


Figure 7. Comparison of the images of ice cream frozen with a cryostat with US at frequencies of 21.5 kHz, not stabilized (a) and stabilized (b), after 3 months of storage at the temperature of $-18\text{ }^{\circ}\text{C}$ (magnification $50\times$).

They were characterized by an irregular shape. Some were fused together, while the single ones were oval, but also elongated. In the samples with stabilizers, the crystals were small and significantly separated from each other (Figure 7b). This may be due to the influence of stabilizers and also cavitation activities, scattering the crystal nuclei, because compared to the photo of the sample using the method without ultrasound (Figure 6), these distances are larger and the crystals are significantly smaller, more evenly distributed and oval. What was surprising is that after the longest storage time, the ice crystals, apart from their size, did not change significantly in shape and retained their distinct form for the most part.

Kamińska-Dwórznicza et al. (2023) also determined the crystal structure of mango sorbet after treatment with ultrasound at a frequency of 21.5 kHz as more uniform, made of smaller and similar crystals in terms of shape and size [10]. Taking into account the amount of fat in milk ice cream, which is much higher than in ice cream based on fruits or a plant-based drink, it can be said that in this case, milk ice cream will have a much better crystalline structure and a smooth texture [31]. Moreover, it was already proved that a lower frequency (around 21 kHz) was better in application to the solutions containing more air and the cavitation effect could strengthen the formation of ice nuclei, thus reducing the rate of ice crystals growth [27].

Additionally, this was confirmed by the research of Zhang et al. (2019). After 180 days of storage after immersion-freezing using ultrasonic waves, ice crystals in pork samples

retained their regular shape and uniform distribution in the material. The authors found that sonication with a power of 180 W can effectively keep ice crystals smaller in meat fibers and cause their smaller deformations during freezing and storage (frequency of sonication was not provided) [32]. Hu et al. (2013) observed a similar nature of crystals in samples of wheat dough frozen with the assistance of ultrasonic waves at a frequency of 25 kHz. This method has been found to promote the formation of fine ice crystals in large numbers inside the frozen cake and improve heat transfer [33].

It could be assumed that ultrasonic waves with a frequency of 40 kHz can enable the production of non-stabilized ice cream with evenly distributed crystals and a similar shape, just after production, as already concluded. Unfortunately, it was not possible to limit the merging of the crystals after about 90 days as it was after treatment with 21.5 kHz waves, and connected oval and larger crystals are visible in the photo (Figure 8).

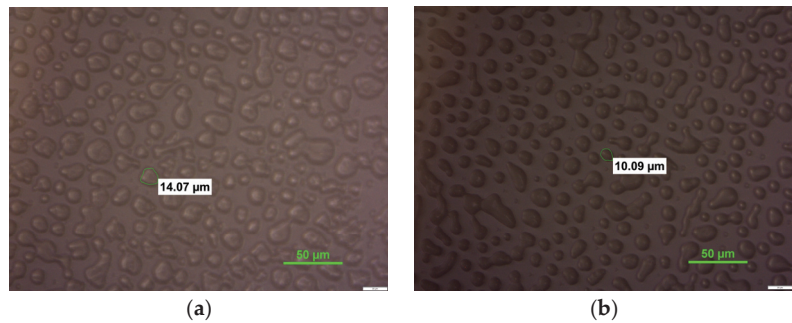


Figure 8. Comparison of the images of ice cream frozen with a cryostat with US at frequencies of 40 kHz, not stabilized (a) and stabilized (b), after 3 months of storage at the temperature of $-18\text{ }^{\circ}\text{C}$ (magnification $50\times$).

However, it should be taken into account that compared to the reference samples, these are small particles and barely perceptible during consumption, because the average diameter of the crystals did not reach $25\text{ }\mu\text{m}$, and this ultimately positive effect on the recrystallization inhibition was found regardless of the presence of the selected hydrocolloids.

4. Conclusions

The present study showed how to use US-assisted freezing in milky ice cream production in order to obtain fine-quality products with more beneficial crystal structures.

The addition of the chosen stabilizers (LBG, xanthan gum and ι -carrageenan) did not significantly influence the density and pH of ice cream; however, the freezing point for the stabilized samples was lower and the overrun was higher.

The US-assisted freezing method with a frequency of 21.5 and 40 kHz allowed for reducing the freezing time by 21 and 23%, respectively, in the case of non-stabilized ice cream compared to the reference sample, while the freezing of stabilized ice cream was shortened by 20 and 21%.

The use of sonication resulted in the limited growth of ice crystals, and the smallest size was achieved at a wave frequency of 21.5 kHz and the addition of stabilizers, which was $7.8\text{ }\mu\text{m}$. Therefore, the proper selection of stabilizers combined with sonication leads to the formation of a more desirable crystal structure. Recrystallization was inhibited more effectively by ultrasonic waves with a frequency of 21.5 kHz than waves with a frequency of 40 kHz. Under the influence of the application of waves with a frequency of 21.5 kHz, the crystal structure consisted of a large number of smaller and more uniform ice crystals.

In conclusion, it has been shown that sonication allows for shortening the freezing time and obtaining a product of better quality, and thanks to its use, the addition of stabilizers can be reduced. This effect could show the manufacturers a new path for ice cream production based only on natural ingredients and with a decreased number of additives. However, it

could also create some problems with process organization and be a big challenge for new freezer constructors.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pr12091957/s1>: Figure S1. Plot of the freezing curves and corresponding linear regression curves with their equations. This applies to the conventional method and stabilized and non-stabilized samples. The designation “a” or “b” corresponds to two different repetitions. Figure S2. Plot of the linear regression curves with corresponding equations based on the temperature and freezing time dependence. This applies to the immersion method and non-stabilized samples. The designation “a” or “b” corresponds to two different repetitions. Figure S3. Plot of the linear regression curves with corresponding equations based on the temperature and freezing time dependence. This applies to the immersion method and stabilized samples. The designation “a” or “b” corresponds to two different repetitions.

Author Contributions: Conceptualization, A.K.-D. and A.K.; methodology, A.K.-D.; software, A.K.; validation, A.K.-D. and A.K.; formal analysis, A.K.; investigation, A.K.-D.; resources, A.K.; data curation, A.K.-D.; writing—original draft preparation, A.K.-D.; writing—review and editing, A.K.; visualization, A.K.-D.; supervision, A.K.D. and A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data generated or analyzed during this study are available from the corresponding author on reasonable request.

Acknowledgments: Research equipment was purchased as part of the “Food and Nutrition Centre—modernisation of the WULS campus to create a Food and Nutrition Research and Development Centre (CZiŻ)”, co-financed by the European Union from the European Regional Development Fund under the Regional Operational Programme of the Mazowieckie Voivodeship for 2014–2020 (project no. RPMA.01.01.00-14-8276/17. Special thanks to Emilia Bogucka for her help with experiment implementation.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kamińska-Dwórznicza, A.; Łaba, S.; Jakubczyk, E. The effects of selected stabilizers addition on physical properties and changes in crystal structure of whey ice cream. *LWT* **2022**, *154*, 112841. [CrossRef]
2. Innocente, N.; Comparin, D.; Corradini, C. Proteose-peptone whey fraction as emulsifier in ice-cream preparation. *Int. Dairy J.* **2002**, *12*, 69–74. [CrossRef]
3. Kamińska-Dwórznicza, A.; Gondek, E.; Laba, S.; Jakubczyk, E.; Samborska, K. Characteristics of Instrumental Methods to Describe and Assess the Recrystallization Process in Ice Cream Systems. *Foods* **2019**, *8*, 117. [CrossRef]
4. Akdeniz, V.; Akalin, A.S. New approach for yoghurt and ice cream production: High-intensity ultrasound. *Trends Food Sci. Technol.* **2019**, *86*, 392–398. [CrossRef]
5. Dadan, M.; Matys, A.; Kamińska-Dwórznicza, A.; Lammerskitten, A.; Toepfl, S.; Parniakov, O. Improvement of freezing processes assisted by ultrasound. In *Design and Optimization of Innovative Food Processing Techniques Assisted by Ultrasound*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 217–273.
6. Zhu, Z.; Zhang, P.; Sun, D.-W. Effects of multi-frequency ultrasound on freezing rates and quality attributes of potatoes. *Ultras. Sonochem.* **2020**, *60*, 104733. [CrossRef]
7. Kot, A.; Kamińska-Dwórznicza, A.; Jakubczyk, E. Study on the Influence of Ultrasound Homogenisation on the Physical Properties of Vegan Ice Cream Mixes. *Appl. Sci.* **2022**, *12*, 8492. [CrossRef]
8. Kot, A.; Jakubczyk, E.; Kamińska-Dwórznicza, A. The Effectiveness of Combination Stabilizers and Ultrasound Homogenization in Milk Ice Cream Production. *Appl. Sci.* **2023**, *13*, 7561. [CrossRef]
9. Nazarewicz, S.; Kozłowicz, K.; Kobus, Z.; Gładyszewska, B.; Matwijczuk, A.; Ślusarczyk, L.; Skrzypek, T.; Sujka, M.; Kozłowicz, N. The Use of Ultrasound in Shaping the Properties of Ice Cream with Oleogel Based on Oil Extracted from Tomato Seeds. *Appl. Sci.* **2022**, *12*, 9165. [CrossRef]
10. Kamińska-Dwórznicza, A.; Kot, A.; Jakubczyk, E.; Buniowska-Olejnik, M.; Nowacka, M. Effect of ultrasound-assisted freezing on the crystal structure of mango sorbet. *Crystals* **2023**, *13*, 396. [CrossRef]
11. Dłużewska, E.; Gazda, B.; Leszczyński, K. Wpływ wybranych hydrokoloidów polisacharydowych na jakość koncentratów lodów owocowych. *Acta Sci. Pol. Technol. Aliment.* **2003**, *2*, 97–107.
12. Góral, M.; Kozłowicz, K.; Pankiewicz, U.; Góral, D.; Kluzka, F.; Wójtowicz, A. Impact of stabilizers on the freezing process, and physicochemical and organoleptic properties of coconut milk-based ice cream. *LWT* **2018**, *92*, 516–522. [CrossRef]

13. Kamińska-Dwórnicka, A.; Janczewska-Dupczyk, A.; Kot, A.; Laba, S.; Samborska, K. The impact of iota- and kappa-carrageenan addition on freezing process and ice crystals structure of strawberry sorbet frozen by various methods. *J. Food Sci.* **2020**, *85*, 50–56. [CrossRef] [PubMed]
14. Goff, H.D.; Hartel, R.W. Ice cream structure. In *Ice Cream*; Springer: Boston, MA, USA, 2013; pp. 313–352.
15. Romulo, A.; Meindrawan, B. Marpietylie Effect of Dairy and Non-Dairy Ingredients on the Physical Characteristic of Ice Cream: Review. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *794*, 012145. [CrossRef]
16. Kamińska-Dwórnicka, A. Wpływ stabilizatorów na ograniczenie rekrytalizacji w lodach typu sorbet. *Przemysł Spożywczy* **2016**, *1*, 34–37. [CrossRef]
17. Ndoye, F.T.; Alvarez, G. Characterization of ice recrystallization in ice cream during storage using the focused beam reflectance measurement. *J. Food Eng.* **2015**, *148*, 24–34. [CrossRef]
18. Syed, Q.A.; Anwar, S.; Shukat, R.; Zahoor, T. Effects of different ingredients on texture of ice cream. *J. Nutr. Health Food Eng.* **2018**, *8*, 422–435. [CrossRef]
19. Brown, J.R.; Seymour, J.D.; Brox, T.I.; Skidmore, M.L.; Wang, C.; Christner, B.C.; Luo, B.H.; Codd, S.L. Recrystallization inhibition in ice due to ice binding protein activity detected by nuclear magnetic resonance. *Biotechnol. Rep.* **2014**, *3*, 60–64. [CrossRef]
20. Raymond, J.A.; Knight, C.A. Ice binding, recrystallization inhibition, and cryoprotective properties of ice-active substances associated with Antarctic sea ice diatoms. *Cryobiology* **2003**, *46*, 174–181. [CrossRef]
21. Mortazavi, A.; Tabatabaie, F. Study of ice cream freezing process after treatment with ultrasound. *World Appl. Sci. J.* **2008**, *4*, 188–190.
22. Jambak, A.; Lerda, D.; Mirčeta, R.; Šimunek, M.; Lelas, V.; Chemat, F.; Herceg, Z.; Batur, V. Experimental design and optimization of ultrasound treatment: Functional and physical properties of sonicated ice cream model mixtures. *J. Food Process. Technol.* **2012**, *3*, 145. [CrossRef]
23. Adhikari, B.M.; Truong, T.; Prakash, S.; Bansal, N.; Bhandari, B. Impact of incorporation of CO₂ on the melting, texture and sensory attributes of soft-serve ice cream. *Int. Dairy J.* **2020**, *109*, 104789. [CrossRef]
24. Astráin-Redín, L.; Abad, J.; Rieder, A.; Kirkhus, B.; Raso, J.; Cebrián, G.; Álvarez, I. Direct contact ultrasound assisted freezing of chicken breast samples. *Ultrason. Sonochem.* **2021**, *70*, 105319. [CrossRef] [PubMed]
25. Lomolino, G.; Zannoni, S.; Zabara, A.; Da Lio, M.; De Iseppi, A. Ice recrystallisation and melting in ice cream with different proteins levels and subjected to thermal fluctuation. *Int. Dairy J.* **2020**, *100*, 104557. [CrossRef]
26. Xu, B.-g.; Zhang, M.; Bhandari, B.; Sun, J.; Gao, Z. Infusion of CO₂ in a solid food: A novel method to enhance the low-frequency ultrasound effect on immersion freezing process. *Innov. Food Sci. Emerg. Technol.* **2016**, *35*, 194–203. [CrossRef]
27. Ohsaka, K.; Trinh, E. Apparatus for measuring the growth velocity of dendritic ice in undercooled water. *J. Cryst. Growth* **1998**, *194*, 138–142. [CrossRef]
28. Cheng, J.; Ma, Y.; Li, X.; Yan, T.; Cui, J. Effects of milk protein-polysaccharide interactions on the stability of ice cream mix model systems. *Food Hydrocoll.* **2015**, *45*, 327–336. [CrossRef]
29. Shi, Z.; Zhong, S.; Yan, W.; Liu, M.; Yang, Z.; Qiao, X. The effects of ultrasonic treatment on the freezing rate, physicochemical quality, and microstructure of the back muscle of grass carp (*Ctenopharyngodon idella*). *LWT* **2019**, *111*, 301–308. [CrossRef]
30. Kiran-Yildirim, B.; Gaukel, V. Ice Crystal Growth in Sucrose Solutions Containing Kappa- and Iota-Carrageenans. *Chem. Eng. Technol.* **2020**, *43*, 1040–1047. [CrossRef]
31. Aslan Türker, D.; Dogan, M. Effects of ultrasound homogenization on the structural and sensorial attributes of ice cream: Optimization with Taguchi and data envelopment analysis. *J. Food Meas. Charact.* **2021**, *15*, 4888–4898. [CrossRef]
32. Zhang, M.; Xia, X.; Liu, Q.; Chen, Q.; Kong, B. Changes in microstructure, quality and water distribution of porcine longissimus muscles subjected to ultrasound-assisted immersion freezing during frozen storage. *Meat Sci.* **2019**, *151*, 24–32. [CrossRef]
33. Hu, S.-Q.; Liu, G.; Li, L.; Li, Z.-X.; Hou, Y. An improvement in the immersion freezing process for frozen dough via ultrasound irradiation. *J. Food Eng.* **2013**, *114*, 22–28. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Review

Use of Triboelectric Nanogenerators in Advanced Hybrid Renewable Energy Systems for High Efficiency in Sustainable Energy Production: A Review

Van-Long Trinh ¹ and Chen-Kuei Chung ^{2,*}

¹ School of Mechanical and Automotive Engineering, Hanoi University of Industry, 298 Caudien Street, Hanoi 100000, Vietnam

² Department of Mechanical Engineering, National Cheng Kung University, Tainan 701, Taiwan

* Correspondence: ckchung@mail.ncku.edu.tw

Abstract: Renewable energy is the best choice for clean and sustainable energy development. A single renewable energy system reveals an intermittent disadvantage during the energy production process due to the effects of weather, season, day/night, and working environment. A generally hybrid renewable energy system (HRES) is an energy production scheme that is built based on a combination of two or more single renewable energy sources (such as solar energy, wind power, hydropower, thermal energy, and ocean energy) to produce electrical energy for energy consumption, energy storage, or a power transmission line. HRESs feature the outstanding characteristics of enhancing energy conversion efficiency and reducing fluctuations during the energy production process. Triboelectric nanogenerator (TENG) technology transduces wasted mechanical energies into electrical energy. The TENG can harvest renewable energy sources (such as wind, water flow, and ocean energy) into electricity with a sustainable working ability that can be integrated into an HRES for high power efficiency in sustainable renewable energy production. This article reviews the recent techniques and methods using HRESs and triboelectric nanogenerators (TENGs) in advanced hybrid renewable energy systems for improvements in the efficiency of harvesting energy, sustainable energy production, and practical applications. The paper mentions the benefits, challenges, and specific solutions related to the development and utilization of HRESs. The results show that the TENG is a highly potential power source for harvesting energy, renewable energy integration, application, and sustainable energy development. The results are a useful reference source for developing HRES models for practical applications and robust development in the near future.

Keywords: triboelectric nanogenerator; renewable energy; sustainable energy development; hybrid energy system; energy applications

Citation: Trinh, V.-L.; Chung, C.-K. Use of Triboelectric Nanogenerators in Advanced Hybrid Renewable Energy Systems for High Efficiency in Sustainable Energy Production: A Review. *Processes* **2024**, *12*, 1964. <https://doi.org/10.3390/pr12091964>

Academic Editor: Massimo Caruso

Received: 3 August 2024

Revised: 30 August 2024

Accepted: 4 September 2024

Published: 12 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Energy is an indispensable element of our social development, from cooking to industrial consumption. Traditional energy is increasingly exhausted because fossil energy resources are limited, and their exploitation causes harmful effects to our environment. Renewable energy (RE) has a crucial mission in facing the problems of a lack of energy and protecting the natural environment all over the world [1–7]. Renewable energy contributes to technology, solutions, and national policies leading toward the sustainability of energy utilization, economic–social development, environmental protection, and industrialization strategies [8–20]. Traditional renewable energy resources include solar energy, hydropower, ocean energy, wind power, bioenergy, geothermal energy, and hydrogen energy, as shown in Figure 1. Renewable energy (RE) is the best solution for exploiting, developing, and consuming energy to repel carbon dioxide (CO₂) emissions to the index of zero carbon for sustainable development [21–25]. The global carbon dioxide emission shows a worrying

number with the high data of 37×10^9 tonnes (t) recorded in 2022. With the rapid development of renewable energy, the Net-Zero Roadmap (NZR) was planned by the International Energy Agency (IEA) with the strategy of maintaining the global warming temperature of 1.5°C until 2030 and attaining net-zero carbon dioxide emissions by 2050. With the net-zero emissions (NZE) scenario, emissions will be reduced by about 80% by 2030 with the speeding up of renewable energy development, energy efficiency, methane emission reductions, and energy conversion technologies. Carbon dioxide emissions will decrease by about 35% by 2030 in comparison to the value in 2022 in the NZE scenario, with the appearance of clean energy and renewable energy. The share of the global implementation of clean energy technologies from China and other advanced economies was recorded during the period of 2010 to 2022, with a contribution of over 95% from global electric vehicles and heat pumps, as well as about 85% from the contribution of wind and solar energy capacity from 2015 to 2022 [26]. Table 1 shows the rapid increase in renewable generation (RG) around the world, with 4209 terawatt hours (TWh) in 2010, 7964 TWh in 2021, and 8599 TWh in 2022, as well as estimations of 19295 TWh in 2030 and about 55057 TWh in 2050 with the NZE scenario, respectively. The table shows increases in renewable generation of about 89.21%, 104.3%, 358.42%, and 1208.08% in comparison to the value of renewable energy generation in 2010, respectively. The table also shows renewable generation in critical areas of North America, Central and South America, Europe, Africa, Eurasia, and the Asia Pacific with outstanding aggregations. North America shows impressive numbers with 856 TWh in 2010, 1374 TWh in 2021, and 1497 TWh in 2022, as well as estimations of 3538 TWh in 2030 and about 9261 TWh in 2050 with the NZE scenario, respectively. The increases in renewable generation in this area are statistical values of about 60.51%, 74.88%, 313.32%, and 981.89% in comparison with the value of renewable energy generation in 2010, respectively. Central and South America shows recorded values of about 752 TWh in 2010, 896 TWh in 2021, and 1018 TWh in 2022, as well as estimations of 1428 TWh in 2030 and about 3768 TWh in 2050 with the NZE scenario, respectively. In the area, renewable generation increases are counted at about 19.15%, 35.37%, and 89.89% and then rapidly increased to 401.06% in comparison with the value of renewable energy generation in 2010, respectively. Europe has recorded positive values with 954 TWh in 2010, 1601 TWh in 2021, and 1620 TWh in 2022, as well as estimations of 3438 TWh in 2030 and about 6834 TWh in 2050 with the NZE scenario, respectively. The RG increases in Europe were evaluated with numbers of about 67.82%, 69.81%, and 260.38%, and a remarkable number of 616.35% in comparison with the value of renewable energy generation in 2010, respectively. Africa has the effective development of renewable generation with 116 TWh in 2010, 201 TWh in 2021, and 210 TWh in 2022, as well as estimations of 711 TWh in 2030 and about 3453 TWh in 2050 with the NZE scenario, respectively. The area receives growth values of about 73.28%, 81.03%, and 512.93%, and an intense increase of 2876.72% in comparison with the value of renewable energy generation in 2010, respectively. Eurasia shows good RG values with 226 TWh in 2010, 287 TWh in 2021, and 277 TWh in 2022, as well as estimations of about 380 TWh in 2030 and 844 TWh in 2050 with the NZE scenario, respectively. The area achieved growth numbers of about 26.99%, 22.56%, 68.14%, and 273.45% in comparison with the value of renewable energy generation in 2010, respectively. The Asia Pacific contributes to the outstanding RG numbers with 1287 TWh in 2010, 3568 TWh in 2021, and 3932 TWh in 2022, as well as estimations of 9568 TWh in 2030 and about 28321 TWh in 2050 with the NZE scenario, respectively. This area records excellent increases of about 177.23%, 205.52%, 643.43%, and 2100.54% in comparison with the value of renewable energy generation in 2010, respectively. The table also shows remarkable results of renewable energy generation from different countries and combines those which have positive strategies in the development of renewable energy generation such as China, the European Union, the United States, and Brazil [27].

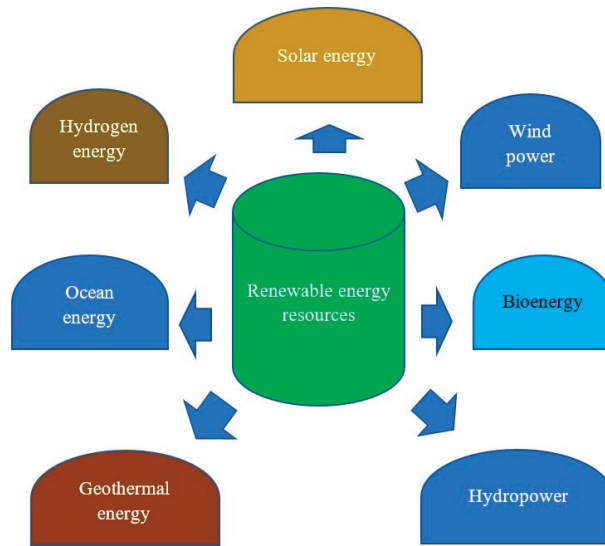


Figure 1. Traditional renewable energy resources include solar energy, wind power, bioenergy, hydropower, geothermal energy, ocean energy, and hydrogen energy.

Table 1. Renewable energy generation (TWh) in the world and some critical areas.

		2010	2021	2022	2030 with the NZEs Scenario	2050 with the NZEs Scenario
In the world	RG (TWh)	4209	7964	8599	19,295	55,057
	Comparison with 2010	-	89.21%	104.3%	358.42%	1208.08%
North America	RG (TWh)	856	1374	1497	3538	9261
	Comparison with 2010	-	60.51%	74.88%	313.32%	981.89%
Central and South America	RG (TWh)	752	896	1018	1428	3768
	Comparison with 2010	-	19.15%	35.37%	89.89%	401.06%
Europe	RG (TWh)	954	1601	1620	3438	6834
	Comparison with 2010	-	67.82%	69.81%	260.38%	616.35%
Africa	RG (TWh)	116	201	210	711	3453
	Comparison with 2010	-	73.28%	81.03%	512.93%	2876.72%
Eurasia	RG (TWh)	226	287	277	380	844
	Comparison with 2010	-	26.99%	22.56%	68.14%	273.45%
Asia Pacific	RG (TWh)	1287	3568	3932	9568	28,321
	Comparison with 2010	-	177.23%	205.52%	643.43%	2100.54%
China	RG (TWh)	782	2448	2681	6419	14836
	Comparison with 2010	-	213.04%	242.84%	720.84%	1797.19%
European Union	RG (TWh)	653	1081	1085	2407	4720
	Comparison with 2010	-	65.54%	66.16%	268.61%	622.82%
United States	RG (TWh)	441	867	973	2087	7683
	Comparison with 2010	-	96.60%	120.63%	373.24%	1642.18%
Brazil	RG (TWh)	437	508	594	732	1378
	Comparison with 2010	-	16.25%	35.93%	67.51%	215.33%

Renewable energy utilization contributes to the protection of our living environment and fossil energy resources [28]. The Renewables 2024 Global Status Report (GSR 2024) was conducted by the Renewable Energy Policy Network for the 21st Century (REN21) with the impressive number of renewable power capacity of 473 gigawatts (GW) corresponding to the increase of 36% in power capacity added in 2023 and the contribution of 151 countries to the development strategy of net-zero targets [29]. The global energy demand increased by about 1.3% in 2022 according to the International Energy Agency (IEA). The Stated

Policies Scenario (STEPS) also gives an overview of the latest policies regarding areas such as energy, industry, and climate for countries over the world. The STEPS suggests that the contribution of renewable energy to power capacity should be increased of about eighty per cent by 2030 in the effort of reducing carbon dioxide emissions [27].

The renewable energy (RE) has been received many attentions for excavation, development, and application such as planning environmental protection strategies using renewable energy [30], constructing pathways of deep decarbonization with the benefits of environment protection and economic control [31], developing energy legislation for countries to promote the replacement of traditional energy with renewable energy to help solve the problem of global warming and the lack of energy [32], investigating the evidence of a positive influence on economic growth with the use of renewable energy [33], and using renewable energy sources to achieve net-zero energy for buildings [34].

A triboelectric nanogenerator (TENG) is an emerging technology for transforming wasted mechanical energy into electricity for many practical applications [35]. The TENG shows many advantages of modern renewable energy technology with outstanding characteristics of green energy, lightweight, simple structure, and sustainable energy source [36]. TENG technology can convert wasted mechanical energies by, for example, transducing water wave energy, motions, oscillations, and vibrations [37–43], kinetic energy from the ocean [44–47], and biomechanical energy from bio-object movement [48], into electrical energy. TENG technology is a strong candidate for renewable and sustainable energy generation as it converts renewable energy sources into electricity such as wind energy, water wave energy, and ocean energy. The TENG has outstanding features, including that it produces no gas emissions, is eco-friendly, and is a form of clean energy [49–51]. The TENG meets the following criteria of sustainable development goal 7 (SDG-7): providing everyone access to energy services using modern, affordable, and reliable energy resources by 2030 (Target 7.1), increasing the share of global renewable energy (Target 7.2), doubling the energy efficiency improvement (Target 7.3), promoting access to clean energy via research actions, technology applications, and capital investment (Target 7a), and expanding sustainable energy access for developing countries via services, technology, and infrastructure development (Target 7b) [52].

A single renewable energy system (SRES) is generally represented as a renewable energy conversion technology that utilizes a single renewable energy source such as solar energy, hydropower, ocean energy, wind power, bioenergy, geothermal energy, or hydrogen energy. Intermittent energy generation is the biggest disadvantage of SRESs which rely on natural environment conditions such as weather, season, day/night, and working environment [53]. This directly influences the energy conversion performance of renewable energy systems such as their efficiency, power, and continuity features. A hybrid renewable energy system (HRES) is an energy production model that is built by a combination of two or more single renewable energy systems. HRESs convert renewable energy sources into electrical energy for electrical consumption devices, energy storage systems, and power transmission lines. HRESs are the best solution to the intermittent energy supply problem; because they consist of different single renewable energy systems, they can harvest many types of renewable energy sources at difference times [54]. HRESs show outstanding characteristics of high power efficiency and sustainable renewable energy production [55]. HRESs are one of the best methods to enhance output performance and reduce fluctuation during the energy generation process. HRESs are an effective supporting method for energy generation in remote communities with wonderful characteristics such as low energy cost and environmental protection. HRESs have been studied by many research groups who have applied techniques and methods such as artificial intelligence, hybrid algorithms, and computer tools to develop HRES models for practical applications [56–59].

Hybrid renewable energy systems can be integrated into a hybrid energy storage system to improve the benefits of a power system by, for example, decreasing the capital cost, enhancing the efficiency, extending the system life, or balancing the output power [60]. Figure 2 shows the structure of a hybrid energy system integrated with energy storage

modules to improve the performance of a power system [61]. The model consists of basic components including nonrenewable energy resources, renewable energy resources, converters, energy storage modules, an energy management hub, alternative current (AC)–direct current (DC), AC–AC, DC–DC, and DC–AC converters, a bidirectional converter, AC–DC buses, AC–DC loads, and an electrical grid. The energy resources include nonrenewable and renewable energy types that supply the input power to the system. The energy storage module has the role of storing electrical energy for long-term use and balancing the input signal of electricity with the electrical equipment. The AC–DC, AC–AC, DC–DC, and DC–AC converters are used to change an alternative current (AC) to a direct current (DC), an AC to an AC, a DC to an AC, and a DC to an AC, respectively. The bidirectional converter is used to make an interface between a low-voltage storage unit and a high-voltage bus. The AC and DC buses are used to couple the power sources integrated into the hybrid energy system. The electrical load uses the AC or DC power to drive the actions of the electrical equipment. The power grid is used to transmit electricity to areas that lack electrical energy.

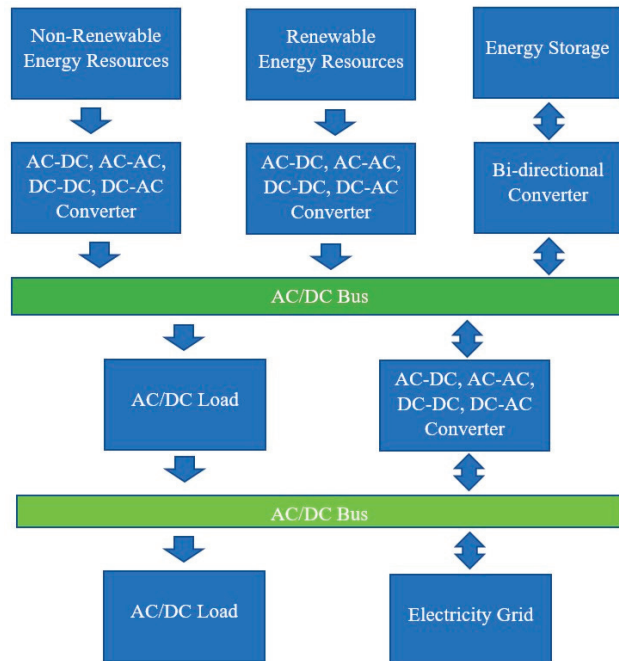


Figure 2. A hybrid energy model with some basic parts including nonrenewable energy sources, renewable energy sources, energy storage, electrical converters, electrical loads, and an electricity grid.

HRESs have contributed to a reduction in net present costs and carbon dioxide emissions [62,63]. HRESs have led to the development of technologies and methods to expand the practical applications of renewable energy; these include the development of algorithms to optimize the autonomous service of powering residential buildings [64], the introduction of thermochemical conversion technology to turn waste into electricity [65], the use of algorithm-based fuzzy logic tools to select a suitable location for HRES facilities [66], and the creation of charging stations for electric vehicles [67]. TENG technology is emerging as an effective solution which can be integrated into HRESs to boost the efficiency, power output, and integrity of the power system [68,69]. This paper reviews the recent techniques and methods of developing hybrid renewable energy systems to improve the energy conversion efficiency, output performance, and continuity of renewable energy conversion systems. The integration of a triboelectric nanogenerator (TENG) in an advanced hybrid renewable

energy system (HRES) is proposed for harvesting renewable energy with the ability to improve efficiency and produce sustainable energy, among other practical applications. This paper addresses the benefits, challenges, and solutions for strengthening the efficiency, utilization, and development of HRESs. The results reveal that triboelectric nanogenerators have great potential as a harvesting technology for energy generation, renewable energy integration, application, and sustainable energy development, eliminating the disadvantages of intermittency and variability that single renewable energy systems have stumbled upon. The results hope to support the development of HRES models for practical applications and robust energy development in the near future.

2. Single Renewable Energy Systems

A single renewable energy system (SRES) is an energy conversion system that is used to convert a renewable energy resource into useful energy, such as photovoltaic systems converting solar energy into electricity, wind energy systems changing wind power into electricity, triboelectric nanogenerator systems turning wasted mechanical energy into electrical energy, and hydropower systems transforming hydropower into electricity. Single renewable energy systems are developed using mathematical models to address the output performance of the system.

2.1. Photovoltaic System

A photovoltaic system (PVS) is a solar energy system that converts solar energy into electricity. The total solar radiation is the input energy for PVSs. The total solar radiation can be estimated using Equation (1) [70]:

$$I_T = I_b R_b + I_d R_d + (I_b + I_d) R_r \quad (1)$$

where

I_T is the total solar radiation (kW h/m^2);

I_b is the direct normal solar radiation;

I_d is the diffuse solar radiation;

R_d is the tilt factor of the diffuse of the solar radiation;

R_r is the tilt factor of the reflection of the solar radiation.

The power output of PVSs can be calculated using Equation (2):

$$P_{si} = I_T \eta A_{PVS} \quad (2)$$

where

P_{si} is the power output of the PVS;

A_{PVS} is the area of the PVS (m^2);

η is the system efficiency.

η can be calculated using Equation (3):

$$\eta = \eta_m \eta_{pc} P_f \quad (3)$$

where

η_{pc} is the power condition efficiency;

P_f is a packing factor;

η_m is module efficiency.

This value can be calculated by Equation (4):

$$\eta_m = \eta_r (1 - \beta (T_c - T_r)) \quad (4)$$

where

η_r is the module reference efficiency;

β is the temperature coefficient;

T_c is the monthly average temperature;

T_r is the reference temperature.

The monthly average temperature can be calculated using Equation (5):

$$T_c = T_a + \frac{\alpha\tau}{U_L} I_T \quad (5)$$

where

T_a is the instantaneous ambient temperature.

$$\frac{U_L}{\alpha\tau} = \frac{I_{T,NT}}{(NT - T_{a,NT})} \quad (6)$$

NT is the normal operation cell temperature. In normal conditions, $T_{a,NT}$ is 20 degrees centigrade, $I_{T,NT}$ is 800 watts (W), and the wind speed is 1 m per second (m/s).

2.2. The Wind Energy System (WES)

Wind energy systems produce electricity by using a wind turbine generator to change wind energy into electric energy with a power output P calculated by Equation (7) [71,72]:

$$P_w = \frac{1}{2} \rho A_w V^3 \quad (7)$$

where

P_w is the power output of the WES;

A_w is the swept area of the wind turbine;

V is the velocity of the air.

2.3. Triboelectric Nanogenerator (TENG)

The triboelectric nanogenerator (TENG) has recently emerged as a new renewable and clean energy source. The TENG can convert wasted mechanical energy into electrical energy. The TENG's working mechanism is based on the triboelectrification effect of coupled tribomaterials during contact-separate cycles [73]. TENGs have more outstanding characteristics, including being lightweight and low cost, using easy-to-find materials, having a simple structure [74], and generating sustainable power [75,76]. Due to these outstanding characteristics, TENGs have received much attention and have been used for numerous practical applications such as transforming blue energy into electricity [77–80], developing biomedical sensors, healthcare devices, therapeutic applications, and implantable biomedical applications [81–85], constructing micro-electro-mechanical systems (MEMSs) equipment [86], self-powered sensing in temperature sensors, healthcare sensors, pressure sensors, humidity sensors, force sensors, accelerating sensors, self-powered active sensors for hydrogen detection, multifunctional sensors, human–machine interface sensing, and self-powered sensing devices [87–95], harvesting biomechanical energy to sustainably power wearable bioelectronics systems and self-powered wearable electronics [96–98], monitoring the marine pipeline, the ocean wave, and the environmental potential of hydrogen [99,100], introducing cellulose materials to drive self-powered sensors [101], converting mechanical energy for self-powered electronics, smart devices, Internet of Things (IoT), and lighting LEDs [102–104], constructing electrocatalytic systems with a self-powered source [105], developing multifunctional self-powered electronics, sensors, portable electronics devices, functional devices, and biomedical devices [106–109], developing sustainable power [110,111], creating human–machine interface applications [112], combining with wood to create self-powered sensors [113], and developing a self-powered device for lighting purposes [114]. Many other methods have been researched to develop advanced triboelectric nanogenerators (ATENGs) such as introducing carbon material with low-dimensional features to improve output performance [115], developing chemical modification methods to enhance output performance [116], and constructing ATENGs using technologies from electrochemical systems [117].

Figure 3 shows the working mechanism of the TENG with five basic stages (initial, contacting, separating, released, and contacting again states) to produce electric energy. Figure 3a shows noncharges in the initial state. Figure 3b shows the adverse charges that were generated as the two triboelectric surfaces made contact with each other. Figure 3c shows the potential unbalance as the two tribomaterials are released from each other. An electric current moves via the external load. Figure 3d shows the neutral state as the two tribomaterials are released from each other. Figure 3e shows the potential unbalance as the two tribomaterials are pressed together again. An electrical current flows through the external load. The TENG produces an electric current with the open-circuit voltage (V_{OC}) expressed by Equation (8) [118]:

$$V_{OC} = -\frac{\sigma d}{\epsilon_0} \tag{8}$$

where

- V_{OC} is the open-circuit voltage;
- σ is the triboelectric charge density;
- d is the distance between the two contact surfaces;
- ϵ_0 is the vacuum permittivity.

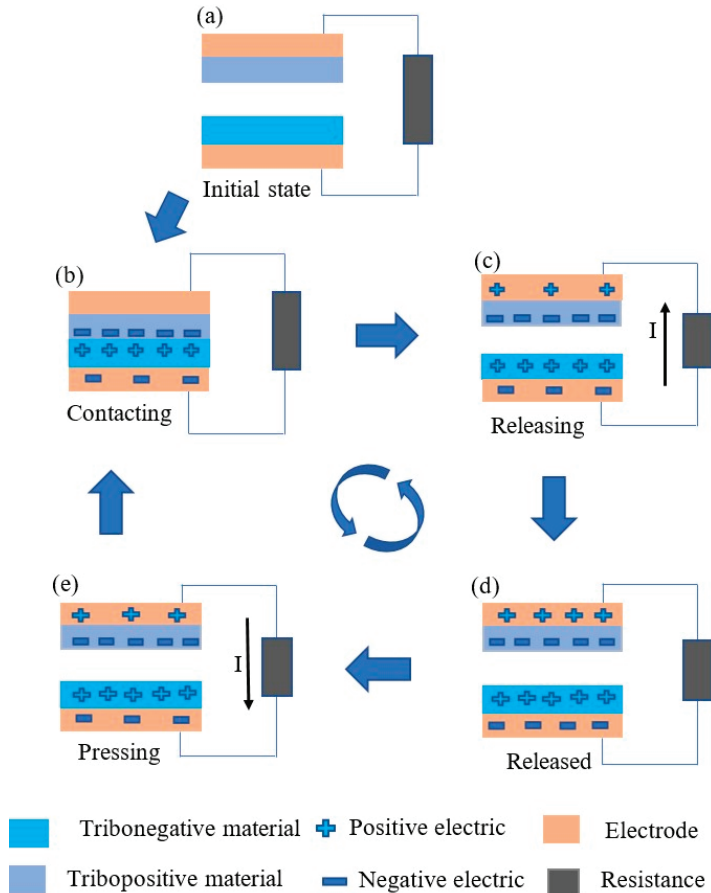


Figure 3. The working mechanism of the TENGs. (a) The initial state. (b) The contacting state. (c) The releasing state, in which a current (I) flows via an external load. (d) The released state. (e) The pressing state, in which a current (I) runs through an external load.

The TENG shows good serving ability in hybrid energy systems, improving the efficiency of the hybrid energy system with a TENG and a piezoelectric nanogenerator [119], providing a self-powering feature for a smart system based on hybrid energy using an ATENG and an electromagnetic generator [120] and a self-powering source for an Internet of Things (IoT) system using a hybrid energy system comprising a TENG and a pyroelectric nanogenerator [121], and integrating a TENG, solar cells, and an electromagnetic generator into an HRES to harvest the ocean's energy for practical applications [122].

3. Hybrid Renewable Energy System (HRES)

A hybrid renewable energy system (HRES) is constructed from two or more types of single renewable energy systems. Advanced hybrid renewable energy systems support the production of energy, especially in remote areas, with many advantages such as their use of cheap and sustainable energy sources. HRESs can penetrate into the power grid or act alone to supply energy to a specific area. New methods and technologies have been focused on using HRESs to convert renewable energy sources into electricity such as investigating an HRES model to transduce hydropower and thermal energy into electricity in Ukraine [123] and developing an HRES based on triboelectric nanogenerators, thermoelectric nanogenerators, piezoelectric nanogenerators, electromagnetic nanogenerators, and solar cells to improve the efficiency of energy conversion [124]. Figure 4 shows the proposal of a hybrid renewable energy system with the contribution of a photovoltaic energy system, a wind energy system, and a triboelectric nanogenerator to produce electric energy for electrical lines, electric consumption equipment, and energy storage systems [125–127]. Figure 4a shows a solar photovoltaic energy system with photovoltaic energy conversion technology that is used to change solar energy into electricity. Figure 4b shows a wind power system that uses wind power technology to convert wind energy to electrical energy for energy consumption, storage, and integration. Figure 4c shows a triboelectric nanogenerator system that uses triboelectric generator technology to convert mechanical energy into electricity for the purposes of electricity integration, energy storage, and providing power to electric consumption equipment. Figure 4d shows an electrical line that has the duty of transmitting electric energy to distant areas. Figure 4e shows how energy from HRESs is used in electrical consumption devices. Figure 4f shows the energy storage system that is used to store energy for long-time use, transport energy to remote areas, and power portable devices. The model consists of three different kinds of energy transduction systems to improve renewable energy conversion efficiency. Specifically, solar photovoltaic energy systems produce electricity via the photovoltaic effect principle. Photovoltaic cells are generally semiconductor devices that convert solar energy into electricity [128,129]. Figure 5 shows an equivalent circuit representing a solar photovoltaic cell. I_{ph} is the current source of the photocurrent of the PV cell. The current source is linear dependency relation on the solar irradiance. I_d represents the reverse saturating current of the diode. R_{sh} is the PV intrinsic shunt resistance. R_s is the PV series resistance related to energy losses via the assembly technique of solder bonds, junctions, and wires. Single PV cells can be connected to form an array of PVs for more generation efficiency. The output current of the PV array can be expressed by Equation (9):

$$I = N_p I_{pn} - N_p I_d \left[\exp\left(\frac{qV}{kTAN_s} - 1\right) \right] \quad (9)$$

where I is the output current of the PV array. N_p represents the module numbers connected in parallel. I_{pn} is the photocurrent. I_d represents the reverse saturation current of the PV cell. q represents the electron charge. V represents the output voltage of the PV array. k , T , A , and N_s are the Boltzmann constant, the temperature of the PV cell, the deviation of the pn junction feature of the PV cell, and the number of PV cells connected in series, respectively. The wind power system produces electricity from wind energy via wind turbines, as mentioned in Section 2.2. The triboelectric nanogenerator produces electricity from mechanical energy via triboelectric effect, as mentioned in Section 2.3. By

using different energy sources of solar, wind, and mechanical energy, HRESs can produce electricity at all times for practical applications. This ensures a high efficiency of energy transduction in HRESs. In the case of a lack of input triggers of solar, wind, or mechanical energy, the energy storage unit will become an electrical generator to supply electricity to the consumption equipment.

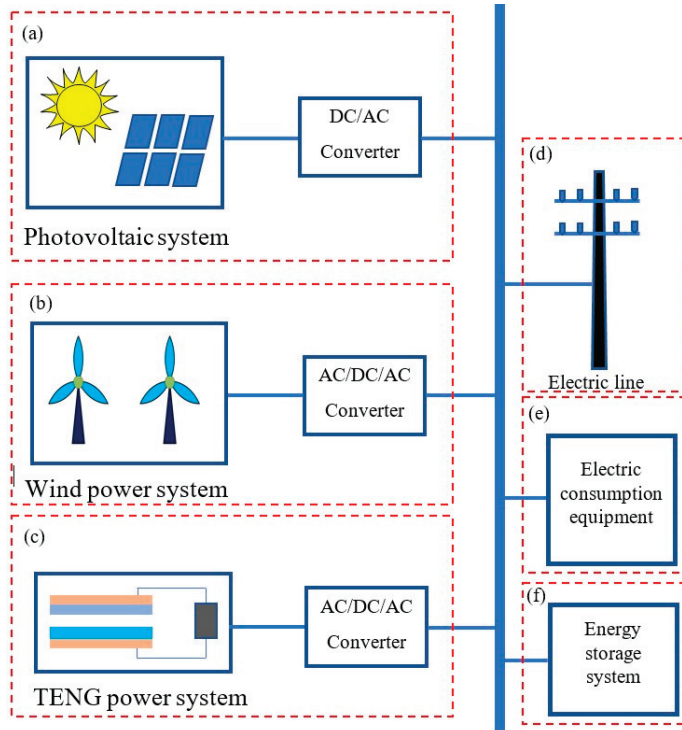


Figure 4. A proposal for a hybrid renewable energy system; (a) a photovoltaic energy system; (b) a wind power system; (c) a TENG power system; (d) an electric line; (e) electrical consumption equipment; and (f) an energy storage system.

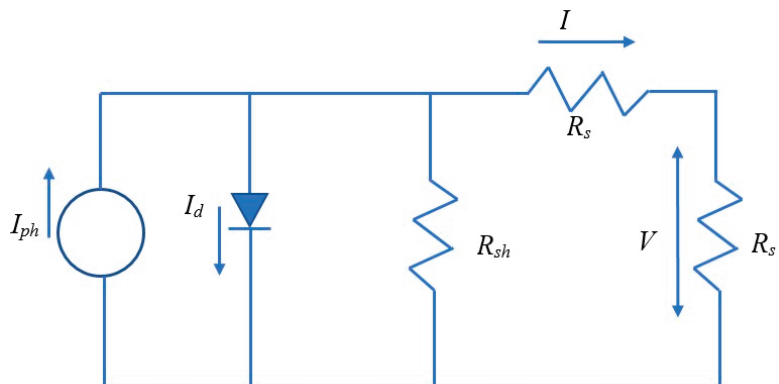


Figure 5. The equivalent circuit for a solar photovoltaic cell with arrows representing the electric signals of the output current (I) of the PV array, the photocurrent (I_{ph}), the reverse saturation current (I_d) of the PV cell, and the output voltage (V) of the PV array.

The TENG can be assembled with other harvesting energy systems to form an HRES by, for example, integrating the TENG with the photovoltaic effect to form a hybrid energy harvester for sustainable power generation [130], developing a hybrid harvesting energy system-based TENG and a piezoelectric nanogenerator to improve the energy harvesting efficiency [131], or designing a hybrid energy system using a TENG and glucose biofuel energy to convert energy from multiple sources of biomechanical and biochemical energies into electricity [132]. TENGs can effectively harvest many wasted energies from our environment when it is used in hybrid energy systems such as harvesting vibration energy by a HES of TENG and an electromagnetic generator [133], collecting vibration energy by a HRES of TENGs and EMGs [134].

Many research groups have been developed the advanced hybrid renewable energy systems for different purposes such as producing energy by a HRES for remote areas and small communities [135], improving efficiency, and conductivity by introducing thermal energy storage techniques [136], promoting sustainable development by implementing optimization and control methods to improve the energy efficiency conversion [137], and improving the power quality for isolated areas using an optimization technique for hydropower and photovoltaic energy systems [138].

4. High Energy Conversion Efficiency, Renewable Energy Integration, Application, and Sustainable Energy Production

4.1. High Energy Conversion Efficiency (HECE)

Energy conversion efficiency (ECE) is the most important index of RESs [139–141]. Many solutions are applied to enhance the ECE of single renewable energy systems, such as applying new materials to improve the ECE of photovoltaic systems [142], building theoretical efficiency relationships to improve the ECE [143], investigating the effect of the structure of energy harvesting systems to improve the ECE of the output performance [144,145], and optimizing parameters to improve the ECE of the energy conversion [146,147]. However, single renewable energy systems have faced many problems that constrain the melioration of the energy conversion efficiency such as seasons, day/night time, rain time, economy, and the degree of input triggers [148,149]. The hybrid renewable energy system (HRES) is a good solution to these issues, with outstanding characteristics that improve the energy conversion efficiency of the system [150,151]. HRESs have high energy conversion efficiency performance as they avoid the limitations of single renewable energy systems by, for example, reducing the effects of temperature and solar radiation to improve the energy conversion efficiency of HRESs [152], enabling cost-effectiveness by optimizing the size of HRESs [153], and achieving high energy conversion efficiency in HRESs by incorporating an energy storage unit [154].

The high energy conversion efficiency (HECE) is really important throughout all steps of the process, including the ideas of HRESs, the designing procedure, the construction stage, the operating procedure, and during other estimations such as economic and environmental impacts [155,156]. The economy is a crucial factor for obtaining capital investment during the development of the HRESs. Net present value (NPV) has generally been used to assess capital investment into HRESs. NPV can be determined using Equation (10) [157]:

$$\hat{E} = \sum_t (R_t - Z_t)(1 + d)^{-t} \quad (10)$$

where

R_t is the outcome of the hybrid project;

Z_t is the cost of the hybrid system in a year;

d is the discount rate.

The energy cost is crucial when evaluating the capital investment efficiency factor. A hybrid renewable energy system has the highest cost efficiency if it has the minimum energy cost. The energy cost can be estimated using Equation (11):

$$p = \frac{\sum_t Z_t (1+d)^{-t}}{\sum_t Q_t (1+d)^{-t}} \quad (11)$$

where

$$Q_t = R_t / p_t;$$

Q_t is the sale volume in one year;

p_t is the energy price.

Many research groups have been focused on improving the efficiency of energy harvesting by applying technologies, techniques, and methods. Ensuring high efficiency and sustainable energy production in hybrid renewable energy systems has been the source of much concern for certain research groups, who have developed a THRES using a TENG and an EMG to achieve high energy conversion efficiency [158], developed a THRES using a TENG and photoelectric conversion to boost the efficiency of the energy conversion [159], constructed a THRES using a TENG and solar cells to achieve high energy conversion efficiency for sustainable agricultural development [160], composed THRESs using triboelectric nanogenerators and piezoelectric nanogenerators to enhance the power conversion efficiency for sensing applications [161], and improved power conversion efficiency by introducing a THRES composed of TENG and EMG technologies for a marine monitoring sensor [162].

The energy conversion efficiency (ECE) of a single renewable energy system is generally governed by Equation (12) [163]:

$$\eta_{sl} = \frac{E_{OUT(total)}}{E_{IN(sl)}} \quad (12)$$

where

η_{sl} is the energy conversion efficiency of a single renewable energy system;

E_{in} is the input energy of the SRES;

E_{out} is the output energy of the SRES.

The efficiency of the SRES is strongly dependent on the input trigger. For example, the efficiency of a solar photovoltaic system increases as the sun rises and decreases as the sun goes down. The energy conversion is intermittent as the night time comes. The energy conversion efficiency of a single renewable energy system can also be disadvantaged by dust accumulation [164]. To improve the energy conversion efficiency of renewable energy systems, many research groups have focused on developing new technologies, methods, materials, and conversion models, such as using $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ material to enhance the efficiency of a solar cell system [165], and integrating some renewable energy conversion technologies utilizing wind power, solar cells, and rainwater to enhance the energy conversion efficiency of an HRES [166]. However, single renewable energy systems still undergo prominent issues due to input trigger conditions such as weather conditions, day/night time, and sunny/raining conditions that negatively influence the efficiency of the system. Hybrid renewable energy is the best solution to the challenge of weather conditions as it absorbs all of the input triggers to achieve high energy conversion efficiency. High energy conversion efficiency refers to the ability of a system to convert as much input energy as possible into output energy. Almost all hybrid renewable energy systems have the feature of high energy conversion efficiency because HRESs consist of more than one single renewable energy system, so they can harvest more renewable energy resources in different weather conditions. To demonstrate this more clearly, a mathematical model is proposed to describe the total energy conversion efficiency of hybrid renewable energy systems:

$$\eta_{HRES} = \frac{E_{OUT(total)}}{E_{IN(total)}} \quad (13)$$

where

η_{HRES} is the energy conversion efficiency of a hybrid renewable energy system;
 $E_{OUT(total)}$ is the total output energy that is produced by the HRES;
 $E_{IN(total)}$ is the total input energy that will be processed by the HRES.
 $E_{IN(total)}$ can be expressed by Equation (14):

$$E_{IN(total)} = \sum_1^n E_i \quad (14)$$

where

n is the number of single renewable energy systems that are assembled into the HRES;
 E_i is the energy of the i th single renewable energy system.

The total input renewable energy ($E_{IN(total)}$) includes n single renewable energy sources that are available in the survey environment for the proposed Equation (14). The total input renewable energy can be expressed by Equation (15):

$$E_{IN(total)} = E_{1(IN)} + E_{2(IN)} + E_{3(IN)} + E_{others(IN)} \quad (15)$$

where it is estimated that

$E_{1(IN)}$ is solar energy (J);
 $E_{2(IN)}$ is wind energy (J);
 $E_{3(IN)}$ is the wasted mechanical energy (J) that is used by the TENG to produce electricity;
 $E_{others(IN)}$ is other energy (J) gathered from $E_{others(IN)i}$ energies when i th = 1 to n .
The total output energy can be expressed by Equation (16):

$$E_{OUT(total)} = E_{1(OUT)} + E_{2(OUT)} + E_{3(OUT)} + E_{others(OUT)} \quad (16)$$

where it is estimated that

$E_{1(OUT)}$ is solar photovoltaics energy (J);
 $E_{2(OUT)}$ is wind energy (J);
 $E_{3(OUT)}$ is the wasted mechanical energy (J) that is used by the TENG to produce electricity;
 $E_{others(OUT)}$ is other output energy (J) that is produced by $E_{others(OUT)i}$ energies when i th = 1 to n .

Therefore, Equation (10) can be expressed by Equation (17):

$$\eta_{HRES} = \sum_1^n \eta_i = \eta_1 + \eta_2 + \eta_3 + \eta_{others} \quad (17)$$

where

η_i is the energy conversion efficiency of the i th single renewable energy system;
 η_1 is the energy conversion efficiency of the solar photovoltaic energy system;
 η_2 is the energy conversion efficiency of the wind power system;
 η_3 is the energy conversion efficiency of the triboelectric nanogenerators;
 η_{others} is the energy conversion efficiency of the other single renewable energy systems.

This reveals that a single renewable energy system is used to convert a part of the total input renewable energy; the single energy conversion efficiency is smaller than that of the HRES. The equations prove that HRESs have high energy conversion efficiency (HECE).

4.2. Renewable Energy Integration (REI)

Renewable energy integration (REI) is a process of energy processing used to incorporate renewable energy into the power grid and other industrial applications [167]. REI is the best solution to reducing carbon dioxide emissions that are produced by the transformation of fossil sources into other energies [168]. REI has been developed for multigeneration

purposes such as electricity, cooling, drying, and heating products [169–171]. REI shows good service ability in energy grids for smart sites [172]. Many research groups have been focused on developing technologies and methods to integrate hybrid renewable energy systems into the power grid, smart grid, and multigeneration systems by, for example, optimizing a sustainable HRES model (from wind, biomass, and solar energies) to integrate into the power grid to improve the sustainability indexes of economy and environment in a petroleum refinery plant [173], developing HRESs that integrate hydrogen energy to meet the objectives of economy, technology, environmental, and social development [174], integrating HRESs into the microgrid for applications in industrial manufacturing and residential consumption with the contribution of photovoltaic power [175], improving the energy sustainability of the power grid by integrating HRESs into the electrical grid with the contribution of hydrogen energy storage units [176], using a hybrid model of the genetic algorithm particle and swarm optimization algorithm (GA-PSO) to optimize the design and management of an HRES for energy cost reduction and energy loss avoidance [177], applying artificial intelligence (AI) to integrate hybrid renewable energy systems into the microgrid to enhance the performance of the power grid [178], penetrating the HRES into the microgrid system in a high electrical consumption area with the aim of having positive impacts on the economy, environment, and technology [179], analyzing strategies to improve REI for developing countries [148], developing compensation technologies to decrease the influence of transmission line compensation on REI in the power grid to improve stability, power balance, and voltage regulation of the system [180], building an REI system for a small area of a university campus to achieve sustainable development with collective self-consumption ability [181], and incorporating a neural network (NN) to manage the power of REI in a direct current microgrid for effective operation of the system [182].

4.3. Applications

The TENG is a clean and renewable energy harvester that converts wasted mechanical energy into electrical energy. TENGs can stand alone when generating electricity or integrate into a hybrid renewable energy system. The TENG-based HRES (THRES) shows many crucial applications. Figure 6 shows applications of TENG-based HRESs such as self-charging power systems, self-powered biomedical complexes, self-powered wearable electronics, self-powered monitoring systems, smart electronics, human healthcare monitoring, and self-powered sensors. Table 2 shows some successful hybrid renewable energy systems constructed to harvest energy that have real-life daily applications. Some successful applications include energy harvesting, self-charging power systems, self-powered biomedical systems, self-powered monitor systems, self-powered wearable electronics, smart electronics, and self-powered sensors.

Energy harvesting (EH) is the biggest duty of THRESs; they convert renewable energy sources into useful energy for many further applications. Researchers have developed a THRES by combining a TENG and a piezoelectric nanogenerator (PENG) to harvest energy from rotational and axial motion types [183], constructed a THRES using a TENG and a PENG to harvest energy for a walking sensor [184], fabricated a THRES using a TENG and an EMG to harvest biomechanical energy [185], developed a THRES using a TENG and an electromagnetic generator (EMG) to harvest biomechanical energy [186], developed a THRES to harvest biomechanical energy [187], and established a THRES using a TENG and an EMG to garner energy [188].

Self-charging power systems (SCPSs) are an outstanding structure in THRESs that integrate a TENG and other pyroelectric, photovoltaic, thermoelectric, and piezoelectric harvesters for many practical applications; these include sustainably driving electronic devices [189], building TENGs and THRESs for self-charging electronic devices [38], supplying sustainable power sources for essential applications of portable electronics and the Internet of Things (IoT) [190], developing THRES-based textiles to provide self-charging power for artificial intelligence applications [191], constructing green THRESs for self-

powering portable electronic devices [192,193], composing a THRES using a TENG and an electromagnetic generator to support the high performance efficiency of a self-charging power system for Internet of Things (IoT) applications [194], and developing a THRES using a TENG and photovoltaic panels to set up an SCPS for charging and driving the electrolysis of seawater by converting hybrid energies from the ocean and solar power [195].

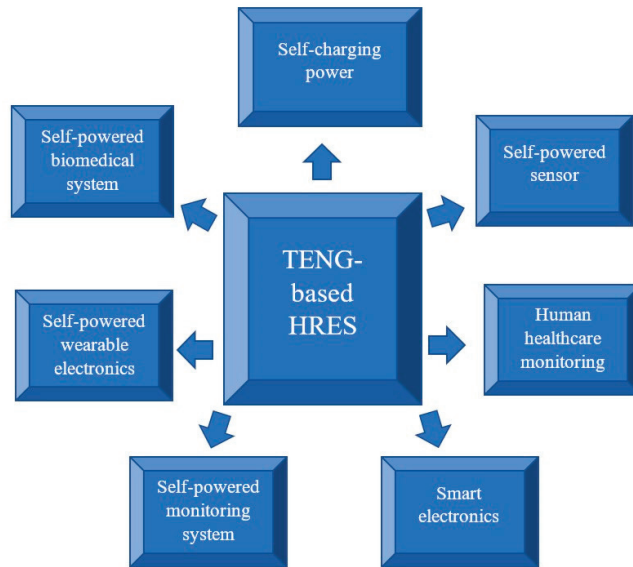


Figure 6. The applications of TENG-based HRESs such as self-charging power systems, self-powered biomedical complexes, self-powered wearable electronics, self-powered monitoring systems, smart electronics, human healthcare monitoring, and self-powered sensors.

Self-powered biomedical systems (SPBSs) are an important application of THRESs that use renewable energy to monitor human healthcare and help treat diseases by powering the medical electronic devices; for example, researchers have built a THRES using a TENG and an electromagnetic generator (EMG) to monitor healthcare systems [196], constructed a THRES using a TENG and a nonlinear electromagnetic generator to harvest biomechanical energy to power portable healthcare monitor machines and portable electronics [197], developed THRESs using a TENG and a piezoelectric generator (PEG) for implantable biomedical applications [198], and composed a THRES by combining a TENG and a PEG for a flexible biosensor device [199,200].

Self-powered monitor systems (SPMSs) have many meaningful practical applications as they have the outstanding features of automation and sustainability; for example, researchers have developed a THRES using a TENG and a PENG for sport monitoring operations [201], constructed a THRES by combining a TENG-EMG and a solar cell to create a self-powered compact sensor network to monitor natural disasters [202], and controlled air quality using a THRES constructed using a TENG and an EMG [203].

Self-powered wearable electronics (SPWEs) are also effective applications of the THRESs which are mobile, lightweight, and sustainable; for example, researchers have developed a THRES using a TENG and an EMG to power sustainable wearable electronics [204], and structured a THRES by combining a TENG and solar cells for wearable electronics [205]. Smart electronics (SEs) also receive a lot of attention from research groups, who have developed a THRES using a TENG and an EMG to scavenge biomechanical energy for smartphones, smartwatches, and Bluetooth devices [206].

Self-powered sensors (SPSs) use TENGs and THRESs in sensing applications and they are lightweight, have tiny dimensions, and are simple structures. Researchers have

developed a THRES using a PEG and a TENG for self-powered sensors [207], built THRESs for biosensors [208], used a THRES comprising a TENG and an EMG to drive a self-powered speed sensor [209], developed a THRES using a TENG and a piezoelectric nanogenerator (PENG) for a healthcare monitoring sensor [210], used a THRES comprising a TENG and a PENG for a self-powered sensing network [211], and composed a THRES by combining a TENG and a piezoelectric nanogenerator for a self-powered human behavior sensor [212].

Table 2. Some successful hybrid renewable energy systems and their applications.

THRES	Hybrid Type	Performance	Application	Ref.
TCO-HG	PENG–TENG	Driving 60 LEDs	Harvesting energy, charging, and lighting.	[183]
PTNG	TENG–PENG	70 μ W	Harvesting energy, and walking sensor.	[184]
THRES	TENG–EMG	630 mA	Harvesting energy, charging, and wireless power transmission.	[188]
UHO-TEHG	TENG and EMG	1.02 W	Human healthcare monitoring and self-powering portable electronics.	[196]
EINR-HG	TENG and EMG	131.4 mW	Portable healthcare monitoring machines and portable electronics.	[197]
PEDOT	TENG and PENG	1.71 mW	Sport monitoring operations, healthcare applications, and smart home devices.	[201]
THRES	TENG–EMG	6 W	Harvesting biomechanical energy and sustainable development.	[186]
HMI-HBNG	TENG–EMG	185 W/m ²	Harvesting biomechanical energy, self-powered systems, and smart electronics.	[206]
RSHG	TENG–EMG	48 V, 1 mA	Harvesting energy, lighting LEDs, and driving electric watch.	[158]
FHNG	TENG–PENG	-	Healthcare monitoring sensor and self-powered devices.	[210]
HTEPENG	TENG–PENG	3.7 W/m ²	Self-powered sensing network and portable electronics.	[211]
PSC/TENG	TENG–Solar cells	2.62 Wm ⁻²	Renewable power generation and agricultural application.	[160]

4.4. Sustainable Energy Production (SEP)

Sustainable energy production (SEP) plays a large role in achieving net-zero goals and sustainable development in our society. Figure 7 shows a basic diagram of sustainable energy development accompanied by environmental and economic development. The figure demonstrates the relationship between energy, economy, and social development for the purpose of sustainable energy production [213]. There have been a lot of studies concerned with developing technologies, methods, and national policies that focus on building sustainable energy production strategies; these studies have synthesized nanomaterials for sustainable energy production [214], developed national energy policies to enable SEP in Turkey [215], developed green SEP in China by tracing garden waste biomass sources [216], constructed a theoretical framework for hybrid renewable energy using photovoltaic energy and hydro-energy for SEP [217], projected long-term SEP in Saudi Arabia until 2030 [218], developed an SEP strategy by pursuing sustainable development goal 7 (SDG 7) criteria including access to affordable modern energy services, sustainable energy consumption, sustainable energy supply, and one more criterion of energy security [219], applied an energy optimization form to accomplish SEP with reduction strategies to alleviate the environmental impact of pollution and emissions in Pakistan [220], provided national policies to help decrease emissions and planned strategies to accomplish SEP in Turkey [221], developed solar energy and biomass energy for SEP in Nigeria [222], outlined a plan to achieve sustainable energy development by using renewable energy sources and making

national policies for SEP in Azerbaijan [213], and determined the impact of globalization on sustainable energy production by developing and using strategies of renewable energy.

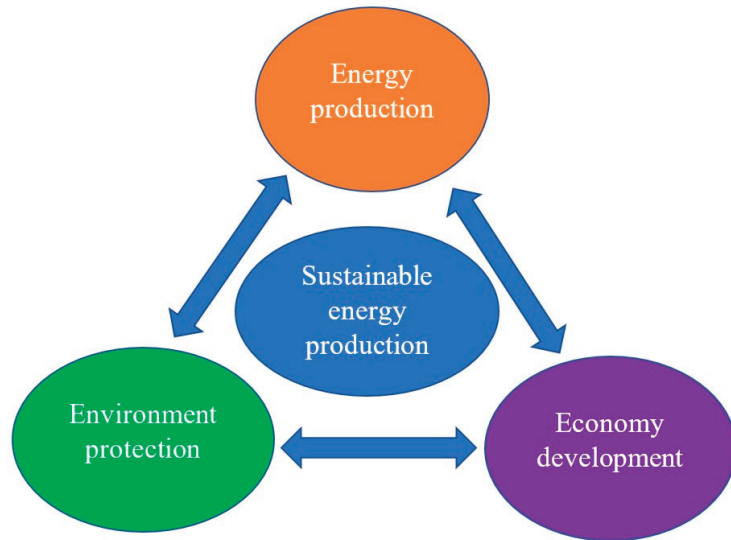


Figure 7. The diagram of sustainable energy development with the roles of energy production, environment development, and economic development.

5. Benefits, Challenges, and Solutions

5.1. Benefits

Hybrid renewable energy system have many benefits such as providing clean energy for everyone with high availability and low cost [223], increasing power penetration approaches, decreasing the intermittency of renewable energy sources, enhancing the reliability of research and development into renewable energy resources, promoting the electrification of remote and rural areas, encouraging the exploitation of new energy harvesting techniques [224,225], bringing the circular economy benefit to rural areas [226], achieving the rural health benefits of cost-effectiveness and greenhouse gas emissions reduction [227], contributing to the net-zero energy area by using solar power, thermal energy storage, and heat pump units [228], enhancing the efficiency and stability of energy systems [229], accomplishing cost reduction and efficiency improvement [230], improving the reliability of power systems [231], decreasing carbon dioxide emissions [232], improving socioeconomic development issues regarding healthcare, education, and economy in local areas [233], combining single renewable energy systems into one HRES to increase the electrification of countries and territories [234], driving a multigeneration system [235], contributing to sustainable development all over the world [236], effectively using waste for net-zero purposes [237], having outstanding characteristics of reliable, clean, and affordable energy for sustainable development [238], avoiding the intermittence resulting from a single renewable energy [239], and providing an affordable energy source for low-income households [240].

There are tremendous benefits to introducing triboelectric nanogenerators into hybrid renewable energy systems because of the outstanding characteristics of TENGs such as their flexible structure [241], self-powered source [242], lightweight nature, use of easy-to-find materials, simple structure, renewable energy resource [243], and use as a sustainable energy source [244]. A triboelectric nanogenerator can easily integrate into the power grid. In summary, renewable energy is the answer for reducing carbon dioxide emissions, protecting the environment, and pushing sustainable development. HRESs bring

about many benefits, for example, decreasing fluctuations in RE, improving the output performance of energy systems, enhancing energy conversion efficiency, and optimizing energy transduction.

5.2. Challenges

There are lots of challenges that HRESs face during all the procedures of designing, building, and operating the system, such as technology and technique problems. Renewable energy systems need a large area. RES sources are often affected by the environment. Climate conditions, such as the day/night period, sun/raining time, and windy conditions, have directly affected RES performance [245]. By integrating HRESs into the power grid, there are many challenges that research groups have had to overcome, such as enhancing economic effectiveness, improving energy storage technologies [246], finding methods and techniques to solve problems of reliability in the systems [247], seeking technologies to solve the problems of optimization and effectiveness of the system [248], finding a way to improve the reliability of the power system in an HRES [249], researching methods and technologies to achieve multi-objective optimization of HRESs [250], facing the problems of computation, finance, and environment in rural areas when deploying HRESs [251], facing prominent problems from policies, regulations, and institutions when implementing HRESs [252], controlling excess electricity from HRESs [253], facing technology differences during the integration of HRESs with SRESs [254], determining the parameters required to penetrate a single renewable energy into an HRES [255], problem of high costs when constructing, exploiting, and operating HRESs [256], building mathematical models to optimize the size of HRESs [257], and facing many problems with the stability, quality, reliability, and operation of HRESs [258]. To briefly summarize, there still remain challenges that TENGs will encounter before they can be used in HRESs because the input sources are exposed to natural fluctuation and weather conditions. The output performance of TENGs and HRESs are small in comparison with the energy needed for human development. The limitations of technologies, materials, and working mechanisms are big problems during the integration of TENGs into HRESs and the power grid. The different distribution of renewable energy sources, which can cause disadvantage problems for establishing the infrastructure of the HRESs.

5.3. Solutions

To overcome these challenges and gain even more benefits from HRESs, there are many solutions that have been developed by research groups, such as applying an optimization method in the management and modeling of HRESs to control the output performance of HRESs [259], using a distributed energy resource customer adoption tool to optimize the size of HRESs for a microgrid [260], promoting the benefit sharing strategy to obtain maximal benefits of HRESs with minimal electrical cost and minimal carbon dioxide emissions [261], using a hybrid optimization tool to optimize HRESs for the criteria of efficiency, electrical cost, environmental protection, emission problems, and lifetime of the hybrid system [262], using innovative technologies of machine learning and the advanced inverter to solve challenges of sustainability, intermittency, optimization, power storage, and management of the system [263], building frameworks for integrating RESs into the power grid with the benefits of effective energy management, real-time power management, and power forecasting ability [264], using the HOMER Pro tool to optimize HRESs and enhance their advantages, such as decreased gas emissions, reduced electrical costs, utilization of RES resources, and effective combination of renewable energy units of biogas, photovoltaic energy, and energy storage systems [265], focusing on designing and optimizing HRESs to achieve sustainable development, green transportation, clean hydrogen energy, cost-effectiveness, and rural electrification [266], supporting frameworks to help with the selection of an HRES for sustainable development [267], using a hybrid optimization model for electric renewable software to achieve investment reduction and size optimization for HRESs [268], using a receding horizon optimization method with pre-

dictive control algorithms to optimize the operation of HRESs [269], using battery storage systems to enhance the reliability of HRESs [270,271], using a multiobjective optimization model to deal with the problems of economy, environmental impact, and carbon dioxide emission reductions [272], introducing a particle swarm optimization (PSO) to optimize HRESs with the objectives of social-political improvements [273], applying a mixed-integer linear programming tool to determine the high energy demand area [274], using artificial intelligence technology (AI) to solve problems with efficiency, complexity, and reliability in HRESs [275], using a hybrid meta-heuristic optimization technology to ensure reliability and accomplish the multiobjective goals of HRESs [276], using artificial neural networks (ANNs) to predict the energy generation needed for sustainable energy production [277], and using simulation tools to optimize HRESs with reliability indicators of financial cost optimization, excess energy minimization, and load protection [278]. In brief, there are solutions to fight the challenges of HRESs that include using energy storage units to store energy, reusing energy during periods where there is a lack of input trigger or bad weather conditions, and using more and more renewable energy transduction systems to optimize the energy conversion efficiency. The development of new technologies, methods, and materials to improve the output performance of HRESs are still needed. The utilization of optimization methods when identifying the distribution of renewable energy sources for siting infrastructure of HRESs is encouraged to achieve the highest possible energy conversion efficiency.

6. Conclusions

Renewable energy is crucial for sustainable energy development given the rapidly increasing energy demand in the world. A single renewable energy system has limitations to its stable generation due to the influences of weather conditions, seasons, day/night time, and working conditions. Triboelectric nanogenerators produce electricity by converting mechanical energies that exist prominently in the surrounding environment such as motion, walking, vibrations, water flow, sounds, and ocean energy. Triboelectric nanogenerators are potential for renewable energy transduction systems because they can convert many renewable energy sources (such as wind and ocean energy) into electricity. Hybrid renewable energy systems can improve output performance, enhance efficiency, and decrease fluctuation during energy generation procedures. This paper reviews recent methods and techniques of developing HRESs to enhance energy transduction efficiency, output performance, and sustainability of HRESs. Introducing triboelectric nanogenerators into HRESs brings many benefits such as energy conversion efficiency improvement, sustainable energy generation, and other practical applications. This paper mentions the advantages, challenges, and solutions offered by HRESs to improve its efficiency, development, and applications. The results show that TENGs are a prospective technology that can be integrated into HRESs to solve the problems of intermittency and variability in RE systems. The results hope to motivate the development of hybrid renewable energy systems for practical applications and sustainable energy development in the near future.

Author Contributions: Conceptualization, V.-L.T. and C.-K.C.; methodology, V.-L.T. and C.-K.C.; validation, V.-L.T.; formal analysis, V.-L.T.; investigation, V.-L.T. and C.-K.C.; resources, V.-L.T. and C.-K.C.; data curation, V.-L.T.; writing—original draft preparation, V.-L.T.; writing—review and editing, V.-L.T. and C.-K.C.; visualization, V.-L.T.; supervision, C.-K.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially sponsored by the National Science and Technology Council (NSTC), Taiwan, under grant No. NSTC113-222-1-E006-099MY3. It was also supported in part by the School of Mechanical and Automotive Engineering, Hanoi University of Industry, Hanoi 100000, Vietnam.

Data Availability Statement: Data are presented in the coauthors' research results and schematic drawings and are available on request.

Acknowledgments: We thank the Core Facility Center in National Cheng Kung University, Taiwan, for their equipment support.

Conflicts of Interest: The authors declare no conflicts of interest.

Symbol and Acronym:

CO ₂	Carbon dioxide
t	Tonne
NZR	Net-Zero Roadmap
NZEs	Net-zero emissions
IEA	International Energy Agency
°C	Degrees Celsius
RG	Renewable generation
TWh	Terawatt-hours
WEO	World Energy Outlook
GSR2024	2024 Global Status Report
STEPS	Stated Policies Scenario
GW	Gigawatts
TENG	Triboelectric nanogenerator
SDG-7	Sustainable development goal 7
SRES	Single renewable energy system
PVS	Photovoltaic system
HRES	Hybrid renewable energy system
AC	Alternative current
DC	Direct current
MEMS	Micro-electro-mechanical system
Voc	Open-circuit voltage
SEP	Sustainable energy production
ECE	Energy conversion efficiency
REI	Renewable energy integration
PSO	Particle swarm optimization
ANN	Artificial neural network
AI	Artificial intelligence
EH	Energy harvesting
SCPS	Self-charging power system
SPBS	Self-powered biomedical system
SPMS	Self-powered monitor system
SPWE	Self-powered wearable electronic
SE	Smart electronic
SPS	Self-powered sensor
LED	Light-emitting diode
mA	Milliamperere
μW	Microwatt
W	Watt
Wm ⁻²	Watt per square meter
W/m ²	Watt per square meter
V	Voltage
J	Joule

References

1. Ang, T.-Z.; Salem, M.; Kamarol, M.; Das, H.S.; Nazari, M.A.; Prabakaran, N. A comprehensive study of renewable energy sources: Classifications, challenges and suggestions. *Energy Strategy Rev.* **2022**, *43*, 100939. [CrossRef]
2. Strielkowski, W.; Civiń, L.; Tarkhanova, E.; Tvaronavičienė, M.; Petrenko, Y. Renewable Energy in the Sustainable Development of Electrical Power Sector: A Review. *Energies* **2021**, *14*, 8240. [CrossRef]
3. Hassan, Q.; Algburi, S.; Sameen, A.Z.; Al-Musawi, T.J.; Al-Jiboory, A.K.; Salman, H.M.; Ali, B.M.; Jaszczur, M. A comprehensive review of international renewable energy growth. *Energy Built Environ.* **2024**, *in press*. [CrossRef]
4. Langer, J.; Quist, J.; Blok, K. Review of Renewable Energy Potentials in Indonesia and Their Contribution to a 100% Renewable Electricity System. *Energies* **2021**, *14*, 7033. [CrossRef]

5. Johannsen, R.M.; Mathiesen, B.V.; Kermeli, K.; Crijns-Graus, W.; Østergaard, P.A. Exploring pathways to 100% renewable energy in European industry. *Energy* **2023**, *268*, 126687. [CrossRef]
6. Marks-Bielska, R.; Bielski, S.; Pik, K.; Kurowska, K. The Importance of Renewable Energy Sources in Poland's Energy Mix. *Energies* **2020**, *13*, 4624. [CrossRef]
7. Trinh, V.L.; Chung, C.K. Renewable energy for SDG-7 and sustainable electrical production, integration, industrial application, and globalization: Review. *Clean. Eng. Technol.* **2023**, *15*, 100657. [CrossRef]
8. Salah, S.I.; Eltaweel, M.; Abeykoon, C. Towards a sustainable energy future for Egypt: A systematic review of renewable energy sources, technologies, challenges, and recommendations. *Clean. Eng. Technol.* **2022**, *8*, 100497. [CrossRef]
9. Gao, C.; Chen, H. Electricity from renewable energy resources: Sustainable energy transition and emissions for developed economies. *Util. Policy* **2023**, *82*, 101543. [CrossRef]
10. Adebayo, T.S.; Oladipupo, S.D.; Adeshola, I.; Rjoub, H. Wavelet analysis of impact of renewable energy consumption and technological innovation on CO2 emissions: Evidence from Portugal. *Environ. Sci. Pollut. Res.* **2022**, *29*, 23887–23904. [CrossRef]
11. Adebayo, T.S. Renewable Energy Consumption and Environmental Sustainability in Canada: Does Political Stability Make a Difference? *Environ. Sci. Pollut. Res.* **2022**, *29*, 61307–61322. [CrossRef] [PubMed]
12. Adebayo, T.S.; Ullah, S.; Kartal, M.T.; Ali, K.; Pata, U.K.; Ağa, M. Endorsing sustainable development in BRICS: The role of technological innovation, renewable energy consumption, and natural resources in limiting carbon emission. *Sci. Total Environ.* **2023**, *859*, 160181. [CrossRef] [PubMed]
13. Adebayo, T.S.; Kartal, M.T.; Ağa, M.; Al-Faryan, M.A.S. Role of country risks and renewable energy consumption on environmental quality: Evidence from MINT countries. *J. Environ. Manag.* **2023**, *327*, 116884. [CrossRef] [PubMed]
14. Kirikkaleli, D.; Güngör, H.; Adebayo, T.S. Consumption-based carbon emissions, renewable energy consumption, financial development and economic growth in Chile. *Bus. Strategy Environ.* **2022**, *31*, 1123–1137. [CrossRef]
15. Lo, K.-Y.; Yeoh, J.H.; Hsieh, I.Y.L. Towards Nearly Zero-Energy Buildings: Smart Energy Management of Vehicle-to-Building (V2B) Strategy and Renewable Energy Sources. *Sustain. Cities Soc.* **2023**, *99*, 104941. [CrossRef]
16. Raihan, A.; Tuspekova, A. Role of economic growth, renewable energy, and technological innovation to achieve environmental sustainability in Kazakhstan. *Curr. Res. Environ. Sustain.* **2022**, *4*, 100165. [CrossRef]
17. Adebayo, T.S.; Kirikkaleli, D. Impact of renewable energy consumption, globalization, and technological innovation on environmental degradation in Japan: Application of wavelet tools. *Environ. Dev. Sustain.* **2021**, *23*, 16057–16082. [CrossRef]
18. Shan, S.; Genç, S.Y.; Kamran, H.W.; Dinca, G. Role of green technology innovation and renewable energy in carbon neutrality: A sustainable investigation from Turkey. *J. Environ. Manag.* **2021**, *294*, 113004. [CrossRef]
19. Zhu, D.; Mortazavi, S.M.; Maleki, A.; Aslani, A.; Yousefi, H. Analysis of the robustness of energy supply in Japan: Role of renewable energy. *Energy Rep.* **2020**, *6*, 378–391. [CrossRef]
20. Samour, A.; Baskaya, M.M.; Tursoy, T. The Impact of Financial Development and FDI on Renewable Energy in the UAE: A Path towards Sustainable Development. *Sustainability* **2022**, *14*, 1208. [CrossRef]
21. Thellufsen, J.Z.; Lund, H.; Sorknaes, P.; Østergaard, P.A.; Chang, M.; Drysdale, D.; Nielsen, S.; Djørup, S.R.; Sperling, K. Smart energy cities in a 100% renewable energy context. *Renew. Sustain. Energy Rev.* **2020**, *129*, 109922. [CrossRef]
22. Zheng, H.; Song, M.; Shen, Z. The evolution of renewable energy and its impact on carbon reduction in China. *Energy* **2021**, *237*, 121639. [CrossRef]
23. Adebayo, T.S.; Ullah, S. Towards a sustainable future: The role of energy efficiency, renewable energy, and urbanization in limiting CO emissions in Sweden. *Sustain. Dev.* **2024**, *32*, 244–259. [CrossRef]
24. Yuan, X.; Su, C.-W.; Umar, M.; Shao, X.; Lobon, O.-R. The race to zero emissions: Can renewable energy be the path to carbon neutrality? *J. Environ. Manag.* **2022**, *308*, 114648. [CrossRef] [PubMed]
25. Su, C.-W.; Pang, L.-D.; Tao, R.; Shao, X.; Umar, M. Renewable energy and technological innovation: Which one is the winner in promoting net-zero emissions? *Technol. Forecast. Soc. Chang.* **2022**, *182*, 121798. [CrossRef]
26. IEA. *Net Zero Roadmap: A Global Pathway to Keep the 1.5 °C Goal in Reach*; IEA: Paris, France, 2021. Available online: <https://www.iea.org/reports/net-zero-roadmap-a-global-pathway-to-keep-the-15-0c-goal-in-reach> (accessed on 12 June 2024).
27. IEA. *World Energy Outlook 2023*; IEA: Paris, France, 2023; Available online: <https://www.iea.org/reports/world-energy-outlook-2023> (accessed on 13 June 2024).
28. Zoundi, Z. CO₂ emissions, renewable energy and the Environmental Kuznets Curve, a panel cointegration approach. *Renew. Sustain. Energy Rev.* **2017**, *72*, 1067–1075. [CrossRef]
29. REN21. *Renewables 2024 Global Status Report Collection*. 2023. Available online: <https://www.ren21.net/gsr-2024/> (accessed on 5 May 2024).
30. Olabi, A.G.; Abdalkareem, M.A. Renewable energy and climate change. *Renew. Sustain. Energy Rev.* **2022**, *158*, 112111. [CrossRef]
31. Yang, X.; Pang, J.; Teng, F.; Gong, R.; Springer, C. The environmental co-benefit and economic impact of China's low-carbon pathways: Evidence from linking bottom-up and top-down models. *Renew. Sustain. Energy Rev.* **2021**, *136*, 110438. [CrossRef]
32. Liu, Y.; Feng, C. Promoting renewable energy through national energy legislation. *Energy Econ.* **2023**, *118*, 106504. [CrossRef]
33. Gyimah, J.; Yao, X.; Tachega, M.A.; Sam Hayford, I.; Opoku-Mensah, E. Renewable energy consumption and economic growth: New evidence from Ghana. *Energy* **2022**, *248*, 123559. [CrossRef]
34. Ahmed, A.; Ge, T.; Peng, J.; Yan, W.-C.; Tee, B.T.; You, S. Assessment of the renewable energy generation towards net-zero energy buildings: A review. *Energy Build.* **2022**, *256*, 111755. [CrossRef]

35. Cheng, T.; Gao, Q.; Wang, Z.L. The Current Development and Future Outlook of Triboelectric Nanogenerators: A Survey of Literature. *Adv. Mater. Technol.* **2019**, *4*, 1800588. [CrossRef]
36. Trinh, V.-L.; Chung, C.-K. Advances in Triboelectric Nanogenerators for Sustainable and Renewable Energy: Working Mechanism, Tribo-Surface Structure, Energy Storage-Collection System, and Applications. *Processes* **2023**, *11*, 2796. [CrossRef]
37. Kim, Y.; Kim, I.; Im, M.; Kim, D. Shear Thickening Fluid and Sponge-Hybrid Triboelectric Nanogenerator for a Motion Sensor Array-Based Lying State Detection System. *Materials* **2024**, *17*, 3536. [CrossRef]
38. Gołabek, J.; Strankowski, M. A Review of Recent Advances in Human-Motion Energy Harvesting Nanogenerators, Self-Powering Smart Sensors and Self-Charging Electronics. *Sensors* **2024**, *24*, 1069. [CrossRef]
39. Huang, X.; Wei, B.; Ling, Z.; Yang, F.; Pang, H. A Low-Frequency Vibration Sensor Based on Ball Triboelectric Nanogenerator for Marine Pipeline Condition Monitoring. *Sensors* **2024**, *24*, 3817. [CrossRef]
40. Wu, L.; Ren, Z.; Wang, Y.; Tang, Y.; Wang, Z.L.; Yang, R. Miniaturized and High Volumetric Energy Density Power Supply Device Based on a Broad-Frequency Vibration Driven Triboelectric Nanogenerator. *Micromachines* **2024**, *15*, 645. [CrossRef]
41. Yan, F.; Zhao, J.; Li, F.; Chu, Y.; Du, H.; Sun, M.; Xi, Z.; Du, T.; Xu, M. High-Performance Coaxial Counter-Rotating Triboelectric Nanogenerator with Lift-Drag Hybrid Blades for Wind Energy Harvesting. *Nanomaterials* **2024**, *14*, 598. [CrossRef]
42. Yan, J.; Tang, Z.; Mei, N.; Zhang, D.; Zhong, Y.; Sheng, Y. Research Progress on the Application of Triboelectric Nanogenerators for Wind Energy Collection. *Micromachines* **2023**, *14*, 1592. [CrossRef]
43. Cai, X.; Liu, Z.; Dong, J.; Li, H.; Han, J.; Huang, J.; Chen, H. U-Shaped Tube Based Liquid-Solid Triboelectric Nanogenerator for Harvesting Unutilized Compressed Air Energy. *Micromachines* **2023**, *14*, 2057. [CrossRef]
44. Li, X.; Xu, L.; Wang, Z.L. Networking Strategies of Triboelectric Nanogenerators for Harvesting Ocean Blue Energy. *Nanoenergy Adv.* **2024**, *4*, 70–96. [CrossRef]
45. Huang, Y.; Yang, S.; Fan, J.; Du, Z.; Lin, B.; Tu, Y.; Pan, L. Study of a Center Pipe Oscillating Column Wave Energy Converter Combined with a Triboelectric Nanogenerator Device. *J. Mar. Sci. Eng.* **2024**, *12*, 100. [CrossRef]
46. Matin Nazar, A.; Idala Egbe, K.-J.; Abdollahi, A.; Hariri-Ardebili, M.A. Triboelectric Nanogenerators for Energy Harvesting in Ocean: A Review on Application and Hybridization. *Energies* **2021**, *14*, 5600. [CrossRef]
47. Liang, X.; Liu, S.; Yang, H.; Jiang, T. Triboelectric Nanogenerators for Ocean Wave Energy Harvesting: Unit Integration and Network Construction. *Electronics* **2023**, *12*, 225. [CrossRef]
48. Liu, B.; Du, T.; Xu, X.; Liu, J.; Zhu, P.; Guo, L.; Li, Y.; Wang, T.; Zou, Y.; Wang, H.; et al. An Underwater Triboelectric Biomechanical Energy Harvester to Power the Electronic Tag of Marine Life. *J. Mar. Sci. Eng.* **2023**, *11*, 1766. [CrossRef]
49. Walden, R.; Kumar, C.; Mulvihill, D.M.; Pillai, S.C. Opportunities and Challenges in Triboelectric Nanogenerator (TENG) based Sustainable Energy Generation Technologies: A Mini-Review. *Chem. Eng. J. Adv.* **2022**, *9*, 100237. [CrossRef]
50. Jiang, L.; Liu, X.; Lv, J.; Li, G.; Yang, P.; Ma, Y.; Zou, H.; Wang, Z.L. Fluid-based triboelectric nanogenerators: Unveiling the prolific landscape of renewable energy harvesting and beyond. *Energy Environ. Sci.* **2024**, *17*, 3700–3738. [CrossRef]
51. Wu, Y.; Qu, J.; Chu, P.K.; Shin, D.-M.; Luo, Y.; Feng, S.-P. Hybrid photovoltaic-triboelectric nanogenerators for simultaneously harvesting solar and mechanical energies. *Nano Energy* **2021**, *89*, 106376. [CrossRef]
52. United Nations Department of Social and Economic Affairs. *Sustainable Development Goal 7: Ensure Access to Affordable, Reliable, Sustainable and Modern Energy for All*; UN DESA: New York, NY, USA, 2017. Available online: <https://sdgs.un.org/goals/goal7> (accessed on 5 May 2024).
53. Ammari, C.; Belatrache, D.; Touhami, B.; Makhloufi, S. Sizing, optimization, control and energy management of hybrid renewable energy system—A review. *Energy Built Environ.* **2022**, *3*, 399–411. [CrossRef]
54. Olatomiwa, L.; Mekhilef, S.; Ismail, M.S.; Moughavemi, M. Energy management strategies in hybrid renewable energy systems: A review. *Renew. Sustain. Energy Rev.* **2016**, *62*, 821–835. [CrossRef]
55. Akarsu, B.; Serdar Genç, M. Optimization of electricity and hydrogen production with hybrid renewable energy systems. *Fuel* **2022**, *324*, 124465. [CrossRef]
56. Khare, V.; Nema, S.; Baredar, P. Solar-wind hybrid renewable energy system: A review. *Renew. Sustain. Energy Rev.* **2016**, *58*, 23–33. [CrossRef]
57. Al-Othman, A.; Tawalbeh, M.; Martis, R.; Dhoun, S.; Orhan, M.; Qasim, M.; Ghani Olabi, A. Artificial intelligence and numerical models in hybrid renewable energy systems with fuel cells: Advances and prospects. *Energy Convers. Manag.* **2022**, *253*, 115154. [CrossRef]
58. Chong, L.W.; Wong, Y.W.; Rajkumar, R.K.; Rajkumar, R.K.; Isa, D. Hybrid energy storage systems and control strategies for stand-alone renewable energy power systems. *Renew. Sustain. Energy Rev.* **2016**, *66*, 174–189. [CrossRef]
59. Kallio, S.; Siroux, M. Hybrid renewable energy systems based on micro-cogeneration. *Energy Rep.* **2022**, *8*, 762–769. [CrossRef]
60. Bocklich, T. Hybrid Energy Storage Systems for Renewable Energy Applications. *Energy Procedia* **2015**, *73*, 103–111. [CrossRef]
61. Shivarama Krishna, K.; Sathish Kumar, K. A review on hybrid renewable energy systems. *Renew. Sustain. Energy Rev.* **2015**, *52*, 907–916. [CrossRef]
62. Al Afif, R.; Ayed, Y.; Maaitah, O.N. Feasibility and optimal sizing analysis of hybrid renewable energy systems: A case study of Al-Karak, Jordan. *Renew. Energy* **2023**, *204*, 229–249. [CrossRef]
63. Jia, K.; Liu, C.; Li, S.; Jiang, D. Modeling and optimization of a hybrid renewable energy system integrated with gas turbine and energy storage. *Energy Convers. Manag.* **2023**, *279*, 116763. [CrossRef]

64. Cheraghi, R.; Hossein Jahangir, M. Multi-objective optimization of a hybrid renewable energy system supplying a residential building using NSGA-II and MOPSO algorithms. *Energy Convers. Manag.* **2023**, *294*, 117515. [CrossRef]
65. Lee, J.; Lin, K.-Y.A.; Jung, S.; Kwon, E.E. Hybrid renewable energy systems involving thermochemical conversion process for waste-to-energy strategy. *Chem. Eng. J.* **2023**, *452*, 139218. [CrossRef]
66. Aghaloo, K.; Ali, T.; Chiu, Y.-R.; Sharifi, A. Optimal site selection for the solar-wind hybrid renewable energy systems in Bangladesh using an integrated GIS-based BWM-fuzzy logic method. *Energy Convers. Manag.* **2023**, *283*, 116899. [CrossRef]
67. Allouhi, A.; Rehman, S. Grid-connected hybrid renewable energy systems for supermarkets with electric vehicle charging platforms: Optimization and sensitivity analyses. *Energy Rep.* **2023**, *9*, 3305–3318. [CrossRef]
68. Sivasubramanian, R.; Aravind Vaithilingam, C.; Indira, S.S.; Paiman, S.; Misron, N.; Abubakar, S. A review on photovoltaic and nanogenerator hybrid system. *Mater. Today Energy* **2021**, *20*, 100772. [CrossRef]
69. Razaghi, Z.; Zamanpour, F.; Vafaiee, M.; Letafati, M.; Gholami, M.; Taghavinia, N.; Khalaj, B.H.; Mohammadpour, R.; Irajizad, A. Versatile hybrid transparent triboelectric nanogenerator-solar cell device enables efficient energy harvesting and personal handwriting recognition. *Compos. Part B Eng.* **2024**, *283*, 111615. [CrossRef]
70. Deshmukh, M.K.; Deshmukh, S.S. Modeling of hybrid renewable energy systems. *Renew. Sustain. Energy Rev.* **2008**, *12*, 235–249. [CrossRef]
71. Wilberforce, T.; Olabi, A.G.; Sayed, E.T.; Alalmi, A.H.; Abdelkareem, M.A. Wind turbine concepts for domestic wind power generation at low wind quality sites. *J. Clean. Prod.* **2023**, *394*, 136137. [CrossRef]
72. Schubel, P.J.; Crossley, R.J. Wind Turbine Blade Design. *Energies* **2012**, *5*, 3425–3449. [CrossRef]
73. Trinh, V.L.; Chung, C.K. A Facile Method and Novel Mechanism Using Microneedle-Structured PDMS for Triboelectric Generator Applications. *Small* **2017**, *13*, 11. [CrossRef]
74. Trinh, V.L.; Chung, C.K. Harvesting mechanical energy, storage, and lighting using a novel PDMS based triboelectric generator with inclined wall arrays and micro-topping structure. *Appl. Energy* **2018**, *213*, 353–365. [CrossRef]
75. Yoon, H.-J.; Ryu, H.; Kim, S.-W. Sustainable powering triboelectric nanogenerators: Approaches and the path towards efficient use. *Nano Energy* **2018**, *51*, 270–285. [CrossRef]
76. Wang, Y.; Li, Z.; Fu, H.; Xu, B. Sustainable triboelectric nanogenerators based on recycled materials for biomechanical energy harvesting and self-powered sensing. *Nano Energy* **2023**, *115*, 108717. [CrossRef]
77. Shi, Z.; Zhang, Y.; Gu, J.; Liu, B.; Fu, H.; Liang, H.; Ji, J. Triboelectric Nanogenerators: State of the Art. *Sensors* **2024**, *24*, 4298. [CrossRef] [PubMed]
78. Tahir, H.R.; Malengier, B.; Sujana, S.; Van Langenhove, L. Embroidery Triboelectric Nanogenerator for Energy Harvesting. *Sensors* **2024**, *24*, 3782. [CrossRef]
79. Liu, L.; Hu, T.; Zhao, X.; Lee, C. Recent Progress in Blue Energy Harvesting Based on Triboelectric Nanogenerators. *Nanoenergy Adv.* **2024**, *4*, 156–173. [CrossRef]
80. Wang, H.; Xu, L.; Wang, Z. Advances of High-Performance Triboelectric Nanogenerators for Blue Energy Harvesting. *Nanoenergy Adv.* **2021**, *1*, 32–57. [CrossRef]
81. Wu, Y.; Li, Y.; Zou, Y.; Rao, W.; Gai, Y.; Xue, J.; Wu, L.; Qu, X.; Liu, Y.; Xu, G.; et al. A multi-mode triboelectric nanogenerator for energy harvesting and biomedical monitoring. *Nano Energy* **2022**, *92*, 106715. [CrossRef]
82. Khandelwal, G.; Maria Joseph Raj, N.P.; Kim, S.-J. Triboelectric nanogenerator for healthcare and biomedical applications. *Nano Today* **2020**, *33*, 100882. [CrossRef]
83. Xiao, X.; Chen, G.; Libanori, A.; Chen, J. Wearable Triboelectric Nanogenerators for Therapeutics. *Trends Chem.* **2021**, *3*, 279–290. [CrossRef]
84. Wang, C.; Shi, Q.; Lee, C. Advanced Implantable Biomedical Devices Enabled by Triboelectric Nanogenerators. *Nanomaterials* **2022**, *12*, 1366. [CrossRef]
85. Zheng, Q.; Zhang, H.; Shi, B.; Xue, X.; Liu, Z.; Jin, Y.; Ma, Y.; Zou, Y.; Wang, X.; An, Z.; et al. In Vivo Self-Powered Wireless Cardiac Monitoring via Implantable Triboelectric Nanogenerator. *ACS Nano* **2016**, *10*, 6510–6518. [CrossRef] [PubMed]
86. Zhang, X.-S.; Su, M.; Brugger, J.; Kim, B. Pencil-ing a triboelectric nanogenerator on paper for autonomous power MEMS applications. *Nano Energy* **2017**, *33*, 393–401. [CrossRef]
87. Shao, B.; Chen, Z.; Su, H.; Peng, S.; Song, M. The Latest Advances in Ink-Based Nanogenerators: From Materials to Applications. *Int. J. Mol. Sci.* **2024**, *25*, 6152. [CrossRef] [PubMed]
88. Sun, X.; Wei, Y.; Sun, Y.; Yuan, J.; Chen, H.; Chen, Z.; Wang, M.; Luo, L. Preparation of Polyaniline-Modified Cellulose/PDMS Composite Triboelectric Material and Application of Its Pretreatment in MOW Pulp. *Polymers* **2024**, *16*, 1413. [CrossRef] [PubMed]
89. Saqib, M.; Khan, S.A.; Khan, M.; Iqbal, S.; Rehman, M.M.; Kim, W.Y. Self-Powered Humidity Sensor Driven by Triboelectric Nanogenerator Composed of Bio-Wasted Peanut Skin Powder. *Polymers* **2024**, *16*, 790. [CrossRef]
90. Munirathinam, P.; Anna Mathew, A.; Shanmugasundaram, V.; Vivekananthan, V.; Purusothaman, Y.; Kim, S.-J.; Chandrasekar, A. A comprehensive review on triboelectric nanogenerators based on Real-Time applications in energy harvesting and Self-Powered sensing. *Mater. Sci. Eng. B* **2023**, *297*, 116762. [CrossRef]
91. Wang, J.; Zhang, H.; Xie, Y.; Yan, Z.; Yuan, Y.; Huang, L.; Cui, X.; Gao, M.; Su, Y.; Yang, W.; et al. Smart network node based on hybrid nanogenerator for self-powered multifunctional sensing. *Nano Energy* **2017**, *33*, 418–426. [CrossRef]
92. Ding, Z.; Nguyen, D.C.; Kim, H.; Wang, X.; Choi, K.; Lee, J.; Choi, D. Self-Powered Acceleration Sensor for Distance Prediction via Triboelectrification. *Sensors* **2024**, *24*, 4021. [CrossRef]

93. Duan, Z.; Cai, F.; Chen, Y.; Chen, T.; Lu, P. Advanced Applications of Porous Materials in Triboelectric Nanogenerator Self-Powered Sensors. *Sensors* **2024**, *24*, 3812. [CrossRef]
94. Ke, K.-H.; Chung, C.-K. High-Performance Al/PDMS TENG with Novel Complex Morphology of Two-Height Microneedles Array for High-Sensitivity Force-Sensor and Self-Powered Application. *Small* **2020**, *16*, 2001209. [CrossRef]
95. Uddin, A.; Chung, G.S. Wide-ranging impact-competent self-powered active sensor using a stacked corrugated-core sandwich-structured robust triboelectric nanogenerator. *Sens. Actuators B Chem.* **2017**, *245*, 1–10. [CrossRef]
96. Zou, Y.; Raveendran, V.; Chen, J. Wearable triboelectric nanogenerators for biomechanical energy harvesting. *Nano Energy* **2020**, *77*, 105303. [CrossRef]
97. Wang, Z.; Li, N.; Zhang, Z.; Cui, X.; Zhang, H. Hydrogel-Based Energy Harvesters and Self-Powered Sensors for Wearable Applications. *Nanoenergy Adv.* **2023**, *3*, 315–342. [CrossRef]
98. Kwak, S.S.; Yoon, H.-J.; Kim, S.-W. Textile-Based Triboelectric Nanogenerators for Self-Powered Wearable Electronics. *Adv. Funct. Mater.* **2019**, *29*, 1804533. [CrossRef]
99. Du, X.; Zhang, H.; Cao, H.; Hao, Z.; Nakashima, T.; Tanaka, Y.; Jiao, P.; Mutsuda, H. Double-Swing Spring Origami Triboelectric Nanogenerators for Self-Powered Ocean Monitoring. *Energies* **2024**, *17*, 2981. [CrossRef]
100. Sun, W.; Dong, J.; Li, W.; Gao, X.; Liu, J.; Nan, D. An Anthocyanin-Based Eco-Friendly Triboelectric Nanogenerator for pH Monitoring and Energy Harvesting. *Molecules* **2024**, *29*, 1925. [CrossRef]
101. Huang, J.; Zhang, Y.; Yu, H.; Han, G.; Cheng, W. Cellulose-Based Triboelectric Nanogenerator Prepared by Multi-Fluid Electrospinning for Respiratory Protection and Self-Powered Sensing. *Actuators* **2024**, *13*, 178. [CrossRef]
102. Ke, K.-H.; Lin, L.; Chung, C.-K. Low-cost micro-graphite doped polydimethylsiloxane composite film for enhancement of mechanical-to-electrical energy conversion with aluminum and its application. *J. Taiwan Inst. Chem. Eng.* **2022**, *135*, 104388. [CrossRef]
103. Bochu, L.; Potu, S.; Navaneeth, M.; Khanapuram, U.K.; Rajaboina, R.K.; Kodali, P. Innovative Integration of Triboelectric Nanogenerators into Signature Stamps for Energy Harvesting, Self-Powered Electronic Devices, and Smart Applications. *Eng* **2024**, *5*, 958–966. [CrossRef]
104. Zou, Y.; Sun, M.; Yan, F.; Du, T.; Xi, Z.; Li, F.; Zhu, C.; Wang, H.; Zhao, J.; Sun, P.; et al. A High-Performance Flag-Type Triboelectric Nanogenerator for Scavenging Wind Energy toward Self-Powered IoTs. *Materials* **2022**, *15*, 3696. [CrossRef]
105. Xia, K.; Wu, D.; Fu, J.; Xu, Z. A pulse controllable voltage source based on triboelectric nanogenerator. *Nano Energy* **2020**, *77*, 105112. [CrossRef]
106. Song, W.; Yin, X.; Liu, D.; Ma, W.; Zhang, M.; Li, X.; Cheng, P.; Zhang, C.; Wang, J.; Wang, Z.L. A highly elastic self-charging power system for simultaneously harvesting solar and mechanical energy. *Nano Energy* **2019**, *65*, 103997. [CrossRef]
107. Li, J.; Chen, J.; Guo, H. Triboelectric Nanogenerators for Harvesting Wind Energy: Recent Advances and Future Perspectives. *Energies* **2021**, *14*, 6949. [CrossRef]
108. Wang, S.; Liu, S.; Zhou, J.; Li, F.; Li, J.; Cao, X.; Li, Z.; Zhang, J.; Li, B.; Wang, Y.; et al. Advanced triboelectric nanogenerator with multi-mode energy harvesting and anti-impact properties for smart glove and wearable e-textile. *Nano Energy* **2020**, *78*, 105291. [CrossRef]
109. Huang, Y.-J.; Chung, C.-K. Design and Fabrication of Polymer Triboelectric Nanogenerators for Self-Powered Insole Applications. *Polymers* **2023**, *15*, 4035. [CrossRef] [PubMed]
110. Cao, M.; Chen, Y.; Sha, J.; Xu, Y.; Chen, S.; Xu, F. All-Cellulose Nanofiber-Based Sustainable Triboelectric Nanogenerators for Enhanced Energy Harvesting. *Polymers* **2024**, *16*, 1784. [CrossRef]
111. Chen, B.; Wang, Z.L. Toward a New Era of Sustainable Energy: Advanced Triboelectric Nanogenerator for Harvesting High Entropy Energy. *Small* **2022**, *18*, 2107034. [CrossRef]
112. Chung, C.-K.; Huang, Y.-J.; Wang, T.-K.; Lo, Y.-L. Fiber-Based Triboelectric Nanogenerator for Mechanical Energy Harvesting and Its Application to a Human–Machine Interface. *Sensors* **2022**, *22*, 9632. [CrossRef]
113. Cheng, T.; Cao, K.; Jing, Y.; Wang, H.; Wu, Y. Transparent and Efficient Wood-Based Triboelectric Nanogenerators for Energy Harvesting and Self-Powered Sensing. *Polymers* **2024**, *16*, 1208. [CrossRef]
114. Chung, C.K.; Ke, K.H. High contact surface area enhanced Al/PDMS triboelectric nanogenerator using novel overlapped microneedle arrays and its application to lighting and self-powered devices. *Appl. Surf. Sci.* **2020**, *508*, 145310. [CrossRef]
115. Cheng, K.; Wallaert, S.; Ardebili, H.; Karim, A. Advanced triboelectric nanogenerators based on low-dimension carbon materials: A review. *Carbon* **2022**, *194*, 81–103. [CrossRef]
116. Yu, Y.; Wang, X. Chemical modification of polymer surfaces for advanced triboelectric nanogenerator development. *Extrem. Mech. Lett.* **2016**, *9*, 514–530. [CrossRef]
117. Xuan, N.; Song, C.; Cheng, G.; Du, Z. Advanced triboelectric nanogenerator based self-powered electrochemical system. *Chem. Eng. J.* **2024**, *481*, 148640. [CrossRef]
118. Yang, Y.; Lin, L.; Zhang, Y.; Jing, Q.; Hou, T.C.; Wang, Z.L. Self-powered magnetic sensor based on a triboelectric nanogenerator. *ACS Nano* **2012**, *6*, 10378–10383. [CrossRef]
119. Xie, L.; Zhai, N.; Liu, Y.; Wen, Z.; Sun, X. Hybrid Triboelectric Nanogenerators: From Energy Complementation to Integration. *Research* **2021**, *2021*, 9143762. [CrossRef]
120. Chen, X.; Ren, Z.; Han, M.; Wan, J.; Zhang, H. Hybrid energy cells based on triboelectric nanogenerator: From principle to system. *Nano Energy* **2020**, *75*, 104980. [CrossRef]

121. Shi, Q.; Sun, Z.; Zhang, Z.; Lee, C. Triboelectric Nanogenerators and Hybridized Systems for Enabling Next-Generation IoT Applications. *Research* **2021**, *2021*, 6849171. [CrossRef]
122. Shao, H.; Wen, Z.; Cheng, P.; Sun, N.; Shen, Q.; Zhou, C.; Peng, M.; Yang, Y.; Xie, X.; Sun, X. Multifunctional power unit by hybridizing contact-separate triboelectric nanogenerator, electromagnetic generator and solar cell for harvesting blue energy. *Nano Energy* **2017**, *39*, 608–615. [CrossRef]
123. Sabishchenko, O.; Rebilas, R.; Sczygiol, N.; Urbański, M. Ukraine Energy Sector Management Using Hybrid Renewable Energy Systems. *Energies* **2020**, *13*, 1776. [CrossRef]
124. Wang, Y.; Wang, N.; Cao, X. From Triboelectric Nanogenerator to Hybrid Energy Harvesters: A Review on the Integration Strategy toward High Efficiency and Multifunctionality. *Materials* **2023**, *16*, 6405. [CrossRef]
125. Ali, S.; Jang, C.-M. Optimum Design of Hybrid Renewable Energy System for Sustainable Energy Supply to a Remote Island. *Sustainability* **2020**, *12*, 1280. [CrossRef]
126. Ali, S.; Lee, S.-M.; Jang, C.-M. Statistical analysis of wind characteristics using Weibull and Rayleigh distributions in Deokjeok-do Island—Incheon, South Korea. *Renew. Energy* **2018**, *123*, 652–663. [CrossRef]
127. Al-Rawashdeh, H.; Al-Khashman, O.A.; Al Bdour, J.T.; Gomaa, M.R.; Rezk, H.; Marshli, A.; Arrfou, L.M.; Louzazni, M. Performance Analysis of a Hybrid Renewable-Energy System for Green Buildings to Improve Efficiency and Reduce GHG Emissions with Multiple Scenarios. *Sustainability* **2023**, *15*, 7529. [CrossRef]
128. Singh, G.K. Solar power generation by PV (photovoltaic) technology: A review. *Energy* **2013**, *53*, 1–13. [CrossRef]
129. Gontean, A.; Lica, S.; Bularka, S.; Szabo, R.; Lascu, D. A Novel High Accuracy PV Cell Model Including Self Heating and Parameter Variation. *Energies* **2018**, *11*, 36. [CrossRef]
130. Cho, Y.; Lee, S.; Hong, J.; Pak, S.; Hou, B.; Lee, Y.-W.; Jang, J.E.; Im, H.; Sohn, J.I.; Cha, S.; et al. Sustainable hybrid energy harvester based on air stable quantum dot solar cells and triboelectric nanogenerator. *J. Mater. Chem. A* **2018**, *6*, 12440–12446. [CrossRef]
131. Zhao, C.; Zhang, Q.; Zhang, W.; Du, X.; Zhang, Y.; Gong, S.; Ren, K.; Sun, Q.; Wang, Z.L. Hybrid piezo/triboelectric nanogenerator for highly efficient and stable rotation energy harvesting. *Nano Energy* **2019**, *57*, 440–449. [CrossRef]
132. Li, H.; Zhang, X.; Zhao, L.; Jiang, D.; Xu, L.; Liu, Z.; Wu, Y.; Hu, K.; Zhang, M.-R.; Wang, J.; et al. A Hybrid Biofuel and Triboelectric Nanogenerator for Bioenergy Harvesting. *Nano-Micro Lett.* **2020**, *12*, 50. [CrossRef]
133. Quan, T.; Wu, Y.; Yang, Y. Hybrid electromagnetic–triboelectric nanogenerator for harvesting vibration energy. *Nano Res.* **2015**, *8*, 3272–3280. [CrossRef]
134. Quan, T.; Yang, Y. Fully enclosed hybrid electromagnetic–triboelectric nanogenerator to scavenge vibrational energy. *Nano Res.* **2016**, *9*, 2226–2233. [CrossRef]
135. Shezan, S.K.A.; Al-Mamoon, A.; Ping, H.W. Performance investigation of an advanced hybrid renewable energy system in indonesia. *Environ. Prog. Sustain. Energy* **2018**, *37*, 1424–1432. [CrossRef]
136. Ding, Z.; Wu, W.; Leung, M. Advanced/hybrid thermal energy storage technology: Material, cycle, system and perspective. *Renew. Sustain. Energy Rev.* **2021**, *145*, 111088. [CrossRef]
137. Atawi, I.E.; Al-Shetwi, A.Q.; Magableh, A.M.; Albalawi, O.H. Recent Advances in Hybrid Energy Storage System Integrated Renewable Power Generation: Configuration, Control, Applications, and Future Directions. *Batteries* **2023**, *9*, 29. [CrossRef]
138. Meshram, S.; Agnihotri, G.; Gupta, S. Advanced photovoltaic/hydro hybrid renewable energy system for remote areas. *J. Renew. Sustain. Energy* **2014**, *6*, 013140. [CrossRef]
139. Malinowski, M.; Milczarek, A.; Kot, R.; Goryca, Z.; Szuster, J.T. Optimized Energy-Conversion Systems for Small Wind Turbines: Renewable energy sources in modern distributed power generation systems. *IEEE Power Electron. Mag.* **2015**, *2*, 16–30. [CrossRef]
140. Sheng, W. Wave energy conversion and hydrodynamics modelling technologies: A review. *Renew. Sustain. Energy Rev.* **2019**, *109*, 482–498. [CrossRef]
141. Hung, Y.-H.; Yang, F.-C. Northern Lights: Prospecting Efficiency in Europe’s Renewable Energy Sector. *Processes* **2024**, *12*, 618. [CrossRef]
142. Green, M.A.; Bremner, S.P. Energy conversion approaches and materials for high-efficiency photovoltaics. *Nat. Mater.* **2017**, *16*, 23–34. [CrossRef]
143. Cullen, J.M.; Allwood, J.M. Theoretical efficiency limits for energy conversion devices. *Energy* **2010**, *35*, 2059–2069. [CrossRef]
144. Yabin, L.; Sodano, H.A. Structural Effects and Energy Conversion Efficiency of Power Harvesting. *J. Intell. Mater. Syst. Struct.* **2008**, *20*, 505–514. [CrossRef]
145. Zhang, Z.; Wu, T.; Sun, E.; Chen, Y.; Wang, N. Ferroelectric Material in Triboelectric Nanogenerator. *Materials* **2024**, *17*, 2834. [CrossRef] [PubMed]
146. Melis, A. Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency. *Plant Sci.* **2009**, *177*, 272–280. [CrossRef]
147. Brown, A.S.; Green, M.A. Impurity photovoltaic effect: Fundamental energy conversion efficiency limits. *J. Appl. Phys.* **2002**, *92*, 1329–1336. [CrossRef]
148. Khan, N.S.; Kazmi, S.A.; Anwar, M.; Mughal, S.U.; Ullah, K.; Rathi, M.K.; Salal, A. Enhancing Renewable Energy Integration in Developing Countries: A Policy-Oriented Analysis of Net Metering in Pakistan Amid Economic Challenges. *Sustainability* **2024**, *16*, 6034. [CrossRef]
149. Toplicean, I.-M.; Datcu, A.-D. An Overview on Bioeconomy in Agricultural Sector, Biomass Production, Recycling Methods, and Circular Economy Considerations. *Agriculture* **2024**, *14*, 1143. [CrossRef]

150. Wu, B.; Li, S.; Liu, Y.; Smedley, K.M. A New Hybrid Boosting Converter for Renewable Energy Applications. *IEEE Trans. Power Electron.* **2016**, *31*, 1203–1215. [CrossRef]
151. Anoune, K.; Bouya, M.; Ghazouani, M.; Astito, A.; Abdellah, A.B. Hybrid renewable energy system to maximize the electrical power production. In Proceedings of the 2016 International Renewable and Sustainable Energy Conference (IRSEC), Marrakech, Morocco, 14–17 November 2016; pp. 533–539.
152. Zaitsev, R.V.; Kirichenko, M.V.; Khrypunov, G.S.; Prokopenko, D.S.; Zaitseva, L.V. Development of hybrid solar generating module for high-efficiency solar energy station. In Proceedings of the 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kyiv, Ukraine, 29 May–2 June 2017; pp. 360–364.
153. Erdinc, O.; Uzunoglu, M. Optimum design of hybrid renewable energy systems: Overview of different approaches. *Renew. Sustain. Energy Rev.* **2012**, *16*, 1412–1425. [CrossRef]
154. Bocklisch, T. Hybrid energy storage approach for renewable energy applications. *J. Energy Storage* **2016**, *8*, 311–319. [CrossRef]
155. Marchenko, O.V.; Solomin, S.V. System studies for analyzing the efficiency of renewable energy sources. *Therm. Eng.* **2010**, *57*, 919–924. [CrossRef]
156. Sharafi, M.; ElMekkawy, T.Y.; Bibeau, E.L. Optimal design of hybrid renewable energy systems in buildings with low to high renewable energy ratio. *Renew. Energy* **2015**, *83*, 1026–1042. [CrossRef]
157. Marchenko, O.V.; Solomin, S.V. Efficiency Assessment of Renewable Energy Sources. *E3S Web Conf.* **2019**, *114*, 05001. [CrossRef]
158. Cao, R.; Zhou, T.; Wang, B.; Yin, Y.; Yuan, Z.; Li, C.; Wang, Z.L. Rotating-Sleeve Triboelectric–Electromagnetic Hybrid Nanogenerator for High Efficiency of Harvesting Mechanical Energy. *ACS Nano* **2017**, *11*, 8370–8378. [CrossRef]
159. Zhao, L.; Duan, J.; Liu, L.; Wang, J.; Duan, Y.; Vaillant-Roca, L.; Yang, X.; Tang, Q. Boosting power conversion efficiency by hybrid triboelectric nanogenerator/silicon tandem solar cell toward rain energy harvesting. *Nano Energy* **2021**, *82*, 105773. [CrossRef]
160. Liu, T.; Zheng, Y.; Xu, Y.; Liu, X.; Wang, C.; Yu, L.; Fahlman, M.; Li, X.; Murto, P.; Chen, J.; et al. Semitransparent polymer solar cell/triboelectric nanogenerator hybrid systems: Synergistic solar and raindrop energy conversion for window-integrated applications. *Nano Energy* **2022**, *103*, 107776. [CrossRef]
161. Zhang, J.; He, Y.; Boyer, C.; Kalantar-Zadeh, K.; Peng, S.; Chu, D.; Wang, C.H. Recent developments of hybrid piezo–triboelectric nanogenerators for flexible sensors and energy harvesters. *Nanoscale Adv.* **2021**, *3*, 5465–5486. [CrossRef]
162. Wang, Q.; Yu, G.; Lou, Y.; Li, M.; Hu, J.; Li, J.; Cui, W.; Yu, A.; Zhai, J. Elastic Self-Recovering Hybrid Nanogenerator for Water Wave Energy Harvesting and Marine Environmental Monitoring. *Sensors* **2024**, *24*, 3770. [CrossRef]
163. Ebrahimi, A.; Attar, S.; Farhang-Moghaddam, B. A multi-objective decision model for residential building energy optimization based on hybrid renewable energy systems. *Int. J. Green Energy* **2021**, *18*, 775–792. [CrossRef]
164. Fan, S.; Wang, X.; Cao, S.; Wang, Y.; Zhang, Y.; Liu, B. A novel model to determine the relationship between dust concentration and energy conversion efficiency of photovoltaic (PV) panels. *Energy* **2022**, *252*, 123927. [CrossRef]
165. Contreras, M.A.; Mansfield, L.M.; Egaas, B.; Li, J.; Romero, M.; Noufi, R.; Rudiger-Voigt, E.; Mannstadt, W. Wide bandgap Cu(In,Ga)Se₂ solar cells with improved energy conversion efficiency. *Prog. Photovolt. Res. Appl.* **2012**, *20*, 843–850. [CrossRef]
166. Chong, W.T.; Naghavi, M.S.; Poh, S.C.; Mahlia, T.M.I.; Pan, K.C. Techno-economic analysis of a wind–solar hybrid renewable energy system with rainwater collection feature for urban high-rise application. *Appl. Energy* **2011**, *88*, 4067–4077. [CrossRef]
167. Zemite, L.; Kozadajevs, J.; Jansons, L.; Bode, I.; Dzelzitis, E.; Palkova, K. Integrating Renewable Energy Solutions in Small-Scale Industrial Facilities. *Energies* **2024**, *17*, 2792. [CrossRef]
168. Shafiullah, G.M. Hybrid renewable energy integration (HREI) system for subtropical climate in Central Queensland, Australia. *Renew. Energy* **2016**, *96*, 1034–1053. [CrossRef]
169. Suleman, F.; Dincer, I.; Agelin-Chaab, M. Development of an integrated renewable energy system for multigeneration. *Energy* **2014**, *78*, 196–204. [CrossRef]
170. Ahmadi, P.; Dincer, I.; Rosen, M.A. Development and assessment of an integrated biomass-based multi-generation energy system. *Energy* **2013**, *56*, 155–166. [CrossRef]
171. Ratlamwala, T.A.H.; Dincer, I. Development of a geothermal based integrated system for building multigenerational needs. *Energy Build.* **2013**, *62*, 496–506. [CrossRef]
172. Kaygusuz, A.; Keles, C.; Alagoz, B.B.; Karabiber, A. Renewable energy integration for smart sites. *Energy Build.* **2013**, *64*, 456–462. [CrossRef]
173. Toghiani, M.; Saadat, A. From challenge to opportunity: Enhancing oil refinery plants with sustainable hybrid renewable energy integration. *Energy Convers. Manag.* **2024**, *305*, 118254. [CrossRef]
174. Eriksson, E.L.V.; Gray, E.M. Optimization and integration of hybrid renewable energy hydrogen fuel cell energy systems—A critical review. *Appl. Energy* **2017**, *202*, 348–364. [CrossRef]
175. Bartolucci, L.; Cordiner, S.; Mulone, V.; Rocco, V.; Rossi, J.L. Hybrid renewable energy systems for renewable integration in microgrids: Influence of sizing on performance. *Energy* **2018**, *152*, 744–758. [CrossRef]
176. Arsal, A.Z.; Hannan, M.A.; Al-Shetwi, A.Q.; Mansur, M.; Muttaqi, K.M.; Dong, Z.Y.; Blaabjerg, F. Hydrogen energy storage integrated hybrid renewable energy systems: A review analysis for future research directions. *Int. J. Hydrogen Energy* **2022**, *47*, 17285–17312. [CrossRef]
177. Mellouk, L.; Ghazi, M.; Aaroud, A.; Boulmalf, M.; Benhaddou, D.; Zine-Dine, K. Design and energy management optimization for hybrid renewable energy system- case study: Laayoune region. *Renew. Energy* **2019**, *139*, 621–634. [CrossRef]

178. Talaat, M.; Elkholy, M.H.; Alblawi, A.; Said, T. Artificial intelligence applications for microgrids integration and management of hybrid renewable energy sources. *Artif. Intell. Rev.* **2023**, *56*, 10557–10611. [CrossRef]
179. Alhawsawi, E.Y.; Habbi, H.M.D.; Hawsawi, M.; Zohdy, M.A. Optimal Design and Operation of Hybrid Renewable Energy Systems for Oakland University. *Energies* **2023**, *16*, 5830. [CrossRef]
180. Mazibuko, N.; Akindeji, K.T.; Moloi, K. A Review on the Impact of Transmission Line Compensation and RES Integration on Protection Schemes. *Energies* **2024**, *17*, 3433. [CrossRef]
181. Araújo, I.; Nunes, L.J.R.; Vilas, D.P.; Curado, A. Integrating Renewable Energy Produced by a Library Building on a University Campus in a Scenario of Collective Self-Consumption. *Energies* **2024**, *17*, 3405. [CrossRef]
182. Jouili, K.; Jouili, M.; Mohammad, A.; Babqi, A.J.; Belhadj, W. Neural Network Energy Management-Based Nonlinear Control of a DC Micro-Grid with Integrating Renewable Energies. *Energies* **2024**, *17*, 3345. [CrossRef]
183. Chung, J.; Song, M.; Chung, S.-H.; Choi, W.; Lee, S.; Lin, Z.-H.; Hong, J.; Lee, S. Triangulated Cylinder Origami-Based Piezoelectric/Triboelectric Hybrid Generator to Harvest Coupled Axial and Rotational Motion. *Research* **2021**, *2021*, 7248579. [CrossRef]
184. Matin Nazar, A.; Egbe, K.-J.I.; Jiao, P. Hybrid Piezoelectric and Triboelectric Nanogenerators for Energy Harvesting and Walking Sensing. *Energy Technol.* **2022**, *10*, 2200063. [CrossRef]
185. Vivekananthan, V.; Kim, W.J.; Alluri, N.R.; Purusothaman, Y.; Abisegapriyan, K.S.; Kim, S.-J. A sliding mode contact electrification based triboelectric-electromagnetic hybrid generator for small-scale biomechanical energy harvesting. *Micro Nano Syst. Lett.* **2019**, *7*, 14. [CrossRef]
186. Islam, E.; Abdullah, A.M.; Chowdhury, A.R.; Tasnim, F.; Martinez, M.; Olivares, C.; Lozano, K.; Uddin, M.J. Electromagnetic-triboelectric-hybrid energy tile for biomechanical green energy harvesting. *Nano Energy* **2020**, *77*, 105250. [CrossRef]
187. Zheng, J.; Cao, Z.; Han, C.; Wei, X.; Wang, L.; Wu, Z. A Hybrid Triboelectric-Electromagnetic Nanogenerator Based on Arm Swing Energy Harvesting. *Nanoenergy Adv.* **2023**, *3*, 126–137. [CrossRef]
188. Chen, Y.; Jie, Y.; Zhu, J.; Lu, Q.; Cheng, Y.; Cao, X.; Wang, Z.L. Hybridized triboelectric-electromagnetic nanogenerators and solar cell for energy harvesting and wireless power transmission. *Nano Res.* **2022**, *15*, 2069–2076. [CrossRef]
189. Wang, J.; Wen, Z.; Zi, Y.; Lin, L.; Wu, C.; Guo, H.; Xi, Y.; Xu, Y.; Wang, Z.L. Self-Powered Electrochemical Synthesis of Polypyrrole from the Pulsed Output of a Triboelectric Nanogenerator as a Sustainable Energy System. *Adv. Funct. Mater.* **2016**, *26*, 3542–3548. [CrossRef]
190. Luo, J.; Wang, Z.L. Recent advances in triboelectric nanogenerator based self-charging power systems. *Energy Storage Mater.* **2019**, *23*, 617–628. [CrossRef]
191. Dong, K.; Wang, Z.L. Self-charging power textiles integrating energy harvesting triboelectric nanogenerators with energy storage batteries/supercapacitors. *J. Semicond.* **2021**, *42*, 101601. [CrossRef]
192. Zhang, Q.; Zhang, Z.; Liang, Q.; Gao, F.; Yi, F.; Ma, M.; Liao, Q.; Kang, Z.; Zhang, Y. Green hybrid power system based on triboelectric nanogenerator for wearable/portable electronics. *Nano Energy* **2019**, *55*, 151–163. [CrossRef]
193. Khan, A.A.; Mahmud, A.; Ban, D. Evolution from Single to Hybrid Nanogenerator: A Contemporary Review on Multimode Energy Harvesting for Self-Powered Electronics. *IEEE Trans. Nanotechnol.* **2019**, *18*, 21–36. [CrossRef]
194. Zhang, Y.; Gao, X.; Zhang, Y.; Gui, J.; Sun, C.; Zheng, H.; Guo, S. High-efficiency self-charging power systems based on performance-enhanced hybrid nanogenerators and asymmetric supercapacitors for outdoor search and rescue. *Nano Energy* **2022**, *92*, 106788. [CrossRef]
195. Zhang, Q.; Liang, Q.; Nandakumar, D.K.; Qu, H.; Shi, Q.; Alzakia, F.I.; Tay, D.J.J.; Yang, L.; Zhang, X.; Suresh, L.; et al. Shadow enhanced self-charging power system for wave and solar energy harvesting from the ocean. *Nat. Commun.* **2021**, *12*, 616.
196. Rana, S.M.S.; Salauddin, M.; Sharifuzzaman, M.; Lee, S.H.; Shin, Y.D.; Song, H.; Jeong, S.H.; Bhatta, T.; Shrestha, K.; Park, J.Y. Ultrahigh-Output Triboelectric and Electromagnetic Hybrid Generator for Self-Powered Smart Electronics and Biomedical Applications. *Adv. Energy Mater.* **2022**, *12*, 2202238. [CrossRef]
197. Rahman, M.T.; Rana, S.M.S.; Salauddin, M.; Maharjan, P.; Bhatta, T.; Park, J.Y. Biomechanical Energy-Driven Hybridized Generator as a Universal Portable Power Source for Smart/Wearable Electronics. *Adv. Energy Mater.* **2020**, *10*, 1903663. [CrossRef]
198. Khan, A.; Joshi, R.; Sharma, M.K.; Ganguly, A.; Parashar, P.; Wang, T.-W.; Lee, S.; Kao, F.-C.; Lin, Z.-H. Piezoelectric and triboelectric nanogenerators: Promising technologies for self-powered implantable biomedical devices. *Nano Energy* **2024**, *119*, 109051. [CrossRef]
199. Zou, Y.; Liao, J.; Ouyang, H.; Jiang, D.; Zhao, C.; Li, Z.; Qu, X.; Liu, Z.; Fan, Y.; Shi, B.; et al. A flexible self-arched biosensor based on combination of piezoelectric and triboelectric effects. *Appl. Mater. Today* **2020**, *20*, 100699. [CrossRef]
200. Zheng, Q.; Shi, B.; Li, Z.; Wang, Z.L. Recent Progress on Piezoelectric and Triboelectric Energy Harvesters in Biomedical Systems. *Adv. Sci.* **2017**, *4*, 1700029. [CrossRef]
201. Zhu, M.; Shi, Q.; He, T.; Yi, Z.; Ma, Y.; Yang, B.; Chen, T.; Lee, C. Self-Powered and Self-Functional Cotton Sock Using Piezoelectric and Triboelectric Hybrid Mechanism for Healthcare and Sports Monitoring. *ACS Nano* **2019**, *13*, 1940–1952. [CrossRef]
202. Qian, J.; Jing, X. Wind-driven hybridized triboelectric-electromagnetic nanogenerator and solar cell as a sustainable power unit for self-powered natural disaster monitoring sensor networks. *Nano Energy* **2018**, *52*, 78–87. [CrossRef]
203. Chen, J.; Gong, S.; Gong, T.; Yang, X.; Guo, H. Stackable Direct Current Triboelectric-Electromagnetic Hybrid Nanogenerator for Self-Powered Air Purification and Quality Monitoring. *Adv. Energy Mater.* **2023**, *13*, 2203689. [CrossRef]

204. Zhang, K.; Wang, X.; Yang, Y.; Wang, Z.L. Hybridized Electromagnetic–Triboelectric Nanogenerator for Scavenging Biomechanical Energy for Sustainably Powering Wearable Electronics. *ACS Nano* **2015**, *9*, 3521–3529. [CrossRef]
205. Wen, Z.; Yeh, M.-H.; Guo, H.; Wang, J.; Zi, Y.; Xu, W.; Deng, J.; Zhu, L.; Wang, X.; Hu, C.; et al. Self-powered textile for wearable electronics by hybridizing fiber-shaped nanogenerators, solar cells, and supercapacitors. *Sci. Adv.* **2016**, *2*, e1600097. [CrossRef]
206. Rana, S.M.S.; Rahman, M.T.; Salauddin, M.; Maharjan, P.; Bhatta, T.; Cho, H.; Park, J.Y. A human-machine interactive hybridized biomechanical nanogenerator as a self-sustainable power source for multifunctional smart electronics applications. *Nano Energy* **2020**, *76*, 105025. [CrossRef]
207. Rayegani, A.; Saberian, M.; Delshad, Z.; Liang, J.; Sadiq, M.; Nazar, A.M.; Mohsan, S.A.; Khan, M.A. Recent Advances in Self-Powered Wearable Sensors Based on Piezoelectric and Triboelectric Nanogenerators. *Biosensors* **2023**, *13*, 37. [CrossRef] [PubMed]
208. Mariello, M. Recent Advances on Hybrid Piezo-Triboelectric Bio-Nanogenerators: Materials, Architectures and Circuitry. *Nanoenergy Adv.* **2022**, *2*, 64–109. [CrossRef]
209. Wang, P.; Pan, L.; Wang, J.; Xu, M.; Dai, G.; Zou, H.; Dong, K.; Wang, Z.L. An Ultra-Low-Friction Triboelectric–Electromagnetic Hybrid Nanogenerator for Rotation Energy Harvesting and Self-Powered Wind Speed Sensor. *ACS Nano* **2018**, *12*, 9433–9440. [CrossRef] [PubMed]
210. Lee, T.; Kim, I.; Kim, D. Flexible Hybrid Nanogenerator for Self-Powered Weather and Healthcare Monitoring Sensor. *Adv. Electron. Mater.* **2021**, *7*, 2100785. [CrossRef]
211. Rasel, M.S.; Maharjan, P.; Park, J.Y. Hand clapping inspired integrated multilayer hybrid nanogenerator as a wearable and universal power source for portable electronics. *Nano Energy* **2019**, *63*, 103816. [CrossRef]
212. Jian, G.; Yang, N.; Zhu, S.; Meng, Q.; Ouyang, C. A Mousepad Triboelectric-Piezoelectric Hybrid Nanogenerator (TPHNG) for Self-Powered Computer User Behavior Monitoring Sensors and Biomechanical Energy Harvesting. *Polymers* **2023**, *15*, 2462. [CrossRef]
213. Vidadili, N.; Suleymanov, E.; Bulut, C.; Mahmudlu, C. Transition to renewable energy and sustainable energy development in Azerbaijan. *Renew. Sustain. Energy Rev.* **2017**, *80*, 1153–1161. [CrossRef]
214. Liu, C.-j.; Burghaus, U.; Besenbacher, F.; Wang, Z.L. Preparation and Characterization of Nanomaterials for Sustainable Energy Production. *ACS Nano* **2010**, *4*, 5517–5526. [CrossRef]
215. Benli, H. Potential of renewable energy in electrical energy production and sustainable energy development of Turkey: Performance and policies. *Renew. Energy* **2013**, *50*, 33–46. [CrossRef]
216. Shi, Y.; Ge, Y.; Chang, J.; Shao, H.; Tang, Y. Garden waste biomass for renewable and sustainable energy production in China: Potential, challenges and development. *Renew. Sustain. Energy Rev.* **2013**, *22*, 432–437. [CrossRef]
217. Margeta, J.; Glasnovic, Z. Theoretical settings of photovoltaic-hydro energy system for sustainable energy production. *Sol. Energy* **2012**, *86*, 972–982. [CrossRef]
218. Amran, Y.H.A.; Amran, Y.H.M.; Alyousef, R.; Alabduljabbar, H. Renewable and sustainable energy production in Saudi Arabia according to Saudi Vision 2030; Current status and future prospects. *J. Clean. Prod.* **2020**, *247*, 119602. [CrossRef]
219. Gunnarsdottir, I.; Davidsdottir, B.; Worrell, E.; Sigurgeirsdottir, S. Sustainable energy development: History of the concept and emerging themes. *Renew. Sustain. Energy Rev.* **2021**, *141*, 110770. [CrossRef]
220. Ur Rehman, S.A.; Cai, Y.; Siyal, Z.A.; Mirjat, N.H.; Fazal, R.; Kashif, S.U. Cleaner and Sustainable Energy Production in Pakistan: Lessons Learnt from the Pak-TIMES Model. *Energies* **2020**, *13*, 108. [CrossRef]
221. Yuksel, I.; Kaygusuz, K. Renewable energy sources for clean and sustainable energy policies in Turkey. *Renew. Sustain. Energy Rev.* **2011**, *15*, 4132–4144. [CrossRef]
222. Giwa, A.; Alabi, A.; Yusuf, A.; Olukan, T. A comprehensive review on biomass and solar energy for sustainable energy generation in Nigeria. *Renew. Sustain. Energy Rev.* **2017**, *69*, 620–641. [CrossRef]
223. Lian, J.; Zhang, Y.; Ma, C.; Yang, Y.; Chaima, E. A review on recent sizing methodologies of hybrid renewable energy systems. *Energy Convers. Manag.* **2019**, *199*, 112027. [CrossRef]
224. Mohammed, Y.S.; Mustafa, M.W.; Bashir, N. Hybrid renewable energy systems for off-grid electric power: Review of substantial issues. *Renew. Sustain. Energy Rev.* **2014**, *35*, 527–539. [CrossRef]
225. Jamal, T.; Salehin, S. 5—Hybrid renewable energy sources power systems. In *Hybrid Renewable Energy Systems and Microgrids*; Kabalci, E., Ed.; Academic Press: Cambridge, MA, USA, 2021; pp. 179–214.
226. Li, S.; Zhang, L.; Liu, X.; Zhu, C. Collaborative operation optimization and benefit-sharing strategy of rural hybrid renewable energy systems based on a circular economy: A Nash bargaining model. *Energy Convers. Manag.* **2023**, *283*, 116918. [CrossRef]
227. Lemence, A.L.G.; Tamayao, M.-A.M. Energy consumption profile estimation and benefits of hybrid solar energy system adoption for rural health units in the Philippines. *Renew. Energy* **2021**, *178*, 651–668. [CrossRef]
228. Kim, M.-H.; Kim, D.; Heo, J.; Lee, D.-W. Techno-economic analysis of hybrid renewable energy system with solar district heating for net zero energy community. *Energy* **2019**, *187*, 115916. [CrossRef]
229. Roy, P.; He, J.; Zhao, T.; Singh, Y.V. Recent Advances of Wind-Solar Hybrid Renewable Energy Systems for Power Generation: A Review. *IEEE Open J. Ind. Electron. Soc.* **2022**, *3*, 81–104. [CrossRef]
230. Sawle, Y.; Gupta, S.C.; Bohre, A.K. Review of hybrid renewable energy systems with comparative analysis of off-grid hybrid system. *Renew. Sustain. Energy Rev.* **2018**, *81*, 2217–2235. [CrossRef]

231. Sorrenti, I.; Harild Rasmussen, T.B.; You, S.; Wu, Q. The role of power-to-X in hybrid renewable energy systems: A comprehensive review. *Renew. Sustain. Energy Rev.* **2022**, *165*, 112380. [CrossRef]
232. Ma, Z.; Dong, F.; Wang, J.; Zhou, Y.; Feng, Y. Optimal design of a novel hybrid renewable energy CCHP system considering long and short-term benefits. *Renew. Energy* **2023**, *206*, 72–85. [CrossRef]
233. Kayima, P.; Semakula, H.M.; Wasswa, H.; Mugagga, F.; Mukwaya, P.I. Analysis of the socio-economic benefits of on-grid hybrid solar energy system on Bugala island in Uganda. *Energy Sustain. Dev.* **2023**, *77*, 101332. [CrossRef]
234. Babatunde, O.M.; Munda, J.L.; Hamam, Y. A Comprehensive State-of-the-Art Survey on Hybrid Renewable Energy System Operations and Planning. *IEEE Access* **2020**, *8*, 75313–75346. [CrossRef]
235. Mohammadi, K.; Khanmohammadi, S.; Khorasanizadeh, H.; Powell, K. A comprehensive review of solar only and hybrid solar driven multigeneration systems: Classifications, benefits, design and prospective. *Appl. Energy* **2020**, *268*, 114940. [CrossRef]
236. Ebhota, W.S.; Jen, T.-C. Fossil Fuels Environmental Challenges and the Role of Solar Photovoltaic Technology Advances in Fast Tracking Hybrid Renewable Energy System. *Int. J. Precis. Eng. Manuf.-Green Technol.* **2020**, *7*, 97–117. [CrossRef]
237. Islam, S.; Dincer, I.; Yilbas, B.S. A novel renewable energy-based integrated system with thermoelectric generators for a net-zero energy house. *Int. J. Energy Res.* **2020**, *44*, 3458–3477. [CrossRef]
238. Javed, M.S.; Ma, T.; Jurasz, J.; Ahmed, S.; Mikulik, J. Performance comparison of heuristic algorithms for optimization of hybrid off-grid renewable energy systems. *Energy* **2020**, *210*, 118599. [CrossRef]
239. Neves, D.; Silva, C.A.; Connors, S. Design and implementation of hybrid renewable energy systems on micro-communities: A review on case studies. *Renew. Sustain. Energy Rev.* **2014**, *31*, 935–946. [CrossRef]
240. Babatunde, O.M.; Munda, J.L.; Hamam, Y. Selection of a Hybrid Renewable Energy Systems for a Low-Income Household. *Sustainability* **2019**, *11*, 4282. [CrossRef]
241. Li, F.; Yin, A.; Zhou, Y.; Liu, T.; Liu, Q.; Ruan, W.; Bu, L. Stiffness Modulation in Flexible Rotational Triboelectric Nanogenerators for Dual Enhancement of Power and Reliability. *Nanomaterials* **2024**, *14*, 380. [CrossRef]
242. Xu, Z.; Chang, Y.; Zhu, Z. A Triboelectric Nanogenerator Based on Bamboo Leaf for Biomechanical Energy Harvesting and Self-Powered Touch Sensing. *Electronics* **2024**, *13*, 766. [CrossRef]
243. Freire, A.L.; Lima, L.R.; Candido, I.C.M.; Silva, L.G.; Ribeiro, S.J.L.; Carrilho, E.; Oliveira, T.L.; de Oliveira, L.F.C.; Barud, H.S.; de Oliveira, H.P. Metal-Free, Bio-Triboelectric Nanogenerator Based on a Single Electrode of Bacterial Cellulose Modified with Carbon Black. *Nanoenergy Adv.* **2024**, *4*, 110–121. [CrossRef]
244. Sripadmanabhan Indira, S.; Aravind Vaithilingam, C.; Oruganti, K.S.; Mohd, F.; Rahman, S. Nanogenerators as a Sustainable Power Source: State of Art, Applications, and Challenges. *Nanomaterials* **2019**, *9*, 773. [CrossRef]
245. Arefin, M.A.; Islam, M.T.; Rashid, F.; Mostakim, K.; Masuk, N.I.; Islam, M.H.I. A Comprehensive Review of Nuclear-Renewable Hybrid Energy Systems: Status, Operation, Configuration, Benefit, and Feasibility. *Front. Sustain. Cities* **2021**, *3*, 723910. [CrossRef]
246. Nehrir, M.H.; Wang, C.; Strunz, K.; Aki, H.; Ramakumar, R.; Bing, J.; Miao, Z.; Salameh, Z. A Review of Hybrid Renewable/Alternative Energy Systems for Electric Power Generation: Configurations, Control, and Applications. *IEEE Trans. Sustain. Energy* **2011**, *2*, 392–403. [CrossRef]
247. León Gómez, J.C.; De León Aldaco, S.E.; Aguayo Alquicira, J. A Review of Hybrid Renewable Energy Systems: Architectures, Battery Systems, and Optimization Techniques. *Eng* **2023**, *4*, 1446–1467. [CrossRef]
248. Siddaiah, R.; Saini, R.P. A review on planning, configurations, modeling and optimization techniques of hybrid renewable energy systems for off grid applications. *Renew. Sustain. Energy Rev.* **2016**, *58*, 376–396. [CrossRef]
249. Wang, R.; Xiong, J.; He, M.-f.; Gao, L.; Wang, L. Multi-objective optimal design of hybrid renewable energy system under multiple scenarios. *Renew. Energy* **2020**, *151*, 226–237. [CrossRef]
250. Memon, S.A.; Patel, R.N. An overview of optimization techniques used for sizing of hybrid renewable energy systems. *Renew. Energy Focus* **2021**, *39*, 1–26. [CrossRef]
251. Natividad, L.E.; Benalcazar, P. Hybrid Renewable Energy Systems for Sustainable Rural Development: Perspectives and Challenges in Energy Systems Modeling. *Energies* **2023**, *16*, 1328. [CrossRef]
252. Rathod, A.A.; Subramanian, B. Scrutiny of Hybrid Renewable Energy Systems for Control, Power Management, Optimization and Sizing: Challenges and Future Possibilities. *Sustainability* **2022**, *14*, 16814. [CrossRef]
253. Vaziri Rad, M.A.; Kasaian, A.; Niu, X.; Zhang, K.; Mahian, O. Excess electricity problem in off-grid hybrid renewable energy systems: A comprehensive review from challenges to prevalent solutions. *Renew. Energy* **2023**, *212*, 538–560. [CrossRef]
254. Basit, M.A.; Dilshad, S.; Badar, R.; Sami ur Rehman, S.M. Limitations, challenges, and solution approaches in grid-connected renewable energy systems. *Int. J. Energy Res.* **2020**, *44*, 4132–4162. [CrossRef]
255. Malik, P.; Awasthi, M.; Sinha, S. Biomass-based gaseous fuel for hybrid renewable energy systems: An overview and future research opportunities. *Int. J. Energy Res.* **2021**, *45*, 3464–3494. [CrossRef]
256. Bourennani, F.; Rahnamayan, S.; Naterer, G.F. Optimal Design Methods for Hybrid Renewable Energy Systems. *Int. J. Green Energy* **2015**, *12*, 148–159. [CrossRef]
257. Khan, A.A.; Minai, A.F.; Pachauri, R.K.; Malik, H. Optimal Sizing, Control, and Management Strategies for Hybrid Renewable Energy Systems: A Comprehensive Review. *Energies* **2022**, *15*, 6249. [CrossRef]
258. Eltamaly, A.M.; Alotaibi, M.A.; Alolah, A.I.; Ahmed, M.A. IoT-Based Hybrid Renewable Energy System for Smart Campus. *Sustainability* **2021**, *13*, 8555. [CrossRef]

259. Bajpai, P.; Dash, V. Hybrid renewable energy systems for power generation in stand-alone applications: A review. *Renew. Sustain. Energy Rev.* **2012**, *16*, 2926–2939. [CrossRef]
260. Jung, J.; Villaran, M. Optimal planning and design of hybrid renewable energy systems for microgrids. *Renew. Sustain. Energy Rev.* **2017**, *75*, 180–191. [CrossRef]
261. Zhang, L.; Li, S.; Nie, Q.; Hu, Y. A two-stage benefit optimization and multi-participant benefit-sharing strategy for hybrid renewable energy systems in rural areas under carbon trading. *Renew. Energy* **2022**, *189*, 744–761. [CrossRef]
262. Khosravani, A.; Safaei, E.; Reynolds, M.; Kelly, K.E.; Powell, K.M. Challenges of reaching high renewable fractions in hybrid renewable energy systems. *Energy Rep.* **2023**, *9*, 1000–1017. [CrossRef]
263. Hassan, Q.; Algburi, S.; Sameen, A.Z.; Salman, H.M.; Jaszczur, M. A review of hybrid renewable energy systems: Solar and wind-powered solutions: Challenges, opportunities, and policy implications. *Results Eng.* **2023**, *20*, 101621. [CrossRef]
264. Falope, T.; Lao, L.; Hanak, D.; Huo, D. Hybrid energy system integration and management for solar energy: A review. *Energy Convers. Manag. X* **2024**, *21*, 100527. [CrossRef]
265. Ennemiri, N.; Berrada, A.; Emrani, A.; Abdelmajid, J.; El Mrabet, R. Optimization of an off-grid PV/biogas/battery hybrid energy system for electrification: A case study in a commercial platform in Morocco. *Energy Convers. Manag. X* **2024**, *21*, 100508. [CrossRef]
266. Allouhi, A. A hybrid PV/wind/battery energy system to assist a run-of-river micro-hydropower for clean electrification and fuelling hydrogen mobility for young population in a rural Moroccan site. *J. Clean. Prod.* **2024**, *442*, 140852. [CrossRef]
267. Ma, W.; Xue, X.; Liu, G. Techno-economic evaluation for hybrid renewable energy system: Application and merits. *Energy* **2018**, *159*, 385–409. [CrossRef]
268. Bahramara, S.; Moghaddam, M.P.; Haghifam, M.R. Optimal planning of hybrid renewable energy systems using HOMER: A review. *Renew. Sustain. Energy Rev.* **2016**, *62*, 609–620. [CrossRef]
269. Wang, X.; Palazoglu, A.; El-Farra, N.H. Operational optimization and demand response of hybrid renewable energy systems. *Appl. Energy* **2015**, *143*, 324–335. [CrossRef]
270. Mazzeo, D.; Oliveti, G.; Baglivo, C.; Congedo, P.M. Energy reliability-constrained method for the multi-objective optimization of a photovoltaic-wind hybrid system with battery storage. *Energy* **2018**, *156*, 688–708. [CrossRef]
271. Memon, S.A.; Upadhyay, D.S.; Patel, R.N. Optimal configuration of solar and wind-based hybrid renewable energy system with and without energy storage including environmental and social criteria: A case study. *J. Energy Storage* **2021**, *44*, 103446. [CrossRef]
272. Mayer, M.J.; Szilágyi, A.; Gróf, G. Environmental and economic multi-objective optimization of a household level hybrid renewable energy system by genetic algorithm. *Appl. Energy* **2020**, *269*, 115058. [CrossRef]
273. Eriksson, E.L.V.; Gray, E.M. Optimization of renewable hybrid energy systems—A multi-objective approach. *Renew. Energy* **2019**, *133*, 971–999. [CrossRef]
274. Kelechi, O.; Tokos, H. An MILP Model for the Optimization of Hybrid Renewable Energy System. In *Computer Aided Chemical Engineering*; Kravanja, Z., Bogataj, M., Eds.; Elsevier: Amsterdam, The Netherlands, 2016; pp. 2193–2198.
275. Maghami, M.R.; Mutambara, A.G.O. Challenges associated with Hybrid Energy Systems: An artificial intelligence solution. *Energy Rep.* **2023**, *9*, 924–940. [CrossRef]
276. Gusain, C.; Mohan Tripathi, M.; Nangia, U. Study of Meta-heuristic Optimization Methodologies for Design of Hybrid Renewable Energy Systems. *Therm. Sci. Eng. Prog.* **2023**, *39*, 101711. [CrossRef]
277. Rahman, M.M.; Shakeri, M.; Tiong, S.K.; Khatun, F.; Amin, N.; Pasupuleti, J.; Hasan, M.K. Prospective Methodologies in Hybrid Renewable Energy Systems for Energy Prediction Using Artificial Neural Networks. *Sustainability* **2021**, *13*, 2393. [CrossRef]
278. Kavadias, K.A.; Triantafyllou, P. Hybrid Renewable Energy Systems' Optimisation. A Review and Extended Comparison of the Most-Used Software Tools. *Energies* **2021**, *14*, 8268. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Review

Robots for the Energy Transition: A Review

Sergio Taraglio ¹, Stefano Chiesa ¹, Saverio De Vito ², Marco Paoloni ¹, Gabriele Piantadosi ², Andrea Zanela ¹ and Girolamo Di Francia ^{2,*}

¹ ENEA C.R. Casaccia, Via Anguillarese 301, 00123 Roma, Italy; sergio.taraglio@enea.it (S.T.); stefano.chiesa@enea.it (S.C.); marco.paoloni@enea.it (M.P.); andrea.zanela@enea.it (A.Z.)

² ENEA C.R. Portici, Piazzale Enrico Fermi 1, Portici, 80055 Napoli, Italy; saverio.devito@enea.it (S.D.V.); gabriele.piantadosi@enea.it (G.P.)

* Correspondence: girolamo.difranca@enea.it; Tel.: +39-081-7722277

Abstract: The energy transition relies on an increasingly massive and pervasive use of renewable energy sources, mainly photovoltaic and wind, characterized by an intrinsic degree of production uncertainty, mostly due to meteorological conditions variability that, even if accurately estimated, can hardly be kept under control. Because of this limit, continuously monitoring the operative status of each renewable energy-based power plant becomes relevant in order to timely face any other uncertainty source such as those related to the plant operation and maintenance (O&M), whose effect may become relevant in terms of the leveled cost of energy. In this frame, the use of robots, which incorporate fully automatic platforms capable of monitoring each plant and also allow effective and efficient process operation, can be considered a feasible solution. This paper carries out a review on the use of robots for the O&M of photovoltaic, wind, hydroelectric, and concentrated solar power, including robot applications for controlling power lines, whose role can in fact be considered a key complementary issue within the energy transition. It is shown that various robotic solutions have so far been proposed both by the academy and by industries and that implementing their use should be considered mandatory for the energy transition scenario.

Keywords: robot; renewable energy; operation and maintenance; OPEX; photovoltaic; wind; CSP; hydroelectric; power lines

Citation: Taraglio, S.; Chiesa, S.; De Vito, S.; Paoloni, M.; Piantadosi, G.; Zanela, A.; Di Francia, G. Robots for the Energy Transition: A Review. *Processes* **2024**, *12*, 1982. <https://doi.org/10.3390/pr12091982>

Academic Editor: Sheng Du

Received: 27 July 2024

Revised: 8 September 2024

Accepted: 12 September 2024

Published: 14 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In 2022, the global electricity demand has reached approximately 30,000 TWh, representing almost one-fifth of overall world energy consumption [1]. Even considering the best efficiency improvement in energy use, the expected energy transition evolution will lead to a more than double increase in current electric energy demand between 2025 and 2050 [2]. And, while at present about 2/3 of the electricity consumed is still produced by means of fossil sources (mainly gas and oil), the 2050 target will have to be achieved entirely using renewable energy (REN) sources and therefore, by introducing into the power grid huge amounts of photovoltaic, wind, and hydroelectric energies [1].

In this respect, it is of value to underline that in some countries where the use of renewables is more advanced, problems of the electricity grid not adapting to new scenarios have been reported [2]. From this point of view, the increasing use of photovoltaic and wind energies poses a very demanding challenge. Both are, in fact, definitely volatile energy sources (1) because of their production capacity being strongly dependent on meteorological conditions and, therefore, due to the uncertainty of these, increasing as the desired time horizon for the required production forecast moves away [3,4], and (2) because their low power density per unit of surface area compared to fossil fuels imposes ordinary and extraordinary operation and maintenance schemes, which can become very critical and economically costly if we consider the intervention times required and the fact that

these plants can be distributed over very large areas or, as in the case of wind farms, they can also be installed offshore [5,6].

Both of these problems translate into Capital expenditures, CAPEX and operative expenses, and OPEX costs increase. In the first case, the need to limit the effect of forecast errors or system malfunctions is coped with via the massive use of electricity storage systems and probably, for very high levels of renewable penetration, the adoption of conversion power-to-gas systems, which are today still in the development phase and impose the limitation of “labour-intensive” costs, presently making up around one-third of the CAPEX costs, by means of the use of automatic systems for the construction of the production plants themselves [7,8]. As regards to plants operation and maintenance costs (O&M), the most demanding effort is to make such phases increasingly automated, limiting as much as possible the “labour-intensive” costs, which today constitute the dominant part for this category of costs [9].

In both cases, robotics now plays an essential role. In the first case, standardization in industrial processes involving wind and photovoltaic energy has led to the possibility of designing and implementing automated systems for the construction of large plants for the production of renewable energy [10]. In the second case, studies on the development and field testing of robotic systems for the inspection and maintenance of these same systems are increasingly frequent [11].

In this work, we will essentially deal with this latter issue by reviewing the use of robots for the O&M of photovoltaic, wind, hydroelectric, and concentrated solar system plants, also including the robotic systems necessary for the control of power lines which are essential parts of the energy transition. This work is divided into seven sections. In addition to the Introduction, Section 2 contextualizes the issue under consideration within the broader theme of the necessary digitalisation of the energy sector in terms of renewable sources; Section 3 is structured into various sub-paragraphs each dedicated to robot applications for a particular REN (3.1 CSP, 3.2 photovoltaic, 3.3 wind, 3.4 hydroelectric, and 3.5 electricity lines); Section 4 discusses the robot machines from the point of view of their operational SW; Section 5 is dedicated to the database required for ML approaches to be implemented; Sections 6 and 7 are, respectively, devoted to the discussion of the reported data and the conclusions.

2. The Energy Sector Digitalisation

The increasing use of information and communication technology (ICT) in various social, economic, industrial, and cultural sectors is at the origin of the so-called phenomenon of digital disruption, which currently also involves more and more behaviour of individuals in every part of the world and that, according to Skog and co-authors’ analysis, has to be interpreted in terms of a creative and, at the same time, destructive process that crosses every socio-economic sector [12]. Even if in a more gradual manner, according to Lea Myllykallio in [13], this is now occurring, with visible effects, in the energy sector, where the effective strong support of the pervasive and relevant use of digital technologies can be said to have actually been accelerated by the recent increasing and large adoption of renewable energies in the energy mix of most Western countries.

According to the IEA world outlook 2022, renewable energy use worldwide increased from about 15% in 2010 to 20% in 2020 and is now expected to overcome 60% in the 2050 mix [14]. Beyond the well-known environmental benefits, the most important change related to the massive use of renewable energies for electricity production is connected to the unavoidable changes in the energy trading methods as a result of their production variability due to intrinsic and site-specific weather condition forecast uncertainty. Although this issue has greatly advanced in terms of accuracy over time, it still has an intrinsic level of volatility increasing as the time horizon increases [15,16]. This has led to the introduction of a growing level of complexity in the management of energy markets, which has, in turn, forced and increasingly obliged the involved stakeholders (TSOs, producers, distributors, etc.) to move financial transactions related to this type of energy to closer and closer time

horizons, such as the day-after market and, increasingly too, even the intraday market, with negotiations that tend to be increasingly characterized by correspondingly short time horizons of even just a few minutes [17]. It is clear that this type of financial market can only operate digitally, and it is also clear that as the penetration of renewable energy in a country's energy mix increases, this type of operational methodology will become increasingly relevant. It is important to underline that the development of blockchain technologies has been and is functional within this context, and that without which, in fact, the certified volumes of energy that transactions handle would not be able to be exchanged, especially in a market increasingly made up of prosumers [18,19].

The digitalisation of energy trading has simultaneously led to an equally rapid evolution in the application of advanced machine learning techniques both to improve the precision in renewable energy production estimates, thus allowing the strict time financial contracting constraints to be relaxed, and to improve the reliability of energy production systems in terms of preventive maintenance [20]. Both such issues significantly affect the energy trading process' reliability. In the case of ML techniques applied to forecasting, the aim is to improve precision not only by means of the use of data from satellite providers, but also taking into account data obtained from ground systems which, using appropriate training procedures, allow for significant improvements in the individual site production forecast [21]. As reported by Benti and coauthors, these techniques are characterized by a series of problems, with one of the main ones being the lack of high-quality data for the training and validation steps. Data for renewable energy forecasting are often sparse, noisy, and incomplete, which makes it difficult to build accurate models. Another challenge is the lack of transparency and interpretability in ML and DL models or the unavoidable lack of training data for new systems which poses the problem of the so-called transferring of calibration among different systems, with obvious reliability effects [22]. The ML techniques related to the optimization of the plant maintenance phase are in some ways similar and are based on very detailed knowledge of the individual renewable production plants [23,24]. These are increasingly heavily equipped with sensors that, with appropriate IoT architectures and by means of SCADA systems, allow for precise knowledge of the system state and of its evolution over time [25]. In fact, since the O&M costs of renewable plants today constitute a significant and increasing percentage of the total cost of the plants, the boost towards the use of digital technologies in the O&M field is as relevant as the trading requests in terms of supporting the diffusion of renewable digitalization [26]. From this point of view, the energy systems of countries that are implementing the transition towards non-climate-changing productions are archetypes of IoT pyramid systems, in which the lower part is made up of highly pervasive sensor arrays and the upper part is the development of automatic tools for the operation and maintenance of the systems which are the subject of this review [27].

Finally, it should be recalled that the exponential increase in information produced and exchanged in real time during the operative lifetime of a plant requires the implementation of security management techniques (cybersecurity), which are also part of the range of methodologies that are under development in terms of the digitalisation of energy systems. For example, A. Rekerahoet and coauthors have recently summarized the various cyber-attack systems on the renewable energy production plants security, some of which are extremely complex to cope with because of the high granularity of REN production systems [28].

3. Robot Applications

In the following, the use of robotics in the field of renewable energies has been divided into five different sections, namely concentrated solar power (CSP), photovoltaic (PV) energy production, wind farms, hydroelectrical plants, and finally power lines, that match five different areas of the production and dispatching of energy.

3.1. Robotics in Concentrating Solar Power (CSP)

A concentrating solar power (CSP) plant is usually composed of an array of mirror surfaces directing and concentrating the sun radiation towards a collecting device with the aim to heat a given substance to a high temperature (at least in the range of hundreds of °C) to then be subsequently used to produce electricity, usually through a steam turbine connected to an electrical power generator. Currently, the cumulative CSP capacity around the world is of about 6.5 GW, with most of the installations being in Spain and the US [29].

There are mainly four different designs:

1. the parabolic trough, where a linear parabolic reflector concentrates sunlight onto a receiver positioned along the reflector's focal line;
2. the Fresnel reflectors, where the reflector is composed of many flat mirror strips arranged to reflect sunlight onto an overhanging tube;
3. the solar tower, where an array of heliostats, which are dual-axis tracking reflectors, concentrate the radiation on a central receiver on top of a tower;
4. the Dish Stirling, composed of a parabolic reflector concentrating light onto a receiver in the focal point of the parabola, similarly to a radio telescope.

From the robotics point of view, it is possible to hypothesize several tasks that may be operated by service robots in these plants:

1. system deployment, i.e., the physical placing of the mirrors;
2. mirror cleaning, in order to maximize energy production;
3. plant patrolling, for security purposes;
4. plant predictive monitoring, for the O&M of plant functionality;
5. plant monitoring and maintenance, for issues not directly linked to the plant functionality (e.g., grass mowing or bird shooing away).

The research covered by the materials in this section spans from approximately 2011 to 2024. This range reflects the growing interest in integrating robotics into CSP plants, driven by advances in automation and the need for more efficient and sustainable energy solutions. The keywords that were used in identifying the materials are as follows: "Concentrating Solar Power Robotics" provided a comprehensive research scope to encompass all aspects of robotics in CSP plants; "Cleaning Robots" was selected because cleaning is a critical task for maintaining CSP plant efficiency; "CSP Plant Maintenance Automation" captured studies on the automation of maintenance tasks, which is vital for reducing operational costs and enhancing plant reliability; "UAV Inspection in Solar Power Plants" was included to address the increasing use of drones for inspection tasks, as highlighted by recent research; and "Robotic Systems in Solar Tower Plants" ensured that studies focused on solar towers, a prevalent CSP project, were considered. These keywords were essential for exploring how robotics and automation are applied to improve the efficiency and reliability of CSP plants.

Search fields included academic databases such as IEEE Xplore, ScienceDirect, and Google Scholar, where topics on renewable energy, automation, and robotics are frequently published. Additionally, industry reports and conference proceedings on renewable energy technologies were explored to capture the latest developments in this industry niche. An analysis of patents and research project deliverables was also conducted to fully understand the technological evolution and innovation in this field. Inclusion criteria for articles focused on studies that explicitly discuss the application of robotics in CSP plants, with a preference for empirical research and case studies that provide concrete results or prototypes. Articles were included if they addressed specific robotic tasks such as cleaning, inspection, or maintenance in CSP plants. Exclusion criteria filtered out materials that only tangentially mentioned CSP or robotics without providing significant insights into their integration, as well as those that focused on other forms of solar power generation, such as PV or robotic applications unrelated to CSP, and simple laboratory prototypes. Additionally, studies published before 2011 or those without peer review were excluded to ensure the relevance and quality of research.

In Table 1, the relation between the type of CSP plant and conceivable robotic tasks is presented. References are shown per task in the last column and per plant type in last row. As it happens, in the photovoltaic realm, the largest number of applications is in the field of mirror cleaning, which is more directly linked to its specific Levelized Cost of Electricity (LCOE).

In [30], HECTOR is described (Heliostat Cleaning Team-Oriented Robot), one of the first robotic solutions for the cleaning of heliostats in solar tower CSP plants. Its efficiency with respect to water and power permits it to be fully autonomous during the cleaning of one heliostat. The transport of the robot from one heliostat to the next is not autonomous, relying on human intervention. For a truly efficient cleaning, a team of robots should work in parallel. The motion of the robot is wheeled; thus, it is concocted for heliostats with almost horizontal poses, the guidance is performed exploiting the boundaries of the mirrors, and it is capable of avoiding broken parts of mirrors. Tested at the Gemasolar plant in Spain, it is presently commercialized by the Sener company.

An inspection climbing robot is described in [31]. This can perform inspections and maintenance on the vertical surfaces of the receiver in a solar tower CSP plant. The non-destructive testing (NDT) inspections are performed by means of eddy current measurements for corrosion, erosion, and cracking, as well as a visual one. The robot carries the sensors and is kept in contact with the tower via suction. The vertical movement is performed with the help of a crane atop the tower. Servos ensure the proper positioning of the sensors, and the load capacity is only limited by the crane. The robot weights around 300kg. The robot moves on the panels of the receiver in the tower and aligns the sensors correctly for inspection.

In a review [32] dedicated to the overall cleaning solutions for CSP plants, cleaning methods and cleaning systems are extensively covered, and three robotic solutions are cited. Besides the quoted robot [30], the PARIS robot for trough reflectors is presented [33,34], as developed by the Sener company, and a UAV-based system is reviewed [35].

The cleaning robot for Fresnel mirrors, Soltibot, developed by Soltigua, is presented in [36], a result of the EU project MinWaterCSP (2016–20) [37]. The operation of the robot is illustrated in Figure 1.



Figure 1. Cleaning event; **left:** the selected collector on dirty state, **right:** operating cleaning robot on the day of cleaning [36].

It is worth noting that in [36], the authors measure the advantages, in terms of time and cost, of the use of the robot in a working environment.

Reference [38] is a comprehensive review of integrating UAV-based systems for the planning, operation, and maintenance of CSP plants. The main fields of application of UAV-based technologies are the characterization, aimpoint control, and calibration of the light concentrating system both for heliostats and parabolic troughs. The UAV flight path is optimized within the CSP field to collect the needed images for the planned task and to

mitigate the time-of-flight problem due to limited power. It is important to note that the reviewed approaches make use of RTK GPS for the UAV positioning. Even with such RTK technology, which drops the accuracy to a few centimetres, the reconstruction of the mirrors can be insufficient since actual deviations on mirrors are usually below the millimetre level.

An interesting approach to the field inspections of heliostats is presented in [39], where the authors developed a polarimetric imaging drone, carrying out field tests at Sandia's National Solar Thermal Test Facility and showing that Degree of Linear Polarization and Angle of Polarization images greatly enhance the edge detection results compared with the conventional visible images, supporting the fast and accurate detection of heliostat mirror edges and cracks. It also foresees possible future UAV applications in CSP plants such as the use of UAV swarms to save time, inspections for faults, land surveys for plant designs, mirror cleaning with water or air jets or with ultrasound methods, wildlife protection, and security and surveillance.

A detailed study focusing on the maintenance of the mirrors of a CSP plant is presented in [40]. This employs images taken by drones flying over the plant and shows that mirror inspection from a drone using high resolution images can yield data about soiling and mirror defects.

A cleaning robot for Fresnel mirrors and collectors, named Frenell, by the Frenell company is presented in [41]. In the same paper, another robot for Fresnel mirrors, produced and commercialized by SUNCNIM, is also presented. It is interesting to note that the SUNCNIM company designs, builds, and operates solar plants with energy storage for the production of electricity and thermal energy; in this sense, the robotic aspect is a part of a whole, even if the company is currently in judicial liquidation [42].

Table 1. Possible intersection of robotic tasks with type of CSP plant.

Robotic task	Parabolic Trough	Fresnel	Solar Tower	Dish Stirling	References
System deployment		X	X		[42]
Mirror cleaning	X	X	X		[30,33,34,38,39,41,42]
Security patrolling	X	X	X	X	
Monitoring O&M	X	X	X	X	[31,35–37]
Non energy monitoring	X	X	X	X	
References	[33]	[34,38,39]	[30,31,35–37,41,42]		

The EU's ongoing project Durable [43] will apply disruptive aerospace, robotic, non-destructive inspection and additive manufacturing technologies to evolve towards better development in the operation and maintenance of wind and solar energy parks. In WP5, approaches for control and surveillance, such as non-destructive testing by robots (UAVs or UGVs), contact inspection (ultrasonic, thermographic), and autonomous and intelligent navigation, will be investigated and applied.

The Heliogen company (Pasadena, CA, USA) has developed ChariotAV, which combines sensors, robotics, and autonomous control software to service an entire field of heliostats independently for cleaning [44]. The follow-up has been the ICARUS robot (Installation & Cleaning Autonomous Robot & Utility Solution) [45], an autonomous vehicle with mounted LIDAR and high precision RTK GPS, able to perform the cleaning of heliostats by towing an equipped trailer and to deploy small dimension octagonal heliostats with high precision.

An innovative non-intrusive optical (NIO) approach is proposed in [46] to measure mirror surface slope error, mirror facet canting error, and heliostat tracking error based on reflection images using the natural target, the tower, in a heliostat field. The approach adopts various techniques in photogrammetry, reflectometry, and geometrical optics to determine the relative positions of the camera, tower, and heliostat. Snell's law is applied

to derive the mirror surface slope error, mirror facet canting error, and heliostat tracking error. All the images are collected with unmanned UAVs.

Briefly summarizing the above, it is possible to state that the selected papers present mainly cleaning or drone-based aerial mirror inspection systems. In addition, several robotic systems have been realized and are advertised on the world wide web, but often with outdated pages, hinting to working prototypes that are still not ready for the market. This may be due to an overly limited market. Figure 2 highlights the distribution of the research focus across different robotic tasks, showing that most of the studies are concentrated on monitoring for operation and maintenance and mirror cleaning applications, making up about 31% and 36% of the published papers; the remaining published papers focus on system deployment—9%, non energy monitoring—16%, and security patrolling—8%.

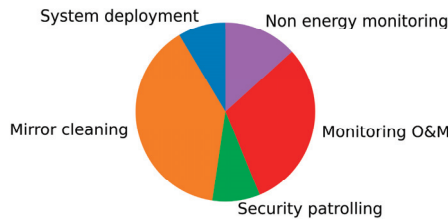


Figure 2. Distribution of CSP research focus across different robotic tasks.

3.2. Robotics in Photovoltaic

At present, most of the applications of robotics for photovoltaic solar power plants is related to the cleaning of the panels, and various papers report a range of related innovative technologies. The exploration of the papers begins with an in-depth look at cutting-edge cleaning methods, including dry-cleaning systems, water-free solutions, and specialized robots for desert environments. Robots for inspection and maintenance, exploiting visual servoing, real-time object detection, and predictive maintenance via unmanned aerial vehicles (UAVs) and machine learning algorithms, are also receiving increasing interest. Sophisticated and intelligent robotic systems have been employed, from quadruped robotic platforms facilitating non-destructive evaluation of modular constructs, which are adept at accommodating various panel configurations. The implementation of robot details such as multi-suspension units, expertly crafted to dampen vibrations and minimize their impact, are presented. Also, cutting-edge control strategies aimed at boosting autonomy and overall effectiveness, infrared-based dust mitigation systems, and advanced navigation and control features are addressed. The aim here is to offer readers a broad overview of the diverse robotic solutions proposed to transform solar panel maintenance.

The research covered by the articles reviewed in this section was conducted between the years 2014 and 2024. This period marks significant advancements in robotics and renewable energy, especially in the application of autonomous systems in PV plants, and reflects the rapid technological progress and increasing focus on photovoltaic energy worldwide. The choice of search keywords was guided by the need to explore the intersection of robotics and PV systems, covering a range of relevant applications and functions. The keywords that were instrumental in identifying the articles selected in this review include the following: “Robotics in Photovoltaic Systems”—directly targeting the integration of robotic technology within PV systems, capturing papers that discuss the use of robots for various tasks such as cleaning, inspection, and maintenance; “Solar Panel Cleaning Robots”—cleaning panels is a key issue, so this keyword ensured the inclusion of research focused on developing robotic systems for the efficient cleaning of solar panels; “UAV Inspection for Solar Panels”—this was essential for identifying articles that investigate the use of UAVs for inspecting solar panels, reflecting a growing interest in this area of photovoltaic system maintenance; “Predictive Maintenance in PV Plants”—this was aimed at the discovery of studies exploring advanced maintenance strategies, particularly those em-

ploying predictive algorithms and robotics to improve the efficiency of photovoltaic plants; “Autonomous Robots in Renewable Energy”—this broader keyword captured innovations in autonomous robotic systems across various renewable energy sectors, with a particular emphasis on their applications within photovoltaic systems. The search focused on titles, abstracts, and keywords within databases like IEEE Xplore, ScienceDirect, and Google Scholar. The inclusion of these fields was crucial for capturing the core focus of papers, ensuring relevance to the overarching theme of robotics in PV systems. The searches were conducted with filters for peer-reviewed journal articles and conference papers to maintain the quality of the selected research. Inclusion criteria were papers published in the last five years that specifically addressed the application of robotics in PV systems, including cleaning, inspection, and maintenance. Articles had to be peer reviewed and accessible in full text to be included in the review. Exclusion criteria were papers that focused on unrelated aspects of robotics or PV systems, such as purely theoretical models without practical application, studies not involving robotic systems, research focused on non-solar renewable energy technologies, or simple laboratory prototypes. Papers not available in English or lacking full-text access were also excluded to maintain consistency and depth in the review. This methodical approach ensured a focused and comprehensive selection of articles that provide valuable insights into the current state and future potential of robotics in PV systems. Table 2 organizes the materials according to the robotic tasks they address and to the application environment in PV systems they focus on.

In [47], the design and creation of a robot engineered for dry-cleaning PV solar panels to cope with efficiency losses due to dust accumulation is described. The robot includes an image processing system for colour analysis to monitor dust density, optimizing the cleaning process. Featuring a cleaning head with two cylindrical brushes powered by a 12 V DC motor, the robot’s frame is made of lightweight aluminium. The robot performs vertical and horizontal cleaning cycles, with each cycle lasting approximately 10 s, and operates on a 12 V battery recharged by a small PV panel. Controlled by an Arduino Uno microcontroller and a Bluetooth module, the robot’s effectiveness is validated by significant increases in panel efficiency post-cleaning across different dust densities.

Ref. [48] reports on the development of an autonomous and smart mobile robot system designed to enhance the maintenance efficiency of solar PV arrays by automating the cleaning process. The mobile robot is equipped with gyroscope and proximity sensors to efficiently navigate and follow the sweep path over the PV array surfaces. Utilizing a proportional integral derivative (PID) control system, the robot stabilizes its movement and maintains its path with an average stabilization time of 5.72 s. Operating in both autonomous and manual modes, it cleans a 56-square-metre PV array in approximately 13 min autonomously and in 20–24 min manually. The robot’s structure features an aluminium frame and an acrylic body and weighs 4.8 kg. It is powered by a four-cell LiPo battery, providing around 40 min of operational time. The drive system includes wheels driven by PG28-type motors, allowing for six degrees of freedom in movement. Each sweeper roller, equipped with GW4632-370 motors and nylon tassels, rotates at 150 RPM to efficiently clean the panels, with proximity sensors preventing falls. This innovative design significantly reduces the effort, cost, and risk associated with the manual cleaning of solar arrays, improving the overall efficiency of PV power plants.

Ref. [49] describes the development and implementation of an autonomous robot designed to clean PV panels in desert environments, addressing the issue of dust and sand accumulation that can reduce efficiency by up to 80%. Using a waterless cleaning method, the robot operates without the need for rails or guides and features two independently motorized helical brushes that rotate to remove dust and sand from the panels. Moving horizontally along the PV array, the robot activates only the brush facing the direction of movement to prevent the re-deposition of cleaned sand. Key components include ultrasonic sensors for real-time speed and position regulation, high-friction rubber belts for smooth movement on uneven surfaces, and an Arduino DUE platform for managing movement and brush activation. The robot’s design innovations, such as the elimination of guide rails

and a controlled brush system, ensure effective dust removal and low power consumption. Experimental tests validate its effectiveness.

Ref. [50] discusses a robotic system optimized for cleaning solar panels, focusing on dust accumulation. This autonomous robot uses a combination of brushes and vacuum mechanisms to remove dust, navigating PV arrays with sensors to detect panel edges and avoid obstacles. It follows a predefined path for systematic coverage and docks at a charging station when necessary. Through multiple testing sessions, parameters such as brush speed, vacuum suction, and movement speed are optimized, resulting in significant improvements in cleaning performance. The optimized robot enhances PV efficiency by maintaining cleanliness, increasing power output and longevity.

Ref. [51] describes an automated robotic system for dry-cleaning solar panels in Thuwal, Saudi Arabia, using a silicone rubber brush, shown in Figure 3. The robot, weighing 36 kg, operates daily at dawn to minimize sunlight obstruction and efficiently removes overnight dust accumulation. Field tests using a 10 kW PV system showed that the robot's daily cleaning outperformed weekly manual cleaning, demonstrating its low-cost, reliable solution for maintaining solar panel cleanliness in arid regions.

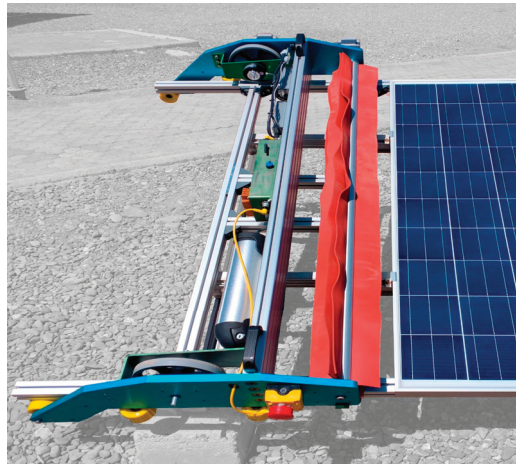


Figure 3. The robotic cleaning system as installed in the test field [51].

Figure 4 shows the comparison of electroluminescence images of a solar panel before and after driving across and brushing the panel surface for 1000 cleaning cycles; the image recorded after the cycles does not reveal any micro-cracks or broken cells, indicating that the quality of the solar cells was preserved.

Ref. [52] presents a novel water-free cleaning robot for distributed PV systems in water-scarce areas. The lightweight robot, equipped with a negative pressure adsorption system and independent wheels, can clean inclined surfaces up to 30 degrees. It uses a rolling brush and negative pressure dust removal system, achieving high cleaning efficiency without secondary pollution. The robot's design includes PWM-controlled motors to prevent vibrations, ensuring effective dust collection and enhancing PV panel performance.

Ref. [53] details the development of the robot "UTU", a compact, mobile device designed to clean solar panels by removing dirt and dust. The UTU robot features a water suction mechanism, roller brushes, tank track wheels, and a 3D-printed cover. It uses IR sensors and a 9-DOF sensor for edge detection and navigation. The robot can be controlled via a mobile application, allowing for remote monitoring and operation. The UTU robot's advanced technology and operating algorithm ensure thorough and efficient cleaning, making it suitable for solar panels in hard-to-reach areas.

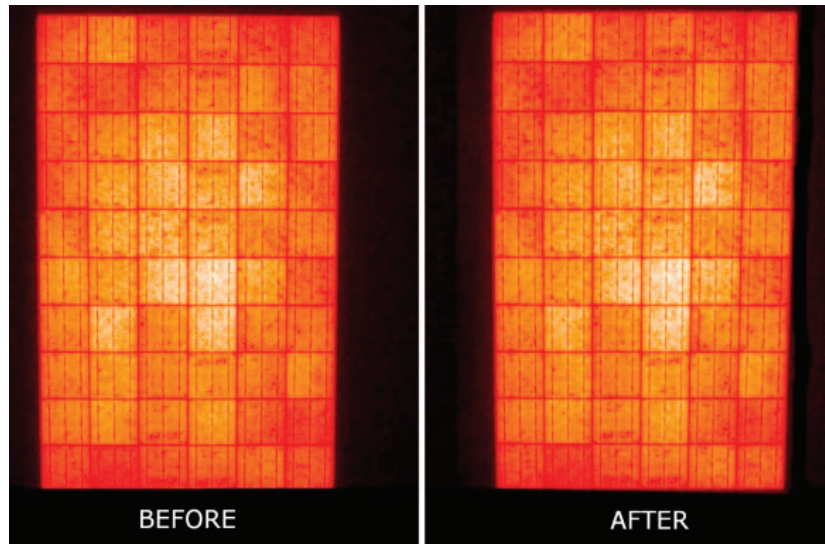


Figure 4. Comparison of electroluminescence of solar panel after cleaning solar module for 1000 cycles (from [51]).

Ref. [54] discusses an autonomous solar panel cleaning robot equipped with rubber wheels for non-slip omnidirectional rotation. It features an air vacuum system to remove dust from panel surfaces and operates without cable or water pipe connections. The robot harnesses its energy through a flexible solar panel and can be controlled manually or operate autonomously. Its efficient dust removal system and autonomous capabilities make it a practical solution for maintaining solar panel efficiency, especially in water-scarce areas.

Ref. [55] focuses on the use of UAVs equipped with a visual serving control system integrated with nonlinear model predictive control (NMPC) for inspecting PV arrays. Using a DJI Matrice 100, the UAV accurately tracks the middle of PV arrays at various velocities and heights, ensuring detailed image acquisition during low-altitude flights. The system extracts features from RGB-D images and uses a Kalman filter to estimate PV array edges, allowing for precise positioning which is critical for high-quality image capturing and necessary for expert analysis or AI-based fault detection. Validated in both simulated and real-world conditions, the system reduces useless data and improves image resolution, enhancing the efficiency of PV array inspections through real-time feature extraction and control execution.

Ref. [56] explores using UAVs equipped with a keypoint-based object detection framework for the real-time inspection of solar farms. Utilizing the NVIDIA AGX Jetson Orin platform, the UAVs achieve up to 60 frames per second at a 1024×1376 resolution, ideal for quick and accurate industrial inspections. The UAVs detect the vertices of solar panels, estimating their six degrees of freedom (6-DOF) pose, and autonomously detect defects and anomalies, facilitating efficient inspections. By integrating UAVs with advanced object detection frameworks, the system enhances accessibility, reduces costs, improves safety, and enables real-time data processing for proactive maintenance, thereby improving solar farm inspections.

Ref. [57] discusses a system for the automatic detection, classification, and localization of defects in large PV plants using UAVs equipped with IR and RGB imaging. The Mavic 2 Enterprise Advanced UAV captures thermal images with a 640×512 resolution and RGB images with an 8000×6000 resolution. The system captures IR images to detect thermal defects and RGB images for surface defects on PV modules. Tasks include image capturing, defect detection, image stitching and processing, defect classification using a convolutional

neural network (CNN), defect localization, and data cross-validation between IR and RGB images. The UAV-based system enhances image features through techniques like Max Pooling and local response normalization, achieving high classification accuracy and facilitating maintenance by accurately identifying and recording defect locations. This system proves to be a valuable tool for efficiently maintaining large-scale PV plants.

Ref. [58] discusses the development of a UAV inspection system for the preventive maintenance of solar farms using a DJI Tello quadrotor. Equipped with various sensors and camera vision technologies, this cost-effective platform enables several machine vision capabilities. Tasks performed by the UAV include autonomous navigation using image filtering, edge detection through OpenCV, and visual odometry for trajectory detection. It employs TensorFlow's object detection API with MobileNets and Single Shot Detector architecture to detect and classify solar panel defects in real-time, such as glass breakage, dust shading, bird droppings, snow, and leaves. The system also classifies solar panel types and performs precision landing on a wireless charging pad using augmented reality markers. Additionally, the system provides real-time monitoring, logs detected defects, and maps the inspection route, displaying crucial parameters through a GUI interface. This UAV inspection system offers enhanced mobility, reduced time and costs, and improved coverage and efficiency in solar farm maintenance.

Ref. [59] discusses a water-free cleaning robot designed for dust removal from PV panels in water-scarce areas. The robot uses a wheeled mechanism and negative pressure adsorption system to adhere to and clean panels with a tilt angle of up to 30 degrees. It employs a rolling brush and high-speed centrifugal fan for efficient dust collection. Weighing no more than 8 kg, the robot can traverse obstacles and cover a cruising distance of up to 200 m. Field tests demonstrated a significant improvement in PV system efficiency, making it a practical solution for maintaining optimal power generation.

Ref. [60] examines the deployment of the Boston Dynamics Spot robot for the non-destructive evaluation (NDE) of renewable energy infrastructures. Equipped with tools such as thermographic cameras, LIDAR, ultrasonic transducers, and acoustic cameras, the quadruped robot is capable of inspecting not only PV installations but also other infrastructures, such as hydroelectric plants and wind turbines. The robot performs the periodic monitoring and inspection of critical components in hazardous or hard-to-reach areas, enhancing safety and efficiency. It autonomously inspects hydro plant components and performs detailed analysis using thermographic and LIDAR measurements. The integration of IoT and machine learning technologies enables predictive maintenance and fault prevention, and it has been conceived to incorporate, in the future, drones and humanoid robots to further enhance capabilities and eliminate human presence in hazardous zones.

Ref. [61] addresses the challenge of vibrations caused by solar panel cleaning robots (SPCRs), which can lead to microcracks in silicon cells and deform PV panels, reducing their lifespan and efficiency. The study proposes a multi-suspension unit designed to dampen vibrations and shocks, including dampers and shock absorbers to minimize dynamic loads. Inspired by tank track suspensions, the unit features adjustable spring mechanisms ensuring proper contact between the robot's track wheels and the PV panels. Experimental setups show a significant decrease in panel deflections and improved PV system durability. This development enhances SPCR design, addressing vibration and pressure impacts, thereby extending the longevity and operational efficiency of PV panels.

Ref. [62] discusses the development of a modular robot designed to clean various lengths of solar panels, addressing the need for an efficient and adaptable cleaning solution. The robot uses a universal module for transmission and manipulation, adapting to different solar panel configurations and sizes, ranging from 1 to 4 m. The primary cleaning mechanism involves water, a spiral brush, and a rubber sweeper to remove dust and dirt, enhancing power generation efficiency. The modular design provides flexibility for different types of solar panel installations, including solar farms, rooftops, and floating solar panels. In Thailand, where air pollution and agricultural activities contribute to significant

dust build-up, the robot's cleaning method improves efficiency by up to 10%. Operating autonomously, it reduces the need for manual labour and the associated costs, offering a cost-effective and efficient alternative to traditional cleaning methods such as using telescopic poles and mops. The robot's adaptability and thorough cleaning capabilities make it particularly beneficial in environments with prevalent dust and bird droppings.

Ref. [63] discusses the design and application of an SPCR to enhance PV system efficiency by removing dust and debris. The SPCR features a dual-motor and crawler system for horizontal movement across panels and a vertical cleaning brush for thorough cleaning. An Arduino board controls the robot, utilizing limit and distance sensors to maintain designated operation areas. The SPCR includes a charging station for automatic recharging and a pulley pallet system for smooth panel transitions without additional rails, reducing material and labour costs. Constructed from lightweight aluminium sigma profiles, the SPCR is portable and easy to install, providing a cost-effective and efficient solution for maintaining solar panel cleanliness and optimizing energy output.

Ref. [64] focuses on a robotic vacuum cleaner designed for solar panels, addressing environmental challenges like panel inclination and dust removal. The robot features a two-stage cleaning mechanism, consisting of the following: a rolling brush for dust agitation and a vacuum motor for dust collection. Equipped with accelerometers and ultrasonic sensors, the robot controls orientation, detects edges, and prevents falls. It follows a reciprocating motion path for maximum coverage and docks for recharging when battery levels are low. Gripper wheels allow for the traversal of inclined surfaces, and obstacle handling capabilities enable smooth movement over panel junctions. The robot's efficient cleaning and minimal power consumption make it a viable solution for solar panel maintenance.

Ref. [65] details IR sensor-based dust mitigation system for solar panels, operated by a robotic arm. This system addresses dust accumulation by initiating early morning cleaning, monitoring panel temperatures, and moving autonomously across panels. The robotic arm uses a silicone rubber wiper and water sprinkler for effective cleaning, achieving high efficiency with minimal water consumption. Field tests showed an average increase of 11.88% in energy output and 13.02% in module efficiency, making it a cost-effective and scalable solution for solar panel maintenance.

Ref. [66] discusses the design of an intelligent cleaning robot for solar panels, identifying two main types as follows: small-scale trackless walking robots and large-scale crossing track walking robots. The robot features two strip brushes and one rolling brush for enhanced cleaning, capable of climbing slopes and handling height differences between panels. It includes replaceable soft and hard brushes for different types of dirt. The robot's design addresses limitations such as battery power, transportation convenience, and wind resistance, making it a versatile and effective solution for maintaining solar panel cleanliness and efficiency. Figure 5 highlights the distribution of the research focus across different robotic tasks within the articles, showing that most of the studies are concentrated on robotic cleaning solutions for PV systems. In detail, about 33% were related to cleaning issues, followed by inspection—18%, general maintenance—13%, control strategies—9%, predictive maintenance—9%, vibration mitigation—9%, and modular adaption—9%.

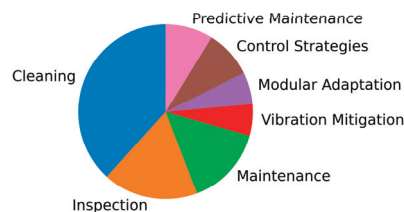


Figure 5. Distribution of PV research focus across different robotic tasks.

Table 2. Relationship between robotic tasks and application environments in PV systems, with references to the papers.

Robotic Task	Desert Env.	Water-Scarce Areas	Large PV Plants	Distributed PV Systems	Hard-to-Reach Areas	General/Multiple Env.	References
Cleaning	X	X	X	X	X	X	[47–55,57,59,62,64,66]
Inspection	X		X			X	[53–58,60]
Maintenance			X	X		X	[51,57,58,60,63,66]
Vibration Mitigation						X	[50,61]
Modular Adaptation						X	[56,62]
Control Strategies						X	[58,61,65]
Predictive Maintenance						X	[60,63]
References	[49,55,62]	[52,54,59]	[48,53,57,58,60,63,64]	[51,52,63,66]	[53,66]	[47,50,55–58,60–63,65]	

3.3. Robotics for the Wind Farms

The present global wind power capacity is 743 GW (2021) [67], relying on wind farms that can be installed either onshore or in shallow water. Therefore, different robotic tasks may be considered as follows:

- (a) the system monitoring and maintenance: made with aerial unmanned vehicles, crawlers/climbers, or underwater robots;
- (b) the system manufacturing or assembly: made with the help of robots for rotor blade inspection or to assist the assembly of parts into a wind tower;
- (c) system security and surveillance.

The research across the articles of this section was conducted within a specific time-frame, predominantly between 2020 and 2023. This period reflects recent advancements in robotics, particularly in response to the growing demand for efficient and reliable wind farm operations.

The keywords that were instrumental in identifying the articles include the following: “Robotic Inspection”, a key task that enhances efficiency and accuracy in detecting defects, corrosion, or other issues, crucial for maintaining the safety and performance of wind farms; “Wind Turbine Maintenance”, which helped narrow down research focused on maintaining turbines, especially in challenging environments such as offshore locations; “Autonomous Systems”, the use of autonomous robots is a growing trend in wind farm operations, with these systems able to perform inspections, maintenance, and even some manufacturing tasks independently; “Offshore Wind Farms”, which facilitated the identification of studies related to the operation and maintenance of offshore turbines; “Blade Inspection”, which captured studies discussing techniques and robots specifically designed to inspect large and often hard-to-reach surfaces, where accurate inspection is essential for extending the lifespan of turbines and preventing costly failures; “Non-Destructive Testing (NDT)”, which is crucial for inspecting wind turbine components without causing damage, allowing robots using NDT to provide detailed assessments of turbine components while they remain in operation, making this keyword highly relevant for capturing articles on advanced inspection technologies.

Searches were conducted in scientific databases such as IEEE Xplore, ScienceDirect, SpringerLink, and Google Scholar to ensure the inclusion of both academic and applied research. The inclusion criteria focused on papers published within the last five years to ensure relevance to current technology trends. Only peer-reviewed articles were considered to maintain a high standard of quality. The focus was on research explicitly addressing

robotic applications in wind farms, excluding papers that discussed wind energy in general without a robotics focus. Additionally, articles were excluded if they present simple laboratory prototypes or if they did not provide the experimental validation or practical application of the discussed technologies, ensuring that the selected papers contributed tangible insights to the field. Table 3 presents a categorization of robotic tasks in wind farms across different applications and technological approaches.

In [68], an imaginative view in which heterogeneous robotic assets, underpinned by AI agent technology, coordinate their behaviour to autonomously inspect, maintain, and repair offshore wind farms over long periods of time and unstable weather conditions is discussed. They cooperate with onshore human operators, who supervise the mission at a distance via the use of shared deliberation techniques. Several challenging research directions in this context are highlighted, and ambitious ideas to tackle them as well as initial solutions are offered. The foreseen robotic solutions are autonomous surface vessels, unmanned aerial systems, and crawling robots.

The development and experimental validation of a complex maintenance system consisting of multiple robotic platforms for a variety of tasks, such as a wind turbine tower and rotor blade service, is described in [69]. As sketched in Figure 6, drones equipped with vision and LIDARs are used for global inspection and to guide slower climbing robots. Localization was enabled by adapting odometry for conical-shaped surfaces considering additional navigation sensors. A prototype of a magnetic robot for NDT and manipulation is also presented. All the robotic platforms run the Robot Operating System (ROS), and a neural network AI approach has been implemented for the detection of the corrosion of welding lines in images.



Figure 6. Wind farm maintenance vision, flying and climbing robots [69].

In [70], a novel framework for acquiring visual data around 3D infrastructures by establishing a team of fully autonomous Micro Aerial Vehicles (MAVs) with robust localization, planning, and perception capabilities is presented. The proposed aerial system reaches a high level of autonomy on a large scale while nearing the real-life deployment of aerial robotics. It mainly employs cameras to take visual data for a structure from a motion pipeline for the 3D reconstruction of the wind turbines. The problem of accurate position measurements of MAVs has been tackled with the use of Ultra-Wide Band (UWB) distance measurements and IMU sensor fusion.

The research work on a new shearography design for integration with a robotic climber for on-site wind turbine blade inspection is presented in [71]. Shearography is an optical technique in the field of non-destructive testing (NDT) of various materials. Its main advantages are that it is noncontact and can cover a large area in a single inspection. It has been widely acknowledged as an effective technique, particularly for the NDT of composite materials to detect subsurface defects such as delamination, disbond, cracks, and impact damages. The sensor is mounted on board a climbing robot that is available on the market and is tested on site.

A robotic platform used for the inspection of rotor blades during manufacturing is presented in [72]. The autonomous robot is equipped with a robotic arm carrying the

following sensors needed for blade inspection: an RGB-D camera and a radar module that works at 80 GHz with a 24 GHz bandwidth. The focus of the paper is on the control of the autonomous mobile manipulator. It provides insights into related research fields, including autonomous navigation and surface orthogonal motion planning. The presented methods are applicable to various tasks related to large-scale inspection.

The design and development of a scaled down prototype of a climbing robot for wind turbine maintenance to perform critical tower operations is presented in [73]. This maintenance robot's interesting feature is a winding mechanism, which uses a tension force to grip onto the tower surface either in static or dynamic situations. The robot is composed of two parenthesis-like halves connected with two wires at the tips, with the above winding mechanism whose tension allows the necessary grip to perform a straight up–down locomotion.

In [74], an autonomous multirobot system which can transport, deploy, and retrieve a wind turbine blade inspection robot using an unmanned aerial vehicle (UAV) is presented. The proposed solution is a fully autonomous system including a robot deployment interface for deployment, a mechatronic link-hook module (LHM) for retrieval, both installed on the underside of a UAV, a mechatronic on-load attaching module installed on the robotic payload, and an intelligent global mission planner. The mentioned robotic payload is the BladeBug MKII, an inspection, maintenance, and repair (IMR) blade crawler robot. Field tests are provided for the different aspects of deployment, retrieval, and operation.

In [75], a crawler-type climbing robot system for measuring the paint film thickness of offshore wind turbine towers is presented. The robot is magnetically linked to the tower and moving on tracks. An interesting feature is its being waterproof, allowing for the monitoring of the submerged parts of the towers. Besides the theoretical model and dynamic simulation, tests and experiments with a working prototype are also provided.

In [76], the use of mobile robotic assistants for the partial automation of wind turbines manufacturing is proposed. The robotic assistant can result in reduced production costs and better working conditions. The article presents the development of a robot assistant for human operators to effectively perform the assembly of wind turbines. The case is from a world's leading wind turbine manufacturer. The developed system is also applicable to other cases of large component manufacturing involving intensive manual effort.

In [77], several methods for robot-based damage detection and evaluation are reviewed. The pinpointed robot technologies are the UAVs and crawling and underwater vehicles. Depending on the payload of the robot, different NDT techniques can be used to assess the different types of damage, including surface damages (e.g., cracks, corrosion, erosion) and hidden damages (e.g., delamination). In general, UAVs have the least payload but the highest efficiency and flexibility, whereas climbing and underwater robots have a higher payload and can support larger and heavier devices but are less flexible. On the side of data analysis, the use of intelligent algorithms reduces the dependence on human experts to perform labour-intensive tasks and improves the quality of inspection by standardizing analysis algorithms and procedures, with very high accuracies.

In [78], the design, test, and in-service operation of a hexapod robot, along with options for its deployment, is described (see Figure 7). The robot can deploy a wide range of measuring equipment and maintenance tools. The hexapod gait employed enables the chassis to be an end-effector, avoiding the use of traditional robot arms, and the multiple degrees of freedom associated with the robot legs enable the robot to walk on both convex and concave surfaces, as well as to straddle a leading-edge profile on a wind turbine aerofoil. It has been tested through multiple blade walks and tasks such as lightning protection system verification, inspections, and blade surface treatments.

Figure 8 highlights the distribution of the research focus across different robotic tasks within the articles, showing that most of the studies are concentrated on robotic solutions for system monitoring and maintenance. Here, system monitoring and maintenance accounts for most of the published research papers, reaching about 56%, followed by system manufacturing and assembly, at 27%, and system security and surveillance, at 17%.

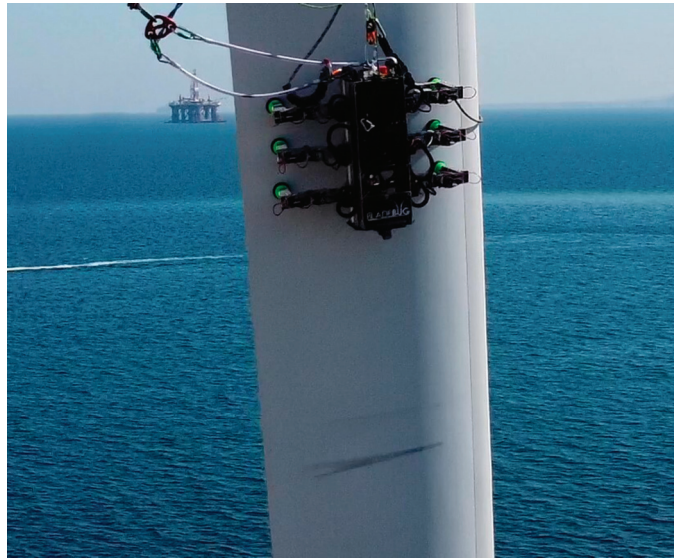


Figure 7. On-blade deployment of a crawling inspection robot [78].

Table 3. Categorization of the considered robotic tasks in wind farms across different applications and technological approaches.

Robotic Task	Onshore Wind Farms	Offshore Wind Farms	Autonomous Operation	Robotic Integration with Human Operators	Advanced Inspection Techniques	References
System Monitoring and Maintenance	X	X	X	X	X	[68–72,74,75,78]
System Manufacturing or Assembly	X	X	X	X	X	[69,72,76]
System Security and Surveillance	X	X	X	X	X	[68,77]
References	[69,72,75–78]	[68,69,74,75]	[68,70,72,74,76–78]	[68,72,74,76,77]	[70–72,77,78]	

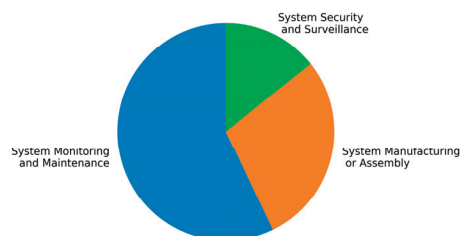


Figure 8. Distribution of wind farm research focus across different robotic tasks.

3.4. Robotics for the Hydroelectric Generation

Dams play an important role today, contributing to the production of a significant portion of electricity from renewable sources. Additionally, they allow for the control of water flows, serve as water reservoirs, and contribute to the reduction in disasters.

In this paragraph, we will focus mainly on dams for hydroelectric power generation. The articles reviewed range mainly from 2016 to 2024, ensuring a comprehensive

overview of the most recent developments in the field. The literature search was conducted using a set of carefully selected keywords, including “robot dam inspection”, “robot dam maintenance”, and “underwater dams concrete monitoring”, which were applied across various search fields, such as titles, abstracts, and keywords, to find relevant studies. To refine the selection, the inclusion criteria focused on peer-reviewed journal articles and conference papers that presented original research or significant reviews within this period. Studies were included if they provided empirical data, detailed methodologies, or in-depth analyses of the technologies under consideration. Exclusion criteria were used to omit non-peer-reviewed sources, studies not available in English, and those lacking sufficient methodological detail or empirical evidence.

Dams are generally made of concrete, and their structure can present problems over time, such as deformations, cracks, spalling, erosion, and wear. Besides the structure, other aspects to monitor include debris that can obstruct the water outlet channels, the penstocks, the turbines, and the general evolution of sediments. Once the damage is identified and maintenance is performed, monitoring is important. But, according to [79], the monitoring and evaluation of crack repair is an important task rarely reported.

The use of robotics in the operation and maintenance of hydroelectric dams enhances safety, efficiency, and data collection capabilities. Robotics technologies are pivotal in performing inspections and maintenance tasks that are otherwise dangerous or challenging for human operators. This type of plant requires the intervention of diverse types of robots depending on the component to be monitored. For underwater monitoring, underwater robots are needed, while for those above water, aerial drones or climbing robots are usually used.

The operations can be divided into those to be carried out above the water level and those below.

3.4.1. Above-Surface Monitoring

The analysis of structural integrity is of fundamental importance to prevent disasters and must be carried out regularly. Automating this process introduces advantages such as cost and risk reduction and the reliability of the results. An advantageous strategy is to use photogrammetry because of its reduced duration of field work, the capacity to record simultaneous three-dimensional coordinates, and, in principle, the ability to monitor an unlimited number of points [80].

Regarding operations above the water, plenty of work has been carried out on the use of drones and climbing robots for the analysis of the dam surface. In [81], a climbing robot is used to detect automatically, count defect instance numbers, and reconstruct the surface of dam spillways by incorporating the deep learning method with a visual 3D reconstruction method. The ORB-SLAM method is used to extract keyframes from the acquired colour and depth images and to obtain the pose transform matrices of all images. Then, a dense environment point cloud model is generated by bundle adjustment.

In [82], the methods and results from a robotic crawler-based ground-penetrating radar (GPR) inspection of rehabilitated concrete from a hydroelectric dam spillway are presented. Standard practise would have left these repairs uninspected or put people at risk when deploying them on the spillway for manual inspections. The paper shows that high-quality data can be recorded from repairs without the need for manual inspection. A remotely controlled robotic crawler system assures repair quality efficiently and effectively (see Figure 9).

A different more holistic approach is described in [83], where an organized, scalable, and decentralized architecture is described, to support decision-making, such as interpretation and goal-oriented reasoning from the cooperation between high-level autonomous cognition systems and expert knowledge. The aim is the inspection of dams through a cognitive-based architecture that creates a high-level decision-making process through decentralized multi-objective reasoning. For instance, it may analyze the current 3D point cloud and, in order to enhance quality, change the originally planned path.



Figure 9. (a) Crawler with GPR being lowered onto spillway; (b) Crawler with GPR collecting data across repaired concrete surface; (c) Data collection with handheld device [82].

3.4.2. Below-Surface Level Monitoring

The presence of currents and poor visibility pose a danger for a human diver. Therefore, the use of robotic systems is a safer and cheaper alternative, avoiding the need to dewater the plant. Underwater robotic vehicles can pinpoint and measure the exact locations of cracks, spalling, erosion, voids, and wear, providing critical data for immediate and long-term maintenance planning. These robots are equipped with advanced imaging and sensor technologies that allow them to operate in zero-visibility conditions and turbulent waters, ensuring detailed and accurate inspections of dam structures and tunnels.

In [84], the use of remotely operated underwater vehicles (ROVs) to carry sensors for the conventional ultrasonic pulse velocity (UPV) technique is discussed for underwater applications. The paper shows that UPV can identify zones of inhomogeneity within the concrete structure and discusses the influence of water saturation, marine growth, and the inspection surface, presenting experimental studies both in the laboratory and in the field.

The development of a dam inspection robot is described in [85]. The robot employs visual data for the inspection of dams. The paper is focused on the solution of adhering to the concrete walls by a negative pressure effect plate due to thrusters, a solution that opposes the common approach with ROVs, which are prone to external disturbances on the stability of the robot and positioning performance and difficulties on the side of untrained operators.

A hybrid robot composed of a crawler robot and an ROV is proposed in [86] to perform reservoir visual inspection. The tracked crawler carries the ROV and inspects the bottom with a 3D sonar; the ROV performs visual inspection on the tank walls and above the water. The focus of the paper is on the control of the distance and orientation to the reservoir surfaces to recognize and register defects (mainly cracks).

Regarding visual techniques for the inspection of concrete in underwater environments, recent methods are heavily based on an artificial intelligence approach centred on neural networks [87,88]. Since the algorithms are usually trained by examples, the availability of datasets plays a crucial role [87]. Another important issue is represented by the fact that low light, due to depth and turbidity, and biofouling may make video processing difficult. In [89], several techniques for the detection of underwater cracks on concrete dam surfaces are reviewed.

In [90], a method is proposed that can overcome issues, such as uneven illumination, colour distortion, and detail blurring, and can effectively detect and localize cracks in underwater optical images with low illumination, a low signal-to-noise ratio, and low contrast. In [91], a method able to handle images which are full of uneven illumination, even when the crack is tiny and interfered with by other factors, is described.

In hydropower, typical situations that involve in situ maintenance include cavitation damage, fatigue-induced cracking, and corrosion. The SCOMPI robot [92] is a robot dedicated for in situ fixing. It is a six-axis manipulator which can deposit weld metal on complex shapes or reconstruct profiles through grinding. It is mainly applied to turbine repair.

In [93], a three-linear axis grinding robot and a test bench designed for the study of underwater grinding processes for the rectification of underwater hydroelectric steel structures are presented.

The review of recent works has shown an ever-increasing interest in the automated monitoring of the concrete structure of the dam. Refs. [84,86] have also integrated solutions for cleaning marine growth. Most of the works have used ROVs for underwater operations, while climbing robots are used both for above- and below-water operations. Drones are obviously used for external analysis, typically image acquisition.

The use of deep learning techniques for the analysis of captured images in order to classify and localize defects is of increasing interest both above [81,83,87] and below the water surface [88]. Transfer learning is a technique to reuse training in other contexts, given the difficulty of finding specific datasets [87].

For the analysis of the inside of the structure, only techniques such as GPR or UPV, for which it is required that the sensor has a stable contact with the structure, are used.

Table 4 categorizes tasks based on the sensors used and the type of monitoring performed. In particular, for the pinpointing of the defect, the localisation of the robot on the dam is of fundamental importance. Ref. [82] noted difficulties in using a GPS to accurately measure a vertical displacement so close to a concrete wall and suggests the use of a robotic total station.

Table 4. Tasks based on the sensors used and on the type of monitoring performed.

Sensors	Robotic Task	Concrete Surface Inspection		Concrete Interior Inspection	
		Above Water Surface	Below Water Surface	Above Water Surface	Below Water Surface
GPR				[82]	
UPV					[84]
Optical camera		[81,83,87]	[85,86,88,89]		

The underwater environment has always been a challenging environment to operate in. The water, often turbid, makes video analysis difficult; the currents complicate the robot's control actions, and radio communications with the ground station are not feasible. In fact, all the papers describe the use of remotely controlled ROVs with direct human intervention.

3.5. Robotics for Power Lines

The search for robotics applications in power line O&M was conducted simply using Google with no specific parameters. This approach aimed to discover not only academic papers but also content from electric companies, providing insight into the working environment for these robots, particularly the inspection and maintenance tasks currently performed by humans that could be taken over by robots.

In addition to understanding the work environment, as described in the following lines, another result from this first phase was the emergence of two main application branches: robots operating suspended from power lines and drones. For these, two separate bibliographic searches were conducted, resulting in two comprehensive reviews recently published (in December 2023 [94] and March 2024 [95], respectively). As new bibliographic searches since those dates have not yielded further results, the contents of these two reviews are detailed further below.

Power lines transport electrical energy from production centres to end users by utilizing wires of various types and sections, operating at different voltage levels depending on the amount of energy to be transported and the distance between production centres

and users. To ensure stable operation and prevent downtime, damage, and malfunctions, periodic monitoring and maintenance programmes have to be implemented.

Power lines span long distances, often crossing thousands of kilometres through suburban areas and inaccessible regions such as mountains, deserts, bodies of water, forests, and other challenging environments. These environments are frequently subject to adverse climatic conditions that expose the power lines to demanding working conditions, such as thermal fluctuations, rain, ice, wind-induced vibrations, and extreme temperatures, all of which can induce significant losses, corrosion, and fatigue ruptures. Additionally, lightning strikes can cause severe damage, including strand ruptures and wire melting. Insulators may develop cracks due to environmental changes and electrical arcs, directly impacting the safety of transmission lines since they provide electrical isolation and mechanical support to the system. In coastal regions, salt accumulation on insulators can lead to a loss of dielectric strength and electrical current leakage between the lines and the ground. Conductors may experience thinning and punctures caused by leakage currents and corona discharge events [96,97].

Power lines frequently encounter aerodynamic instability caused by wind and rain. Like other oscillations, this type of vibration increases the fatigue of conductors, particularly at line supports or clamps. On rainy days, raindrops strike the power line, adhere to the high-voltage conductor, and, under adverse wind conditions, can lead to the collapse of the supporting towers [98–100].

To minimize maintenance costs and to prevent power outages, regular inspections are conducted to identify components requiring maintenance and to detect faults in transmission equipment. Inspection operations focus on components such as insulators, vibration dampers, conductors, spacers, and clamps [101,102]. Over time, two primary methods have been used for inspecting power transmission lines: the first method involves manual inspection by workers on the ground, using telescopes, on gondolas suspended from overhead ground wires, or even climbing the metal towers. This method has significant drawbacks, including high labour intensity, lengthy inspection cycles, high costs, and risks to personal safety. Additionally, ensuring inspection quality in challenging environments such as highlands, meadows, and forests is particularly difficult.

The second method involves using helicopters equipped with various cameras: a visual (daylight) camera, an infrared camera to detect hot spots (as damage to cables and insulators usually increases electrical resistance and local temperature), and an ultraviolet camera to detect corona discharge (defective components cause a local increase in the electrical field, leading to the ionization of air and the emission of UV radiation). Helicopters offer greater efficiency for inspections in high-altitude, cold, and non-dedicated landing areas. However, this method is costly, and the inspection quality decreases under adverse weather conditions [103,104]. Furthermore, some critical defects, such as the internal corrosion of steel-reinforced aluminium conductors, require precise inspection that should preferably be performed from close distances.

Cost, time efficiency, and human safety are critical during the inspection process. To replace human workers, power companies and scientific institutions have developed specialized robots. These robots, which move suspended from the wires, can perform inspection tasks at close distances from the power lines. Various reviews of these types of robots are available in the literature [104–113], but the most recent and comprehensive review of existing and developing robots for power line inspection is presented in [94]. This work details forty-nine robot designs, distinguishing between those that move on ground wires (ten designs) and those that move on live wires, further differentiating between those moving on single wires (thirty-three designs) and multiple wires (six designs).

Table 5 shows, for each type of movement, the number of robots considered in [94] that have reached a certain stage of development. At the first stage, the examine papers illustrate only 3D CAD drawings; at the second, the functionality of the proposed robots has been validated through computer simulations; at the third stage, the robots' functionality has been tested in laboratory experiments (using real parts of power lines or a mock-up);

and at the final stage, the robots have been tested on real power lines (de-energized in some cases).

Table 5. Summary of different stages of development for the robot considered in [94].

Type of Movement	3D CAD Drawings	Computer Simulations	Laboratory Experiments	Field Testing
Multiple wires			3	3
Single wires	5	3	19	6
Ground wires	1		5	4

Robots operating on power lines must overcome various obstacles encountered along their path. These obstacles include dampers, insulators, clamps (such as tension and suspension clamps), spacers, splicing sleeves, towers, aircraft warning lights, and other electrical equipment (see Figure 10).

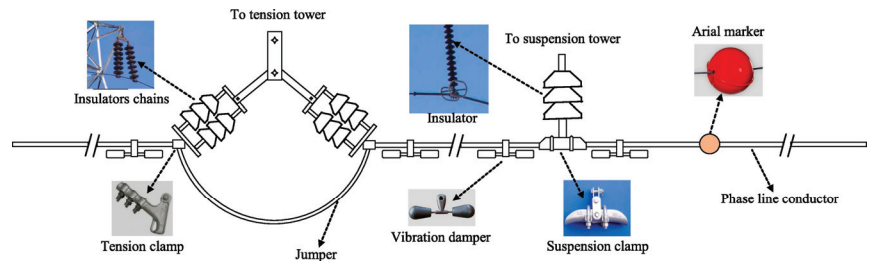


Figure 10. A sketch of a typical power line system showing the obstacles that should be bypassed by any inspecting robot [105].

While some obstacles are small and relatively easy to overcome, others are larger and present significant challenges. As well as overcoming obstacles, the robots should be able to move quickly along the line, halt when it encounters an obstacle, and maintain stability throughout the obstacle-crossing process [105,114,115]. Various approaches to obstacle crossing have been documented, ranging from multi-arm structures to gibbon-inspired crawling structures. Some assistance comes from the fact that the positions of many of these obstacles are known in advance from the line construction plans.

Figure 11a shows Expliner, the robot proposed in [116] and field-tested. Its mobility is based on two pulley units placed on the cables of a bunch line and driven by the electric motors M1 and M2; each of the two pulley units is connected to a horizontal base by another two motors, M3 and M5, and a vertical base is rigidly connected to the horizontal one, with a 2-DOF manipulator connected to its lower end and actuated by the motors M4 and M6; on the tip of this manipulator, there is a counterweight with batteries and electronics.

Therefore, through an automated sequence of coordinated movements, shown in Figure 11b, it is possible to change the position of the center of mass of the device to lift one of the two pulley units and to overcome large obstacles.

As inspection robots are attached to power lines, they add to the overall weight of the lines. One challenge in minimizing the robots' weight is the size of the motors and the need for appropriate electromagnetic shielding for the motors and other electronic components to ensure they function correctly near the fields generated by live lines [117,118].

The inspection process can also be affected by wind, which may cause oscillations in transmission lines. This not only complicates obstacle avoidance but also blurs inspection imaging. Reference [99] introduces a robot designed to withstand oscillations and conduct inspections under windy conditions.

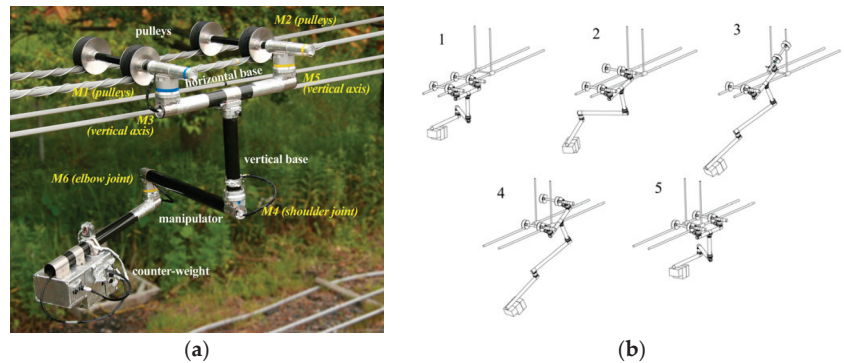


Figure 11. (a) Concept of Expliner with main components; (b) Sequence of motions for overcoming clamp suspender (first half of motion) [116].

In 1991, Sawada et al. from Japan were among the first researchers to develop and design a practical mobile power line inspection robot [119]. This robot navigated a ground wire located above the live lines, maneuvering over obstructions created by subsidiary equipment. When the robot encountered a tower, it unfolded an arc-shaped arm that acted as a guide rail attached to the ground wire on the opposite side of the tower. It then travelled along the arm to pass around to the other side of the tower, and once firmly affixed to the wire on the other side, it detached the arm and folded it up until needed again.

Robots using ground wires were the first to be proposed, as ground wires are always located a few metres above the live lines to ensure lightning protection and are supported at the tops of the towers. Generally, the only obstacles to the robot's travel are the signalling spheres, as ground wires have no insulators. Another advantage of travelling on ground wires is that they do not cause electromagnetic interference with the robot's motors and cameras; thus, electromagnetic shielding is not necessary, reducing the robot's weight.

Conversely, the detailed inspection of live lines requires powerful cameras and specialized control methods. Consequently, researchers began exploring the possibility of robots moving directly on live lines to gather more detailed data [120]. Since these robots pass over or very close to the components that need inspection, the cameras and sensors do not require high-quality sensing. To prevent the motors, cameras, or sensors on the robots from being affected by the magnetic field generated by the live lines, the outer surface of the robot must be designed as an electromagnetic shield.

The utilization of live lines has also prompted the development of new sensors for inspections. For instance, a new laser sensor introduced in study [121] directly measures the diameter of the cable on which the robot's pulleys travel.

Although the robots that move along power lines have the considerable advantage of being able to inspect conductors and other components closely, their inspection processes are considered slow due to the time required to navigate obstacles and the time needed for their installation on the wires [95]. For the latter, special cranes mounted on mobile platforms are necessary, and for some robots, the installation can take about an hour.

For this reason, aerial inspections are often utilized by power companies and research institutions; the widespread commercial availability of drones and other unmanned aerial vehicles (UAVs) makes the development of new flying devices unnecessary, shifting the research focus to image processing instead. A very recent and comprehensive review of the use of unmanned aerial systems for power line inspection is presented in [95].

Hybrid robot designs [122,123] have been introduced to combine the advantages of an inspection process conducted by moving along power lines, with the flying capabilities used to position the platform on the line and bypass obstacles. The stability of landing on power lines, especially under windy conditions, necessitates a complex control strategy

for these robots [124]. To compute their pose relative to the power lines, these robots use cameras and LIDAR together.

4. Machine Learning

Effective operations and maintenance (O&M) practises, including automated robotic approaches, are essential to manage energy production plants, to minimize downtime, and to optimize their performance [125].

Robotic systems, equipped with advanced machine learning (ML) and computer vision (CV) techniques, are currently deeply changing the way O&M tasks are implemented since, once integrated into the plant, they can autonomously navigate and inspect any kind of energy infrastructure, such as wind turbine blades, solar panel arrays, hydroelectric dams, and detecting faults with high accuracy and precision. The integration of efficient ML and CV allows these robotic systems to process vast amounts of data in real-time both in place (edge AI) or remotely exploiting the cloud infrastructure, identifying patterns and anomalies that may indicate potential issues [126,127]. Moreover, autonomous robotic systems can already carry aboard embedded high computing power systems for in-field AI-driven decision-making. A recent example is given in [128], where a drone is shown to autonomously locate faulty PV modules, deciding when to trigger O&M field teams.

Advanced deep learning (DL) vision-based techniques and UAV technologies present a multitude of opportunities for automating the following four primary tasks in power line inspection: mapping and inspecting power line components, detecting and measuring icing, monitoring vegetation encroachment, and disaster monitoring. For example, cutting-edge object detectors powered by deep convolutional neural networks (e.g., Faster R-CNN [129], SSD [130], YOLO [131], and R-FCN [132]) combined with deep neural networks for image classification, such as ResNet [133], Inception-v4, and Inception-ResNet [134], are capable of detecting, classifying, and mapping any power line component.

Moreover, deep learning-based semantic segmentation methods (e.g., DPN [135] and Mask R-CNN [136]), along with traditional background removal techniques like colour-based suppression [137] and pulse-coupled neural filters [138], can effectively eliminate backgrounds from detected components. Subsequently, inspections on the segmented images can be performed using techniques such as texture analysis and vision-based anomaly detection to identify faults. These methods are also advantageous for disaster monitoring. For instance, a tree detection model can be integrated with a power line detection model to identify trees fallen across or against power lines following natural disasters like storms, hurricanes, and earthquakes.

In the field of wind and solar renewable energies, drones and ground robots equipped with high-resolution cameras and infrared sensors can capture detailed images and thermal data from PV panels, CSP collectors, and wind turbine blades [39]. Using CV algorithms, these robots can analyze the images to detect cracks, corrosion, soiling/dirt, and thermal anomalies that are indicative of faults. ML models can then classify these anomalies and predict the likelihood of component failures, enabling proactive maintenance and reducing both OPEX and the risk of unexpected downtime [32,139].

In hydroelectric production infrastructures, ML algorithms have been successfully used for the identification of cracks and spallation in dams. Modern deep learning architectures like YOLO and Faster R-NN have made automatic inspection easier to implement [91] and have substantially improved performance using a limited number of training samples which are particularly difficult to acquire in this specific field [140,141].

Taking advantage of that, robotic systems can now be trained to perform complex tasks that would be challenging or dangerous for human workers. For example, they can conduct inspections in hard-to-reach areas, such as the tops of wind turbine towers or the submerged sections of hydroelectric dams. By automating these tasks, robots not only enhance the safety and efficiency of O&M operations, but also ensure more consistent and reliable fault detection [142,143].

Knowledge extraction relies on a number of machine learning tools, as listed below, that have to be carefully tuned onto the specific energy plant:

Image segmentation: to extract the relevant parts to be monitored, e.g., solar panels in aerial drone imagery, blades and turbine components, dam structural components and water pathways to monitor integrity and leaks, power line components and surrounding vegetation.

Object detection: to identify defects such as cracks hotspots and dirt or foreign objects affecting solar panel performance, detect blade damage, corrosion, or misalignment, identify components and detect faults such as broken insulators or damaged conductors or encroaching vegetation, and identify any dam structural damage (crack, spalling, etc.).

Simultaneous Localisation and Mapping (SLAM): to create detailed maps of solar or wind power plants for precise inspection and maintenance and to map internal structures for maintenance and inspection in hydroelectric plants.

Signal processing for anomaly detection: to analyze thermal and electrical signals to detect anomalies and predict failures in solar panels or power lines and to process vibration and acoustic signals to identify potential equipment failures in wind turbines.

Environmental monitoring: to monitor shading on solar panels from nearby objects or vegetation, track environmental conditions that may affect turbine performance, monitor surrounding areas for vegetation growth and potential fire hazards to power lines, as well as monitoring their impacts in terms of gas pollutants (e.g., ozone) and particulate emissions [144].

Predictive maintenance: to use historical operational data by online anomaly detection systems to forecast infrastructure maintenance requirements.

3D reconstruction: to generate 3D models for comprehensive monitoring and maintenance planning.

4.1. Localization, Segmentation, and Detection

Image localization refers to the task of identifying the specific location of objects within an image. This is a critical step in more complex tasks such as object detection and image segmentation. Key models and techniques for image localization include region-based CNNs (R-CNNs), You Only Look Once (YOLO), and Single Shot MultiBox Detector (SSD). The R-CNN family [145], including Fast R-CNN and Faster R-CNN, represents a significant leap in object detection and localization. These models first generate region proposals and then classify each region to detect objects. Faster R-CNN integrates region proposal generation into the network, significantly speeding up the process and improving accuracy. YOLO is known for its speed and real-time object detection capabilities. Unlike R-CNNs, which process region proposals separately, YOLO frames object detection as a single regression problem, predicting both the bounding boxes and class probabilities directly from full images in one evaluation. This results in much faster detection speeds, making YOLO suitable for real-time applications [131]. SSD also aims at real-time object detection, but takes a different approach by combining predictions from multiple feature maps of different resolutions. This allows SSD to handle objects of various sizes more effectively and efficiently compared to other models.

4.2. Classification

Image classification involves assigning a label to an image based on its content. This task forms the backbone of many CV applications and has been revolutionized by the advent of deep learning. The key models used for image classification include convolutional neural networks (CNNs), vision transformers (ViTs), and, more recently, hybrid models. CNNs: CNNs have been the cornerstone of image classification for nearly a decade. They are designed to automatically and adaptively learn spatial hierarchies of features from input images. Architectures such as AlexNet, VGGNet, Inception, and ResNet have set benchmarks for image classification tasks. ResNet introduced the concept of residual learning, enabling the training of very deep networks by addressing the vanishing gradient

problem. ViTs [146]: Inspired by the success of transformers in natural language processing, vision transformers have recently emerged as a powerful tool for image classification. ViTs treat an image as a sequence of patches and employ self-attention mechanisms to model long-range dependencies. This approach has shown promising results, particularly in scenarios where large-scale pre-training data are available. Hybrid models: Combining the strengths of CNNs and ViTs, hybrid models aim to leverage the local feature extraction capabilities of CNNs with the global context understanding of transformers. These models often outperform their pure counterparts in various benchmarks, offering a balanced trade-off between computational efficiency and accuracy.

4.3. Three-Dimensional Reconstruction

Three-dimensional reconstruction involves creating three-dimensional models from two-dimensional images. This task is essential in applications such as virtual reality, medical imaging, and autonomous navigation. Techniques and models for 3D reconstruction include structure from motion (SfM), multi-view stereo (MVS), and deep learning-based approaches. SfM [147]: SfM is a photogrammetric technique that estimates 3D structures from 2D image sequences. By analyzing the motion of feature points across multiple images, SfM constructs a sparse 3D point cloud of the scene. MVS [148]: MVS takes the sparse point clouds generated by SfM and densifies them, producing detailed 3D models. It uses multiple images from different viewpoints to infer depth information and reconstruct the scene with higher precision. Deep learning-based approaches: Recent advances in deep learning have introduced novel methods for 3D reconstruction, such as volumetric CNNs and differentiable rendering. These models can learn to generate 3D shapes directly from images, leveraging large datasets and powerful computational resources to produce high-fidelity reconstructions.

5. Datasets

The development and training of computer vision models for tasks such as image classification, localization, 3D reconstruction, and predictive maintenance require extensive and diverse datasets for training CV models. Several public datasets have been realized to support research and advancements in these areas. For image classification, popular datasets include ImageNet [149], with millions of labelled images across thousands of categories, and CIFAR-10/100 [150], which provides smaller, more manageable datasets for quick experimentation. For object detection and localization, the COCO (Common Objects in Context) [151] dataset offers a vast array of annotated images with object segmentations and labels, while Pascal VOC [152] provides a comprehensive collection of images with annotated objects for benchmarking detection algorithms. In the realm of 3D reconstruction, datasets such as Middlebury [153] and DTU [154] provide multi-view stereo images and ground truth 3D models, facilitating the evaluation of reconstruction algorithms. Additionally, there are predictive maintenance benefits from datasets like NASA's Turbofan Engine Degradation Simulation Data [155], which includes time-series data and failure records from simulated engines, and the MIMII dataset [156], which contains audio and image data from industrial machines under normal and anomalous conditions. These datasets collectively support the continuous improvement and validation of computer vision models across various applications.

Hong et al. [81] created a multi-class semantic segmentation dataset of 1711 images (with resolutions of 848×480 and 1280×720 pixels) acquired by a wall climbing robot, including cracks, erosion, spots, patched areas, and power safety cable relative to a concrete surface of a dam.

The specific training or fine tuning of pre-trained models for the specific context of the fault detection of renewable energy plants requires vertical datasets on the domain of interest. Below is a list of available public datasets:

The CPLID dataset is the Chinese Power Line Insulator Dataset [157]. (<https://github.com/InsulatorData/InsulatorDataSet> (accessed on 10 September 2024)) It consists of aerial

images, providing 600 images representing normal insulators and of 248 synthetic images of defective insulator (e.g., missing cap). In all the images, insulators are labelled.

Tomaszewski et al. [158] provided the Insulator dataset to the public (<https://github.com/heitorcfelix/public-insulator-datasets/tree/master>, accessed on 10 September 2024). It includes outdoor images taken from the ground, providing 2630 images representing a long rod electrical insulator under varying lighting conditions and against different backgrounds. In all the images, insulators are labelled.

Bian et al. [159] provided the Tower dataset in their research paper (<https://drive.google.com/drive/folders/1UyP0fBNUqFeoW5nmPVGzyFG5lQZcqlc5>, accessed on 10 September 2024). It consists of images collected from inspection videos and the internet, offering 1300 images representing towers in different backgrounds. In all the images, towers are labelled.

Yetgin et al. [160] made the Conductor dataset publicly available in their research paper with the cooperation of the Turkish Electricity Transmission Company (TEIAS) (<https://data.mendeley.com/datasets/n6wrv4ry6v/8>, accessed on 10 September 2024) and <https://data.mendeley.com/datasets/twpxp8xccsw/9>, accessed on 10 September 2024). It contains images collected from inspection videos using infrared and visible cameras, providing 4400 images representing power lines and 4400 images without power lines. In each category, infrared images and visible ones are equal in number. To achieve multiscale recognition, images were acquired at different distances from the target. Also, to allow weakly supervised learning, the dataset is separated in two subsets according to different annotations. Subset 1, made of 8000 images, is labelled with image-level annotations, while the remaining images are labelled with pixel-level annotations.

The Duke California Solar Array Dataset [161] is composed of over 400 km² of imagery and 16,000 hand-labelled solar arrays.

The Desert Knowledge Australia Center Dataset [162] contains multiple real-life data of PV technologies ranging many types, ages, models, and configurations.

The ESOLMET-IER Dataset [163] is made available by The Institute of Renewable Energies UNAM and contains solar metric and meteorological data taken from the station “ESOLMET-IER”.

The National Solar Radiation Data Base (NSRDB) [164] is a publicly available dataset that consists of solar radiation and meteorological data from the US and surrounding countries over the last 23 years.

Finally, the Photovoltaic Thermal Images Dataset [165,166] is composed of aerial thermal images of PV arrays with the presence of one or more anomaly cells.

6. Discussion

Deterioration and faults in renewable energy plants can arise from mechanical failures, electrical issues, environmental factors, and human errors. Some common faults for wind turbines include blade damage or erosion, gearbox failures, generator malfunctions, and yaw system errors. Concerning the photovoltaic plants, faults may include panel degradation or cracking, inverter failures, hotspots and shading issues, and connection and wiring faults, including line-to-line and line-to-ground issues. Concentrated solar plants are mainly affected by several issues at the collector level, including, for example, soiling. In hydroelectric plants, common faults include turbine wear and tear, generator problems, dam and spillway structural issues, and control system errors. Finally, for the power lines, problems may arise from vegetation growth and associated fire hazards, broken insulators, damaged conductors, and missing components. This work has focused on reviewing the papers focused on the investigation, development, and field testing of robotic systems for the inspection and maintenance of such power plants. It has been observed that the number of robotic solutions reported in the scientific literature is proportional to the pervasiveness of the specific type of renewable energy. The largest number of applications is therefore found in the photovoltaic sector, with numerous systems for cleaning panels in different weather and environmental conditions. The applications to CSP or hydroelectric plants is

much more limited. Another evident aspect is that of the ubiquity of the use of unmanned aerial vehicles in photovoltaic plants, CSPs, wind farms, and power lines. The flying systems can carry only lightweight sensors, but since most of the inspection is performed by cameras, either visual or infrared, this does not appear to be a major limitation.

Another important aspect is the development of several hybrid solutions in which the capabilities of different robots are coordinated to tackle the O&M of plants from different points of view at the same time. As an example, flying drones and crawling robots used to inspect wind turbines or even flying drones to deploy and retrieve crawling robots on the blades of a wind generator are increasingly investigated. Automated inspections with drone-based CV systems can detect panel defects, dust accumulation, and shading problems in the case of solar power plants. ML algorithms can now predict turbine failures by analyzing sensor data, while CV can also identify blade damage or misalignment through aerial imagery in the case of wind power. But CV systems can also monitor structural integrity and detect blockages or leaks in real time for the hydropower plants. Finally, monitoring power lines with CV and ML can enhance the detection of faults, icing, and vegetation encroachment, as well as improve disaster response by identifying and assessing damage from natural events.

Besides more traditional NDT such as laser metrology, eddy current measurement, and shearography, the use of visual or infrared information is of paramount importance. This information is processed through machine learning and computer vision techniques typically based on AI and deep learning approaches. Several vision-based methods for anomaly detection have been reviewed and many different training datasets have been presented. Moreover, it is worth noting that some solutions can be found on the market in terms of services for power plants, especially in the field of panel cleaning, but, at the same time, the solutions proposed seem to still be in a very early stage.

Finally, two limitations that characterize this work need to be underlined. First, the contribution of machine learning techniques may be considered, in general, an integral part of the technological transformation, enhancing the capabilities of O&M processes and reducing human intervention by automating selected processes that are integrated with robotic operations. However, the literature investigated in this article has shown that the role of robotics in the operation and maintenance of renewable energy technologies remains somewhat fragmented. While there has been a notable proliferation of such techniques in the inspection sectors of photovoltaic power, wind power, and power lines, their application in other renewable energy sectors, such as management or actuation, remains relatively limited, an issue that will deserve more careful attention from the scientific community.

A second relevant gap in the examined scientific literature and, therefore, a significant cultural deficiency highlighted by this work is the absence of specific technical analyses relating the economic impact of the use of robotics on the renewable energy sector. As a matter of fact, a positive correlation between the increasing use of robotic systems and the increase in the incidence of the unemployment rate has been already highlighted [167]. This general trend should also be reflected in the energy sector, which actually employs around 3% of the total worldwide workforce [168]. On the other hand, the use of robotic systems is expected to increase the standardization level of production processes, leading, in turn, to a decrease in energy production costs, with effects that should overall be positive on worldwide economic growth. Unfortunately, this type of study is currently completely neglected in scientific research.

7. Conclusions

In the renewable energy sector, the use of autonomous machines to carry out the plants operation and maintenance tasks must no longer be considered only as one of the possible feasible operational options. Indeed, it must now be considered an unavoidable choice if the energy transition process, which is designed to drastically reduce the use of fossil fuels by 2050, has to avoid being slowed down or even being interrupted. The underlying reason lies in the fact that as renewable energies spread, their OPEX and CAPEX costs

decrease significantly, and since there is not a fuel cost to take into account, they tend to be mainly limited by “labour-intensive” components. The way to get around this bottleneck is therefore the use of robots that must replace humans both in the installation phase of the systems and, even more so, in the operation and maintenance phase. For example, just to clarify these ideas, in the case of photovoltaic or wind energy, the labour-intensive components now account, respectively, for over 60% and over 40% of the maintenance yearly cost. The range of robots that can be used for these monitoring functions, as this review has shown, is already wide; these include drones, ground robots, and aquatic robots which, however, must be specialized for the particular energy domain in which they are required to operate. The human–machine connection is therefore still central to this area of research and development, especially in relation to areas such as that of renewable energy, characterized by a high granularity of systems with power densities generally 10–1000 times lower than those using fossil fuels and in which local environmental conditions make each system particular and different from any other. A joint and rapid effort is therefore desirable to meet the demands of the industrial sector, which will become increasingly pressing, driven by an increasingly highly competitive energy market.

Author Contributions: Conceptualization, G.D.F.; methodology, S.T., S.C., S.D.V., M.P., G.P. and A.Z.; investigation, S.T., S.C., S.D.V., M.P., G.P. and A.Z.; writing—original draft preparation, S.T., S.C., S.D.V., M.P., G.P. and A.Z.; writing—review and editing, S.T., M.P., G.D.F., A.Z. and S.C.; funding acquisition, G.D.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in the text:

CAPEX	Capital expenditure
CNN	Convolutional neural network
CSP	Concentrated solar power
CV	Computer vision
DL	Deep learning
GPR	Ground-penetrating radar
LIDAR	Light detection and ranging
MAV	Micro aerial vehicle
ML	Machine learning
MVS	Multi-view stereo
NDT	Non-destructive testing
O&M	Operations and maintenance
OPEX	Operative expense
PV	Photovoltaic
R-CNN	Region-based CNN
REN	Renewable energies
ROV	Remotely operated vehicle
SfM	Structure from motion
SPCR	Solar panel cleaning robot
SSD	Single shot detector
UAV	Unmanned aerial vehicle
UPV	Ultrasonic pulse velocity
ViT	Vision transformers

References

- Electricity Production by Source, World. Available online: <https://ourworldindata.org/grapher/electricity-prod-source-stacked> (accessed on 28 June 2024).
- Electricity Grids and Secure Energy Transitions—Analysis—IEA. Available online: <https://www.iea.org/reports/electricity-grids-and-secure-energy-transitions> (accessed on 28 June 2024).
- Wang, J.; Chen, L.; Tan, Z.; Du, E.; Liu, N.; Ma, J.; Sun, M.; Li, C.; Song, J.; Lu, X.; et al. Inherent Spatiotemporal Uncertainty of Renewable Power in China. *Nat. Commun.* **2023**, *14*, 5379. [CrossRef] [PubMed]
- Payne, H.J.; Hemming, S.; Van Rens, B.A.P.; Van Henten, E.J.; Van Mourik, S. Quantifying the Role of Weather Forecast Error on the Uncertainty of Greenhouse Energy Prediction and Power Market Trading. *Biosyst. Eng.* **2022**, *224*, 1–15. [CrossRef]
- Sayed, A.; El-Shimy, M.; El-Metwally, M.; Elshahed, M. Reliability, Availability and Maintainability Analysis for Grid-Connected Solar Photovoltaic Systems. *Energies* **2019**, *12*, 1213. [CrossRef]
- Zheng, Y.; Wei, J.; Zhu, K.; Dong, B. Reliability Analysis Assessment of the Wind Turbines System under Multi-Dimensions. *Adv. Compos. Lett.* **2020**, *29*, 2633366X2096633. [CrossRef]
- Limpens, G.; Jeanmart, H. Electricity Storage Needs for the Energy Transition: An EROI Based Analysis Illustrated by the Case of Belgium. *Energy* **2018**, *152*, 960–973. [CrossRef]
- European Commission; Directorate-General for Energy; Badouard, T.; Moreira de Oliveira, D.; Yearwood, J.; Torres, P.; Altman, M. *Cost of Energy (LCOE)—Energy Costs, Taxes and the Impact of Government Interventions on Investments—Final Report*; Publications Office: Luxembourg, 2020.
- Ramasamy, V.; Zuboy, J.; Woodhouse, M.; O’Shaughnessy, E.; Feldman, D.; Desai, J.; Walker, A.; Margolis, R.; Basore, P. *U.S. Solar Photovoltaic System and Energy Storage Cost Benchmarks, with Minimum Sustainable Price Analysis: Q1 2023*; National Renewable Energy Laboratory (NREL): Golden, CO, USA, 2023.
- Solar Construction Automation Solutions | Terabase Energy. Available online: <https://www.terabase.energy/products-services/construction> (accessed on 28 June 2024).
- Now Publishers—Chapter 1. Introduction to Inspection and Maintenance with Robotics. Available online: <https://nowpublishers.com/article/Chapter/9781638282822?cId=978-1-63828-283-9.ch1> (accessed on 28 June 2024).
- Skog, D.A.; Wimelius, H.; Sandberg, J. Digital Disruption. *Bus. Inf. Syst. Eng.* **2018**, *60*, 431–437. [CrossRef]
- The Disruptive Potential of Energy Digitalization: A Comparative Analysis of German and French Policy-Mixes | LUP Student Papers. Available online: <https://lup.lub.lu.se/student-papers/search/publication/9102203> (accessed on 28 June 2024).
- World Energy Outlook 2022—Analysis—IEA. Available online: <https://www.iea.org/reports/world-energy-outlook-2022> (accessed on 28 June 2024).
- Weather Forecasts Have Become Much More Accurate; We Now Need to Make Them Available to Everyone—Our World in Data. Available online: <https://ourworldindata.org/weather-forecasts> (accessed on 28 June 2024).
- AI: Weather Forecasting. Available online: <https://www.research.hsbc.com/C/1/1/320/9xhMRr6> (accessed on 28 June 2024).
- Knaut, A.; Obermueller, F. *How to Sell Renewable Electricity—Interactions of the Intraday and Day-Ahead Market Under Uncertainty*; EWI Working Papers 2016-4; Energiewirtschaftliches Institut an der Universitaet zu Koeln (EWI): Cologne, Germany, 2016.
- Das, S.; Swapnarekha, H.; Vimal, S. Integration of Blockchain Technology with Renewable Energy for Sustainable Development: Issues, Challenges and Future Direction. In *Computational Intelligence in Pattern Recognition*; Das, A.K., Nayak, J., Naik, B., Vimal, S., Pelusi, D., Eds.; Springer Nature Singapore: Singapore, 2023; pp. 595–607.
- Khare, V.; Bhatia, M. Renewable Energy Trading: Assessment by Blockchain. *Clean. Energy Syst.* **2024**, *8*, 100119. [CrossRef]
- Parmiggiani, E.; Mikalef, P. The Case of Norway and Digital Transformation over the Years. In *Digital Transformation in Norwegian Enterprises*; Mikalef, P., Parmiggiani, E., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 11–18, ISBN 978-3-031-05276-7. [CrossRef]
- Hajjaj, C.; El Ydrissi, M.; Azouzoute, A.; Oufadel, A.; El Alani, O.; Boujoudar, M.; Abraim, M.; Ghennioui, A. Comparing Photovoltaic Power Prediction: Ground-Based Measurements vs. Satellite Data Using an ANN Model. *IEEE J. Photovolt.* **2023**, *13*, 998–1006. [CrossRef]
- Benti, N.E.; Chaka, M.D.; Semie, A.G. Forecasting Renewable Energy Generation with Machine Learning and Deep Learning: Current Advances and Future Prospects. *Sustainability* **2023**, *15*, 7087. [CrossRef]
- Gong, L.; Chen, Y. Machine Learning-Enhanced IoT and Wireless Sensor Networks for Predictive Analysis and Maintenance in Wind Turbine Systems. *Int. J. Intell. Netw.* **2024**, *5*, 133–144. [CrossRef]
- Del Rio, A.M.; Ramirez, I.S.; Marquez, F.P.G. Photovoltaic Solar Power Plant Maintenance Management Based on IoT and Machine Learning. In Proceedings of the IEEE 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Zallaq, Bahrain, 29 September 2021; pp. 423–428.
- Ferlito, S.; Ippolito, S.; Santagata, C.; Schiattarella, P.; Di Francia, G. A Study on an IoT-Based SCADA System for Photovoltaic Utility Plants. *Electronics* **2024**, *13*, 2065. [CrossRef]
- Steffen, B.; Beuse, M.; Tautorat, P.; Schmidt, T.S. Experience Curves for Operations and Maintenance Costs of Renewable Energy Technologies. *Joule* **2020**, *4*, 359–375. [CrossRef]
- Xing, L.; Sizov, G.; Gunderson, O.E. Digital Transformation in Renewable Energy: Use Cases and Experiences from a Nordic Power Producer. In *Digital Transformation in Norwegian Enterprises*; Mikalef, P., Parmiggiani, E., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 63–89, ISBN 978-3-031-05276-7. [CrossRef]

28. Rekeraho, A.; Cotfas, D.T.; Cotfas, P.A.; Bălan, T.C.; Tuyishime, E.; Acheampong, R. Cybersecurity Challenges in IoT-Based Smart Renewable Energy. *Int. J. Inf. Secur.* **2024**, *23*, 101–117. [CrossRef]
29. Rodat, S.; Thonig, R. Status of Concentrated Solar Power Plants Installed Worldwide: Past and Present Data. *Clean. Technol.* **2024**, *6*, 365–378. [CrossRef]
30. Hardt, M.; Martínez, D.; González, A.; Garrido, C.; Aladren, S.; Villa, J.R.; Saenz, J. Hector—Heliostat Cleaning Team-Oriented Robot. In Proceedings of the Solar-PACES 2011 Conference, Granada, Spain, 20–23 September 2011.
31. Felsch, T.; Strauss, G.; Perez, C.; Rego, J.; Maurtua, I.; Susperregi, L.; Rodríguez, J. Robotized Inspection of Vertical Structures of a Solar Power Plant Using NDT Techniques. *Robotics* **2015**, *4*, 103–119. [CrossRef]
32. Bouaddi, S.; Fernández-García, A.; Sansom, C.; Sarasua, J.A.; Wolfertstetter, F.; Bouzekri, H.; Sutter, F.; Azpitarte, I. A Review of Conventional and Innovative Sustainable Methods for Cleaning Reflectors in Concentrating Solar Power Plants. *Sustainability* **2018**, *10*, 3937. [CrossRef]
33. Vicente Peña, C.J.; Massabe Muñoz, F.; Villa Navarro, J.R.; Garrido Ortiz, S. Cleaning System for Cleaning Parabolic Trough Collector Plants and Cleaning Method Using Said System. Patent WO 2012/126559 A2, 27 September 2012.
34. PARIS “Bufs up” Solar Fields. Available online: https://www.renewableenergymagazine.com/solar_thermal_electric/paris-buffs-up-solar-fields (accessed on 21 June 2024).
35. Pardell, R. Agcfd: Automated Glass Cleaning Flying Drone System. U.S. Patent Appl. 15/301,516, 27 July 2017.
36. Azouzoute, A.; Zitouni, H.; El Ydrissi, M.; Hajjaj, C.; Garoum, M.; Bennouna, E.G.; Ghennioui, A. Developing a Cleaning Strategy for Hybrid Solar Plants PV/CSP: Case Study for Semi-Arid Climate. *Energy* **2021**, *228*, 120565. [CrossRef]
37. Cleaning Robot Soltibot® Successfully Tested on Linear Fresnel Collectors. Available online: <https://cordis.europa.eu/article/id/124143-cleaning-robot-soltibot-successfully-tested-on-linear-fresnel-collectors> (accessed on 21 June 2024).
38. Milidonis, K.; Eliades, A.; Grigoriev, V.; Blanco, M.J. Unmanned Aerial Vehicles (UAVs) in the Planning, Operation and Maintenance of Concentrating Solar Thermal Systems: A Review. *Sol. Energy* **2023**, *254*, 182–194. [CrossRef]
39. Tian, M.; Desai, N.; Bai, J.; Brost, R.; Small, D.; Novick, D.; Yellowhair, J.; Rafique, M.Z.E.; Pisharam, V.; Yao, Y. Toward Autonomous Field Inspection of CSP Collectors with a Polarimetric Imaging Drone. In Proceedings of the SolarPACES Conference Proceedings, Sydney, Australia, 10–13 October 2023; Volume 1. [CrossRef]
40. Coventry, J.; Asselineau, C.-A.; Salahat, E.; Raman, M.A.; Mahony, R. A Robotic Vision System for Inspection of Soiling at CSP Plants. In Proceedings of the SOLARPACES 2011, Daegu, Republic of Korea, 1–4 October 2019; p. 100001.
41. EU—Inship Project DELiverable 3.4 Survey of Existing Automated Mirror Cleaning Technologies 2017. Ref. Ares(2016)6703785—30/11/2016. Available online: <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5b6442653&appId=PPGMS> (accessed on 10 September 2024).
42. SUNCNIM | Suncnim. Available online: <https://www.suncnim.com/en> (accessed on 21 June 2024).
43. Home—DURABLE. Available online: <https://www.durableproject.eu/> (accessed on 21 June 2024).
44. Solar Thermal Technology | Heliogen. Available online: <https://www.heliogen.com/Technology/#autonomy-and-Robotics> (accessed on 21 June 2024).
45. Heliogen Demonstrates AI-Powered Autonomous Robot, ICARUS. Available online: <https://www.heliogen.com/press-releases/heliogen-demonstrates-ai-powered-autonomous-robot-icarus-designed-to-lower-installation-and-maintenance-costs-of-full-scale-concentrated-solar-plants/> (accessed on 21 June 2024).
46. Mitchell, R.A.; Zhu, G. A Non-Intrusive Optical (NIO) Approach to Characterize Heliostats in Utility-Scale Power Tower Plants: Methodology and in-Situ Validation. *Solar Energy* **2020**, *209*, 431–445. [CrossRef]
47. Amin, A.; Wang, X.; Alroichdi, A.; Ibrahim, A. Designing and Manufacturing a Robot for Dry-Cleaning PV Solar Panels. *Int. J. Energy Res.* **2023**, *2023*, 7231554. [CrossRef]
48. Megantoro, P.; Abror, A.; Syahbani, M.A.; Anugrah, A.W.; Perkasa, S.D.; Setiadi, H.; Awalim, L.J.; Vigneshwaran, P. Autonomous and Smart Cleaning Mobile Robot System to Improve the Maintenance Efficiency of Solar Photovoltaic Array. *Bull. EEI* **2023**, *12*, 3288–3297. [CrossRef]
49. Antonelli, M.G.; Beomonte Zobel, P.; De Marcellis, A.; Palange, E. Autonomous Robot for Cleaning Photovoltaic Panels in Desert Zones. *Mechatronics* **2020**, *68*, 102372. [CrossRef]
50. Cai, S.; Bao, G.; Ma, X.; Wu, W.; Bian, G.-B.; Rodrigues, J.J.P.C.; De Albuquerque, V.H.C. Parameters Optimization of the Dust Absorbing Structure for Photovoltaic Panel Cleaning Robot Based on Orthogonal Experiment Method. *J. Clean. Prod.* **2019**, *217*, 724–731. [CrossRef]
51. Parrott, B.; Carrasco Zanini, P.; Shehri, A.; Kotsovos, K.; Gereige, I. Automated, Robotic Dry-Cleaning of Solar Panels in Thuwal, Saudi Arabia Using a Silicone Rubber Brush. *Sol. Energy* **2018**, *171*, 526–533. [CrossRef]
52. Fan, S.; Liang, W.; Wang, G.; Zhang, Y.; Cao, S. A Novel Water-Free Cleaning Robot for Dust Removal from Distributed Photovoltaic (PV) in Water-Scarce Areas. *Sol. Energy* **2022**, *241*, 553–563. [CrossRef]
53. Muhammed, S.; Laheb, T.; Ma’ad, H.; Maarroof, O.W. “UTU” Compact Solar Panel Cleaning Robot. *Int. J. Adv. Nat. Sci. Eng. Res.* **2023**, *7*, 217–226.
54. Ölmez, B.B.; Ergezer, Ö.; Güğül, G.N. Autonomous Solar Panel Cleaning Robot with Rubber Wheeled and Air-Absorbing Motor. *Int. J. Energy Appl. Technol.* **2021**, *8*, 182–187. [CrossRef]
55. Velasco-Sánchez, E.; Recalde, L.F.; Guevara, B.S.; Varela-Aldás, J.; Candelas, F.A.; Puente, S.T.; Gandolfo, D.C. Visual Servoing NMPC Applied to UAVs for Photovoltaic Array Inspection. *IEEE Robot. Autom. Lett.* **2024**, *9*, 2766–2773. [CrossRef]

56. Rodriguez-Vazquez, J.; Prieto-Centeno, I.; Fernandez-Cortizas, M.; Perez-Saura, D.; Molina, M.; Campoy, P. Real-Time Object Detection for Autonomous Solar Farm Inspection via UAVs. *Sensors* **2024**, *24*, 777. [CrossRef]
57. Jeffrey Kuo, C.-F.; Chen, S.-H.; Huang, C.-Y. Automatic Detection, Classification and Localization of Defects in Large Photovoltaic Plants Using Unmanned Aerial Vehicles (UAV) Based Infrared (IR) and RGB Imaging. *Energy Convers. Manag.* **2023**, *276*, 116495. [CrossRef]
58. Al-Gawda, M.G.; Lau, C.Y.; Lai, N.S. Autonomous Uav Inspection System For Preventive Maintenance Of Solar Farms. *J. Eng. Sci. Technol.* **2022**, 132–156.
59. Nabti, M.; Bybi, A.; Chater, E.A.; Garoum, M. Machine Learning for Predictive Maintenance of Photovoltaic Panels: Cleaning Process Application. *E3S Web Conf.* **2022**, *336*, 00021. [CrossRef]
60. Tsenis, T.; Tromaras, A.; Kappatos, V. NDE Using Quadrupeled Robotic Platform for Renewable Energy Producing Infrastructures. In Proceedings of the 2024 IEEE 8th Energy Conference ENERGYCON, Doha, Qatar, 4 March 2024; pp. 1–6.
61. Le, N.T.; Nguyen, M.D.; Phan, T.D.; Truong, C.T.; Duong, V.T.; Nguyen, H.H.; Nguyen, T.T. Development of a Multi-Suspension Unit for Solar Cleaning Robots to Mitigate Vibration Impact on Photovoltaic Panels. *Appl. Sci.* **2023**, *13*, 12104. [CrossRef]
62. Chailoet, K.; Pengwang, E. Assembly of Modular Robot for Cleaning Various Length of Solar Panels. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *639*, 012014. [CrossRef]
63. Akyazi, Ö.; Şahin, E.; Özsoy, T.; Algül, M. A Solar Panel Cleaning Robot Design and Application. *Eur. J. Sci. Technol.* **2019**, 343–348. [CrossRef]
64. Aravind, G.; Vasam, G.; Kumar, T.S.B.G.; Balaji, R.N.; Ilango, G.S. A Control Strategy for an Autonomous Robotic Vacuum Cleaner for Solar Panels. In Proceedings of the IEEE 2014 Texas Instruments India Educators' Conference (TIIEC), Bangalore, India, 4–5 April 2014; pp. 53–61.
65. Ghodki, M.K. An Infrared Based Dust Mitigation System Operated by the Robotic Arm for Performance Improvement of the Solar Panel. *Sol. Energy* **2022**, *244*, 343–361. [CrossRef]
66. Yan, S.; Jian, Y.; Xu, L.J. Research on Design of Intelligent Cleaning Robot for Solar Panel. In Proceedings of the the 20th International Conference on Electronic Business, Hong Kong, China, 5 December 2020.
67. GWEC Global Wind Report. Available online: <https://gwec.net/wp-content/uploads/2021/03/GWEC-Global-Wind-Report-2021.pdf> (accessed on 10 June 2024).
68. Bernardini, S.; Jovan, F.; Jiang, Z.; Watson, S.; Weightman, A.; Moradi, P.; Richardson, T.; Sadeghian, R.; Sareh, S. A Multi-Robot Platform for the Autonomous Operation and Maintenance of Offshore Wind Farms. In Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, 9 May 2020.
69. Franko, J.; Du, S.; Kallweit, S.; Duelberg, E.; Engemann, H. Design of a Multi-Robot System for Wind Turbine Maintenance. *Energies* **2020**, *13*, 2552. [CrossRef]
70. Kanellakis, C.; Fresk, E.; Mansouri, S.S.; Kominiak, D.; Nikolakopoulos, G. Towards Visual Inspection of Wind Turbines: A Case of Visual Data Acquisition Using Autonomous Aerial Robots. *IEEE Access* **2020**, *8*, 181650–181661. [CrossRef]
71. Li, Z.; Tokhi, M.O.; Zhao, Z.; Zheng, H. A Compact Laser Shearography System for On-Site Robotic Inspection of Wind Turbine Blades. *JAIT* **2021**. [CrossRef]
72. Engemann, H.; Cönen, P.; Dawar, H.; Du, S.; Kallweit, S. A Robot-Assisted Large-Scale Inspection of Wind Turbine Blades in Manufacturing Using an Autonomous Mobile Manipulator. *Appl. Sci.* **2021**, *11*, 9271. [CrossRef]
73. Liu, J.-H.; Padrigalan, K. Design and Development of a Climbing Robot for Wind Turbine Maintenance. *Appl. Sci.* **2021**, *11*, 2328. [CrossRef]
74. Jiang, Z.; Jovan, F.; Moradi, P.; Richardson, T.; Bernardini, S.; Watson, S.; Weightman, A.; Hine, D. A Multirobot System for Autonomous Deployment and Recovery of a Blade Crawler for Operations and Maintenance of Offshore Wind Turbine Blades. *J. Field Robot.* **2023**, *40*, 73–93. [CrossRef]
75. Yang, P.; Zhang, M.; Sun, L.; Li, X. Design and Control of a Crawler-Type Wall-Climbing Robot System for Measuring Paint Film Thickness of Offshore Wind Turbine Tower. *J. Intell. Robot. Syst.* **2022**, *106*, 50. [CrossRef]
76. Malik, A.A. Development of a Mobile Robot Assistant for Wind Turbines Manufacturing. *arXiv* **2022**, arXiv:2208.03584. Available online: <http://arxiv.org/abs/2208.03584> (accessed on 10 September 2024).
77. Liu, Y.; Hajj, M.; Bao, Y. Review of Robot-Based Damage Assessment for Offshore Wind Turbines. *Renew. Sustain. Energy Rev.* **2022**, *158*, 112187. [CrossRef]
78. Cieslak, C.; Shah, A.; Clark, B.; Childs, P. Wind-Turbine Inspection, Maintenance and Repair Robotic System. In Proceedings of the Volume 14: Wind Energy, Boston, MA, USA, 26 June 2023; American Society of Mechanical Engineers; p. V014T37A004. Available online: <https://asmedigitalcollection.asme.org/GT/proceedings/GT2023/87127/V014T37A004/1168596> (accessed on 10 September 2024).
79. Zhang, H.; Li, J.; Kang, F.; Zhang, J. Monitoring and Evaluation of the Repair Quality of Concrete Cracks Using Piezoelectric Smart Aggregates. *Constr. Build. Mater.* **2022**, *317*, 125775. [CrossRef]
80. Erol, S.; Erol, B.; Ayan, T. A General Review Of The Deformation Monitoring Techniques And A Case Study: Analysing Deformations Using Gps/Levelling. In Proceedings of the XXth ISPRS Congress Technical Commission VII, Istanbul, Turkey, 12 July 2004; pp. 622–627.
81. Hong, K.; Wang, H.; Yuan, B.; Wang, T. Multiple Defects Inspection of Dam Spillway Surface Using Deep Learning and 3D Reconstruction Techniques. *Buildings* **2023**, *13*, 285. [CrossRef]

82. Bigman, D.P. Evaluating the Feasibility of Robotic Crawler Deployed Ground Penetrating Radar to Assess Repairs of a Concrete Hydroelectric Dam Spillway in Alabama, USA. *Water* **2023**, *15*, 1858. [CrossRef]
83. Pinto, M.F.; Honorio, L.M.; Melo, A.; Marcato, A.L.M. A Robotic Cognitive Architecture for Slope and Dam Inspections. *Sensors* **2020**, *20*, 4579. [CrossRef]
84. Patil, S.; Jacob, A.A.; Upadhyay, V.; Jhunjhunwala, T.; Balasubramanian, K. Robotic Ultrasonic Pulse Velocimetry for Submerged Concrete Assets. *J. Non Destr. Test. Eval.* **2023**, *20*, 58–66.
85. Sakagami, N.; Yumoto, Y.; Takebayashi, T.; Kawamura, S. Development of Dam Inspection Robot with Negative Pressure Effect Plate. *J. Field Robot.* **2019**, *36*, 1422–1435. [CrossRef]
86. Kohut, P.; Giergiel, M.; Cieslak, P.; Ciszewski, M.; Buratowski, T. Underwater Robotic System for Reservoir Maintenance. *J. Vibroeng.* **2016**, *18*, 3757–3767. [CrossRef]
87. Li, Y.; Bao, T.; Xu, B.; Shu, X.; Zhou, Y.; Du, Y.; Wang, R.; Zhang, K. A Deep Residual Neural Network Framework with Transfer Learning for Concrete Dams Patch-Level Crack Classification and Weakly-Supervised Localization. *Measurement* **2022**, *188*, 110641. [CrossRef]
88. Li, X.; Sun, H.; Song, T.; Zhang, T.; Meng, Q. A Method of Underwater Bridge Structure Damage Detection Method Based on a Lightweight Deep Convolutional Network. *IET Image Process.* **2022**, *16*, 3893–3909. [CrossRef]
89. Chen, D.; Huang, B.; Kang, F. A Review of Detection Technologies for Underwater Cracks on Concrete Dam Surfaces. *Appl. Sci.* **2023**, *13*, 3564. [CrossRef]
90. Qi, Z.; Liu, D.; Zhang, J.; Chen, J. Micro-Concrete Crack Detection of Underwater Structures Based on Convolutional Neural Network. *Mach. Vis. Vision. Appl.* **2022**, *33*, 74. [CrossRef]
91. Shi, P.; Fan, X.; Ni, J.; Wang, G. A Detection and Classification Approach for Underwater Dam Cracks. *Struct. Health Monit.* **2016**, *15*, 541–554. [CrossRef]
92. Hazel, B.; Côté, J.; Laroche, Y.; Mongenot, P. Field Repair and Construction of Large Hydropower Equipment with a Portable Robot. *J. Field Robot.* **2012**, *29*, 102–122. [CrossRef]
93. Thuot, D.; Liu, Z.; Champliand, H.; Beaudry, J.; Richard, P.-L.; Blain, M. Remote Robotic Underwater Grinding System and Modeling for Rectification of Hydroelectric Structures. *Robot. Comput. -Integr. Manuf.* **2013**, *29*, 86–95. [CrossRef]
94. Ekren, N.; Karagöz, Z.; Şahin, M. A Review of Line Suspended Inspection Robots for Power Transmission Lines. *J. Electr. Eng. Technol.* **2024**, *19*, 2549–2583. [CrossRef]
95. Ahmed, F.; Mohanta, J.C.; Keshari, A. Power Transmission Line Inspections: Methods, Challenges, Current Status and Usage of Unmanned Aerial Systems. *J. Intell. Robot. Syst.* **2024**, *110*, 54. [CrossRef]
96. Jing, Y.Z.; Liao, H.J.; Zhang, C.H.; Peng, T.; Liao, Z.Z.; Liu, F. Design of auto detection system for high-voltage insulator inspection robot. In Proceedings of the 2017 International Conference on Robotics and Automation Sciences (ICRAS), Hong Kong, China, 26–29 August 2017. [CrossRef]
97. Wale, P.B.; Kamal, S.K. Maintenance of transmission line by using robot. In Proceedings of the 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, India, 9–10 September 2016. [CrossRef]
98. Zhou, C.; Liu, Y.; Rui, X. Mechanism and characteristic of rain-induced vibration on high-voltage transmission line. *J. Mech. Sci. Technol.* **2012**, *26*, 2505–2510. [CrossRef]
99. Zhou, C.; Liu, Y.; Ma, Z. Investigation on aerodynamic instability of high-voltage transmission lines under rain-wind condition. *J. Mech. Sci. Technol.* **2015**, *29*, 131–139. [CrossRef]
100. Zhou, C.; Liu, Y. Modeling and mechanism of rain-wind induced vibration of bundled conductors. *Shock. Vib.* **2016**, *2016*, 1038150. [CrossRef]
101. Wu, G.; Xu, X.; Xiao, H.; Dai, J.; Xiao, X.; Huang, Z.; Ruan, L. A novel self-navigated inspection robot along high-voltage power transmission line and its application. In Proceedings of the Intelligent Robotics and Applications First International Conference (ICIRA 2008), Wuhan, China, 15–17 October 2008. [CrossRef]
102. Wu, G.; Cao, H.; Xu, X.; Xiao, H.; Li, S.; Xu, Q.; Liu, B.; Wang, Q.; Wang, Z.; Ma, Y. Design and application of inspection system in a self-governing mobile robot system for high voltage transmission line inspection. In Proceedings of the 2009 Asia-Pacific Power and Energy Engineering Conference, Wuhan, China, 27–31 March 2009. [CrossRef]
103. Zhu, X.; Wang, H.; Fang, L.; Zhao, M.; Zhou, J. Algorithm research of inspection robot for searching for pose of overhead ground wires. In Proceedings of the 6th World Congress on Intelligent Control and Automation, Dalian, China, 21–23 June 2006. [CrossRef]
104. Toussaint, K.; Pouliot, N.; Montambault, S. Transmission line maintenance robots capable of crossing obstacles: State-of-the-art review and challenges ahead. *J. Field Robot.* **2009**, *26*, 477–499. [CrossRef]
105. Alhassan, A.B.; Zhang, X.; Shen, H.; Xu, H. Power transmission line inspection robots: A review, trends and challenges for future research. *Int. J. Electr. Power Energy Syst.* **2020**, *118*, 105862. [CrossRef]
106. CIGRE. The use of robotics in assessment and maintenance of overhead lines. In Proceedings of the Technical Brochure of 13th annual CIGRE Canada Conference on Power Systems, Calgary, AB, Canada, 15–18 October 2018.
107. Nayerloo, M.; Chen, X.; Wang, W.; Chase, J. Cable-climbing robots for power transmission lines inspection. In *Mobile Robots—State of the Art in Land, Sea, Air, and Collaborative Missions*; Chen, X.Q., Chen, Y.Q., Chase, J.G., Eds.; IntechOpen: London, UK, 2009; Chapter 4; pp. 63–84. [CrossRef]

108. Elizondo, D.; Gentile, T.; Candia, H.; Bell, G. Overview of robotic applications for energized transmission line work—Technologies, field projects and future developments. In Proceedings of the 2010 1st International Conference on Applied Robotics for the Power Industry, Montreal, QC, Canada, 5–7 October 2010. [CrossRef]
109. Gonçalves, R.S.; Carvalho, J.C.M. Review and latest trends in mobile robots used on power transmission lines. *Int. J. Adv. Robot. Syst.* **2013**, *10*, 1–14. [CrossRef]
110. Pagnano, A.; Höpf, M.; Teti, R. A roadmap for automated power line inspection. Maintenance and repair. *Procedia CIRP* **2013**, *12*, 234–239. [CrossRef]
111. Seok, K.H.; Kim, Y.S. A state of the art of power transmission line maintenance robots. *J. Electr. Eng. Technol.* **2016**, *11*, 1412–1422. [CrossRef]
112. Li, L.; Li, N.; Li, Y.; Zhang, B.; Zhao, J.; Zhang, C.; Dai, Z. A state-of-the-art survey of the robotics applied for the power industry in China. In Proceedings of the 2016 4th International Conference on Applied Robotics for the Power Industry (CARPI), Jinan, China, 11–13 October 2016. [CrossRef]
113. Phillips, A.; Engdahl, E.; McGuire, D.; Major, M.; Bartlett, G. Autonomous overhead transmission line inspection robot (TI) development and demonstration. In Proceedings of the 2012 2nd International Conference on Applied Robotics for the Power Industry (CARPI), Zurich, Switzerland, 11–13 September 2012. [CrossRef]
114. Zheng, L.; Yi, R. Obstacle negotiation control for a mobile robot suspended on overhead ground wires by optoelectronic sensors. In Proceedings of the 2009 International Conference on Optical Instruments and Technology: Advanced Sensor Technologies and Applications, Shanghai, China, 19–22 October 2009. [CrossRef]
115. Guanghong, T.; Lijin, F. A multi-unit serial inspection robot for power transmission lines. *Ind. Robot.* **2019**, *46*, 223–234. [CrossRef]
116. Debenest, P.; Guarnieri, M.; Takita, K.; Fukushima, E.F.; Hirose, S.; Tamura, K.; Kimura, A.; Kubokawa, H.; Iwama, N.; Shiga, F.; et al. Toward a Practical Robot for Inspection of High-Voltage Lines. In *Field and Service Robotics*; Springer Tracts in Advanced Robotics; Howard, A., Iagnemma, K., Kelly, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 62, pp. 45–55. [CrossRef]
117. Chen, Y.; Fang, Z.; Liu, S.; Wang, Y.; Zhong, C.; Cai, C.; Zhang, Y.; Wei, Y.; Wang, Z. A soft-robotic gripper for ultra-high-voltage transmission line operations. In Proceedings of the 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2), Wuhan, China, 30 October–1 November 2020. Available online: <https://ieeexplore.ieee.org/document/9347270> (accessed on 10 September 2024).
118. Nie, Y.; Jiao, C.; Fan, Y. Active shielding design of patrol robot wireless charging system. In Proceedings of the 2019 IEEE 3rd International Electrical and Energy Conference (CIEEC), Beijing, China, 7–9 September 2019. [CrossRef]
119. Sawada, J.; Kusumoto, K.; Maikawa, Y.; Munakata, T.; Ishikawa, Y. A mobile robot for inspection of power transmission lines. *IEEE Trans. Power Deliv.* **1991**, *6*, 309–315. [CrossRef]
120. Zhu, X.; Wang, H.; Fang, L.; Zhao, M.; Zhou, J. Dual arms running control method of inspection robot based on obliquitous sensor. In Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, 9–15 October 2006. [CrossRef]
121. Dian, S.; Wen, X.; Dong, H.; Weng, T. Development of a self-balance dual-arm robot for inspection of high-voltage power transmission lines. In Proceedings of the 2012 IEEE International Conference on Mechatronics and Automation, Chengdu, China, 5–8 August 2012. [CrossRef]
122. Katrašnik, J.; Pernuš, F.; Likar, B. A climbing-flying robot for power line inspection. In *Climbing and Walking Robots*; Miripour, B., Ed.; IntechOpen: London, UK, 2010; Chapter 6; pp. 95–110. [CrossRef]
123. Chang, W.; Yang, G.; Yu, J.; Liang, Z.; Cheng, L.; Zhou, C. Development of a power line inspection robot with hybrid operation modes. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 24–28 September 2017. [CrossRef]
124. Mirallès, F.; Hamelin, P.; Lambert, G.; Lavoie, S.; Pouliot, N.; Montfrond, M.; Montambault, S. LineDrone technology: Landing an unmanned aerial vehicle on a power line. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018. [CrossRef]
125. Iftikhar, H.; Sarquis, E.; Branco, P.J.C. Why Can Simple Operation and Maintenance (O&M) Practices in Large-Scale Grid-Connected PV Power Plants Play a Key Role in Improving Its Energy Output? *Energies* **2021**, *14*, 3798. [CrossRef]
126. Keisang, K.; Bader, T.; Samikannu, R. Review of operation and maintenance methodologies for solar photovoltaic microgrids. *Front. Energy Res.* **2021**, *9*, 730230. [CrossRef]
127. Spencer, B.F.; Hoskere, V.; Narazaki, Y. Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. *Engineering* **2019**, *5*, 199–222, ISSN 2095-8099. [CrossRef]
128. Zou, J.T.; Rajveer, G.V. Drone-based solar panel inspection with 5G and AI Technologies. In Proceedings of the 2022 8th International Conference on Applied System Innovation (ICASI), Nantou, Taiwan, 22–23 April 2022; pp. 174–178. [CrossRef]
129. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, Canada, 7–12 December 2015. Available online: https://papers.nips.cc/paper_files/paper/2015 (accessed on 18 July 2024).
130. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016. [CrossRef]

131. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [CrossRef]
132. Dai, J.; Li, Y.; He, K.; Sun, J. R-fcn: Object detection via region-based fully convolutional networks. In Proceedings of the Conference on Advances in Neural Information Processing Systems 29 (NIPS 2016), Barcelona, Spain, 5–10 December 2016. Available online: https://papers.nips.cc/paper_files/paper/2016 (accessed on 18 July 2024).
133. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [CrossRef]
134. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI conference on artificial intelligence (AAAI-17), San Francisco, CA, USA, 4–9 February 2017. Available online: <https://aaai.org/papers/11231-aaai-31-2017/> (accessed on 22 July 2024).
135. Liu, Z.; Li, X.; Luo, P.; Loy, C.C.; Tang, X. Semantic image segmentation via deep parsing network. In Proceedings of the IEEE International conference on computer vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1377–1385. [CrossRef]
136. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2961–2969. [CrossRef]
137. Bhujade Rajeev, M.; Adithya, V.; Hrishikesh, S.; Balamurali, P. Detection of power-lines in complex natural surroundings. In Proceedings of the Second International Conference on Information Technology Convergence and Services (ITCSE 2013), Bangalore, India, 20–21 December 2013. Available online: <https://airccj.org/CSCP/vol3/csit3910.pdf> (accessed on 22 July 2024).
138. Li, Z.; Liu, Y.; Hayward, R.; Zhang, J.; Cai, J. Knowledge-based power line detection for UAV surveillance and inspection systems. In Proceedings of the 2008 23rd International Conference Image and Vision Computing New Zealand, Christchurch, New Zealand, 26–28 November 2008. Available online: <https://ieeexplore.ieee.org/document/4762118> (accessed on 10 September 2024).
139. Peeters, C.; Guillaume, P.; Helsen, J. Vibration-based bearing fault detection for operations and maintenance cost reduction in wind energy. *Renew. Energy* **2018**, *116*, 74–87. [CrossRef]
140. Huang, B.; Zhao, S.; Kang, F. Image-based automatic multiple-damage detection of concrete dams using region-based convolutional neural networks. *J. Civil. Struct. Health Monit.* **2023**, *13*, 413–429. [CrossRef]
141. Li, L.; Zhang, H.; Pang, J.; Huang, J. Dam surface crack detection based on deep learning. In Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence, Shanghai, China, 20–22 September 2019; pp. 738–743.
142. Osman, A.I.; Chen, L.; Yang, M.; Msigwa, G.; Farghali, M.; Fawzy, S.; Rooney, D.W.; Yap, P.S. Cost, environmental impact, and resilience of renewable energy under a changing climate: A review. *Environ. Chem. Lett.* **2023**, *21*, 741–764. [CrossRef]
143. Zhao, S.Z.; Kang, F.; Li, J.J.; Ma, C.B. Structural health monitoring and inspection of dams based on UAV photogrammetry with image 3D reconstruction. *Automat. Constr.* **2021**, *130*, 103832. [CrossRef]
144. De Vito, S.; Del Giudice, A.; Di Francia, G. Electric Transmission and Distribution Network Air Pollution. *Sensors* **2024**, *24*, 587. [CrossRef]
145. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014. [CrossRef]
146. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 17), Long Beach, California, USA, 4–9 December 2017. Available online: <https://dl.acm.org/doi/10.5555/3295222.3295349> (accessed on 10 September 2024).
147. Schonberger, J.L.; Frahm, J.M. Structure-from-Motion Revisited. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [CrossRef]
148. Seitz, S.M.; Curless, B.; Diebel, J.; Scharstein, D.; Szeliski, R. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006. Available online: <https://ieeexplore.ieee.org/document/1640800> (accessed on 10 September 2024).
149. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009. [CrossRef]
150. Krizhevsky, A. Learning Multiple Layers of Features from Tiny Images. Technical Report TR-2009, University of Toronto. 2009. Available online: <https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf> (accessed on 19 July 2024).
151. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the 13th European Conference, Zurich, Switzerland, 6–12 September 2014. [CrossRef]
152. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
153. Scharstein, D.; Szeliski, R. High-accuracy Stereo Depth Maps Using Structured Light. In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Madison, WI, USA, 18–20 June 2003. Available online: <https://ieeexplore.ieee.org/document/1211354> (accessed on 10 September 2024).
154. Jensen, R.; Dahl, A.L.; Vogiatzis, G.; Tola, E.; Aanæs, H. Large Scale Multi-view Stereopsis Evaluation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014. [CrossRef]

155. Saxena, A.; Goebel, K.; Simon, D.; Eklund, N. Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation. In Proceedings of the 1st International Conference on Prognostics and Health Management (PHM08), Denver, CO, USA, 6–9 October 2008. [CrossRef]
156. Purohit, H.; Tanabe, R.; Ichige, T.; Endo, T.; Nikaido, Y.; Suefusa, K.; Kawaguchi, Y. MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection. In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019), New York, NY, USA, 25–26 October 2019. Available online: <https://zenodo.org/record/3384388> (accessed on 19 July 2024). [CrossRef]
157. Tao, X.; Zhang, D.; Wang, Z.; Liu, X.; Zhang, H.; Xu, D. Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks. *IEEE Trans. Syst. Man. Cybern. Syst.* **2020**, *50*, 1486–1498. [CrossRef]
158. Tomaszewski, M.; Ruszczak, B.; Michalski, P. The collection of images of an insulator taken outdoors in varying lighting conditions with additional laser spots. *Data Brief.* **2018**, *18*, 765–768. [CrossRef]
159. Bian, J.; Hui, X.; Zhao, X.; Tan, M. A monocular vision-based perception approach for unmanned aerial vehicle close proximity transmission tower inspection. *Int. J. Adv. Robot. Syst.* **2019**, *16*, 1729881418820227. [CrossRef]
160. Yetgin, Ö.E.; Gerek, Ö.N.; Nezir, Ö. Ground truth of powerline dataset (infrared-ir and visible light-vl). *Mendeley Data* **2017**, *8*.
161. Bradbury, K.; Saboo, R.; Johnson, T.L.; Malof, J.M.; Devarajan, A.; Zhang, W.; Collins, L.M.; Newell, R.G. Distributed Solar Photovoltaic Array Location and Extent Data Set for Remote Sensing Object Identification. *Sci. Data* **2016**, *3*, 160106. [CrossRef]
162. DKASC; Alice Springs; DKA Solar Centre. Available online: <http://dkasolarcentre.com.au/locations/alice-springs> (accessed on 19 July 2024).
163. ESOLMET-IER Instituto de Energías Renovables. Available online: http://esolmet.ier.unam.mx/Tipos_consulta.php (accessed on 19 July 2024).
164. Sengupta, M.; Xie, Y.; Lopez, A.; Habte, A.; Maclaurin, G.; James, S. The national solar radiation data base (NSRDB). *Renew. Sustain. Energy Rev.* **2018**, *89*, 51–60. [CrossRef]
165. Pierdicca, R.; Paolanti, M.; Felicetti, A.; Piccinini, F.; Zingaretti, P. Automatic Faults Detection of Photovoltaic Farms: SolAIr, a Deep Learning-Based System for Thermal Images. *Energies* **2020**, *13*, 6496. [CrossRef]
166. The Photovoltaic Thermal Images Dataset. Available online: <https://vrai.dii.univpm.it/content/photovoltaic-thermal-images-dataset> (accessed on 19 July 2024).
167. Jung, J.H.; Lim, D.G. Industrial robots, employment growth, and labor cost: A simultaneous equation analysis. *Technol. Forecast. Soc. Chang.* **2020**, *159*, 120202. [CrossRef]
168. IEA. *World Energy Employment 2022*; IEA: Paris, France, 2022. Available online: <https://www.iea.org/reports/world-energy-employment-2022> (accessed on 10 September 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

A Novel Nonsingular Fast Terminal Sliding Mode Control with Sliding Mode Disturbance Observer for Permanent Magnet Synchronous Motor Servo Control

Difen Shi ¹, Kai Bodemann ², Yao Wang ^{3,*}, Changliang Xu ¹, Lulu Liu ¹ and Chungui Feng ⁴

¹ College of Electronic Engineering, Nanjing Xiaozhuang University, Nanjing 211171, China; difenshi@njjust.edu.cn (D.S.); xuchangliang@nuua.edu.cn (C.X.); liululu@njxzc.edu.cn (L.L.)

² Laboratory for Metrology, Measurement and Instrumentation, Faculty of Engineering and Computer Science, Osnabrück University of Applied Sciences, 49076 Osnabrück, Germany; kai.bodemann@hs-osnabrueck.de

³ School of Electrical & Automation, Nantong University, Nantong 226236, China

⁴ School of Mechanical, Electrical and Automotive Engineering, Liuzhou City Vocational College, Liuzhou 545036, China; fcg8823@126.com

* Correspondence: wy_ntu@ntu.edu.cn

Abstract: This article proposes a novel nonsingular fast terminal sliding mode control (N-NFTSMC) with a sliding mode disturbance observer (SDOB) for permanent magnet synchronous motor (PMSM) servo control. Firstly, to reduce the chattering issue, a new sliding mode reaching law (NSRL) is proposed for the N-NFTSMC. Secondly, to further improve the dynamic tracking accuracy, we introduce a sliding disturbance observer to estimate unknown disturbances for feedforward compensation. Comparative simulations via Matlab/Simulink 2018 are conducted using the traditional NFTSMC and N-NFTSMC; the step simulation results show that the chattering phenomenon was suppressed well via the N-NFTSMC scheme. The sine wave tracking simulation proves that the N-NFTSMC has better dynamic tracking performance when compared with traditional NFTSMC. Finally, we carry out experiments to validate that the N-NFTSMC adequately suppresses the chattering issue and possesses better anti-disturbance performance.

Keywords: PMSM; nonsingular terminal sliding mode control; sliding mode reaching law; sliding mode disturbance observer

Citation: Shi, D.; Bodemann, K.; Wang, Y.; Xu, C.; Liu, L.; Feng, C. A Novel Nonsingular Fast Terminal Sliding Mode Control with Sliding Mode Disturbance Observer for Permanent Magnet Synchronous Motor Servo Control. *Processes* **2024**, *12*, 1986. <https://doi.org/10.3390/pr12091986>

Academic Editors: Ján Pitel' and Daniel Fodorean

Received: 25 July 2024

Revised: 6 September 2024

Accepted: 11 September 2024

Published: 14 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Permanent magnet synchronous motors (PMSMs) have been applied in many AC servo systems due to their simple structure and fast response time [1–3]. Therefore, PMSMs are widely used in modern industrial fields like robot manipulators, pneumatic servo systems [4], electric vehicles (EVs) and spacecraft. The traditional proportional–integral–derivative (PID) controller is usually used in PMSMs because its parameters are easily adjustable [5]. However, the PMSM system contains many unknown dynamics that will worsen the control performance of the PMSM. The PID controller has difficulties meeting the requirements needed for the high-precision control performance of PMSMs. Therefore, it is important to develop advanced control methods to deal with these problems. To tackle these issues, plenty of novel control methods have been proposed, for example, active disturbance control (ADRC) [6], sliding mode control (SMC) [7] and model predictive control [8].

Among these approaches, SMC has a performance advantage due to its robustness [9–13]. However, there are still some problems that need to be solved. The system error for the typical linear SMC is not capable of finite-time convergence, and the chattering phenomenon must be suppressed. Y. Sun's novel nonsingular terminal sliding mode control (NTSMC) is an approach that guarantees system stabilization within a bounded time interval [14]. A full-order nonsingular terminal sliding mode control method is proposed in [15] for dynamic

systems subject to both matched and mismatched disturbances. The chattering problem may cause negative effects and destabilizes the system in real applications [16]. Different strategies have been developed to reduce the chattering phenomenon, such as high-order SMC [17,18] and advanced sliding mode reaching law in [19–23]. Among these methods, the advanced sliding mode reaching law can decrease the chattering issue effectively. It is contradictory to overcome heavy chattering and obtain a fast response at same time. Therefore, an adaptive neural network nonsingular fast terminal sliding mode control (ANNNFTSMC), which has advantages that include a fast response speed and a small static error, is investigated in [24].

To enhance the anti-disturbance performance, the disturbance observer has been widely studied [25–27]. A time-varying nonlinear disturbance observer (TVNDO), as specified in [28], was designed to evaluate unknown disturbances. There are many nonlinear disturbances in practical engineering applications; however, PI control cannot meet the demands of dynamic performance. Based on this problem, an improved model-free continuous super-twisting NFTSM is applied for IPMSM in [29]. Considering that disturbances are usually changing in engineering environments, a DOB, which deals with time-varying disturbances effectively, is designed and analyzed in [30,31].

Although NFTSMC has a high convergent speed, it still has chattering issues. Based on the analysis above, an effective way to decrease the chattering issue is to modify the sliding mode reaching law. We have devised an NSMRL to reduce chattering. When applied to the motion control of PMSMs, the tracking accuracy under complicated working conditions can be improved. In this paper, we propose a novel NFTSMC with a sliding mode observer (SDOB); the control structure can be seen in Figure 1. The contribution of this study can be summarized as follows:

- (1) To reduce the chattering problem in the traditional NFTSMC, an NSMRL is designed for the N-NFTSMC. The NSMRL with a smooth switch function eliminates chattering behavior effectively.
- (2) An SDOB is proposed to estimate the unknown disturbances of the PMSM system. The N-NFTSMC with the SDOB is designed to improve the anti-disturbance performance of the system.
- (3) Simulation and experimental verification are conducted to prove the effectiveness of the N-NFTSMC and to compare it with the classic NFTSMC.

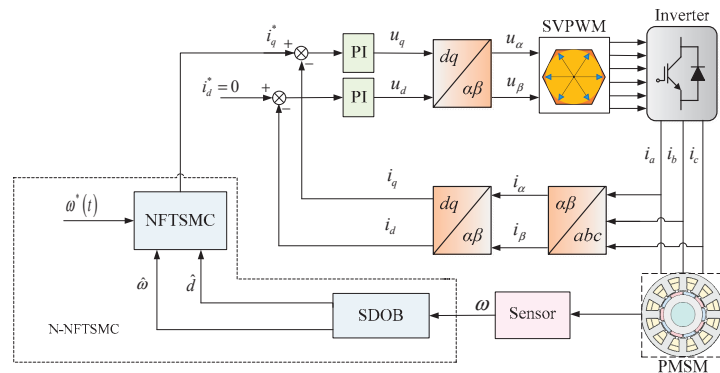


Figure 1. Control structure of the PMSM.

The remainder of this paper is organized as follows: Section 2 states the traditional NFTSMC design and problem formulation. Section 3 states the proposed N-NFTSMC. Section 4 details the design of the SDOB and its convergence analysis. Section 5 verifies the effectiveness of the proposed control method using a Simulink simulation. Section 6 describes experiments to validate the proposed N-NFTSMC.

2. NFTSMC Design

2.1. Modelling of PMSM

Based on the theoretical speculations regarding the control of permanent magnet synchronous motors (PMSMs), with the d-axis current set at 0, the d–q axis mathematical model is presented as follows:

$$\begin{cases} \dot{i}_q = -\frac{R}{L}i_q - n\omega i_d - \frac{n\psi_f}{L} + \frac{u_q}{L} \\ \dot{\omega} = \frac{1}{J}(K_e i_q - T_L - B\omega) \end{cases} \quad (1)$$

where u_q is the stator voltage; i_d and i_q denote the currents within the stator; L is the stator inductance; ψ_f stands for the magnetic flux; R signifies the stator resistance; n denotes the pole-pair counts of the PMSM; ω is rotational speed; J indicates the rotor's inertial moment; K_e represents the motor's torque coefficient ($T_e = \frac{3}{2}n\psi_f i_q = K_e i_q$); T_L represents the torque applied by the load; and B denotes the friction coefficient.

2.2. Definition of State Variables

If the reference speed is defined as ω^* and the real speed is defined as ω , then the error between ω^* and ω can be obtained by

$$\begin{cases} e = \omega^* - \omega \\ \dot{e} = -\dot{\omega} \end{cases} \quad (2)$$

In the experimental verification, T_L is also set at a constant value, which can be obtained by $\dot{T}_L = 0$. Then, the derivative of Equation (2) is given as

$$\begin{bmatrix} \dot{e} \\ \ddot{e} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{b_m}{J} \end{bmatrix} \begin{bmatrix} \dot{e} \\ \ddot{e} \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{1.5p_n\psi}{J} \end{bmatrix} \dot{i}_q \quad (3)$$

The state variables can be summarized as

$$\begin{bmatrix} e \\ \dot{e} \\ \ddot{e} \end{bmatrix} = \begin{bmatrix} \omega^* - \omega \\ -\frac{1.5p_n\psi}{J} \dot{i}_q + \frac{b_m}{J} \omega + \frac{T_f}{J} \\ -\frac{1.5p_n\psi}{J} \dot{i}_q + \frac{b_m}{J} \dot{\omega} \end{bmatrix} \quad (4)$$

2.3. Classic Nonsingular Fast Terminal Sliding Mode Control

The classic nonsingular fast terminal sliding mode surface is set forth as

$$s = e + k_1|e|^{\sigma_1}\text{sign}(e) + k_2|\dot{e}|^{\sigma_2}\text{sign}(\dot{e}) \quad (5)$$

The appropriate sliding mode reaching law ensures that the system variable states convergence to the designed sliding surface in a finite time. The exponential sliding mode control law is similarly formulated as follows:

$$\dot{s} = -\varepsilon\text{sign}(s) - ks, \quad \varepsilon > 0, k > 0. \quad (6)$$

The derivation of the sliding mode surface of (5) can be written as:

$$\begin{aligned} \dot{s} &= \dot{e} + \sigma_1 k_1 |e|^{\sigma_1-1} \text{sign}(e) \dot{e} + k_2 \sigma_2 |\dot{e}|^{\sigma_2-1} \text{sign}(\dot{e}) \\ &= \left(1 + \sigma_1 k_1 |e|^{\sigma_1-1} \text{sign}(e)\right) \dot{e} + k_2 \sigma_2 |\dot{e}|^{\sigma_2-1} \text{sign}(\dot{e}) \end{aligned} \quad (7)$$

Combined with (4), (5) and (7), the output of the classic NFTSMC is generated according to the subsequent formula:

$$\dot{i}_q = \frac{J}{1.5p_n\psi} \cdot \frac{1}{k_2\sigma_2|\dot{e}|^{\sigma_2-1}\text{sign}(\dot{e})} \int_0^t \left[\left(1 + \sigma_1 k_1 |e|^{\sigma_1-1} \text{sign}(e) - \frac{b_m}{J} k_2 \sigma_2 |\dot{e}|^{\sigma_2-1} \text{sign}(\dot{e})\right) \dot{e} + \varepsilon \text{sign}(s) + ks \right] dt \quad (8)$$

3. N-NFTSMC Design

The overall block diagram of the N-NFTSMC controller is shown in Figure 2. It depicts the connections between the mathematical models.

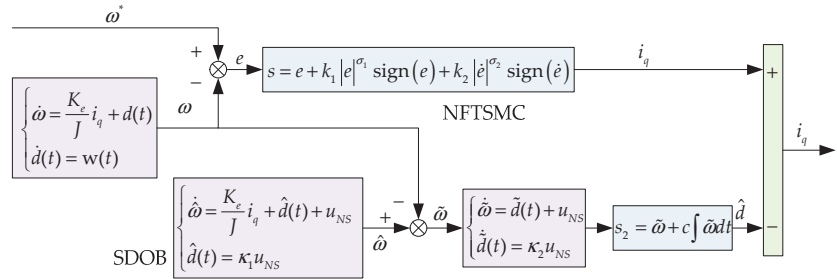


Figure 2. The block diagram of N-NFTSMC.

3.1. Controller Design

Equation (6) indicates that an increase in the coefficients leads to an improved system convergence speed. On the other hand, the process of convergence will also cause chattering. Therefore, it is difficult to maintain the balance between the convergence speed and the amount of chattering using the traditional NFTSMC. The main reason is that the reaching law includes a switching function $\text{sign}(\cdot)$. To cope with this issue, we replace $\text{sign}(\cdot)$ with $\tanh(\cdot)$, which can provide a smooth transition. A comparison of the two functions is shown in Figure 3.

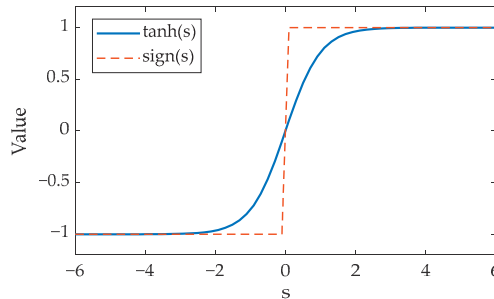


Figure 3. Function $\tanh(s)$ and $\text{sign}(s)$.

To reduce the chattering, a new sliding mode reaching law (NSMRL) based on the previously mentioned sliding mode reaching law in Equation (6) is designed as follows:

$$\begin{cases} \dot{s} = -\varepsilon_1 \frac{s}{1+|s|} - \varepsilon_2 |s|^{\alpha_1} \tanh(s) - \varepsilon_3 |s|^{\alpha_2} \tanh(s) \\ \varepsilon_1 > 0, \varepsilon_2 > 0, \varepsilon_3 > 0, 0 < \alpha_1 < 1, \alpha_2 > 1. \end{cases} \quad (9)$$

It can be learned from Equation (9) that when the system state is far away from the sliding mode surface s , the terms $\varepsilon_1 \frac{s}{1+|s|}$ and $\varepsilon_3 |s|^{\alpha_2} \tanh(s)$ will accelerate the state convergence to the sliding mode. When the state is close to the sliding mode surface, the term $\varepsilon_2 |s|^{\alpha_1} \tanh(s)$ works and makes the states reach the sliding mode surface. In Figure 4, there are three sets of reaching laws. As depicted in Figure 4, the red line, which denotes the NSMRL, achieves convergence with the greatest speed despite being distant from equilibrium.

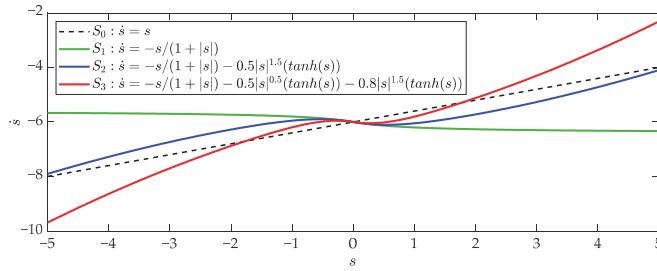


Figure 4. Comparison of the variants of the different sliding mode reaching laws.

The NSMRL in (9) is equal to the derivation of (5), combined with (7). The control law of the N-NFTSMC is presented below:

$$i_q = \frac{J}{1.5p_n\psi} \cdot \frac{1}{k_2\sigma_2|\dot{e}|^{\sigma_2-1}\text{sign}(\dot{e})} \int_0^t \left[\left(1 + \sigma_1 k_1 |e|^{\sigma_1-1} \text{sign}(e) - \frac{b_m}{J} \sigma_2 k_2 |\dot{e}|^{\sigma_2-1} \text{sign}(\dot{e}) \right) \dot{e} + \varepsilon_1 \frac{s}{1+|s|} + \varepsilon_2 |s|^{\alpha_1} \tanh(s) + \varepsilon_3 |s|^{\alpha_2} \tanh(s) \right] dt \quad (10)$$

3.2. Stability Analysis for Novel NFTSMC

The Lyapunov function Equation (11) is used to analyze the stability of the N-NFTSMC:

$$V_1 = \frac{1}{2} s^2 \quad (11)$$

Combining (5), (9) and (10), the derivation of (11) can be obtained as

$$\begin{aligned} \dot{V}_1 &= s \left(\dot{e} + \sigma_1 k_1 |e|^{\sigma_1-1} \text{sign}(e) \dot{e} + k_2 \sigma_2 |\dot{e}|^{\sigma_2-1} \text{sign}(\dot{e}) \dot{e} \right) \\ &= s \left\{ \dot{e} + \sigma_1 k_1 |e|^{\sigma_1-1} \text{sign}(e) \dot{e} + k_2 \sigma_2 |\dot{e}|^{\sigma_2-1} \text{sign}(\dot{e}) \left[-\frac{1}{k_2 \sigma_2 |\dot{e}|^{\sigma_2-1} \text{sign}(\dot{e})} \left(1 + \sigma_1 k_1 |e|^{\sigma_1-1} \text{sign}(e) \dot{e} - \frac{b_m}{J} k_2 \sigma_2 |\dot{e}|^{\sigma_2-1} \text{sign}(\dot{e}) - \dot{s} \right) + \frac{b_m}{J} \dot{\omega} \right] \right\} \quad (12) \\ &= s \left(-\varepsilon_1 \frac{s}{1+|s|} - \varepsilon_2 |s|^{\alpha_1} \tanh(s) - \varepsilon_3 |s|^{\alpha_2} \tanh(s) \right) \\ &= -\varepsilon_1 \frac{s^2}{1+|s|} - \varepsilon_2 |s|^{\alpha_1} s \cdot \tanh(s) - \varepsilon_3 |s|^{\alpha_2} s \cdot \tanh(s) < 0 \end{aligned}$$

4. Sliding Mode Disturbance Observer

To overcome the chattering problem with sliding mode control, this section proposes a new sliding mode disturbance observer. The convergence of the SDOB is proved in Section 4.2.

4.1. SDOB Design

By defining $d(t) = \frac{1}{J}(-T_L - B\omega)$, the mathematical model of PMSM (Equation (1)) can be simplified as

$$\begin{cases} \dot{\omega} = \frac{K_e}{J} i_q + d(t) \\ \dot{d}(t) = w(t) \end{cases} \quad (13)$$

According to the motion equation of the PMSM, we can model the sliding mode observer (SDOB) as follows:

$$\begin{cases} \dot{\hat{\omega}} = \frac{K_e}{J} i_q + \hat{d}(t) + u_{NS} \\ \dot{\hat{d}}(t) = \kappa_1 u_{NS} \end{cases} \quad (14)$$

Equation (14) minus (13) results in the error equation of the SDOB:

$$\begin{cases} \dot{\tilde{\omega}} = \tilde{d}(t) + u_{NS} \\ \dot{\tilde{d}}(t) = \kappa_2 u_{NS} \end{cases} \quad (15)$$

where $\kappa_2 = \kappa_1 - w(t)$. The sliding mode of the SDOB is designed as follows:

$$s_2 = \tilde{\omega} + c \int \tilde{\omega} dt \tag{16}$$

where $\tilde{\omega} = \hat{\omega} - \omega$ and $\tilde{d}(t) = \hat{d}(t) - d(t)$ represent the errors in the estimation of speed and disturbance, respectively. The course of the hyperbolic tangent over time is shown in Figure 2. The reaching law is thus designed as

$$\dot{s}_2 = -\lambda \tanh(s_2) \tag{17}$$

Combined with Equation (15), (16) and (17), u_{NS} can be obtained:

$$u_{NS} = -\lambda \tanh(s_2) - \tilde{d}(t) \tag{18}$$

4.2. Stability Analysis for SDOB

The Lyapunov function is used to prove the stability of the designed SDOB:

$$V_2 = \frac{1}{2} s_2^2 \tag{19}$$

After differentiating V_2 , we can obtain

$$\dot{V}_2 = s_2 \dot{s}_2 = s_2 (-\lambda \tanh(s_2)) \leq 0 \tag{20}$$

where $\lambda > 0$.

5. Simulation Verification

To verify the control performance of the proposed N-NFTSMC, simulation experiments are conducted in this section. Firstly, step response simulation experiments are conducted to verify whether the response time of the system is improved using the novel method. Secondly, sine wave tracking simulation is used to verify the proposed method.

5.1. Step Response Experiment

To verify the improved response time of the N-NFTSMC method, it is compared with the SMC method through a step response simulation. The simulation results can be seen in Figure 5. The SMC method converges to the reference signal at 0.016 s, while the novel NFTSMC method converges to the reference signal at 0.013 s. These results show that the N-NFTSMC method converges faster to the equilibrium state.

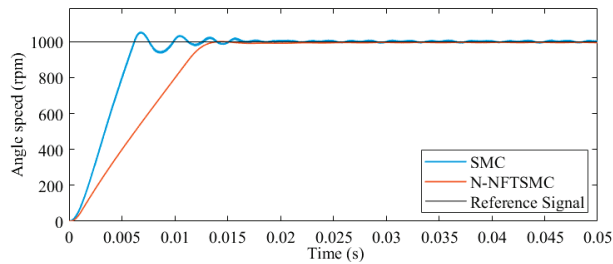


Figure 5. Comparison of the step responses for SMC and N-NFTSMC.

A step response simulation is conducted to verify the suppression of chattering using the N-NFTSMC proposed in this paper. The sliding mode surface parameters are set as $k_1 = 2$ and $k_2 = 0.8$ for the NFTSMC and N-NFTSMC. For the NFTSMC, the reaching law parameters are listed in Table 1. The NFTSMC uses the traditional reaching law (6), and the N-NFTSMC uses the NSMRL (9). The simulation results are shown in Figures 6 and 7.

Table 1. The sliding mode reaching law parameters.

Parameter	Value
ϵ_1	400
ϵ_2	0
α_1	1/2
α_2	3/2

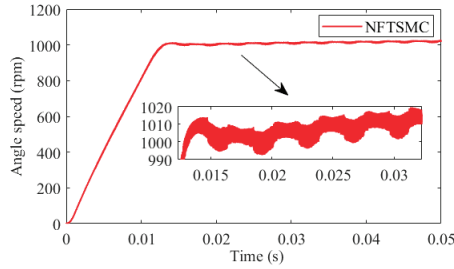


Figure 6. Step response by traditional NFTSMC.

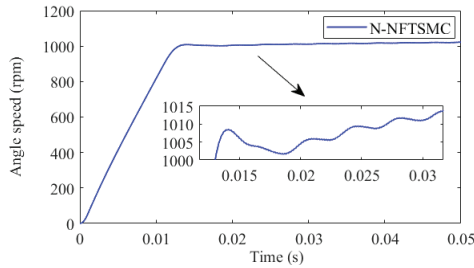


Figure 7. Step response by N-NFTSMC.

Chattering can be seen in Figure 6. Compared with the traditional NFTSMC, the chattering problem is solved by the N-NFTSMC, as shown in Figure 7. This proves the effectiveness of the NSMRL to suppress chattering.

5.2. Sinusoidal Tracking Simulation

To further verify the effectiveness of the N-NFTSMC approach, firstly, a sine wave tracking simulation is conducted in this section using the NFTSMC. A sine wave with an amplitude of 300 is set as the reference signal. The simulation results can be seen in Figure 8. The chattering phenomenon is more significant near the two poles of the sine curve.

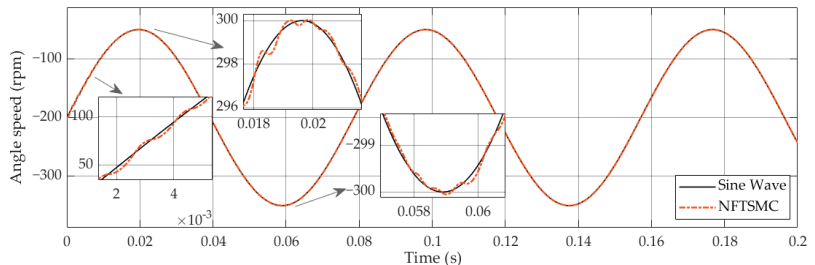


Figure 8. Sine wave tracking by traditional NFTSMC.

Secondly, the same simulation experiment is repeated using the N-NFTSMC approach. The tracking curve using the N-NFTSMC is shown in Figure 9. The tracking curve of N-NFTSMC is smoother and closer to the sine wave compared with the NFTSMC curve.

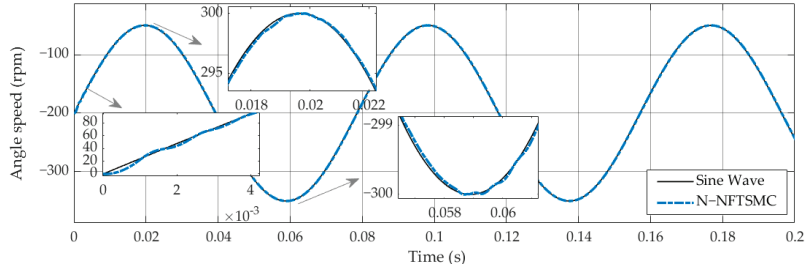


Figure 9. Sine wave tracking by N–NFTSMC.

To compare the simulation results of the two sets of experiments more intuitively, the tracking error curves are shown in Figure 10. The maximum error using the N-NFTSMC method is about 0.61 rpm, while the maximum error using the traditional NFTSMC method is about 0.98 rpm. The sine wave tracking simulation results show that the N-NFTSMC method can track the reference signal more accurately than the NFTSMC method.

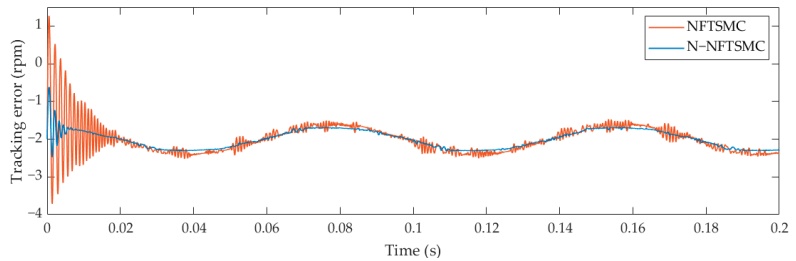


Figure 10. Sine wave tracking error.

The electromagnetic torque curve is also given in Figure 11. It is obvious that the blue line representing the N-NFTSMC is smoother. The sinusoidal tracking simulation result shows that the N-NFTSMC proposed in this paper has a better dynamic tracking performance than the traditional NFTSMC method.

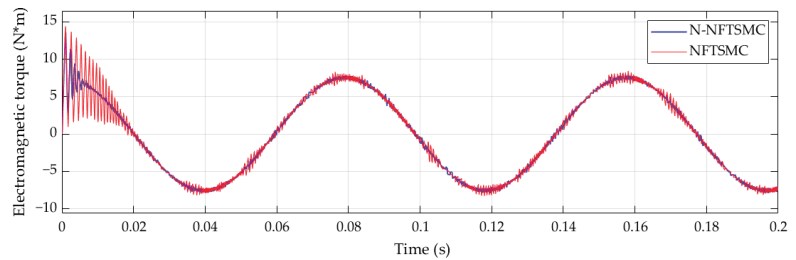


Figure 11. Electromagnetic torque curve.

6. Experimental Verification

To prove the competence of the N-NFTSMC method, semi-physical experiments are implemented on a PMSM servo platform, as portrayed in Figure 12.



1-Computer 2-Controller 3-Tension controller
 4-Driving circuit board 5-Current measurement circuits
 6-Inverter 7-PMSM 8-Position sensor 9-Magnetic particle brake

Figure 12. Experimental setup of PMSM servo system.

The PMSM in this servo system is a four-phase surface-mounted PMSM. The specialized parameters of the PMSM are detailed in Table 2.

Table 2. Parameters of the PMSM.

Parameter	Value
Rated power	200 W
Line resistance	0.33 Ω
Line inductance	9×10^{-4} H
Number of pole pairs	4
Torque coefficient	0.087 Nm/A
Rated voltage	36 V
Rated current	7.5 A

For experimental comparisons, the comparative transient response tests for the NFTSMC and N-NFTSMC control strategies are carried out at 500 rpm and 1000 rpm, respectively. The experimental results of the reference speed, 500 rpm, are shown in Figure 13a. It can be seen that when the motor runs at 500 rpm, after reaching the steady state, the maximum speed fluctuations based on the NFTSMC and N-NFTSMC are about 10 rpm and 6 rpm, respectively. Changing the reference speed to 1000 rpm increases the maximum speed fluctuations based on the NFTSMC and N-NFTSMC to 20 rpm and 12 rpm, as shown in Figure 13b. These results demonstrate that the N-NFTSMC reduces the chattering problem.

In the PMSM servo control, the d-axis is set at 0, and the q-axis is used for output torque control. The chattering of the q-axis voltage directly influences the velocity chattering. Figure 14 shows a comparison of the q-axis voltage using the NFTSMC and N-NFTSMC. With no load torque, the PMSM starts to speed up from 0 to 500 rpm and from 0 to 1000 rpm. It can be learned from the two sets of experiments that the q-axis voltage chattering range is smaller using the N-NFTSMC.

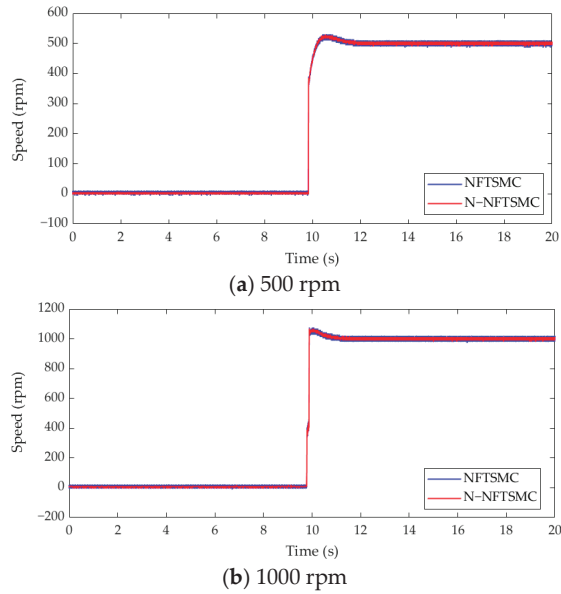


Figure 13. Startup responses of PMSM with NFTSMC and N-NFTSMC: (a) 500 rpm and (b) 1000 rpm.

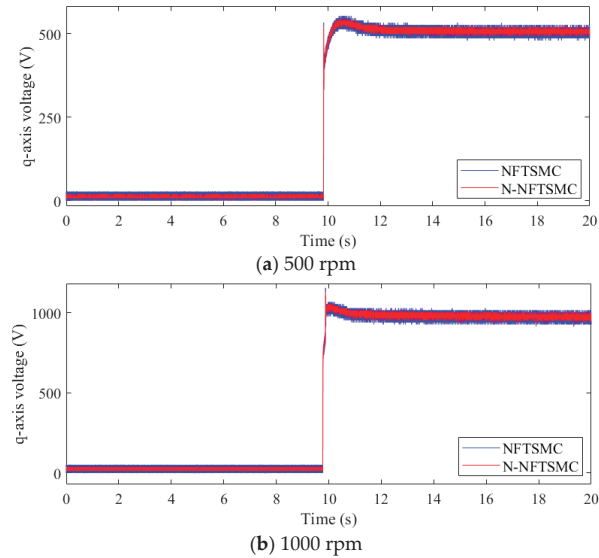


Figure 14. Q-axis voltage for step response: (a) 500 rpm and (b) 1000 rpm.

To demonstrate that the N-NFTSMC possesses better robustness than the NFTSMC, loading experiments are conducted under a speed command of 500 rpm. When the PMSM stably rotates at 500 rpm, we add a sudden load at 10 s using the magnetic powder brake and observe its influence on the PMSM velocities. As can be seen in Figure 15, the N-NFTSMC showed a better anti-disturbance performance than the NFTSMC.

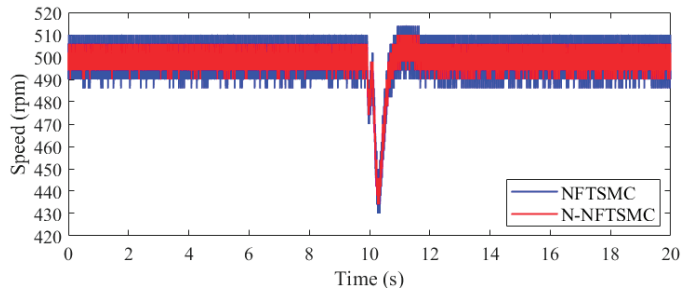


Figure 15. Experimental results at 500 rpm using a sudden load.

In order to further verify the robustness of N-NFTSMC, a loading experiment is conducted at a reference speed of 1000 rpm. From the experimental results in Figure 16, the chattering range of the N-NFTSMC is smaller than that of the NFTSMC. The N-NFTSMC also exhibits an anti-disturbance performance that is stronger than that of the NFTSMC.

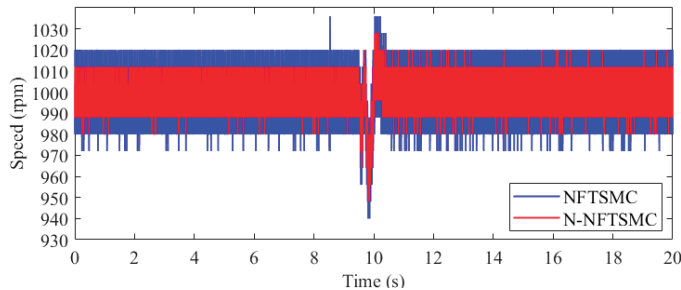


Figure 16. Experimental results at 1000 rpm using a sudden load.

The experimental results using different control methods are listed in Table 3 for clearer comparisons. We can see that, compared with the NFTSMC, the proposed N-NFTSMC shows smaller speed chattering. Moreover, when a sudden load is applied, the N-NFTSMC has a smaller maximum chattering value. These results verify that the N-NFTSMC can effectively suppress chattering and improve the anti-disturbance of the PMSM motion control system.

Table 3. Comparisons of experimental results using different control methods.

Control Method	Maximum Chattering Value (without Sudden Load)	Maximum Chattering Value (with Sudden Load)
NFTSMC (500 rpm)	10 rpm	70 rpm
N-NFTSMC (500 rpm)	6 rpm	66 rpm
NFTSMC (1000 rpm)	20 rpm	60 rpm
N-NFTSMC (1000 rpm)	12 rpm	52 rpm

7. Conclusions

In this work, we combined the nonsingular fast terminal sliding mode control with a sliding mode disturbance observer to enhance the chattering suppression and anti-disturbance capacity of the PMSM servo system. We conducted step response simulations that showed that our approach improves the response time and adequately suppresses chattering compared with the traditional FTSMC. Furthermore, we conducted sine tracking

simulations that showed that our approach has a lower maximum error compared with the traditional FTSMC.

In addition, we conducted the experiments using different speeds. The experimental results show that the N-NFTSMC improves the anti-disturbance performance and alleviates chattering. The results from the simulations and experiments indicate that the N-NFTSMC strengthens the servo system's dynamic performance and its disturbance mitigation. In the future, second-order sliding-mode algorithms will be considered for high-performance PMSM servo control.

Author Contributions: All authors made significant contributions to the work presented in this manuscript. Conceptualization, D.S. and Y.W.; methodology, D.S. and K.B.; investigation, D.S. and Y.W.; resources, C.X.; writing—original draft preparation, D.S. and K.B.; writing—review and editing, C.X. and L.L.; supervision, C.F.; project administration, Y.W.; funding acquisition, D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Foundation of Nanjing Xiaozhuang University (Grant No. 2023NXY23). This work was also supported by the Foundation of Basic Science (Natural Science) Research Program for Higher Education Institutions in Jiangsu Province (Grant No. 24KJD460005) and the Youth Fund of the Jiangsu Natural Science Foundation (Grant No. BK20220610). The National Natural Science Foundation of China (Grant No. 12404021), the Foundation of Basic Science (Natural Science) Research Program for Higher Education Institutions in Jiangsu Province (Grant No. 23KJD140003), the Foundation of Nanjing Xiaozhuang University (Grant No. 2022NXY25).

Data Availability Statement: The data presented in this study are available upon request from the first author.

Conflicts of Interest: The authors declare no conflicts of interest.

Nomenclature

Abbreviation	Full Name
PMSM	Permanent magnet synchronous motor
NFTSMC	Nonsingular fast terminal sliding mode control
N-NFTSMC	Novel nonsingular fast terminal sliding mode control
NSRL	New sliding mode reaching law
SDOB	Sliding mode disturbance observer
PID	Proportional–integral–derivative
ADRC	Active disturbance control
TVNDO	Time-varying nonlinear disturbance observer
ANNNFTSMC	Adaptive neural network nonsingular fast terminal sliding mode control
IPMSM	Interior permanent magnet synchronous motor
SMC	Sliding mode control
NTSMC	Nonsingular terminal sliding mode control

References

- Dong, Z.; Liu, Y.; Wen, H.; Feng, K.; Yu, F.; Liu, C. A Novel Winding Connection Sequence of Dual Three-Phase Series-End Winding PMSM Drive for Speed Range Extension. *IEEE Trans. Magn.* **2023**, *59*, 8202905. [CrossRef]
- Chen, Y.; Zang, B.; Wang, H.; Liu, H.; Li, H. Composite PM Rotor Design and Alternating Flux Density Harmonic Component Analysis of a 200 kW High-Speed PMSM Used in FESS. *IEEE Trans. Ind. Appl.* **2023**, *59*, 1469–1480. [CrossRef]
- Chung, S.U.; Kim, J.W.; Chun, Y.D.; Woo, B.C.; Hong, D.K. Fractional Slot Concentrated Winding PMSM with Consequent Pole Rotor for a Low-Speed Direct Drive: Reduction of Rare Earth Permanent Magnet. *IEEE Trans. Energy Convers.* **2015**, *30*, 103–109. [CrossRef]
- Zhao, L.; Xia, Y.; Yang, Y.; Liu, Z. Multicontroller positioning strategy for a pneumatic servo system via pressure feedback. *IEEE Trans. Ind. Electron.* **2017**, *64*, 4800–4809. [CrossRef]
- Li, J.; Li, W. On-Line PID Parameters Optimization Control for Wind Power Generation System Based on Genetic Algorithm. *IEEE Access* **2020**, *8*, 137094–137100. [CrossRef]
- Wang, B.; Tian, M.; Yu, Y.; Dong, Q.; Xu, D. Enhanced ADRC with Quasi-Resonant Control for PMSM Speed Regulation Considering Aperiodic and Periodic Disturbances. *IEEE Trans. Transp. Electrification* **2020**, *8*, 3568–3577. [CrossRef]

7. Wang, Y.; Feng, Y.; Zhang, X.; Liang, J. A New Reaching Law for Antidisturbance Sliding-Mode Control of PMSM Speed Regulation System. *IEEE Trans. Power Electron.* **2020**, *35*, 4117–4126. [CrossRef]
8. Xu, B.; Jiang, Q.; Ji, W.; Ding, S. An Improved Three-Vector-Based Model Predictive Current Control Method for Surface-Mounted PMSM Drives. *IEEE Trans. Transp. Electr.* **2022**, *8*, 4418–4430. [CrossRef]
9. Ding, S.; Zheng, W.X. Nonsingular terminal sliding mode control of nonlinear second-order systems with input saturation. *Int. J. Robust Nonlinear Control* **2016**, *26*, 1857–1872. [CrossRef]
10. Yang, J.; Li, S.; Su, J.; Yu, X. Continuous nonsingular terminal sliding mode control for systems with mismatched disturbances. *Automatica* **2013**, *49*, 2287–2291. [CrossRef]
11. Feng, Y.; Han, F.; Yu, X. Chattering free full-order sliding-mode control. *Automatica* **2014**, *50*, 1310–1314. [CrossRef]
12. Yang, L.; Yang, J. Nonsingular fast terminal sliding-mode control for nonlinear dynamical systems. *Int. J. Robust Nonlinear Control* **2011**, *21*, 1865–1879. [CrossRef]
13. Yu, D.; Chen, C.L.P.; Xu, H. Fuzzy swarm control based on sliding-mode strategy with self-organized omnidirectional mobile robots system. *IEEE Trans. Syst. Man Cybern. Syst.* **2022**, *52*, 2262–2274. [CrossRef]
14. Sun, Y.; Gao, Y.; Zhao, Y.; Liu, Z.; Wang, J.; Kuang, J.; Yan, F.; Liu, J. Neural network-based tracking control of uncertain robotic systems: Predefined-time nonsingular terminal sliding-mode approach. *IEEE Trans. Ind. Electron.* **2022**, *69*, 10510–10520. [CrossRef]
15. Du, X.; Fang, X.; Liu, F. Continuous Full-Order Nonsingular Terminal Sliding Mode Control for Systems with Matched and Mismatched Disturbances. *IEEE Access* **2019**, *7*, 130970–130976. [CrossRef]
16. Le, D.V.; Ha, C. Finite-Time Fault-Tolerant Control for a Stewart Platform Using Sliding Mode Control with Improved Reaching Law. *IEEE Access* **2022**, *10*, 43284–43302. [CrossRef]
17. Utkin, V. Discussion Aspects of High-Order Sliding Mode Control. *IEEE Trans. Autom. Control* **2016**, *61*, 829–833. [CrossRef]
18. Rath, J.J.; Defoort, M.; Karimi, H.R.; Veluvolu, K.C. Output Feedback Active Suspension Control with Higher Order Terminal Sliding Mode. *IEEE Trans. Ind. Electron.* **2017**, *64*, 1392–1403. [CrossRef]
19. Gao, W.; Hung, J.C. Variable structure control of nonlinear systems: A new approach. *IEEE Trans. Ind. Electron.* **1993**, *40*, 45–55.
20. Wang, T.; Zhao, M.; Li, Y.; Liu, K. Double-power reaching law sliding mode control for spacecraft decline based on radial basis function networks. In Proceedings of the 2017 29th Chinese Control and Decision Conference (CCDC), Chongqing, China, 28–30 May 2017; pp. 5396–5401.
21. Tao, M.; Chen, Q.; He, X.; Sun, M. Adaptive fixed-time fault-tolerant control for rigid spacecraft using a double power reaching law. *International. J. Robust Nonlinear Control* **2019**, *29*, 4022–4040. [CrossRef]
22. Fallaha, C.J.; Saad, M.; Kanaan, H.Y.; Al-Haddad, K. Sliding-Mode Robot Control with Exponential Reaching Law. *IEEE Trans. Ind. Electron.* **2011**, *58*, 600–610. [CrossRef]
23. Yang, G.-Y.; Chen, S.-Y. Piecewise fast multi-power reaching law: Basis for sliding mode control algorithm. *Meas. Control* **2020**, *53*, 1929–1942. [CrossRef]
24. Liu, W.; Chen, S.; Huang, H. Adaptive Nonsingular Fast Terminal Sliding Mode Control for Permanent Magnet Synchronous Motor Based on Disturbance Observer. *IEEE Access* **2019**, *7*, 153791–153798. [CrossRef]
25. Xu, J.; Yu, X.; Qiao, J. Hybrid Disturbance Observer-Based Anti-Disturbance Composite Control with Applications to Mars Landing Mission. *IEEE Trans. Syst. Man Cybern. Syst.* **2021**, *51*, 2885–2893. [CrossRef]
26. Fu, Y.; Wang, D. Higher-Order State and Disturbance Observer with $O(T^3)$ Errors for Linear Systems. *IEEE Access* **2019**, *7*, 102812–102819. [CrossRef]
27. Yang, Y.; Zhang, J.; Huang, M.; Tan, X. Disturbance Observer-Based Event-Triggered Control of Switched Positive Systems. *IEEE Trans. Circuits Syst. II Express Briefs* **2024**, *71*, 1191–1195. [CrossRef]
28. Zhang, Z.; Liu, X.; Yu, J.; Yu, H. Time-Varying Disturbance Observer Based Improved Sliding Mode Single-Loop Control of PMSM Drives with a Hybrid Reaching Law. *IEEE Trans. Energy Convers.* **2023**, *38*, 2539–2549. [CrossRef]
29. Liu, J.; Yang, Y.; Li, X.; Zhao, K.; Yi, Z.; Xin, Z. Improved Model-Free Continuous Super-Twisting Non-Singular Fast Terminal Sliding Mode Control of IPMSM. *IEEE Access* **2023**, *11*, 85361–85373. [CrossRef]
30. Zhang, R.; Xu, B.; Shi, P. Output Feedback Control of Micromechanical Gyroscopes Using Neural Networks and Disturbance Observer. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 962–972. [CrossRef]
31. Wang, Z.; Li, S.; Li, Q. Continuous Nonsingular Terminal Sliding Mode Control of DC–DC Boost Converters Subject to Time-Varying Disturbances. *IEEE Trans. Circuits Syst. II Express Briefs* **2022**, *67*, 2552–2556. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Characteristic Canonical Analysis-Based Attack Detection of Industrial Control Systems in the Geological Drilling Process

Mingdi Xu ¹, Zhaoyang Jin ¹, Shengjie Ye ¹ and Haipeng Fan ^{2,*}

¹ Wuhan Institute of Digital Engineering, Wuhan 430074, China; mingdixu@163.com (M.X.); jinzhaoyang37@163.com (Z.J.); shengjieily@hust.edu.cn (S.Y.)

² School of Automation, China University of Geosciences, Wuhan 430074, China

* Correspondence: fanhaipeng@cug.edu.cn

Abstract: Modern industrial control systems (ICSs), which consist of sensor nodes, actuators, and buses, contribute significantly to the enhancement of production efficiency. Massive node arrangements, security vulnerabilities, and complex operating status characterize ICSs, which lead to a threat to the industrial processes' stability. In this work, a condition-monitoring method for ICSs based on canonical variate analysis with probabilistic principal component analysis is proposed. This method considers the essential information of the operating data. Firstly, the one-way analysis of variance method is utilized to select the major variables that affect the operating performance. Then, a concurrent monitoring model based on probabilistic principal component analysis is established on both the serially correlated canonical subspace and its residual subspace, which is divided by canonical variate analysis. After that, monitoring statistics and control limits are constructed. Finally, the effectiveness and superiority of the proposed method are validated through comparisons with actual drilling operations. The method has better sensitivity than traditional monitoring methods. The experimental result reveals that the proposed method can effectively monitor the operating performance in a drilling process with its highest accuracy of 92.31% and a minimum monitoring delay of 11 s. The proposed method achieves much better effectiveness through real-world process scenarios due to its distributed structural division and the characteristic canonical analysis conducted in this paper.

Citation: Xu, M.; Jin, Z.; Ye, S.; Fan, H. Characteristic Canonical Analysis-Based Attack Detection of Industrial Control Systems in the Geological Drilling Process. *Processes* **2024**, *12*, 2053. <https://doi.org/10.3390/pr12092053>

Academic Editor: Iqbal M. Mujtaba

Received: 30 July 2024

Revised: 9 September 2024

Accepted: 15 September 2024

Published: 23 September 2024

Keywords: industrial control systems; performance monitoring; canonical variate analysis; principal component analysis

1. Introduction

Industrial control systems (ICSs) are increasingly critical in modern infrastructure and significant projects such as the hydraulic facility, transport, energy, and chemical industries. In this sense, ICS security is also directly linked to the smooth operation of critical infrastructures [1]. When the ICSs are attacked, it will directly harm the physical world by causing environmental pollution, power outages, oil leaks, and explosions. With the acceleration of the digitalization process of ICSs, the integration of industrialization and informatization has been gradually strengthened. Due to the increasing openness of industrial control systems, there are an increasing number of threats to the systems. Hence, timely and accurate anomaly detection in ICSs is essential in reflecting the security status of the production process and determining the vulnerability of the industrial control systems. Maintaining secure operation of industrial control networks is increasingly critical in improving production efficiency and safety [2,3].

ICSs are a series of control systems, which include supervisory control and data acquisition systems, distributed control systems and programmable logic controllers (PLCs), and other control systems and control units. An ICS ensures the safe, reliable, and secure operation of industrial processes. In ICSs, malicious attacks are possible due to the inherent



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

loopholes in communication protocols. Recently, industrial control networks have faced constant threats, such as the Stuxnet virus attacking PLC codes to achieve such an attack, thereby destroying the centrifuge's regular operation. Thus, numerous researchers have devoted themselves to constructing state models for ICSs to enable anomaly detection for different attacks.

The ICSs' layers interact and communicate with one another via the network while carrying out their specific assigned tasks. The ICS is vulnerable in both the network along physical layers due to its close coupling between cyberspace and physical space. An attacker may launch a cyberattack, which could result in malicious software and data asset theft or tampering with the equipment, leading to the loss of crucial control information and failure of crucial control commands.

The threat of attack has caused worldwide concern about the cyber security of ICSs. Given the numerous attack threats faced by ICSs, Teixeira et al. proposed an ICS pass-through attack model based on three-dimensional information and physical space to characterize the various attack means in different spaces and to illustrate the characteristics of multiple types of attacks [4]. Accordingly, Adepu et al. proposed a framework for describing physical attacks, cyberattacks, and other types of attacks by dividing them into domain, attacker, and attack models [5]. In light of the wide variety of attack types, complex attack paths, and variable attack strategies facing ICSs, it is challenging to construct a mathematical model covering all scenarios.

Currently, data-driven methods of extracting information from process data and modeling monitoring have become a hotspot in anomaly detection research. The advancement of sensor technology has allowed almost all industrial objects to be equipped with various types of sensors, which has resulted in a great deal of data being collected in industrial processes. By merging the data from various sources and examining the correlation between the information, data-driven anomaly detection methods can detect whether a system is under attack. A relational model that captures the intruder's identity, velocity, level of threat, and target of intrusion was developed, which serves as a foundation for continuous cyberspace state monitoring [6]. Lu et al. proposed a security monitoring method for industrial control networks based on an improved C-SVC (C-Support Vector Classifier), which can effectively identify multiple types of abnormal states and form situational awareness results [7]. A hidden Markov model-based attack detection for Stuxnet has been proposed in the industrial control system subject to random packet dropouts [6]. Despite being based on mechanistic models of attack-induced abnormal states, the methods above have inherent limitations when applied to large-scale complex industrial processes.

Considering the large scale and complexity of the system in question, as opposed to complex processes mechanisms, researchers have monitored network security status by analyzing the process data in industrial control networks. Multivariate statistical process monitoring (MSPM) methods have been widely studied and applied over the past few decades [8]. Rather than modeling a particular attack model, MSPM depicts the operational state of the system. Attack detection on ICSs is achieved by comparing the deviations from the operational state. Among the most well-known representative branches of statistical process monitoring is principal component analysis (PCA), which is regarded as an effective means of dimension reduction. PCA identifies the major changes in data by decomposing multiple related variables into several orthogonal principal components [9,10]. The PCA-based MSPM approach enables monitoring by modeling the variable space of the system where two different monitoring statistics, Hotelling T^2 and Squared prediction error Q , are viewed as the monitoring statistics [11,12].

Although PCAs are widely used to detect anomalies, they do not perform as well when their assumptions are incorrect. The underlying Gaussian assumption in the calculation of control limits of monitoring statistics in PCA makes it a poor monitoring tool for non-Gaussian processes. A variety of PCA variants have been proposed for nonlinear processes, including probability PCA (PPCA) [13] and kernel PCA (KPCA) [14], in which the data are projected into a high-dimensional space. In essence, KPCA remains a linear

dimensionality reduction method, and its effectiveness is heavily influenced by the choice of kernel function, which is not appropriate for systems with nonlinear or stochastic perturbations. Within the maximum likelihood framework, PPCA measures the similarity between new data points according to their probability density functions [15,16]. Canonical variate analysis (CVA), which provides a more accurate description of the process by maximizing the correlation between mainly dependent and quality variables [17,18], is another valid method for incorporating both static and dynamic process characteristics. Zhang et al. developed a CVA-based modeling and monitoring method for simultaneous static and dynamic analysis in three-phase flow processes [19,20]. A fault information-aided canonical variate analysis and a structured monitoring strategy has been proposed to improve anomaly detection rate [17]. However, the process is usually assumed to operate under one condition, whereas industrial processes always operate in multiple modes.

For plant-wide processes, multimodal methods were introduced as a solution to these problems. Generally, block division is the key step in sub-block modeling. These methods can be classified into two main categories: data-driven and knowledge-based. Based on field experience and prior process knowledge, knowledge-based methods usually divide process variables into blocks. A hierarchical multiblock total projection to latent structures (T-PLS) based on an operating performance assessment scheme was proposed to identify the anomalies in operating statuses [21]. Using prior process knowledge, Zhu et al. proposed the distributed parallel PCA process monitoring framework to decompose the high-dimensional process variables [22]. When there is a lack of accurate prior knowledge, monitoring and anomaly detection performance may be less than optimal if the process variables are not correctly divided.

Data-driven methods have also been extensively used to divide variable blocks in distributed process monitoring using the process measurements from industrial historians. The data-driven approach clusters variables into sub-blocks by evaluating the correlations between variables. For instance, Hu et al. used mutual information (MI) analysis to extract the complex relationships between each possible process variable and the burn-through point in the sintering process [23]. Zhang et al. investigated an improved mixture probability principal component analysis with clustering for nonlinear process monitoring where the k -means is subsequently utilized as a clustering algorithm to divide the variables into optimal sub-blocks [24]. Minimal redundancy maximal relevance was used to divide the most related variables into the same block and form a dynamic multiblock monitoring framework [25]. With mutual information-spectral clustering, the measured variables were automatically divided into sub-blocks on which a Bayesian inference-based multiblock KPCA monitoring model was established [26]. Combining knowledge-based and data-driven approaches, Cao et al. developed a hierarchical hybrid, distributed PCA for the plant-wide monitoring of chemical processes with two-layer manner sub-block division.

Although the aforementioned monitoring strategies have been demonstrated as effective, the monitoring performance may not be optimal when faced with sophisticated cyberattacks. On the one hand, network layer attacks such as data injection present more randomness and uncertainty than faults in the system. The above characteristics lead to traditional monitoring methods failing to identify the dynamic characteristics caused by attacks when modeling with normal samples. Specifically, PCA-based monitoring methods cannot fully extract the state-by-state characteristics of the system in the principal metric space, leading to omissions and false alarms in the monitoring results. Similarly, when confronted with large-scale complex systems, the traditional centralized modeling approach cannot adequately reconstruct the system's state characteristics.

Motivated by the above research status, a concurrent distributed monitoring method was proposed to tackle the ICS attack detection tasks. redUsing a two-stage distributed modeling approach, we can extract all the state characteristics of the system. By using the MI method, the decision variables are selected and the distributed structure is realized. Then, the PPCA models compute both the serially correlated subspace and its residual

subspace based on canonical variate analysis, which makes a complete interpretation of process dynamics under ICSs possible.

In the proposed framework, all detection variables were selected into the first layer by one-way variance analysis, and the detection variables were further divided into sub-blocks using a combination of general knowledge-based strategies with mutual information. Then, CVA-PPCA monitoring models were established for each sub-block, in which CVA was used to explore the serial correlations, and PPCA-based monitoring models were constructed for the variables of subspace. Finally, Bayesian inference was used to obtain comprehensive statistical indicators of the ICSs, which can realize plant-wide anomaly detection. Thus, the dynamic characteristics of the ICSs were restored, allowing for a deeper understanding of its security status. The main contributions of the present work are as follows.

1. An adaptive process variable selection and blocking method for distributed monitoring was implemented with combined knowledge-based strategies with mutual information.
2. Both linear and non-linear behaviors were analyzed and monitored, which can provide a meaningful interpretation for fine-scale identifying ICS attacks.

The rest of this paper is organized as follows. The problem description and monitoring framework are given in Section 2. Section 3 outlines the proposed concurrent distributed CVA-PPCA-based monitoring method in detail. Section 4 details a validation of the effectiveness of the proposed method on actual drilling processes. Finally, conclusions are made in Section 5.

2. Problem Description and Modeling Framework

In this section, the problems of ICS security monitoring are summarized. Based on these, a framework of monitoring model was designed.

2.1. Problem Description

ICS is an umbrella term for various network-connected control systems in the industrial field. Over the past few decades, ICSs have greatly enhanced the degree of industrial process automation and brought certain security risks. Figure 1 shows a typical industrial control network architecture for the geological drilling processes. A controller employs a communication network to regulate the operation of the controlled process by measurements from geographically dispersed sensors.

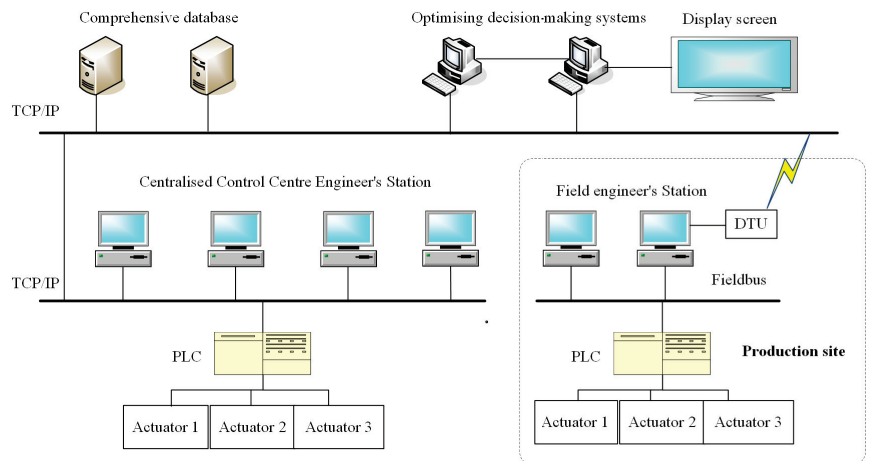


Figure 1. A typical industrial control system structure for the drilling process.

During the drilling process, the PLC is responsible for controlling the industrial control system in order to read the data from the field sensors. Additionally, the Profibus communication protocol was utilized in order to facilitate communication between the PLC and the industrial control machine. For the purpose of reading the data from the PLC over the OLE for Process Control (OPC) protocol, the WinCC configuration software was utilized. MVC (Model–View–Controller) architecture was utilized by the system, which enables intelligent optimization control, as well as complicated logic operations.

A system failure results from an attacker’s deliberate destruction or manipulation of actuators, control units, etc., which is another manifestation of the ICS vulnerability in the physical layer. Network attacks and instrument malfunction both appear as anomalies in the data sampled by the sensors. The difference is that network attacks cause equipment failure, so the data usually show a causal relationship between them. Additionally, network attacks tend to maintain the statistical characteristics of the data sparingly, whereas equipment failures often result in outliers, missing values, and other easily observable changes. Due to the complexity of physical layer attacks, the attack detection algorithms in this paper only address attacks suffered at the network layer.

False data injection is a common network layer attack. In the event that sensors transmit sensing data to the PLC, the data may be tampered with, leading to the instability of the control system. In this attack, the original correct measurement value $z_i(t)$ of moment t will be tampered with, resulting in the measurement value $\tilde{z}_i(t)$ deviating from the normal value $z_i(t)$, which causes the feedback control system to perform incorrect responses. The attack process can be expressed as [7]

$$z_i(t) = \begin{cases} 0, & t \notin T_{\text{atc}}, \\ z_i(t-1) + \tau\varphi^t, & t \in T_{\text{atc}}, \end{cases} \quad (1)$$

where τ and φ are the impact index, which is usually a constant; and T_{atc} is the attack period. This paper assumed that the anomalous state of the system was caused by fake data that were imposed by the attacker.

In general, false data injection attacks include the manipulation of system measures while the attacker is aware of the setup of the system. These attacks are difficult to monitor directly since they are difficult to detect. The three primary types of attacks that fall under the category of fake data injection assaults are known as surge attacks, deviation attacks, and geometry attacks. To varied degrees and at varying rates, the normal operation of the system is disrupted in each of these instances, and, when it is severe, it is likely to result in serious accidents. Figure 2 presents histogram plots of the partial variables in the geological drilling process, such as the rate of permeation (ROP) as an example. Clearly, the distribution of data that is not ideal (shown by the red area) is mostly contained within the distribution of data that is optimal (represented by the blue area). Since this is the case, one of the most important concerns in ICS security monitoring is how to further parse data features. Monitoring the current status of network security can assist decision makers in determining whether or not an attacker intends to launch an attack. The operation of the system will be guaranteed to be stable and secure as a result of this.

In a data tampering attack, the attacker tampers with measured values of a system since he knows the system configuration and cannot be detected intuitively. Therefore, the following challenges need to be faced when investigating ICS-oriented attack detection methods.

1. Complexity: The number of current cyberattacks on ICSs is increasing, with attackers exploiting ICS vulnerabilities to deliver different types of attacks and threats.
2. Crypticity: There are insufficient means of identifying attack behavior, and the attack detection false alarm rate is high due to attackers deliberately confusing the attack with the normal operation of the control system.

Therefore, an essential component of achieving ICS attack detection involves developing a monitoring model that accurately captures the dynamic aspects of the attack behavior.

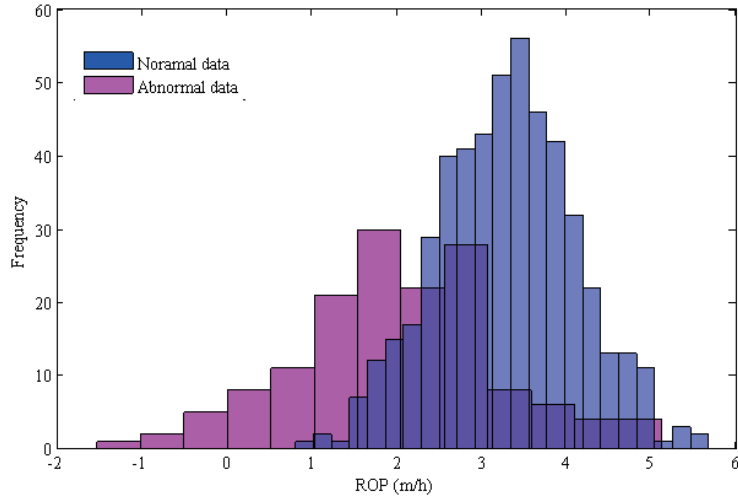


Figure 2. Histograms of the drilling data under optimal and non-optimal modes.

2.2. Modeling Framework

The objective of this study was to detect the abnormalities of ICSs by constructing a process monitoring model based on the sufficient normal data of related detection variables. A novel CVA-PPCA-based monitoring method was presented to overcome the shortcomings and improve the performance of network anomaly identification. The framework of the proposed network condition monitoring scheme is shown in Figure 3.

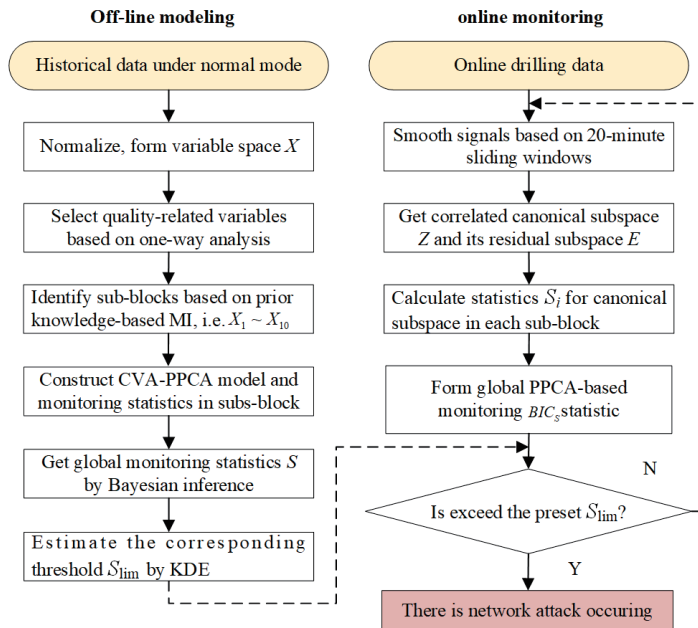


Figure 3. The framework of the proposed CVA-PPCA-based monitoring method.

The monitoring model consists of two parts: offline modeling and online monitoring. According to one-way analysis, the ICS performance quality-related detection variables were chosen; these were then further divided into reasonable sub-blocks by the MI analysis, which were added with prior knowledge. Within each sub-block, the CVA method was used to classify the variables according to their correlation into their correlated canonical subspace and residual subspace. Then, the PPCA-based monitoring model was established in canonical subspace. Finally, Bayesian inference was used to obtain comprehensive statistical indicators of the whole process, which can realize anomaly detection.

For online monitoring, real-time monitoring statistics can be compared with historical data to determine the overall performance of the integrated monitoring system and to define the detection thresholds according to attack type. Anomalies can then be detected by comparing the monitoring statistics to see if the limits have been exceeded.

3. Implementation of the Monitoring Model

In this section, the ICS security monitoring model is established. Firstly, sub-block division was carried out using one-way analysis and mutual information analysis. Using the CVA method, the original variable space was divided, and the PPCA monitoring model with preset control limits was constructed. To achieve online monitoring, the online data are used to calculate the monitoring statistics and compare them to the detection threshold.

3.1. Sub-Block Division Based on One-Way Analysis and Mutual Information Analysis

There are usually multiple industrial controls and multiple systems within ICSs. The whole process contains a number of detection variables. The multi-block modeling approach is an effective way to deal with the anomaly detection problem of large-scale processes. To fully extract the correlations between variables, sub-block division is necessary before offline modeling.

A two-stage delineation method was used in this study to create a multi-sub-block structure, with one-way analysis of variance being selected in the first stage to determine the operational state-related decision variables, which was followed by mutual information analysis and process knowledge for sub-block delineation.

In the first phase, one-way analysis of variance (ANOVA) can be used to determine the effect of the different operating modes on the distribution of variable data. By measuring the difference in the variance fluctuations caused by different operating conditions and random errors, ANOVA determines if changes in the operating conditions are a major factor in system operation.

There are five normal geological drilling conditions: drill up and down, rotary drilling, back reaming, hole sweeping, and sliding drilling. Assuming that the number of samples for each operating condition is selected as n_1, n_2, n_3, n_4 , and n_5 , then the drilling data for each condition is recorded as $x_{1,j}, x_{2,j}, \dots, x_{n_j,j} (j = 1, 2, \dots, 5)$. The degree of variation V_T between the drilling data can be calculated as follows:

$$\begin{aligned} V_T &= \sum_{j=1}^5 \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2, \\ \bar{x} &= \frac{1}{N} \sum_{j=1}^h \sum_{i=1}^{n_j} x_{ij} = \frac{1}{n} \sum_{j=1}^5 n_j \bar{x}_j, \\ \bar{x}_j &= \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}, \end{aligned} \quad (2)$$

where \bar{x} is the mean value of data collected for the variable, and \bar{x}_j is the mean value of the variable in a data set for a mode. Furthermore, V_T can be decomposed into the sum of its

error sum of squares and effect sum of squares, which is denoted as $V_T = V_E + V_F$, and V_E and V_F are relative-independent, the details of which can be defined as

$$\begin{aligned} V_E &= \sum_{j=1}^h \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2, \\ V_F &= \sum_{j=1}^h \sum_{i=1}^{n_j} (\bar{x}_j - \bar{x})^2. \end{aligned} \quad (3)$$

According to the above definition, it is clear that V_T measures the distributional differences within different drilling conditions and also globally. Thus, it is possible to select the operating status-related variables related to the ICSs effectively. The degree of influence of a variable is measured by constructing a test statistic F_T and its test probability ρ :

$$F_T = \frac{(V_F)/(h-1)}{(S_E)/(N-h)} \sim F(h-1, N-h), \quad (4)$$

where F_T obeys F -distribution, $h = 5$, and the test probability is $\rho(F(h-1, N-h) \geq F_T)$. The smaller the test probability, the greater the effect of the parameter on the operating conditions.

Table 1 presents the test probability of each parameter based on 1800 samples of data collected from the industrial control network in the drilling process. Clearly, the probability of testing parameters d_{11} and d_{12} is significantly higher than those of the other variables, which is also consistent with the process knowledge. A total of 10 variables can be selected for $\rho \leq 0.001$, i.e., X_1, X_2, \dots, X_{10} .

Table 1. Results of the one-way analysis of variance.

Parameter	Description	ρ
d_1	Rate of penetration (km/h)	6.27×10^{-8}
d_2	Weight on bit (kN)	0.43×10^{-10}
d_3	Rotation speed (r/min)	1.84×10^{-8}
d_4	Mud flow in (out) (L/s)	3.27×10^{-6}
d_5	Tank volume (m ³)	4.17×10^{-15}
d_6	Standpipe pressure (Mpa)	9.38×10^{-12}
d_7	Hookload (kN)	2.86×10^{-14}
d_8	Hook height (m)	1.22×10^{-18}
d_9	Rotary torque (kN·m)	8.86×10^{-18}
d_{10}	Depth (m)	1.09×10^{-9}
d_{11}	Bit dept h (m)	5.36×10^{-2}
d_{12}	Bit diameter (mm)	6.25×10^{-3}

In the second stage, the detection variable blocking is based on MI combined with prior knowledge. MI involves determining whether a detection parameter's data distribution and a performance indicator's distribution are interdependent. When several variables interact, MI is the entropy that was initially contained as it decays. It suggests that information entropy is not constant but rather varies with the number of events that occur. MI is commonly interpreted as a metric that quantifies the degree of dependence and strength between two variables. Specifically, given two random variables x_1 and x_2 , the mutual information between them is defined as

$$I(X, Y) = \sum_X \sum_Y p(X, Y) \log \frac{p(X, Y)}{p(X)p(Y)}, \quad (5)$$

where $p(x)$ and $p(X)$ are the marginal probability density functions of X and Y , and $p(X, Y)$ is the joint probability of X and Y .

As this equation represents the uncertainty in x_2 after removing x_1 , it confirms the intuitive meaning of MI as the amount of information one variable provides about another. By analyzing the physical mechanism of the drilling production, it can be seen that d_4 , d_5 , and d_6 are part of the mud system, and d_2 and d_3 are also one of the d_1 -influencing parameters. Then, according to the blocking criterion [13], these variables were divided into three sub-blocks: $[X_1, X_2, X_3, X_9]$, $[X_4, X_5, X_6]$, and $[X_7, X_8, X_{10}]$.

Hence, the detection variables were blocked according to their interrelationships using the MI combined with prior knowledge, and the CVA-PPCA anomaly detection model is then applied on a distributed sub-block structure.

3.2. Canonical Subspace Identification Based on CVA

The drilling detection variables $d_1 \sim d_{10}$ are categorized into distinct sub-blocks based on current correlations. Then, state monitoring models would be constructed within each sub-block by parsing the data characteristics to accomplish anomaly detection for various attack methods.

Canonical variate analysis (CVA) is a dimension reduction algorithm that maximizes the alignment between two sets of variables. By maximizing the correlation between the “past” values and the “future” values of the system, the CVA-based approach generates state-space models from time-related data. Thus, CVA can be used to establish the relationships between process variables and quality variables, and the trained CVA model can be used for process monitoring related to quality.

In CVA, linear dimension reduction is used to reduce the size of variables so that it can be used to determine the most significant correlation between qualitative and primary dependent variables, as well as dynamic processes [20]. This study addresses the auto-correlation challenge of modeling the operational state of industrial control networks.

The past and future drilling data matrix is constructed using drilling data $x_k = [X_1, X_2, \dots, X_n]^T (k = 1, 2, \dots, N; n = 12)$. Assume that, at moment k , the past vector $x_{p,k}$, comprising the past data, and the future vector $x_{f,k}$, containing the present and future observations, are defined as

$$\begin{aligned} x_{p,k} &= \left[x_{(k-1)}^T, x_{(k-2)}^T, \dots, x_{(t-l)}^T \right]^T, \\ x_{f,k} &= \left[x_{(k)}^T, x_{(k+1)}^T, \dots, x_{(k+l)}^T \right]^T, \end{aligned} \tag{6}$$

where the two vectors, i.e., $x_{p,k}$ and $x_{f,k}$, should first be normalized to a zero mean and with unit variance. To define the past and future matrices, vectors were arranged in the following Hankel matrix:

$$\begin{aligned} X_p &= \begin{bmatrix} x_{p(l+1)}, x_{p(l+2)}, \dots, x_{p(l+N_1)} \end{bmatrix}, \\ X_f &= \begin{bmatrix} x_{f(l+1)}, x_{f(l+2)}, \dots, x_{f(l+N_1)} \end{bmatrix}, \end{aligned} \tag{7}$$

where $N_1 = N - 2l + 1$ for a dataset with N samples.

The aim of CVA is to reveal the remarkable features of the ICS operating conditions by identifying the projection matrix L and J in order to identify a linear combination of the future and past observations that have the optimal linear performance. The problem of solving the projection matrix is defined as follows:

$$\begin{aligned} \max_{J,L} & J^T \Sigma_{pf} L, \\ \text{s.t.} & J^T \Sigma_{pp} J = I, \\ & L^T \Sigma_{ff} L = I. \end{aligned} \tag{8}$$

The projection matrix J and L can be calculated by singular-value decomposition (SVD) on the Hankel matrix H as follows:

$$H = \Sigma_{ff}^{-1/2} \Sigma_{fp} \Sigma_{pp}^{-1/2} = U \Lambda V^T, \quad (9)$$

where the sample covariances $\Sigma_{ff}^{-1/2}$ and $\Sigma_{pp}^{-1/2}$ and the cross-covariance of Σ_{fp} of the past vector $x_{p,k}$ and the future vector $x_{f,k}$ are defined as follows:

$$\begin{bmatrix} \Sigma_{pp} & \Sigma_{pf} \\ \Sigma_{fp} & \Sigma_{ff} \end{bmatrix} = \frac{1}{N_1 - 1} \begin{bmatrix} X_p X_p^T & X_p X_f^T \\ X_f X_p^T & X_f X_f^T \end{bmatrix}, \quad (10)$$

where U and V consist of singular vectors that are orthogonal and only pairwise-correlated, and Λ is a diagonal matrix containing the canonical correlation coefficients. Thus, the projection matrices J and L can be calculated by taking the first r columns of U and V , respectively.

For the k moments of the ICS operation, the transformation matrices J and L are as follows:

$$\begin{aligned} J_r &= V_r^T \Sigma_{pp}^{-1/2}, \\ L_r &= U_r^T \Sigma_{pp}^{-1/2}. \end{aligned} \quad (11)$$

The canonical state subspace Z and its residual subspace E of the drilling data matrix x can be defined as

$$\begin{aligned} Z &= J_r X_p \in R^{r \times N_1}, \\ E &= F_r X_p \in R^{n_1 \times N_1}, \end{aligned} \quad (12)$$

where the residual projection matrix $F_r = (I - V_r V_r^T) \Sigma_{pp}^{-1/2}$.

Therefore, the space of the primary and dependent variables Z , which are canonically correlated with the ICSs' operational performance, is extracted within each sub-block. Then, a PPCA-based monitoring model is built on it to detect cyberattacks.

3.3. Overall Monitoring Model

According to CVA, the ICS variable space for drilling processes consists of a correlated canonical and residual subspace. It is necessary to establish a model for monitoring subspace in order to implement the proposed scheme.

PPCA-based monitoring model: The PPCA method is a representation of PCA in probability space, where probability density functions measure the degree of the novelty of new data points. While PCA is a linear down-scaling method, PPCA can take into account the nonlinear and dynamic characteristics of the system fully. When dealing with non-linear characteristics, PCA is vastly improved by the incorporation of probability. Data x is believed to be generated by the latent variable z when viewed from the perspective of probability. In order to produce the standard PPCA, the following pattern is utilized [15]:

$$x = f(z, w) + \xi, \quad (13)$$

where $x \in R^d$ is the process observation variable, $z \in R^p$ is the vector of latent variables, $w \in R^{n \times q}$ is the associated model parameter vector like loading matrix, ξ is an independent noise vector, and $f(\cdot)$ describes the unknown function, which can be interpreted by a linear model in general.

$$X = WZ + \mu + \xi, \quad (14)$$

where $X \in R^{d \times n}$, $Z \in R^{q \times n}$, and μ is the monitoring delay. The model parameters are then determined using a maximum-likelihood technique given a set of observational data.

According to the canonical subspace $Z \in \mathbb{R}^{n \times m}$ acquired in the previous section. The PPCA algorithm seeks the projection matrix $W \in \mathbb{R}^m$ to further reveal both the static and

dynamic process variations in which the linear transformation $Z_c = ZW$ has the maximal variance. Like PCA, the problem of matrix projection can be expressed mathematically as

$$\arg \max_W \frac{1}{n-1} Z_c^T Z_c = \arg \max_W \frac{1}{n-1} W^T Z^T Z W. \quad (15)$$

The transformed goal of the PPCA is to map the original m -dimensional data into a d -dimensional space, whose principal element model T can be expressed as

$$Z = \sum_{i=1}^p z_{c,i} w_i^T + \sum_{i=p+1}^m z_{c,i} w_i^T = Z_c W^T + E_c, \quad (16)$$

where W is the load matrix; Z_c is the scoring matrix; P is the number of principal components retained, which is commonly determined by a rule known as the cumulative percentage variance (CPV) [27]; and $E_c = Z - Z_c$ is the residual matrix, which represents process noise interference.

In general, the principal element is associated with a multivariate standard-normal distribution, while the noise residual is associated with a multivariate normal distribution, where $Z_c \sim \mathcal{N}(0, I)$, $E_c \sim \mathcal{N}(0, \sigma^2 I)$ and σ^2 is the noise variance. Then, the distribution of sample Z with respect to principal element Z_c is $Z|Z_c \sim G(Z_c W^T, \sigma^2 I)$. According to Bayes' theorem, the distribution of the sample data X is $X \sim G(0, C)$, and $C = WW^T + \sigma^2 I$.

Thus, the problem solved by the PPCA algorithm can be seen as forming observations Z from the distribution $G(0, C)$ by the hidden variable Z_c . The problem to be addressed translates into the estimation of the distribution parameters W and σ from the measurement samples [24]. This paper solves the probability distribution using the maximum-likelihood estimation problem. Expectation maximization (EM) is a powerful method for estimating the parameters of hidden variable models, which uses an expectation maximization algorithm that iterates repeatedly to find the parameters.

Online attack detection: To monitor the state of the ICSs online, the monitoring threshold must first be determined. Traditionally, PCA-based monitoring methods calculate two types of statistics, T^2 and Q , as well as the corresponding control charts. Specifically, the T^2 statistic is designed to monitor the data variations in the principal component space (PCS), while the Q statistic is used to monitor the data changes in the residual space. Observations of large deviations in the monitoring statistics may indicate an abnormal state of the industrial control network.

On the basis of the PPCA algorithm, the principal component space Z_c , contains systematic variation information and will be used to construct the T^2 statistic, while the residual E_c will form the Q statistic. The monitoring statistics are defined as

$$\begin{aligned} T^2 &= z_c^T \Lambda^{-1} z_c, \\ Q &= \|z - z_c\|^2 = (z - wz_c)^T (z - wz_c). \end{aligned} \quad (17)$$

In the case of a multivariate normal distribution for the process variables, the detection threshold for T^2 can be obtained using the F -distribution with α as the significance factor:

$$T^2 \sim \frac{r(n^2 - 1)}{n(n - p)} F_{r, n-p, \alpha}, \quad (18)$$

where p is the number of PCSs. As with the residual subspace, a weighted Chi-squared distribution can approximate the confidence limit of Q , such as

$$Q \sim d\chi_{g, \alpha}^2, \quad (19)$$

where $d = v_c/2m_q$ and $g = 2m_q^2/v_c$, in which m_q is the mean value of Q , and v_c is the corresponding variance.

As the PPCA exclusively employs the Martensian paradigm for the detection of principal elements and noise [28], the comprehensive monitoring statistics, which consist of T^2 and Q , can be directly generated from the whitened values of the statistics. The following formats were used to calculate the comprehensive monitoring statistic S :

$$S = \left\| \left(WW^T + \sigma^2 I \right)^{-0.5} z \right\|^2 = z_c^T \left(\sigma_c^2 I + WW^T \right)^{-1} z_c. \quad (20)$$

As a result of the proposed monitoring model, which effectively detects the data injection attacks on the ICSs, S_{lim} is the threshold determined by kernel density estimation (KDE) [29], which is the measurement of the degree of deviation from the normal operating conditions. Additionally, S is the monitoring statistic based on the PPCA, and S_{lim} is the threshold determined by the kernel density estimation (KDE). The threshold S_{lim} is given by

$$P(S \leq S_{lim}) = \int^{S_{lim}} \hat{\phi}(s|W, \sigma) ds = 1 - \alpha, \quad (21)$$

where $\hat{\phi}(s|W, \sigma)$ is the probability density function of S estimated by KDE. If the corresponding detection logic satisfies, for example, $S \leq S_{lim}$, the operating performance is optimal; otherwise, it is non-optimal.

According to the previous discussion, there are several sub-blocks formed here. There is a need to integrate local statistics to construct comprehensive surveillance indicators for the whole process. This study used Bayesian inference to integrate the monitoring results of multiple sub-blocks into the overall monitoring results due to its excellent performance in sub-block decision fusion. Conceptually, the probability of each sub-model being under attack can be expressed as

$$P_S(F|x_i) = \frac{P_S(x_i|F)P_S(F)}{P_Q(x_i)}, \quad (22)$$

where the prior probability of x_i is calculated as

$$P_S(x_i) = P_S(x_i|N)P_S(N) + P_S(x_i|F)P_S(F), \quad (23)$$

and the conditional probabilities $P_S(x_i|N)$ and $P_S(x_i|F)$ are defined as

$$P_S(x_i|N) = e^{-S_i/S_{i,lim}}, \quad (24)$$

$$P_S(x_i|F) = e^{-S_{i,lim}/S_i}, \quad (25)$$

where S_i represents the statistic in the i -th sub-block and $S_{i,lim}$ represents the control limits in the i -th mode blocks; N and F denote the optimal and non-optimal operating performance, respectively; $P_S(N)$ and $P_S(F)$ represent the prior probabilities under the confidence level α and $1-\alpha$; and $P_S(N) + P_S(F) = 1$. The intuitive interpretation is that the operating status expressed by sampling data is either normal or non-optimal in the drilling process.

After that, in the modeling phase, it is possible to obtain comprehensive monitoring indicators by integrating the PPCA sub-models for various operating modes based on Bayesian inference.

$$BIC_S = \sum_{i=1}^m \left\{ \frac{P_S(x_i|F)P_S(F|x_i)}{\sum_{i=1}^m P_S(x_i|F)} \right\}. \quad (26)$$

During the actual monitoring process, it can be determined that the ICSs have received an attack when the monitoring indicator exceeds the preset threshold.

4. Experimental Results and Analysis

This section verifies the validity of the methodology through practical examples, which are derived from the geological drilling process, and is divided into processes.

4.1. Geological Drilling Process

Geological exploration and resource extraction are contingent upon the successful completion of a geological drilling project. The drilling process is primarily conducted by drill rigs that are equipped with alternative current frequency conversion electric motors. Figure 4 illustrates the schematic of a typical geological drilling process. A few of the components that were used in the drilling process included the crown blocks, moving blocks, derrick, driller's residence, rotary table, drilling control system, mud pump, mud pit, sedimentation pit, drill string, bottom hole assembly, and drill bit. Figure 5 shows a geothermal well construction site with an on-site industrial control system.

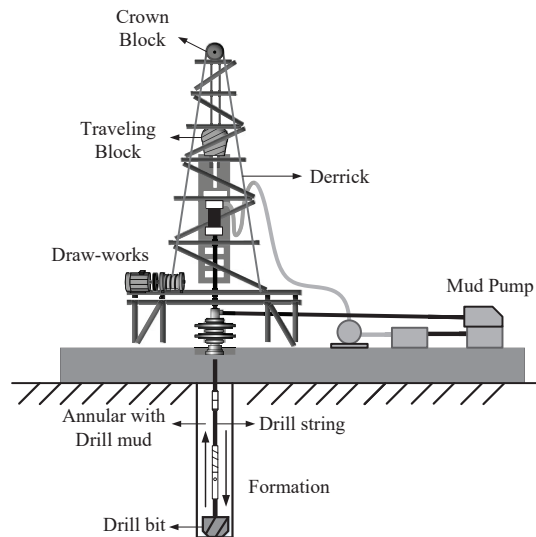


Figure 4. Schematic of a geological drilling process.



Figure 5. The drilling system of a real geological exploration well.

4.2. Overall Results of the ICS Attack Detection

In this paper, real-life case studies with drilling data from a geothermal well demonstrated the effectiveness and superiority of the proposed operating performance monitoring method. The selected running data contains the 12 process variables mentioned in Table 1 from 1052 m to 1058 m, with an interval of 1 s, totaling 2826 data samples. Figure 6 demonstrates time-series data of the actual running process of the ICSs during drilling. Despite the fact that the data injection attack on the network began at 160 s, no significant change was observed in the data curves of the detected variables. Therefore, more in-depth analyses of the data generated in the ICSs are needed to obtain a more accurate portrayal of the ICS operating state.

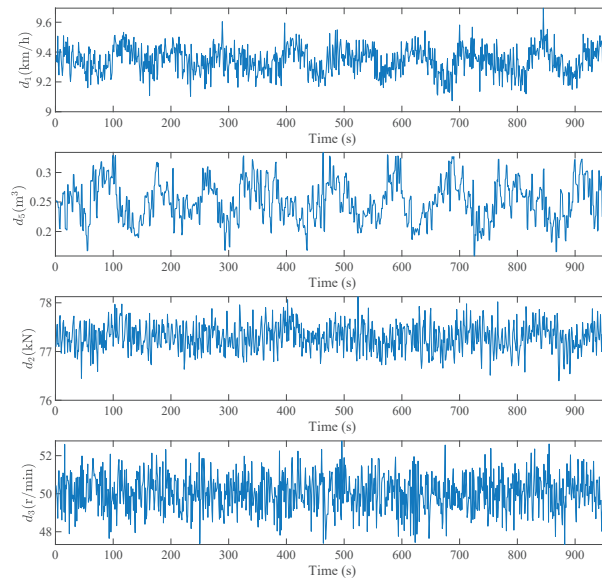


Figure 6. Time series plots of the drilling process under production.

Before constructing the ICS monitoring model, the data set under normal operations was obtained. A standard data matrix was created by selecting 10 decision variables based on a one-way analysis (ANOVA), i.e., X_1, X_2, \dots, X_{10} . According to the blocking MI-based criterion, these variables were divided into three sub-blocks, $[X_1, X_2, X_3, X_9]$, $[X_4, X_5, X_6]$, and $[X_7, X_8, X_{10}]$. For each sub-block, the CVA-PPCA offline monitoring model is established on their canonical subspace, and the calculation of the composite discriminatory indicators and discriminatory thresholds are performed.

During the online monitoring phase, online data are collected according to a window of 20 min, and the monitoring statistic S_{new} is calculated to identify the attack conditions in comparison with the detection threshold. The length of the monitoring window has some effect on the quality of the monitoring. A long window may not detect the fluctuations caused by dual-use attacks, such as, for instance, when there is too short of a window, which may cause frequent alarms and may interfere with the driller's normal operation. Using the industrial control system at the drilling site and manual experience, this study specified a 20-min monitoring window, leading to better results.

Specifically, the principal components of the variables with $\text{CPV} = 98.2\%$ were selected to construct the monitoring model. In all of the monitoring charts, the KDE algorithm was adopted to preset the control limits at a confidence level of $\alpha = 0.05$ and monitoring statics $S_{\text{lim}} = 1.2961$.

In this paper, the anomalous state of the ICSs was the result of two categories of data tampering: surge attacks and biased attacks [30,31]. During a surge attack, a single piece of data is manipulated in order to provide the greatest amount of damage in the shortest amount of time, and it exhibits a step change. Contrary to this, a biased attacker adds non-zero constants to numerous parts of data in a sequence and shows a slow process of change. The monitoring model in this paper was intended to detect the assaults that the system has received by analyzing the monitoring statistics that had been generated by the attacks relevant to the change. Figure 7 illustrates the ICS attack detection results obtained through the proposed method.

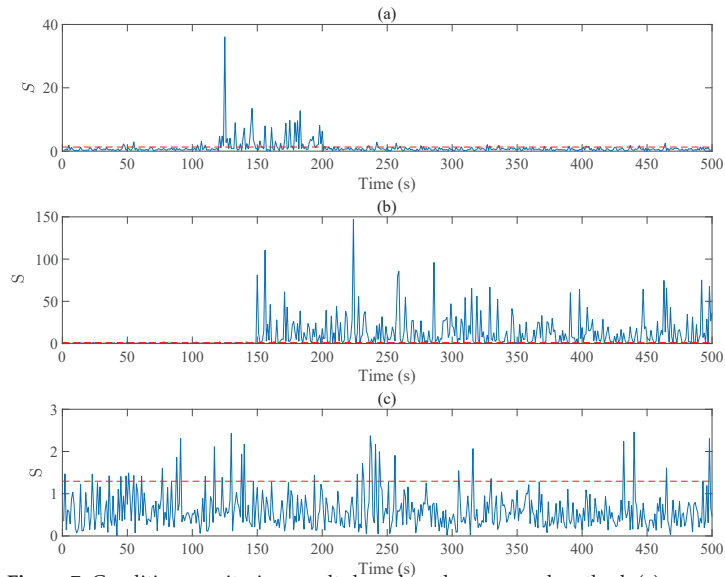


Figure 7. Condition monitoring results based on the proposed method: (a) surge attacks; (b) biased attacks; and (c) the normal conditions.

As shown in Figure 7, the red dashed line indicates the preset control limits, whereas the blue line represents the monitoring statistics calculated from the online data. The surge attack and deviation attack were performed at the 110th seconds of each experiment, as shown in Figure 7a and Figure 7b, respectively. In addition, Figure 7c shows the monitoring results under normal operating conditions. Based on the attacking records, the model successfully identified the impact of the step-wise and slowly varying deviations from the normal operating state. The experimental results revealed that the proposed method can effectively identify anomalies due to attacks with 92.31% accuracy and 12 s monitoring delay.

For greater clarity, the PCA-based process monitoring method was chosen to perform the comparative experiments as a monitoring strategy [32]. To realize the comparison, the integrated monitoring statistics of S_t , achieved by combining T^2 and Q , were adopted in the attack detection task [33]. The control limit was set as $S_t = 7.9127$. It can be seen from Figure 8 that the PCA failed to detect the attacks because there was no significant change in the monitoring statistics. In both cases, the PCA method was less susceptible to the operational instability caused by assaults. As a result of the initial data structure being altered, the anomalies caused by data injection-type attacks did not rapidly accumulate and did not significantly affect the detection data. Consequently, the original PCA method was unable to extract the features that were related to operating conditions, resulting in unsatisfactory monitoring results. The monitoring process also suffered from more misses, false alarms, and longer anomaly detection delays than the method proposed in this study.

To effectively showcase the effectiveness of the proposed method in the monitoring processes, there were some sophisticated process monitoring methods that were selected for comparison such as the original PPCA [15] and mRMR-PCA [32]. The monitoring delay (μ) refers to the period between the incidence of attack performance and the detection of its reasons. Evaluating the performance monitoring involves assessing the non-detection rate (η) and false alarm rate (γ) according to specific criteria. The following matrix proves the definitions of the above indicators

$$\eta = \frac{n_{FP}}{n_{TN} + n_{FP}} \times 100\%, \quad (27)$$

$$\gamma = \frac{n_{FN}}{n_{TP} + n_{FN}} \times 100\%. \quad (28)$$

The variable n_{FN} represents the count of samples that are incorrectly classified into non-optimal modes when they should have been classified into optimal modes. The variable n_{TP} represents the count of samples that were correctly classified into optimal modes. The variable n_{FP} represents the count of samples that are incorrectly classified into optimal modes when they should have been classified into non-optimal modes. Lastly, the variable n_{TN} represents the count of samples that are correctly classified into non-optimal modes. Lower values for η and γ suggest a superior monitoring performance.

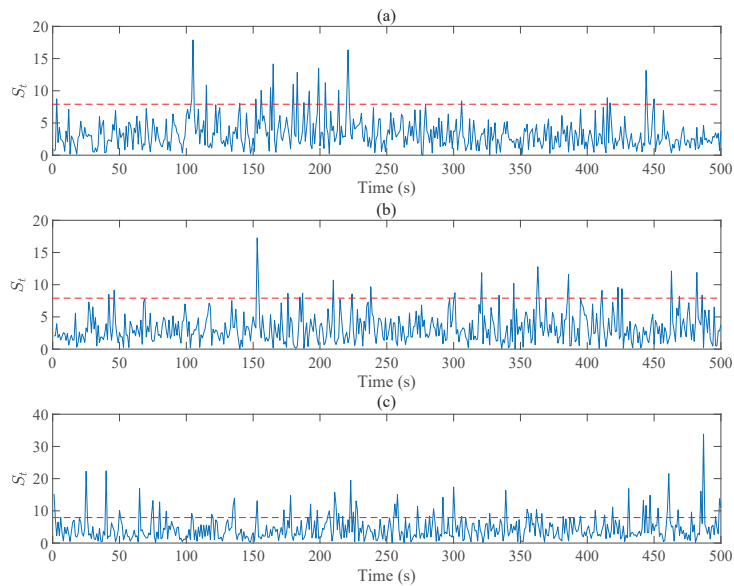


Figure 8. Condition monitoring results based on the PCA method: (a) surge attacks; (b) biased attacks; (c) the normal conditions.

The detection results of the different methods for monitoring data injection attacks are shown in Tables 2 and 3. It is essential to clarify that the typical PCA approach failed to detect both attacks because of its γ for the two statistics, which went up to 74.31% and 83.34%. The PPCA method is inadequate due to its failure to include the non-linear attributes of the data, rendering it unsuccessful in detecting abnormalities. The η of the Q -statistic calculated by mRMR-PCA was 6.05%, but it was 90.64% for the T^2 -statistic in Case 1, which did not meet the needs of field applications. The mRNR-PCA based-monitoring method utilizes a distributed architecture, and, while it did not successfully identify the local attacks, its efficacy was attributed to the singular PCA model. The results show that our method has a comparatively better monitoring performance than the other methods.

In terms of statistical metrics, the maximum enhancement of η and γ reached 69.17 % and 9.67%, respectively, and the shortest detection delays of 11 s and 20 s were achieved in both cases.

Table 2. Attack detection for the different methods.

Type	Indexes	PCA [32]		PPCA [15]	mRMR-PCA [32]		Proposed Method
		T^2	Q	T_s^2	T^2	Q	S
Case 1	η (%)	74.31	21.43	15.2	90.64	6.05	5.14
	γ (%)	3.02	8.21	9.67	14.74	1.93	5.07
	μ (s)	64	42	35	-	40	11
Case 2	η (%)	72.16	83.34	5.14	75.13	5.84	7.76
	γ (%)	2.76	11.13	8.22	3.66	6.16	5.77
	μ (s)	-	48	20	-	104	21

Table 3. Attack detection results obtained with PCA, the original PPCA, mRMR-PCA, and the proposed method.

Type	Indexes	PCA [32]		PPCA [15]	mRMR-PCA [32]		Proposed Method
		T^2	Q	T_s^2	T^2	Q	S
Case 1	False Alarms	252	71	51	306	20	18
	Missed Alarms	10	27	31	48	7	17
	Accuracy(s)	5.33	10.08	77.65	4.41	80.33	90.35
Case 2	False Alarms	245	282	17	255	17	26
	Missed Alarms	7	37	27	10	21	20
	Accuracy(s)	5.21	4.87	83.55	5.67	85.86	94.3

In intuitive terms, the distributed structure ensures that the monitoring model can effectively extract the local and global features with finer-grained precision. In contrast, the typical correlation space combined with the data feature approach captures the latent data features of the ICSs and more accurately portrays the operational state of the process as a whole.

In summary, the proposed approach takes into account the relationship between variable spaces and residual spaces for online monitoring, whereas PCA just evaluates the interaction between variables. The findings suggest that an enhancement in performance monitoring can be achieved by partitioning the initial dataset using PPCA and CVA-based variable reconstruction.

5. Conclusions

This paper proposed a concurrent distributed ICS monitoring method for network attack detection using prior knowledge-based mutual information (MI) and canonical variate analysis with probabilistic principal component analysis (CVA-PPCA). While other centralized process monitoring methods treat all variables as a uniform modeling space, MI-based variable division is capable of probing the underlying local and global characteristics of ICSs comprehensively. Additionally, the CVA-PPCA method established in each sub-block can then more closely reflect and detect the external attack from different aspects.

Due to the complexity and variability of the stratum during geological drilling, as well as the randomness of the network attacks, it was necessary to improve the method's adaptability further by, for example, setting control limits and selecting monitoring windows. Aspects of anomaly tracing and small sample modeling are also important to consider for ICS security when dealing with unknown attack backgrounds. As the study progresses, it will be applied to a variety of industrial processes and recommendations will be provided in the decision-making phase. Further research will also focus on developing a monitoring scheme that takes into account the dynamic nature of variables.

Author Contributions: Conceptualization, M.X. and H.F.; methodology, H.F.; software, Z.J. and S.Y.; validation, M.X., Z.J., S.Y. and H.F.; formal analysis, Z.J.; investigation, S.Y.; resources, M.X.; data curation, H.F.; writing—original draft preparation, M.X., Z.J., S.Y. and H.F.; writing—review and editing, Z.J. and H.F.; visualization, S.Y.; supervision, M.X.; project administration, M.X.; funding acquisition, H.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National defense basic scientific research program (Grant JKCY2022211C007); National defense basic scientific research program (Grant JKCY2021206B104); the “CUG Scholar” Scientific Research Funds at the China University of Geosciences (Wuhan) (Project No. 2023095); the National Natural Science Foundation of China (Grant 6227021554); and the Fundamental Research Funds for the Central Universities, China University of Geosciences.

Data Availability Statement: Data are contained within the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper, and authors of this manuscript also have no conflicts of interest with any organization.

Abbreviations

The following abbreviations are used in this manuscript:

ICSs	Industrial Control Systems
MSPM	Multivariate Statistical Process Monitoring
CVA	Canonical Variate Analysis
PPCA	Probability Principal Component Analysis
ROP	Rate of Permeation
MI	Mutual Information
KDE	Kernel Density Estimate
ANOVA	One-Way Analysis Of Variance
CPV	Cumulative Percentage Variance

References

- Lin, I.C.; Tseng, P.C.; Chang, Y.S.; Weng, T.C. IOTA Data Preservation Implementation for Industrial Automation and Control Systems. *Processes* **2023**, *11*, 2160. [CrossRef]
- Gao, M.; Feng, D. Stochastic stability analysis of networked control systems with random cryptographic protection under random zero-measurement attacks. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 1098–1111. [CrossRef]
- Jia, C.Q.; Feng, D.Q. Industrial control system devices security assessment with multi-objective decision. *Acta Autom. Sin.* **2016**, *42*, 706–714.
- Teixeira, A.; Pérez, D.; Sandberg, H.; Johansson, K.H. Attack models and scenarios for networked control systems. In Proceedings of the 1st international conference on High Confidence Networked Systems, Beijing, China, 17–18 April 2012; pp. 55–64.
- Adepu, S.; Mathur, A. Generalized attacker and attack models for cyber physical systems. In Proceedings of the 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), Atlanta, GA, USA, 10–14 June 2016; Volume 1, pp. 283–292.
- Lu, G.; Feng, D.; Huang, B. Hidden Markov model-based attack detection for networked control systems subject to random packet dropouts. *IEEE Trans. Ind. Electron.* **2020**, *68*, 642–653. [CrossRef]
- Lu, G.; Feng, D. Industrial control network security situation awareness based on improved C-SVC. *Control Decis.* **2017**, *32*, 1223–1228.
- Chaouk, H.; Obeid, E.; Halwani, J.; Arayro, J.; Mezher, R.; Amine, S.; Gazo Hanna, E.; Mouhtady, O.; Younes, K. Application of Principal Component Analysis for the Elucidation of Operational Features for Pervaporation Desalination Performance of PVA-Based TFC Membrane. *Processes* **2024**, *12*, 1502. [CrossRef]
- Huang, K.; Zhang, L.; Wu, D.; Yang, C.; Gui, W. Nonstationary industrial process monitoring based on stationary projective dictionary learning. *IEEE Trans. Control Syst. Technol.* **2022**, *31*, 1122–1132. [CrossRef]
- Yang, C.; Zhang, J.; Wu, D.; Huang, K.; Gui, W. Variable partition based parallel dictionary learning for linearity and nonlinearity coexisting dynamic process monitoring. *Control Eng. Pract.* **2024**, *142*, 105750. [CrossRef]
- Ji, H.; Hou, Q.; Wu, D. Modified performance-enhanced PCA for incipient fault detection of dynamic industrial processes. *J. Process Control* **2023**, *131*, 103107. [CrossRef]
- Li, G.; Qin, S.J.; Ji, Y.; Zhou, D. Reconstruction based fault prognosis for continuous processes. *Control Eng. Pract.* **2010**, *18*, 1211–1219. [CrossRef]

13. Fan, H.; Lai, X.; Du, S.; Yu, W.; Lu, C.; Wu, M. Distributed monitoring with integrated probability PCA and mRMR for drilling processes. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–13. [CrossRef]
14. Kaib, M.T.H.; Kouadri, A.; Harkat, M.F.; Bensmail, A.; Mansouri, M. Improving kernel PCA-based algorithm for fault detection in nonlinear industrial process through fractal dimension. *Process Saf. Environ. Prot.* **2023**, *179*, 525–536. [CrossRef]
15. Zhang, J.; Chen, M.; Hong, X. Nonlinear process monitoring using a mixture of probabilistic PCA with clusterings. *Neurocomputing* **2021**, *458*, 319–326. [CrossRef]
16. Lu, C.; Zeng, J.; Dong, Y.; Xu, X. Streaming variational probabilistic principal component analysis for monitoring of nonstationary process. *J. Process Control* **2024**, *133*, 103134. [CrossRef]
17. Lou, S.; Wu, P.; Yang, C.; Xu, Y. Structured fault information-aided canonical variate analysis model for dynamic process monitoring. *J. Process Control* **2023**, *124*, 54–69. [CrossRef]
18. Li, L.; Dong, F.; Zhang, S. Manifold regularized deep canonical variate analysis with interpretable attribute guidance for three-phase flow process monitoring. *Expert Syst. Appl.* **2024**, *251*, 124015. [CrossRef]
19. Zhang, S.; Bao, X.; Wang, S. Common canonical variate analysis (CCVA) based modeling and monitoring for multimode processes. *Chem. Eng. Sci.* **2023**, *271*, 118581. [CrossRef]
20. Zhang, S.; Zhao, C.; Huang, B. Simultaneous static and dynamic analysis for fine-scale identification of process operation statuses. *IEEE Trans. Ind. Inform.* **2019**, *15*, 5320–5329. [CrossRef]
21. Liu, Y.; Wang, F.; Gao, F.; Cui, H. Hierarchical multiblock T-PLS based operating performance assessment for plant-wide processes. *Ind. Eng. Chem. Res.* **2018**, *57*, 14617–14627. [CrossRef]
22. Zhu, J.; Ge, Z.; Song, Z. Distributed parallel PCA for modeling and monitoring of large-scale plant-wide processes with big data. *IEEE Trans. Ind. Inform.* **2017**, *13*, 1877–1885. [CrossRef]
23. Hu, J.; Wu, M.; Cao, W.; Pedrycz, W. Soft-Sensing of Burn-Through Point Based on Weighted Kernel Just-in-Time Learning and Fuzzy Broad-Learning System in Sintering Process. *IEEE Trans. Ind. Inform.* **2024**, *20*, 7316–7324. [CrossRef]
24. Zhang, J.; Chen, H.; Chen, S.; Hong, X. An improved mixture of probabilistic PCA for nonlinear data-driven process monitoring. *IEEE Trans. Cybern.* **2017**, *49*, 198–210. [CrossRef] [PubMed]
25. Zhong, K.; Ma, D.; Han, M. Distributed dynamic process monitoring based on dynamic slow feature analysis with minimal redundancy maximal relevance. *Control Eng. Pract.* **2020**, *104*, 104627. [CrossRef]
26. Melo, A.; Câmara, M.M.; Pinto, J.C. Data-Driven Process Monitoring and Fault Diagnosis: A Comprehensive Survey. *Processes* **2024**, *12*, 251. [CrossRef]
27. Xiao, B.; Li, Y.; Sun, B.; Yang, C.; Huang, K.; Zhu, H. Decentralized PCA modeling based on relevance and redundancy variable selection and its application to large-scale dynamic process monitoring. *Process Saf. Environ. Prot.* **2021**, *151*, 85–100. [CrossRef]
28. Kim, D.; Lee, I.B. Process monitoring based on probabilistic PCA. *Chemom. Intell. Lab. Syst.* **2003**, *67*, 109–123. [CrossRef]
29. Chen, X.; Zheng, J.; Zhao, C.; Wu, M. Full decoupling high-order dynamic mode decomposition for advanced static and dynamic synergetic fault detection and isolation. *IEEE Trans. Autom. Sci. Eng.* **2022**, *21*, 226–240. [CrossRef]
30. Moudoud, H.; Mlika, Z.; Khoukhi, L.; Cherkaoui, S. Detection and prediction of fdi attacks in iot systems via hidden markov model. *IEEE Trans. Netw. Sci. Eng.* **2022**, *9*, 2978–2990. [CrossRef]
31. Wang, S.; Wang, Y.; Yang, B.; Mo, F.; Zhang, Z. Variational Bayesian Learning with reliable likelihood approximation for accurate Process Quality Evaluation. *IEEE Trans. Ind. Inform.* **2023**, *20*, 815–823. [CrossRef]
32. Xu, C.; Zhao, S.; Liu, F. Distributed plant-wide process monitoring based on PCA with minimal redundancy maximal relevance. *Chemom. Intell. Lab. Syst.* **2017**, *169*, 53–63. [CrossRef]
33. Yue, H.H.; Qin, S.J. Reconstruction-based fault identification using a combined index. *Ind. Eng. Chem. Res.* **2001**, *40*, 4403–4414. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Research on Dimension Reduction Method for Combustion Chamber Structure Parameters of Wankel Engine Based on Active Subspace

Liangyu Li ¹, Yaoyao Shi ¹, Ye Tian ², Wenyan Liu ³ and Run Zou ^{4,*}

¹ School of Mechanical and Electrical Engineering, North University of China, Taiyuan 030051, China; zoyelly@163.com (L.L.); shiyaoyao0522@163.com (Y.S.)

² School of Semiconductors and Physics, North University of China, Taiyuan 030051, China; tianye080t@163.com

³ School of Information and Communication Engineering, North University of China, Taiyuan 030051, China; lwy911218@163.com

⁴ School of Energy and Power Engineering, North University of China, Taiyuan 030051, China

* Correspondence: zourun@nuc.edu.cn

Abstract: The combustion chamber structure of a rotary engine involves a combination of interacting parameters that are simultaneously constrained by engine size, compression ratio, machining, and strength. It is more difficult to study the weight of the effect of the combustion chamber structure on the engine performance using traditional linear methods, and it is not possible to find the combination of structural parameters that has the greatest effect on the engine performance under the constraints. This makes it impossible to optimize the combustion chamber structure of a rotary engine by focusing on important structural parameters; it can only be optimized based on all structural parameters. In order to solve the above problems, this paper proposes a method of dimensionality reduction for the structural parameters of a combustion chamber based on active subspace and combining a probability box and the EDF (Empirical Distribution Function). This method uses engine performance indexes such as explosion pressure, maximum cylinder temperature, and indicated average effective pressure as the influence proportion analysis targets and quantitatively analyzes the influence proportion of combustion chamber structure parameters on engine performance. Eight main structural parameters with an influence of more than 85% on the engine performance indexes were obtained, on the basis of which three important structural parameters with an influence of more than 45% on the engine performance indexes and three adjustable structural parameters with an influence of less than 15% on the engine performance indexes were determined. This quantitative analysis work provides an optimization direction for the further optimization of the combustion chamber structure in the future.

Keywords: rotary engine; combustion chamber structure; active subspace; probability box; EDF; parameter dimension reduction

Citation: Li, L.; Shi, Y.; Tian, Y.; Liu, W.; Zou, R. Research on Dimension Reduction Method for Combustion Chamber Structure Parameters of Wankel Engine Based on Active Subspace. *Processes* **2024**, *12*, 2238. <https://doi.org/10.3390/pr12102238>

Academic Editor: Cheng-Yuan Lin

Received: 6 September 2024

Revised: 4 October 2024

Accepted: 8 October 2024

Published: 14 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As a reliable power system, the internal combustion engine is widely used in many fields. Nowadays, inexpensive and reliable small engines are widely used as propulsion or auxiliary systems in motorcycles, all-terrain vehicles, boats, and small airplanes [1,2]. Although small piston engines are heavily used, they are less efficient compared to large piston engines [3–5]. And unlike a four-stroke reciprocating engine, where the piston stops briefly four times per cycle as the direction of motion changes, the moving parts of a rotary engine maintain a continuous, unidirectional rotary motion, which results in smoother work and less vibration [6–8]. At the same time, the rotary engine also has the advantages of higher power density, simple design, compact structure, and light weight [9–11]. These advantages make rotary engines strong competitors to reciprocating piston engines in areas such as supercharged vehicles and small unmanned aerial vehicles [12–14].

However, due to the narrow and flat combustion chamber shape, high surface area, and low compression ratio, it is difficult for rotary engines to achieve the fuel economy of reciprocating engines [15–18], which restricts the application of rotary engines in various fields [19,20]. In order to improve fuel economy and optimize rotary engine performance, scholars in various countries have conducted a series of investigations on engine structure [21–24]. For example, Kuo et al. investigated the effect of rotor profile on the compression flow of an engine by numerical analysis and found that increasing the shape factor K can benefit the compression efficiency and mixture formation [25]. Tartakovsky et al. started with the in-cylinder flow field and improved the engine performance by optimizing the rotor combustion chamber structure [26]. Jeng et al. further took the degree of mixing between fuel and air as the entry point and improved the engine performance by optimizing the rotor combustion chamber structure [27]. Wei et al. noticed that the position of the rotor combustion chamber has an effect on flame propagation and improved the engine performance by optimizing the rotor combustion chamber position [28–30]. Shi Cheng et al. investigated the role of turbulent blade position in the combustion process of a rotary engine and found that the closer the turbulent blade is to the spark plug chamber, the higher the turbulence velocity and turbulence dissipation rate are in the spark region. It was shown that the combustion rate of the mixture in the combustion chamber with a turbulent blade was accelerated, and it was able to exhibit better combustion characteristics and emission performance [31–34]. Zeng et al. [35] studied the influence of a turbulent blade on flow and combustion performance in the combustion chamber by numerically analyzing the total pressure loss, combustion efficiency, in-cylinder flow, and cylinder temperature changes of the combustion chamber under different spoiler structure parameters. It was found that adding a turbulent blade to the combustion chamber is beneficial to stabilize the combustion and gas mixing, which can further improve the combustion efficiency, improve the outlet temperature distribution, and reduce the pollutant emissions.

However, the above study only focuses on the influence of a certain structural parameter on engine performance and does not quantitatively analyze the proportions of the influence of all combustion chamber structural parameters on engine performance. This leads to the problem of dimensional catastrophe due to too many parameters to be optimized and the problem of not being able to determine the direction of optimization due to unclear parameter optimization priorities when optimizing the combustion chamber structure. Aiming at the combustion chamber structure, which is a complex physical system involving multiple disciplines and multidimensional uncertainties, domestic and foreign scholars have proposed a number of methods for dimensionality reduction. Among them, the gradient solver represented by Sensitivity Analysis (SA) [36] achieves dimensionality reduction by determining the magnitude of the influence of design and state variables on the objective or constraint functions and then filtering out the inputs with less influence. Linear projection methods represented by Classical Multidimensional Scaling (MDS) [37], Principal Component Analysis (PCA) [38], Linear Discriminant Analysis (LDA) [39], and Random Projection (RP) [40] complete the dimensionality reduction and sensitivity analysis processing by mapping the high-latitude problem to the low-latitude space. However, the above methods are not ideal for the dimensionality reduction and quantification of complex nonlinear problems. In this case, a nonlinear method is needed for dimensionality reduction. Common nonlinear methods include Kernel Principal Component Analysis (KPCA) [41], Local Linear Embedding (LLE) [42], Hessian Locally Linear Embedding (HLLE) [43], Laplacian Eigenmaps (LE) [44], and so on. However, the above methods are less computationally efficient in the face of complex, high-dimensional problems, especially in the face of mixed uncertainty problems, and are more sensitive to the selection of samples required for dimensionality reduction, which leads to the limited application of the above methods. In order to solve the problem of mixed uncertainty dimensionality reduction for complex systems, Paul et al. proposed a special dimensionality reduction structure, the Active Subspace (AS), in 2013 [45]. An AS is a low-dimensional action subspace defined by the sensitive directions of the input space, along which input perturbations maximize

the effect on the output. By identifying and utilizing the AS, the dimensionality reduction process can be accelerated while ensuring the analysis accuracy.

In this paper, on the basis of previous research, based on the AS, by combining the probability box with the interval variable Empirical Distribution Function (EDF) [46,47], a method of dimensionality reduction of the structural parameters of the combustion chamber of a rotary engine is proposed. A total of 14 combustion chamber structural parameters, including the turbulent blade, were subjected to eigenvalue estimation and dimensionality reduction, and the weights of the effects of different structural parameters on the maximum cylinder pressure, the maximum cylinder temperature, and the average effective pressure were obtained. Eight main structural parameters, including the angle θ_1 between the bottom edge of the middle part of the combustion chamber and the Y-Z plane were further obtained, and the influence of these structural parameters on the engine performance indexes accounted for more than 85%. Three important structural parameters, including the angle θ_2 between the bottom edge of the rear part of the combustion chamber and the Y-Z plane, were obtained on this basis, and the influence of these structural parameters on the engine performance indexes accounted for more than 45%. Three adjustable structural parameters, including the length l of the bottom edge of the front of the combustion chamber, were obtained. The influence of these three structural parameters on the engine performance index accounts for a relatively small percentage, below 15%, but their influence on the compression ratio and other structural requirements is larger and can be adjusted according to the compression ratio, process, and strength requirements after determining the main structural parameters. This quantitative analysis work provides an optimization direction and theoretical support for the further optimization of the combustion chamber structure in the future. At the same time, this study provides a reference basis for eigenvalue estimation and dimension reduction methods for multidimensional complex engineering problems.

2. Uncertainty Analysis of Combustion Chamber Structure Parameters

In practical engineering, the probability distribution is divided into random uncertainty and cognitive uncertainty [48,49]. The former has a known distribution function, and the latter lacks such a function.

This study investigates how structural parameter uncertainty in the combustion chamber impacts single-cylinder gasoline Wankel engine performance. Table 1 lists the engine's structural parameters.

Table 1. Rotor engine structure parameters.

Parameter	Value
Creation radius (mm)	60
Eccentricity (mm)	10
Rotor width (mm)	42
Translational distance (mm)	1.45
Speed (r/min)	8000
Single cylinder volume (cc)	133
Compression ratio	10
Ignition advance angle	30° BTDC
Ignition source	Single spark plug
Fuel type	Gasoline
Jet strategy	Pre-mixed intake duct
Air-fuel ratio	1
Ignition diameter (mm)	8
Ignition energy (J)	0.08
Engine power (kw)	12.3
Torque (N.m)	15.66
Oil consumption (g/kw.h)	402.3

The rotor engine combustion is categorized into four parts: leading, middle, trailing, and spoiler. Rounded corners smooth the bottom surfaces. The cross sections of the leading, middle, and trailing parts of the combustion chamber are ovals formed by two semicircles and a rectangle, while the spoiler plate is a cuboid with a height below the rotor profile (Figure 1).

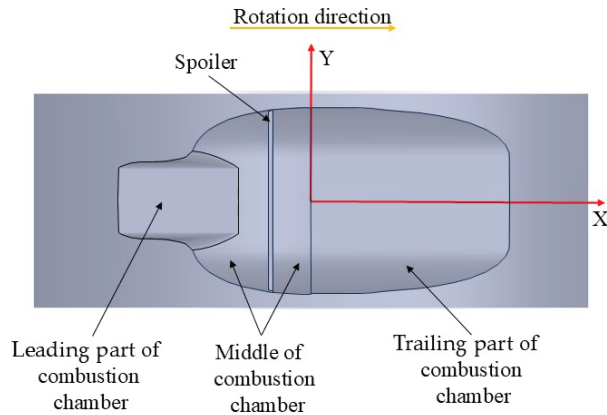


Figure 1. Wankel engine combustion chamber structure.

The Wankel engine's combustion chamber consists of three elliptical cross sections and an oblong spoiler plate. The middle and trailing parts share the same cross-sectional dimensions but differ in inclination angles. Figures 2 and 3 depict the size schematics, and Table 2 compares the variables and structural parameters.

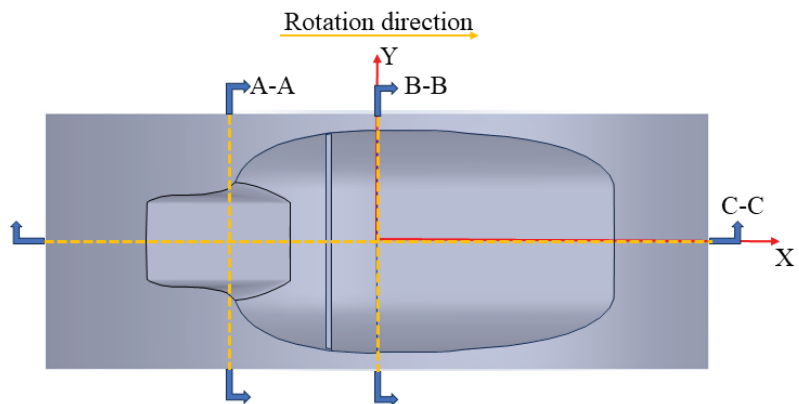


Figure 2. Variable comparison table.

To maintain the same compression ratio, the rotor combustion chamber must meet eight structural, process, and performance requirements:

- (1) The distance from the deepest part of the rotor combustion chamber to its center of gravity on the x - z plane must not exceed h_{α} , considering structural strength, heat dissipation structure arrangement, and spindle assembly.
- (2) The maximum width of the chamber must not exceed l_{α} , addressing seal arrangement and structural strength.
- (3) The distance between the chamber ends and the rotor tip on the y - z plane must not exceed k_{α} , addressing radial seal arrangement and structural strength.

- (4) The width of the rotor spoiler plate must not exceed c_a , ensuring spoiler plate strength.
- (5) According to research, the spoiler plate height should be at least $\frac{4b_a}{5}$ [26–30], where b_a signifies the vertical distance from the chamber bottom to the rotor profile at the spoiler plate position.
- (6) To simplify processing, the rotor spoiler should be positioned centrally or toward the rear of the combustion chamber, avoiding overlap between chamber sections.
- (7) The leading and middle parts of the combustion chamber must be connected, with a single recess on the rotor surface.
- (8) The arc radius of the transition section must not exceed q_a .

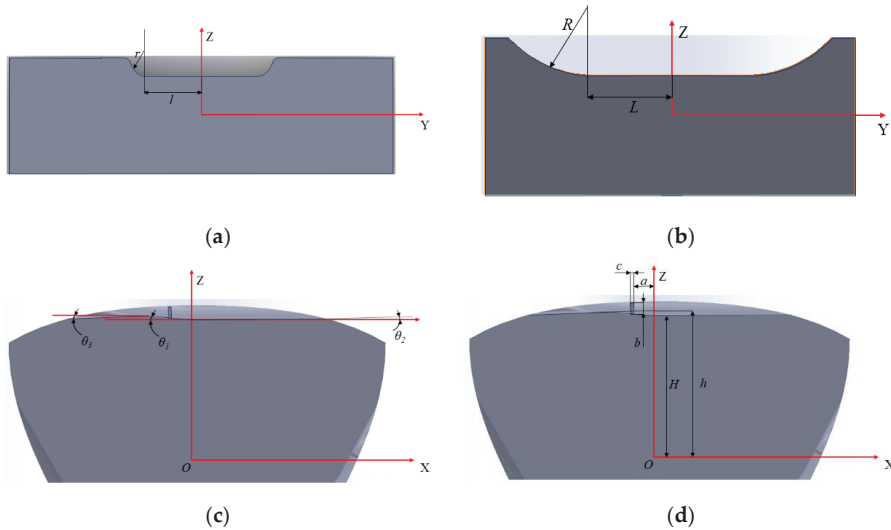


Figure 3. Combustion chamber dimensions. (a) A-A section size. (b) B-B section size. (c) C-C section angle. (d) C-C section size.

Table 2. Variable comparison.

Variable	Structure Name
r	Combustion chamber leading arc radius
l	Length of the leading bottom edge of the combustion chamber
h	The distance from the bottom of the combustion chamber's leading section to the rotor's center of gravity at the x - z plane
θ_3	The angle between the combustion chamber's leading bottom edge and the y - z plane
R	Trailing arc radius in the combustion chamber
L	Trailing bottom edge length of the combustion chamber
H	The distance from the bottom of the trailing section of the combustion chamber to the rotor's center of gravity at the x - z plane
θ_1	The angle between the middle bottom edge of the combustion chamber and the y - z plane
θ_2	The angle between the trailing bottom edge of the combustion chamber and the y - z plane
a	The vertical distance between the spoiler plate and the rotor x - z section
b	The distance from the spoiler's top center to the combustion chamber's bottom edge
c	The spoiler plate's width
q_1	The excessive arc radius from the leading part to the middle of the combustion chamber
q_2	The excessive arc radius from the middle to the trailing part of the combustion chamber

To ensure that the generated structural parameter combination meets the above requirements, the entire combustion chamber structure is parameterized first, so that a series of formulas can be used to calculate whether a certain structural combination meets the

requirements of compression ratio, size, processing, and strength. When generating structural combinations, the values of several structural parameters are randomly generated within the required range of size, processing, and strength. Based on this, the compression ratio is used as the optimization objective; the size, processing, and strength requirements are used as the range; and the remaining structural parameters are used as the solving parameters for optimization iteration using optimization algorithms.

Using the parameters in Table 3, 100 sets of structural parameter combinations were generated, including the 14 parameters listed. Figures 4–17 depict their distribution and probabilities. From these figures, c , q_1 , and q_2 follow a uniform distribution, while a follows a normal distribution. The distributions of r , l , h , θ_3 , R , L , H , θ_1 , θ_2 , and b are unclear, representing cognitive uncertainty. Consequently, the dimension reduction problem involves 14 dimensions with 4 random and 10 cognitive uncertainty variables.

Table 3. Constraints.

Constraint Name	Variable Name	Value
The minimum distance from the bottom of the combustion chamber to the center of gravity of the rotor	h_α	42 mm
Maximum width of the combustion chamber	l_α	40 mm
The projection distance between the two ends of the rotor combustion chamber and the rotor tip on the y - z section	k_α	10 mm
Maximum width of the spoiler plate	c_α	3 mm
The maximum radius of the excessive arc	q_α	2 mm

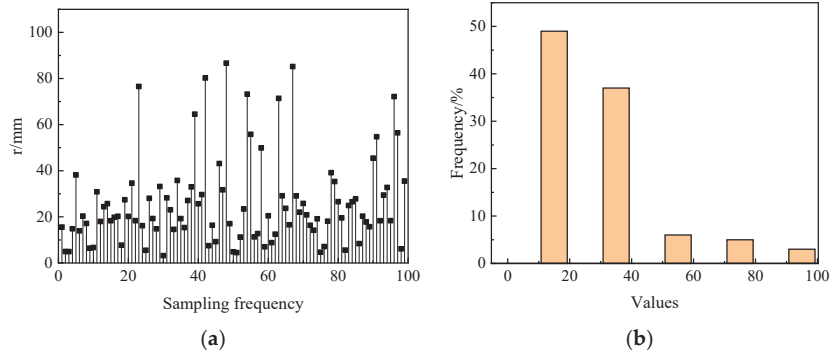


Figure 4. Distribution of the arc radius (r) at the leading edge of the combustion chamber. (a) Distribution conditions. (b) Distribution probability.

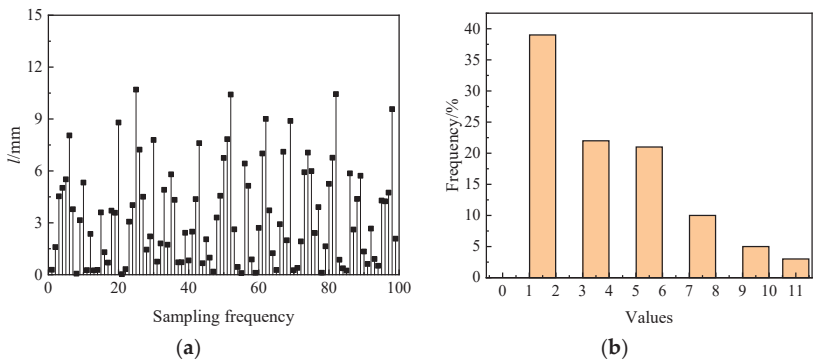


Figure 5. Distribution of the length (l) at the leading bottom edge of the combustion chamber. (a) Distribution conditions. (b) Distribution probability.

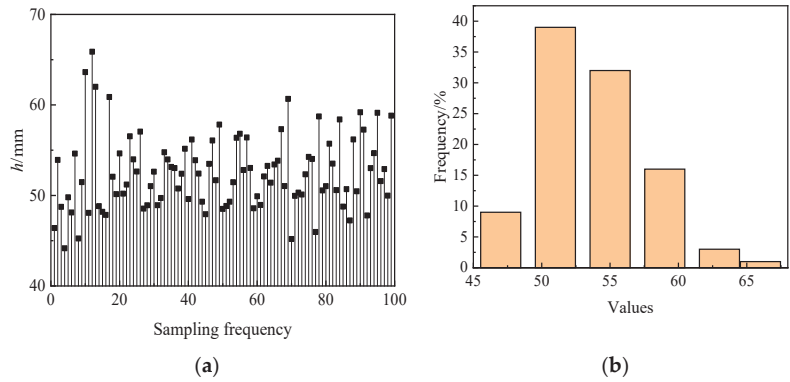


Figure 6. Distribution of the distance (h) from the front section bottom of the combustion chamber to the rotor's center of gravity in the x - z plane. (a) Distribution conditions. (b) Distribution probability.

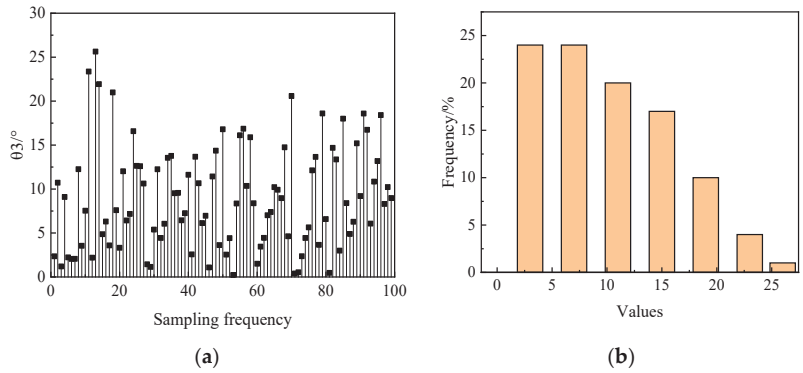


Figure 7. Distribution of the angle (θ_3) between the leading bottom edge of the combustion chamber and the y - z plane. (a) Distribution conditions. (b) Distribution probability.

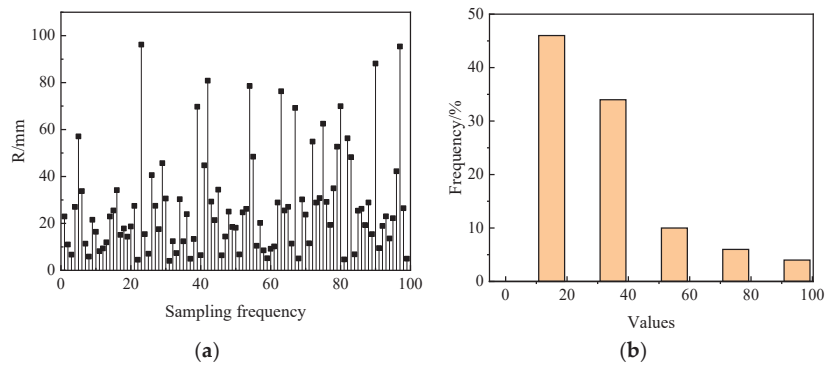


Figure 8. Distribution of the length (R) at the trailing bottom of the combustion chamber. (a) Distribution conditions. (b) Distribution probability.

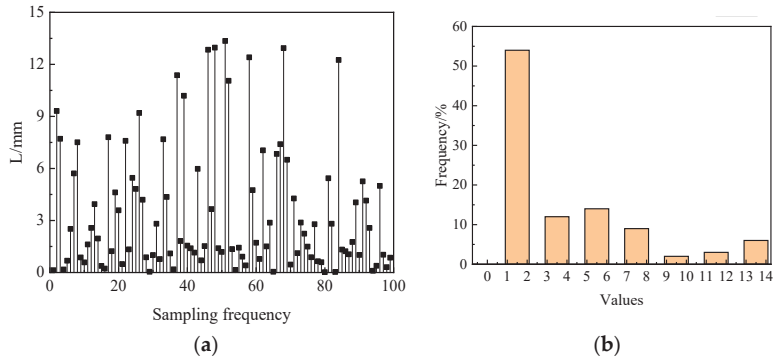


Figure 9. Distribution of the length (L) at the trailing bottom of the combustion chamber. (a) Distribution conditions. (b) Distribution probability.

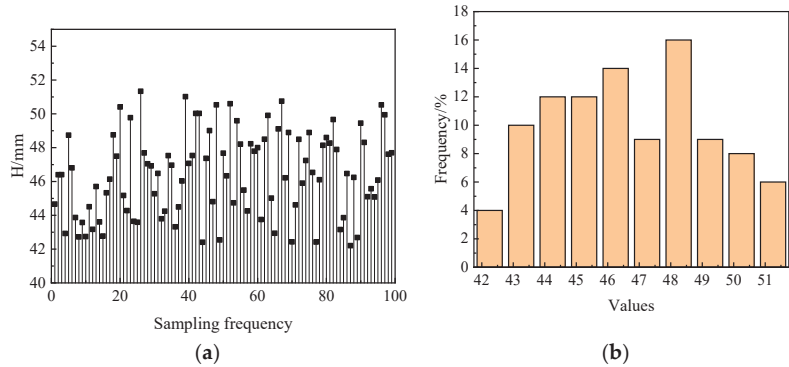


Figure 10. Distribution of the distance (H) from the trailing chamber bottom to the rotor's center of gravity in the x - z plane. (a) Distribution conditions. (b) Distribution probability.

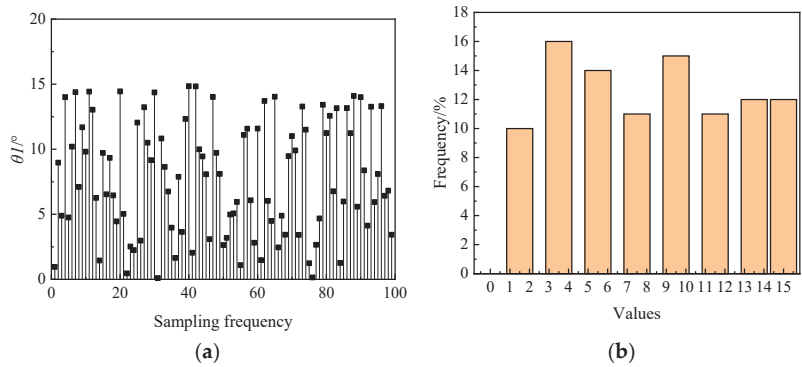


Figure 11. Distribution of the angle (θ_1) between the combustion chamber's middle bottom edge and the y - z plane. (a) Distribution conditions. (b) Distribution probability.

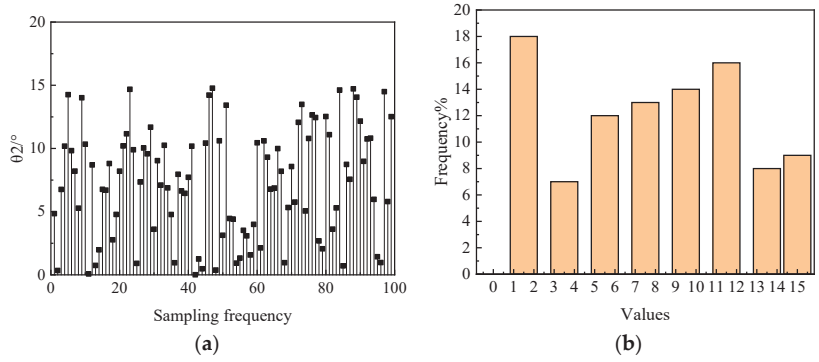


Figure 12. Distribution of the angle (θ_2) between the trailing bottom of the combustion chamber and the y-z plane. (a) Distribution conditions. (b) Distribution probability.

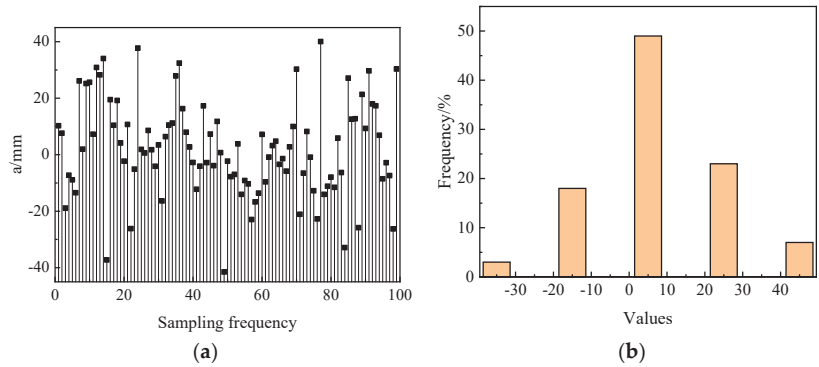


Figure 13. Distribution of the vertical distance (a) between the spoiler plate and the rotor x-z section. (a) Distribution conditions. (b) Distribution probability.

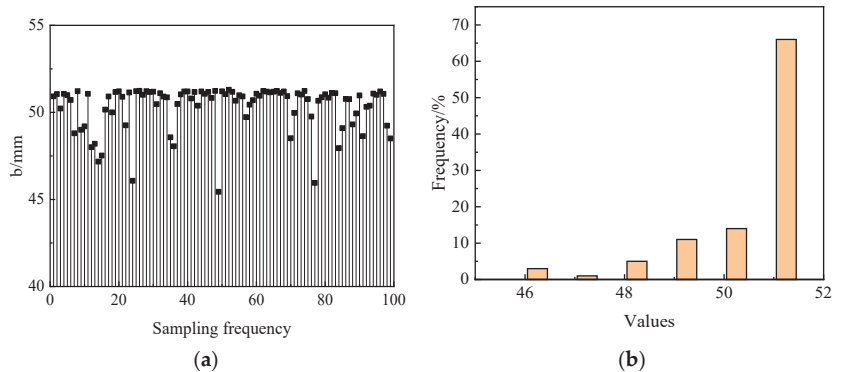


Figure 14. Distribution of the distance (b) from the spoiler top center to the chamber bottom edge. (a) Distribution conditions. (b) Distribution probability.

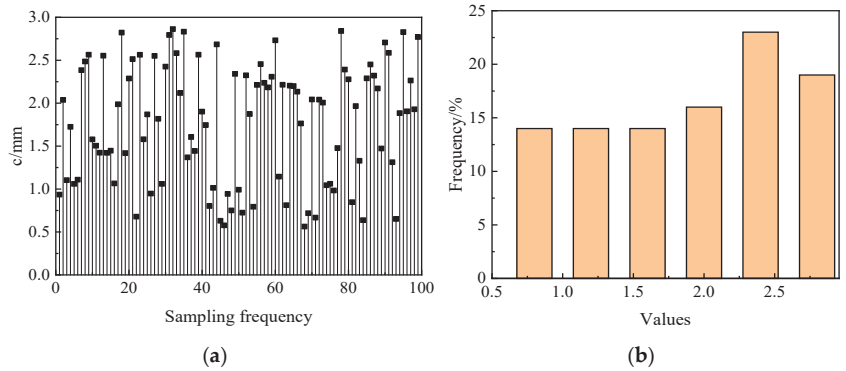


Figure 15. The distribution of the spoiler plate width (c). (a) Distribution conditions. (b) Distribution probability.

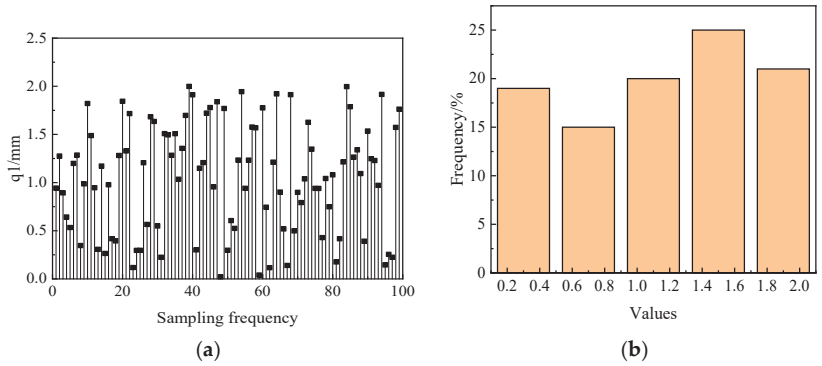


Figure 16. Distribution of the excessive arc radius (q_1) between the leading and middle chamber sections. (a) Distribution conditions. (b) Distribution probability.

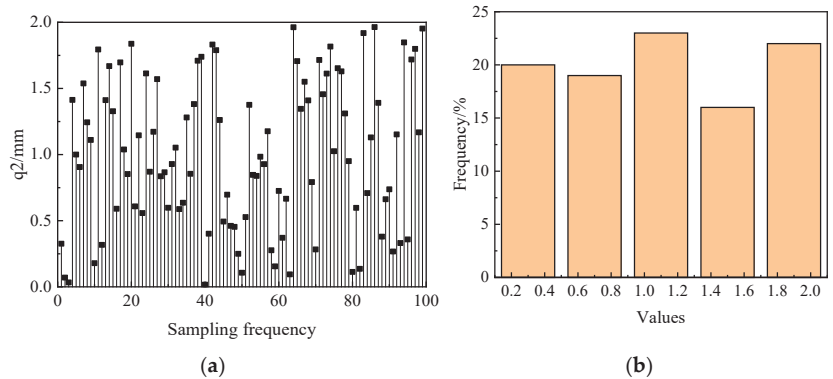


Figure 17. Distribution of the excessive arc radius (q_2) between the middle and trailing chamber sections. (a) Distribution conditions. (b) Distribution probability.

3. Estimation Method of Active Subspace Eigenvalues Based on Probability Box–EDF

Recent advancements in dimensionality reduction techniques such as principal component analysis, factor analysis, topological mapping regression, and random projection have been significant [50–52]. Building on principal component analysis, the active subspace method evaluates input parameters through the output covariance matrix. This

technique has been used in transonic wing design optimization, hydrological model construction, and satellite optimization, demonstrating benefits for complex system problems at high latitudes.

Unlike principal component analysis, which focuses on eigenvalue size impact on the covariance, the active subspace method constructs a new subspace by adjusting eigenvectors to significantly alter the system output direction [53,54]. For a system f with initial dimension n , its input x , output f , output function gradient $f\nabla$, and reduced dimension input v can be expressed as follows:

$$\begin{cases} x = [x_1, x_2, x_3, \dots, x_n]^T \\ f = f(x) \\ f\nabla = f\nabla(x) \\ v' = [v'_1, v'_2, v'_3, \dots, v'_m]^T \\ n > m \end{cases} \tag{1}$$

Normalizing each boundary of the standard parameter space to $[0, 1]$ allows the non-standard parameter space to be transformed through regularization. The input parameter x with N samples can be expressed as follows:

$$X = \{x_{i,j}; 1 \leq i \leq n, 1 \leq j \leq N\} \tag{2}$$

Any member of matrix X can be regularized as follows:

$$\begin{cases} x'_{i,j} = \frac{x_{i,j} - \hat{\mu}_i}{\hat{\sigma}_i} \\ \hat{\mu}_i = \frac{1}{N} \sum_{j=1}^N x_{i,j} \\ \hat{\sigma}_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (x_{i,j} - \bar{x}_i)^2} \end{cases} \tag{3}$$

The gradient covariance matrix of f is as follows:

$$\begin{cases} C = \int f\nabla(x)f\nabla(x)^T \rho(x)dx = W\Lambda W^T \\ \Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n) \end{cases} \tag{4}$$

where W signifies an orthogonal matrix, Λ represents a non-negative eigenvalue matrix, and the eigenvalue's relative size indicates the variables' contribution to the objective function. In practical engineering, obtaining the transfer function $f(x)$ for the gradient covariance matrix is challenging, so the limited difference computation method is often used to calculate the sample outputs to derive $f\nabla$ and C [55,56]. The gradient covariance matrix can be expressed as follows:

$$C = \frac{1}{N} \sum_{i=1}^N f\nabla_i(x)f\nabla_i(x)^T \tag{5}$$

The eigenvalues in matrix Λ are arranged in descending order, yielding $\Lambda' = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n)$. The first m eigenvalues are chosen as the new active subspace dimensions, with the selection criteria shown in Equation (6).

$$\begin{cases} \frac{\lambda_m}{\sum_{i=1}^m \lambda_i} \geq k \\ \frac{\lambda_{m+1}}{\sum_{i=1}^m \lambda_i} \leq k \end{cases} \tag{6}$$

where k denotes the proportional coefficient. The eigenvalue matrix Λ' is divided into the first m order Λ'_1 and the last $n-m$ order Λ'_2 .

$$\Lambda' = \begin{bmatrix} \Lambda'_1 & \\ & \Lambda'_2 \end{bmatrix} \tag{7}$$

Equation (4) can be expressed as

$$C = W\Lambda W^T = [U \ V] \begin{bmatrix} \Lambda'_1 & \\ & \Lambda'_2 \end{bmatrix} [U \ V]^T \tag{8}$$

where U denotes the basis vector for the active eigenvalue. After dimension reduction, the input v can be expressed as follows:

$$v = U^T x \tag{9}$$

The objective function can be approximated as follows:

$$f(x) \approx g(U^T x) \tag{10}$$

The error between the original and the dimension-reduced objective functions can be expressed as shown below:

$$\int f(x) - g(U^T x) \rho dx \tag{11}$$

The active subspace method, widely used across various fields, faces limitations due to input variable uncertainty. From Equation (4), the gradient covariance matrix of f can only be calculated if the probability distribution $\rho(x)$ of all input variables is known. Once this distribution is clear, the gradient covariance matrix can be calculated using the transfer function or samples, leading to the eigenvalue matrix Λ . Consequently, the subspace method for dimension reduction is applicable only if all input independent variables are random with a clear probability distribution. This limitation restricts the active subspace method's use in engineering. Further research is needed on dimensionality reduction for cognitive and mixed uncertainty problems in engineering [57].

An objective function f with p random uncertainty variables, q cognitive uncertainty input independent variables, and n mixed uncertainty input independent variables can be expressed as follows:

$$\begin{cases} f = f(x_a, x_e) \\ x_a \sim \rho(x_a) \\ x_e \in [x_e^L, x_e^U] \\ x_a = \{x_{a1}, x_{a2}, x_{a3}, \dots, x_{ap}\} \\ x_e = \{x_{e1}, x_{e2}, x_{e3}, \dots, x_{eq}\} \end{cases} \tag{12}$$

The gradient of each independent variable can be expressed as

$$\nabla_x f^I = \left[\frac{\partial f}{\partial x_{a1}} \dots \frac{\partial f}{\partial x_{ap}} \frac{\partial f}{\partial x_{e1}} \dots \frac{\partial f}{\partial x_{eq}} \right]^T \tag{13}$$

where the real interval vector $\nabla_x f^I$, its upper limit \bar{c}_i , and its lower limit \underline{c}_i can be expressed as follows:

$$\begin{cases} \bar{c}_i = \max \left\{ \frac{\partial f}{\partial x_i}, x_a \sim \rho(x_a), x_e \in [x_e^L, x_e^U] \right\} \\ \underline{c}_i = \min \left\{ \frac{\partial f}{\partial x_i}, x_a \sim \rho(x_a), x_e \in [x_e^L, x_e^U] \right\} \end{cases} \tag{14}$$

Combining Equations (13) and (14) indicates that when $q = 0$, $\nabla_x f^I = \nabla_x f$. When $q \neq 0$, the gradient covariance matrix C in Equation (4) can be expressed as

$$C = \hat{C}^I = \frac{1}{M} \sum_{k=1}^M (\nabla_x f_k^I) \cdot (\nabla_x f_k^I)^T \tag{15}$$

where \hat{C}^I is a real symmetric interval matrix with M samples. Each element in \hat{C}^I can be expressed as

$$\begin{aligned} \hat{C}_{ij}^I &= \frac{1}{M} \sum_{k=1}^M \left(\frac{\partial f_k^I}{\partial x_i} \right) \cdot \left(\frac{\partial f_k^I}{\partial x_j} \right) \\ &= [\hat{C}_{ij}^k, \bar{C}_{ij}^k] \\ &= \left[\frac{1}{M} \sum_{k=1}^M c_{ij}^k, \frac{1}{M} \sum_{k=1}^M \bar{c}_{ij}^k \right] \\ &\quad i, j = 1, \dots, n \end{aligned} \tag{16}$$

where $[c_{ij}^k, \bar{c}_{ij}^k]$ can be expressed as follows:

$$\begin{aligned} [c_{ij}^k, \bar{c}_{ij}^k] &= [c_i^k, \bar{c}_i^k] \times [c_j^k, \bar{c}_j^k] \\ &= \left[\min(c_i^k \cdot c_j^k, c_i^k \cdot \bar{c}_j^k, \bar{c}_i^k \cdot c_j^k, \bar{c}_i^k \cdot \bar{c}_j^k), \max(c_i^k \cdot c_j^k, c_i^k \cdot \bar{c}_j^k, \bar{c}_i^k \cdot c_j^k, \bar{c}_i^k \cdot \bar{c}_j^k) \right] \end{aligned} \tag{17}$$

Each feature pair in \hat{C}^I can be expressed as shown below:

$$\begin{aligned} &(\lambda_i^I, w_i^I) \\ &i = 1, \dots, n \end{aligned} \tag{18}$$

Based on the analysis, the mixed uncertainty dimension reduction problem is transformed into the eigenvalue interval problem of \hat{C}^I expressed as follows:

$$\begin{aligned} \lambda^I &= [\underline{\lambda}, \bar{\lambda}] \\ &= [\underline{\lambda}_i, \bar{\lambda}_i] \\ &i = 1, \dots, n \end{aligned} \tag{19}$$

According to the Deif theorem, if a matrix retains the eigenvector component's sign within the interval, the upper and lower bounds of the eigenvalues can be determined. The sign matrix that remains unchanged on \hat{C}^I is expressed as follows [58].

$$\begin{aligned} S^i &= \text{diag}(\text{sgn}(w_i)) \\ &i = 1, \dots, n \end{aligned} \tag{20}$$

$$\begin{cases} (\hat{C}^c - S^i \Delta \hat{C} S^i) \underline{w}_i = \underline{\lambda}_i \underline{w}_i \\ (\hat{C}^c + S^i \Delta \hat{C} S^i) \bar{w}_i = \bar{\lambda}_i \bar{w}_i \\ i = 1, \dots, n \end{cases} \tag{21}$$

where \hat{C}^c signifies the median matrix of \hat{C}^I , and $\Delta \hat{C}$ represents the radius matrix of \hat{C}^I , expressed as

$$\begin{cases} \hat{C}^c = (\hat{C}_{ij}^c) = \frac{\hat{C}_{ij} + \bar{C}_{ij}}{2} \\ \hat{C}_{ij}^c = \frac{\hat{C}_{ij} + \bar{C}_{ij}}{2} \\ i = 1, \dots, n \\ \Delta \hat{C} = (\Delta \hat{C}_{ij}^c) = \frac{\hat{C}_{ij} - \bar{C}_{ij}}{2} \\ \Delta \hat{C}_{ij}^c = \frac{\hat{C}_{ij} - \bar{C}_{ij}}{2} \\ i = 1, \dots, n \end{cases} \tag{22}$$

where $sgn(w_i)$ denotes the vector of the symbol function for each element of w_i . If all elements in w_i have the same sign, then the following is true:

$$\begin{cases} \hat{C} w_i = \lambda_i w_i \\ \hat{C} \bar{w}_i = \bar{\lambda}_i \bar{w}_i \\ i = 1, \dots, n \end{cases} \tag{23}$$

The maximum and minimum values of $|w_i|$ can be computed as follows, where I is the unit matrix.

$$\begin{cases} \lambda_i - \Delta \hat{C} |w_i| \leq (\lambda_i I - S^i \hat{C}^c S^i) |w_i| \leq \Delta \hat{C} |w_i| \\ \begin{bmatrix} \lambda_i I - S^i \hat{C}^c S^i - \Delta \hat{C} \\ \hat{C}^c S^i - \lambda_i I - \Delta \hat{C} \end{bmatrix} |w_i| \leq 0 \\ i = 1, \dots, n \end{cases} \tag{24}$$

A positive semi-definite real symmetric matrix C has an orthogonal matrix W , such that

$$W^T \hat{C} W = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \tag{25}$$

For $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, we have

$$\begin{cases} (W w_i)^T \hat{C} W w_i = w_i^T \Lambda w_i = \lambda_1 w_1^2 + \lambda_2 w_2^2 + \dots + \lambda_n w_n^2 \geq 0 \\ w_i = [w_1, w_2, \dots, w_n]^T \\ \lambda_i \geq 0 \\ i = 1, \dots, n \end{cases} \tag{26}$$

where

$$\begin{aligned} \lambda_i &= w_i^T \hat{C} w_i \\ &= w_i^T \left[\frac{1}{M} \sum_{k=1}^M (\nabla_x f_k^l) \cdot (\nabla_x f_k^l)^T \right] w_i \\ &= \frac{1}{M} \sum_{k=1}^M [(\nabla_x f_k)^T w_i]^2 \geq 0 \\ & \quad i = 1, \dots, n \end{aligned} \tag{27}$$

where w_i denotes the effect of the i -th input variable on the objective function. Combined with Equations (4)–(11), it yields dimensionality reduction of mixed uncertainty.

In deriving a mixed uncertainty dimension reduction problem from Equations (20)–(27), a key requirement is that the gradient covariance matrix \hat{C}^l must satisfy the Deif theorem, maintaining the sign of eigenvector components within the interval. In practice, it is challenging to verify without the objective function (f) or sufficient data. To tackle large computational load, dimensionality, and prior condition challenges in mixed uncertainty reduction, eigenvalue estimation techniques and gradient response estimation approaches have been proposed. The former include the disc method, perturbation method, and spectral radius method. The latter represent direct optimization, eigenvalue analysis, and modal decomposition. Eigenvalue methods only approximate ranges and cannot obtain upper and lower bounds, while gradient response methods lose accuracy with high system uncertainty [50].

This study proposes a mixed uncertainty eigenvalue estimation method using the EDF and probability boxes [59–61] for dimension reduction, as illustrated in Figure 18.

The EDF describes variable distributions by approximating sample frequencies to the probability distribution of random variables as a step function, resulting in the Cumulative Distribution Function (CDF). This method converts the mixed uncertainty dimension reduction into a random uncertainty problem. Figure 19 shows a CDF with an EDF distribution, where the black line shows the EDF fit and the histogram indicates variable probabilities [62,63].

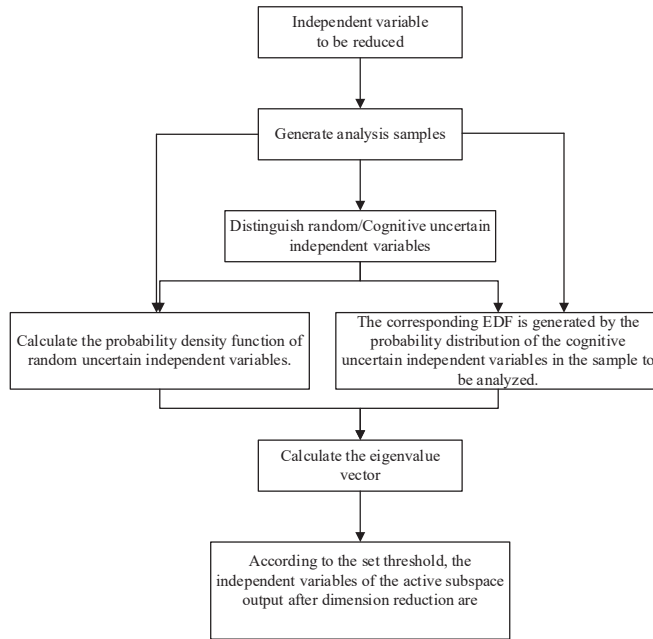


Figure 18. Eigenvalue estimation of the mixed uncertainty active subspace using the EDF and probability box.

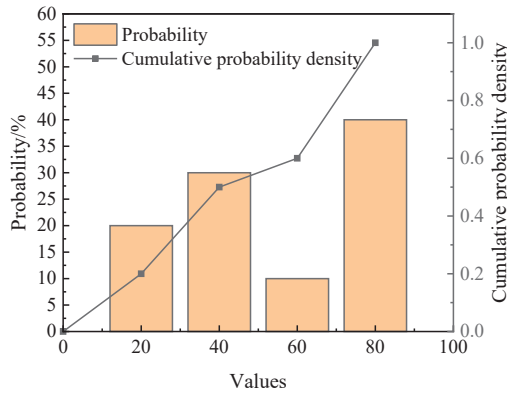


Figure 19. EDF distribution.

In practice, due to randomness in input variables, the empirical density is fitted using an equal distribution model. The empirical density function for mixed uncertainty is

$$\hat{F}_n(x) = \frac{\sum_{i=1}^n I(X_i \leq x)}{n} \tag{28}$$

where the EDF is fitted by $\hat{F}_n(x)$; X_i represents an independent and identically distributed random variable; x_i denotes the function value of the empirical distribution function at x ; and n signifies the number of samples. The accuracy of the empirical density function increases with more samples.

All cognitive uncertainty variables can be expressed as random uncertainty variables.

$$\begin{aligned}
 y &= \hat{W}_1^T(x_a, x_e) \\
 x_a &\sim \rho_{x_a}, x_e \sim \rho_{x_e}
 \end{aligned}
 \tag{29}$$

where ρ_{x_a} and ρ_{x_e} denote the sampling weights of random and cognitive uncertainties based on the distribution function and fitted differential EDF, respectively. Integrating these variables into the equation yields the following:

$$C = \frac{1}{N} \sum_{i=1}^N y \nabla_i(x) y \nabla_i(x)^T
 \tag{30}$$

Eigenvalues for the mixed uncertainty active subspace are determined using Equation (27), enabling mixed uncertainty dimensionality reduction. Probability boxes representing variable distributions are used to obtain EDF curves. These boxes incorporate interval concepts for both random and cognitive uncertainties. Figure 20 illustrates a probability box variable x with cumulative distribution function $\hat{F}_n(x)$, upper bound $\overline{\hat{F}_n}(x)$, and lower bound $\underline{\hat{F}_n}(x)$. The probability box requires only upper and lower distribution bounds to split the boundary around the variable distribution function, eliminating the need for specific distribution forms and ensuring that the cumulative distribution function lies within the defined boundaries.

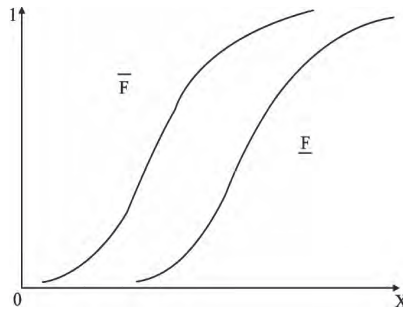


Figure 20. Probability box.

To obtain the probability box of variables, we use the D-S theory composed of multiple focal elements. For a sample in space R , any D-S structure can be expressed as

$$\begin{cases} m : 2^R \rightarrow [0, 1] \\ m(\emptyset) \neq 0, \sum_{A \subseteq R} m(A) = 1 \end{cases}
 \tag{31}$$

where each interval A is a focal element; $m(A)$ denotes the reliability value corresponding to this focal element; and 2^R signifies the power set of R . The trust function Bel and the likelihood function Pl can be expressed as follows:

$$\begin{cases} Bel(A) = \sum \{m(B) | B \subseteq A, B \neq \emptyset\} \\ Pl(A) = \sum \{m(B) | A \cap B, B \neq \emptyset\} \end{cases}
 \tag{32}$$

Probability boxes consist of multiple D-S structures, the upper and lower bounds of which are accumulated to attain the left boundary of the upper and lower bounds of the probability box. The D-S structure represents a discretized probability box. Figure 21 illustrates the relationship between the probability box and the structure. EDF fitting using probability boxes is conducted as follows:

- Step 1: Resample N samples into N groups, and divide each group into M segments from largest to smallest.

- Step 2: Use each segment’s maximum and minimum values as upper and lower bounds, respectively. If the distribution aligns with a certain distribution model, the upper and lower bounds of each group’s probability are obtained based on the distribution model; otherwise, the bounds are obtained based on the EDF probability distribution in Equation (28).
- Step 3: Average the maximum upper bounds and minimum lower bounds across N groups to obtain the fitted cumulative probability curve.

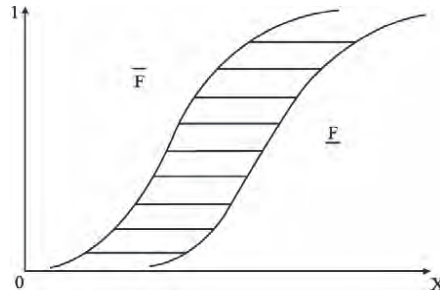


Figure 21. Probability box and structure.

To verify the accuracy of the above methods, this study evaluates an objective function with random and cognitive uncertainties, as shown in Equation (33), where x_1 denotes a random variable uniformly distributed in $[0, 1]$, and x_2 represents a cognitive variable in $[0, 3]$. Figure 22 shows the distribution and probabilities from 100 samples, which do not fit any standard distribution model.

$$f(x_1, x_2) = e^{x_1} + x_2 + x_1x_2 \tag{33}$$

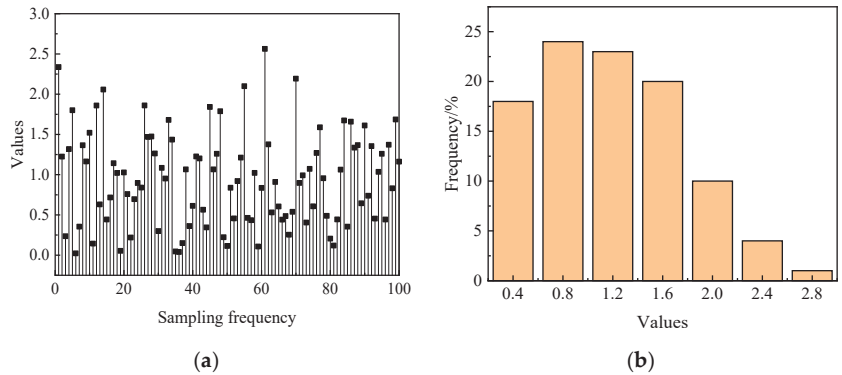


Figure 22. Distribution of the independent variable x_2 . (a) Distribution conditions. (b) Distribution probability.

Equation (4) shows that with M samples, the gradient covariance matrix C of $f(x_1, x_2)$ can be expressed as follows:

$$C = \hat{C}^I = \frac{1}{M} \sum_{k=1}^M (\nabla_x f_k^I) \cdot (\nabla_x f_k^I)^T \tag{34}$$

$$= \frac{1}{M} \sum_{k=1}^M \begin{bmatrix} [e^{2x_1}, e^{2x_1} + 6e^{x_1} + 9] & [(x_1 + 1)e^{x_1}, (x_1 + 1)e^{x_1} + 3(x_1 + 1)e^{x_1} + 3] \\ [(x_1 + 1)e^{x_1}, (x_1 + 1)e^{x_1} + 3(x_1 + 1)e^{x_1} + 3] & [x_1^2 + 2x_1 + 1] \end{bmatrix}$$

The EDF obtained by probability box fitting is shown in Figure 23.

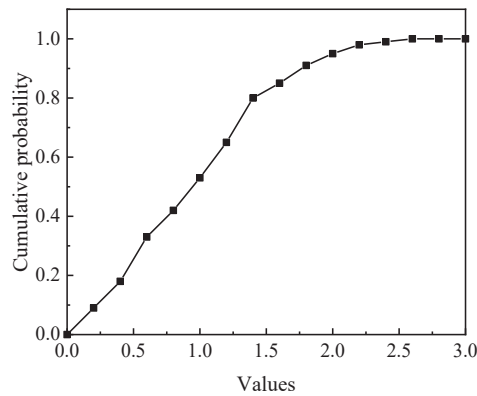


Figure 23. Fit EDF depicting the EDF fitted using the probability box.

With $M = 100$ samples, eigenvalues λ_1 , λ_2 , and their proportions, obtained using computation Equations 33 and 34 probability box–EDF estimation, are shown in Table 4. The proposed method achieves high accuracy, with a deviation of only 1.2931%, making it effective for eigenvalue calculation and dimensionality reduction in rotor engine combustion chambers.

Table 4. Eigenvalue calculation results.

Eigenvalue Calculation Method	λ_1	λ_1 Proportion	λ_2	λ_2 Proportion
Formula calculation	12.3058	62.1078%	7.5078	37.8922%
Probability box–EDF estimation	12.8423	60.8148%	8.2748	39.1852%

4. Dimensionality Reduction Analysis of Combustion Chamber Structure Parameters

This paper uses peak cylinder pressure and temperature as performance indicators. With Table 5 showing the boundary conditions, the RNG K- ϵ model calculates cylinder airflow [64–66]. The PRF skeleton mechanism with 41 components and 124 chemical reactions is combined with the SAGE chemical kinetic model for combustion calculations [67,68]. Figure 24 depicts solid model parameters [58].

Table 5. Rotor engine boundary conditions.

Boundary	Type	Temperature (K)	Pressure (MPa)
Inlet	Inflow	300	0.101325
Intake port	Fixed wall	300	/
Outlet	Outflow	570	0.101325
Exhaust port	Fixed wall	550	/
Rotor	Moving wall	400	/
Rotor flank 1	Fixed wall	624	0.117210
Rotor flank 2	Fixed wall	600	/

Grid number affects calculation results. To eliminate this influence, ensure accuracy, and select the optimal grid number to reduce computation time, grid independence analysis of the simulation model is necessary. This study compares in-cylinder pressure changes under four different grid sizes with constant boundary conditions set as shown in Table 5. Figure 25 shows pressure change curves for different grid sizes. Adaptive mesh refinement (AMR) [69–71] is applied to a 2 mm mesh, producing results consistent with a 1 mm mesh size and achieving stable calculations. Considering efficiency, accuracy, and calculation time, a 2 mm mesh is selected, resulting in approximately 200,000 hexahedral element grids,

as shown in Figure 26. The 100 parameter combinations obtained in Section 2 are calculated with performance indices (Figure 27).

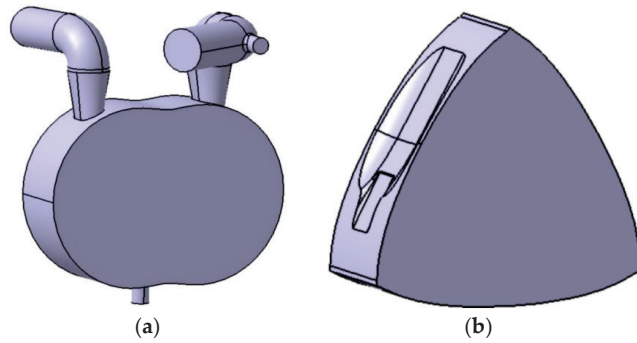


Figure 24. Three-dimensional model. (a) Cylinder fluid domain. (b) Rotor entity.

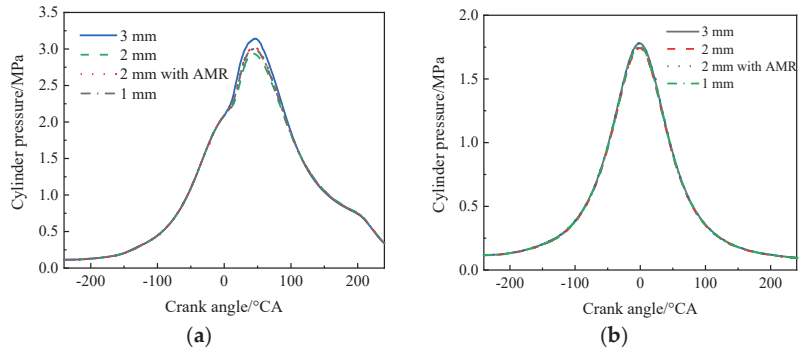


Figure 25. The cylinder pressure change curve under different grid numbers. (a) Ignition state. (b) Non-ignition state.

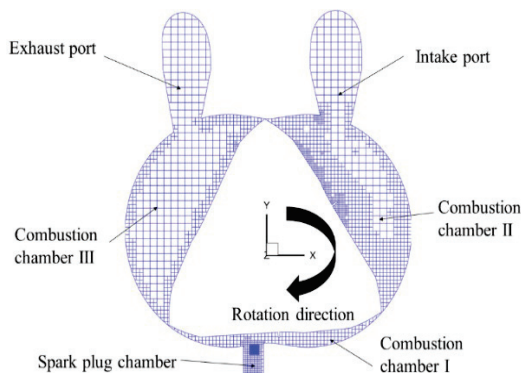
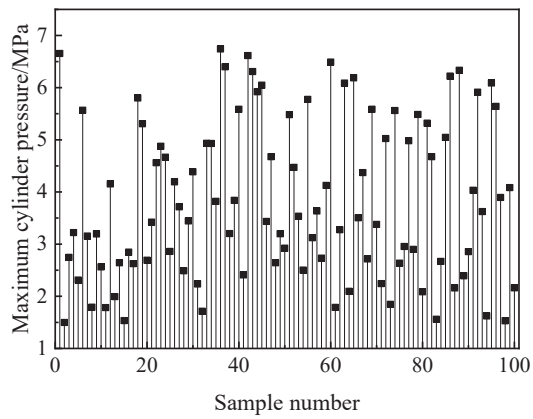
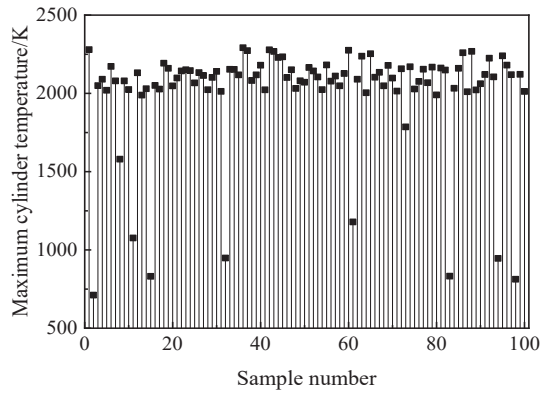


Figure 26. Rotor engine computational grid.

The eigenvalue estimation method from the earlier section is applied to the generated 100-size combinations. Figures 28–37 present the EDFs obtained by fitting all 10 cognitive uncertainty variables using a probability box.



(a)



(b)

Figure 27. Performance index. (a) Maximum cylinder pressure. (b) Maximum cylinder temperature.

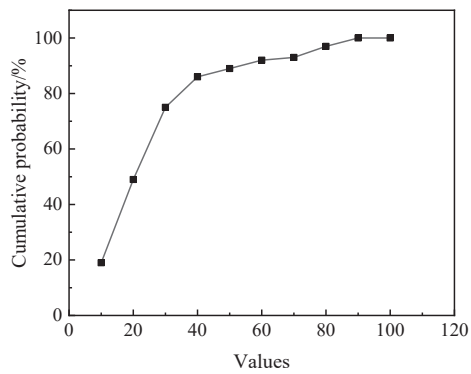


Figure 28. EDF fitting of the arc radius r in the combustion chamber.

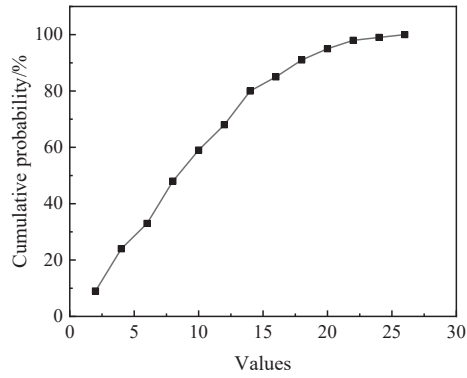


Figure 29. EDF fitting for the combustion chamber leading bottom length l .

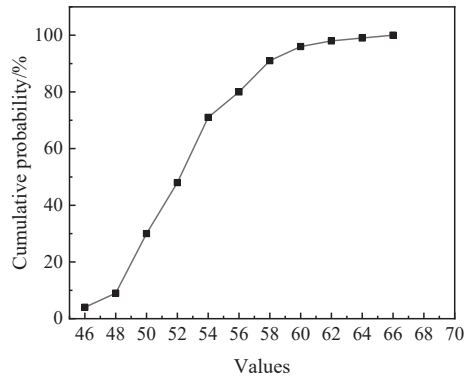


Figure 30. EDF fitting for the distance (h) between the combustion chamber bottom and the rotor's center of gravity in the x - z plane.

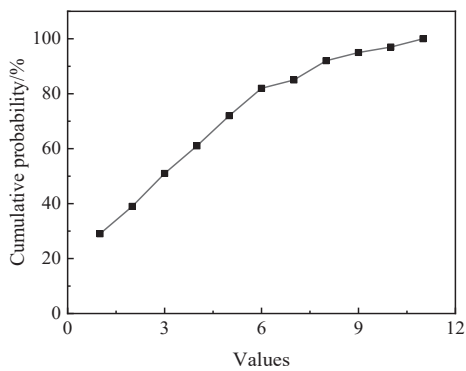


Figure 31. EDF fitting for the angle (θ_3) between the bottom edge of the combustion chamber's leading section and the y - z plane.

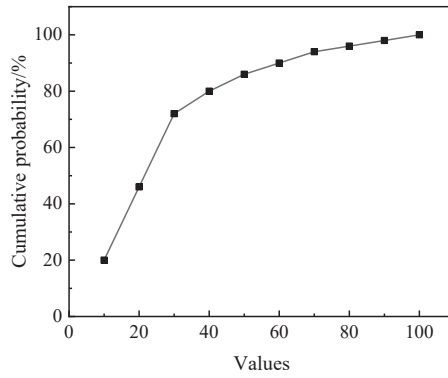


Figure 32. EDF fitting of the arc radius (R) in the combustion chamber's trailing section.

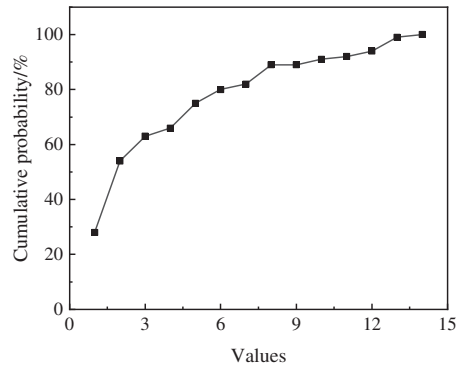


Figure 33. EDF fitting of the trailing bottom edge length (L) of the combustion chamber.

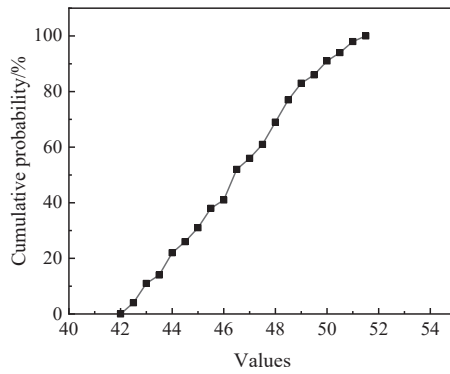


Figure 34. EDF fitting of the distance (H) from the combustion chamber's trailing bottom to the rotor's center of gravity in the x - z plane.

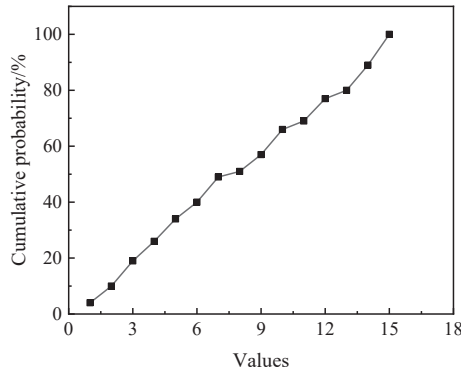


Figure 35. EDF fitting of the angle (θ_1) between the bottom edge of the combustion chamber's middle section and the y - z plane.

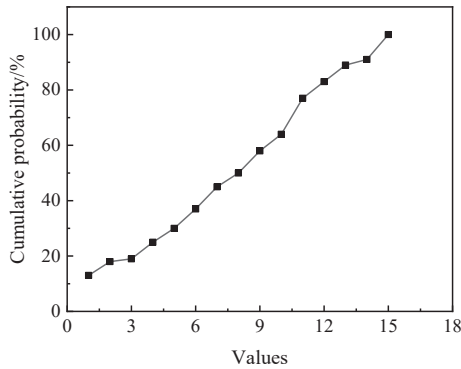


Figure 36. EDF fitting of the angle (θ_2) between the rear bottom edge of the combustion chamber and the y - z plane.

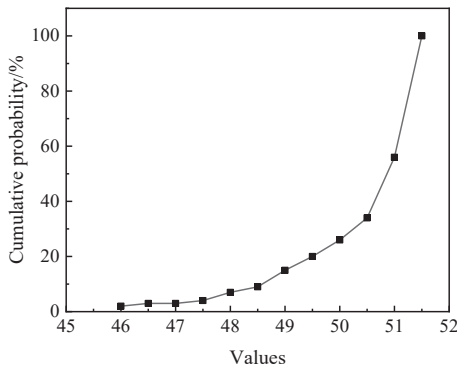


Figure 37. EDF fitting of the distance (b) between the center of the spoiler plate top and the bottom of the combustion chamber.

Taking the maximum cylinder pressure, the maximum cylinder temperature, and the indicated average effective pressure as the performance indexes, the EDF and eigenvalue estimation methods of each parameter obtained above are used to calculate the eigenvalue matrix of the function composed of 14 structural parameters, including the

radius r of the front arc of the combustion chamber. The calculation results are shown in Table 6. The larger the eigenvalues in the table, the greater the impact on the corresponding performance indicators.

Table 6. The calculated eigenvalues.

Structural Parameter	Performance Index		
	Maximum Cylinder Pressure	Maximum Cylinder Temperature	Indicated Average Effective Pressure
r	2.1339	3.3178	2.5761
l	3.1528	4.7906	1.6218
h	4.6812	6.3345	3.7114
$\theta 3$	2.3218	8.2688	8.0427
R	2.8055	6.7792	4.1701
L	2.2952	4.3389	4.0759
H	14.2391	16.9644	14.4644
$\theta 1$	18.1673	18.1360	13.5508
$\theta 2$	16.6057	17.0970	14.9315
a	8.5642	9.3406	7.9712
b	16.8614	17.7883	8.6039
c	1.3363	1.2064	0.8551
$q1$	0.5012	0.7131	0.5182
$q2$	0.3877	0.9580	0.0224

Figure 38 illustrates the radar plot of each structural parameter's impact on the performance index using the eigenvalue ratio. The coordinate axis indicates the influence percentage, and the larger structural parameters have a greater effect on the performance index.

From the figure, it can be seen that the effect of each structural parameter on the same performance index and that of the same structural parameter on different performance indexes is slightly different. However, whether the performance index is based on the maximum cylinder pressure, the maximum cylinder temperature, or the indicated average effective pressure, the three structural parameters that have the greatest influence on it are the angle $\theta 2$ between the rear bottom of the combustion chamber and the Y-Z plane, the distance b between the center of the top of the turbulent blade and the bottom of the combustion chamber, and the angle $\theta 1$ between the bottom of the middle of the combustion chamber and the Y-Z plane. The three structural parameters that have the least influence on the performance index are the spoiler width c , the radius of the excessive arc between the front of the combustion chamber and the middle of the combustion chamber $q1$, and the radius of the excessive arc between the middle of the combustion chamber and the rear of the combustion chamber $q2$. The vertical distance a between the spoiler and the rotor X-Z section and the distance b between the top center of the spoiler and the bottom edge of the combustion chamber have a greater influence on the maximum cylinder pressure and the maximum cylinder temperature but have a smaller influence on the indicated average effective pressure.

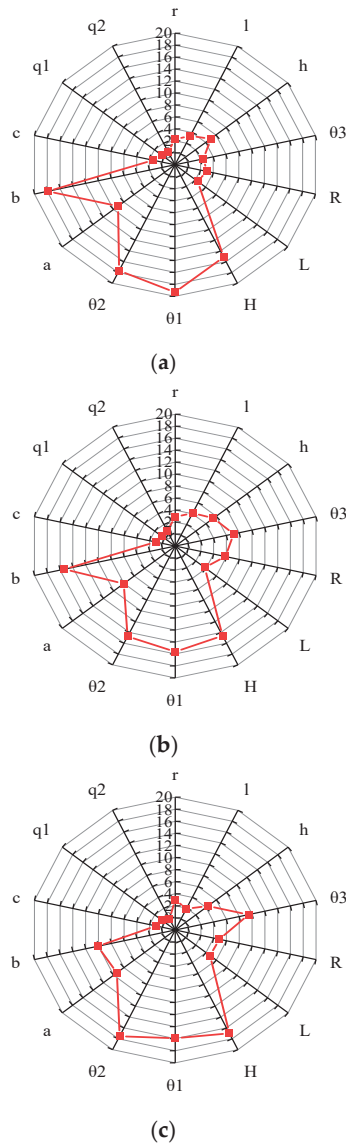


Figure 38. Impact of structural parameters on various performance metrics. (a) Impact of numerous structural parameters on maximum cylinder pressure. (b) Effect of different structural parameters on peak cylinder temperature. (c) Effect of each structural parameter on the indicated mean effective pressure.

Through the above chart, the main structural parameter combination can be obtained, which is composed of the angle θ_1 between the bottom edge of the middle part of the combustion chamber and the Y-Z plane, the distance b between the center of the spoiler top and the bottom edge of the combustion chamber, the angle θ_2 between the bottom edge of the rear part of the combustion chamber and the Y-Z plane, the distance H between the bottom of the middle and rear part of the combustion chamber at the X-Z section of the rotor and the center of gravity of the rotor, the vertical distance a between the spoiler and the X-Z section of the rotor, the angle θ_3 between the bottom edge of the front part

of the combustion chamber and the Y-Z plane, the radius R of the middle and rear part of the combustion chamber, and the distance h between the bottom of the front part of the combustion chamber at the X-Z section of the rotor. The influence of this parameter combination on the maximum cylinder pressure, the maximum cylinder temperature, and the indicated average effective pressure is 89.5728%, 86.7924%, and 89.0677%, respectively. The influence of structural parameters such as the bottom edge length l of the front part of the combustion chamber, the bottom edge length L of the middle and rear part of the combustion chamber, the arc radius r of the front part of the combustion chamber, the width c of the spoiler, the excessive arc radius q_2 between the middle part of the combustion chamber and the rear part of the combustion chamber, and the excessive arc radius q_1 between the front part of the combustion chamber and the middle part of the combustion chamber is small and can be ignored in the study of combustion chamber structure optimization. This is consistent with the current research trend regarding the structural parameters of the combustion chamber of the rotary engine. On this basis, three structural parameters can be obtained, which are the angle θ_2 between the bottom edge of the combustion chamber and the Y-Z plane, the distance H between the bottom of the middle and rear section of the combustion chamber at the X-Z section of the rotor and the center of gravity of the rotor, and the angle θ_1 between the bottom edge of the middle of the combustion chamber and the Y-Z plane. These three structural parameters should be used as important structural parameters for the study of the combustion chamber structure of the rotary engine. These three parameters have a great influence on each performance index, and the influence on the maximum cylinder pressure, the maximum cylinder temperature, and the indicated average effective pressure are 52.1109%, 45.9847%, and 50.4569%, respectively.

Since the design of the combustion chamber structure demands a variety of requirements, including compression ratio, process, and structure, after optimizing the above structural parameters that have the greatest impact on performance, it is necessary to change some structural parameters at the same time to make them meet many requirements, including compression ratio. It can be seen from the above charts that the three structural parameters, such as the length l of the front bottom of the combustion chamber, the length L of the rear bottom of the combustion chamber, and the radius r of the front arc of the combustion chamber, have a small proportion of influence on each performance index. The influence of these three parameters on the maximum cylinder pressure, the maximum cylinder temperature, and the indicated mean effective pressure is 8.7753%, 13.7104%, and 11.5934%, respectively, and the range of parameters is large. Because the change in parameters has a great influence on the compression ratio, process, and strength requirements, the above three structural parameters can be used as adjustable structural parameters in the study of combustion chamber structure optimization. When the main structural parameters are determined, the adjustable structural parameters are adjusted to make the combustion chamber meet many requirements, including compression ratio, process, and structure.

5. Conclusions

In order to reduce the dimension in the optimization process of the combustion chamber structure of the rotary engine and clarify the optimization priority and direction, this paper proposes a dimension reduction method for the structural parameters of the combustion chamber of the rotary engine based on the AS, referring to the probability box and EDF. The main research contents of this paper are as follows:

1. By analyzing the generation process of the active subspace and the composition of rotor structure parameters, based on active subspace, combined with the probability box and EDF, a dimension reduction method for rotary engine combustion chamber structure parameters is proposed, and the accuracy of the method is verified. The results show that the deviation between the calculated eigenvalues and the actual eigenvalues is only 1.2931%, and the estimation accuracy is high, which can be used

- for eigenvalue calculation and parameter dimension reduction of high-dimensional mixed uncertainty problems.
2. Using the dimension reduction method proposed above, the dimension combination composed of 14 structural parameters is reduced to an important structural parameter composed of 8 structural parameters, including θ_1 , b , θ_2 , H , a , θ_3 , R , and h . The effects of the above eight structural parameters on the maximum cylinder pressure, the maximum cylinder temperature, and the indicated mean effective pressure are 89.5728%, 86.7924%, and 89.0677%, respectively.
 3. On this basis, three main structural parameters, with the influence of θ_2 accounting for more than 45%, are obtained. And three adjustable structural parameters, including l , are obtained. The influence of the latter on the total proportion of each performance index is small, and the influence on the maximum cylinder pressure, the maximum cylinder temperature, and the indicated average effective pressure is 8.7753%, 13.7104%, and 11.5934%, respectively. Moreover, the change in parameters has a great influence on the compression ratio, process, and strength requirements. Therefore, the above three structural parameters can be used as adjustable structural parameters in the optimization of combustion chamber structure. When the main structural parameters are determined, the adjustable structural parameters are adjusted to make the combustion chamber meet many requirements, including compression ratio, process, and structure.

In the future, based on the research in this paper, the research group will deeply study the influencing law of combustion chamber structure parameters on engine performance and quantitatively analyze it, hoping to obtain the influence proportions of different combustion chamber structure parameters and different levels of the same structure parameters on engine performance. Furthermore, the dimension reduction method based on AS proposed in this paper will be applied to the research of other structure or control parameters of rotary engines, which provides a reference and basis for further improving the performance of rotary engines and broadening the application field of rotary engines.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pr12102238/s1>, Figure S1: Sectional View of Combustion Chamber Pit; Figure S2: Schematic diagram of the segmented structure of the front, middle, and rear parts of the combustion chamber; Figure S3: Schematic diagram of segmented structure of spoiler; Figure S4: Parametric modeling flow chart; Figure S5: Macro Command; Figure S6: Ompression ratio of solid model; Figure S7: Distribution of parameters for the radius R of the rear arc in the combustion chamber and the length L of the bottom edge in the rear of the combustion chamber; Figure S8: Solid model of rotor.

Author Contributions: Validation, Y.S.; formal analysis, L.L.; investigation, Y.T. and W.L.; resources, R.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Postdoctoral Fellowship Program of CPSF under Grant Number (GZC20241576) and the Applied Basic Research Programs of Shanxi Province in China (202403021212341, 202203021222038, 202203021222045).

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, S. The Status of the Power Plants for UAVs in China. *Aerosp. Power* **2019**, *2*, 9–12.
2. Dong, Y.; Huang, M.; Li, R. Overview of the Development of General Aviation Engines. *J. Xi'an Aeronaut. Univ.* **2017**, *35*, 8–13.
3. Yin, Z.; Li, S.; Li, G. Current State and Development of the Unmanned Aerial Vehicle Power plant. *Aeroengine* **2007**, *33*, 10–15.
4. Lu, D.; Zheng, J.; Hu, C.; Bian, S. Research on the development status of general aviation piston engine. *Intern. Combust. Engine Parts* **2019**, *8*, 64–66.
5. Feng, G.; Zhou, M. Assessment of heavy fuel aircraft piston engine type. *J. Tsinghua Univ.* **2016**, *56*, 1114–1121.

6. Jiao, H.; Liu, J.; Zou, R.; Wang, N. Combined influence of ignition chamber volume and spark plug channel diameter on the performance of small-scale natural gas Wankel rotary engine. *Eng. Appl. Comput. Fluid Mech.* **2021**, *15*, 1775–1791. [CrossRef]
7. Kong, X.; Liu, H. Research Progress of Key Technologies of Aviation Piston Engine for UAV. *Small Intern. Combust. Engine Veh. Tech.* **2021**, *50*, 79–87.
8. Johar, T.; Hsieh, C.F. Design Challenges in Hydrogen-Fueled Rotary Engine—A Review. *Energies* **2023**, *16*, 607. [CrossRef]
9. Wu, S. Research on Technical Characteristics and Application of Unmanned Aerial Vehicle Power Unit. *Shanghai Energy Sav.* **2022**, 1536–1540.
10. Wang, X.; Liu, J. GE Aerospace’s Roadmap for Next-Generation Aerospace Power Technology. *Aerosp. Power* **2023**, 24–27.
11. Li, H.; Sun, F. The Application Development and Key Technologies of Wankel engine. *Small Intern. Combust. Engine Veh.* **2023**, *52*, 68–74.
12. Kotsiopoulos, P.; Yfantis, E.; Lois, E.; Hountalas, D. Diesel and JP-8 fuel performance on a Petter AVI diesel engine. In Proceedings of the 39th Aerospace Sciences Meeting and Exhibit (AIAA), Reno, NV, USA, 8–11 January 2001. [CrossRef]
13. Dutczak, J. Heavy fuel engines. *Combust. Engines* **2015**, *163*, 34–46. [CrossRef]
14. Wu, H.; Wang, L.L.; Wu, Y.; Sun, B.; Zhao, Z.; Liu, F. Spray performance of air-assisted kerosene injection in a constant volume chamber under various in-cylinder GDI engine conditions. *Appl. Therm. Eng.* **2019**, *150*, 762–769. [CrossRef]
15. Li, Z. Simulation Analysis on Cooperative Control of Ignition and Knock of Light-Weight, Spark Ignition, Heavy Oil Aeroengine. Master’s Thesis, Beijing Jiao Tong University, Beijing, China, 2016.
16. Wang, Z. Experimental Study on Performance of 15kW Two-Stroke Ignited PFI Kerosene Piston Engine. Master’s Thesis, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 2018.
17. Attard, W.P.; Blaxill, H.; Anderson, E.K.; Litke, P. Knock limit extension with a gasoline fueled pre-chamber jet igniter in a modern vehicle powertrain. *SAE Int. J. Engines* **2012**, *5*, 1201–1215. [CrossRef]
18. Wang, C.Y.; Zhang, F.J.; Wang, E.H.; Yu, C.; Gao, H.; Liu, B.; Zhao, Z.; Zhao, C. Experimental study on knock suppression of spark-ignition engine fueled with kerosene via water injection. *Appl. Energy* **2019**, *242*, 248–259. [CrossRef]
19. Chen, L.F.; Raza, M.; Xiao, J.H. Combustion analysis of an aviation compression ignition engine burning pentanol kerosene blends under different injection timings. *Energy Fuel* **2017**, *31*, 9429–9437. [CrossRef]
20. Ning, L.; Duan, Q.M.; Wei, Y.H.; Zhang, X.; Yu, K.; Yang, B.; Zeng, K. Effects of injection timing and compression ratio on the combustion performance and emissions of a two-stroke DISI engine fuelled with aviation kerosene. *Appl. Therm. Eng.* **2019**, *161*, 114–124.
21. Lu, Y.; Pan, J.F.; Fan, B.W.; Otchere, P.; Chen, W.; Cheng, B. Research on the application of aviation kerosene in a direct injection wankel engine -Part 1: Fundamental spray characteristics and optimized injection strategies. *Energy Convers. Manag.* **2019**, *195*, 519–532. [CrossRef]
22. Yang, J.; Ji, C.; Wang, S.; Wang, D.; Ma, Z.; Zhang, B. Numerical investigation on the mixture formation and combustion processes of a gasoline wankel engine with direct injected hydrogen enrichment. *Appl. Energy* **2018**, *224*, 34–41. [CrossRef]
23. Cao, B.W.; Liu, J.X. Design and optimization of orifice plates in the air-assisted kerosene injection system applied in the wankel rotary engine. *Proc. Inst. Mech. Eng. Part C-J. Mech. Eng. Sci.* **2023**, *237*, 2945–2953. [CrossRef]
24. Zhao, Y.; Deng, X.; Feng, Z.; Zhu, R.; Su, X. Effects of Combustion Chamber Pit Shape on Combustion Process for Gasoline Wankel engine. *Veh. Engine* **2022**, 40–46.
25. Kuo, C.H.; Ma, H.L.; Chen, C.C. Chamber contour design and compression flow calculations of wankel engine. *J. CCIT* **2010**, *39*, 35–50.
26. Tartakovskiy, L.; Baibikov, V.; Gutman, M.; Veinblat, M.; Reif, J. Simulation of Wankel engine performance using commercial software for piston engines. In Proceedings of the 2012 Small Engine Technology Conference & Exhibition, Madison, WI, USA, 16–18 October 2012. SAE Technical Paper 2012–32–0098.
27. Jeng, D.Z.; Hsieh, M.J.; Lee, C.C.; Han, Y. The numerical investigation on the performance of wankel engine with leakage, different fuels and recess sizes. In Proceedings of the JSAE/SAE 2013 Small Engine Technology Conference, Taipei, Taiwan, 8–10 October 2013. SAE Technical Paper 2013–32–9160.
28. Wei, F.B.; Feng, P.J.; Xian, L.Y.; Rui, C.; Di, W.U. Effect of pocket location on combustion process in natural gas-fueled wankel engine. *Appl. Mech. Mater.* **2013**, *316–317*, 73–79. [CrossRef]
29. Fan, B.W.; Pan, J.F.; Pan, Z.H.; Tang, A.K.; Zhu, Y.J.; Xue, H. Effects of pocket shape and ignition slot locations on the combustion processes of a wankel engine fueled with natural gas. *Appl. Therm. Eng.* **2015**, *89*, 11–27. [CrossRef]
30. Ji, C.W.; Su, T.; Zhang, S.F.; Yu, M.; Cong, X. Effect of hydrogen addition on combustion and emissions performance of a gasoline wankel engine at part load and stoichiometric conditions. *Energy Convers. Manag.* **2016**, *121*, 272–280. [CrossRef]
31. Amrouche, F.; Varnhagen, S.; Erickson, P.A.; Park, J. An experimental study of a hydrogen-enriched ethanol fueled Wankel wankel engine at ultra lean and full load conditions. *Energy Convers. Manag.* **2016**, *123*, 174–184. [CrossRef]
32. Su, T.; Ji, C.; Wang, S.; Shi, L.; Cong, X. Effect of ignition timing on performance if a hydrogen-enriched n-butanol wankel engine at lean condition. *Energy Convers. Manag.* **2018**, *161*, 27–34. [CrossRef]
33. Chen, W.; Pan, J.; Fan, B.; Liu, Y.; Peter, O. Effect of injection strategy on fuel-air mixing and combustion process in a direct injection diesel wankel engine (DI-DRE). *Energy Convers. Manag.* **2017**, *154*, 68–80. [CrossRef]
34. Shi, C.; Zhang, P.; Ji, C. Understanding the role of turbulence-induced blade configuration in improving combustion process for hydrogen-enriched wankel engine. *Fuel* **2022**, *319*, 123807. [CrossRef]

35. Bienner, A.; Gloerfelt, X.; Cinnella, P. Leading-Edge Effects on Freestream Turbulence Induced Transition of an Organic Vapor. *Flow Turbul. Combust.* **2024**, *224*, 345–373. [CrossRef]
36. Li, M.; Hamel, J.; Azarm, S. Optimal uncertainty reduction for multi-disciplinary multi-output systems using sensitivity analysis. *Struct. Multidiscip. Optim.* **2010**, *40*, 77–96. [CrossRef]
37. Davey, K.; Sadeghi, H.; Darvizeh, R. The theory of scaling. *Contin. Mech. Thermodyn.* **2023**, *35*, 471–496. [CrossRef]
38. Jolliffe, I. *Principal Component Analysis*; Wiley Online Library: Hoboken, NJ, USA, 2002.
39. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
40. Wang, J. *Geometric Structure of High-Dimensional Data and Dimensionality Reduction*; Springer: Berlin/Heidelberg, Germany, 2012.
41. Schölkopf, B.; Smola, A.; Müller, K.-R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **1998**, *10*, 1299–1319. [CrossRef]
42. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [CrossRef]
43. Donoho, D.L.; Grimes, C. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5591–5596. [CrossRef]
44. Belkin, M.; Niyogi, P. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Comput.* **2003**, *15*, 1373–1396. [CrossRef]
45. Constantine, P.G. *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*; SIAM: Philadelphia, PA, USA, 2015.
46. Scott, D.W. The Curse of Dimensionality and Dimension Reduction. In *Multivariate Density Estimation*; Wiley: Hoboken, NJ, USA, 2008.
47. Wang, B.X.; Huang, X.Z.; Chang, M.X. Regional reliability sensitivity analysis based on dimension reduction technique. *Probabilistic Eng. Mech.* **2023**, *74*, 1394–1419. [CrossRef]
48. Russi, T.M. Uncertainty Quantification with Experimental Data and Complex System Models. Ph.D. Thesis, University of California, Berkeley, CA, USA, 2010.
49. Stewart, G.W. Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM Rev.* **1973**, *15*, 727–764. [CrossRef]
50. Davey, K.; Abd Malek, M.I.; Ali, Z.; Sadeghi, H.; Darvizeh, R. The theory of scaled electromechanics. *Int. J. Eng. Sci.* **2024**, *203*, 1041–1068. [CrossRef]
51. Fujimori, K.; Goto, Y.; Liu, Y.; Taniguchi, M. Sparse principal component analysis for high-dimensional stationary time series. *Scand. J. Stat.* **2023**, *50*, 1953–1983. [CrossRef]
52. Wang, M.; Lu, Y.; Pan, W. An improved pattern-based prediction model for a class of industrial processes. *Trans. Inst. Meas. Control.* **2022**, *44*, 1410–1423. [CrossRef]
53. Chib, S.; Greenberg, E. Understanding the metropolis-hastings algorithm. *Am. Stat.* **1995**, *49*, 327–335. [CrossRef]
54. Ray, T. Golinski's speed reducer problem revisited. *AIAA J.* **2003**, *41*, 556–558. [CrossRef]
55. Allen, M.; Maute, K. Reliability-based design optimization of aeroelastic structures. *Struct. And. Multidiscip. Optim.* **2004**, *27*, 228–242. [CrossRef]
56. Lillacci, G.; Khammash, M. A distribution-matching method for parameter estimation and model selection in computational biology. *Int. J. Robust Nonlinear Control* **2012**, *22*, 1065–1081. [CrossRef]
57. Seshadri, P.; Constantine, P.; Iaccarino, G.; Parks, G. Aggressive design: A density-matching approach for optimization under uncertainty. *arXiv* **2014**, arXiv:1409.7089.
58. Ferson, S.; Kreinovich, V.; Ginzburg, L.; Myers, D.S.; Sentz, K. *Constructing Probability Boxes and Dempster-Shafer Structures*; SAND2003-4015; Sandia National Laboratories: Albuquerque, NM, USA, 2003.
59. Yang, X. Research on Highly Efficient and Precise Methods for Reliability Analysis with Epistemic Uncertainties. Master's Thesis, Northwestern Polytechnical University, Xi'an, China, 2016.
60. Huang, X.; Wang, Q.; Ding, J. Faculty of Information Engineering and Automation: Index Uncertainty Modeling in Grid Planning Based on Probability Box Theory. *Inf. Control* **2016**, *45*, 272–277.
61. Ferson, S.; Tucker, W.T. *Sensitivity in Risk Analyses with Uncertain Numbers*; Sandia National Laboratories: Albuquerque, NM, USA, 2006; pp. 156–159.
62. Crespo, L.G.; Kenny, S.P.; Giesy, D.P. Reliability analysis of polynomial systems subject to p-box uncertainties. *Mech. Syst. Signal Process.* **2013**, *37*, 121–136. [CrossRef]
63. Beer, M.; Ferson, S.; Kreinovich, V. Imprecise probabilities in engineering analyses. *Mech. Syst. Signal Process.* **2013**, *37*, 4–29. [CrossRef]
64. Jiao, H.; Ye, X.; Zou, R.; Wang, N.; Liu, J. Comparative study on ignition and combustion between conventional spark-ignition method and near-wall surface ignition method for small-scale Wankel rotary engine. *Energy* **2022**, *255*, 1049–1072. [CrossRef]
65. Zou, R.; Liu, J.; Jiao, H.; Zhao, J.; Wang, N. Combined effect of intake angle and chamber structure on flow field and combustion process in a small-scaled rotary engine. *Appl. Therm. Eng.* **2022**, *203*, 1497–1521. [CrossRef]
66. Hege, J. *The Wankel Wankel Engine: A History*; McFarland and Company Inc.: Jefferson, NC, USA, 2006.
67. Wladyslaw, M. Modelling and Simulation of Working Processes in Wankel Engine with Direct Hydrogen Injection System. *Combust. Engines* **2015**, *161*, 42–52.
68. Peden, M. Study of Direct Injection Limitations on a Wankel Engine. Master's Thesis, University of Bath, Bath, UK, 2017.
69. Wendeker, M.; Grabowski, L.; Pietrykowski, K.; Margryta, P. *Phenomenological Model of a Wankel Engine*; Lublin University of Technology: Lublin, Poland, 2012.

70. Tomlinson, A. Modelling of Wankel Engine Performance in Commercial Piston Engine Software. Master's Thesis, University of Bath, Bath, UK, 2016.
71. Georgios, Z. Mathematical and Numerical Modelling of Flow and Combustion Processes in a Spark Ignition Engine. Master's Thesis, Department of Applied Mathematics, University of Wisconsin, Madison, WI, USA, 2005.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Processes Editorial Office
E-mail: processes@mdpi.com
www.mdpi.com/journal/processes



Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-2440-3