sensors

Special Issue Reprint

Sensing Technologies in Industrial Defect Detection

Edited by
Peng Liu

mdpi.com/journal/sensors

MDPI

Sensing Technologies in Industrial Defect Detection

Sensing Technologies in Industrial Defect Detection

Guest Editor

Peng Liu

MDPI

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

Guest Editor
Peng Liu
School of Mechanical and
Aerospace Engineering
Jilin University
Changchun
China

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Sensors* (ISSN 1424-8220), freely accessible at: https://www.mdpi.com/journal/sensors/special_issues/3TT019L154.

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. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-9209-9 (Hbk)
ISBN 978-3-7258-9210-5 (PDF)
https://doi.org/10.3390/books978-3-7258-9210-5

© 2026 by the authors. Articles in this reprint are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The reprint as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

About the Editor . . . . . **vii**

Preface . . . . . **ix**

Peng Liu
Sensing Technologies in Industrial Defect Detection
Reprinted from: *Sensors* **2026**, *26*, 5706, https://doi.org/10.3390/s26185706 . . . . . **1**

Anying Chai, Zhaobo Fang, Mengjia Lian, Ping Huang, Chenyang Guo, Wanda Yin, et al.
Hi-MDTCN: Hierarchical Multi-Scale Dilated Temporal Convolutional Network for Tool Condition Monitoring
Reprinted from: *Sensors* **2025**, *25*, 7603, https://doi.org/10.3390/s25247603 . . . . . **5**

Liangxia Zhang, Yali Yang, Hao Chen and Shusheng Lv
Correlation Between Meso-Defect and Fatigue Life Through Representing Feature Analysis for 6061-T6 Aluminum Alloys
Reprinted from: *Sensors* **2026**, *26*, 631, https://doi.org/10.3390/s26020631 . . . . . **31**

Bangcheng Zhang, Boyu Shen, Zhi Gao, Yubo Shao, Zaixiang Pang and Xiaojing Yin
Composite Fault Feature Index-Guided Variational Mode Decomposition with Dynamic Weighted Central Clustering for Bearing Fault Detection
Reprinted from: *Sensors* **2026**, *26*, 1394, https://doi.org/10.3390/s26041394 . . . . . **44**

Saiyan Saiyod, Woottichai Nonsakhoo, Zhengping Li and Piyanat Sirisawat
Defect-Intent Ambiguity Addressing for Training-Free Deterministic PCB Defect Localization via Template Selection and Dissimilarity Mapping
Reprinted from: *Sensors* **2026**, *26*, 1541, https://doi.org/10.3390/s26051541 . . . . . **63**

Peng Liu, Xinze Li, Chenlong Zhang, Yanru Kang, Jun Qian and Weizheng Chen
Production System Monitoring Based on Petri Nets Enhanced with Multi-Source Information
Reprinted from: *Sensors* **2026**, *26*, 1785, https://doi.org/10.3390/s26061785 . . . . . **93**

Gleb Solovev, Evgenii Klokov, Dmitrii Krasnov and Mikhail Sokolov
Advancing Defect Detection in Laser Welding: A Machine Learning Approach Based on Spatter Feature Analysis
Reprinted from: *Sensors* **2026**, *26*, 1825, https://doi.org/10.3390/s26061825 . . . . . **108**

Zhendong Qu, Jialong He, Yan Liu, Song Mao and Xiaowu Han
Interval Prediction of Remaining Useful Life Based on Uncertainty Quantification with Bayesian Convolutional Neural Networks Featuring Dual-Output Units
Reprinted from: *Sensors* **2026**, *26*, 2592, https://doi.org/10.3390/s26092592 . . . . . **128**

Jiazhi Dai, Mario Rotea and Nasser Kehtarnavaz
A Two-Stage Classification Method for Improved Fault Detection in Wind Turbines Based on SCADA Data
Reprinted from: *Sensors* **2026**, *26*, 3865, https://doi.org/10.3390/s26123865 . . . . . **147**

Peng Liu, Guotai Huang, Jialu Xi and Jiaqi Wu
Multi-Component Joint Maintenance Decision for Electro-Hydraulic Servo Fatigue Testing Machine Based on Multi-Head Deep Reinforcement Learning
Reprinted from: *Sensors* **2026**, *26*, 4087, https://doi.org/10.3390/s26134087 . . . . . **166**

Xiaopeng Chang, Bangcheng Zhang, Zhi Gao, Siyu Chen and Jingru Liu
A Span-Prior-Guided Explainable Multimodal Neural Network Method for Final-State Quality Inspection of Hairpin Windings
Reprinted from: *Sensors* **2026**, *26*, 4857, https://doi.org/10.3390/s26154857 . . . . . **187**

About the Editor

Peng Liu

Peng Liu is a professor and doctoral supervisor at Jilin University. He currently serves as Deputy Director of the Engineering Training Center and Vice Dean of the School of Mechanical and Aerospace Engineering, Jilin University, and as Deputy Director of the Jilin Provincial Innovation Center for Material Service Performance Testing Technology and Intelligent Equipment. His research interests focus on intelligent instrumentation, equipment operation and maintenance, and intelligent manufacturing systems. He has led more than ten research projects funded by the National Natural Science Foundation of China, the Ministry of Industry and Information Technology, major science and technology programs of Jilin Province, and key research and development programs of Jilin Province, including projects undertaken in collaboration with industry. He has also participated as a key researcher in more than ten major national research programs, including National Science and Technology Major Projects and National Key Research and Development Programs. He has published more than forty papers as first or corresponding author, with over 2000 citations, including one ESI Highly Cited Paper. As the first inventor, he has been granted more than 30 Chinese invention patents and registered more than 40 software copyrights. His honors include the First Prize of the Sixth National University Teachers Self-Developed Experimental Instrument Competition, the Second Prize of the China Machinery Industry Technology Invention Award, the First Prize of the China Machinery Industry Education Integration Teaching Innovation Competition, and the First Prize of the Jilin Provincial Technology Invention Award.

Preface

Industrial defect detection is fundamental to product quality, equipment reliability, and safe operation in modern manufacturing. With the development of intelligent manufacturing, industrial inspection is progressing from conventional offline examination toward automated sensing and data-driven assessment. Advances in measurement technologies, signal and image processing, and artificial intelligence are providing new means to observe defects, identify abnormal conditions, and evaluate degradation processes. This Reprint brings together recent studies that reflect this progression. Its scope begins with the acquisition of physical information from industrial components and processes, continues with the extraction and interpretation of defect- and fault-related features, and extends to condition assessment, degradation prediction, and maintenance support. The contributions therefore represent different stages of a common technical chain rather than isolated inspection tasks. The motivation for preparing this Reprint is the growing need to connect sensing technologies with reliable engineering assessment. Effective industrial inspection depends not only on obtaining high-quality measurements, but also on transforming sensing data into meaningful information about defect states, equipment condition, and future performance. This Reprint aims to present current progress in this area and to promote closer integration between sensing, intelligent analysis, reliability assessment, and practical industrial applications. The Reprint is intended for researchers, engineers, graduate students, and practitioners working on industrial inspection, nondestructive evaluation, equipment health assessment, and intelligent maintenance. As Guest Editors, we sincerely thank all authors for their contributions, the reviewers for their constructive evaluations, and the editorial team for their professional support throughout the preparation of the Special Issue and this Reprint.

Peng Liu
Guest Editor

sensors

MDPI

Editorial

Sensing Technologies in Industrial Defect Detection

Peng Liu [1,2]

[1] Key Laboratory of CNC Equipment Reliability, Ministry of Education, Jilin University, Changchun 130025, China; liupeng@jlu.edu.cn
[2] School of Mechanical and Aerospace Engineering, Jilin University, Changchun 130022, China

1. Introduction

Industrial defect detection is an essential component of modern manufacturing and plays a critical role in ensuring product quality, operational reliability, and production safety. With the continuing development of intelligent manufacturing, industrial inspection is evolving from conventional offline and operator-dependent examination toward automated, continuous, and data-driven sensing. Machine vision has become one of the most widely adopted approaches for surface and appearance inspection, while non-destructive evaluation techniques such as ultrasonic testing and infrared thermography provide complementary capabilities for detecting internal or subsurface defects [1–3]. These sensing technologies have substantially expanded the observability of manufacturing processes and engineering components, enabling defects to be detected, localized, and characterized with increasing spatial, temporal, and quantitative resolution.

At the same time, advances in sensor networks, data acquisition, signal processing, and artificial intelligence are extending industrial sensing beyond isolated defect identification. Multi-sensor data fusion allows complementary information from different sensing modalities to be integrated for more reliable condition assessment and fault diagnosis [4]. In parallel, prognostics and health management has established a broader framework in which sensing data are used not only to detect abnormal conditions but also to diagnose degradation, predict future system states, and support condition-based or predictive maintenance [5]. Nevertheless, the transition from laboratory demonstrations to reliable industrial deployment remains challenging. Variations in operating conditions and data distributions can substantially reduce the generalization performance of data-driven diagnostic models [6]. The increasing use of deep learning also raises concerns regarding model transparency and the physical credibility of diagnostic decisions, particularly in safety- and quality-critical applications [7]. For prognostic tasks, uncertainty, robustness, interpretability, and practical feasibility are increasingly recognized as important requirements in addition to prediction accuracy alone [8].

Against this background, the Special Issue "Sensing Technologies in Industrial Defect Detection" was established to present recent developments in sensing, inspection, condition monitoring, fault diagnosis, prognosis, and intelligent maintenance for industrial applications. The contributions collected in this Special Issue span multiple sensing modalities and analytical paradigms, including machine vision, infrared thermography, X-ray computed tomography, vibration and multi-source sensing, process and equipment monitoring, machine learning, uncertainty-aware prediction, and intelligent maintenance decision making. Although these studies address different industrial systems and defect or fault scenarios, they collectively reflect a common evolution in the field: industrial sensing is progressing from the acquisition of defect-related signals toward integrated frameworks capable of extracting physically meaningful information, evaluating equipment or product

condition, predicting degradation, and supporting operational decisions. This Special Issue therefore provides a timely perspective on the increasingly close integration of sensing technologies, signal and image processing, artificial intelligence, reliability assessment, and intelligent manufacturing.

2. Overview of Published Papers

Several studies focus directly on defect and quality inspection. Solovev et al. (6) present an infrared-thermography-based framework for in-process laser welding monitoring, in which spatter and weld-zone thermal features are combined with machine learning to identify multiple welding defects. Saiyod et al. (4) address printed circuit board inspection using a training-free, reference-based image-processing framework, highlighting the continued value of interpretable deterministic methods when labeled data are limited or auditability is required. Chang et al. (10) propose an explainable multimodal neural network for final-state quality inspection of hairpin windings, combining image information with span-related geometric priors. These studies demonstrate that industrial defect detection increasingly benefits from the integration of physical or geometric knowledge with data-driven models rather than relying solely on generic image classifiers.

Other papers extend the Special Issue from direct defect localization to degradation assessment and condition monitoring. Chai et al. (1) develop a hierarchical multi-scale temporal convolutional network for tool condition monitoring using multi-sensor signals, while Zhang et al. (3) propose an adaptive signal-decomposition and clustering strategy for bearing fault detection. Both studies emphasize the importance of extracting weak, multi-scale, and temporally evolving features from complex sensor signals. Zhang et al. (2) link mesoscopic defect characteristics obtained through X-ray computed tomography to fatigue life in 6061-T6 aluminum alloy, illustrating how defect sensing can support quantitative assessment of subsequent structural performance.

The Special Issue also includes studies that move beyond diagnosis toward prognosis and operational decision making. Qu et al. (7) address remaining useful life prediction through Bayesian convolutional neural networks with dual-output units, explicitly considering aleatoric and epistemic uncertainty. Dai et al. (8) propose a two-stage wind turbine fault detection strategy for highly imbalanced SCADA data, combining anomaly screening with supervised classification. Liu et al. (9) further connect state information with maintenance planning through a multi-head deep reinforcement learning approach for joint maintenance decisions in an electro-hydraulic servo fatigue testing machine. At the system level, Liu et al. (5) integrate wearable sensing and multi-source information with enhanced Petri nets to improve production system monitoring and real-time state representation.

Taken together, these studies show that the scope of industrial defect detection is becoming broader and more interconnected. Nevertheless, several challenges remain. First, sensing performance under variable industrial conditions, including illumination changes, environmental interference, operating-condition drift, sensor noise, and domain shifts, still limits model transferability. Second, high-quality labeled defect data are often scarce, imbalanced, or expensive to obtain, particularly for rare but safety-critical defects. Third, multimodal sensing can improve observability but introduces new issues in calibration, synchronization, feature alignment, and information redundancy. Fourth, increasingly complex learning models require stronger interpretability, uncertainty quantification, and traceable decision mechanisms before they can be trusted in safety- or quality-critical production environments.

Future research should therefore place greater emphasis on physics-informed and knowledge-guided learning, robust multimodal fusion, self-supervised and few-shot learning, uncertainty-aware diagnosis, and explainable artificial intelligence. Equally important

https://doi.org/10.3390/s26185706

is the transition from laboratory demonstrations to long-term industrial validation. Sensor systems and algorithms should be evaluated across machines, materials, production batches, and operating conditions using standardized metrics and reproducible protocols. Edge computing and lightweight models will also be important for real-time deployment, while digital twins and closed-loop manufacturing systems may provide a pathway for converting sensing results into adaptive process control and maintenance actions.

3. Conclusions

The papers collected in this Special Issue demonstrate the continuing evolution of industrial sensing from defect observation toward intelligent condition understanding. The contributions cover multiple sensing modalities and analytical paradigms, ranging from deterministic image comparison and thermal monitoring to multi-sensor temporal learning, uncertainty-aware prognosis, and reinforcement-learning-based maintenance decisions. Their diversity reflects the fact that no single sensor or algorithm is universally optimal for industrial defect detection; successful solutions depend on the physical characteristics of the target defect, the manufacturing environment, data availability, and the required level of interpretability and real-time performance.

Looking ahead, progress will increasingly depend on the coordinated development of sensing hardware, data analytics, domain knowledge, and industrial deployment strategies. Robustness, generalization, data efficiency, multimodal integration, uncertainty quantification, and explainability remain central research priorities. By addressing these challenges, sensing technologies can support a transition from reactive inspection toward continuous, predictive, and eventually adaptive quality assurance. We hope that the contributions in this Special Issue will provide useful references for researchers and engineers and stimulate further interdisciplinary research in intelligent industrial inspection and smart manufacturing.

Funding: This research was supported by the National Natural Science Foundation of China (grant no. 52475559).

Conflicts of Interest: The author declares no conflicts of interest.

List of Contributions:

1. Chai, A.; Fang, Z.; Lian, M.; Huang, P.; Guo, C.; Yin, W.; Wang, L.; He, E.; Li, S. Hi-MDTCN: Hierarchical Multi-Scale Dilated Temporal Convolutional Network for Tool Condition Monitoring. *Sensors* **2025**, *25*, 7603. https://doi.org/10.3390/s25247603.
2. Zhang, L.; Yang, Y.; Chen, H.; Lv, S. Correlation Between Meso-Defect and Fatigue Life Through Representing Feature Analysis for 6061-T6 Aluminum Alloys. *Sensors* **2026**, *26*, 631. https://doi.org/10.3390/s26020631.
3. Zhang, B.; Shen, B.; Gao, Z.; Shao, Y.; Pang, Z.; Yin, X. Composite Fault Feature Index-Guided Variational Mode Decomposition with Dynamic Weighted Central Clustering for Bearing Fault Detection. *Sensors* **2026**, *26*, 1394. https://doi.org/10.3390/s26041394.
4. Saiyod, S.; Nonsakhoo, W.; Li, Z.; Sirisawat, P. Defect-Intent Ambiguity Addressing for Training-Free Deterministic PCB Defect Localization via Template Selection and Dissimilarity Mapping. *Sensors* **2026**, *26*, 1541. https://doi.org/10.3390/s26051541.
5. Liu, P.; Li, X.; Zhang, C.; Kang, Y.; Qian, J.; Chen, W. Production System Monitoring Based on Petri Nets Enhanced with Multi-Source Information. *Sensors* **2026**, *26*, 1785. https://doi.org/10.3390/s26061785.
6. Solovev, G.; Klokov, E.; Krasnov, D.; Sokolov, M. Advancing Defect Detection in Laser Welding: A Machine Learning Approach Based on Spatter Feature Analysis. *Sensors* **2026**, *26*, 1825. https://doi.org/10.3390/s26061825.

https://doi.org/10.3390/s26185706

7. Qu, Z.; He, J.; Liu, Y.; Mao, S.; Han, X. Interval Prediction of Remaining Useful Life Based on Uncertainty Quantification with Bayesian Convolutional Neural Networks Featuring Dual-Output Units. *Sensors* **2026**, *26*, 2592. https://doi.org/10.3390/s26092592.
8. Dai, J.; Rotea, M.; Kehtarnavaz, N. A Two-Stage Classification Method for Improved Fault Detection in Wind Turbines Based on SCADA Data. *Sensors* **2026**, *26*, 3865. https://doi.org/10.3390/s26123865.
9. Liu, P.; Huang, G.; Xi, J.; Wu, J. Multi-Component Joint Maintenance Decision for Electro-Hydraulic Servo Fatigue Testing Machine Based on Multi-Head Deep Reinforcement Learning. *Sensors* **2026**, *26*, 4087. https://doi.org/10.3390/s26134087.
10. Chang, X.; Zhang, B.; Gao, Z.; Chen, S.; Liu, J. A Span-Prior-Guided Explainable Multimodal Neural Network Method for Final-State Quality Inspection of Hairpin Windings. *Sensors* **2026**, *26*, 4857. https://doi.org/10.3390/s26154857.

References

1. Gao, Y.; Li, X.; Wang, X.V.; Wang, L.; Gao, L. A Review on Recent Advances in Vision-Based Defect Recognition towards Industrial Intelligence. *J. Manuf. Syst.* **2022**, *62*, 753–766. [CrossRef]
2. Cantero-Chinchilla, S.; Wilcox, P.D.; Croxford, A.J. Deep Learning in Automated Ultrasonic NDE—Developments, Axioms and Opportunities. *NDT E Int.* **2022**, *131*, 102703. [CrossRef]
3. Pugazhendi, P.; Viswanathan, V.; Sengottaiyan, S.; Nesamony, J.V.P.; Madheswaran, S.K.; Victor, S. Infrared Thermography for Defect Detection in Fiber-Reinforced Polymer Composites: A Physics-Informed Analytical Synthesis and Review. *Measurement* **2026**, *287*, 122422. [CrossRef]
4. Kibrete, F.; Woldemichael, D.E.; Gebremedhen, H.S. Multi-Sensor Data Fusion in Intelligent Fault Diagnosis of Rotating Machines: A Comprehensive Review. *Measurement* **2024**, *232*, 114658. [CrossRef]
5. Zio, E. Prognostics and Health Management (PHM): Where Are We and Where Do We (Need to) Go in Theory and Practice. *Reliab. Eng. Syst. Saf.* **2022**, *218*, 108119. [CrossRef]
6. Xiao, Y.; Shao, H.; Yan, S.; Wang, J.; Peng, Y.; Liu, B. Domain Generalization for Rotating Machinery Fault Diagnosis: A Survey. *Adv. Eng. Inform.* **2025**, *64*, 103063. [CrossRef]
7. Chen, G.; Yuan, J.; Zhang, Y.; Zhu, H.; Huang, R.; Wang, F.; Li, W. Enhancing Reliability Through Interpretability: A Comprehensive Survey of Interpretable Intelligent Fault Diagnosis in Rotating Machinery. *IEEE Access* **2024**, *12*, 103348–103379. [CrossRef]
8. Salinas-Camus, M.; Goebel, K.; Eleftheroglou, N. A Comprehensive Review and Evaluation Framework for Data-Driven Prognostics: Uncertainty, Robustness, Interpretability, and Feasibility. *Mech. Syst. Signal Process.* **2025**, *237*, 113015. [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.

sensors

MDPI

Article

Hi-MDTCN: Hierarchical Multi-Scale Dilated Temporal Convolutional Network for Tool Condition Monitoring

Anying Chai [1], Zhaobo Fang [1], Mengjia Lian [2,*], Ping Huang [1], Chenyang Guo [1], Wanda Yin [1], Lei Wang [1], Enqiu He [3] and Siwen Li [1]

[1] College of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China; chaianying@sut.edu.cn (A.C.); fangzhaobo2023@gmail.com (Z.F.); huangping0809@sut.edu.cn (P.H.); guo_cy@smail.sut.edu.cn (C.G.); yinwanda20@gmail.com (W.Y.); wanglei20250901@gmail.com (L.W.); lisiwen@sut.edu.cn (S.L.)
[2] School of Mathematics and Information Engineering, Longyan University, No. 1 Dongxiao North Road, Xinluo District, Longyan 364012, China
[3] School of Chemical Equipment, Shenyang University of Technology, No. 30, Guanghua Street, Hongwei District, Liaoyang 111003, China; heq12345@gmail.com
* Correspondence: aioteam@sut.edu.cn

Abstract

Accurate identification of tool wear conditions is of great significance for extending tool life, ensuring processing quality, and improving production efficiency. Current research shows that signals collected by a single sensor have limited dimensions and cannot comprehensively capture the degradation process of tool wear, while multi-sensor fusion recognition methods cannot effectively handle the complementarity and redundancy between heterogeneous sensor data in feature extraction and fusion. To address these issues, this paper proposes Hi-MDTCN (Hierarchical Multi-scale Dilated Temporal Convolutional Network). In the network, we propose a hierarchical signal analysis framework that processes the signal in segments. When processing intra-segment signals, we design a Multi-channel one-dimensional convolutional network with attention mechanism to capture local wear features at different time scales and fuse them into a unified representation. When processing signal segments, we design a Bi-TCN module to further capture long-term dependencies in wear evolution, mining the overall trend of tool wear over time. Hi-MDTCN adopts a dilated convolution mechanism, which can achieve an extremely large receptive field without building an overly deep network structure, effectively solving problems faced by recurrent neural networks in long sequence modeling such as gradient vanishing, low training efficiency, and poor parallel computing capability, achieving efficient parallel capture of long-range dependencies in time series. Finally, the proposed method is applied to the PHM2010 milling data. Experimental results show that the model's tool condition recognition accuracy is higher than traditional methods, demonstrating its effectiveness for practical applications.

Keywords: tool condition monitoring; multi-modal; multi-sensor fusion; deep learning

1. Introduction

Milling, as an efficient cutting process, utilizes rotating milling cutters to perform intermittent cutting on workpieces, capable of machining planes, grooves, threads, and complex geometric components, and is widely applied in industrial manufacturing and

mass production [1]. Cutting tools are essential components in milling processes, and tool breakage can easily cause unexpected downtime of milling equipment. Accumulated tool wear leads to breakage, resulting in time and cost losses, with tool failures accounting for 7–20% of milling machine downtime and tool replacement costs accounting for 3–12% of total processing costs [2]. Although preventive replacement of milling tools is performed before breakage, replacement strategies rely on operator experience and lack accuracy. Research shows that the effective life of milling tools is typically only utilized at 50–80%; therefore, real-time monitoring of tool conditions can maximize tool life utilization and save processing costs [3].

As artificial intelligence advances, monitoring systems have gradually been applied to tool condition monitoring. Tool Condition Monitoring (TCM) systems are divided into offline and online types. Offline TCM systems require interruption of the machining process and use optical microscopes to inspect tool health during downtime periods [4]. Downtime inspection adds additional labor and time costs, and due to limited inspection frequency, it cannot timely detect tool wear changes. In contrast, in online TCM systems, machining operations continue without interruption, and data can be processed and tool conditions monitored in real time within specified cycles [5].

At present, techniques for identifying tool health status are divided into two categories: physics-based models and data-driven models. Physics-based models rely on physical laws or semi-empirical formulas to predict tool wear. These methods usually depend on mathematical descriptions of cutting physics, but due to the complexity and high nonlinearity of the cutting process, such methods are difficult to adapt to real-world conditions and have limited prediction accuracy [6]. In contrast, data-driven models can be trained with sufficient historical data and model the nonlinear dependencies between features extracted from sensor signals and cutting conditions to indirectly infer tool wear. Compared with physics-based models, data-driven models perform better in handling complex working conditions and large-scale data, and can achieve real-time monitoring based on sensor data. Therefore, data-driven models demonstrate broad application prospects in the field of tool wear monitoring [7]. However, current data-driven models struggle to capture both historical and future context in temporal data, and when facing multi-source sensor data, their ability to extract key wear features is limited, resulting in insufficient robustness under complex working conditions and making it difficult to further improve recognition accuracy.

To address these issues, this paper proposes a hierarchical multi-scale dilated temporal convolutional network (Hi-MDTCN) for multi-sensor fusion-based online TCM method. It employs a parallel branch structure to separately perform convolutional extraction and channel attention weighting on multi-source signals, significantly enhancing the representation quality of various types of signal features; secondly, it models temporal dependencies between segments through multi-scale dilated convolutional modules and combines residual connection mechanisms to enhance deep feature transmission capabilities; furthermore, it proposes a lightweight channel attention mechanism for dynamic weighting of intrasegment and intersegment features, effectively improving the model's sensitivity to key wear features. Compared with traditional methods, Hi-MDTCN possesses higher parallel computing capability and stronger modeling stability, allowing the training complexity to be reduced while maintaining high prediction accuracy, thereby achieving the purpose of monitoring tool conditions in real time. The contributions of this paper are as follows

(a) This paper proposes a Bidirectional Temporal Convolutional Module (Bi-TCN), which employs a symmetric padding strategy to achieve non-causal modeling, enabling feature extraction to simultaneously capture both historical and future contextual

information. Compared to the causal relationship constraints of traditional TCN, this module significantly enhances the temporal feature expression capability.

(b) For multi-source sensor data feature extraction, a lightweight intra-segment attention mechanism is designed, dynamically enhancing key wear features through channel attention weights. This mechanism can improve feature expression capability without increasing model complexity, effectively enhancing model robustness under complex working conditions.

(c) The paper proposes a Hierarchical Multi-scale Temporal Network Hi-MDTCN, and designs a two-stage hierarchical processing strategy for intra-segment and inter-segment in the network. Intra-segment processing employs three-way parallel convolutions to process multi-source signals and introduces channel attention mechanisms to strengthen key features. Inter-segment processing adopts multi-scale dilated convolutional mechanisms to capture multi-scale patterns in the tool wear evolution process. The architecture achieves collaborative modeling of local features and global trends through the integration of time series slices.

The remaining parts are organized as follows. Section 2 reviews some related work on current TCM methods, Section 3 briefly introduces the basic theory related to TCN, Section 4 elaborates in detail the proposed Hi-MDTCN method framework. Section 5 demonstrates the effectiveness of the proposed method in tool milling datasets and compares it with other methods. Finally, Section 5 provides the conclusion.

2. Related Work

Deploying TCM systems can effectively improve production quality and reduce energy consumption [8]. At present, data-driven TCM methods are divided into direct methods and indirect methods [9]. Direct methods measure tool geometry based on vision systems; however, due to the high cost of high-speed cameras and the large amount of data generated that require substantial computational resources to process, the predictability of online monitoring processes is usually lower [10].

Indirect sensing methods determine tool wear by analysing multi-source signals, enabling tool condition monitoring with minimal interference during actual cutting processes [11]. Compared to direct methods, indirect methods are more practical and cost-effective [12,13]. However, indirect methods are typically related to the machine tool's response to the cutting process, are susceptible to noise interference, and therefore often less accurate than direct methods [14,15]. TCM using indirect methods relies on traditional signal processing techniques, and due to the highly non-linear and multi-variate coupling characteristics of the cutting processes, it is difficult to extract effective features from complex multi-source sensor data [16,17].

In recent years, as artificial intelligence advances rapidly, machine learning approaches have substantially enhanced tool condition recognition accuracy in indirect monitoring methods by leveraging automated feature extraction and pattern recognition [18–20]. Machine learning algorithms such as Support Vector Machine (SVM) [21], Random Forest (RF) [22], and K-Nearest Neighbor (KNN) [23] have been used for tool wear prediction, but these methods still rely on manual feature engineering and have limited expression capability for high-dimensional data. To address these limitations, deep learning techniques have increasingly emerged as research focal points owing to their robust automatic feature extraction capabilities and superior performance in handling complex data [24,25]. Zhu et al. [26] utilized fast neural network methods to extract features from cutting force signals, achieving effective monitoring of tool wear in micro-milling processes. Huang et al. [27] proposed a tool wear monitoring method combining Short-Time Fourier Transform with

deep convolutional neural networks, implementing adaptive feature extraction and wear state prediction by modeling the time-frequency characteristics of vibration signals. Wang et al. [28] extracted time-frequency features of acoustic emission signals through Fourier transform and wavelet packet decomposition, and used BP neural networks for modeling and analysis; experimental results showed that the tool wear monitoring method based on BP neural networks outperformed SVM. Although the above studies have confirmed the effectiveness of deep learning for tool condition monitoring, they only analyze single-source signals. Single sensor signals often can only reflect certain aspects of tool conditions and struggle to comprehensively cover the actual working conditions of tools in complex machining processes.

Compared to single-source signal monitoring methods, multi-sensor fusion technology can significantly improve the accuracy and robustness of tool wear prediction, reducing the impact of environmental interference and signal uncertainty [29,30]. Li et al. [31] proposed a residual variational autoencoder transfer learning model based on multi-sensor fusion, optimising the encoding structure of variational autoencoders through residual networks to achieve fault diagnosis between different tools under parameter transfer strategies. Lin et al. [32] effectively improved the accuracy of tool wear state recognition by establishing Stacked Sparse Autoencoders (SSAE) and BP neural networks for multi-sensor signals, respectively. Song et al. [33] proposed a tool wear monitoring method that fuses multi-kernel Gaussian process regression with stacked multi-layer denoising autoencoders, achieving high precision feature representation learning and reliable wear state prediction by optimising kernel function parameters and denoising force signal features. The above autoencoder-based methods can effectively extract deep features from multi-source sensor data and enhance feature representation capabilities through optimisation mechanisms such as residual networks and stacked sparse structures, providing feasible solutions for tool condition monitoring in actual machining processes. However, traditional AE structures struggle to capture the temporal changes during the tool wear process and cannot effectively model the dynamic process of wear evolution.

To address this issue, recurrent neural network-based methods have become a new research direction. Recurrent neural networks, through memory units and gating mechanisms, can better preserve historical information and temporal dependencies of tool wear, further improving the accuracy of monitoring and real-time performance. Sayyad et al. [34] combined LSTM with feature selection to optimize and sequentially model sensor signals, providing reliable monitoring means for tool wear control. Zhou et al. [35] designed a dual parallel BiLSTM architecture to process the multi-domain features of vibration and cutting force signals separately, achieving an accurate prediction of tool wear states with temporal dependency patterns. Xiao et al. [36] proposed a tool wear prediction method based on Bayesian optimized Gated Recurrent Units (GRU), achieving wear state prediction with high generalization capability by modeling time domain, frequency domain, and time-frequency domain features of multi-working condition signals with hyperparameter optimization. Liu et al. [37] effectively improved the accuracy of tool wear state recognition by deep mining temporal feature dependencies through BiGRU networks combined with variational mode decomposition for signal optimisation processing. Although RNNs can effectively model the dynamic evolution process of tool wear, their current state prediction must be based on the calculation results of the previous moment, causing errors to accumulate during temporal propagation, leading to a decrease in the prediction accuracy as time steps increase [38].

To overcome these problems, one-dimensional convolutional neural networks (1D-CNN), with their local receptive fields and parameter-sharing mechanisms, can efficiently

extract local patterns and features from temporal data while reducing computational complexity. Wang et al. [39] proposed a multi-sensor tool wear diagnosis method combining 1D-CNN with Deep Generalized Canonical Correlation Analysis (DGCCA), achieving accurate prediction of tool wear states by using DGCCA with attention mechanisms to fuse output features from various networks and remove redundant information. Dong et al. [40] proposed an improved CaAt-ResNet-1d model, effectively enhancing the accuracy of tool wear states in complex cutting environments by adding CaAt1 channel attention mechanisms and CaAt5 mechanisms for automatically learning features of different channels in residual network blocks. Yang et al. [41] significantly improved the stability of tool wear state prediction under varying working conditions by using multi-scale entropy values as conditional indicators to construct one-dimensional feature vectors for input into 1D CNN. Furthermore, by converting temporal signals into two-dimensional image representations, two-dimensional convolutional neural networks (2D-CNN) can simultaneously capture spatial structural features in the time-frequency domain, providing more comprehensive feature expression capabilities for tool wear state monitoring and further improving model prediction accuracy. Zhang et al. [42] enhanced the model's capability to identify various tool wear stages by transforming force signals into two-dimensional image formats using time-frequency analysis methods and implementing convolution kernels with varying sizes to capture multi-scale features simultaneously. Li et al. [43] proposed an integrated convolutional neural network architecture, achieving real-time monitoring of tool conditions during cutting processes by combining Short-Time Fourier Transform techniques to map one-dimensional acoustic signals to two-dimensional representations in the time-frequency domain. Liu et al. [44] proposed a real-time milling chatter detection method based on Continuous Wavelet Transform (CWT) scalograms and deep convolutional neural networks (CNN), providing a reliable solution for real-time monitoring of tool wear in milling processes. Although CNNs have achieved good performance in tool wear monitoring, there are still some limitations. For example, CNNs often use fixed-size convolution kernels, resulting in limited receptive fields and difficulties in modeling long-distance dependencies [45]. Furthermore, extracting features from the time and frequency domain from multi-sensors increases model complexity; CNN channels need to share convolution kernels, and excessive information exacerbates the difficulty of extracting and fusing redundant features, easily introducing unnecessary information and affecting the quality of the feature [46]. When using 2D-CNN for tool wear monitoring, temporal signals need to be converted to time-frequency images as input, with preprocessing steps that are computationally expensive and complex, bringing additional computational overhead to the model and introducing extra delay [47]. Therefore, 2D-CNN methods are often not efficient enough in tool monitoring applications with high real-time requirements.

To compensate for the shortcomings of the above research methods, the Hi-MDTCN proposed in this paper breaks through the limitations of traditional methods through its hierarchical design and bidirectional temporal convolutional mechanism. Its advantages lie in the ability to simultaneously capture contextual information from both past and future, adaptively filter important features through lightweight channel attention mechanisms, and implement long dependency relationship modeling using multi-scale dilated convolutions. This architecture does not require complex signal conversion preprocessing, directly processes original temporal data, significantly reduces computational delay while maintaining high prediction accuracy and generalization capability, providing a real-time reliable solution for real-time tool condition monitoring.

3. Preliminaries

During the tool wear process, the evolution of wear states exhibits strong historical dependence, with the current evolution significantly influenced by previous processing conditions. This temporal evolution characteristic can be formally expressed as Equation (1):

$$s_t = f(s_{t-1}, s_{t-2}, \ldots, s_{t-n}) + \epsilon_t \tag{1}$$

where s_t is the tool wear state at time t, and ϵ_t is random noise.

Therefore, TCM is essentially a sequence modeling problem, which requires the capture of temporal dependencies to achieve accurate wear prediction and optimize tool replacement strategies and production efficiency. TCN addresses the sequence modeling problem by ensuring temporal causality through causal convolutions, meaning current predictions rely only on historical data, avoiding future information leakage. Simultaneously, it employs dilated convolutions to exponentially expand the receptive field, efficiently capturing long-term evolution patterns through sparse sampling. These architectural characteristics enable TCN to possess both local feature extraction and global trend modeling capabilities, providing an effective solution for complex time series modeling problems in tool condition monitoring.

3.1. Causal Convolutions

In the time series modeling domain, causal convolution is a core component of TCN. This structure strictly enforces temporal causality constraints, meaning that the output at the current time step can only depend on input information from the current and previous time steps, and cannot access data from future time steps. Its mathematical expression is given by Equation (2).

$$y_t = \sum_{i=0}^{k-1} w_i \cdot x_{t-i} \tag{2}$$

where, y_t represents the output feature at time step t, w_i represents the i-th weight parameter of the convolution kernel, k represents the size of the convolution kernel in the time dimension, and x_{t-i} represents the input feature at relative position i. This approach ensures the model cannot access future information, satisfying real-time prediction requirements.

3.2. Dilated Convolution

The historical information capacity of causal convolution only increases linearly with network depth, making it difficult to process sequence tasks requiring long-term dependencies. Dilated convolution, as the second core component of TCN, enhances the network's modeling capability for long sequence dependencies by introducing sampling strategies with gaps. In contrast to standard causal convolution, dilated convolution can exponentially increase the receptive field while preserving parameter efficiency, successfully addressing the difficulty of capturing long-term dependencies. For 1-D sequence input $x \in \mathbb{R}^n$, and filter $f : \{0, \cdots, k-1\} \to \mathbb{R}$, the dilated convolution operation F on sequence element s is defined by Equation (3):

$$F(s) = (x *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \tag{3}$$

where d is the dilation factor, k is the filter size, and $s - d \cdot i$ represents the direction toward past time steps. Therefore, dilated convolution is equivalent to introducing a fixed step length between each tap of the neighboring filter. When $d = 1$, dilated convolution reduces

to regular convolution. Using larger dilation factors enables perception of inputs from more distant positions, effectively expanding the receptive field of the convolutional neural network. The TCN combined with causal dilated convolution is shown in Figure 1.

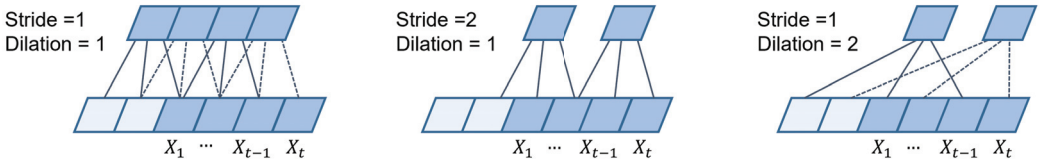


Figure 1. Shows examples of TCN combining causal convolution and atrous convolution at different values of stride and atrous rate, applying a 1d convolution with kernel sizes k = [3; 3; 3], dilation d = [1; 1; 2] and strides s = [2; 1; 1].

4. Methodology and Proposed Model

The proposed Hi-MDTCN architecture consists of three stages: signal preprocessing, intra-segment feature extraction, and inter-segment temporal modeling. The specific framework is shown in Figure 2.

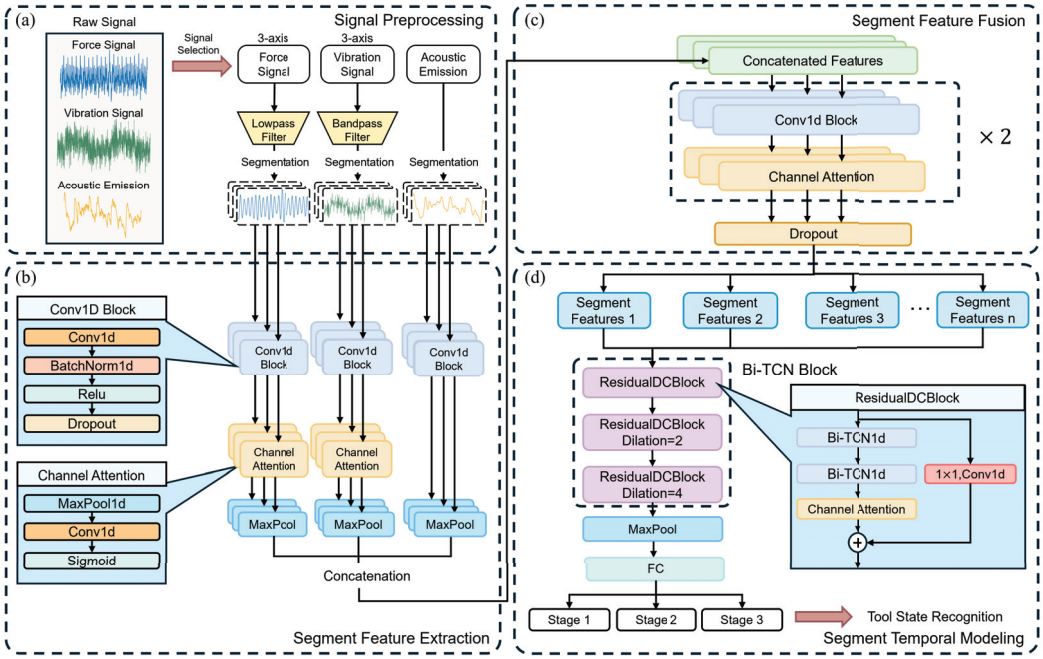


Figure 2. Framework of the proposed approach. (**a**) illustrates signal preprocessing, including filtering and segmentation. (**b**,**c**) represent segment feature extraction and fusion, enhancing intra-segment features. (**d**) shows temporal modeling across segments to monitor tool condition.

First, in the signal preprocessing stage, a Butterworth filtering strategy is adopted for different types of sensor signals to eliminate noise and extract representative features. Next, in the intra-segment feature extraction stage, a Bi-TCN block is introduced, which employs a symmetric padding strategy to achieve non-causal modeling. This allows feature extraction to simultaneously capture both historical and future contextual information, significantly enhancing the temporal feature representation capability compared to the causality constraints of traditional TCNs. Meanwhile, for the feature extraction of multi-source sensor data, a lightweight intra-segment attention mechanism is designed, which dynamically enhances key wear features through channel attention weights, improving feature representation capability without increasing model complexity, and effectively enhancing the robustness of the model under complex working conditions. Finally, in the

inter-segment temporal modeling stage, we propose a hierarchical multi-scale temporal network architecture, which achieves an exponentially expanded receptive field by designing combinations of dilated convolutional layers with increasing dilation rates, thereby enabling efficient and parallel capture of long-term temporal dependencies. The proposed model adopts a hierarchical processing strategy that first extracts local intra-segment features and then aggregates them into global representations.

This approach addresses the issues of gradient vanishing and inefficient parallel computation in traditional recurrent structures for long sequence modeling, while also avoiding the computational overhead of complex time-frequency transformations required by 2D-CNN methods. As a result, the model achieves better training stability and inference capability while maintaining high prediction accuracy.

4.1. Signal Preprocessing

By analyzing the first-order difference features of the force sensor signal, the end time of the feed and the start time of the tool retraction during the milling process can be detected, enabling precise extraction of the effective cutting segment. The original force signal usually contains high-frequency noise, which can be effectively reduced by applying a moving average window for smoothing. Mathematically, this is expressed as Equation (4)

$$f_{\text{smooth}}[i] = \frac{1}{W} \sum_{j=i-\frac{W-1}{2}}^{i+\frac{W-1}{2}} f_{\text{raw}}[j] \tag{4}$$

where $f_{\text{raw}}[j]$ denotes the amplitude of the original force signal at sampling point j, and $f_{\text{smooth}}[i]$ denotes the amplitude of the smoothed signal at sampling point i. W represents the width of the moving average window, and the summation range for j is from $i - \frac{W-1}{2}$ to $i + \frac{W-1}{2}$, with the window centered at the current point.

To accurately detect feed end and retraction start, the force signal is smoothed to suppress high-frequency noise while preserving transient trends. The moving-average window size W is defined as Equation (5)

$$W = \left\lfloor \frac{f_s \times 60}{k \times n \times Z} \right\rceil, \tag{5}$$

where f_s is the sampling frequency (Hz), n is the spindle speed (rpm), and Z is the number of tool teeth. The parameter k denotes the smoothing coefficient, and $\lfloor \cdot \rceil$ indicates rounding toward the nearest integer with a lower bound. This formulation exploits the physical periodicity of the tooth-passing frequency $f_{\text{TPF}} = n \times Z/60$. The window should span multiple tooth-passing periods to suppress periodic impacts and harmonics while avoiding excessive smoothing that would blur transients. Here, $k \in [0.1, 0.3]$ corresponds to the number of tooth-passing periods covered by the window, balancing noise suppression and transient preservation.

After smoothing, a first-order difference is applied to calculate the difference between adjacent sampling points, thereby detecting abrupt changes in the signal. Mathematically, this is expressed as Equation (6)

$$diff[i] = f_{\text{smooth}}[i+1] - f_{\text{smooth}}[i] \tag{6}$$

where $diff[i]$ denotes the first-order difference at sampling point i, and $f_{\text{smooth}}[i]$ and $f_{\text{smooth}}[i+1]$ denote the amplitudes of the smoothed signal at two adjacent sampling points.

A positive value indicates that the signal is rising, which may correspond to the end of the tool feed, with the tool entering a stable cutting state. A negative value indicates that the signal is falling, which may correspond to the start of tool retraction.

Based on the statistical characteristics of the difference, the weighted sum of the mean and standard deviation of the absolute value of the difference sequence is used as the threshold to ensure good adaptability for signals of different strengths, as shown in Equation (7):

$$\theta = \bar{d} + k \cdot \sigma_{|d|} \tag{7}$$

here, θ is the threshold, $\bar{d}$ represents the mean of the difference sequence, k is an adjustable weighting coefficient, and $\sigma_{|d|}$ is the standard deviation of the absolute value of the difference sequence. Based on this threshold, continuous running windows are used to detect the start and end points. When the absolute value of the difference in the window is consistently greater than the threshold and the sign is consistent, it is determined as the end of the feed or the start of tool retraction. The signal segment near the end of the feed or the start of tool retraction reflects the state characteristics of the tool approaching the wear limit. To further enhance the effective signal features, it is necessary to filter out noise and remove interference from high-frequency noise in the signal segment.

The filtering method is selected according to the frequency characteristics of the signal, including low-pass, high-pass, and band-pass. For different types of signals, different filtering strategies are adopted to effectively remove noise and improve the accuracy of feature extraction. Therefore, a Butterworth filter is designed to perform low-pass filtering, retaining the low-frequency components of the signal and smoothing the signal. The transfer function of an n-order Butterworth filter with cutoff frequency ω_c can be mathematically expressed as Equation (8):

$$H(\omega) = \frac{1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}} \tag{8}$$

Here, ω is the frequency, ω_c is the cutoff frequency, and n is the filter order. For force signals, low-pass filtering is used so that the low-frequency components of the signal can pass through without attenuation, while the high-frequency components are attenuated according to the filter order, suppressing high-frequency noise in the signal and vibration features. Vibration signals are processed with band-pass filtering, allowing the signal within a specific frequency band to pass through without attenuation, thereby suppressing drift and low-frequency noise, while retaining frequency bands related to vibration. The Butterworth filter has the characteristic of the maximum flatness in the passband among all types of filters, ensuring that the phase and amplitude of the signal are not distorted by filtering, as shown in Figure 3.

Traditional long-term sequence signal modeling methods have high computational complexity and memory usage, making them difficult to train. In addition, it is difficult for a single time step sampling point to directly represent the tool wear state; local feature extraction through continuous time windows is required to effectively reveal the wear evolution pattern. To address these issues, a hierarchical two-stage processing strategy is designed. By uniformly dividing the stable cutting segment into multiple time slices, multi-sensor fusion features are extracted for each short segment within the segment, and temporal associations between different segments are established to capture the wear evolution trend. The hierarchical architecture effectively solves the problem of exponential growth in computational complexity caused by direct modeling of long sequences; second, the complementary fusion of multi-segment features can significantly improve the

robustness of the model to noise. In addition, parallel processing of each segment greatly improves computational efficiency. Through the hierarchical design of the time series, both local detail changes and global trend information are preserved, providing a more comprehensive and reliable basis for accurate monitoring of tool wear state.

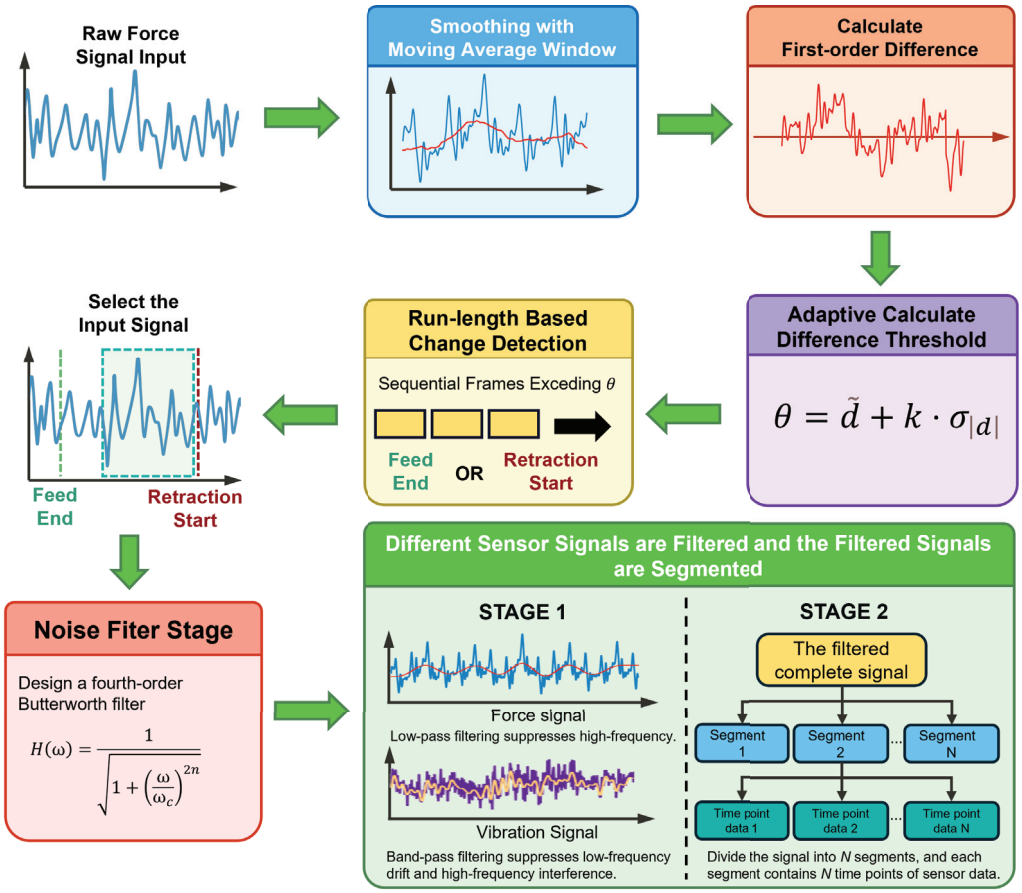


Figure 3. Signal preprocessing procedure.

4.2. Intra-Segment Feature Extraction

In order to capture short-term local features during the tool wear process, an intra-segment feature extractor is designed to extract features in parallel from signals of multiple sensors within each time segment. Specifically, after preprocessing and segmentation, the original multi-channel sensor signals are input into a three-branch parallel lightweight convolutional network, with each branch corresponding to one type of sensor signal: triaxial force sensing, triaxial vibration sensing, and acoustic emission sensing signals. Each branch adopts the same structure: first, a lightweight one-dimensional convolutional block is used to extract local temporal features, then a channel attention mechanism is applied to weight the features of each channel, and finally, global pooling is used to compress the temporal information into a feature vector. Although the number of channels of the original input needs to be increased during feature extraction, since the acoustic emission signal is single-channel at the input, the effect of setting a channel attention module is not obvious. Therefore, after feature extraction of the acoustic emission signal is completed, it is directly pooled and then concatenated with other features for fusion. The intra-segment feature extraction architecture is shown in Figure 4.

A lightweight one-dimensional convolutional module, Conv1dBlock, is used to extract the features of each branch. Each module contains a convolution kernel of length k, BN,

ReLU and Dropout. For the c_m channels, with input features $X \in \mathbb{R}^{c_m \times L}$ of temporal length L, the output at channel j and time t is given by Equation (9):

$$y_{j,t} = \sum_{i=1}^{c_m} \sum_{k=0}^{K-1} w_{j,i,k}\, x_{i,t-k} + b_j \tag{9}$$

here, $w_{j,i,k}$ are the convolution kernel parameters, and b_j is the bias term. After convolution, batch normalization and ReLU activation are applied, followed by Dropout, which can effectively extract local temporal features and suppress overfitting. After feature extraction by the one-dimensional convolutional module, the output features are input into the channel attention module to focus on important channel information. The attention module first performs global pooling on each channel to obtain the channel description vector $z, z \in \mathbb{R}^c$, and then passes it through a one-dimensional convolution and sigmoid function to calculate the channel weights, as mathematically expressed in Equation (10):

$$\begin{cases} \alpha = \sigma(f_{1D}(z)) \\ \sigma(x) = \dfrac{1}{1+e^{-x}} \end{cases} \tag{10}$$

Here, f_{1D} denotes a learnable 1D convolution operation with kernel size k, and $\sigma(\cdot)$ is the Sigmoid activation function. Specifically, for the c-th channel, as shown in Equation (11):

$$\begin{cases} \alpha_c = \sigma\left(\displaystyle\sum_{j=-\frac{k-1}{2}}^{\frac{k-1}{2}} u_j\, z_{c+j} \right) \\ \hat{x}_{c,t} = \alpha_c\, x_{c,t} \end{cases} \tag{11}$$

Here, u_j is the 1D convolution kernel parameter, $\alpha_c \in (0,1)$ is the attention weight of channel c after attention weighting. The channel attention mechanism can achieve cross-channel interaction through local convolution, and introducing a small number of parameters can significantly improve the effectiveness of channel attention. After attention weighting, the temporal feature maps of each branch highlight the key channel information. For each branch, after global pooling and attention weighting, the temporal feature map is compressed to a feature vector of length C by max pooling. For example, if the output feature dimension is $C \times L$, it is pooled to $\mathbb{R}^C$, where each element can be used as the maximum response of the corresponding channel. The feature vectors output by the force sensing branch, vibration sensing branch, and acoustic emission branch are denoted as $f_{dyn}, f_{acc}, f_{ae} \in \mathbb{R}^C$. After obtaining the feature vectors of each branch, the three branch features are concatenated along the channel dimension to form a feature vector of length $3C$, $f = [f_{dyn}; f_{acc}; f_{ae}] \in \mathbb{R}^{3C}$. In order to further integrate multi-source information and enhance joint features, the features extracted by the three parallel convolution branches are concatenated, and then input into a fusion convolution layer. The feature vector concatenated after the first Conv1dBlock and its attention layer is reshaped to a tensor of size $3C \times 1$, which is input into a convolution kernel of size K in the Conv1dBlock. This fusion convolution reduces the number of channels from $3C$ to C, followed by an attention layer to reweight the fused features. Then, a second Conv1dBlock and its attention layer are used for further feature integration, and finally, a fused feature vector of length C is output. The entire process can be regarded as first performing feature extraction and weighting at the branch level, and then obtaining a unified representation through convolution and fusion at the segment level. Finally, a Dropout layer is added at the output to prevent overfitting.

The design of parallel branches enables the network to extract its own key patterns for different types of signals, while also improving computational efficiency. The parallel convolutional structure and attention weighting of multi-source signals significantly enhance the expressive power of intra-segment features and the quality of multi-modal information fusion, providing a richer and more reliable representation for subsequent temporal modeling.

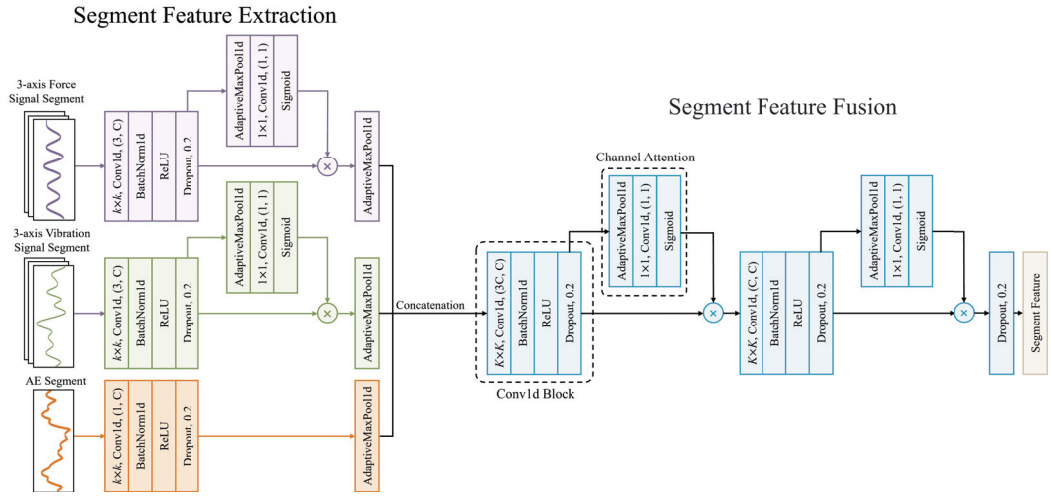


Figure 4. Intra-segment feature extraction process.

4.3. Inter-Segment Temporal Modeling

According to the regulations of numerical control machining programs, tool changes must be performed at the safe and fixed position of the machine tool. Therefore, the judgment of tool status needs to be based on the comprehensive information of the machining stage signals, rather than relying solely on the local signal at a certain moment. Therefore, we propose the design of Bi-TCN, which can utilize both the past and future temporal information within the current stage, instead of only using past information like TCN.

Therefore, in the inter-segment temporal modeling stage, a Bi-TCN1d layer is designed to capture long-term dependencies between different time segments. The forward convolution branch processes the past-direction information of the time series, while the backward convolution branch processes the future-direction information. In this way, the output of the Bi-TCN1d layer at each moment can take into account the contextual dependencies before and after that moment, achieving bidirectional temporal modeling, as shown in Figure 5.

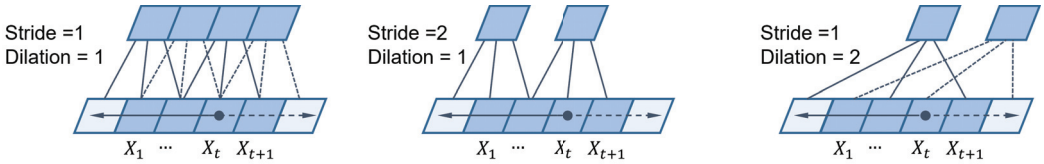


Figure 5. Shows examples of Bi-TCN1d combining causal convolution and atrous convolution at different values of stride and atrous rate, applying a 1d convolution with kernel sizes k = [3; 3; 3], dilation d = [1; 1; 2] and strides s = [2; 1; 1].

At the same time, we adopt a multi-layer stacked residual dilated convolutional network (Residual Dilated Convolutional Block, ResidualDCBlock) to capture long-term dependencies between different time segments. Each ResidualDCBlock module contains two consecutive Bi-TCN1d layers. By using symmetric padding, the output sequence length

is kept the same as the input, achieving non-causal convolution. After each convolution, BN, ReLU activation, and dropout are applied in sequence to enhance the stability and robustness of feature extraction. The overall process is shown in Figure 6.

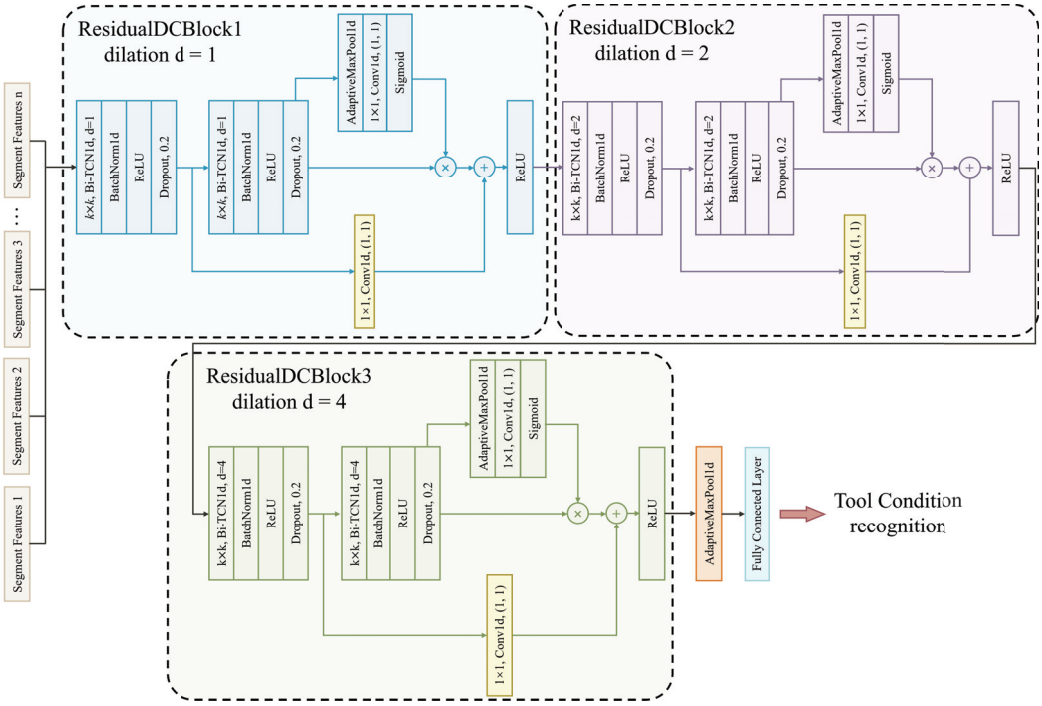


Figure 6. Inter-segment temporal modeling process for identifying tool states. The Bi-TCN Block consists of three stacked ResidualDCBlocks with dilation rates of 1, 2, and 4.

The first Bi-TCN1d first performs feature transformation on the input sequence, and the output is passed through a second Bi-TCN1d layer with the same configuration, then enters the channel attention module, where the features from the forward and backward branches are fused. The attention module performs global pooling on the output features, followed by a 1×1 one-dimensional convolution and Sigmoid activation to generate channel attention weights, highlighting the channel information that more effectively reflects the tool wear state. The final output of the module is the result of the residual connection, that is, the sum of the output from the convolutional layer and the channel attention module with the input. If the number of output channels does not match, a 1×1 one-dimensional convolution is used to match the dimensions, and then ReLU activation is applied to obtain the final features. Specifically, the module takes the segment feature sequence $X \in \mathbb{R}^{c_m \times L}$ as input, where c_m is the number of input channels and L is the length of the sequence. The output of the module is defined by Equation (12):

$$y = \mathrm{ReLU}(\mathcal{F}(x) + \mathcal{H}(x)) \tag{12}$$

here, $\mathcal{F}(x)$ represents the composite mapping in the residual block, including two layers of learnable convolution with dilation, batch normalization, ReLU, and dropout operations, followed by a channel attention module. $\mathcal{H}(x)$ is the identity mapping; when the number of input and output channels is inconsistent, a 1×1 convolution is introduced to match the dimensions, as mathematically expressed in Equation (13):

$$\mathcal{H}(x) = \begin{cases} x, & c_m = c_n \\ \mathrm{Conv}_{1\times1}(x), & c_m \neq c_n \end{cases} \tag{13}$$

here, c_n is the number of output channels. The residual structure allows each module to directly "skip" the input signal to the output, thus enhancing the advantage of increasing the depth of the network in a hierarchical manner. In Bi-TCN, we stack multiple ResidualDCBlock layers, with each layer's dilation rate increasing exponentially, which exponentially expands the receptive field of the network. Therefore, after L convolutional layers, the size of the final receptive field can be calculated similarly to Equation (14):

$$R = (k-1)(2^L - 1) + 1 \tag{14}$$

here, k is the kernel size of each layer, and L is the number of convolutional layers. By increasing the dilation rate, the network can maintain a wide temporal coverage with both shallow and deep layers, effectively obtaining global trend information across multiple segments, as shown in Figure 7.

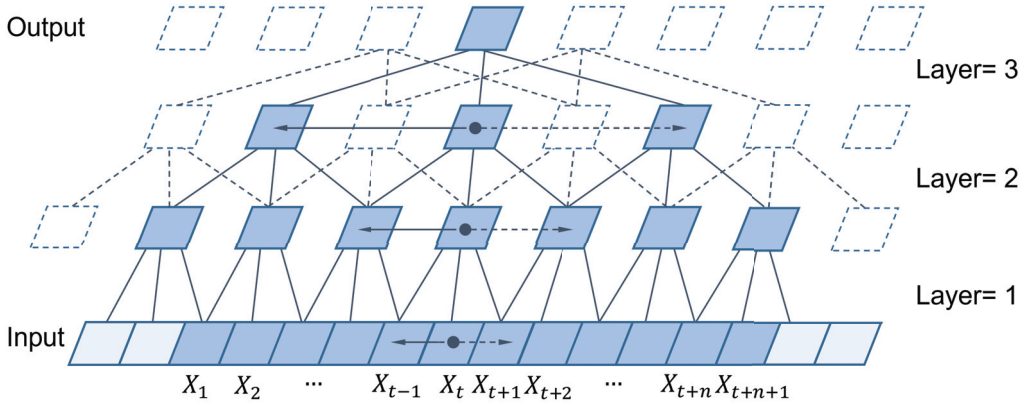


Figure 7. The Bi-TCN module with exponentially expanding receptive fields through stacked layers of different dilation rates. Shown are three Bi-TCN1d layers with kernel sizes k = [3; 3; 3], dilation rates d = [1; 1; 2] and strides s = [2; 1; 1].

The residual connections in each layer provide shortcuts for information propagation in deep networks, allowing gradients to flow directly across layers and greatly alleviating the common problem of gradient vanishing in deep networks. In addition, the attention module dynamically adjusts channel weights in each layer, so that key wear characteristics are enhanced during propagation, improving the model sensitivity and robustness to subtle signals.

Compared with traditional recurrent neural networks, Bi-TCN can compute the output at each position of the sequence in parallel. Therefore, when processing long sequences, Bi-TCN usually has faster training and inference speed. Parallelism and non-causality enable Bi-TCN to utilize both intra-segment and inter-segment information simultaneously. Thus, Bi-TCN not only inherits the parallelism and stability advantages of TCN, but also more fully integrates temporal context through its bidirectional design, achieving sequence modeling of multi-stage, non-stationary tool wear.

The overall description of the proposed method is shown in Algorithm A1 in Appendix A for details., including signal preprocessing (lines 1 to 7), intra-segment feature extraction (lines 8 to 15), and inter-segment temporal modeling (lines 16 to 20).

5. Experimental Study

We conducted experiments on the PHM2010 dataset to test the performance of the proposed method and verify its effectiveness. The code for this paper is available at: https://github.com/2023AIOT/Hi-MDTCN/, accessed on 11 December 2025.

5.1. Dataset Description

The PHM2010 dataset was collected under dry milling conditions using a high-speed CNC machine Roders Tech RFM760 throughout the entire tool life cycle, and contains a total of six tool life cycle experiments. Among them, three cutting experiments, C1, C4, and C6, recorded the tool wear value after each machining stage. The materials of the milling cutter and workpiece are ball-end tungsten carbide tool and stainless steel HRC52, respectively. The workpiece surface was machined row by row along the x-axis using a 6-mm three-flute cutter. After completing one stroke along the x-axis, the tool was retracted to start a new stroke. At the end of each stroke, the flank wear of the three tool teeth was measured offline using a LEICA MZ12 microscope. The specific experimental conditions are shown in Table 1.

Table 1. Experimental conditions for PHM 2010.

Equipment Type	Experimental Equipment	Experimental Projects	Parameters
CNC milling machine	Roders Tech RFM760	Spindle speed (r/min)	10,400
Tool type	3-tooth ball nose milling cutter	Feed rate (mm/min)	1555
Workpiece material	Stainless steel HRC52	Cutting width (mm)	0.125
Force sensor	Kistler 9265B dynamometer	Cutting depth (mm)	0.2
Charge amplifier	Kistler 5019A charge amplifier	Tool feeding amount (mm)	0.001
Vibration sensor	Kistler 8636c acceleration sensor	Sampling frequency (kHz)	50
AE sensor	Kistler acoustic emission sensor	Milling method	Climb milling
Wear measuring device	LEICA MZ12 microscope	Cooling method	Dry cutting
Data acquisition card	NI DAQ data acquisition card		

During the milling process, a Kistler 9265B three-component dynamometer was installed between the worktable and the workpiece to collect force signals in the x, y, and z directions during the cutting process. At the same time, three Kistler 8636c acceleration sensors were directly fixed on the workpiece to measure the machine tool vibration data in the x, y and z directions, respectively. In addition, a Kistler AE sensor was installed on the workpiece to monitor high-frequency stress waves. All sensor signals were collected using an NI DAQ PCI1200 board with a sampling frequency of 50 kHz. Therefore, this dataset contains seven types of signals: force signals in the x, y, and z axes (N), vibration signals in the x, y and z axes (g), and the root mean square value of the acoustic emission signal AE-RMS (V). The experimental setup is shown in Figure 8.

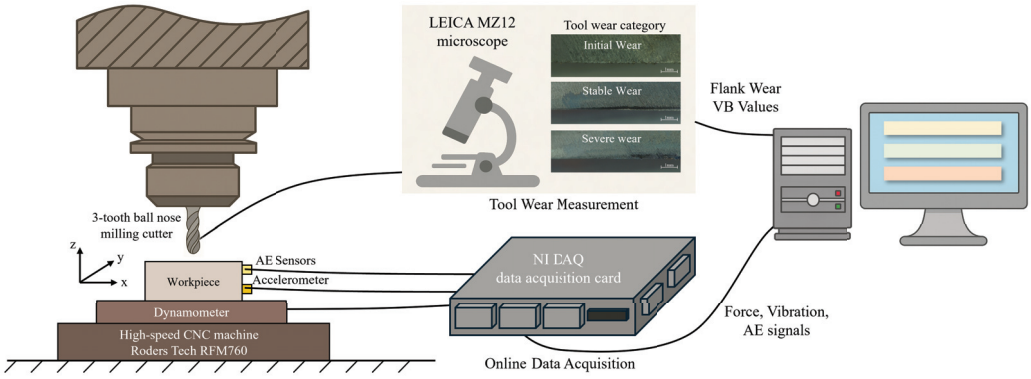


Figure 8. Experimental equipment and configuration.

5.2. Data Preprocessing

Since experiments C1, C4, and C6 documented tool-wear measurements following each machining operation, these three experiments served as the dataset for this research. After each machining operation, three flank-wear values are recorded for the tool. In this

study, the arithmetic mean of these three flank-wear measurements is taken as the final flank-wear value for that experimental record, and the tool-wear stages are evaluated based on the trend of these averaged wear values. The tool wear curves throughout the entire operational cycles of experiments C1, C4, and C6 are shown in Figure 9.

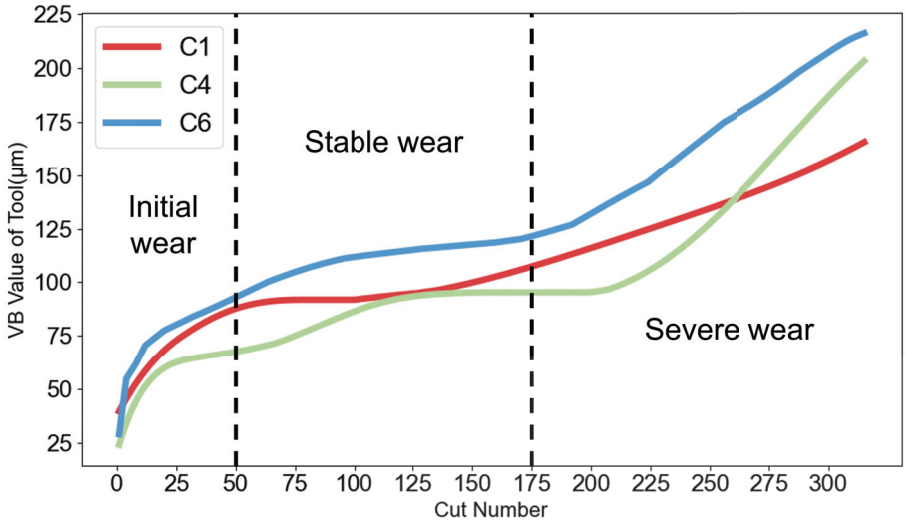


Figure 9. Experimental equipment and configuration.

All cutting tools across the three experiments entered the stable wear phase following the 50th milling operation; nevertheless, the commencement timing of the severe wear phase varied across different experiments. Given that experiment C6 consistently exhibited higher tool wear levels, the initiation point of severe wear in experiment C6 served as the benchmark for differentiating between stable wear and severe wear stages. Consequently, tool wear data spanning operations 1–50 were categorized as initial wear, operations 51–175 as stable wear, and operations 176–315 as severe wear. Information from all seven sensor channels was integrated into the preprocessing workflow. To guarantee reliable assessment of model performance, the experimental dataset was partitioned into three groups for pairwise cross-validation, with comprehensive details presented in Table 2.

Table 2. Cross-validated tool set distribution for T dataset.

Datasets	Training Datasets	Test Datasets
T1	C1–C4	C6
T2	C4–C6	C1
T3	C1–C6	C4

The data acquisition card has a sampling frequency of up to 50 kHz, and each cutting test takes 15 to 30 s. For each cut, the data samples from each sensor, including the feed and retraction stages, contain more than 100,000 data points. Selecting data from the stable cutting stage close to the retraction phase best reflects the tool wear condition. By analyzing the first-order difference features of the force sensor signal, the specific positions of the end of the feed and the start of the retraction can be obtained, as shown in Figure 10. Finally, 50,000 sensor records generated during 1 second of stable cutting closest to the start of retraction were selected, and the original signals were filtered according to the filtering method described in Section 3.1.

The architectural parameters are detailed in Table 3. The parameter selection was designed based on comprehensive considerations of the feature dimensions of multi-sensor data, temporal dependency length, and computational efficiency. Experimental validation shows that this parameter configuration achieves favorable results in experiments.

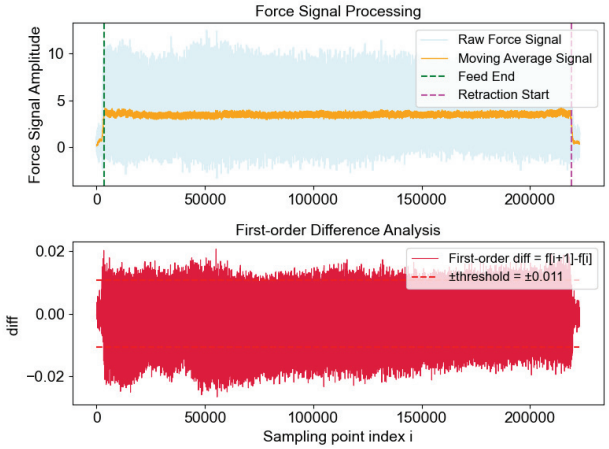


Figure 10. Force signal processing, analysis of the end of the feed and start position of the retraction.

Table 3. Hi-MDTCN Model Architecture Parameter Table.

No.	Stage	Module Name	Input Shape	Output Shape	Input Channels	Output Channels	Kernel Size	Dilation
1	Signal Preprocessing	Lowpass Filter (Force)	[N, 3]	[N, 3]	-	-	-	-
2		Bandpass Filter (Vibration)	[N, 3]	[N, 3]	-	-	-	-
3		None Filter (AE)	[N, 1]	[N, 1]	-	-	-	-
4		Segmentation	[50,000, 7]	[50, 7, 1000]	-	-	-	-
5	Force Signal Branch	Conv1d Block	[B × 50, 3, 1000]	[B × 50, 32, 1000]	3	32	3	1
6		Channel Attention	[B × 50, 32, 1000]	[B × 50, 32, 1000]	32	32	3	-
7		MaxPool	[B × 50, 32, 1000]	[B × 50, 32, 1]	32	32	-	-
8	Vibration Signal Branch	Conv1d Block	[B × 50, 3, 1000]	[B × 50, 32, 1000]	3	32	3	1
9		Channel Attention	[B × 50, 32, 1000]	[B × 50, 32, 1000]	32	32	3	-
10		MaxPool	[B × 50, 32, 1000]	[B × 50, 32, 1]	32	32	-	-
11	Acoustic Emission Branch	Conv1d Block	[B × 50, 1, 1000]	[B × 50, 32, 1000]	1	32	3	1
12		MaxPool	[B × 50, 32, 1000]	[B × 50, 32, 1]	32	32	-	-
13	Segment Feature Fusion	Concatenation	[B × 50, 32, 1] × 3	[B × 50, 96, 1]	96	96	-	-
14		Conv1d Block #1	[B × 50, 96, 1]	[B × 50, 32, 1]	96	32	3	1
15		Channel Attention #1	[B × 50, 32, 1]	[B × 50, 32, 1]	32	32	3	-
16		Conv1d Block #2	[B × 50, 32, 1]	[B × 50, 32, 1]	32	32	3	1
17		Channel Attention #2	[B × 50, 32, 1]	[B × 50, 32, 1]	32	32	3	-
18		Reshape to Segments	[B × 50, 32, 1]	[B, 32, 50]	32	32	-	-
19	Segment Temporal Modeling	ResidualDCBlock (Dilation = 1)	[B, 32, 50]	[B, 32, 50]	32	32	3	1
20		ResidualDCBlock (Dilation = 2)	[B, 32, 50]	[B, 32, 50]	32	32	3	2
21		ResidualDCBlock (Dilation = 4)	[B, 32, 50]	[B, 32, 50]	32	32	3	4
22		MaxPool	[B, 32, 50]	[B, 32, 1]	32	32	-	-
23		Fully Connected	[B, 32]	[B, 3]	32	3	-	-

Note: B: batch size—20; N: signal length—50.

To determine the filtering bands for force and vibration signals, spectral analyses were performed on the three-axis force signals under different wear states, as shown in Figure 11. The effective components of the x- and y-axis forces are mainly concentrated in the low-frequency range and share similar distributions. Therefore, only the x-axis and z-axis spectra are plotted for analysis. As shown in Figure 11a,c,e, the x-axis force signal exhibits pronounced peaks at 150–200 Hz, 500–550 Hz, 1050–1100 Hz, and 1500–1550 Hz, with significant variations in peak amplitudes across different wear states. As illustrated in Figure 11b,d,f, the z-axis force signal shows a prominent peak in the 500–650 Hz range, while minor peak variations are still present around 1500–1550 Hz across different wear stages. During milling, tool wear induces changes in the force signal at the tooth passing frequency and its harmonics, which typically

fall within the low-frequency band. Given that both x-axis and z-axis force signals contain effective components in the 1500–1550 Hz frequency range, and this band may contain feature information related to wear states, the low-pass cutoff frequency for the x-, y-, and z-axis force signals is uniformly set to 1600 Hz to ensure complete retention of these effective signal components while suppressing noise in higher frequency bands.

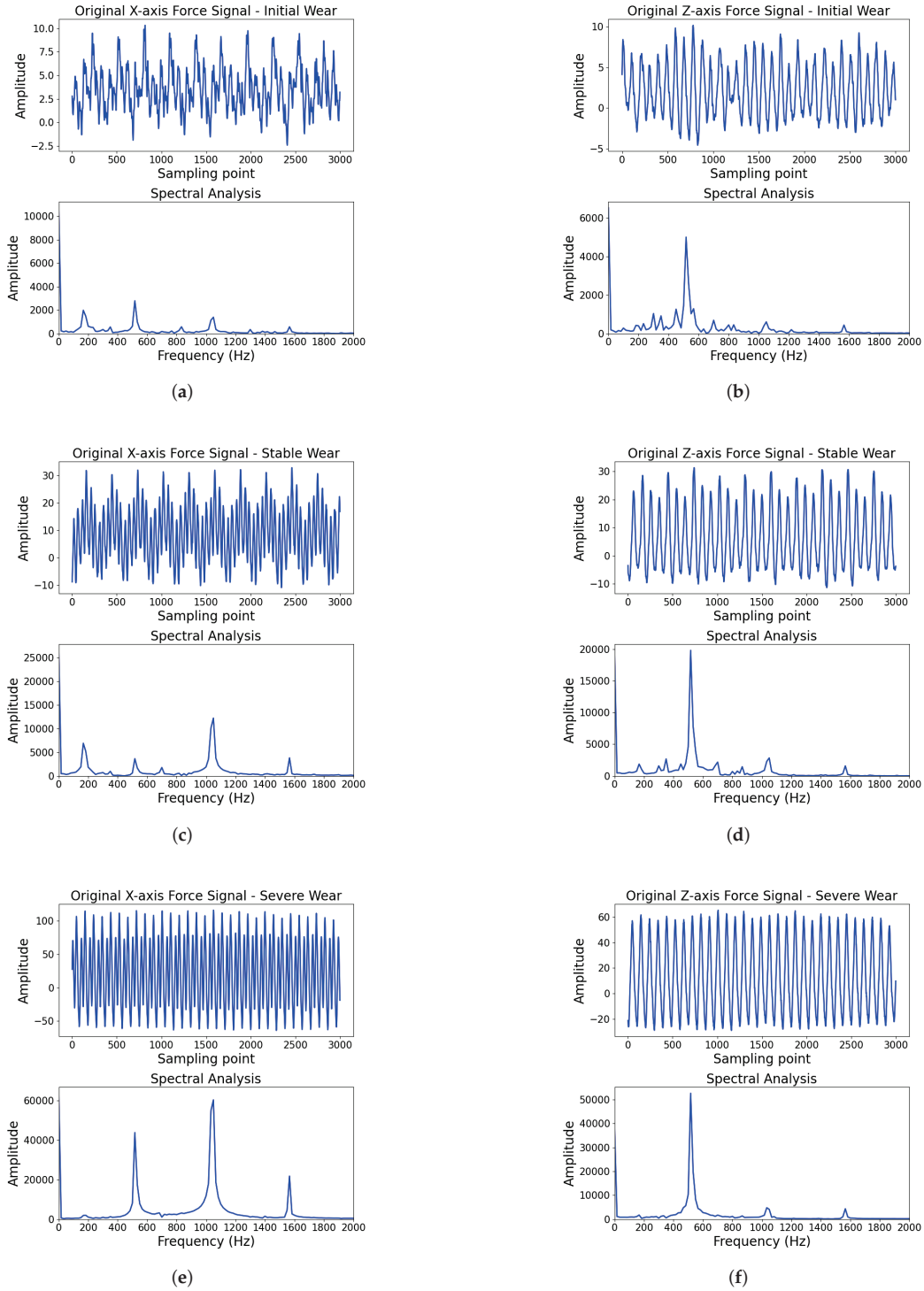
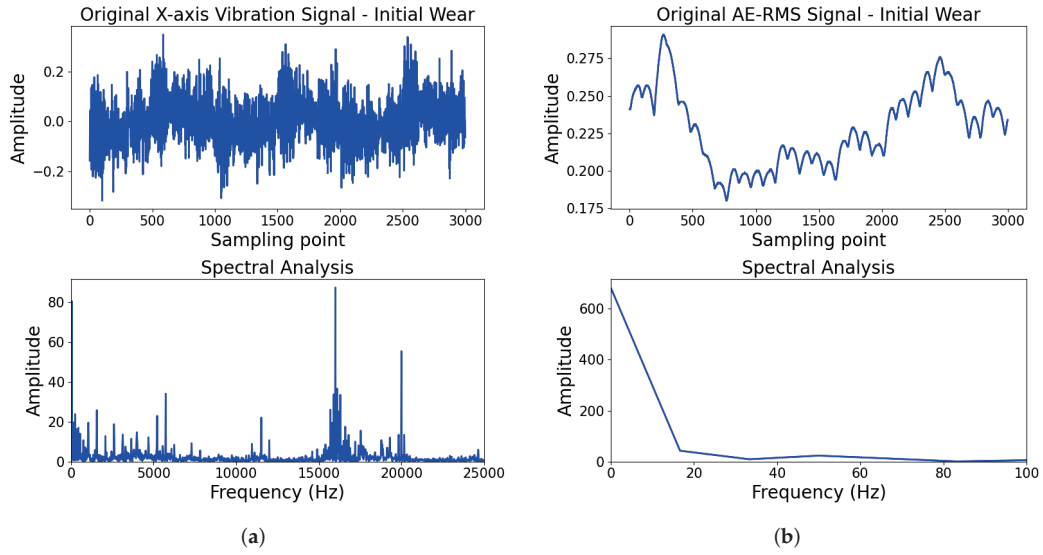


Figure 11. The filtering analysis of force signal. Panels (**a**,**c**,**e**) show x-axis force spectral analysis for initial, stable, and severe wear, respectively, while (**b**,**d**,**f**) show the corresponding z-axis force spectral analysis.

Similarly, spectral analyses of the three-axis vibration signals under different wear states show comparable distributions across x, y, and z axes, as shown in Figure 12. In the initial wear state, as shown in Figure 12a, the x-axis vibration exhibits prominent peaks at 2000 Hz and 4000 Hz, with effective components primarily below 5000 Hz. In the stable wear state, as illustrated in of Figure 12c, major peaks cluster near 2000–2500 Hz and 4000–4500 Hz, with effective components similarly concentrated below 5000 Hz. In the severe wear state, as shown in Figure 12e, the x-axis vibration signal exhibits the most prominent peak around 150–200 Hz, with marked peaks also appearing at 500–550 Hz, 1050–1100 Hz, and 1500–1550 Hz, with energy primarily concentrated below 2000 Hz. Although energy is present in the 15,000–20,000 Hz range, its pattern shows no significant differences across wear states and likely originates from machine tool structural vibrations or electromagnetic interference, contributing little to wear identification. During milling, tool wear-related vibration characteristics are primarily reflected at the tooth passing frequency and its harmonics, which typically fall within the mid-to-low frequency range. Given that signals below 10 Hz typically contain environmental noise, DC drift, and low-frequency machine vibrations unrelated to tool wear, a band-pass filter range of 10–5000 Hz is applied to the x-, y-, and z-axis vibration signals to retain effective frequency components related to wear states while suppressing low-frequency drift and high-frequency noise, thereby improving the signal-to-noise ratio and the reliability of feature extraction. In contrast, AE signals exhibit energy-type waveforms in the time domain with small amplitudes and relatively smooth variations, showing no significant low-frequency drift or broadband high-frequency noise, as illustrated of Figure 12b,d,f. Their temporal evolution primarily reflects the cumulative energy from microcrack initiation and friction processes at the tool-workpiece contact interface, rather than sharp transient pulses. Therefore, no additional filtering is applied to the AE channel; only normalization is performed, and the same time window strategy used for other sensors is employed for segmentation. The Bi-TCN convolutional module and channel attention automatically suppress residual noise and extract wear-sensitive AE energy evolution features in a data-driven manner.



(**a**) (**b**)

Figure 12. *Cont.*

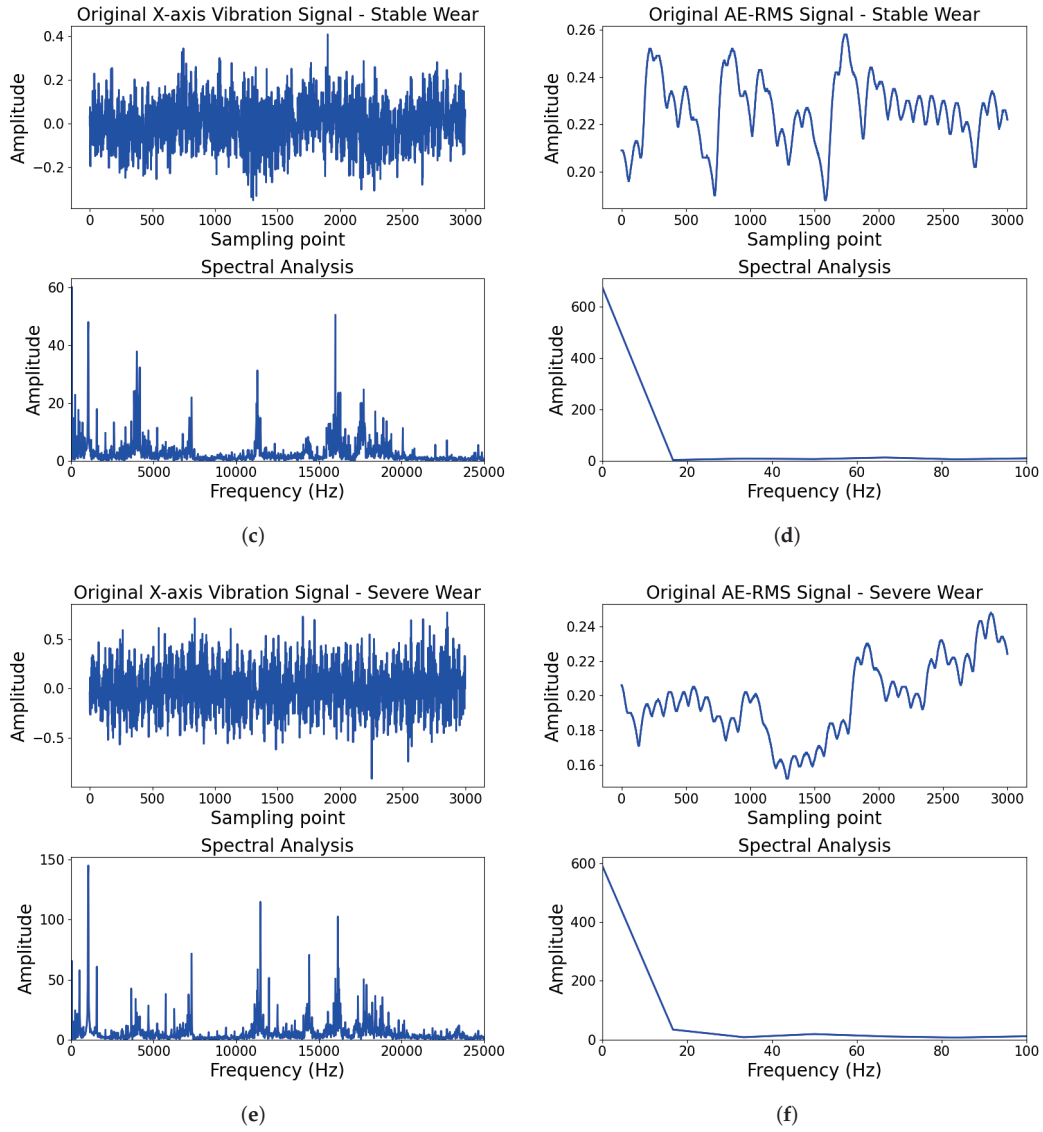


Figure 12. The filtering analysis of vibration and AE-RMS signal. Panels (**a**,**c**,**e**) show x-axis vibration spectral analysis for initial, stable, and severe wear, respectively, while (**b**,**d**,**f**) show the AE-RMS spectral analysis.

After multiple experimental verifications, it was found that dividing the 50,000 sensor data into 50 segments, each with 1000 data points for feature extraction, best reflects the tool wear evolution process at the current stage.

5.3. Results Discussion

The training host uses the Win10 operating system, and the training environment is based on PyTorch 12.1. The model was trained on an NVIDIA RTX4050 GPU, with CUDA version 11.8. The training loss adopts cross-entropy loss, and the Adam optimizer is used for model parameter optimization. The optimizer learning rate is set to 0.001, the weight decay is set to 0.00001, and the dropout is set to 0.2 to reduce network overfitting. A mini-batch training strategy is adopted, with the mini-batch size set to 20. The loss curves and accuracy curves during the training process are shown in Figure 13.

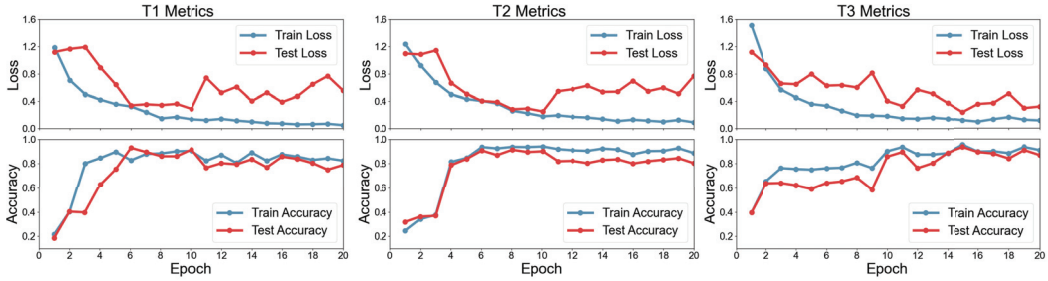


Figure 13. Loss and accuracy curves for training and testing on T dataset.

On the T1 dataset, the training and testing losses decreased synchronously in the early stages, and the model exhibited strong generalization ability. However, after epoch 7, the testing curve fluctuated, indicating that the model began to partially memorize detailed features of the training data, but still retained a certain degree of generalization. By epoch 6, the model learned more generalizable feature representations through parameter fine-tuning, resulting in a peak test accuracy of 93.02%, at which point the generalization ability was optimal.

On the T2 dataset, due to the significant differences in wear characteristics between C1 and C6 at different stages, the training accuracy increased rapidly from 24.44% to 94.13% in the first 10 epochs, while the test accuracy rose from 31.75% to 90.16%, with both metrics increasing synchronously, demonstrating the model's effective learning of features from different tool wear states. The 8th epoch achieved the best test accuracy of 91.43%.

Similarly, on the T3 dataset, due to the significant differences in wear characteristics between C1 and C6 at different stages, the training accuracy increased rapidly from 39.68% to 90.16% in the first 10 epochs, while the test accuracy rose from 39.68% to 85.71%, with both metrics increasing synchronously, demonstrating the model's effective learning of features from different tool wear states. The 15th epoch achieved the best test accuracy of 93.65%. All experiments were performed on GPU, with a total training time of approximately 42 min for 20 epochs and a peak GPU memory usage of 1.6 GB. To evaluate the real-time performance of the model, we measured the inference time on the test sets. The model achieved an average per-sample inference time of 4 ms and a throughput of 270 samples per second, demonstrating that the model can meet the requirements for real-time monitoring applications. In addition, we compared the proposed model with some classical model methods, and the comparison results are shown in Table 4.

Table 4. Comparison of the accuracy of the experimental results.

Models	T1 (%)	T2 (%)	T3 (%)	Mean ± Std (%)
1D-CNN with DGCCA [39]	88.54	90.80	90.80	90.05 ± 1.30
CaAt-ResNet-1d [40]	86.54	90.23	89.50	88.76 ± 1.95
Informer [48]	90.02	89.92	91.52	90.49 ± 0.89
CNN-Transformer [49]	89.52	86.54	90.04	88.70 ± 1.89
Transformer [50]	90.02	85.06	89.52	88.20 ± 2.73
GRU [51]	83.63	85.72	84.81	84.72 ± 1.05
LSTM [52]	83.52	82.49	81.62	82.54 ± 0.95
GHCRBM [53]	87.81	90.92	62.23	80.32 ± 15.75
Proposed model	**93.02**	**91.43**	**93.65**	**92.70 ± 0.93**

The experimental results show that the proposed model demonstrates superior performance on the three cross-validation datasets. Compared with traditional recurrent neural networks such as GRU and LSTM, the proposed model achieves significantly higher classification accuracy on all test datasets. Compared with 1D-CNN-based methods, the proposed model also shows a clear advantage. Compared with 1D-CNN with DGCCA,

the proposed model improves the accuracy by 4.48%, 0.63%, and 2.85% on the T1, T2, and T3 datasets, respectively. Compared with CaAt-ResNet-1d, the proposed model achieves accuracy improvements of 6.48%, 1.20%, and 4.15% on the three datasets, respectively. Compared with the latest Transformer architectures, the proposed model improves the accuracy by 3.00% on the T1 dataset, and by 6.37% and 4.13% on the T2 and T3 datasets, respectively. Compared with CNN-Transformer, the proposed model achieves accuracy improvements of 3.50%, 4.89%, and 3.61% on the three datasets, respectively. Compared with Informer, the proposed model achieves accuracy improvements of 3.00%, 1.51% and 2.13% on the T1, T2 and T3 datasets, respectively, and achieves a higher average accuracy. Considering the average accuracy across the three datasets, the proposed model achieves the best accuracy of 92.70% with a standard deviation of 0.93%, verifying its effectiveness in the tool wear classification task.

To verify the contribution of each module to the model performance, four models were designed for ablation experiments. Model 1 removed the channel attention module; Model 2 replaced the Bi-TCN with TCN while retaining the channel attention module; Model 3 replaced Bi-TCN with TCN and simultaneously removed the channel attention module; The Proposed model represents the complete architecture incorporating all designed modules.

The experimental evaluation was conducted on three datasets, and the results are summarized in Table 5. Model 1 achieved an average accuracy of 90.26% ± 0.65%, which is 2.44% lower than the complete model. Model 2 achieved an average accuracy of 90.74% ± 1.77%, which is 1.96% lower than the complete model. Model 3 achieved an average accuracy of 85.50% ± 0.54%, which is 7.20% lower than the complete model. The poor performance is primarily attributed to the model's lack of adaptive inter-channel weight learning and bidirectional temporal modeling capabilities, resulting in degraded multi-sensor feature fusion and temporal information extraction. The complete model achieved the best performance with an average accuracy of 92.70% ± 0.93%.

Table 5. Ablation experiment results.

Model Description	T1 (%)	T2 (%)	T3 (%)	Mean ± Std (%)
Model 1 (without channel attention)	89.52	90.16	91.11	90.26 ± 0.65
Model 2 (TCN instead Bi-TCN with channel attention)	91.75	92.23	88.25	90.74 ± 1.77
Model 3 (TCN instead Bi-TCN without channel attention)	86.03	84.76	85.71	85.50 ± 0.54
Proposed model	**93.02**	**91.43**	**93.65**	**92.70 ± 0.93**

The results indicate that both channel attention and Bi-TCN contribute significantly to the model performance. Removing channel attention resulted in a 2.44% accuracy drop; replacing Bi-TCN with TCN resulted in a 1.96% accuracy drop; simultaneously removing both modules caused a 7.20% accuracy drop, demonstrating that both channel attention and Bi-TCN contribute to the performance.

To further explore the performance of the model, four evaluation metrics of the confusion matrix are used for quantitative analysis: accuracy, precision, recall, and specificity. Accuracy represents the overall classification ability of the model in all tool wear stages, precision reflects the correctness of recognition for each class, recall measures the ability to identify the positive class, and specificity quantifies the misclassification rate. In tool wear monitoring, precision should be prioritized in the initial wear and stable wear stages to avoid unnecessary tool replacement, while in the severe wear stage, more attention is paid to recall and specificity to ensure timely detection of potential failures and prevent a decline in machining quality. The confusion matrices for each T dataset are shown in

Figure 14, where the accuracy for all stages, precision for the initial and stable wear stages, recall and specificity for the severe wear stage are shown in Table 6.

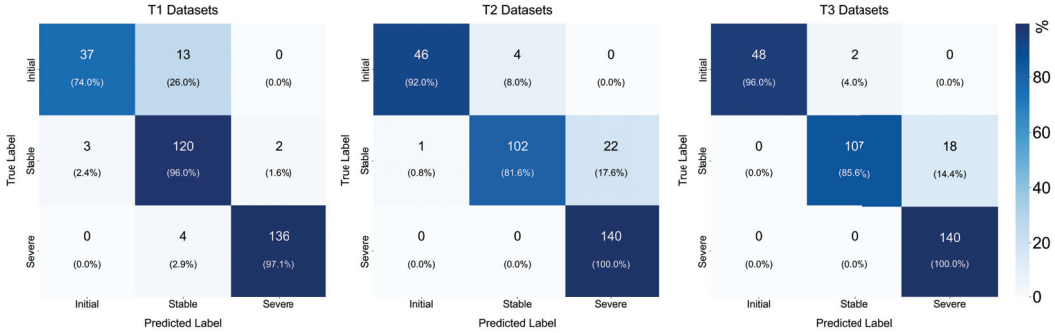


Figure 14. Confusion matrix of experimental results on T dataset.

Table 6. Evaluation indicators on the T dataset.

Indicators	Stage	T1	T2	T3	Average
Accuracy	-	0.9302	0.9143	0.9365	0.9270
Precision	Initial	0.9250	0.9787	1.0000	0.9679
	Stable	0.8759	0.9623	0.9817	0.9399
Recall	Severe	0.9714	1.0000	1.0000	0.9905
Specificity	Severe	0.9886	0.8743	0.8971	0.9200

It can be seen that in the initial and stable wear stages, the average precision of the three datasets reached 96.8% and 94.0%, respectively, while in the severe wear stage, the average recall and specificity reached 99.1% and 92.0%, respectively, indicating that the model is highly effective in identifying the severe tool wear stage. Through cross-dataset comparative experiments and validation with confusion matrices, the proposed model demonstrates strong feature representation capability and generalization adaptability, accurately identifying tool wear evolution patterns. With a lightweight architecture, it achieves real-time, robust, and high-precision condition monitoring without introducing the frequency domain.

6. Conclusions

This work introduces a hierarchical multi-scale dilated temporal convolutional network (Hi-MDTCN) for real-time tool wear monitoring from multi-sensor time series. Within segments, a lightweight channel attention module adaptively amplifies key sensor features; across segments, a bidirectional TCN captures both short- and long-range temporal dependencies, balancing speed and accuracy. Using Butterworth-filtered time-domain inputs only, the approach eliminates extra frequency or time–frequency transforms and reduces preprocessing latency. Experiments show that Hi-MDTCN outperforms baselines in both real-time performance and wear-detection accuracy, offering a reliable solution for tool condition monitoring and intelligent manufacturing.

In future work, we will improve the generalization of the model in diverse tool types and challenging machining conditions, with a focus on wet milling to better reflect industrial practice. Because coolant is commonly used to improve cutting stability, we plan to collect milling data with coolant and, where appropriate, apply transfer learning and domain adaptation to validate robustness in representative shop-floor settings. We will also systematically assess the model's ability to recognize tool-wear states under varying sampling rates, providing guidance for optimizing sensing and sampling strategies.

Author Contributions: Conceptualization, A.C. and M.L.; methodology, A.C. and Z.F.; software, P.H. and C.G.; validation, C.G. and L.W.; formal analysis, L.W. and Z.F.; investigation, W.Y. and Z.F.; writing—original draft preparation, Z.F. and A.C.; writing—review and editing, M.L. and A.C.; supervision, E.H. and S.L.; project administration, M.L.; funding acquisition, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Joint Funds of the Natural Science Foundation of Liaoning Province of China (Grant No. 2023-MSLH-248); the Scientific Research Project of Liaoning Provincial Education Department of China (Grant No. LJ232410142006); Fujian Provincial Natural Science Foundation Project (Grant No.2022J05248); Industry-University-Research Project of Longyan Nonferrous Metals Research Institute (Grant No.PT2024-2-02); and the University-level Doctoral Start-up Project (Grant No. LB2021011).

Data Availability Statement: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

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

Appendix A

Algorithm A1 The Proposed Framework for Hi-MDTCN

Require: Raw multi-sensor signal $X \in \mathbb{R}^{C \times T}$
Ensure: Predicted tool wear class $\hat{y}$

Smooth force signals using sliding window averaging: $f_{smooth}[i] = \frac{1}{W} \sum_{j=i-\frac{W-1}{2}}^{i+\frac{W-1}{2}} f_{raw}[j]$
Calculate first-order difference features: $diff[i] = f_{smooth}[i+1] - f_{smooth}[i]$
Compute adaptive threshold $\theta = \bar{d} + k \cdot \sigma_{|d|}$
Detect tool retraction points when continuous $|diff[i]| > \theta$ with matching sign
Extract original signal segments closest to the identified tool retraction points
Apply 4th-order Butterworth filters with specific cutoff frequencies for different signals
Divide signals into N equal-length segments $\{X_1, X_2, \ldots, X_N\}$ for multi-scale analysis
for $i = 1$ to N **do**
    Extract $X_i^{(dyn)}$, $X_i^{(acc)}$, $X_i^{(ae)}$ from segment X_i
    $h_i^{(dyn)} \leftarrow$ Conv1D + ChannelAttention + MaxPooling on $X_i^{(dyn)}$
    $h_i^{(acc)} \leftarrow$ Conv1D + ChannelAttention + MaxPooling on $X_i^{(acc)}$
    $h_i^{(ae)} \leftarrow$ Conv1D + MaxPooling on $X_i^{(ae)}$
    $h_i \leftarrow \text{Concat}(h_i^{(dyn)}, h_i^{(acc)}, h_i^{(ae)})$
    $f_i \leftarrow$ Fusion Conv1D + ChannelAttention on h_i
end for
$F \leftarrow [f_1, f_2, \ldots, f_N]$
Apply Bi-TCN on F
$z \leftarrow$ GlobalMaxPooling over temporal dimension
$\hat{y} \leftarrow$ FullyConnected(z)
return $\hat{y}$

References

1. Chen, M.; Mao, J.; Fu, Y.; Liu, X.; Zhou, Y.; Sun, W. In-situ tool wear condition monitoring during the end milling process based on dynamic mode and abnormal evaluation. *Sci. Rep.* **2024**, *14*, 12888. [CrossRef]
2. Pimenov, D.Y.; Gupta, M.K.; da Silva, L.R.R.; Kiran, M.; Khanna, N.; Krolczyk, G.M. Application of measurement systems in tool condition monitoring of Milling: A review of measurement science approach. *Measurement* **2022**, *199*, 111503. [CrossRef]
3. Lin, W.-J.; Chen, J.-W.; Jhuang, J.-P.; Tsai, M.-S.; Hung, C.-L.; Li, K.-M.; Young, H.-T. Integrating object detection and image segmentation for detecting the tool wear area on stitched image. *Sci. Rep.* **2021**, *11*, 19938. [CrossRef]
4. Yang, C.; Zhou, J.; Li, E.; Wang, M.; Jin, T. Local-feature and global-dependency based tool wear prediction using deep learning. *Sci. Rep.* **2022**, *12*, 14574. [CrossRef] [PubMed]

5. Wong, S.Y.; Chuah, J.H.; Yap, H.J. Technical data-driven tool condition monitoring challenges for CNC milling: A review. *Int. J. Adv. Manuf. Technol.* **2020**, *107*, 4837–4857. [CrossRef]
6. Li, S.; Li, J.; Zhu, K. Application of physics-guided deep learning model in tool wear monitoring of high-speed milling. *Mech. Syst. Signal Process.* **2025**, *224*, 111949. [CrossRef]
7. Kasilingam, S.; Yang, R.; Singh, S.K.; Farahani, M.A.; Rai, R.; Wuest, T. Physics-based and data-driven hybrid modeling in manufacturing: A review. *Prod. Manuf. Res.* **2024**, *12*, 2305358. [CrossRef]
8. Xue, Z.; Chen, N.; Wu, Y.; Yang, Y.; Li, L. Hierarchical temporal transformer network for tool wear state recognition. *Adv. Eng. Inform.* **2023**, *58*, 102218. [CrossRef]
9. Ntemi, M.; Paraschos, S.; Karakostas, A.; Gialampoukidis, I.; Vrochidis, S.; Kompatsiaris, I. Infrastructure monitoring and quality diagnosis in CNC machining: A review. *Cirp J. Manuf. Sci. Technol.* **2022**, *38*, 631–649. [CrossRef]
10. Zhou, Y.; Xue, W. Review of tool condition monitoring methods in milling processes. *Int. J. Adv. Manuf. Technol.* **2018**, *96*, 2509–2523. [CrossRef]
11. Huang, Z.; Shao, J.; Guo, W.; Li, W.; Zhu, J.; Fang, D. Hybrid machine learning-enabled multi-information fusion for indirect measurement of tool flank wear in milling. *Measurement* **2023**, *206*, 112255. [CrossRef]
12. Mohamed, A.; Hassan, M.; M'Saoubi, R.; Attia, H. Tool Condition Monitoring for High-Performance Machining Systems—A Review. *Sensors* **2022**, *22*, 2206. [CrossRef]
13. Khadka, S.; Rahman Rashid, R.A.; Stephens, G.; Papageorgiou, A.; Navarro-Devia, J.; Hägglund, S.; Palanisamy, S. Predicting cutting tool life: Models, modelling, and monitoring. *Int. J. Adv. Manuf. Technol.* **2025**, *136*, 3037–3076. [CrossRef]
14. Ma, Y.; Yao, C.; Tan, L.; Cui, M.; Ren, J.; Luo, M.; Ding, W.; Wang, J.; Zhao, J. Research progress on intelligent monitoring of machining condition based on indirect method. *Adv. Eng. Inform.* **2025**, *67*, 103518.
15. Xue, Z.; Li, L.; Wu, Y.; Yang, Y.; Wu, W.; Zou, Y.; Chen, N. Study on tool wear state recognition algorithm based on spindle vibration signals collected by homemade tool condition monitoring ring. *Measurement* **2023**, *223*, 113787. [CrossRef]
16. Navarro-Devia, J.H.; Chen, Y.; Dao, D.V.; Li, H. Chatter detection in milling processes—A review on signal processing and condition classification. *Int. J. Adv. Manuf. Technol.* **2023**, *125*, 3943–3980. [CrossRef]
17. Kuntoğlu, M.; Salur, E.; Gupta, M.K.; Sarıkaya, M.; Pimenov, D.Y. A state-of-the-art review on sensors and signal processing systems in mechanical machining processes. *Int. J. Adv. Manuf. Technol.* **2021**, *116*, 2711–2735. [CrossRef]
18. Wang, S.; Tian, Y.; Hu, X.; Wang, J.; Han, J.; Liu, Y.; Wang, J.; Wang, D. Identification of grinding wheel wear states using AE monitoring and HHT-RF method. *Wear* **2025**, *562*, 205668. [CrossRef]
19. Tran, M.-Q.; Doan, H.-P.; Vu, V.Q.; Vu, L.T. Machine learning and IoT-based approach for tool condition monitoring: A review and future prospects. *Measurement* **2023**, *207*, 112351. [CrossRef]
20. Wang, J.; Tian, Y.; Hu, X.; Han, J.; Liu, B. Integrated assessment and optimization of dual environment and production drivers in grinding. *Energy* **2023**, *272*, 127046. [CrossRef]
21. Liang, Y.; Hu, S.; Guo, W.; Tang, H. Abrasive tool wear prediction based on an improved hybrid difference grey wolf algorithm for optimizing SVM. *Measurement* **2022**, *187*, 110247. [CrossRef]
22. Cardoz, B.; Shaikh, H.N.E.; Mulani, S.M.; Kumar, A.; Rajasekharan, S.G. Random forests based classification of tool wear using vibration signals and wear area estimation from tool image data. *Int. J. Adv. Manuf. Technol.* **2023**, *126*, 3069–3081. [CrossRef]
23. Hu, H.; Qin, C.; Guan, F.; Su, H. A Tool Wear Monitoring Method Based on WOA and KNN for Small-Deep Hole Drilling. In Proceedings of the 2021 International Symposium on Computer Technology and Information Science (ISCTIS), Guilin, China, 4–6 June 2021; pp. 284–287.
24. Nasir, V.; Sassani, F. A review on deep learning in machining and tool monitoring: Methods, opportunities, and challenges. *Int. J. Adv. Manuf. Technol.* **2021**, *115*, 2683–2709. [CrossRef]
25. Patil, S.S.; Pardeshi, S.S.; Patange, A.D.; Jegadeeshwaran, R. Deep Learning Algorithms for Tool Condition Monitoring in Milling: A Review. *J. Physics Conf. Ser.* **2021**, *1969*, 012039. [CrossRef]
26. Liu, T.; Zhu, K.; Wang, G. Micro-milling tool wear monitoring under variable cutting parameters and runout using fast cutting force coefficient identification method. *Int. J. Adv. Manuf. Technol.* **2020**, *111*, 3175–3188. [CrossRef]
27. Huang, Z.; Zhu, J.; Lei, J.; Li, X.; Tian, F. Tool wear monitoring with vibration signals based on short-time Fourier transform and deep convolutional neural network in milling. *Math. Probl. Eng.* **2021**, *2021*, 9976939. [CrossRef]
28. Zhang, C.; Wang, J.; Cao, Y.; Jiao, F. Tool wear status monitoring under laser-ultrasonic compound cutting based on acoustic emission and deep learning. *J. Mech. Sci. Technol.* **2024**, *38*, 2411–2421. [CrossRef]
29. Bagga, P.J.; Chavda, B.; Modi, V.; Makhesana, M.A.; Patel, K.M. Indirect tool wear measurement and prediction using multi-sensor data fusion and neural network during machining. *Mater. Today Proc.* **2022**, *56*, 51–55. [CrossRef]
30. Jones, T.; Cao, Y. Tool wear prediction based on multisensor data fusion and machine learning. *Int. J. Adv. Manuf. Technol.* **2025**, *137*, 5213–5225. [CrossRef]

31. Ou, J.; Li, H.; Liu, B.; Peng, D. Deep transfer residual variational autoencoder with multi-sensors fusion for tool condition monitoring in impeller machining. *Measurement* **2022**, *204*, 112028. [CrossRef]
32. Qin, Y.; Liu, X.; Yue, C.; Zhao, M.; Wei, X.; Wang, L. Tool wear identification and prediction method based on stack sparse self-coding network. *J. Manuf. Syst.* **2023**, *68*, 72–84. [CrossRef]
33. Song, G.; Zhang, J.; Zhu, K.; Ge, Y.; Yu, L.; Fu, Z. Tool wear monitoring based on multi-kernel Gaussian process regression and Stacked Multilayer Denoising AutoEncoders. *Mech. Syst. Signal Process.* **2023**, *186*, 109851. [CrossRef]
34. Sayyad, S.; Kumar, S.; Bongale, A.; Kotecha, K.; Selvachandran, G.; Suganthan, P. N. Tool wear prediction using long short-term memory variants and hybrid feature selection techniques. *Int. J. Adv. Manuf. Technol.* **2022**, *121*, 6611–6633. [CrossRef]
35. Zhang, K.; Zhou, D.; Zhou, C.; Hu, B.; Li, G.; Liu, X.; Guo, K. Tool wear monitoring using a novel parallel BiLSTM model with multi-domain features for robotic milling Al7050-T7451 workpiece. *Int. J. Adv. Manuf. Technol.* **2023**, *129*, 1883–1899. [CrossRef]
36. Li, Z.; Xiao, X.; Zhou, W.; Zhang, K.; Fu, H. Identification of tool wear status using multi-sensor signals and improved gated recurrent unit. *Int. J. Adv. Manuf. Technol.* **2025**, *137*, 1249-1260. [CrossRef]
37. Qin, Y.; Liu, X.; Wei, X.; Guo, B.; Jia, R. A variable modal decomposition and BiGRU-based tool wear monitoring method based on non-dominated sorting genetic algorithm II. *Int. J. Comput. Integr. Manuf.* **2025**, *38*, 715–731. [CrossRef]
38. Cao, Y.; Ding, Y.; Jia, M.; Tian, R. A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings. *Reliab. Eng. Syst. Saf.* **2021**, *215*, 107813. [CrossRef]
39. Yin, Y.; Wang, S.; Zhou, J. Multisensor-based tool wear diagnosis using 1D-CNN and DGCCA. *Appl. Intell.* **2023**, *53*, 4448–4461. [CrossRef]
40. Dong, L.; Wang, C.; Yang, G.; Huang, Z.; Zhang, Z.; Li, C. An Improved ResNet-1d with Channel Attention for Tool Wear Monitor in Smart Manufacturing. *Sensors* **2023**, *23*, 1240. [CrossRef]
41. Yang, X.; Yuan, R.; Lv, Y.; Li, L.; Song, H. A Novel Multivariate Cutting Force-Based Tool Wear Monitoring Method Using One-Dimensional Convolutional Neural Network. *Sensors* **2022**, *22*, 8343. [CrossRef] [PubMed]
42. Zhang, Y.; Qi, X.; Wang, T.; He, Y. Tool Wear Condition Monitoring Method Based on Deep Learning with Force Signals. *Sensors* **2023**, *23*, 4595. [CrossRef]
43. Li, Z.; Liu, X.; Incecik, A.; Gupta, M.K.; Królczyk, G.M.; Gardoni, P. A novel ensemble deep learning model for cutting tool wear monitoring using audio sensors. *J. Manuf. Processes* **2022**, *79*, 233–249. [CrossRef]
44. Tran, M.-Q.; Liu, M.-K.; Tran, Q.-V. Milling chatter detection using scalogram and deep convolutional neural network. *Int. J. Adv. Manuf. Technol.* **2020**, *107*, 1505–1516. [CrossRef]
45. Li, Y.; Zou, S.; Zhao, T.; Su, X. MDFA-Net: Multi-Scale Differential Feature Self-Attention Network for Building Change Detection in Remote Sensing Images. *Remote Sens.* **2024**, *16*, 3466. [CrossRef]
46. Wang, Q.; Wang, H.; Hou, L.; Yi, S. Overview of Tool Wear Monitoring Methods Based on Convolutional Neural Network. *Appl. Sci.* **2021**, *11*, 12041. [CrossRef]
47. Wang, P.; Liu, Z.; Gao, R.X.; Guo, Y. Heterogeneous data-driven hybrid machine learning for tool condition prognosis. *CIRP Ann.* **2019**, *68*, 455–458. [CrossRef]
48. Li, W.; Fu, H.; Han, Z.; Zhang, X.; Jin, H. Intelligent tool wear prediction based on Informer encoder and stacked bidirectional gated recurrent unit. *Robot. Comput. Integr. Manuf.* **2022**, *77*, 102368. [CrossRef]
49. Liu, H.; Liu, Z.; Jia, W.; Zhang, D.; Wang, Q.; Tan, J. Tool wear estimation using a CNN-transformer model with semi-supervised learning. *Meas. Sci. Technol.* **2021**, *32*, 125010. [CrossRef]
50. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
51. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv*, **2014**, arXiv:1412.3555. [CrossRef]
52. Graves, A. *Long Short-Term Memory*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 37–45.
53. Li, G.; Wang, Y.; He, J.; Hao, Q.; Yang, H.; Wei, J. Tool wear state recognition based on gradient boosting decision tree and hybrid classification RBM. *Int. J. Adv. Manuf. Technol.* **2020**, *110*, 511–522. [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.

sensors

MDPI

Article

Correlation Between Meso-Defect and Fatigue Life Through Representing Feature Analysis for 6061-T6 Aluminum Alloys

Liangxia Zhang [1], Yali Yang [2], Hao Chen [2,*] and Shusheng Lv [3]

[1] School of Mechanical Engineering, Ningxia Institute of Science and Technology, Shizuishan 753000, China; zlx0524@126.com
[2] School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China; carolyn71@163.com
[3] Chinese Academy of Agricultural Mechanization Sciences, Beijing 100083, China; lsscnq@163.com
* Correspondence: pschenhao@sues.edu.cn; Tel.: +86-13585901312

Abstract

Fatigue strength is vital for engineering applications of aluminum alloys. Accurate models incorporating mesoscopic defect-representing features are one of the issues for accurate fatigue strength prediction. A fatigue life prediction method based on meso-defect-representing features is proposed in this study. Based on staged fatigue damage, meso-defect data was obtained by X-ray CT. After 3D reconstruction and simplification, porosity, shape, and location were selected as the meso-defect-representing features using correlation coefficient analysis. Weights of meso-defect features were determined through FEM simulation. A mesoscopic damage variable incorporating the weights of porosity, shape, and location for meso-defect was defined. Correlation between fatigue life and meso-defect features was established through the mesoscopic damage variable. Experimental verification results showed that the prediction method is an effective method for fatigue life assessment.

Keywords: meso-defect; representing feature; relative weights; mesoscopic damage variable; fatigue life

1. Introduction

Aluminum alloy is widely used in aerospace and vehicle manufacturing due to its good casting properties and light weight. In the process of engineering application, a large number of accidents are related to the defects caused by the fatigue failure of materials. Defects are considered to be one of the main factors that seriously affect the mechanical properties and fatigue properties of materials [1–5].

For damage defect-based fatigue life prediction, efforts have been made to characterize material damage. One of the common approaches is the damage factor (D) based on continuous damage mechanics [6–8]; for non-destructive material, $D = 1$, and for complete loss of load-bearing capacity, $D = 1$. Chaboche proposed the concept of effective stress to establish a one-dimensional nonlinear damage accumulation model, in which the damage variable (D) is a function of the stress state and the current damage state [9]. Lemaitre proposed a fatigue damage evolution model based on cumulative plastic strain [10], which has been continuously improved, and different forms of models have been established for high- and low-cycle fatigue. Subsequent damage accumulation models are basically based on the damage evolution model of Chaboche and Lemaitre.

Meso-defects are the critical factors in determining the fatigue life of aluminum alloys. The characterization of meso-defect features, such as quantity, shape, size, and distribution [11,12], is critically important for understanding and predicting fatigue damage evolution [13]. Due to the enormous numbers, irregular shape, and uneven spatial distribution of meso-defects in the material domain, correlative research between defect features and fatigue damage behavior has been conducted, along with the development of continuum damage mechanics theory and high-resolution three-dimensional inspection technology (such as synchrotron radiation X-ray computed tomography) for the internal structure of materials [14–16]. Quantitative methods have been proposed for fatigue life prediction through defect population [17], porosity [18,19], size [20–22], defect shape [23,24] and location [25–27], influence of defect distance to the surface [28,29], combination of density, shape, and size [30], and so on, through numerical methods [31,32], the finite element method [33], and machine learning methods [34,35].

Due to the multi-scale characteristics of meso-defects and also their evolution during the fatigue process, it is vital to identify the contribution of different meso-defect-representing features to fatigue life. However, most studies are based on one or two characteristics, rather than on complex, heterogeneous meso-defects in real aluminum alloys. The critical research gap lies in moving from defect statistics to a physically graded understanding of fatigue damage initiation [36]. Tammas-Willliams et al. (2015) statistically correlated individual pore features with fatigue performance but did not combine them into a mechanistic damage variable [37]. There are few studies that combine these defect-representing features to characterize the degree of material fatigue damage.

Aiming to address the research gap in fatigue life prediction through an integrated framework for meso-defect features, a mesoscopic damage-representing feature-based fatigue strength evaluation method was proposed in this study. Our objective is to capture meso-defects that have developed sufficiently to significantly affect local mechanical responses and can serve as effective indicators of damage in order to establish a physically informed, mesoscale framework for quantifying the influence of meso-defects on fatigue life. After staged fatigue experiments, meso-defect data was obtained, reconstructed, and simplified. Meso-defect-representing features were selected and incorporated into a mesoscopic damage variable using a weighting method. A fatigue life prediction method based on the mesoscopic damage variable was proposed and experimentally verified, thus providing a relatively comprehensive method combing multiple meso-defect features for quantitative analysis for fatigue behavior.

2. Material and Methods

2.1. Material and Specimen

The material of the test piece is 6061-T6 aluminum alloy. The chemical composition is shown in Table 1. The specimens are made in accordance with the requirements of ASTM E8/E8M-15a for fatigue testing. The model and dimensions of the specimens used are shown in Figure 1.

Table 1. Chemical composition of 6061-T6 alloy used in this study (wt.%).

Al	Mg	Fe	Si	Cr	Cu	Zn	Mn	Ti	Other
95.8	0.8~1.2	0.7	0.6	0.3	0.2~0.4	0.25	0.15	0.15	0.15

https://doi.org/10.3390/s26020631

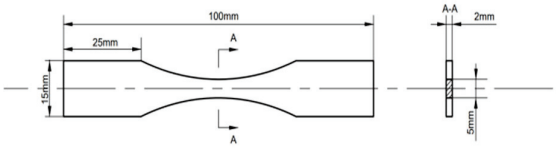


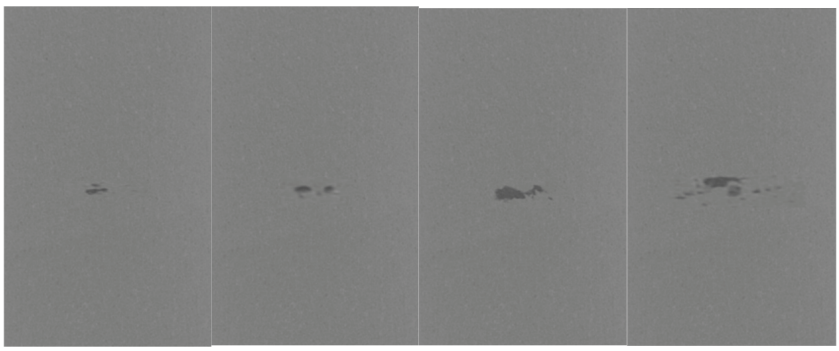
Figure 1. Dimensions of specimens.

2.2. Staged Fatigue Damage Experiment

A tensile fatigue damage experiment was conducted using an MTS electro-hydraulic servo dynamic test system (model 370.10) at room temperature. The test adopted sinusoidal waveform control, and under the axial loading with a frequency of 50 Hz, the stress ratio was set to R = 0.1. Due to the variability of fatigue life testing, the final fracture point from direct physical measurement is not available for all calibration scenarios. A fixed fracture cycle of 100,000 cycles was defined as the rupture point, based on several pretesting, to provide a clear and unambiguous normalization point for comparing damage trajectories across different specimens and conditions in the dataset. Over the entire damage process from zero damage to rupture (100,000 cycles), 5 stages were selected—20,000 cycles, 40,000 cycles, 60,000 cycles, 80,000 cycles, and 100,000 cycles—to determine stage-wise damage behavior. For each stage, 3 replications were obtained.

2.3. Defect Data Acquisition

After the staged fatigue damage experiment, the center sections ($6 \times 5 \times 2$ mm^3) of the specimens were scanned using an X-ray CT system (Pheonix v | tome | xm, GE, Tempe, AZ, USA) to obtain internal meso-defect information, with a resolution of 5 μm. For each specimen, non-destructive internal testing using the CT system scanned 1600 images over 360°. Figure 2 shows four typical Tif slices for different load stages.



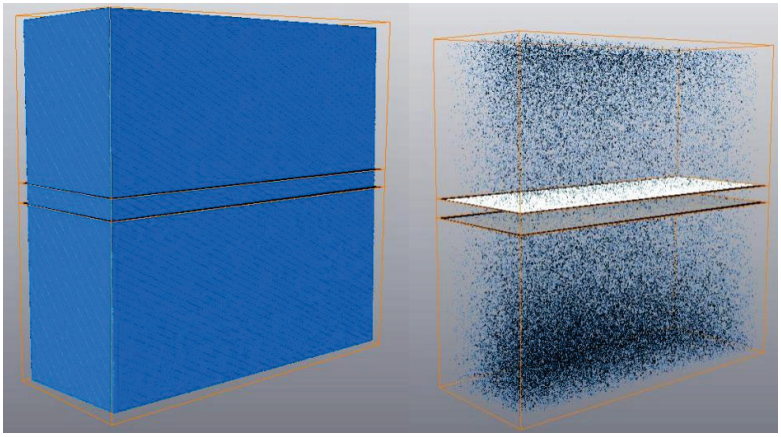
(**a**) 20,000 cycles (**b**) 40,000 cycles (**c**) 60,000 cycles (**d**) 80,000 cycles

Figure 2. Tif diagrams for different loading stage specimens.

To obtain 3D information of meso-defect, AVIZO (2016) software was used to visualize and reconstruct damage data from X-ray CT. Two-dimensional image data were transferred into three-dimensional damage data. Defect data were reconstructed in three dimensions through positioning and display, filtering, defect segmentation, and recognition, as shown in Figure 3.

Then, damage parameters of specimens at different stages were obtained by the Label Analysis module in AVIZO. These parameters included damage characteristic information such as porosity, shape, size, and location of the defects.

https://doi.org/10.3390/s26020631



(**a**) Matrix material model (**b**) Meso-defect model in AVIZO

Figure 3. Visualization of meso-defects.

2.4. Defect Shape Simplification

The shape of a defect can be defined by its sphericity and aspect ratio as representations of shape irregularity and elongation [38]. Generally, the detected defects had irregular shapes with sphericities much less than unity, and both sphericity and aspect ratio typically decrease with increasing defect size. Therefore, the aspect ratio of a defect is commonly used in fracture mechanics-based approaches for life prediction to represent the defect shape. In this study, a rotating ellipsoid was used to simplify defect shape. The spheroid is a geometric shape formed by rotating an ellipse 180° around its short axis. The defect volume V and surface area S were calculated to obtain the semi-major axis a and semi-minor axis b of the spheroid. The calculation formulas are given in Equations (1) and (2):

$$V = \frac{4\pi ab^2}{3}, \tag{1}$$

$$S = \frac{4\pi(2ab + b^2)}{3}. \tag{2}$$

The shape of the ellipsoid can be expressed by the aspect ratio α, as expressed in Equation (3):

$$\alpha = \frac{a}{b}. \tag{3}$$

The shape of the irregular defect was simplified into a regular ellipse, greatly improving the accuracy of defect recognition, as shown in Figure 4.

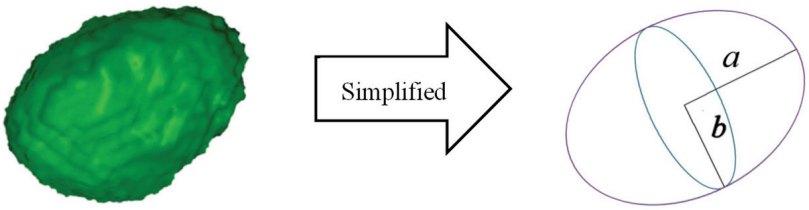


Figure 4. Simplification of defect shape.

2.5. Weight Index Calculation

To identify the critical representing feature for meso-defects, the Pearson correlation coefficient was introduced [39], in which the Pearson correlation coefficient re-

https://doi.org/10.3390/s26020631

flects the stability of relationships along with changes for variables through overall covariance calculation.

$$\gamma_{xy} = \frac{\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}))}{n-1}}{S_x S_y \sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2)}{n-1}} \sqrt{\frac{\sum_{i=1}^{n}(y_i-\overline{y})^2)}{n-1}}}, \tag{4}$$

where x and y represent two sets of data, $\{x_1, x_2, \cdots, x_n\}$ and $\{y_1, y_2, \cdots, y_n\}$. S_x and S_y are the standard deviations of the following data: $S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2)}{n-1}}$, $S_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\overline{y})^2)}{n-1}}$.

The relative weights of the chosen representing indices for meso-defect can be calculated as follows:

$$\omega_i = \frac{k_i}{\sum_{n=1}^{3} k_n}, \tag{5}$$

where ω_i is the weight of the defect feature S_i relative to the other defect features.

2.6. FEM Model

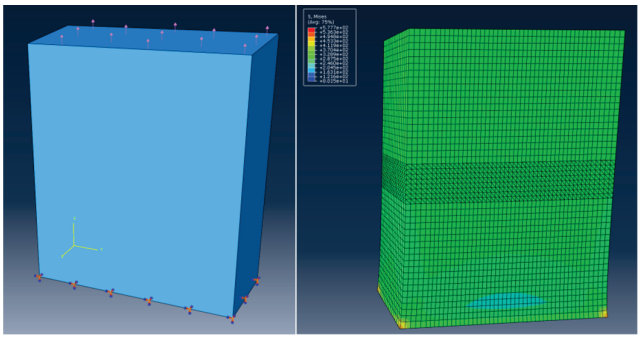
For weight index calculation, the stress concentration factor was used as the main factor. The finite element method was implemented for stress concentration factor calculation along with meso-defect-representing index changes. An FEM model was established for the center section of the specimen in a $6 \times 5 \times 2$ mm^3 domain, the same as the CT scanning section. Meso-defect data after 3D reconstruction (Figure 3) was imported into ABAQUS. The characteristic information of the defects was imported through an ABAQUS subroutine.

The simulation was divided into 3 groups according to defect characteristics, and each group performed 5 simulation experiments by changing the characteristic quantities. The material was modeled as linearly elastic and homogeneous at the mesoscale, a standard assumption for studying the elastic stress concentration caused by isolated meso-defects prior to plasticity or damage evolution. A displacement control method was adopted to carry out the finite element tensile analysis of the damage model. One end of the model was fixed, and stress (p) was applied to the other end on the other end, as shown in Figure 5. From the final tensile simulation results, the value ($p_{\max}$) of the maximum stress point was obtained, as shown in Figure 5b. The stress concentration factor (k) was uniformly introduced as an intermediate variable for each group of simulations to determine the degree of material damage. The stress concentration factor (k) of the damage model can be obtained by Equation (6):

$$k = \frac{p_{\max}}{p}. \tag{6}$$

The stress concentration factor was obtained through simulation using the above model. To establish a transparent and interpretable baseline model to preliminarily identify the main effects of different defect features on damage accumulation, a single-parameter perturbation method was used, in which the characteristic quantity of defect was increased by 10% each time. In order to avoid the mutual influence of defect spacing, when the porosity of defects increased, the defects size was increased proportionally while keeping the original defect positions unchanged. For the aspect ratio, the initial shape was assumed to be 1. For the position index, the volume of defects remained the same, but the distribution and arrangement of defects changed.

https://doi.org/10.3390/s26020631



(**a**) (**b**)

Figure 5. (**a**) Tension model; (**b**) simulation results.

3. Results and Discussions

3.1. Meso-Defect-Representing Features

Table 2 presents the meso-defect-representing feature information at different fatigue stages, including number, porosity, aspect ratio, and position to the surface, after X-ray CT scanning, reconstruction, and simplification. Defects with volumes greater than 10^{-8} mm^3 were taken into consideration for defect number. As meso-defects were simplified into ellipsoid, the representing index for shape was the average aspect ratio. The location-representing index was defined as the average distance to the surface. Defect number decreased with increasing load cycles, whereas porosity, aspect ratio, distance to the surface increased with load cycles.

Table 2. Simplified defect information of staged fatigue test.

Load Cycle ($\times 10^3$)		**20**	**40**	**60**	**80**	**100**
Total number of defects	Group 1	35,106	215,398	125,703	35,684	20,982
	Group 2	36,061	233,821	104,895	30,569	16,456
	Group 3	34,854	190,636	76,695	24,534	18,672
Porosity (%)	Group 1	0.035	0.166	0.268	0.338	0.343
	Group 2	0.076	0.141	0.219	0.288	0.376
	Group 3	0.06	0.176	0.242	0.343	0.439
Aspect ratio	Group 1	0.434	0.659	0.783	0.859	0.887
	Group 2	0.329	0.614	0.681	0.731	0.739
	Group 3	0.404	0.713	0.787	0.798	0.855
Average distance to the surface (mm)	Group 1	0.894	0.852	1.137	1.061	1.233
	Group 2	0.92	1.038	1.204	1.15	1.412
	Group 3	1.045	1.512	1.334	1.435	1.314

3.2. Meso-Defect-Representing Feature Correlation Coefficient

For Pearson correlation calculation, the fatigue cycle number was selected as an intermediate parameter, and the average values in Table 2 for defect number, porosity, aspect ratio (shape), and distance to the surface (position) were used for correlation analysis. As shown in Figure 6, a weak correlation was observed between meso-defect number and damage cycles. For porosity, shape, and position, strong correlations were observed, with coefficients greater than 90%. Therefore, three major meso-defect features—porosity, aspect ratio, and position to surface—were taken as the representing indices for further damage assessment.

https://doi.org/10.3390/s26020631

	numbers	porosity	shape	position	cycles
numbers	1	−0.364	−0.0466	−0.1198	−0.4217
porosity	−0.364	1	0.9363	0.9527	0.9961
shape	−0.0466	0.9363	1	0.9749	0.9309
position	−0.1198	0.9527	0.9749	1	0.9353
cycles	−0.4217	0.9961	0.9309	0.9353	1

Figure 6. Correlation coefficient between representing features and fatigue cycles.

3.3. Weight Index for Representing Features

Based on the above FEM model, the stress concentration factor can be calculated, along with variations in porosity, aspect ratio, and shortest distance to the surface, as shown in Figure 7. The stress concentration factor value increased with changes in porosity and aspect ratio and decreased with decreasing distance to the surface.

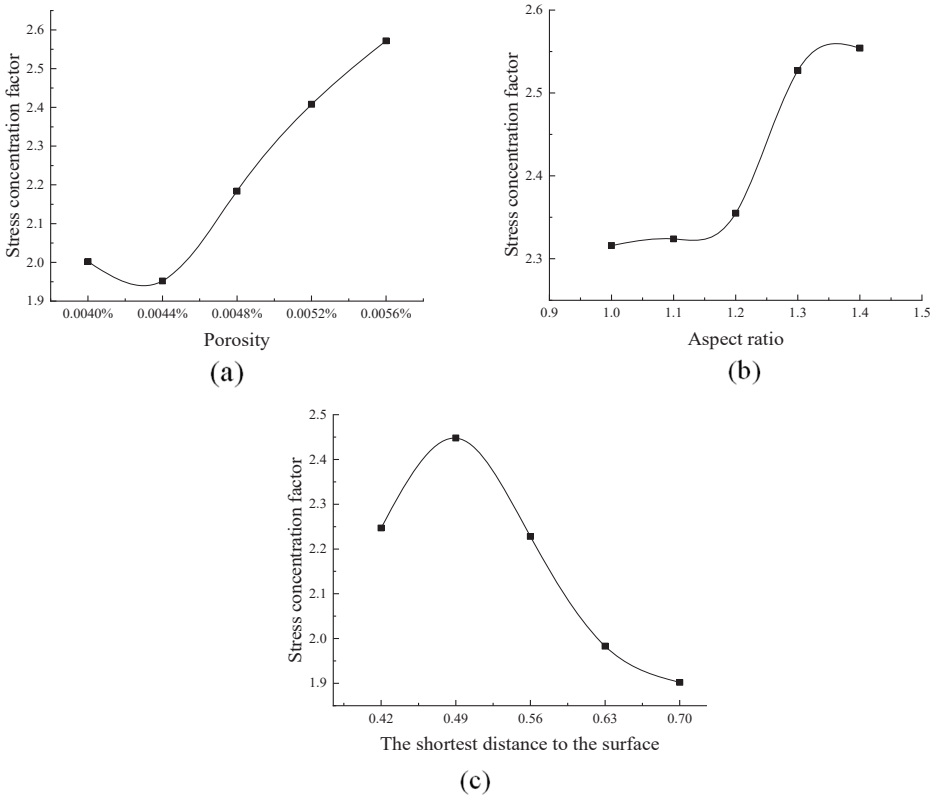


Figure 7. Change in stress concentration factor with defect feature, (**a**) porosity, (**b**) shape, and (**c**) position.

https://doi.org/10.3390/s26020631

These values can be fitted into a linear function using the least-squares method. The function can be described by Equation (7):

$$\begin{cases} y_1 = 39900x_1 + 0.3080 \\ y_2 = 0.6793x_2 + 1.5998 \\ y_3 = -1.649x_3 + 3.0851 \end{cases}, \quad (7)$$

where x_1, x_2, and x_3 represent porosity, aspect ratio, and short distance to surface, respectively, and where y_1, y_2, and y_3 represent the corresponding stress concentrate factors.

Based on the previous hypothesis, the defect characteristic variable x_i increases by 10% and the mesoscopic damage variable x_g will correspondingly increase by k_i%. The elastic coefficients of the defect characteristic indices for porosity, shape, and location can be calculated as k_1 = 8.38, k_2 = 2.98, and k_3 = 5.98. Thus, the relative weights of porosity, shape, and location can be obtained by the above-mentioned weight calculation method Equation (5) as $\omega_1 = 0.483, \omega_2 = 0.172$, and $\omega_3 = 0.345$, respectively.

3.4. Mesoscopic Damage Variable

The mesoscopic damage variable is fundamental for formulating constitutive models that describe the progressive failure process in engineering materials [40]. Due to the uneven size and distribution of meso-defects, a comprehensive damage variable was proposed by combining porosity, shape, and location to represent complex, heterogeneous meso-defects in aluminum alloys.

For accurately describing material failure under the fatigue process, incorporating meso-defect feature indices—porosity, shape, and location—into a comprehensive index is essential. This index, the mesoscopic variable $\boldsymbol{d}$, was obtained from the relative weights of the representing features, as shown in Equation (8). The mesoscopic variable $\boldsymbol{d}$ varies within the range [0, 1]. At the rupture point (100,000 cycles), $\boldsymbol{d}$ = 1.

$$d = \omega_1 \frac{P_i}{P_m} + \omega_2 \frac{S_i}{S_m} + \omega_3 \frac{L_i}{L_m}, \quad (8)$$

where ω_1, ω_2, and ω_3 are the relative weights corresponding to porosity, shape, and position of defects. P_i, S_i, and L_i are the damage equivalents of the three features at a given cycle stage (porosity, shape, and location), as shown in Table 2. P_m, S_m, and L_m are characteristic quantities corresponding to complete damage in terms of porosity, shape, and location.

To verify the effectiveness of the mesoscopic damage variable, the macroscopic damage variable $\boldsymbol{D}$ was introduced for validation of the defined mesoscopic damage variable.

$$D = \frac{k_e - k_{min}}{k_{max} - k_{min}}, \quad (9)$$

where k_e is the stress concentration factor corresponding to the equivalent damage of the material, $k_{\min}$ is the stress concentration factor the undamaged material, and k_{max} is the stress concentration factor corresponding to the maximum damage of the material. As the maximum fatigue life of the test specimen obtained from preliminary experiments was approximately 100,000 cycles, the stress concentration factor at the 100,000-cycle stage was defined as the maximum stress concentration factor.

The relative error between the mesoscopic and macroscopic variables can be calculated, and then the average value can then be used as the similarity δ, as shown in Equation (10):

$$\delta = \sum_{i=1}^{5} \frac{|d_i - D_i|}{D_i} \times 100\%. \quad (10)$$

https://doi.org/10.3390/s26020631

By substituting the relative weights and representing index data for different fatigue stages in Table 2 into Equation (8), the mesoscopic damage variable d and the macroscopic damage variable $\boldsymbol{D}$ can be obtained, as shown in Table 3.

Table 3. Macroscopic damage and mesoscopic damage at each stage.

Cycle ($\times 10^3$)	20	40	60	80	100
D	0.464	0.617	0.783	0.865	1
d	0.384	0.591	0.718	0.823	1
δ	17.2%	4.2%	8.3%	4.9%	0

A quite similar trend was observed for the mesoscopic variable $\boldsymbol{d}$ and the macroscopic damage variable $\boldsymbol{D}$, increasing with load cycles from 0 to 1, which is similar to the normal damage variables reported in previous studies [7,41]. The biggest difference was observed at the 20,000-cycle stage, probably because of the influence of more than 30,000 small meso-defects (Table 2). As load cycles increased, larger meso-defects developed, and the difference between $\boldsymbol{d}$ and $\boldsymbol{D}$ was reduced. The average relative error between the mesoscopic variable $\boldsymbol{d}$ and the macroscopic variable $\boldsymbol{D}$ is 6.9%, showing good agreement in damage evolution. Due to the fusion of 3 representing features (porosity, shape, and location) in the mesoscopic variable $\boldsymbol{d}$, the calculation conditions may be closer to the real fatigue conditions of aluminum alloys.

The reported average error of 6.9% is a global metric calculated across all stress amplitudes within the calibration dataset. This error is not constant and is expected to vary with stress amplitude due to underlying damage mechanisms. For broader application, this method can be utilized for both low-cycle and high-cycle fatigue. For low-cycle fatigue, the error may increase. Due to the linear-elastic matrix assumption, the macroscopic damage variable integrates significant plastic dissipation and micro-crack coalescence—nonlinear processes that are not explicitly captured by the current mesoscale model. For high-cycle fatigue, the error might be lower, as damage evolution is primarily elastic and the linearized, elasticity-based mesoscopic damage variable better approximates macroscopic stiffness degradation.

Furthermore, as normalization of the mesoscopic damage variable $\boldsymbol{d}$ was a key model assumption, fracture was enforced to occur at 100,000 cycles, and d was consequently normalized to unity. Instead of macroscopic elastic-modulus-based normalization, fracture was defined at a fixed observed cycle number (100,000 cycles), to provide a clear and unambiguous normalization point for comparing damage trajectories across different specimens and conditions in our dataset. This approach is common in continuum damage mechanics when the final fracture point from direct physical measurement is not available for all calibration scenarios. Modulus-based normalization will be included in future work. A plan to measure the critical elastic modulus drop at fracture in subsequent experiments will be outlined to refine the model's terminal boundary condition.

3.5. Fatigue Life Assessment Based on Mesoscopic Damage Variable

The mesoscopic damage variable serves as a critical bridge between microstructural degradation and macroscopic fatigue failure. The mesoscopic damage variable $\boldsymbol{d}$ in Table 4 can be derived separately, as shown in Figure 8. The mesoscopic damage variable $\boldsymbol{d}$ increases with loading cycles from 0 to 1 until fracture. This result is similar to previous studies of damage variable D based on Young's modulus degradation [21] and the multi-scale fatigue damage index (MDFI) incorporating meso-scale defects [42], demonstrating the effectiveness of the proposed mesoscopic damage variable.

https://doi.org/10.3390/s26020631

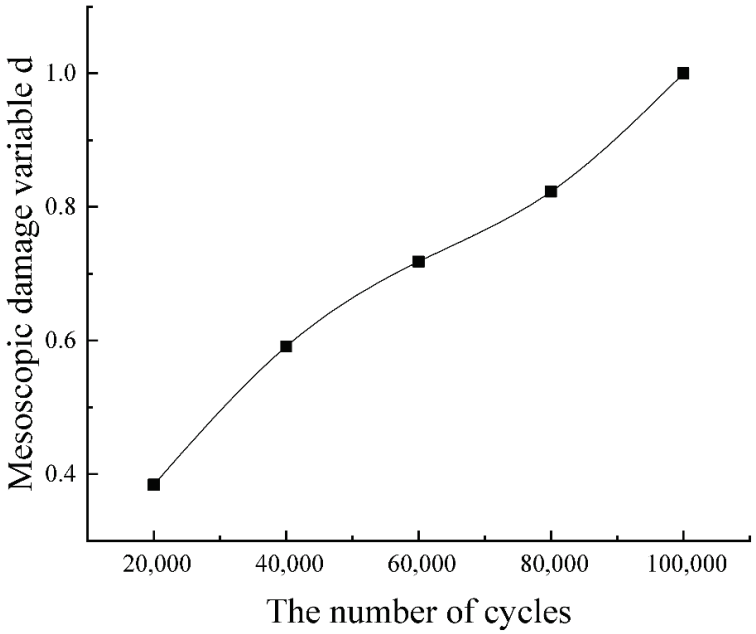


Figure 8. Mesoscopic damage variable in different fatigue stages.

By polynomial regression, the relation between fatigue life (N) and mesoscopic damage (***d***) can be obtained, as shown in Equation (11):

$$N = -64.572d^4 + 129.08d^3 - 72.008d^2 + 17.504d + 0.0004. \quad (11)$$

For continuum damage mechanics, the uncertainty of constitutive relationships tends to increase as they approach the boundaries of the calibration range. To ensure the prediction capability of the proposed model, the validation point was chosen within the experimental testing cycle range [0, 100,000]. For verification, the fatigue stage at 50,000 cycles was selected as a separate fatigue stage.

The fatigue test was carried out using the same experimental parameters settings, with three replications. For comparison, the same procedures were conducted, including fatigue testing, CT scanning, 3D reconstruction, simplification, and so on. Porosity was obtained through AVIZO. For shape, the aspect ratio was calculated after simplification. For location, the shortest distance from the meso-defect coordinate position to the surface was used. Porosity, shape, and location at 50,000 cycles are listed in Table 4.

Table 4. Representing features of meso-defects under 50,000 fatigue cycles.

Meso-Defect Representing Feature	Porosity	Shape	Location
Parameter	0.189	0.834	0.913

Thus, the mesoscopic damage ***d*** for 50,000 cycles was 0.648 from Equation (8). Based on Equation (10), the predicted fatigue life was 48,436 cycles. The error between the prediction method and the actual test was 3.13%, indicating relatively high accuracy. This illustrates that the model's predictions within the primary range of interest are accurate and physically consistent. Therefore, the proposed mesoscopic damage ***d***, incorporating meso-defect porosity, shape, and location, can be used for fatigue life assessment.

In the current study, we primarily focused on validating the model's interpolation performance within the fitted range, which forms the foundation of the model's reliability. A physically informed mesoscale framework for quantifying defect influence was

https://doi.org/10.3390/s26020631

established. The model relies on linear superposition of scalar damage contributions from meso-defects features, thereby neglecting nonlinear interactions. The study is currently based on 6061-T6 aluminum alloy under constant-amplitude loading to establish a proof of concept and calibrate the core defect-weighting methodology. The model reliably predicts fatigue life for meso-defect features and stress levels within the calibrated domain. For other alloys, arbitrary meso-defect distributions, or different production techniques, the baseline model required re-calibration of material-specific parameters for different materials and loading spectra. The generalizability of the proposed model should be further enhanced. Future extrapolation tests are an essential step in comprehensively assessing the model's robustness.

4. Conclusions

Based on staged fatigue experiments, meso-defect features were obtained and processed. A mesoscopic damage variable d incorporating defect porosity, shape, and location was established, along with a corresponding fatigue life prediction method. The following conclusions can be drawn.

(1) Based on simplified meso-defects, porosity, shape, and location were selected as the meso-defect-representing indices using correlation coefficient analysis.
(2) The weights of meso-defect-representing features were determined through FEM simulation based on stress concentration factor calculations.
(3) A mesoscopic damage variable $\boldsymbol{d}$ was determined using the weight method, along with a macroscopic damage variable D derived from the stress concentration factor. The average relative error value between the mesoscopic damage variable $\boldsymbol{d}$ and the macroscopic damage variable $\boldsymbol{D}$ was 6.9%.
(4) The relationship model between the mesoscopic damage variable $\boldsymbol{d}$ and fatigue life is established. Verification using the 50,000-cycle experiment showed an error of 3.13% between the proposed prediction method and the experimental result, validating the effectiveness of this method.

Further exploration of this baseline model should focus on the quantitative interaction relationships among meso-defect-representing features and on model normalization to better elucidate the mechanisms by which meso-defect influence fatigue damage evolution and to enable more accurate fatigue life assessment.

Author Contributions: L.Z.: Data curation, formal analysis, software, validation, visualization, writing—original draft preparation, and writing—review and editing. Y.Y.: data curation, formal analysis, writing—review and editing, conceptualization. H.C.: data curation, methodology, writing—review and editing, conceptualization, supervision, and resources. S.L.: investigation and validation. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ningxia Natural Science Foundation project (2024AAC03343).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to technical confidentiality.

Acknowledgments: Thanks for the help on simulation analysis related works from the Reliability and Lightweight Design Lab, Shanghai University of Engineering Science, Shanghai, China. Thanks for the support from the National First-Class Undergraduate Program in Mechanical Engineering and the First-Class Undergraduate Program in Automotive Service Engineering.

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

https://doi.org/10.3390/s26020631

References

1. Murakami, Y.; Endo, M. Effect of defects, inclusions and inhomogeneities on fatigue strength. *Int. J. Fatigue* **1994**, *16*, 163–182. [CrossRef]
2. Höppel, H.W.; Göken, M.; Mughrabi, H. On the importance of microstructural damage in the fatigue of metallic materials. *Z. Für Met.* **2002**, *93*, 1077–1088.
3. McDowell, D.L.; Dunne, F.P.E. Microstructure-sensitive computational modeling of fatigue crack formation. *Int. J. Fatigue* **2010**, *32*, 1521–1542. [CrossRef]
4. Zhang, W.; Shiozawa, K.; Beretta, S. Competing failure analysis: The role of intrinsic micro-defects vs. fatigue crack growth in additive manufactured Ti-6Al-4V under very high cycle fatigue. *Int. J. Fatigue* **2024**, *178*, 108–123.
5. Wang, P.; Tarafder, M.; Ai, Y. A probabilistic framework for assessing fatigue life of welded joints: Integrating initial defect distribution with multi-scale damage mechanics. *Fatigue Fract. Eng. Mater. Struct.* **2025**, *48*, 150–170.
6. Rabotnov, Y.N. On the equations of State for Creep. In *Progress in Applied Mechanics, the Prager Anniversary*; Macmillan: New York, NY, USA, 1963; pp. 307–315.
7. Murakami, S. *Continuum Damage Mechanics: A Continuum Mechanics Approach to the Analysis of Damage and Fracture*; Springer: Berlin/Heidelberg, Germany, 2012.
8. Li, X.; Zhang, Y.; Wang, C. A unified damage-coupled constitutive model for fatigue life prediction of metallic materials under multiaxial loading. *Int. J. Fatigue* **2024**, *189*, 108–125.
9. Lemaitre, J.; Desmorat, R. *Engineering Damage Mechanics: Ductile, Creep, Fatigue and Brittle Failures*; Springer: Berlin/Heidelberg, Germany, 2005.
10. Chaboche, J.L.; Lesne, P.M. A nonlinear continuous fatigue damage model. *Fatigue Fract. Eng. Mater. Struct.* **1988**, *11*, 1–7. [CrossRef]
11. Zhang, Y.; Liu, B.; Wang, Z.; Li, C. Quantitative characterization of meso-defect clusters and their influence on fatigue crack initiation in 7075-T6 aluminum alloy. *Acta Mater.* **2025**, *275*, 120041.
12. Zhou, J.; Chen, T.; Singh, A. A statistical framework for correlating meso-defect distribution with fatigue variability in additively manufactured AlSi10Mg. *Int. J. Fatigue* **2024**, *184*, 108276.
13. McDowell, D.L. Multiaxial fatigue of metals: Fatigue life and defect tolerance. *Fatigue Fract. Eng. Mater. Struct.* **1999**, *22*, 1083–1096.
14. Wang, K.; McDowell, D.L.; Cai, W. A multiscale correlative framework linking 3D meso-defect statistics to continuum damage evolution in high-strength aluminum alloys. *J. Mech. Phys. Solids* **2025**, *192*, 105718.
15. Martin, G.; Roux, S.; Babout, L. Deep learning-based segmentation and stochastic reconstruction of irregular 3D meso-defect populations for computational damage mechanics. *Model. Simul. Mater. Sci. Eng.* **2024**, *32*, 055024.
16. Ibrahim, R.; Tasan, C.C.; Sheidaei, A. In-situ synchrotron tomography and acoustic emission monitoring of fatigue damage nucleation from complex defect networks in cast Al alloys. *Nat. Commun.* **2024**, *15*, 5214.
17. Ahmad, M.; Hassan, M.; Khan, H.; Wang, L.; Li, W.; Karabal, S.; Withers, P.J. The role of defect population characteristics on the fatigue life of selective laser melted Ti–6Al–4V. *Int. J. Fatigue* **2022**, *165*, 107222.
18. Tammas-Williams, S.; Zhao, H.; Léonard, F.; Prangnell, P.B.; Todd, I.; Withers, P.J. The influence of porosity on fatigue crack initiation in additively manufactured titanium components. *Sci. Rep.* **2015**, *5*, 14988. [CrossRef]
19. Li, Y.; Li, W.; Wang, L.; Wang, H.; Zhu, G.; Ge, P. Machine learning-based fatigue life prediction of laser powder bed fusion manufactured alloys from porosity and defect spatial distribution. *Addit. Manuf.* **2023**, *72*, 103618.
20. Hrabe, N.; Gnäupel-Herold, T.; Quinn, T. Fatigue properties of a titanium alloy (Ti–6Al–4V) fabricated via electron beam melting (EBM): Effects of internal defects and residual stress. *Int. J. Fatigue* **2017**, *94*, 202–210. [CrossRef]
21. Bao, Y.Q.; Yang, Y.L.; Chen, H.; Li, Y.F.; Shen, J.; Yang, S.W. Multiscale Damage Evolution Analysis of Aluminum Alloy Based on Defect Visualization. *Appl. Sci.* **2019**, *9*, 5251. [CrossRef]
22. McLellan, D.L.; McLellan, M.M. An approach to fatigue design for aluminum castings. *AFS Trans.* **1996**, *4*, 445–449.
23. Feng, Q.; Otani, K.; Gokhale, A.M. Shape-sensitive fatigue life prediction model for aluminum alloys incorporating 3D meso-defect morphology. *Mater. Sci. Eng. A* **2024**, *908*, 146540.
24. Dang, L.; He, X.; Tang, D.; Xin, H.; Wu, B. A fatigue life prediction framework of laser-directed energy deposition Ti-6Al-4V based on physics-informed neural network. *Int. J. Struct. Integr.* **2025**, *16*, 327–354. [CrossRef]
25. Leuders, S.; Vollmer, M.; Brenne, F.; Tröster, T.; Niendorf, T. Fatigue strength prediction for titanium alloy TiAl6V4 manufactured by selective laser melting. *Metall. Mater. Trans. A* **2015**, *46*, 3816–3823. [CrossRef]
26. Xue, Y.; Pascu, A.; Horstemeyer, M.F.; Wang, L.; Wang, P.T. Microporosity effects on cyclic plasticity and fatigue of LENS™-processed steel. *Acta Mater.* **2010**, *58*, 4029–4038. [CrossRef]
27. Yadollahi, A.; Shamsaei, N.; Thompson, S.M.; Elwany, A.; Bian, L. Effects of building orientation and heat treatment on fatigue behavior of selective laser melted 17-4 PH stainless steel. *Int. J. Fatigue* **2017**, *94*, 218–235. [CrossRef]

https://doi.org/10.3390/s26020631

28. Chastand, V.; Tezenas, A.; Cadoret, Y.; Quaegebeur, P.; Maia, W.; Charkaluk, E. Fatigue characterization of titanium Ti-6Al-4V samples produced by additive manufacturing. *Procedia Struct. Integr.* **2016**, *2*, 3168–3176. [CrossRef]
29. Yang, S.; Meng, D.; Díaz, A.; Yang, H.; Su, X.; Jesus, A.M.D. Probabilistic modeling of uncertainties in reliability analysis of mid- and high-strength steel pipelines under hydrogen-induced damage. *Int. J. Struct. Integr.* **2025**, *16*, 39–59. [CrossRef]
30. Antonio, R.; Yves, N.; Mickaël, P.; Rémi, A.; Michel, F. Influence of defect morphology and position on the fatigue limit of cast Al alloy: 3D characterization by X-ray microtomography of natural and artificial defects. *Mater. Sci. Eng. A* **2020**, *785*, 139347. [CrossRef]
31. Liu, Y.; Zhang, R.; Fish, J. A phase-field fracture modeling framework for stochastic 3D meso-defect populations in additively manufactured metals. *Comput. Methods Appl. Mech. Eng.* **2025**, *428*, 117113.
32. Chen, Z.; Wicke, M.; Briffod, F. Data-driven reduced-order modeling for fatigue damage initiation at meso-defects using convolutional neural networks and finite element analysis. *Int. J. Numer. Methods Eng.* **2024**, *125*, e7467.
33. Hu, Y.N.; Wu, S.C.; Wu, Z.K.; Zhong, X.L.; Ahmed, S.; Karabal, S.; Xiao, X.H.; Zhang, H.O.; Withers, P.J. A new approach to correlate the defect population with the fatigue life of selective laser melted Ti-6Al-4V alloy. *Int. J. Fatigue* **2020**, *136*, 105584. [CrossRef]
34. Chen, J.; Liu, H.; Singh, R. Machine learning-enhanced damage factor characterization for high-cycle fatigue life assessment. *J. Mater. Sci. Technol.* **2025**, *215*, 45–58. [CrossRef]
35. Wang, H.; Zhang, J.; Li, B.; Xuan, F. Machine learning-based fatigue life prediction of laser powder bed fusion additively manufactured Hastelloy X via nondestructively detected defects. *Int. J. Struct. Integr.* **2025**, *16*, 104–126. [CrossRef]
36. du Plessis, A.; Yadroitsava, I.; Yadroitsev, I. Effects of defects on mechanical properties in metal additive manufacturing: A review focusing on X-ray tomography insights. *Mater. Des.* **2020**, *187*, 108385. [CrossRef]
37. Tammas-Williams, S.; Zhao, H.; Léonard, F.; Derguti, F.; Todd, I.; Prangnell, P.B. XCT analysis of the influence of melt strategies on defect population in Ti–6Al–4V components manufactured by Selective Electron Beam Melting. *Mater. Charact.* **2015**, *102*, 47–61. [CrossRef]
38. Wei, K.; Tjhung, A. Implementation of a continuum damage mechanics model for fatigue life prediction in commercial FE codes: Methodology and validation. *Eng. Fract. Mech.* **2021**, *253*, 107902.
39. Ziółkowski, G.; Chlebus, E.; Szymczyk, P.; Kurzac, J. Application of X-ray CT method for discontinuity and porosity detection in 316L stainless steel parts produced with SLM technology. *Arch. Civ. Mech. Eng.* **2014**, *14*, 608–614. [CrossRef]
40. Chen, G.P.; Wang, H. A personalized recommendation Algorithm on Improving Pearson Correlation Coefficient. *J. Shandong Agric. Univ. (Nat. Sci. Ed.)* **2016**, *47*, 940–944.
41. Lemaitre, J. A continuous damage mechanics model for ductile fracture. *J. Eng. Mater. Technol.* **1985**, *107*, 83–89. [CrossRef]
42. Yang, Y.; Chen, H.; Li, J.; Liu, W.; Shi, X.; He, Q. Digital prognosis of fatigue residual life of lightweight metallic materials. *J. Mater. Eng. Perform.* **2022**, *31*, 8068–8083. [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.

https://doi.org/10.3390/s26020631

sensors

MDPI

Article

Composite Fault Feature Index-Guided Variational Mode Decomposition with Dynamic Weighted Central Clustering for Bearing Fault Detection

Bangcheng Zhang *, Boyu Shen, Zhi Gao, Yubo Shao, Zaixiang Pang and Xiaojing Yin *

School of Mechatronic Engineering, Changchun University of Technology, Changchun 130103, China; shenboyu1209@163.com (B.S.); gaozhi@ccut.edu.cn (Z.G.); shaoyubo@ccut.edu.cn (Y.S.); pangzaixiang@ccut.edu.cn (Z.P.)
* Correspondence: zhangbangcheng@ccut.edu.cn (B.Z.); yinxiaojing@ccut.edu.cn (X.Y.)

Abstract

To address the periodic impacts and amplitude-modulated high-frequency resonance phenomena caused by bearing faults in rotating machinery, this paper proposes a detection method. The core innovation lies in: firstly, constructing a composite fault feature index (CFFI) that integrates normalized kurtosis and fuzzy entropy, which synchronously quantifies the fault impact intensity and periodic structure, and serves as an optimization objective; secondly, defininingа a spectral energy retention rate (SERR) that includes both the full spectrum and characteristic frequency bands to evaluate the denoising effect and fault feature retention, respectively. Based on this, the method adaptively determines the Variational Mode Decomposition (VMD) parameters through the Triangular Topology Aggregation Optimizer (TTAO), and uses Dynamic Weighted Center Clustering (DWCC) to screen key IMFs containing fault-envelope information. On the IMS bearing dataset, the SERR of the reconstructed signal is 0.21356, which is higher than the actual collected signal value of 0.22465, with a relative error of 4.9%, indicating a higher reconstruction accuracy. These quantitative results indicate that CFFI-guided optimization enhances impulsive and periodic fault components while maintaining stable feature-band retention. This approach is suitable for real-world equipment monitoring and exhibits strong engineering applicability.

Keywords: bearing fault; fault characteristic; modal decomposition; clustering algorithm

1. Introduction

As a critical component of rotating machinery, rolling bearing failure often manifests as periodic impacts between rolling elements and defect sites. This process excites high-frequency resonance, generating impulses and amplitude-modulation structures in the envelope domain [1–3]. Under real-world conditions characterized by variable operating states, intense noise, and multi-source interference, fault features are typically weak, non-stationary, and easily masked, posing significant challenges to the sensitivity and robustness of traditional diagnostic methods that rely on fixed parameters and single features [4]. Consequently, how to adaptively select frequency bands, achieve stable signal decomposition, and quantitatively evaluate feature preservation in real-world scenarios has become a key research direction in bearing fault signal processing.

Classical diagnostic procedures typically follow the "band selection–envelope demodulation" principle. Methods such as spectral kurtosis and kurtograms are first used to select

resonance bands that highlight transients, followed by zero-phase bandpass filtering and a Hilbert transform to obtain the envelope, thereby enhancing impulsive and modulated fault information [5]. While effective in many engineering applications, this fixed-parameter approach with a single band-selection criterion is prone to band-selection bias and feature degradation under low signal-to-noise ratio (SNR) and non-stationary conditions [6]. To enhance feature separation, adaptive time–frequency decomposition methods have been widely adopted. Techniques such as EMD, EEMD, and CEEMDAN can decompose signals into intrinsic modes at different scales and capture transients, but they suffer from issues such as mode mixing and end effects [7]. In contrast, Variational Mode Decomposition (VMD) is an adaptive signal decomposition method that recursively extracts intrinsic mode functions (IMFs) with specific center frequencies by constructing and solving a constrained variational problem. VMD iteratively solves for modes and their center frequencies within a variational framework that minimizes bandwidth, offering better noise robustness and decomposition stability, which aligns well with the "impact-resonance" signal mechanism [8]. However, the decomposition quality of VMD is susceptible to its penalty factor (α) and mode number (K). The penalty factor (α) controls the bandwidth of each extracted mode; a larger α yields smoother modes with better noise suppression but may obscure impulsive features, while a smaller α preserves detail at the cost of increased noise sensitivity. The mode number (K) determines how many intrinsic mode functions (IMFs) are generated; an insufficient K leads to under-decomposition (merging of distinct fault components), whereas an excessive K causes over-decomposition (splitting of fault features and introduction of spurious components). In bearing fault diagnosis, an inappropriate penalty factor or mode number (α, K) can directly obscure the impulse-induced resonance signatures targeted for extraction. Without data-driven criteria, over-decomposition or under-decomposition can easily occur.

To address VMD's parameter sensitivity, recent research has employed metaheuristic optimization algorithms, using a health indicator as the fitness function to achieve adaptive parameter tuning. This approach demonstrates a good balance between global and local search in selecting signal-processing parameters [9,10]. Meanwhile, entropy-based features (e.g., sample entropy, permutation entropy, fuzzy entropy) and kurtosis are complementary in characterizing the complexity and impulse strength of time series. They are often used as fault-sensitive indicators or combined into composite criteria to enhance sensitivity and robustness to weak features [11,12]. In the post-decomposition stage, clustering methods can help distinguish noise-dominant from fault-dominant intrinsic mode functions (IMFs), reducing the impact of mode mixing on reconstruction quality and thus preserving fault-related components more effectively. For evaluation, publicly available datasets, such as the IMS bearing data, provide standardized benchmarks and multi-condition coverage for generalized validation [13].

In addition, in terms of constructing health indicators, although there have been studies combining multiple features [14], they are mostly used for later diagnosis and rarely used as a unified optimization criterion for guiding the entire process of signal decomposition [15]. In the stage of modal selection and reconstruction, traditional methods often rely on empirical thresholds or simple rules to screen IMFs, lacking robustness analysis of IMF cluster characteristics, and are susceptible to mode aliasing and noise IMF interference [16].

To tackle the aforementioned challenges, this paper constructs a closed-loop diagnostic framework following the sequence: "indicator, decomposition, clustering, reconstruction, evaluation." Specifically: (1) Guided by spectral kurtosis, perform frequency band selection, followed by zero-phase filtering and envelope demodulation. (2) Use the proposed Composite Fault Feature Indicator (CFFI), which integrates impulse sensitivity with complexity

https://doi.org/10.3390/s26041394

robustness, as the fitness function to optimize VMD's α and K parameters adaptively. This focuses the decomposition on components related to fault envelopes. (3) Apply Dynamic Weighted Central Clustering (DWCC) to the obtained IMFs and reconstruct the fault-dominant components. (4) Quantify noise reduction and feature preservation using a defined Spectral Energy Retention Rate (SERR). The framework's stability and engineering applicability are validated using the IMS bearing dataset. A comparative analysis with baseline methods using various entropy-based fitness functions demonstrates the superior robustness of the proposed approach.

The main contributions of this paper are as follows:

1. CFFI integrating features sensitive to impulses and robust to signal complexity, serving as a unified criterion for both VMD parameter optimization and IMF selection.
2. Employs health-indicator-driven metaheuristic optimization to adaptively select VMD's α and K, mitigating over-/under-decomposition and enhancing focus on fault-envelope information.
3. Adopts DWCC for robust clustering and reconstruction of IMFs, reducing the effects of mode mixing and highlighting fault-dominant components.
4. Introduces the SERR as an evaluation metric, establishing a closed-loop quantitative assessment from band selection to reconstruction, facilitating online monitoring and trend tracking.

2. Research Methodology

The objective of this study is to achieve real-time detection of bearing faults (e.g., spalling on inner/outer rings or rolling elements) by determining optimal VMD parameters. This guides the adaptive selection of the impact and modulated resonance information inherent to such faults [17,18]. This objective is achieved through four key steps: (1) A CFFI specific to bearing faults is proposed to quantitatively measure the strength of fault-related impulses and periodic structures, simultaneously serving as the fitness function for the optimization algorithm. (2) The selected fitness function is used to optimize the two key VMD parameters: the penalty factor α and the mode number K. This optimization ensures that the decomposed IMFs are concentrated within the fault-related resonance bands and envelope components, thereby adaptively amplifying bearing fault signatures. (3) An improved clustering method, DWCC, is applied to the IMFs. Based on the CFFI values, the cluster with the highest average CFFI is selected for signal reconstruction, yielding a signal enriched with bearing fault features. (4) Finally, the feature-band SERR is calculated within the known or estimated fault characteristic frequency band Ω, enabling the identification and trend tracking of bearing fault characteristics.

2.1. Health Indicators

In fault diagnosis systems, health indicators are the key to converting raw vibration signals into interpretable, state-sensitive quantitative measures. Their core function is to detect, quantify, and track early fault characteristics, especially in noisy and non-stationary operating conditions, where simple visual or statistical analysis often fails. To effectively capture the key changing characteristics of bearing fault signals, this paper proposes a novel CFFI.

Traditional health indicators, such as root mean square (RMS), KI, and various entropy indicators, are often designed for the single physical characteristics of signals. Although effective in specific scenarios, their sensitivity and robustness are difficult to balance when facing the complex characteristics of impact and periodic modulation that coexist in bearing faults. For example, kurtosis is sensitive to impact but susceptible to noise and outlier

https://doi.org/10.3390/s26041394

interference; The entropy index has a certain ability to characterize periodic structures, but its peak response to transient shocks is weak.

This makes CFFI not only a better-performing alternative indicator, but also a closed-loop guiding core that embeds diagnostic intelligence into the entire process of "decomposition, optimization, clustering, evaluation". The application of traditional indicators is often open-loop and isolated, while CFFI optimizes and filters signals through the same criteria, improving the cohesion, adaptability, and target consistency of the entire feature extraction process. In addition, in Section 4, this article compares CFFI with the traditional metrics root mean square (RMS) and KI in three evaluation dimensions: trend (Tred), monotonicity (Mono), and scale-adaptive comprehensive feature variability (SAIF).

2.1.1. Construction of the Composite Fault Feature Index

The introduction of the CFFI is based on an understanding of the limitations of existing signal-feature extraction methods, and its construction is crucial to the adaptiveness of the optimization algorithm proposed in this paper. Traditional features such as RMS and KI perform well in specific scenarios. However, when dealing with complex, non-stationary signals exhibiting impact-resonance modulation characteristics typical of bearing faults, a single feature is often insufficient. Bearing faults usually manifest as high-frequency resonance excited by periodic impacts, resulting in pronounced impulses and periodic structures in the envelope domain. To simultaneously capture both impulse sensitivity and periodic structure, the CFFI integrates FE and KI through normalized weighted fusion, defined as follows:

$$\mathrm{CFFI}(IMF_i) = w_k \times \mathrm{KInorm}(IMF_i) + w_e \times [1 - \mathrm{FEnorm}(IMF_i)] \tag{1}$$

where IMF_i represents the i-th intrinsic mode function; KInorm and FEnorm denote the normalized values of KI and FE for that IMF, respectively; w_k and w_e are weighting coefficients (default: $w_k = w_e = 0.5$, adjustable based on operational conditions). This definition ensures that a higher CFFI value indicates stronger impulses and periodic structures in the signal related to bearing faults. The CFFI serves both for selecting and clustering individual IMFs and as the fitness function for the Triangular Topology Aggregation Optimizer (TTAO) in optimizing VMD parameters, enabling the decomposition process to focus on fault-related features adaptively.

Fuzzy entropy [19] is used to quantify the complexity and uncertainty of a time series. Its value tends to be lower when the signal exhibits stronger periodicity. The calculation formula is as follows:

$$\mathrm{FEnorm}(IMF_i) = -\ln\left(\frac{\phi_{m+1}(IMF_i)}{\phi_m(IMF_i)}\right) \tag{2}$$

where ϕ_m and ϕ_{m+1} represent the fuzzy similarity measures at embedding dimensions m and m+1, respectively.

The kurtosis index [20] quantifies impulse strength and the peakedness of a distribution. Impacts from bearing faults and the resulting pulses in the envelope domain typically cause a significant increase in kurtosis. It is calculated using the following formula:

$$\mathrm{KInorm}(IMF_i) = \frac{1}{N}\sum_{k=1}^{N}\left(\frac{IMF_i[k] - \mu}{\sigma}\right)^4 \tag{3}$$

where N denotes the sample size, μ is the mean, and σ represents the standard deviation.

The construction described above ensures that the CFFI is simultaneously sensitive to both impulsive shocks and periodic structures. Consequently, a higher CFFI value indicates

https://doi.org/10.3390/s26041394

a more substantial presence of components associated with bearing faults. Therefore, the CFFI serves a dual purpose: it acts as a criterion for selecting relevant IMFs and functions as the fitness function for the TTAO optimization algorithm, guiding the adaptive search for optimal VMD parameters.

2.1.2. Assessment Metric

To quantitatively evaluate the performance of the proposed novel health indicator (CFFI), this paper employs two standard metrics (Tred and Mono) along with one comprehensive metric, SAIF. Let the total number of sample segments be K, and the health indicator value for each segment be denoted as h_k. The segment index s_k can represent the time series. The definitions and significance of the three types of metrics are provided below.

(1) Tred: This metric measures the linear correlation between the health indicator and time, reflecting the overall change trend of the indicator as the fault evolves over time.

$$\text{Tred} = \frac{\left|\sum_{k=1}^{K}\left(h_k - \overline{H}\right)\left(s_k - \overline{S}\right)\right|}{\sqrt{\sum_{k=1}^{K}\left(h_k - \overline{H}\right)^2 \sum_{k=1}^{K}\left(s_k - \overline{S}\right)^2}} \tag{4}$$

where $\overline{H}$ is the mean value of the health indicator, and $\overline{S}$ is the average value of the time index. To enhance robustness, s_k can be assigned as the rank of the segment index.

(2) Mono: This metric measures the direction of change in the health indicator over time, which is crucial for evaluating the degradation process of equipment health. Monotonicity is defined as follows:

$$\text{Mono} = \frac{\left|N_p - N_n\right|}{K - 1} \tag{5}$$

where N_p and N_n represent the number of health indicators that are greater than 0 and less than 0 after differentiation, respectively, and K denotes the sample size.

(3) SAIF: This is used to comprehensively measure health indicators across multiple time scales, integrating methods such as multi-scale analysis, statistical features, and weighted summation. The definition of SAIF is as follows:

$$\text{SAIF} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} w_i \left(\frac{1}{L}\sum_{j=1}^{L}\left(\frac{\sigma_{ij}}{\mu_{ij} + \epsilon}\right)^2 + \alpha \cdot skew\left(x_{ij}\right) + \beta \cdot kurt\left(x_{ij}\right)\right)} \tag{6}$$

where N is the number of time scales in the signal, L is the number of sub-windows within each time scale, w_i is the weight of the ith time scale, σ_{ij} and μ_{ij} are the standard deviation and mean of the j-th sub window under the i-th time scale, respectively, ϵ is a small constant to prevent division by zero, α and β are weight coefficients for skewness and kurtosis, and $skew\left(x_{ij}\right)$ and $kurt\left(x_{ij}\right)$ are the skewness and kurtosis of the j-th sub window under the i-th time scale, respectively.

2.2. Variational Mode Decomposition

VMD is an adaptive signal-processing method [21,22] that decomposes a complex signal into several IMFs with distinct center frequencies. It addresses issues such as mode mixing and end effects that are present in traditional EMD methods [23] and is widely applied in fields such as fault diagnosis and signal denoising. The core of VMD lies in minimizing the bandwidth of each mode through a variational approach. This process can be summarized in the following steps:

(1) Objective of Mode Decomposition:

https://doi.org/10.3390/s26041394

a. The input signal *f(t)* is assumed to be decomposed into K mode functions $u_{(k)}$, each associated with a center frequency ω_k.
b. The objective is to find these mode functions $u_k(t)$ and their corresponding center frequencies ω_k such that the bandwidth of each mode is minimized.

(2) Formulating the Variational Problem:

The problem of minimizing modal bandwidth is formalized as a constrained variational problem:

$$\begin{gathered}\min_{\{u_k\},\{\omega_k\}} \sum_{k=1}^{K} \| \partial_t \left(\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right) e^{-j\omega_k t} \|_2^2 \\ \text{s.t.} \sum_k u_k = f\end{gathered} \tag{7}$$

where ∂_t denotes the time-derivative operator, $\delta(t)$ is the Dirac delta function, which ensures orthogonality of the modes in the time domain, $*$ represents the convolution operator, j is the imaginary unit, $u_k(t)$ is the kth mode function, and ω_k is its corresponding center frequency.

(3) Lagrangian Multiplier Method:

To handle the reconstruction constraint $\sum_{k=1}^{K} u_k(t) = f(t)$, the Lagrangian multiplier *λ(t)* is introduced:

$$\begin{aligned}L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_{k=1}^{K} \| \partial_t \left(\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right) e^{-j\omega_k t} \|_2^2 + \\ \| f(t) - \sum_{k=1}^{K} u_k(t) \|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle\end{aligned} \tag{8}$$

(4) Solving the Variational Problem: The Alternating Direction Method of Multipliers (ADMM) is employed to solve the aforementioned Lagrangian function iteratively, updating the mode functions $u_k(t)$ and the center frequencies ω_k.

$$\begin{aligned}u_k^{n+1} = \arg\min_{u_k} \| \partial_t \left(\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right) e^{-j\omega_k t} \|_2^2 + \\ \frac{\alpha}{2} \| f(t) - \sum_{i \neq k} u_i(t) - u_k(t) + \frac{\lambda}{2\alpha} \|_2^2\end{aligned} \tag{9}$$

$$\begin{aligned}\omega_k^{n+1} = \frac{\int_0^\infty \omega |u_k(\omega)|^2 d\omega}{\int_0^\infty |u_k(\omega)|^2 d\omega} \\ \lambda^{n+1} = \lambda^n + \tau \left(f(t) - \sum_{k=1}^{K} u_k^{n+1}(t) \right)\end{aligned} \tag{10}$$

(5) Convergence Criterion: It is determined whether the change in the mode functions u_k during each iteration satisfies the convergence condition:

$$\sum_{k=1}^{K} \frac{\| u_k^{n+1} - u_k^n \|_2^2}{\| u_k^n \|_2^2} < \epsilon \tag{11}$$

If satisfied, the iteration stops. The motivation for selecting VMD as the modal decomposition method lies in its high degree of alignment with the mechanism of bearing faults: (1) It effectively captures and separates complex vibration modes. (2) It suppresses noise, thereby enhancing signal clarity. (3) It adapts to non-stationary signals, enabling precise analysis of transient features. (4) It identifies and distinguishes multiple fault modes.

https://doi.org/10.3390/s26041394

2.3. Triangulation Topology Aggregation Optimizer

The Triangulation TTAO is a novel metaheuristic algorithm based on mathematical principles, specifically designed to address complex continuous optimization problems in engineering applications [24]. TTAO leverages the geometric properties of similar triangles to balance exploration and exploitation, thereby improving the efficiency of finding the optimal solution.

The core concept of TTAO is to partition the search space into multiple triangular topology units. Each unit is optimized by updating its vertices and generating interior points within the triangle. These units iteratively evolve through processes of global and local aggregation, gradually converging towards the optimal solution.

The TTAO algorithm is implemented through the following steps:

(1) Initialization: An initial population P is distributed within the search space, forming several triangular units. Each unit consists of three vertices X_{i1}, X_{i2}, X_{i3}, and an additional point X_{i4}

$$\mathrm{P} = \{X_1, X_2, ..., X_N\} \tag{12}$$

where N represents the population size, and each X_i is a d-dimensional vector.

(2) Formation of Triangular Units: The vertices X_{i1}, X_{i2}, and X_{i3} of each triangular unit can be randomly distributed within the search space during the initialization process. The additional point X_{i4} is typically generated inside the triangle to assist in the optimization process.

(3) Global Aggregation: During the Global Aggregation phase, new candidate solutions are generated by performing weighted combinations of the vertices from different triangular units. This step emphasizes exploring potential regions within the search space.

Assuming the vertices of a triangular unit Δ_{ijk} are X_i, X_j, and X_k, a new point X_{new} can be generated using the following formula:

$$\mathrm{X_{new}} = \alpha X_i + \beta X_j + \gamma X_k \tag{13}$$

where α, β, γ are weighting coefficients that satisfy $\alpha + \beta + \gamma = 1$, and are typically determined using random methods.

(4) Local Aggregation: The Local Aggregation phase focuses on refining the search within each triangular unit. A local search around the best vertex is conducted by perturbing its position to find a better solution. For each vertex X_i, its update can be expressed as:

$$\mathrm{X_i^{(t+1)}} = X_i^{(t)} + \lambda \cdot \left(X_{best} - X_i^{(t)}\right) \tag{14}$$

where λ is a step-size coefficient, X_{best} is the best-performing vertex in the current triangular unit, and t denotes the iteration number.

(5) Iteration and Convergence: The algorithm alternates between global and local aggregation phases, updating the individuals in the population until a convergence condition is met (e.g., a maximum number of iterations or an acceptable solution quality). The convergence criterion can typically be defined as the change in the objective function value $\Delta f < \epsilon$, or reaching a predefined maximum number of iterations $t_{\max}$: $\Delta f = \left| f\left(X^{(t)}\right) - f\left(X^{(t-1)}\right) \right| < \epsilon$ or $t \geq t_{max}$.

https://doi.org/10.3390/s26041394

2.4. Dynamic Weighted Centroid Clustering

In signal analysis, decomposing a complex signal into several IMFs with distinct characteristics is a crucial step. Building on this, clustering the IMFs can further reveal underlying patterns and structures within the signal. To better analyze and utilize the decomposed IMFs, this section proposes a DWCC method, which is described in detail alongside actual vibration signals. The specific steps of the DWCC algorithm are as follows:

(1) Data Standardization: Prior to cluster analysis of the IMFs, it is necessary to standardize the CFFI values of the IMFs. This process aims to eliminate the influence of IMFs with different magnitudes and scales on the clustering results, allowing for comparison and analysis of all IMFs on a unified scale. The standardized CFFI value is defined as

$$\mathrm{CFFI}_{\mathrm{normalized}}(IMF_i) = \frac{\mathrm{CFFI}(IMF_i) - \mu_{\mathrm{CFFI}}}{\sigma_{\mathrm{CFFI}}} \tag{15}$$

where $\mathrm{CFFI}_{\mathrm{normalized}}(IMF_i)$ is the CFFI value of the i-th IMF. μ_{CFFI} and σ_{CFFI} are the mean and standard deviation of the CFFI values for all IMFs, respectively.

Through standardization, the impact of differences in magnitude across IMFs is eliminated, thereby improving the accuracy and effectiveness of subsequent cluster analysis.

(2) Cluster Center Initialization: After standardization, the K-means++ algorithm is employed to select K initial cluster centers $\{\mu_1, \mu_2, \cdots, \mu_K\}$, which are determined using the following formula:

$$\mu_j^{(0)} = \arg\min_i \parallel \mathrm{CFFI}_{\mathrm{normalized}}(IMF_i) - \mu_j \parallel^2 \tag{16}$$

where $\mu_j^{(0)}$ denotes the initial center of the j-th cluster. The K-means++ algorithm ensures a reasonable distribution of the initial centers, thereby improving the convergence speed and the effectiveness of the clustering results.

(3) Improved Distance Metric: In DWCC, an improved weighted Euclidean distance metric is used to calculate the similarity between each IMF and the cluster centers, enabling precise measurement of the similarities among IMFs:

$$d_w(IMF_i, \mu_j) = \sqrt{\sum_{k=1}^{m} \mathrm{w}_k \left(\mathrm{CFFI}_{\mathrm{normalized},k}(IMF_i) - \mu_{j,k}\right)^2} \tag{17}$$

where w_k is the weight for each feature. m is the dimensionality of the features.

(4) Adaptive IMF Assignment: An adaptive distribution strategy is employed, which dynamically assigns IMFs to the most suitable cluster center based on the local density and global similarity of each IMF. The adaptive IMF assignment strategy is defined as

$$\mathrm{c_i} = \arg\min_j \left(d_w(IMF_i, \mu_j) + \lambda \frac{1}{\rho(IMF_i)} \right) \tag{18}$$

where c_i is the assigned class of IMF_i. λ is a balancing factor used to adjust the influence of distance and density. $\rho(\mathrm{IMF}_i)$ represents the local density of IMF_i.

(5) Dynamic Cluster Center Update: Cluster centers are updated dynamically using a weighted averaging method to achieve smooth transitions and accelerate convergence. The dynamic update strategy is as follows:

$$\mu_j^{(t+1)} = \alpha \mu_j^{(t)} + (1 - \alpha) \frac{1}{|C_j|} \sum_{i \in C_j} \mathrm{CFFI}_{\mathrm{normalized}}(IMF_i) \tag{19}$$

https://doi.org/10.3390/s26041394

where α is a smoothing coefficient that controls the weight distribution between the new cluster center and the old one. C_j is the set of IMFs currently belonging to the jth cluster. $\mid C_j \mid$ is the number of IMFs in the jth cluster.

(6) Adaptive Convergence Criterion: The convergence criterion is dynamically adjusted based on the stability and trend of the cluster centers to ensure the algorithm achieves optimal clustering results within a reasonable timeframe:

$$\| \mu_j^{(t+1)} - \mu_j^{(t)} \| < \epsilon_t \forall j \tag{20}$$

where ϵ_t is the convergence threshold at the t-th iteration, which can be dynamically adjusted according to the historical changes in the cluster centers.

(7) Optimization Enhancement Strategy: After each iteration, a local search and optimization are performed on the assignment results to reduce intra-class variance further and enhance the robustness and accuracy of the clustering. Specific measures include: reassigning and optimizing IMFs within each cluster, removing IMFs that may belong to multiple categories, and adjusting them to more suitable classes.

2.5. Algorithm Flow

In the method proposed in this paper, the novel health indicator CFFI is used as the fitness function given in Equation (21), and TTAO is employed to optimize VMD. (* indicates specific numerical values used in practical operation.)

$$\begin{cases} fitness = \min\limits_{(K,\alpha)} (CFFI_i) \\ s.t. K \in (*, *) \\ \alpha \in [*, *] \end{cases} \tag{21}$$

$CFFI_i$ is the fitness function value obtained for each candidate. The proposed method is illustrated in Figure 1.

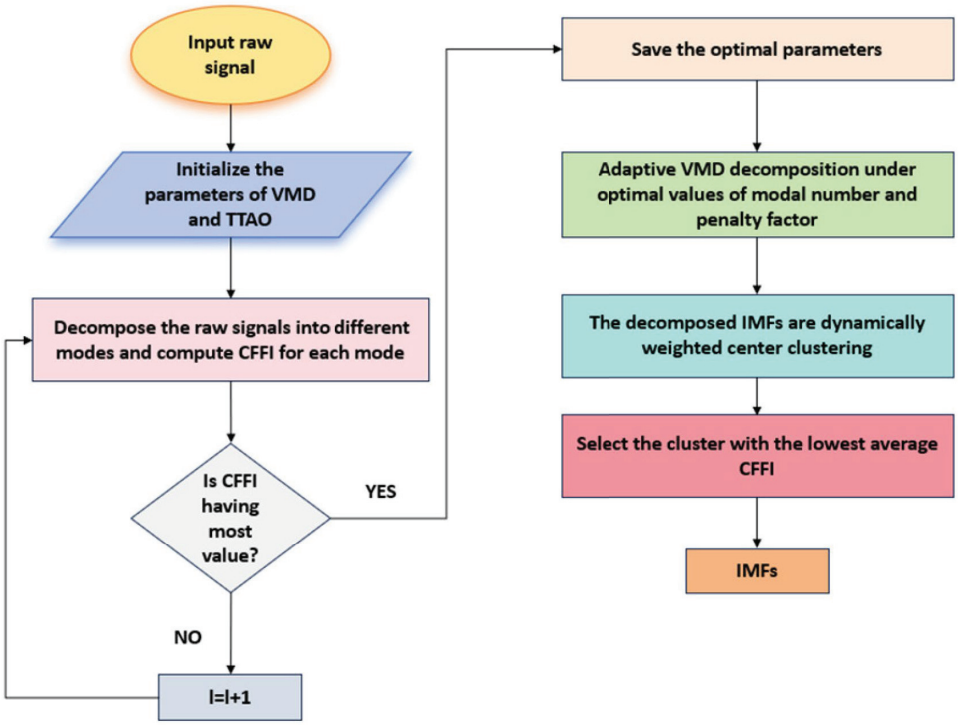


Figure 1. Algorithm flowchart.

https://doi.org/10.3390/s26041394

1. Initialize the parameters of TTAO and VMD.
2. Decompose the original signal into different modes.
3. Calculate the CFFI for each modality.
4. Calculate the fitness function, CFFI, for each modality until the maximum iteration is reached.
5. Obtain the global optimal fitness function and save the optimal values of the mode number K and the penalty factor α.
6. Perform DWCC on the decomposed IMFs.
7. Select the cluster with the highest average CFFI value as the IMFs for the reconstructed signal.

In practical decomposition, low-energy narrowband residual or broadband noise IMF usually exhibits lower CFFI values. DWCC will classify them into low CFFI clusters and remove them in the reconstruction stage, thereby improving the energy concentration and identifiability of fault-related components.

3. Simulation Analysis

To validate the applicability of the proposed method during bearing operation, a non-stationary simulated signal was constructed as a benchmark. This signal is designed to evaluate the algorithm's capabilities for decomposition, optimization, and reconstruction under time-varying frequency and amplitude conditions, as well as in the presence of background noise. It is important to note that this simulated signal is not a physical model of a bearing fault but rather a non-stationary test sequence used to assess the algorithm's robustness. The base signal is defined as a sinusoidal wave whose period and amplitude decrease over time via time attenuation factors. The mathematical expressions for a periodic signal with period T are given as:

$$\mathrm{f}(\mathrm{t}) = \mathrm{f}_0 \cdot \mathrm{e}^{-\mathrm{k_f} t} \tag{22}$$

$$\mathrm{A}(\mathrm{t}) = \mathrm{A}_0 \cdot \mathrm{e}^{-\mathrm{k_A} t} \tag{23}$$

$$\mathrm{x_{sin}}(\mathrm{t}) = A(t) \cdot \sin(2\pi f(t) t) \tag{24}$$

where A is the initial amplitude of the signal, f is the signal frequency, t is time, $f_0 = 2\,\mathrm{Hz}$, $k_f = 0.1\,\mathrm{s}^{-1}$, $A_0 = 1$, and $k_A = 0.2\,\mathrm{s}^{-1}$, as shown in Figure 2a.

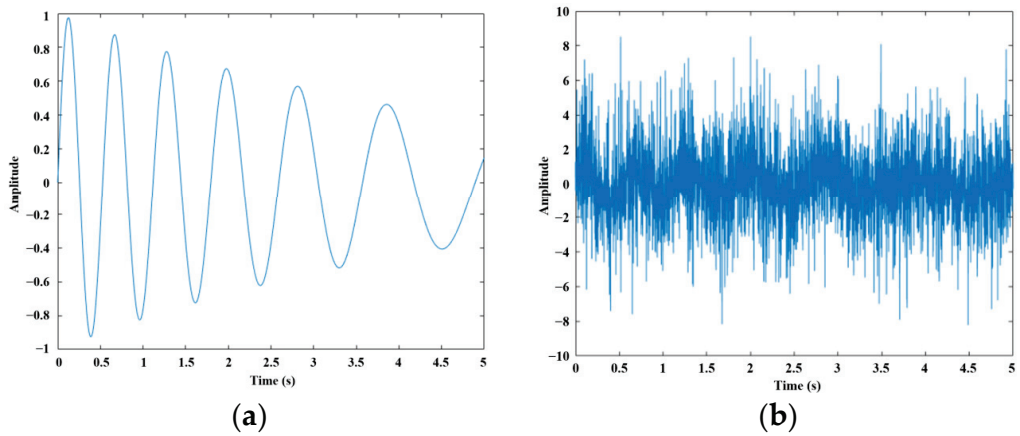


Figure 2. (**a**) Pure periodic decreasing signal; (**b**) periodic decreasing signal perturbed by white noise.

https://doi.org/10.3390/s26041394

On this basis, random white noise is generated and added to the above signal to form the final simulated signal, as shown in Figure 2b, expressed as:

$$x_{\text{final}}(t) = x_{sin}(t) + n(t) \tag{25}$$

where $n(t) = (a_{\min} + (a_{\max} - a_{\min}) \cdot r) \cdot \mathcal{N}(0,1)$, with $a_{\min} = 0.1$, $a_{\max} = 3.0$, r being a random number uniformly distributed in [0, 1], and $\mathcal{N}(0,1)$ representing Gaussian noise with zero mean and unit variance.

First, the parameters for VMD and TTAO were initialized, with CFFI set as the fitness function for TTAO, aiming to maximize fitness. For the VMD, the number of iterations was set to 100, and the maximum number of decomposition layers $K_{\max} = 25$. For the TTAO optimization, the number of iterations was set to 50. The resulting optimal fitness value found by TTAO was 0.11976 (as shown in Figure 3). Subsequently, VMD was used to decompose the original signal into several IMFs with different center frequencies. During each decomposition and parameter search, the CFFI was calculated for each IMF and aggregated (e.g., by taking the maximum value or the mean of the top p values) to form the fitness value. This process iterated until the parameter combination yielding the highest fitness was found, and the optimal parameters were saved (the optimal number of modes $K = 10$, and the penalty factor $\alpha = 200$). Under these parameters, the corresponding IMFs (see Figure 4) and their center frequencies (see Figure 5) were obtained.

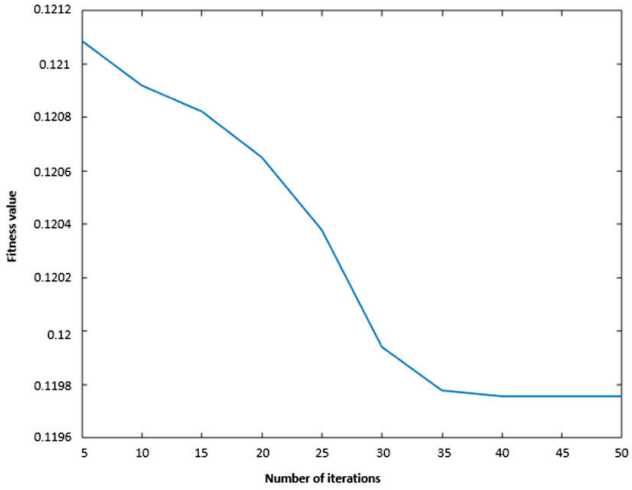


Figure 3. Iterative curve.

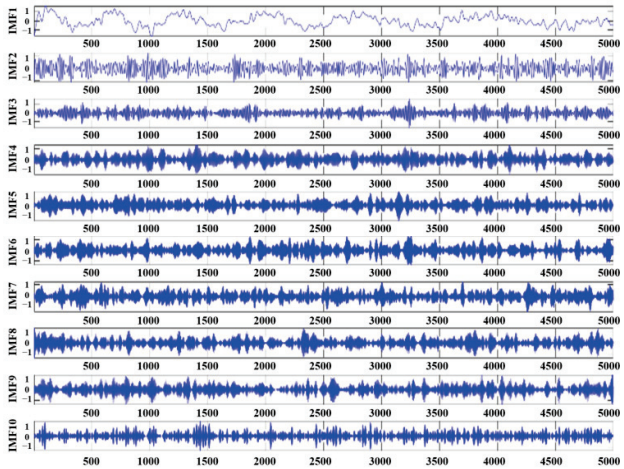


Figure 4. The IMFs after decomposition.

https://doi.org/10.3390/s26041394

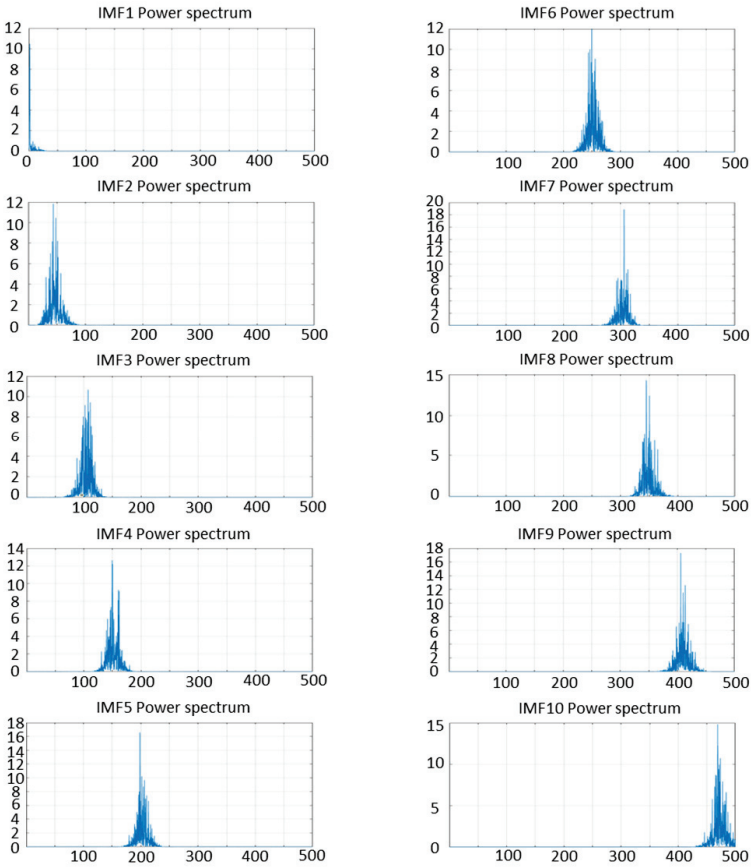


Figure 5. Frequency corresponding to IMFs.

As shown in Figures 4 and 5, the first three IMFs clearly capture the pulse-resonance-dominated shock-attenuation characteristics; the subsequent IMF (IMF4-IMF10) mainly includes low-energy narrowband residual and broadband noise preserved after optimized parameter decomposition (K = 10). Based on the screening of CFFI and DWCC, these low-CFFI modes are classified as noise clusters and discarded in the reconstruction, further ensuring the feature set and denoising effect of the reconstructed signal. DWCC was then performed on all IMFs, followed by CFFI-based screening. The cluster with the highest average CFFI was selected for reconstruction to focus on fault-related components (four components were obtained, as shown in Figure 6).

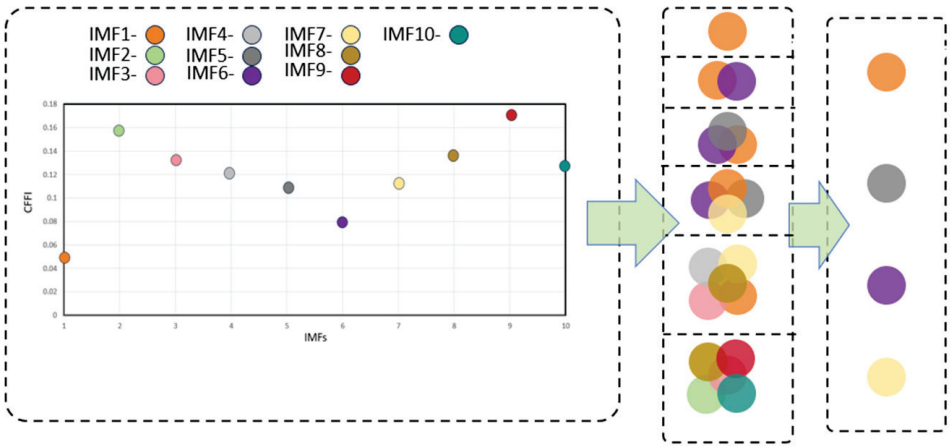


Figure 6. CFFI and DWCC select IMFs.

https://doi.org/10.3390/s26041394

In this example, the number of clusters was set to six. Since a higher CFFI value indicates stronger fault-related impulses and periodic structures, the cluster with the highest average CFFI (in this case, comprising four IMFs) was selected to form the reconstruction components. To evaluate denoising and feature-preservation performance, the SERR is introduced. Without requiring a clean reference signal, SERR assesses the overall quality of noise reduction and feature extraction by comparing the energy changes in the frequency domain before and after denoising. It is defined as

$$E_{original} = \sum_{k} |Y_{original}[k]|^2 \tag{26}$$

$$E_{denoised} = \sum_{k} |Y_{denoised}[k]|^2 \tag{27}$$

$$\text{SERR} = \frac{E_{denoised}}{E_{original}} \tag{28}$$

where $Y_{original}[k]$ is the spectral representation of the signal before denoising, $Y_{denoised}[k]$ is the spectral representation of the signal after denoising, $E_{original}$ represents the total spectral energy of the signal before denoising, and $E_{denoised}$ represents the total spectral energy of the signal after denoising.

Finally, the four obtained components were reconstructed (as shown in Figure 7, which compares the reconstructed signal with the pure signal, demonstrating effective denoising). The spectral energy of the reconstructed signal was calculated, resulting in an SERR value of 0.31254. A lower SERR value indicates a greater reduction in the denoised signal's overall spectral energy, indicating good denoising performance. Figure 8 better illustrates the specific implementation of the signal processing. Figure 8a shows the time-domain plot of the generated noisy signal, and Figure 8b displays the amplitude spectrum and cluster spectrum of the noisy signal. Under the influence of noise, the signal's spectrum appears complex and difficult to analyze directly. Figure 8c presents the amplitude spectrum and cluster spectrum of the denoised signal after processing by the proposed algorithm. Using this algorithm, noise is effectively removed, significantly enhancing signal clarity and feature extraction accuracy.

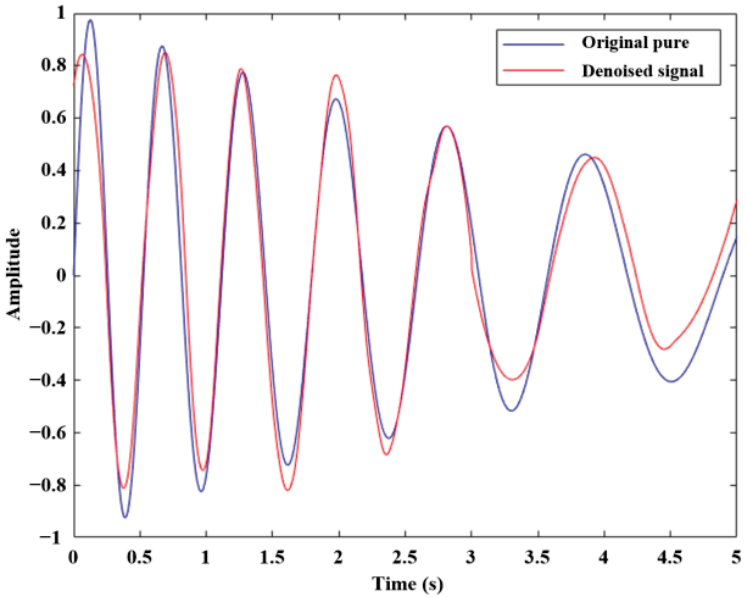


Figure 7. Contrast diagram of pure signal and denoised signal.

https://doi.org/10.3390/s26041394

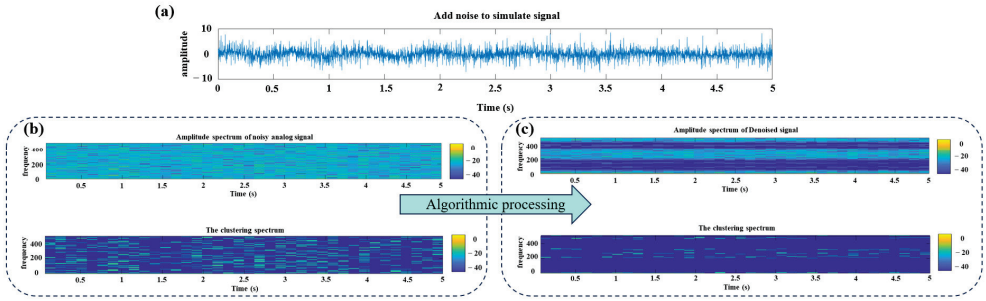


Figure 8. Analog signal spectrum energy change. (**a**) time-domain waveform containing noisy signals; (**b**) amplitude spectrum and clustering spectrum of the noise signal; (**c**) amplitude spectrum and clustering spectrum of the denoised signal.

As shown in Figure 8, the processed signal exhibits more concentrated spectral energy and reduced noise, validating the algorithm's effectiveness for signal denoising and feature extraction.

To verify the algorithm's stability, 10 additional simulation trials were conducted. The parameters for the simulated signals are listed in Table 1. Results from multiple trials demonstrate that the proposed method maintains good consistency and robustness under non-stationary and noisy conditions.

Table 1. Calculation of analog signal SERR.

Number	SERR
1	0.31526
2	0.41265
3	0.34512
4	0.31569
5	0.36965
6	0.37445
7	0.40168
8	0.35487
9	0.30875
10	0.32588

In summary, the signal-processing framework proposed in this paper applies to the analysis of non-stationary signals with decaying periods. It demonstrates good adaptability for feature extraction in bearing vibration signals.

4. Experimental Validation

To verify the practical application advantages of CFFI in bearing fault detection, comparative experiments were conducted based on the IMS public dataset. The overall process includes data preprocessing, indicator calculation and evaluation, decomposition and optimization, IMF screening and reconstruction, as well as ablation and repeated experiment analysis.

Data and Preprocessing:

(1) Frequency band selection and envelope extraction: The failure of the bearing body is usually manifested as periodic impact excitation of high-frequency resonance, and the formation of low-frequency amplitude modulation characteristics in the envelope domain. Combine known or estimated characteristic frequencies, or select bands based on spectral kurtosis to determine the bandpass range, using zero-phase bandpass filtering to preserve fault-related frequency bands; subsequently, Hilbert transform is applied to the bandpass signal to obtain the envelope, and slight detrending is

https://doi.org/10.3390/s26041394

performed as needed to eliminate DC bias. This process can suppress irrelevant noise while highlighting amplitude modulation and pulse information.

(2) Standardization: Standardize the envelope or bandpass signal with z-scores to achieve zero mean and unit variance, improving comparability across different operating conditions/time periods.

(3) Segmented processing: Segment the standardized signal into fixed time windows (approximately 20,480 points per window), and calculate CFFI, RMS, KI, and evaluation metrics (Trend, Monotonity, SAIF) window-by-window to maintain temporal resolution and characterize dynamic changes.

CFFI, root mean square (RMS), and kurtosis index (KI) were calculated window-by-window on IMS data, and evaluated and compared using three metrics: trend, monotonicity, and scale-adaptive integrated feature variability (SAIF). The summary results are shown in Figure 9.

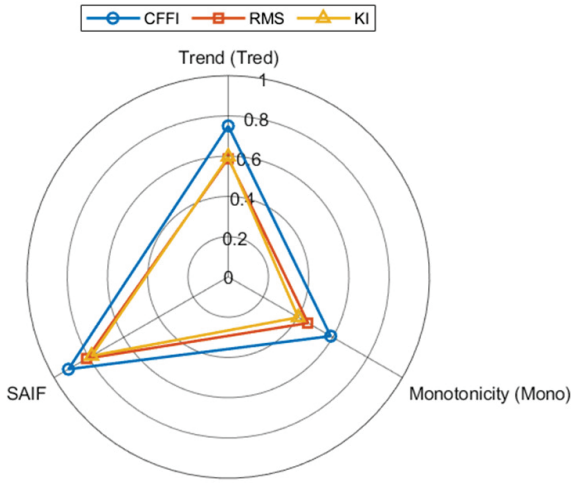


Figure 9. Radar chart comparison of the health indicators across Tred, Mono, and SAIF.

The results showed that CFFI had advantages in all three metrics, especially in the SAIF dimension approaching 1, indicating its stronger significance and stability in fault-related features at multiple scales. Combining the composite characterization of pulse-like (KInorm) and periodic structure complexity (1-FEnorm) in CFFI, it can be concluded that its advantage comes from simultaneously characterizing peak and periodic structures, thereby more robustly highlighting fault-related features under different operating conditions and time scales. This further validates the sensitivity and robustness of CFFI as a health assessment indicator for bearings.

The sampling rate for IMS data recording is 20 kHz, with approximately 20,480 points per file. The preprocessing stage determines the fault-related resonance band based on spectral kurtosis and its harmonics, and performs envelope demodulation and trend removal. Subsequently, the envelope domain signal is decomposed into IMFs with different center frequencies using VMD, with a fixed number of iterations of 100 and a maximum decomposition level of K_{max} = 25. To obtain the optimal decomposition configuration, TTAO is introduced to adaptively search for the penalty factor α and mode number K, and the optimization budget is uniformly set to 50 evaluations; in each candidate parameter evaluation, calculate the CFFI of each IMF and update the search parameters with the goal of maximizing CFFI. The representative results show that the optimal parameters are α = 400 and K = 10. Under this configuration, the SERR of the reconstructed signal is 0.21356, which is lower than the value of 0.22465 for the actual collected signal. This reduction in SERR indicates a decrease in extraneous spectral energy, thereby confirming the effectiveness

https://doi.org/10.3390/s26041394

of the denoising process and a higher quality of decomposition and reconstruction. The relative error between the two is 4.9%. Subsequently, the CFFI mean of each cluster is calculated, and the cluster with the highest mean is selected as the reconstruction set to avoid irrelevant or redundant modes from being mixed in. This process usually selects representative components such as IMF1, IMF5, IMF10 for reconstruction in experiments, which not only preserves the peak and amplitude modulation structure of fault pulses, but also effectively suppresses background noise and non-fault components.

On the premise of ensuring consistency in the data, preprocessing, VMD iteration, and optimization budget, ablation experiments were conducted to quantify the independent contributions of each module. Figure 10 compares the overall performance of five methods on SERR: (1) the complete method (TTAO+CFFI+DWCC); (2) the fixed parameters (K = 10, α = 400) are slightly higher than the complete method, indicating that empirical parameters can approach the optimum under specific operating conditions but lack adaptability; (3) under the same budget, the SERR of the grid search further increases, reflecting the influence of step size and discretization constraints; (4) replacing CFFI with minimum envelope entropy as the fitness resulted in a significant increase in SERR, indicating that entropy indices are less sensitive to spike pulses and complex changes than CFFI; (5) excluding DWCC and directly selecting the top K IMFs for reconstruction results in mixed information and energy dispersion significantly compromised the quality of reconstruction and diagnostic distinguishability.

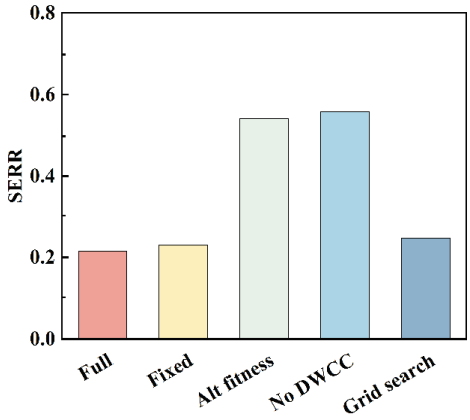


Figure 10. Ablation experiment analysis.

In addition, Figure 11 compares the SERR performance of different fitness functions in 10 repetitions. When using CFFI as the fitness, the 10 results remained stable in the range of approximately 0.20–0.24 with minimal fluctuations; when using minimum envelope entropy, minimum information entropy, minimum sample entropy, and minimum permutation entropy, the overall SERR significantly increases (typical range of about 0.50–0.60), and the dispersion across experiments significantly increases. This result indicates that CFFI can more directly measure the peak characteristics and complex dynamic changes of fault pulses, providing a clear convergence direction for optimizers; entropy-based indicators are susceptible to distribution estimation and window settings in noisy backgrounds and non-stationary pulse-dominated scenes, leading to unstable parameter selection and decreased reconstruction quality. Based on the data from different comparisons and repeated experiments, it can be concluded that TTAO is more efficient in approaching optimal solutions than grid search under limited evaluation budget; CFFI, as a fitness function, is more suitable for characterizing the pulse peaks of mechanical faults than entropy-based indicators; DWCC is a key step in ensuring the quality of IMF screening, which can im-

https://doi.org/10.3390/s26041394

prove the energy concentration of reconstruction and the significance of envelope spectral fault features.

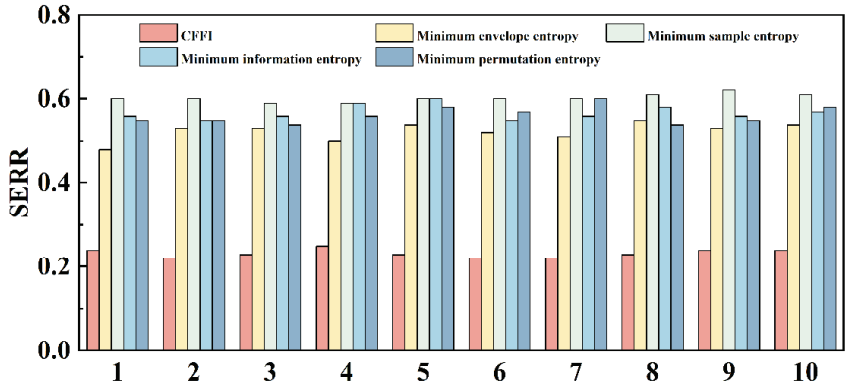


Figure 11. Comparative analysis of SERR under different fitness functions.

5. Conclusions

A health indicator-driven IMF clustering method for bearing fault detection was proposed. To address common issues in non-stationary and noisy environments, such as the susceptibility to mode mixing in traditional methods, difficulty in parameter self-adaptation, and lack of a closed-loop assessment, a framework for adaptive decomposition and selection guided by a CFFI was constructed. Using CFFI as both a sensitive and robust health indicator and an optimization objective, the method employs a TTAO to adaptively search for the optimal penalty factor α and mode number K for VMD. This enabled the decomposition to focus on fault-related impulse and amplitude-modulated resonance modes. Subsequently, DWCC was applied to screen the obtained IMFs, and the cluster with the highest average CFFI value—based on the semantic principle that "higher CFFI indicates stronger fault-related components"—was selected for reconstruction. Finally, a closed-loop assessment is based on the SERR, where the global full-spectrum SERR was used to measure overall noise reduction, and the feature-band SERR measures retention of fault characteristics. The experimental and simulation results based on the IMS public bearing dataset show that the proposed method achieved stable parameter estimation and clear fault modes under various operating conditions. Compared to baseline schemes using RMS, KI, or various entropy-based fitness functions, the proposed method exhibited higher Tred/Mono and SAIF. It achieved a significant reduction in the global SERR while maintaining a stable high retention in the feature-band SERR. This reflected its stronger sensitivity, robustness, and online applicability. It was noted that the current validation primarily relied on the IMS dataset; thus, the method's generalizability across different bearing types and more-variable operational conditions is suggested to warrant further investigation. Future work will focus on automating frequency band selection and CFFI weight adaptation, validating generalization under complex variable-speed conditions, and enabling real-time deployment with computational optimization for embedded platforms.

Author Contributions: Conceptualization, B.Z., B.S., Z.G., Y.S., Z.P. and X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Jilin Provincial Scientific and Technological Development Program (No. 20250306008ZD).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are contained within the article.

https://doi.org/10.3390/s26041394

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

Abbreviations

The following abbreviations are used in this manuscript:

SAIF	Scale-Adaptive Integrated Feature Variability
TTAO	Triangular Topology Aggregation Optimizer
DWCC	Dynamic Weighted Central Clustering
CFFI	Composite Fault Feature Indicator
VMD	Variational Mode Decomposition
SERR	Spectral Energy Retention Rate
IMFs	Intrinsic Mode Functions
SNR	Signal-to-Noise Ratio
RMS	Root Mean Square
Mono	Monotonicity
FE	Fuzzy Entropy
KI	Kurtosis
Tred	Trend

References

1. Antoni, J. The Infogram: Entropic Evidence of the Signature of Repetitive Transients. *Mech. Syst. Signal Process.* **2016**, *74*, 73–94. [CrossRef]
2. Maláková, S.; Sivák, S.; Fedorko, G.; Molnár, V.; Tomek, M. Analysis of Damage Origin of Bevel Gear Wheels. *Eng. Fail. Anal.* **2023**, *146*, 107105. [CrossRef]
3. Hou, Y.; Li, T.; Wang, J.; Ma, J.; Chen, Z. A Lightweight Transformer Based on Feature Fusion and Global–Local Parallel Stacked Self-Activation Unit for Bearing Fault Diagnosis. *Measurement* **2024**, *236*, 115068. [CrossRef]
4. Sun, Z.; Wang, Y.; Zhang, L. Rotating Machinery Health State Assessment under Multi-Working Conditions Based on a Deep Fuzzy Clustering Network. *Measurement* **2023**, *218*, 113172. [CrossRef]
5. He, G.; Ding, K.; Lin, H. Fault Feature Extraction of Rolling Element Bearings Using Sparse Representation. *J. Sound Vib.* **2016**, *366*, 514–527. [CrossRef]
6. Sawalhi, N.; Randall, R.B.; Endo, H. The Enhancement of Fault Detection and Diagnosis in Rolling Element Bearings Using Minimum Entropy Deconvolution Combined with Spectral Kurtosis. *Mech. Syst. Signal Process.* **2007**, *21*, 2616–2633. [CrossRef]
7. Huang, N.; Shen, Z.; Long, S.; Wu, M.L.C.; Shih, H.; Zheng, Q.; Yen, N.-C.; Tung, C.-C.; Liu, H. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. *Proc. R. Soc. London. Ser. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]
8. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
9. Zhu, J.; Wang, C.; Hu, Z.; Kong, F.; Liu, X. Adaptive Variational Mode Decomposition Based on Artificial Fish Swarm Algorithm for Fault Diagnosis of Rolling Bearings. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2016**, *231*, 635–654. [CrossRef]
10. Li, C.; Liu, Y.; Liao, Y. An Improved Parameter-Adaptive Variational Mode Decomposition Method and Its Application in Fault Diagnosis of Rolling Bearings. *Shock Vib.* **2021**, *2021*, 2968488. [CrossRef]
11. He, C.; Wu, T.; Liu, C.; Chen, T. A Novel Method of Composite Multiscale Weighted Permutation Entropy and Machine Learning for Fault Complex System Fault Diagnosis. *Measurement* **2020**, *158*, 107748. [CrossRef]
12. Qian, P.; Ma, E.; Wang, Y.; Li, Q.; Xu, H.; Zhang, D.; Xiang, Z.; Tian, X. A Novel Multi-Band Demodulation Method and Its Application in Bearing Fault Detection. *Measurement* **2025**, *256*, 118405. [CrossRef]
13. Mao, W.; He, L.; Yan, Y.; Wang, J. Online Sequential Prediction of Bearings Imbalanced Fault Diagnosis by Extreme Learning Machine. *Mech. Syst. Signal Process.* **2017**, *83*, 450–473. [CrossRef]
14. Meng, T.; Jiang, X.; Yang, S.; Liu, J.; Gao, H.; Zhu, Z. Spectral Feature-Informed Difference Multi-Modes Decomposition for Compound Bearing Fault Diagnosis. *Expert Syst. Appl.* **2025**, *294*, 128735. [CrossRef]
15. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of Machine Learning to Machine Fault Diagnosis: A Review and Roadmap. *Mech. Syst. Signal Process.* **2020**, *138*, 106587. [CrossRef]
16. Li, X.; Ma, Z.; Kang, D.; Li, X. Fault Diagnosis for Rolling Bearing Based on VMD-FRFT. *Measurement* **2020**, *155*, 107554. [CrossRef]
17. He, Y.; Hu, M.; Feng, K.; Jiang, Z. Bearing Condition Evaluation Based on the Shock Pulse Method and Principal Resonance Analysis. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [CrossRef]

https://doi.org/10.3390/s26041394

18. Li, H.; Wu, X.; Liu, T.; Li, S.; Zhang, B.; Zhou, G.; Huang, T. Composite Fault Diagnosis for Rolling Bearing Based on Parameter-Optimized VMD. *Measurement* **2022**, *201*, 111637. [CrossRef]
19. Lee, H.-M.; Chen, C.-M.; Chen, J.-M.; Jou, Y.-L. An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **2001**, *31*, 426–432. [CrossRef]
20. Dibaj, A.; Hassannejad, R.; Ettefagh, M.M.; Ehghaghi, M.B. Incipient Fault Diagnosis of Bearings Based on Parameter-Optimized VMD and Envelope Spectrum Weighted Kurtosis Index with a New Sensitivity Assessment Threshold. *ISA Trans.* **2021**, *114*, 413–433. [CrossRef]
21. Li, H.; Li, S.; Sun, J.; Huang, B.; Zhang, J.; Gao, M. Ultrasound Signal Processing Based on Joint GWO-VMD Wavelet Threshold Functions. *Measurement* **2024**, *226*, 114143. [CrossRef]
22. Lei, W.; Wang, G.; Wan, B.; Min, Y.; Wu, J.; Li, B. High Voltage Shunt Reactor Acoustic Signal Denoising Based on the Combination of VMD Parameters Optimized by Coati Optimization Algorithm and Wavelet Threshold. *Measurement* **2024**, *224*, 113854. [CrossRef]
23. Wang, J.; Du, G.; Zhu, Z.; Shen, C.; He, Q. Fault Diagnosis of Rotating Machines Based on the EMD Manifold. *Mech. Syst. Signal Process.* **2020**, *135*, 106443. [CrossRef]
24. Zhao, S.; Zhang, T.; Cai, L.; Yang, R. Triangulation Topology Aggregation Optimizer: A Novel Mathematics-Based *Meta*-Heuristic Algorithm for Continuous Optimization and Engineering Applications. *Expert Syst. Appl.* **2024**, *238*, 121744. [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.

https://doi.org/10.3390/s26041394

sensors

MDPI

Article

Defect-Intent Ambiguity Addressing for Training-Free Deterministic PCB Defect Localization via Template Selection and Dissimilarity Mapping

Saiyan Saiyod *, Woottichai Nonsakhoo, Zhengping Li and Piyanat Sirisawat

Hardware-Human Interface and Communications Laboratory (H2I-Comm Lab), Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen 40002, Thailand; nonsakhoo@cassia.kku.ac.th (W.N.); zhengping.l@cassia.kku.ac.th (Z.L.); piyanat_siri@cassia.kku.ac.th (P.S.)
* Correspondence: saiyan@kku.ac.th

Abstract

Automated optical inspection (AOI) for printed circuit boards (PCBs) requires localizing small, sparse defects under illumination drift and minor placement misalignment, while supporting fast, auditable pass/fail decisions. This paper presents a training-free, reference-based digital image processing framework with no learning/training stage that compares each defective query image with a small library of defect-free reference templates (for the same PCB layout/revision) using a small set of interpretable control parameters. A reference is selected by coarse-to-fine matching (fast pre-screening followed by SSIM refinement on a central region), and an optional global alignment is applied only when it increases SSIM to limit defect-driven over-correction. Defects are highlighted by a defect-likelihood field that fuses an SSIM-derived structural dissimilarity map with a normalized absolute-difference map, followed by connected-component extraction to produce confidence-ranked bounding boxes. The method achieves Precision = 0.9663, Recall = 0.9987, and F1 = 0.9822 at the best-F1 operating point (0.149 false positives per image). Under the adopted box-matching protocol, average precision reaches 0.984. Precision–recall and FROC curves are reported to support threshold selection under different false-alarm budgets.

Keywords: printed circuit board (PCB); automated optical inspection (AOI); defect localization; defect detection; image processing; template selection; structural similarity; difference map; correlation; quantile thresholding

1. Introduction

Automated optical inspection (AOI) is a core quality fcontrol stage in printed circuit board (PCB) manufacturing, where decisions must be made reliably at production speed [1,2]. In practice, true defects are sparse and localized (e.g., missing copper, spurious copper, scratches, and pinholes), yet benign variability in illumination and minor board-placement changes is common [3]; therefore, a useful inspection method must suppress nuisance variability while remaining sensitive to subtle, localized structural deviations. AOI is typically positioned inside a broader in-line quality strategy that aims to reduce scrap and rework while keeping false alarms manageable in production [4].

In many industrial quality control (QC) workflows, the primary decision is pass/fail (accept/reject): if any verified defect is present, the board is rejected. Therefore, reliable defect localization and evidence are central, while fine-grained defect-type classification

(e.g., Missing_hole, Mouse_bite, Short) is often optional and can be treated as a downstream step for reporting, process monitoring, or root-cause analysis [5]. At the same time, surveys of PCB defects and inspection practices highlight that defect taxonomies and visual manifestations can vary across processes and products, which complicates universal classification claims and emphasizes the value of interpretable evidence [6].

A key practical disparity is that the notion of a "defect" can be ambiguous without context from the intended design [7]. In PCB engineering, many localized copper features are introduced deliberately to satisfy electromagnetic interference and compatibility (EMI/EMC) and signal/power integrity (SI/PI) constraints rather than to "look clean" in an image [7–10]. For example, an apparent copper bridge may be an unintended short in one location but an intentional connection, net-tie, or controlled strap in another. Likewise, patterns that resemble spurious copper in isolation can be purposeful copper pours and fills (e.g., ground pours for shielding and controlled return-current paths), via stitching and via fences to reduce radiation and crosstalk, guard traces, copper thieving/balancing for manufacturability, or local tuning structures used to manage impedance, ringing, and high-frequency noise.

Spur-like features and short stubs can also be intentional (e.g., test access, reinforcement/teardrops, thermal relief, or current-spreading structures), even though similar geometries may be considered defects in other contexts. Because these intent-driven features may be small and localized, appearance-only inspection can confuse design intent with anomaly [6,11].

This defect-versus-intent ambiguity motivates reference-based inspection: comparing a query board against a defect-free exemplar of the same layout provides the necessary design-context baseline [12]. The template encodes the intended routing, copper distribution, and layout-specific structures used for EMI/EMC and SI/PI, allowing the inspection decision to focus on deviations from the intended reference (true defects) rather than on visually unusual but intentional patterns [7,8].

In production, the primary constraint is often not peak accuracy in a curated benchmark but predictable behavior under everyday drift and change [4,13]. Practical AOI systems are judged by time-to-deploy on a new board, validation burden, compute footprint, and the ability to explain failures when they occur [13]. End-to-end learning-based models are widely used; however, in some deployment settings their internal decision logic can be harder to audit and their lifecycle cost can include labeling, retraining, monitoring, and re-qualification [5,14]. When an inspection decision is disputed, it is valuable to attribute the response to concrete causes such as illumination shift, residual misalignment, template mismatch, or a true localized defect [15,16].

Accordingly, a recurring deployment theme is to reduce manual labeling burden and model maintenance effort while still improving robustness, which motivates both training-free pipelines and low-label learning regimes [17,18].

These constraints motivate a complementary design point: analysis-first, physics- and mathematics-grounded image processing that exposes its decision signal. Classical industrial inspection continues to rely on deterministic components such as thresholding, morphology, and template matching because they are auditable and controllable [19,20]. In this view, a defect-free template functions as a measurement reference standard: inspection becomes change detection against the intended layout rather than category recognition.

The main obstacle is that simple subtraction is sensitive: small geometric mismatch and brightness drift can dominate the difference image and produce widespread false positives [15]. Robust reference-based inspection therefore benefits from structural similarity measures and careful handling of alignment [21]. The structural similarity index (SSIM) [22] emphasizes structure rather than raw intensity and can be evaluated locally

https://doi.org/10.3390/s26051541

to form an SSIM map that highlights localized changes. Alignment can further reduce nuisance differences, but in defect localization it must be used carefully because defects can bias the estimated transform; a pragmatic safeguard is to accept alignment only when it yields a measurable similarity improvement [23].

Related work on PCB AOI spans broad surveys and focused PCB-specific reviews that summarize classical image processing pipelines and, more recently, deep learning-based approaches [1,2,6,11,14,24]. Classical AVI/AOI surveys establish the foundational problem setting and recurring algorithmic motifs (preprocessing, registration/matching, and post-processing for decision rules) [25]. More recent reviews focus on electronics manufacturing and PCB inspection specifically, including the rise of deep learning and the associated practical considerations (data requirements, interpretability, and lifecycle/maintenance cost) [11,14,24]. Across these perspectives, a consistent theme is that production deployment often rewards predictable behavior and actionable diagnostics, which keeps reference-based inspection relevant even when learning-based models are available.

Beyond RGB surface imagery, inspection and monitoring can also involve specialized imaging and task-specific pipelines, for example in solder-joint and electronic assembly defect analysis, where the sensing modality and measurement objective differ from bare-board surface AOI [26,27].

A long-standing and still common formulation is reference/template-based change detection, where the query is compared to a defect-free exemplar and differences are extracted using similarity measures and post-processing [12,20,21]. In this family, normalized cross-correlation and related similarity measures are widely used to compare a query against a template, both for alignment/matching and for highlighting likely change regions [12,28]. Because AOI images can differ by small shifts, rotations, and illumination drift, many works explicitly study robustness of similarity/dissimilarity measures and the sensitivity of subtraction-based signals to misregistration [15,29]. Feature-based matching and alternative similarity measures (beyond raw subtraction) are also widely explored as ways to increase tolerance to nuisance variation while preserving localized change cues [16,30,31]. Rotation or uncertainty-aware designs further emphasize that handling geometric variability is a practical requirement rather than a corner case [32]. This practical issue is directly aligned with our use of structural similarity and alignment safeguards: the goal is to suppress broad, nuisance-driven responses while preserving localized defect evidence [22,23].

Within classical (training-free) inspection, several works explicitly combine subtraction/matching with deterministic operators for categorization and reporting [19,21,33]. For example, an algorithmic scheme for concurrent detection and classification applies fuzzy c-means segmentation after image subtraction and then uses arithmetic/logic operations, the circle Hough transform, morphological reconstruction, and connected-component labeling to assign defect categories; it reports 100% detection and 99.05% classification accuracy in its experimental setting [33]. In a similar spirit, other subtraction-based pipelines target both detection and defect-type decisions by combining difference images with hand-designed operators (e.g., morphology, logical operations, and component analysis), illustrating the long-standing appeal of deterministic, auditable processing in PCB AOI [19].

In parallel, deep learning has been widely adopted for PCB defect detection and classification, commonly borrowing generic object-detection backbones and adapting them for small, low-contrast defects and industrial constraints [24,34–37]. One line of work applies one-stage object detectors (often YOLO-family variants) and proposes architectural changes to improve localization of tiny PCB defects while keeping inference efficient [38–44]. Additional one-stage variants integrate backbone and attention refinements for feature reuse and improved localization under clutter [45–48]. Another line adapts two-stage and feature-

https://doi.org/10.3390/s26051541

pyramid designs (e.g., Faster R-CNN/FPN-style components) to improve multi-scale defect detection where targets are small and the background is cluttered [49–51]. Beyond detector backbones, several studies emphasize lightweight deployment, attention/feature-fusion refinements, and context modeling to trade off accuracy and throughput in practical settings [52–54]. Deep learning is also used for related PCB inspection tasks including defect classification, reconstruction/autoencoder-based inspection signals, and component-level understanding such as PCB segmentation for recognition [55–57]. Transformer- and DETR-inspired detectors are another recent direction, motivated by global context modeling and end-to-end detection design [58,59]. Beyond standard detector framing, learning-based approaches also include alternative objectives and cues (e.g., energy-based or edge-guided signals), domain adaptation strategies to handle distribution shift [60], and PCB-specific designs for bare-board defects [61–63].

Because labeled defect samples can be scarce and production distributions can drift, recent work also explores semi-supervised, unsupervised, and few-shot settings for PCB inspection [17,18,64]. These directions aim to reduce annotation burden and improve adaptability by leveraging unlabeled data, uncertainty modeling, or meta-learning-style transfer from limited support examples [17,18,64].

An important practical gap in PCB inspection is defect-versus-intent ambiguity. Appearance-only cues can confuse intentional layout features with true defects unless a layout-specific reference is available. Motivated by this, this work presents a training-free, reference-based image processing framework that retains the operational advantages of template inspection while improving robustness to everyday nuisance variation. While end-to-end learning-based defect detectors are widely used in industrial inspection across product domains, the proposed pipeline provides an auditable and controllable alternative that does not require a training lifecycle [65,66]. The method is computationally efficient and operationally transparent: it produces interpretable intermediate fields (similarity and difference maps) and relies on controllable operators whose behavior can be inspected and tuned without retraining.

The main contribution of this work lies in a training-free framework design that combines classical operators with explicit robustness safeguards targeting defect-versus-intent ambiguity under everyday nuisance variability. In particular, we emphasize ROI-based template selection, SSIM-gated alignment acceptance, and a fused defect-likelihood mapping that yields confidence-ranked detections for deployment-oriented PR/FROC threshold selection. The key contributions are as follows:

- Coarse-to-fine reference selection with central-region similarity: a fast pre-screening stage plus SSIM refinement computed on a central region to suppress border/outlier effects, with caching and downscaling for throughput.
- SSIM-gated alignment acceptance: an explicit validation rule that applies a global warp only when SSIM improves on the same central region, mitigating defect-driven over-correction.
- Fused, auditable defect-likelihood and operational evaluation: an explicit defect-likelihood field that fuses structural dissimilarity and normalized absolute difference with central-region-based rescaling for stable thresholding, producing confidence-ranked candidates that are characterized via PR/FROC and IoU-sensitivity rather than a single tuned threshold.

The remainder of this paper is organized as follows: Section 2 details the proposed method, Section 3 presents the experimental results and analysis (including the evaluation protocol), Section 4 discusses limitations and practical considerations, and Section 5 concludes this paper.

https://doi.org/10.3390/s26051541

2. Proposed Methods

This section presents a training-free PCB defect-localization pipeline. For each query image, it selects a defect-free reference template and outputs (i) an interpretable defect-likelihood field and (ii) confidence-ranked bounding boxes. At a high level, the pipeline performs coarse-to-fine template selection, constructs the defect-likelihood field by fusing SSIM-based structural dissimilarity with normalized absolute difference, applies global alignment only when it improves SSIM, and extracts sparse candidates via quantile thresholding, morphology, and connected-components analysis (Figure 1).

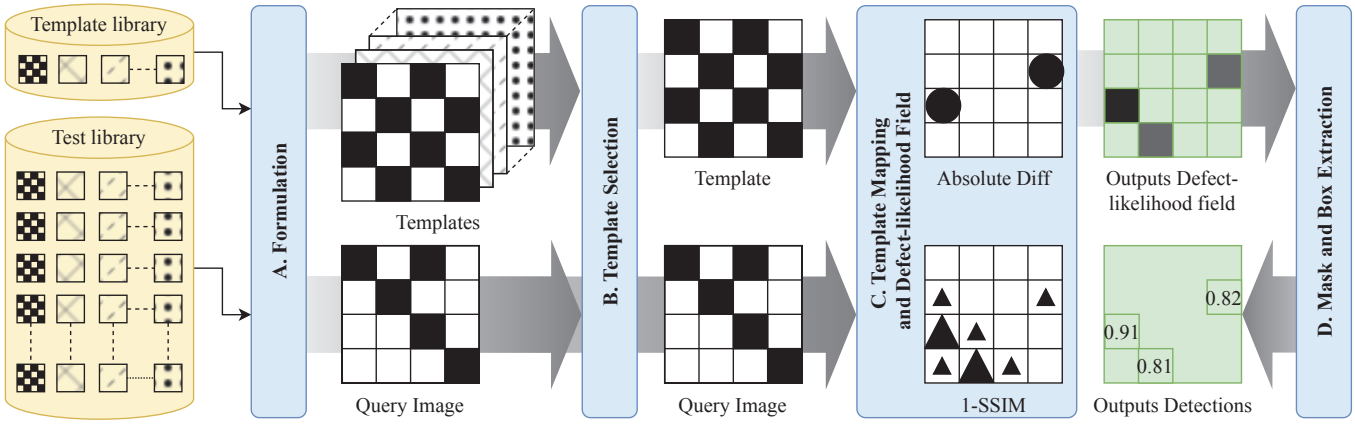


Figure 1. Overview of the proposed image processing framework for PCB defect localization.

2.1. Formulation

This work considers reference-based PCB inspection in which, for each defective query image, one or more defect-free exemplars of the same PCB layout are available. Let I denote a defective query image and $\mathcal{T} = \{T_k\}_{k=1}^{K}$ denote a library of defect-free templates. Here, templates refer to defect-free reference images for the same PCB layout/revision (not defect-type templates that enumerate defect categories). Consequently, the localization stage is class-agnostic across defect categories: any defect that manifests as a localized deviation from a valid defect-free reference can yield a response in the defect-likelihood field.

The goal is to output the following: (i) a spatial defect-likelihood field $M(u, v)$ (higher values indicate more likely defect) and (ii) a set of N axis-aligned bounding boxes with confidence scores $\mathcal{P} = \{(b_i, s_i)\}_{i=1}^{N}$, where each $b_i = [x\ y\ w\ h]$ denotes the top-left corner (x, y) and the box width/height (w, h), and s_i is a confidence score.

To stabilize similarity computation and threshold estimation, statistics are evaluated on a central image region Ω parameterized by a margin with m pixels on all sides. The margin m is computed from an area-avoidance ratio $p \in [0, 1)$, which specifies the fraction of the total image area to ignore.

2.2. Template Selection

All images are converted to grayscale and normalized to $[0, 1]$ by an affine intensity rescaling. Scalar SSIM is denoted by $\mathrm{SSIM}(\cdot, \cdot)$ [22]. Pixel coordinates are (u, v) (column, row) and the image domain is $\mathcal{X} = \{1, \ldots, W\} \times \{1, \ldots, H\}$.

Similarity and thresholds are evaluated on an inner region Ω parameterized by a symmetric margin m determined from an area-avoidance ratio $p \in [0, 0.95]$, where 0.95 is an experimental upper limit to ensure a nontrivial inner region. With a margin m on all four sides, the central area is $(W - 2m)(H - 2m)$ and the remaining area is

$$WH - (W - 2m)(H - 2m) = 2m(W + H) - 4m^2. \quad (1)$$

https://doi.org/10.3390/s26051541

Setting this equal to pWH yields the (smaller) quadratic root

$$m = \frac{(W+H) - \sqrt{(W+H)^2 - 4pWH}}{4}, \tag{2}$$

and $\lfloor m \rfloor$ is used as an integer padding. The inner region is

$$\Omega = \{(u,v) \in \mathcal{X} : 1+m \le u \le W-m,\ 1+m \le v \le H-m\}. \tag{3}$$

In implementation, this continuous expression is converted to an integer ROI with explicit rounding and safety clamping; see Appendix A.1.

Given a query image I and template library $\{T_k\}_{k=1}^{K}$, the template is selected by maximizing SSIM on the inner region at a designated selection resolution. Let $\downarrow_D(\cdot)$ denote isotropic resizing to satisfy $\max(H,W) \le D$ and let $\mathcal{N}(\cdot)$ denote grayscale+normalization. At selection scale D_{sel},

$$I_{\text{sel}} = \mathcal{N}(\downarrow_{D_{\text{sel}}}(I)), \qquad T_{k,\text{sel}} = \mathcal{N}(\downarrow_{D_{\text{sel}}}(T_k)). \tag{4}$$

To reduce computation, similarity is evaluated only within Ω. Denoting cropping by $\mathcal{C}_\Omega(\cdot)$, the selection objective is

$$k^* = \arg\max_{k \in \{1,\dots,K\}} \text{SSIM}\big(\mathcal{C}_\Omega(I_{\text{sel}}),\ \mathcal{C}_\Omega(T_{k,\text{sel}})\big), \tag{5}$$

and the chosen template is $T^* = T_{k^*}$.

To reduce selection time, a two-stage scheme is used: (i) a cheap preselect score at a smaller max dimension D_{pre} to shortlist K_s candidates and (ii) SSIM refinement computed only on that shortlist. Two suitable preselect scores are as follows:

$$\text{corr}(a,b) = \frac{\langle a-\bar{a},\ b-\bar{b}\rangle}{\|a-\bar{a}\|_2\,\|b-\bar{b}\|_2 + \epsilon}, \tag{6}$$

$$\text{mad}(a,b) = -\frac{1}{|\Omega|}\sum_{(u,v)\in\Omega} |a(u,v) - b(u,v)|, \tag{7}$$

where a, b are vectorized values in Ω after mean subtraction (corr) or are directly compared (mad). Correlation-style matching is a standard building block for template-based inspection and fast preselection [12,28,67]. Further details and implementation considerations (including complexity, shortlist logic, and caching of preprocessed templates to amortize decoding/resizing across queries) are provided in Appendix A.2.

2.3. Template Mapping and Defect-Likelihood Field

After selecting T^*, a defect-likelihood field $M(u,v)$ is constructed on a processing grid. To control runtime, expensive stages (alignment, SSIM-map, and morphology) may be downsampled by selecting a target long-side processed length D_{proc} and defining an isotropic scale

$$s = \min\left(1, \frac{D_{\text{proc}}}{\max(H_I, W_I, H_T, W_T)}\right). \tag{8}$$

This yields

$$I_s = \mathcal{N}(\downarrow_s (I_{\text{sel}})), \qquad T_s = \mathcal{N}(\downarrow_s (T^*)). \tag{9}$$

If needed, the query is resized to match the template grid. The implementation also supports resolution presets and scales morphology/area parameters consistently with the processing scale; see Appendix A.3.

https://doi.org/10.3390/s26051541

Although the proposed framework can operate without explicit alignment, in practical AOI settings a lightweight global alignment is highly recommended whenever small placement shifts or mild rotations are expected. Because the defect-likelihood field is computed pixel-wise, even minor global misalignment can produce widespread responses that resemble false defects.

Let τ_θ denote a parametric global warp (translation or rigid) with parameters θ, and let the warped query be $I_s \circ \tau_\theta$. The alignment is estimated by maximizing an intensity-based similarity inside the inner region Ω,

$$\hat{\theta} = \arg\max_{\theta} \ \rho\big(\mathcal{M}_\Omega(I_s \circ \tau_\theta),\ \mathcal{M}_\Omega(T_s)\big), \tag{10}$$

where ρ is a correlation-like similarity and $\mathcal{M}_\Omega$ denotes inner-region masking, i.e., pixels outside Ω are set to zero. More general geometric matching formulations exist (e.g., via epipolar-geometry constraints), but a lightweight global warp is sufficient under the fixture-controlled imaging assumed in AOI lines [68].

Because defect content can bias the estimated warp, the transform is not applied unconditionally. Instead, the candidate warp is validated using SSIM gain on the inner region,

$$\Delta = \mathrm{SSIM}(\mathcal{M}_\Omega(I_s \circ \tau_{\hat{\theta}}), \mathcal{M}_\Omega(T_s)) \tag{11}$$

$$- \mathrm{SSIM}(\mathcal{M}_\Omega(I_s), \mathcal{M}_\Omega(T_s)). \tag{12}$$

The alignment is accepted if $\Delta \geq \Delta_{\min}$ (a user-controlled safeguard); otherwise it is rejected. Accordingly, the aligned (or unaligned) query is defined as

$$\tilde{I} = \begin{cases} I_s \circ \tau_{\hat{\theta}}, & \Delta \geq \Delta_{\min}, \\ I_s, & \text{otherwise.} \end{cases} \tag{13}$$

This acceptance rule makes alignment a practical robustness aid while limiting over-correction in the presence of true defects.

To prevent outliers outside the central region from compressing the dynamic range used for thresholding, a min–max rescaling is used whose extrema are computed over Ω only. Define $X^\Omega_{\min} = \min_{(i,j)\in\Omega} X(i,j)$ and $X^\Omega_{\max} = \max_{(i,j)\in\Omega} X(i,j)$. The operator $\mathcal{R}_\Omega(X)$ therefore performs min–max normalization using extrema computed over pixels in Ω only, which prevents values outside Ω from compressing contrast in the region used for threshold selection. The stabilizer ϵ ensures numerical safety when $\max_\Omega X \approx \min_\Omega X$.

$$\mathcal{R}_\Omega(X)(u,v) = \begin{cases} \dfrac{X(u,v) - X^\Omega_{\min}}{X^\Omega_{\max} - X^\Omega_{\min} + \epsilon}, & (u,v) \in \Omega, \\ 0, & (u,v) \notin \Omega. \end{cases} \tag{14}$$

A local SSIM map is computed between $\tilde{I}$ and T_s. For a window centered at (u,v), let $\mu_{\tilde{I}}, \mu_T$ be local means, let $\sigma^2_{\tilde{I}}, \sigma^2_T$ be local variances, and let $\sigma_{\tilde{I}T}$ be the local covariance. The SSIM map value can be written as

$$S(u,v) = \frac{(2\mu_{\tilde{I}}\mu_T + C_1)(2\sigma_{\tilde{I}T} + C_2)}{(\mu^2_{\tilde{I}} + \mu^2_T + C_1)(\sigma^2_{\tilde{I}} + \sigma^2_T + C_2)}, \tag{15}$$

where C_1, C_2 stabilize the ratio. This is converted to a structural dissimilarity map,

$$D_{\mathrm{ssim}}(u,v) = 1 - S(u,v). \tag{16}$$

https://doi.org/10.3390/s26051541

In our implementation, the SSIM map is computed using MATLAB R2023a's SSIM function with explicitly specified parameters (window size/weighting and stabilizing constants) for reproducibility; the exact values are reported in Appendix A.4. An absolute-difference map is also computed

$$D_{\text{abs}}(u,v) = |\tilde{I}(u,v) - T_s(u,v)|. \tag{17}$$

An additive fusion with weight w is used, followed by rescaling for stable thresholding:

$$M_{\text{raw}}(u,v) = D_{\text{ssim}}(u,v) + w\,\mathcal{R}_\Omega(D_{\text{abs}})(u,v), \tag{18}$$

$$M(u,v) = \mathcal{R}_\Omega\big(M_{\text{raw}}(u,v)\big). \tag{19}$$

In this interpretation, $M(u,v)$ is a defect-likelihood signal: localized defects produce concentrated peaks, while nuisance global mismatch (wrong template) tends to produce broad responses.

2.4. Mask and Box Extraction

Defects are sparse, so a high-quantile threshold is chosen using pixels in Ω only:

$$\theta = Q_q\big(\{M(u,v) : (u,v) \in \Omega\}\big), \tag{20}$$

where $q \in (0,1)$ is a sensitivity parameter.

For a finite set of values $\{z_i\}_{i=1}^n$, the q-quantile Q_q is any value such that at least a fraction q of samples are $\leq Q_q$ and at least a fraction $(1-q)$ are $\geq Q_q$. Using a high quantile to set θ is robust in this setting because true defects typically occupy only a small fraction of pixels within Ω.

The initial binary mask is

$$Y_0(u,v) = \mathbb{1}[M(u,v) > \theta]\,\mathbb{1}[(u,v) \in \Omega]. \tag{21}$$

To suppress speckle noise and consolidate defect blobs, a sequence of set operators is applied to Y_0 within the inner region. Let $B(r)$ denote a disk structuring element of radius r, and let $\text{AreaOpen}(\cdot; A_{\min})$ remove connected components smaller than $A_{\min}$. The sequence is

$$Y_1 = \text{AreaOpen}(Y_0; A_{\min}), \tag{22}$$

$$Y_2 = Y_1 \bullet B(r_c), \tag{23}$$

$$Y_3 = Y_2 \oplus B(r_d), \tag{24}$$

$$Y = Y_3 \wedge \mathbb{1}_\Omega, \tag{25}$$

where $\bullet$ is morphological closing, $\oplus$ is dilation, and $\wedge$ is pixel-wise AND with the inner-region mask $\mathbb{1}_\Omega$.

Connected components of Y are converted to axis-aligned bounding boxes $b = [x\ y\ w\ h]$. Each component is assigned a confidence score using a robust statistic of M inside the component. A stable choice is a high quantile:

$$s_i = Q_{q_s}\big(\{M(u,v) : (u,v) \in \mathcal{C}_i\}\big), \qquad q_s \in (0,1). \tag{26}$$

Boxes are sorted by s_i and mapped back to the original resolution if the processing scale $s < 1$. For the exact connected-component scoring rule and supported alternatives

https://doi.org/10.3390/s26051541

(e.g., quantile/max/mean scoring), see Appendix A.5. Therefore, the final output of this stage is a confidence-ranked set of detections

$$\mathcal{P} = \{(b_i, s_i)\}_{i=1}^{N}, \tag{27}$$

where each $b_i = [x\ y\ w\ h]$ is an axis-aligned bounding box, and s_i is its confidence score derived from the defect-likelihood field. In summary, the mask-and-box stage is controlled by a small set of interpretable parameters: threshold quantile q, minimum area $A_{\min}$, closing radius r_c, dilation radius r_d, and confidence quantile q_s (default $q_s = 0.95$). These parameters jointly control the false-positive/false-negative trade-off and the coarseness of the resulting boxes.

2.5. Algorithm Summary

The proposed framework takes a defective query image and selects a reference template from a small library using a coarse-to-fine strategy (cheap pre-screening followed by SSIM refinement). If enabled, a global alignment is applied with an explicit acceptance test based on SSIM gain on the inner region, limiting transforms biased by defect content. A defect-likelihood field is then constructed by fusing an SSIM-derived structural dissimilarity map with a normalized absolute-difference map and applying an Ω-based rescaling. Finally, high-quantile thresholding and simple morphological post-processing yield connected components that are converted into confidence-ranked bounding boxes. Algorithm 1 summarizes the full procedure, while the mathematical operators and data products are summarized in Figure 2.

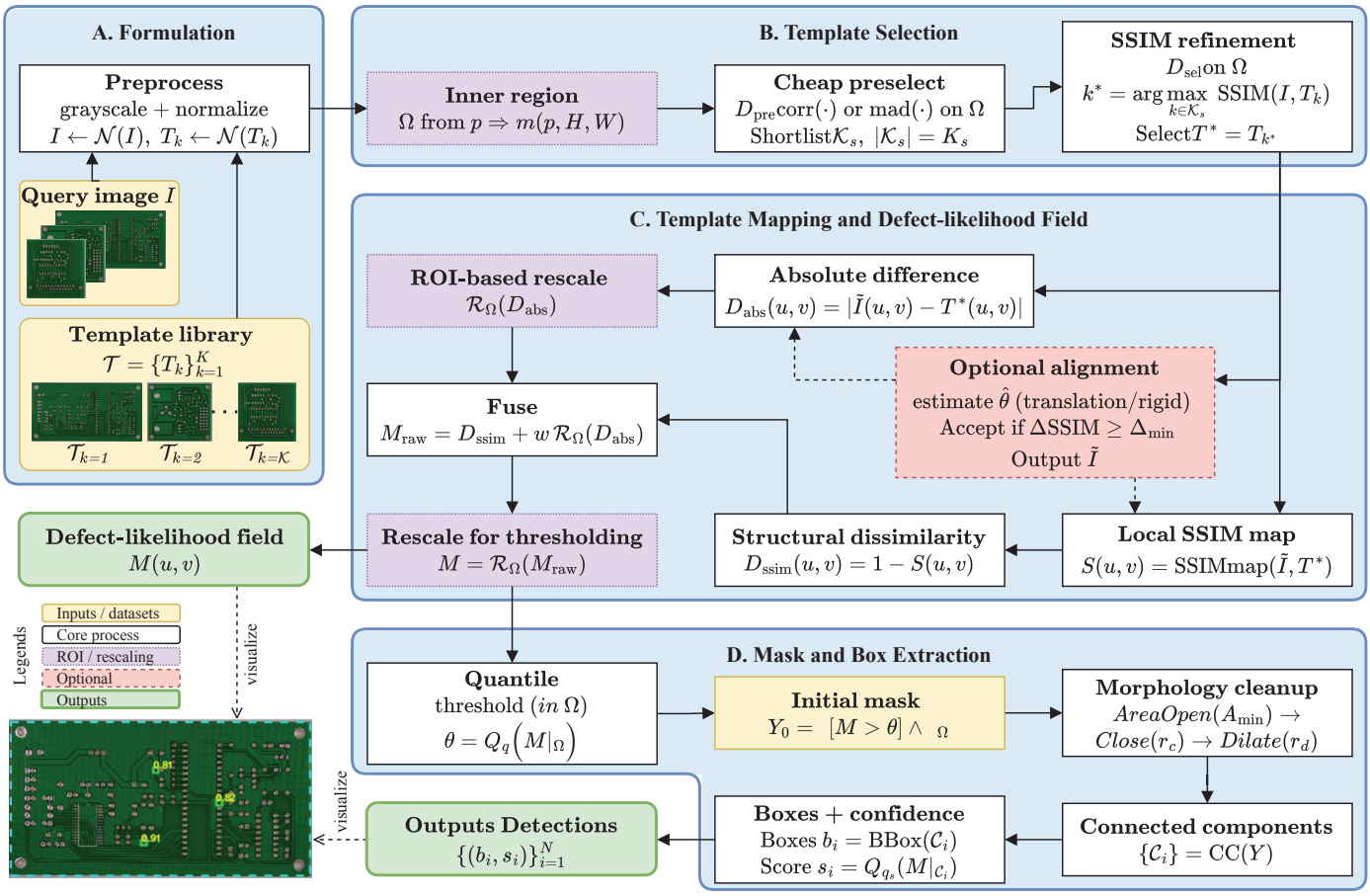


Figure 2. Mathematical overview of the proposed image processing framework.

https://doi.org/10.3390/s26051541

Algorithm 1 Template selection and defect localization

Require: Query image I, templates $\{T_k\}_{k=1}^{K}$,
parameters $(p, w, q, q_s, A_{\min}, r_c, r_d, D_{\text{pre}}, D_{\text{sel}}, D_{\text{proc}}, K_s, \Delta_{\min})$
Ensure: Detections $\mathcal{P} = \{(b_i, s_i)\}$ and defect-likelihood field M
Compute central region Ω from area-avoidance ratio p
Preselect shortlist $\mathcal{K}_s$ (size K_s) using a cheap score at resolution D_{pre}
Select template T^* by maximizing SSIM on Ω over $k \in \mathcal{K}_s$ at resolution D_{sel}
Set processing scale s from D_{proc} and build processed images (I_s, T_s)
Optional alignment: estimate transform $\hat{\theta}$ and compute SSIM gain Δ on Ω; accept if $\Delta \geq \Delta_{\min}$
Define aligned query $\tilde{I}$ (either warped I_s or I_s)
Compute SSIM map $S(u, v)$ between $\tilde{I}$ and T_s
Compute dissimilarity maps $D_{\text{ssim}}(u, v) = 1 - S(u, v)$ and $D_{\text{abs}}(u, v) = |\tilde{I}(u, v) - T_s(u, v)|$
Fuse and rescale to obtain defect-likelihood field M
Threshold: $\theta = Q_q(\{M(u, v) : (u, v) \in \Omega\})$
Initial mask: $Y_0 = \mathbb{1}[M > \theta] \wedge \mathbb{1}_{\Omega}$
Cleanup: $Y = \text{AreaOpen}(Y_0; A_{\min}) \bullet B(r_c) \oplus B(r_d)$
Connected components $\{\mathcal{C}_i\} \to$ boxes $\{b_i\}$ and scores $s_i = Q_{q_s}(\{M(u, v) : (u, v) \in \mathcal{C}_i\})$
Map boxes back to original scale (if $s < 1$)

Computational Complexity

Let K be the number of available templates and $K_s \ll K$ be the shortlist size after preselection. Let $D_{\text{pre}}, D_{\text{sel}}, D_{\text{proc}}$ denote the maximum long-side resolutions used for (i) the cheap preselect score, (ii) SSIM refinement for template selection, and (iii) the downstream processing grid, respectively. Using the common square-grid proxy, the number of evaluated pixels scales as $\mathcal{O}(D^2)$ at resolution D. Restricting computations to the central region Ω (via ROI cropping/masking) primarily changes constants. Per query image, the coarse-to-fine template selection cost is

$$\mathcal{O}\big(K\, D_{\text{pre}}^2 + K_s\, D_{\text{sel}}^2\big), \tag{28}$$

where the first term corresponds to evaluating a cheap preselect score over all templates at D_{pre}, and the second term corresponds to SSIM refinement over the shortlist at D_{sel} (dominant when enabled). After selecting T^*, the downstream stages operate on the processing grid at D_{proc}: (i) optional global alignment uses a correlation-based registration objective (translation/rigid) and in practice scales near-linearly in the number of processed pixels (often with FFT-accelerated correlation), (ii) the SSIM map uses a fixed window size (constant window operations per pixel), and (iii) difference-map construction, Ω-based rescaling, quantile thresholding, morphology with fixed structuring-element radii, and connected-components labeling are all $\mathcal{O}(D_{\text{proc}}^2)$ up to small multiplicative constants. Overall, the per-image runtime scales as

$$\mathcal{O}\big(K\, D_{\text{pre}}^2 + K_s\, D_{\text{sel}}^2 + D_{\text{proc}}^2\big), \tag{29}$$

and the dataset runtime is linear in the number of evaluated images. In practice, template preprocessing (grayscale/normalization/resizing) can be cached, and choosing $D_{\text{pre}} \ll D_{\text{sel}}$ with $K_s \ll K$ keeps selection tractable even for moderate template libraries.

2.6. Experimental Setup and Evaluation Protocol

Experiments use the PKU-Market-PCB dataset [3]. Defect classes are used only for class-wise reporting; the localization pipeline itself is class-agnostic across defect categories (while assuming a valid defect-free reference template for the same layout/revision). The

https://doi.org/10.3390/s26051541

dataset folder organization (templates, defective images, and Pascal Visual Object Classes (VOC) annotations) is described in Appendix A.6. The proposed method is implemented in MATLAB R2023a, and experiments are run on an Apple M1 processor with 8GB RAM. The implementation configuration and hyperparameters are reported in Appendix A.7

2.6.1. Dataset

Although the proposed method employs training-free image processing that produces an explicit defect-likelihood field $M(u, v)$ and interpretable intermediate maps, the final outputs are evaluated using standard box-level localization metrics, which facilitates comparison with established benchmarks in the AOI and object-detection literature. For broader context on automated optical inspection in electronics manufacturing and recent perspectives on PCB defect detection, see the comprehensive surveys in [69,70]. Table 1 summarizes the evaluated dataset, including image resolution and annotation statistics.

Table 1. Dataset summary.

Item	Value
Defective images (run)	693
Evaluated images (with GT XML)	693
Defect-free templates	10
Defect classes	6
GT objects (total)	2953
Median resolution (W × H in pixels)	2868 × 2159
Resolution range (W × H in pixels)	2240–3056 × 1586–2530

2.6.2. Detections and Confidence

For each query image, the method outputs a set of axis-aligned detections $\mathcal{P} = \{(b_i, s_i)\}_{i=1}^{N_P}$, where each $b_i = [x\ y\ w\ h]$ is the bounding box of a connected component extracted from a thresholded version of $M(u, v)$ (Section 2.4). Each detection is assigned a confidence score s_i computed as a robust statistic of M values inside the component (a high-quantile score is used so that a higher s_i indicates stronger evidence of a localized peak). For the exact component-to-box confidence definition used in the implementation, see Appendix A.5.

2.6.3. Ground Truth and Intersection-over-Union (IoU) Matching

Ground-truth annotations are axis-aligned Pascal VOC boxes. For each query image, let $\mathcal{G} = \{g_j\}_{j=1}^{N_G}$ denote the set of ground-truth boxes. Detections are matched to ground truth using a greedy one-to-one assignment within each image: detections are processed in descending confidence order, and a detection b_i is matched to an as-yet-unmatched g_j if $\mathrm{IoU}(b_i, g_j) \geq \tau_{\mathrm{IoU}}$ (each detection matches at most one ground-truth box and vice versa). Unmatched detections are counted as false positives, and unmatched ground-truth boxes are counted as false negatives.

Unless noted otherwise, $\tau_{\mathrm{IoU}} = 0.10$ is used because detections are derived from thresholded connected components rather than tight box regression.

2.6.4. Fixed Operating Point (Counts and F1)

Given a score threshold t, detections with $s_i \geq t$ are retained and dataset-level counts (after per-image one-to-one matching) are computed: true positives $\mathrm{TP}(t)$, false positives $\mathrm{FP}(t)$, and false negatives $\mathrm{FN}(t)$. Precision, recall, and F1 are

https://doi.org/10.3390/s26051541

$$\text{Precision}(t) = \frac{\text{TP}(t)}{\text{TP}(t) + \text{FP}(t) + \epsilon}, \tag{30}$$

$$\text{Recall}(t) = \frac{\text{TP}(t)}{\text{TP}(t) + \text{FN}(t) + \epsilon}, \tag{31}$$

$$\text{F1}(t) = \frac{2\,\text{Precision}(t)\,\text{Recall}(t)}{\text{Precision}(t) + \text{Recall}(t) + \epsilon}. \tag{32}$$

where $\epsilon > 0$ is a small numerical stabilizer (e.g., 10^{-12}) used only to avoid division-by-zero in degenerate cases (e.g., when $\text{TP}(t) + \text{FP}(t) = 0$); it does not affect results when denominators are nonzero.

For completeness, we also report a detection accuracy that does not require defining true negatives (TN), which are not well-defined for variable-length detection outputs:

$$\text{Accuracy}(t) = \frac{\text{TP}(t)}{\text{TP}(t) + \text{FP}(t) + \text{FN}(t) + \epsilon}. \tag{33}$$

2.6.5. Avoiding Threshold Selection on the Reported Set

To avoid optimistic bias from selecting t on the same set used for reporting, operating-point metrics are reported using K-fold cross-validation: in each fold, the score threshold t^* is chosen to maximize $\text{F1}(t)$ on the calibration split (the union of the other $K-1$ folds), and precision/recall/F1 are then computed on the held-out fold using this fixed threshold. This work reports the mean $\pm$ standard deviation across folds. For continuity with the original submission and to aid interpretation of class-wise behavior, this also includes a reference operating point obtained by selecting t^* on the full evaluated set; this reference is explicitly labeled and is not used as the primary estimate of generalization performance. Implementation details for the cross-validated operating point selection are provided in Appendix B.1.

2.6.6. Threshold-Sweep Metrics (PR/AP, AP-Versus-IoU, and FROC)

To summarize performance across confidence thresholds, the precision–recall curve is constructed by ranking all detections across the dataset by confidence s_i and sweeping down this global ranked list. True/false positives are determined using the matching protocol described above at the specified τ_{IoU} [71,72]. Average precision (AP) is computed from this PR curve using a precision envelope and piecewise integration over recall (the VOC07 11-point approximation is disabled in the evaluation script).

To quantify sensitivity to localization stringency, AP is also computed across multiple IoU thresholds. In particular, (i) a custom sweep $\tau_{\text{IoU}} \in \{0.10, 0.20, \ldots, 0.90\}$ and (ii) a COCO-style sweep $\tau_{\text{IoU}} \in \{0.50, 0.55, \ldots, 0.95\}$ are reported. When multiple IoU thresholds are considered, mean Average Precision mAP is reported as the mean AP across the specified thresholds.

For operational analysis, free-response receiver operating characteristic (FROC) curves are reported by sweeping score thresholds. For a threshold t, let N_{img} be the number of evaluated images, let N_{GT} be the total number of ground-truth boxes, and let $N_{\text{img}}^{\text{GT}}$ be the number of images containing at least one ground-truth box. The following quantities are defined:

https://doi.org/10.3390/s26051541

$$\mathrm{FP/image}(t) = \frac{\mathrm{FP}(t)}{N_{\mathrm{img}} + \epsilon}, \tag{34}$$

$$\mathrm{Recall}_{\mathrm{obj}}(t) = \frac{\mathrm{TP}(t)}{N_{\mathrm{GT}} + \epsilon}, \tag{35}$$

$$\mathrm{Recall}_{\mathrm{img}}(t) = \frac{N_{\mathrm{img}}^{\mathrm{TP}}(t)}{N_{\mathrm{img}}^{\mathrm{GT}} + \epsilon}, \tag{36}$$

where $N_{\mathrm{img}}^{\mathrm{TP}}(t)$ is the number of ground-truth-containing images for which at least one detection is a matched true positive (TP) at threshold t. In this paper, FROC plots and operating-point tables use object-level recall $\mathrm{Recall}_{\mathrm{obj}}(t)$ versus $\mathrm{FP/image}(t)$; image-level recall is included for completeness. The exported threshold grid and image-level recall computation used to generate the FROC bundle are described in Appendix B.2.

At the dataset level, $N_{\mathrm{GT}} = \mathrm{TP}(t) + \mathrm{FN}(t)$; therefore, $\mathrm{Recall}(t)$ and $\mathrm{Recall}_{\mathrm{obj}}(t)$ are numerically equivalent. Accordingly, $\mathrm{Recall}_{\mathrm{obj}}(t)$ is used in the FROC definition to emphasize the object-level nature of the curve.

For completeness, explicit definitions for IoU on $[x\ y\ w\ h]$ boxes and the AP computation used in our evaluation are provided in Appendices B.3 and B.4.

3. Experimental Results and Analysis

This section reports qualitative examples and quantitative localization performance under the experimental setup and evaluation protocol described in the previous subsection. The method produces confidence-ranked detections $\mathcal{P} = \{(b_i, s_i)\}$ from an explicit defect-likelihood field $M(u, v)$; therefore, results are presented both at a fixed operating point (counts and precision/recall/F1) and across confidence thresholds (precision–recall, AP, and FROC).

All quantitative results in this section are reported on 693 evaluated query images under the adopted IoU matching protocol (Section 2.6). Threshold-sweep curves (PR/AP/FROC) are computed on the full set. For fixed-threshold operating-point metrics, this work reports K-fold cross-validated results (mean $\pm$ std) to avoid selecting thresholds on the same set used for reporting.

For reference and continuity with prior results, the manuscript also reports descriptive counts at a reference best-F1 threshold selected on the full evaluated set (explicitly labeled as reference).

3.1. Qualitative Results and Failure Modes

Figure 3 shows representative TP cases where the proposed mapping produces localized peaks in $M(u, v)$ that translate into axis-aligned bounding boxes overlapping the annotated defects. While the boxes are not optimized to tightly fit the defect boundary (they are derived from thresholded connected components), they consistently localize the defect region and provide a monotonic confidence score s_i for ranking.

Figure 4 shows representative false positive (FP) cases. In practice, FPs tend to arise from nuisance variability that produces structured responses in the SSIM/difference maps, such as residual global misalignment, illumination drift, or local texture/reflectance changes that are not labeled as defects. A qualitative review of these cases suggests that many FPs stem from two primary sources: (i) structured nuisance variability such as residual global misalignment, illumination drift [73], or local texture changes that are not true defects, and (ii) actual defects that are present in the images but are missing from the corresponding Pascal VOC XML annotation files. Because this work operates strictly under the provided ground-truth labels without resorting to relabeling or manual

https://doi.org/10.3390/s26051541

verification, all reported metrics faithfully reflect the adopted box-matching protocol and evaluation set. The presence of unlabeled defects in the FP set underscores the importance of careful annotation in benchmark datasets and suggests that true recall may be higher than reported. These observations underscore the value of the optional alignment safeguard and careful template selection, which are most influential in reducing nuisance-driven artifacts. Moreover, the confidence-ranked outputs and FROC curves (Section 3.6) enable operational threshold selection based on a tolerable false-alarm budget in production, allowing practitioners to adjust operating points according to their acceptable FP/image rate and trade recall for lower false-alarm burden when necessary.

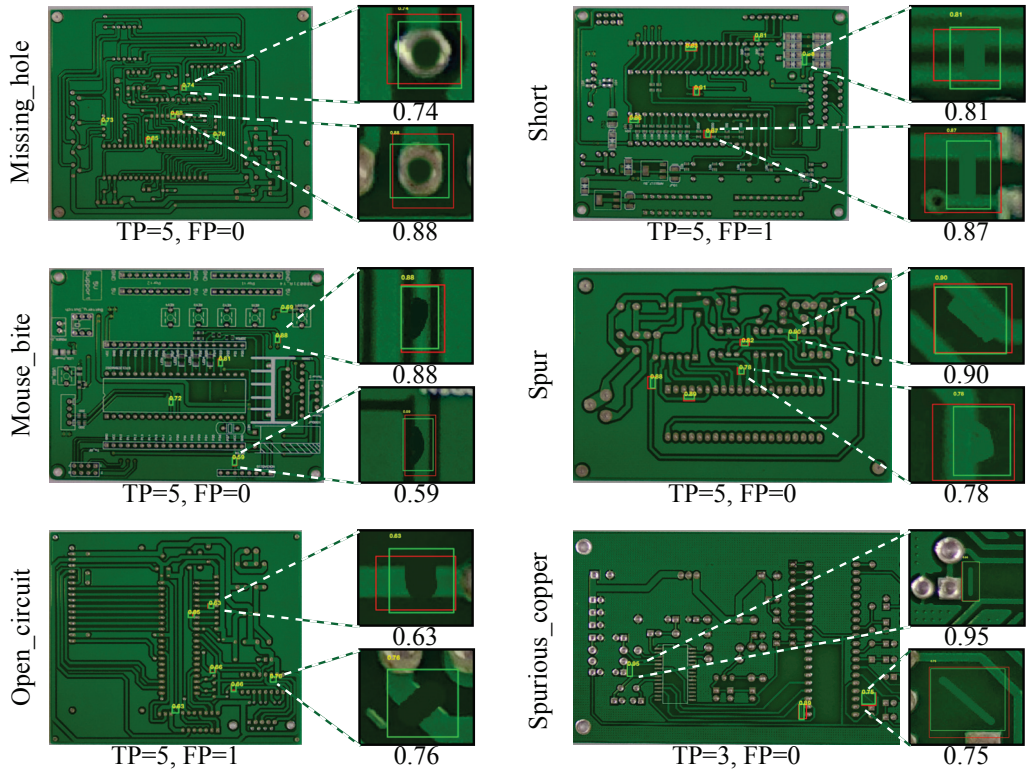


Figure 3. Example true positive cases (red: ground truth; green: proposed detection).

Figure 5 provides representative visual evidence for the observed Spur false negatives (all four FNs in the per-class breakdown). These cases are consistent with thin/fragmented structures producing a weaker or spatially discontinuous response under the current high-quantile thresholding and morphology settings, which can suppress small components or yield boxes that fail the one-to-one IoU matching criterion. Because the pipeline is deterministic and controlled by interpretable parameters $(q, A_{\min}, r_c, r_d, q_s)$, this behavior reflects an explicit operating trade-off that can be adjusted (as also summarized by the FROC curve in Section 3.6).

https://doi.org/10.3390/s26051541

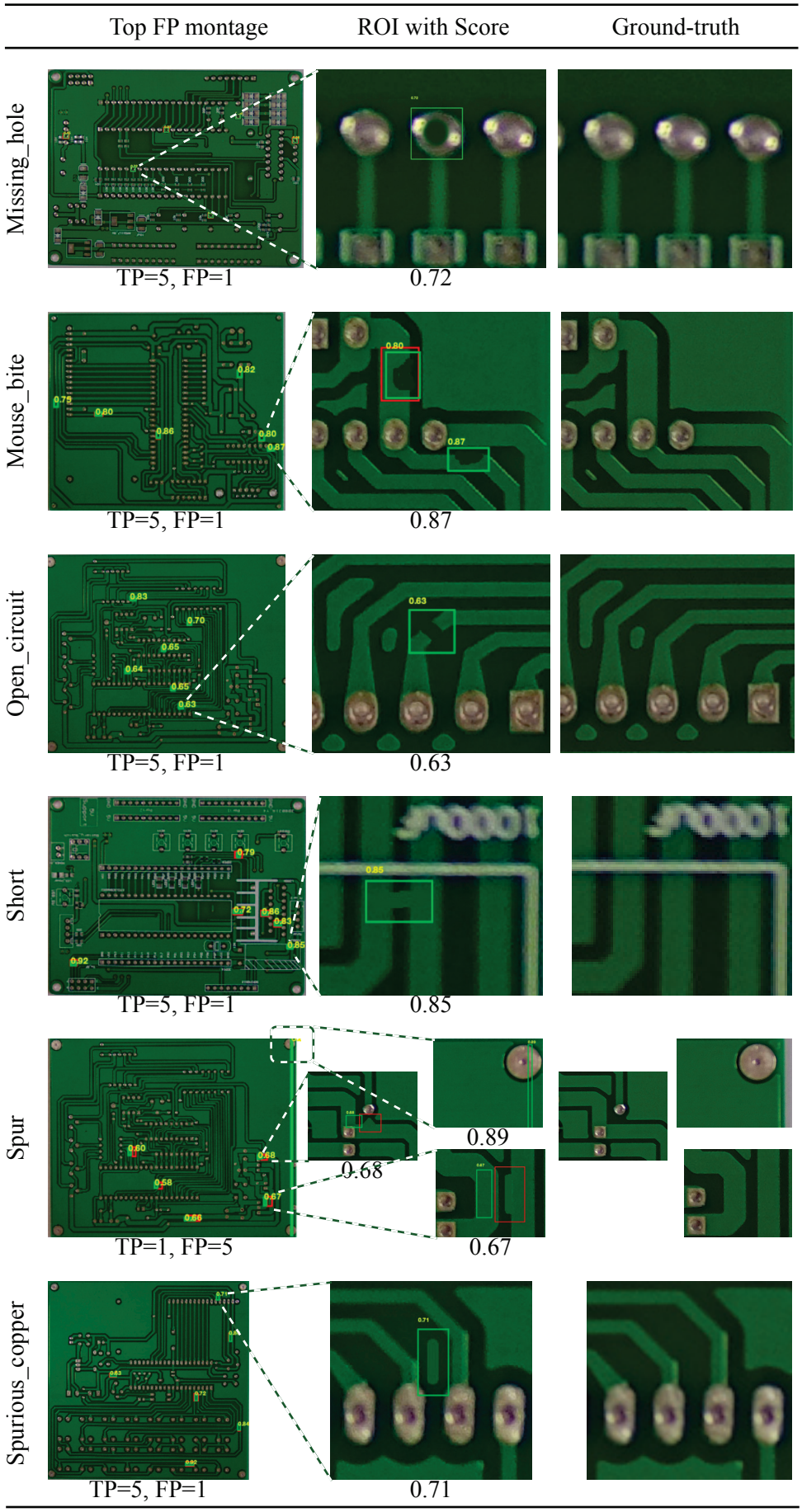


Figure 4. Example false positive cases (red: ground truth; green: proposed detection).

https://doi.org/10.3390/s26051541

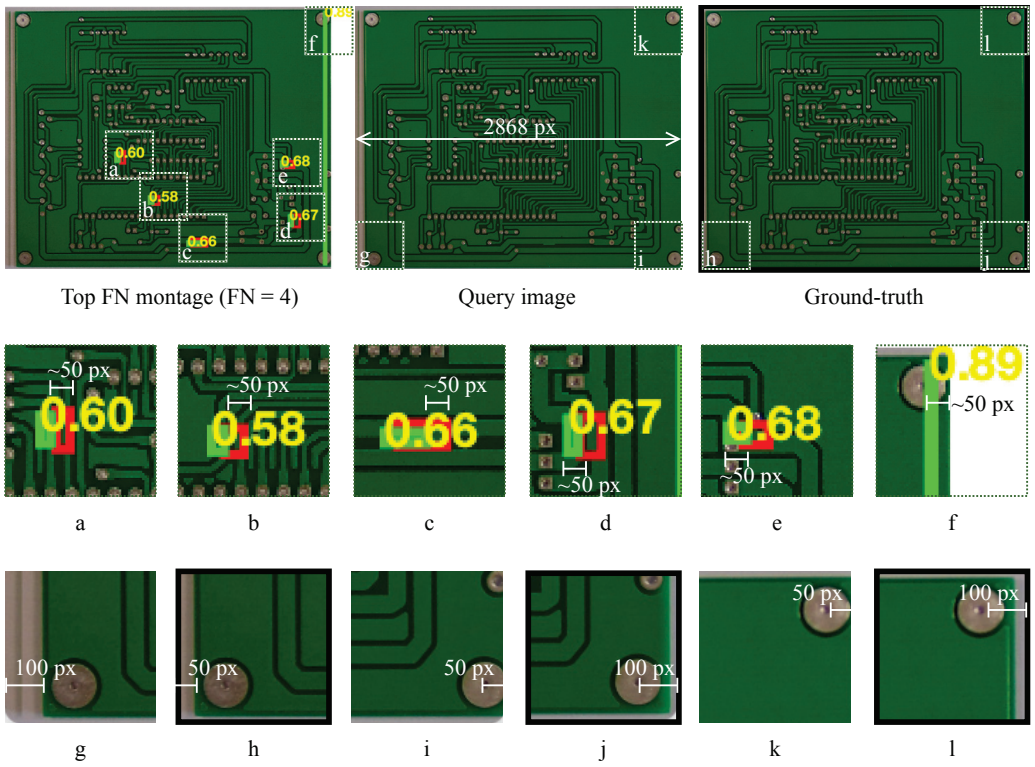


Figure 5. Example false negative case (red: ground truth; green: proposed detection; (**a**–**f**) are corresponding crop regions of the Top FN montage; (**g**,**i**,**k**) are crop regions of the query image; (**h**,**j**,**l**) are crop regions of the reference ground-truth image).

3.2. Template Selection Quality and Practical Failure Causes

The exported per-image diagnostics show that template selection is typically highly stable (template SSIM values are near 1.0 for the majority of images), but rare outliers can occur, which are consistent with occasional template mismatch or nontrivial capture variation. On the 693 evaluated images, the per-image template SSIM has median ≈ 0.9991 (p5 ≈ 0.9976), while a small number of outliers fall below 0.99 (4/693 images) and one extreme case reaches as low as ≈ 0.4967. These outliers are practically important because low similarity can inflate structured responses in $M(u, v)$ and therefore increase false positives under fixed post-processing parameters.

3.3. Quantitative Summary at the Adopted Operating Point

Table 2 reports per-class detection counts (TP/FP/FN) and the resulting precision/recall/F1 under the adopted IoU matching protocol at a reference best-F1 operating point selected on the full evaluated set. Overall, the method achieves TP = 2949, FP = 103, FN = 4, yielding Precision = 0.9663, Recall = 0.9987, and F1 = 0.9822 at this reference threshold. Cross-validated operating-point results (mean $\pm$ std across folds), which avoid threshold selection on the reported set, are provided in Table 3. In this revision, we treat these cross-validated operating-point results as the primary generalization-oriented summary, while the full-set best-F1 operating point is retained only as a clearly labeled reference point.

Across defect types, performance is broadly consistent. The main deviations are informative: all four missed defects (FN = 4) occur in the Spur class, suggesting a small subset of cases where the defect response is weaker or more fragmented under the current post-processing parameters. This behavior is expected in a deterministic pipeline and can be addressed by adjusting the interpretable parameters $(q, A_{\min}, r_c, r_d, q_s)$ for the desired false-positive/false-negative trade-off.

https://doi.org/10.3390/s26051541

Table 2. Reference operating-point detection counts and metrics (threshold selected on the full evaluated set).

Class	TP	FP	FN	Prec.	Rec.	F1	Acc.
Missing_hole	497	22	0	0.9576	1.0000	0.9784	0.9576
Mouse_bite	492	20	0	0.9609	1.0000	0.9801	0.9609
Open_circuit	482	8	0	0.9837	1.0000	0.9918	0.9837
Short	491	18	0	0.9646	1.0000	0.9820	0.9646
Spur	484	19	4	0.9622	0.9918	0.9768	0.9546
Spurious_copper	503	16	0	0.9692	1.0000	0.9843	0.9692
Overall	2949	103	4	0.9663	0.9987	0.9822	0.9650

Table 3. Cross-validated operating-point metrics (mean ± std across folds).

Metric	Mean ± Std	Notes
Threshold	0.322 ± 0.053	K = 5, IoU = 0.10
Precision	0.980 ± 0.008	
Recall	0.997 ± 0.003	
F1	0.988 ± 0.004	
FP per image	0.088 ± 0.035	

3.4. Precision–Recall Characteristics Across Confidence Thresholds

Because detections are confidence-ranked by s_i, varying the acceptance threshold induces a precision–recall (PR) trade-off. Figure 6 reports the overall PR curve and the per-class PR curves, summarizing how precision changes as recall increases and which defect types degrade earlier as the threshold is lowered.

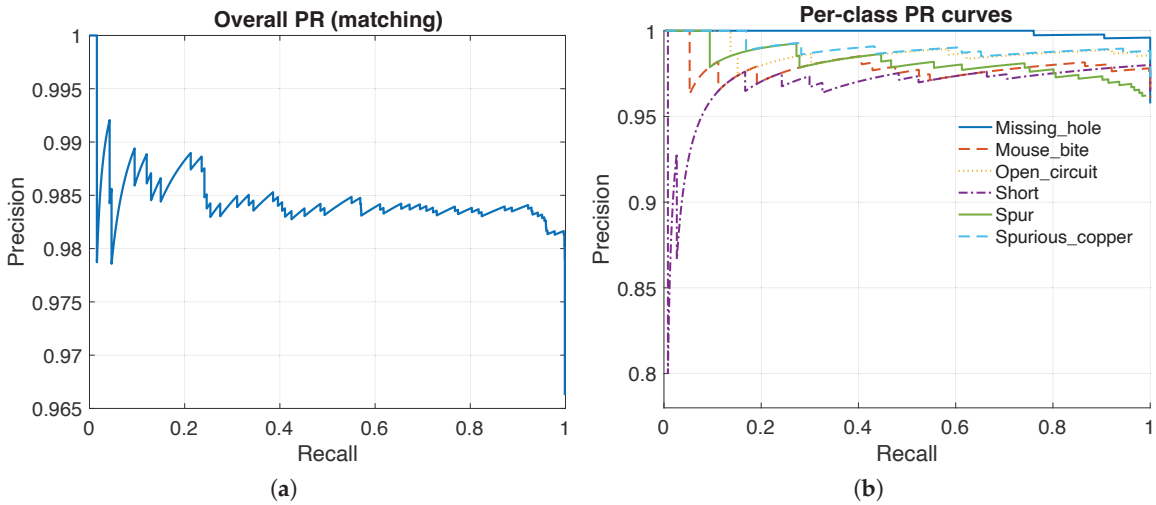


Figure 6. Precision–recall curves under the IoU matching protocol. (**a**) Overall PR curve. (**b**) Per-class PR curves.

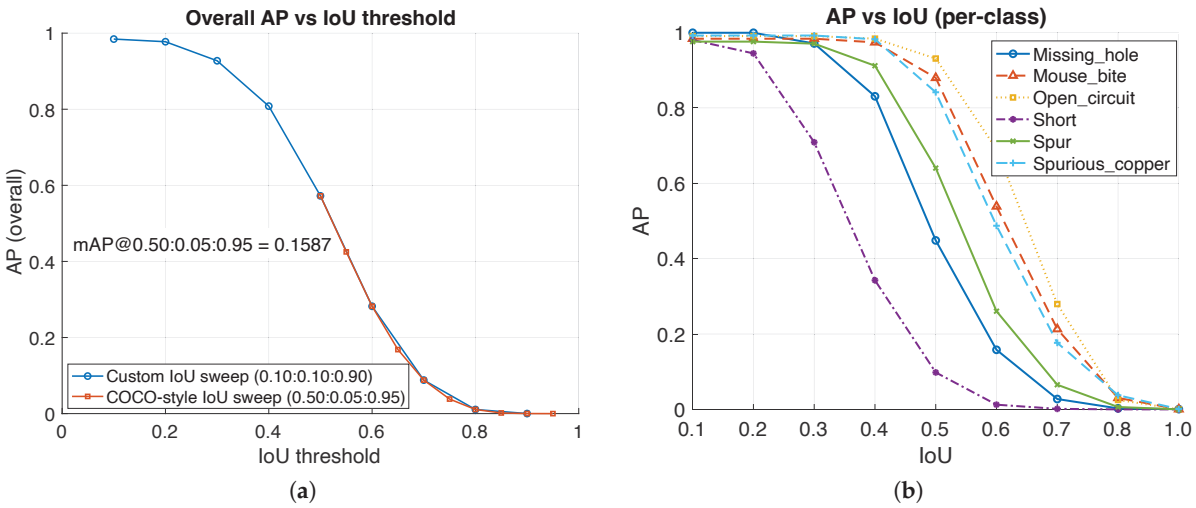
In AOI practice, this view is useful because the operating point is chosen based on the acceptable false-alarm burden in production. In addition, the PR curves validate that the proposed confidence score is informative: a meaningful ranking yields a PR curve that maintains high precision over a wide recall range.

3.5. Sensitivity to Matching Stringency (AP Versus IoU)

Figure 7 reports average precision as a function of the IoU threshold. As the IoU requirement becomes stricter, AP decreases because the proposed detections are derived

https://doi.org/10.3390/s26051541

from thresholded map components and are not explicitly optimized to tightly fit the ground-truth boxes. In particular, because our boxes are obtained by extracting connected components and enclosing them with axis-aligned rectangles, $\tau_{\mathrm{IoU}} = 0.10$ should be interpreted as a permissive matching criterion for region-level localization (component fragmentation/merging can yield looser boxes even when the detected region overlaps the defect). Consequently, the decline in AP at higher IoU thresholds primarily reflects box tightness/placement mismatch rather than an inability to detect defect regions. Quantitatively, the exported AP–IoU sweep shows high AP at the adopted low-stringency regime (overall AP ≈ 0.984 at IoU = 0.10) but a pronounced decline at stricter thresholds (overall AP ≈ 0.572 at IoU = 0.50), approaching near-zero at very strict IoU (e.g., $\geq$0.80). This plot is therefore included to transparently characterize localization-stringency sensitivity rather than to claim pixel-tight bounding-box regression.



(**a**) (**b**)

Figure 7. AP versus IoU threshold. (**a**) Overall AP versus IoU threshold, including custom IoU sweep and optional COCO-style mAP summary. (**b**) Per-class AP versus IoU threshold.

For completeness, Table 4 first summarizes representative overall AP values at selected IoU thresholds, and Figure 7 then visualizes the full AP-versus-IoU behavior, including both the overall curves (custom and COCO-style sweeps) and the per-class curves.

Table 4. AP at representative IoU thresholds (overall).

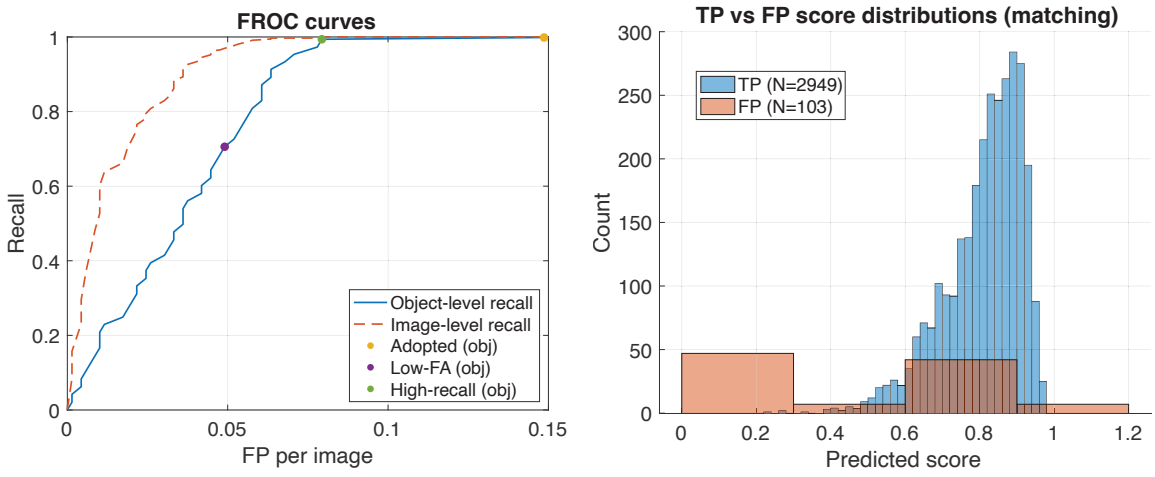
AP	0.05	0.10	0.20	0.30	0.40	0.50
IoU	0.985	0.984	0.977	0.927	0.808	0.572

3.6. False-Alarm Rate Versus Recall (Or FROC)

For deployment, a key question is how many false alarms must be tolerated to achieve a target recall. Figure 8 reports FROC curves, which summarize recall as a function of the false positive rate (e.g., false positives per image) as the confidence threshold varies. This view complements PR curves by emphasizing the operational cost of additional recall.

At the adopted operating point (Table 2), the system attains Recall = 0.9987 with FP/image ≈ 0.149. In practice, Figure 8 is used to choose an operating threshold based on an allowed FP/image budget; the exported FROC curve enables selecting alternative thresholds to trade a small recall reduction for a lower false-alarm burden.

https://doi.org/10.3390/s26051541



(**a**) (**b**)

Figure 8. Operational views for threshold selection. (**a**) Free-response receiver operating characteristic FROC curves. (**b**) True positive and false positive score distributions histograms.

3.7. Score Separability of True and False Detections

Figure 8 compares the confidence-score distributions of true positives and false positives. A clear separation indicates that the confidence score s_i provides a meaningful ranking signal, enabling threshold selection to trade precision for recall. Overlap between TP and FP score distributions highlights ambiguous cases (e.g., residual misalignment or illumination-driven artifacts) that can be addressed by improving template selection, enabling alignment when appropriate, or adjusting the mask-and-box parameters.

3.8. Quantitative Runtime Breakdown

Beyond asymptotic scaling, an empirical runtime breakdown of the reference implementation by stage is reported in Table 5.

Table 5. Runtime breakdown by stage (reference implementation).

Stage	Time (s/Image)
Template selection (total)	0.242
Preselect	0.109
SSIM refine	0.033
Localization (total)	0.793
Preprocess	0.043
Resize	0.059
Align	0.368
Score map	0.048
Mask	0.073
Boxes	0.004
Total	1.035

4. Discussion

The proposed framework targets practical AOI where defect-free references are available and decisions must be reliable, auditable, and deployable without a training lifecycle. Operating-point metrics are summarized using K-fold cross-validation to avoid selecting thresholds on the reported set (Table 3), while a reference best-F1 operating point on the full set is retained for descriptive continuity (Table 2). Under the adopted IoU matching protocol, confidence-ranked outputs support operational interpretation via PR/FROC sweeps (Figures 6 and 8) and IoU-sensitivity via AP-versus-IoU analysis (Figure 7).

https://doi.org/10.3390/s26051541

4.1. Why the Method Works in Practice

Robustness comes from explicitly controlling nuisance variability and using structure-aware comparison: template selection reduces systematic mismatch, optional alignment is gated by SSIM gain to avoid defect-driven over-correction, and the fused defect-likelihood field combines SSIM-derived structural dissimilarity with normalized absolute difference to handle both structural and radiometric changes (e.g., illumination drift). Together, these choices yield a confidence score that is informative for ranking, consistent with the PR and FROC behavior.

4.2. Class-Wise Behavior and Failure Modes

Per-class performance is broadly consistent; the main deviation is that the only missed defects (FN = 4) occur in the Spur class (Table 2), where responses can be weaker or more fragmented under the current post-processing. This is expected in a deterministic connected-components-to-box stage (threshold/area/morphology can suppress fine structures), and recall can be increased by adjusting $(q, A_{\min}, r_c, r_d, q_s)$ with the resulting FP/image quantified by the FROC curve.

4.3. Operating-Point Selection and Localization Stringency

Operating-point metrics are reported using *K*-fold cross-validation to avoid selecting thresholds on the same set used for reporting (Section 2.6). In practice, deployment typically selects thresholds based on an allowable false-alarm budget; Figure 8 and Table 6 make this trade-off explicit. Accordingly, FROC (object-level recall versus FP/image) is recommended as the primary operational view, with PR as a complementary precision summary. The adopted matching threshold $\tau_{\text{IoU}} = 0.10$ reflects that detections are derived from thresholded connected components rather than tight box regression; consequently, AP decreases as IoU becomes stricter (Figure 7), indicating a design choice favoring reliable region localization for QC over pixel-tight boxes. To avoid over-interpreting performance under low-stringency matching, the Results section summarizes AP at representative moderate IoU thresholds and provides the full AP-versus-IoU sweep. If tighter boxes are required, a lightweight post-processing refinement (e.g., contour-based tightening) can be added without changing the detection ranking.

Table 6. Example operating points derived from the exported FROC curve (object-level recall versus FP/image).

Operating Point	FP/Image	Recall	Notes
Adopted threshold	0.149	0.9987	From Table 2
Low-false-alarm	0.049	0.7057	scoreThr = 0.7712 (max recall with FP/image $\leq$ 0.05)
High-recall	0.079	0.9939	scoreThr = 0.4441 (min FP/image with recall $\geq$ 0.99)

4.4. Limitations and Practical Considerations

The method assumes defect-free templates of the same PCB layout/revision, which is realistic in many manufacturing settings and helps resolve defect-versus-intent ambiguity; however, this assumption also defines the primary limitations:

- Layout/revision mismatch or wrong template: if the selected template does not correspond to the same board revision (or if no valid template exists), reference comparison can yield structured false-positive regions that mimic defects. Coarse-to-fine template selection mitigates mild mismatch, but it cannot correct true design changes; in such

https://doi.org/10.3390/s26051541

cases, the correct remedy is updating the template library (or enforcing recipe-based template selection in fixed-product lines).

- Perspective/projective distortion and nonrigid deformation: the SSIM-gated alignment improves robustness to small global shifts/rotations, but larger viewpoint/perspective changes remain out of scope. When the imaging geometry is not fixture-controlled, residual misregistration can dominate difference/SSIM maps and reduce localization quality.
- Heavy illumination non-uniformity and reflectance drift: normalization and the fused SSIM/difference design reduce sensitivity to global brightness drift; however, strong spatially varying illumination (e.g., vignetting, hotspots, and specular highlights) can still introduce structured responses in the defect-likelihood field and increase false positives.
- ROI-based statistics and border effects: evaluating thresholds and similarity on the central region Ω improves robustness to border artifacts, but defects near the excluded border can have reduced influence on thresholding statistics. If border defects are critical, the ROI margin should be reduced (with the expected trade-off of higher nuisance sensitivity).
- Component-derived boxes are not tight boundaries: detections are produced from thresholded connected components, so bounding boxes are intended for reliable region localization rather than pixel-tight boundaries. This is why AP drops at stricter IoU thresholds (Figure 7); if tighter boxes are required, a lightweight contour-based tightening step can be added without changing the detection ranking.

Finally, this revision does not include trained deep learning baselines; because training-based results can be sensitive to split definition and training/engineering choices, we treat such comparisons as out of scope here and instead report explicit PR/AP/FROC and IoU-sensitivity analyses under a fully specified protocol.

4.5. Future Work

Future work includes the following: (i) automatic threshold recommendation for a specified FP/image budget from the exported FROC curve; (ii) lightweight box tightening to improve IoU without retraining; (iii) broader alignment models (e.g., rigid/affine) while retaining an SSIM-gain acceptance gate; (iv) per-template caching and template-library management for scale; and (v) computational/implementation optimizations (early exits, efficient SSIM/difference computation, downsampling-aware parameter scaling, and fixed-point/embedded-friendly variants) with evaluation on single-board computers or microcontrollers.

5. Conclusions

This work addresses defect–intent ambiguity in PCB AOI by presenting a training-free, reference-based defect localization framework designed for deployments where interpretability and controllable false-alarm rates are critical. The method selects a compatible defect-free reference via coarse-to-fine matching, applies global alignment only when it increases SSIM (to avoid defect-driven over-correction), and forms a defect-likelihood field by fusing SSIM-derived structural dissimilarity with normalized absolute differences; confidence-ranked detections are then produced via connected-component extraction.

To support deployment-oriented decision making, we report precision–recall and FROC curves rather than advocating a single universal threshold. At the adopted operating point, the method achieves 0.9987 recall at 0.149 FP/image, and operating-point metrics are additionally reported using K-fold calibration/test thresholding to avoid selecting thresholds on the same set used for reporting (with a full-set best-F1 threshold retained

https://doi.org/10.3390/s26051541

only as a clearly labeled reference point). Under the adopted box-matching protocol with $\tau_{IoU} = 0.10$, average precision reaches 0.984, and AP-versus-IoU sweeps transparently characterize sensitivity to stricter localization requirements. The primary limitation is the need for defect-free templates of the same PCB layout/revision; within this setting, the approach provides a reproducible and auditable alternative to training-based inspection by flagging only statistically and structurally supported deviations for review.

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

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The code supporting the findings of this study will be finalized and released publicly upon acceptance/publication at: https://github.com/nonsakhoo/PCB-defect-localization (accessed on 24 January 2026). During peer review, an unofficial snapshot is made available for reviewers via the same repository URL.

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

Abbreviations

The following abbreviations are used in this manuscript:

Acc.	Accuracy
AI	Artificial Intelligence
AOI	Automated Optical Inspection
AP	Average Precision
EMC	Electromagnetic Compatibility
EMI	Electromagnetic Interference
F1	F1 Score
FN	False Negative
FP	False Positive
FROC	Free-response Receiver Operating Characteristic
GT	Ground Truth
IoU	Intersection over Union
mAP	mean Average Precision
PCB	Printed Circuit Board
PI	Power Integrity
PR	Precision–Recall
QC	Quality Control
RAM	Random-Access Memory
ROI	Region of Interest
SI	Signal Integrity
SSIM	Structural Similarity Index
TN	True Negative
TP	True Positive
VOC	Visual Object Classes
XML	Extensible Markup Language

https://doi.org/10.3390/s26051541

Appendix A. Implementation Details and Parameter Scaling

This appendix describes implementation-faithful computations that affect reproducibility of the proposed method, including template preselection, ROI discretization, component-to-box confidence scoring, and processing-scale parameter scaling.

Appendix A.1. Discrete ROI Construction from the Area-Avoidance Ratio

The manuscript derives a continuous border thickness m from the area-avoidance ratio p. The reference implementation applies an explicit discretization and safety clamp.

Given image size $W \times H$ and $p \in [0, 0.95]$, define the discriminant

$$\Delta = (W + H)^2 - 4pWH, \qquad \Delta \leftarrow \max(0, \Delta), \tag{A1}$$

and compute the smaller root

$$m = \frac{(W + H) - \sqrt{\Delta}}{4}. \tag{A2}$$

The integer padding is then

$$\mathrm{padPx} = \lfloor \max(0, m) \rfloor, \tag{A3}$$

followed by a clamp to prevent degenerate inner regions,

$$\mathrm{padPx} \leftarrow \min\left(\mathrm{padPx},\ \left\lfloor \tfrac{\min(W,H)}{2} \right\rfloor - 1\right). \tag{A4}$$

The resulting ROI is the axis-aligned inner rectangle

$$\Omega = \{(u, v) :\ 1 + \mathrm{padPx} \le u \le W - \mathrm{padPx},\ 1 + \mathrm{padPx} \le v \le H - \mathrm{padPx}\}, \tag{A5}$$

and pixels outside Ω are excluded from thresholding and post-processing.

Appendix A.2. Template Preselection Score and Shortlist Logic

To accelerate template selection, the implementation uses a cheap preselection score computed for all templates at a small working resolution and then refines SSIM only on a shortlist.

In the reference implementation, the preselection images are computed at max long-side D_{pre} and are restricted to the same central region Ω used elsewhere: pixels outside Ω are either set to zero (masking) or excluded by cropping, depending on whether SSIM is evaluated by masking or ROI-cropping.

Let I_{pre} and $T_{k,\mathrm{pre}}$ denote the preprocessed (grayscale, normalized, and resized) query and template images at the preselection resolution. Two preselection scores are supported:

$$\mathrm{mad}(I_{\mathrm{pre}}, T_{k,\mathrm{pre}}) = -\frac{1}{n} \sum_{p=1}^{n} \left| I_{\mathrm{pre}}(p) - T_{k,\mathrm{pre}}(p) \right|, \tag{A6}$$

$$\mathrm{corr}(I_{\mathrm{pre}}, T_{k,\mathrm{pre}}) = \frac{\langle I_{\mathrm{pre}} - \mu_I,\ T_{k,\mathrm{pre}} - \mu_T \rangle}{\| I_{\mathrm{pre}} - \mu_I \|_2 \, \| T_{k,\mathrm{pre}} - \mu_T \|_2 + \epsilon}, \tag{A7}$$

where p indexes vectorized pixels, n is the number of pixels, and μ_I, μ_T are the mean pixel values of the respective preselection images.

Both scores are maximized: corr is a normalized (mean-subtracted) correlation, while mad is defined as the negative mean absolute difference so that smaller pixel-wise devia-

https://doi.org/10.3390/s26051541

tions yield larger scores. The stabilizer ϵ prevents numerical issues when the variance is near zero.

The shortlist is defined as the index set of the top-K_s templates under the chosen score,

$$\mathcal{K}_s = \mathrm{TopK}\big(\{\mathrm{score}(I_{\mathrm{pre}}, T_{k,\mathrm{pre}})\}_{k=1}^{K},\ K_s\big), \tag{A8}$$

and SSIM refinement is then computed only for $k \in \mathcal{K}_s$ on the selection resolution and inner region Ω.

The computational complexity of template preselection can be summarized as follows. Let K be the number of templates, and let P_{pre} denote the number of evaluated pixels in Ω at the preselection resolution; using the common square-grid proxy, $P_{\mathrm{pre}} = \mathcal{O}(D_{\mathrm{pre}}^2)$. Computing either mad or corr for a single template is linear in P_{pre}, so scoring all templates is

$$\mathcal{O}(K\,P_{\mathrm{pre}}) = \mathcal{O}(K\,D_{\mathrm{pre}}^2). \tag{A9}$$

Constructing the shortlist can be implemented either by (i) sorting all K scores, which costs $\mathcal{O}(K \log K)$, or (ii) using a top-K_s selection routine (e.g., partial sort/heap/selection), which costs $\mathcal{O}(K \log K_s)$ (or $\mathcal{O}(K)$ expected time for selection-based methods) when $K_s \ll K$. The subsequent SSIM refinement cost is dominated by evaluating SSIM on K_s candidates at selection resolution, i.e.,

$$\mathcal{O}(K_s\,D_{\mathrm{sel}}^2), \tag{A10}$$

with smaller constants when SSIM is computed on the cropped ROI Ω rather than on the full image. In practice, grayscale conversion and resizing of templates can be cached to amortize per-query preprocessing and reduce I/O overhead.

Appendix A.3. Processing Resolution Presets and Parameter Scaling

To reduce runtime on high-resolution images, the implementation optionally runs the core localization stages on a downsampled processing grid and maps outputs back to the original resolution.

Let L be the maximum long-side length among the query and selected template, i.e.,

$$L = \max(H_I, W_I, H_T, W_T). \tag{A11}$$

Given a preset target long-side length L_{target} (e.g., 1280 for "720p", 854 for "480p"), the isotropic processing scale is

$$s = \min\left(1,\ \frac{L_{\mathrm{target}}}{L}\right). \tag{A12}$$

When $s < 1$, both images are resized by s prior to registration, score-map computation, and mask extraction.

To maintain approximately consistent behavior across scales, morphology and area parameters are scaled when enabled. Denoting original parameters $(A_{\min}, r_c, r_d)$ and scaled parameters $(A'_{\min}, r'_c, r'_d)$, the implementation uses

$$A'_{\min} = \max\left(1, \mathrm{round}(A_{\min}\,s^2)\right), \qquad r'_c = \max\left(1, \mathrm{round}(r_c\,s)\right), \qquad r'_d = \max\left(1, \mathrm{round}(r_d\,s)\right), \tag{A13}$$

where $\mathrm{round}(\cdot)$ denotes rounding to the nearest integer (as in the reference implementation). Finally, predicted boxes are mapped back to the original grid by scaling their $[x\ y\ w\ h]$ coordinates by $1/s$.

Appendix A.4. SSIM-Map Implementation Details

The SSIM map is computed using MATLAB R2023a's SSIM to obtain the per-pixel similarity map (second output). For reproducibility, the reference implementation uses

https://doi.org/10.3390/s26051541

grayscale images normalized to $[0,1]$ (dynamic range $L = 1$) and evaluates SSIM with explicit parameters: `DynamicRange = 1`, $K_1 = 0.01$, $K_2 = 0.03$, `GaussianWeights = true`, and `WindowSize = 11` (square window). Under MATLAB's SSIM definition, $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$, so with $L = 1$ the stabilizing constants are $C_1 = 1 \times 10^{-4}$ and $C_2 = 9 \times 10^{-4}$. For robustness across MATLAB versions, the reference implementation attempts to call SSIM with the above name-value parameters and fall back to the default SSIM(A,B) if a given option (or per-pixel map output) is unsupported; if the per-pixel map output is unavailable, it falls back to using the scalar SSIM value to form a constant map.

Appendix A.5. Detection Confidence Score from Connected Components

The implementation forms detections by connected components of the binary mask and assigns each component a confidence score computed from the defect-likelihood map values within that component.

Let Y be the final binary mask. Let $\{\mathcal{C}_i\}_{i=1}^{N_C}$ denote the connected components of Y. Let M denote the defect-likelihood (score) map used for thresholding. Each component yields an axis-aligned bounding box b_i and a scalar confidence score s_i computed by a configurable statistic over $\{M(u,v) : (u,v) \in \mathcal{C}_i\}$. The default implementation uses a robust quantile score:

$$s_i = Q_{q_s}\big(\{M(u,v) : (u,v) \in \mathcal{C}_i\}\big), \qquad q_s = 0.95. \tag{A14}$$

For completeness, the implementation also supports $s_i = \max_{(u,v)\in\mathcal{C}_i} M(u,v)$ or $s_i = \frac{1}{|\mathcal{C}_i|}\sum_{(u,v)\in\mathcal{C}_i} M(u,v)$.

Appendix A.6. Data Availability

The dataset and code availability are summarized in the Data Availability statement in the main text. If access issues arise during peer review, the code and evaluation scripts can also be provided by the corresponding author upon reasonable request.

For reproducibility, the experiments in this paper use the PKU-Market-PCB dataset [3] with the following dataset organization under `dev/PCB_DATASET/`:

- Defect-free template images: `PCB_USED(Defect-free)/`;
- Defective query images: `images(Defective)/` (organized by defect type, e.g., `Missing_hole`, `Mouse_bite`, `Open_circuit`, `Short`, `Spur`, `Spurious_copper`);
- Bounding-box annotations: Pascal VOC XML files under `Annotations/` with the same defect-type subfolder structure as `images(Defective)/`.

Appendix A.7. Implementation Configuration Used in Experiments

For reproducibility, Table A1 summarizes the key configuration values used to generate the reported results and figures in this manuscript using the mathematical notation introduced in the main text and Appendix B.2. The rightmost column provides the corresponding configuration field in `defaultConfig()` (`dev/v0_14_eval2_release1.m`).

Table A1. Experiment configuration in manuscript notation (mapped to implementation `cfg`).

Symbol	Value Used in Experiments	Implementation Mapping
p	$p = 0.07$ (area-avoidance ratio for ROI border exclusion)	`cfg.roi.avoidAreaRatio` (with `cfg.roi.enable = true`)
q	$q = 0.995$ (threshold quantile on Ω for mask extraction)	`cfg.mask.quantile`
$A_{\min}$	$A_{\min} = 150$ px (minimum connected-component area)	`cfg.mask.minAreaPx` (scaled when resizing is enabled)
r_c	$r_c = 6$ px (morphological closing radius)	`cfg.mask.closeRadiusPx` (scaled when resizing is enabled)
r_d	$r_d = 2$ px (morphological dilation radius)	`cfg.mask.dilateRadiusPx` (scaled when resizing is enabled)
q_s	$q_s = 0.95$ (component confidence quantile)	`cfg.bbox.scoreMethod = 'quantile'`, `cfg.bbox.scoreQuantile = 0.95`

https://doi.org/10.3390/s26051541

Table A1. *Cont.*

Symbol	Value Used in Experiments	Implementation Mapping
w	$w = 0.35$ (absolute-difference weight in $M = (1 - \text{SSIMmap}) + w\,D_{\text{abs}}$)	`cfg.diff.absDiffWeight` (with `cfg.diff.useAbsDiff = true`)
$\Delta_{\min}$	$\Delta_{\min} = 0.0$ (accept alignment iff SSIM gain $\Delta \geq \Delta_{\min}$ on Ω)	`cfg.registration.validateTransform = true`, `cfg.registration.minSSIMGain = 0.0`
τ_{IoU}	$\tau_{\text{IoU}} = 0.10$ (IoU threshold for TP/FP/FN matching)	`cfg.eval.iouThreshold = 0.1`
D_{pre}	$D_{\text{pre}} = 128$ px (max long-side for preselect score)	`cfg.templateSelect.preselect.maxDim`
K_s	$K_s = 3$ (shortlist size after preselect)	`cfg.templateSelect.preselect.topK`
D_{sel}	$D_{\text{sel}} = 426$ px (max long-side for SSIM refinement; preset "240p")	`cfg.templateSelect.downscalePreset = '240p'`
D_{proc}	$D_{\text{proc}} = 1280$ px (max long-side for processing; preset "720p")	`cfg.resize.enable = true`, `cfg.resize.preset = '720p'`
τ_θ	translation-only transform model for alignment	`cfg.registration.type = 'translation'`
ROI eval	SSIM/refinement computed on cropped ROI Ω (faster than masking)	`cfg.templateSelect.useROICropForSSIM = true`

Appendix B. Evaluation and Metric Definitions

This appendix collects explicit metric definitions and evaluation-protocol computations used to produce the reported PR/AP/FROC results under the adopted IoU matching protocol.

Appendix B.1. Operating-Point Threshold Selection (Option A: Best F1)

This paper reports a single reference operating point using Option A (best F1). This operating point is selected from the confidence-ranked detections produced by the pipeline.

Let $\{(s_i, y_i)\}_{i=1}^{N}$ denote all detections across the evaluated dataset, where $s_i \in \mathbb{R}$ is the confidence score, and $y_i \in \{0, 1\}$ indicates whether the detection is a TP under the adopted per-image greedy one-to-one IoU matching (with $y_i = 1$ for TP, and $y_i = 0$ for FP). Let N_{GT} be the total number of ground-truth boxes across the evaluated images.

Sort detections by score in descending order, producing an ordering π such that $s_{\pi(1)} \geq s_{\pi(2)} \geq \cdots \geq s_{\pi(N)}$. Define the cumulative counts

$$\text{TP}(k) = \sum_{i=1}^{k} y_{\pi(i)}, \qquad \text{FP}(k) = \sum_{i=1}^{k} (1 - y_{\pi(i)}), \qquad k = 1, \ldots, N. \tag{A15}$$

The corresponding precision, recall, and F1 sequences are

$$P(k) = \frac{\text{TP}(k)}{\text{TP}(k) + \text{FP}(k) + \epsilon}, \tag{A16}$$

$$R(k) = \frac{\text{TP}(k)}{N_{\text{GT}} + \epsilon}, \tag{A17}$$

$$\text{F1}(k) = \frac{2\,P(k)\,R(k)}{P(k) + R(k) + \epsilon}. \tag{A18}$$

The best-F1 index is $k^* = \arg\max_k \text{F1}(k)$, and the operating-point threshold is chosen as

$$t^* = s_{\pi(k^*)}. \tag{A19}$$

All detections with $s_i \geq t^*$ are retained to compute the reported operating-point counts and metrics.

For the cross-validated operating-point estimates reported in the main text, we use $K = 5$ folds. Evaluated images (those with available Pascal VOC ground-truth XML) are randomly permuted using a fixed implementation seed and assigned to folds in a round-robin manner; in each fold, t^* is selected on the calibration folds and metrics are computed on the held-out fold.

https://doi.org/10.3390/s26051541

Appendix B.2. FROC Computation and Image-Level Recall

The evaluation exports FROC curves by sweeping a finite set of score thresholds derived from the score distribution.

Let $\mathcal{S} = \{s_i\}_{i=1}^{N}$ be the set of all detection scores across the evaluated dataset. This work defines a set of candidate thresholds by sampling quantiles of $\mathcal{S}$. Let $\mathcal{Q} = \{0, \frac{1}{49}, \ldots, 1\}$ (50 evenly spaced quantile levels) and define

$$\mathcal{T} = \text{unique}\big(\{Q_q(\mathcal{S}) : q \in \mathcal{Q}\}\big), \tag{A20}$$

then sort $\mathcal{T}$ in descending order.

For each threshold $t \in \mathcal{T}$, this work retains detections with $s_i \geq t$ and compute dataset-level counts $\text{TP}(t), \text{FP}(t), \text{FN}(t)$ by performing greedy one-to-one matching within each image (each ground-truth box matches at most one detection and vice versa) under the adopted IoU threshold. Let N_{img} denote the number of evaluated images that have available ground-truth annotations, and let N_{GT} denote the total number of ground-truth boxes across those images.

The object-level FROC quantities are

$$\text{FP/image}(t) = \frac{\text{FP}(t)}{N_{\text{img}} + \epsilon}, \qquad \text{Recall}_{\text{obj}}(t) = \frac{\text{TP}(t)}{N_{\text{GT}} + \epsilon}. \tag{A21}$$

Image-level recall is computed by counting images that contain at least one ground-truth box and checking whether any TP occurs in the image at threshold t. Let $N_{\text{img}}^{\text{GT}}$ be the number of images with at least one ground-truth box. Let $N_{\text{img}}^{\text{TP}}(t)$ be the number of those images for which at least one retained detection is matched as a TP at threshold t. Then

$$\text{Recall}_{\text{img}}(t) = \frac{N_{\text{img}}^{\text{TP}}(t)}{N_{\text{img}}^{\text{GT}} + \epsilon}. \tag{A22}$$

In the manuscript, FROC plots use $\text{Recall}_{\text{obj}}(t)$ versus $\text{FP/image}(t)$; $\text{Recall}_{\text{img}}(t)$ is reported for completeness.

Appendix B.3. IoU for [x y w h] Boxes

Let $b = [x\ y\ w\ h]$ and $g = [x'\ y'\ w'\ h']$. Define corners $(x_1, y_1) = (x, y)$, $(x_2, y_2) = (x + w, y + h)$ and similarly for g. The intersection width/height are

$$w_I = \max\big(0, \min(x_2, x_2') - \max(x_1, x_1')\big), \tag{A23}$$

$$h_I = \max\big(0, \min(y_2, y_2') - \max(y_1, y_1')\big), \tag{A24}$$

with intersection area $A_I = w_I h_I$. The IoU is

$$\text{IoU}(b, g) = \frac{A_I}{A(b) + A(g) - A_I + \epsilon}. \tag{A25}$$

Appendix B.4. Average Precision from Confidence-Ranked Detections

Let all detections across the dataset be ranked by confidence s in descending order. For a fixed IoU threshold, each detection is labeled TP or FP by greedy matching within each image (each ground-truth box may match at most one detection). Let $\text{TP}(t)$ and $\text{FP}(t)$

https://doi.org/10.3390/s26051541

be cumulative counts up to rank t, and let N_{GT} be the total number of ground-truth boxes. Precision and recall sequences are

$$P(t) = \frac{\mathrm{TP}(t)}{\mathrm{TP}(t) + \mathrm{FP}(t) + \epsilon}, \qquad R(t) = \frac{\mathrm{TP}(t)}{N_{\mathrm{GT}} + \epsilon}. \tag{A26}$$

AP is the area under the precision–recall curve. This work uses a standard VOC-style discrete approximation based on the precision envelope $\tilde{P}(r) = \max_{\hat{r} \geq r} P(\hat{r})$ and piecewise integration over recall. In our experiments, AP is computed using the all-points VOC interpolation (precision envelope with integration over all recall steps), not the VOC 2007 11-point variant.

References

1. Chin, R.T.; Harlow, C.A. Automated Visual Inspection: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **1982**, *PAMI-4*, 557–573. [CrossRef]
2. Moganti, M.; Ercal, F.; Dagli, C.H.; Tsunekawa, S. Automatic PCB Inspection Algorithms: A Survey. *Comput. Vis. Image Underst.* **1996**, *63*, 287–313. [CrossRef]
3. Peking University Open Lab on Human-Robot Interaction. *PKU-Market-PCB Dataset*. Available online: https://robotics.pkusz.edu.cn/resources/dataset/ (accessed on 26 January 2026).
4. Azamfirei, V.; Psarommatis, F.; Lagrosen, Y. Application of automation for in-line quality inspection, a zero-defect manufacturing approach. *J. Manuf. Syst.* **2023**, *67*, 1–22. [CrossRef]
5. Chen, X.; Wu, Y.; He, X.; Ming, W. A Comprehensive Review of Deep Learning-Based PCB Defect Detection. *IEEE Access* **2023**, *11*, 139017–139038. [CrossRef]
6. Sankar, V.U.; Lakshmi, G.; Sankar, Y.S. A Review of Various Defects in PCB. *J. Electron. Test.* **2022**, *38*, 481–491. [CrossRef]
7. Kim, M. Meander-DGS Effect on Electromagnetic Bandgap Structure for Power/Ground Noise Suppression in High-Speed Integrated Circuit Packages and PCBs. *Electronics* **2022**, *11*, 211. [CrossRef]
8. Ott, H.W. *Electromagnetic Compatibility Engineering*; Wiley: Hoboken, NJ, USA, 2009.
9. Bogatin, E. *Signal and Power Integrity—Simplified*, 2nd ed.; Prentice Hall: Hoboken, NJ, USA, 2010.
10. Johnson, H.W.; Graham, M. *High-Speed Digital Design: A Handbook of Black Magic*; Prentice Hall: Hoboken, NJ, USA, 1993.
11. Ling, Q.; Isa, N.A.M. Printed Circuit Board Defect Detection Methods Based on Image Processing, Machine Learning and Deep Learning: A Survey. *IEEE Access* **2023**, *11*, 15921–15944. [CrossRef]
12. Tsai, D.; Lin, C.; Chen, J. The evaluation of normalized cross correlations for defect detection. *Pattern Recognit. Lett.* **2003**, *24*, 2525–2535. [CrossRef]
13. Li, Y.; Kuo, P.; Guo, J. Automatic Industry PCB Board DIP Process Defect Detection System Based on Deep Ensemble Self-Adaption Method. *IEEE Trans. Compon. Packag. Manuf. Technol.* **2021**, *11*, 312–323. [CrossRef]
14. Park, J.; Kim, Y.; Seo, H.; Cho, Y. Analysis of Training Deep Learning Models for PCB Defect Detection. *Sensors* **2023**, *23*, 2766. [CrossRef]
15. Tsai, D.; Chiang, I.; Tsai, Y. A Shift-Tolerant Dissimilarity Measure for Surface Defect Detection. *IEEE Trans. Ind. Inform.* **2012**, *8*, 128–137. [CrossRef]
16. Gaidhane, V.H.; Hote, Y.V.; Singh, V. An efficient similarity measure approach for PCB surface defect detection. *Pattern Anal. Appl.* **2018**, *21*, 277–289. [CrossRef]
17. Pham, T.T.A.; Thoi, D.K.T.; Choi, H.; Park, S. Defect Detection in Printed Circuit Boards Using Semi-Supervised Learning. *Sensors* **2023**, *23*, 3246. [CrossRef]
18. Chen, C.; Wu, Q.; Zhang, J.; Xia, H.; Lin, P.; Wang, Y.; Tian, M.; Song, R. U2D2PCB: Uncertainty-Aware Unsupervised Defect Detection on PCB Images Using Reconstructive and Discriminative Models. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5017710. [CrossRef]
19. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 4th ed.; Pearson: London, UK, 2018.
20. Crispin, A.J.; Rankov, V. Automated inspection of PCB components using a genetic algorithm template-matching approach. *Int. J. Adv. Manuf. Technol.* **2007**, *35*, 293–300. [CrossRef]
21. Kaur, B.; Kaur, G.; Kaur, A. Detection and classification of Printed circuit board defects using image subtraction method. In Proceedings of the 2014 Recent Advances in Engineering and Computational Sciences (RAECS), Chandigarh, India, 6–8 March 2014; pp. 1–5. [CrossRef]
22. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]

https://doi.org/10.3390/s26051541

23. Debella-Gilo, M.; Kääb, A. Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation. *Remote Sens. Environ.* **2011**, *115*, 130–142. [CrossRef]
24. Zheng, X.; Zheng, S.; Kong, Y.; Chen, J. Recent advances in surface defect inspection of industrial products using deep learning techniques. *Int. J. Adv. Manuf. Technol.* **2021**, *113*, 35–58. [CrossRef]
25. Huang, S.; Pan, Y. Automated visual inspection in the semiconductor industry: A survey. *Comput. Ind.* **2015**, *66*, 1–10. [CrossRef]
26. Dwivedi, G.; Pensia, L.; Lohchab, V.; Kumar, R. Nondestructive Inspection and Quantification of Soldering Defects in PCB Using an Autofocusing Digital Holographic Camera. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–8. [CrossRef]
27. Cai, N.; Lin, J.; Ye, Q.; Wang, H.; Weng, S.; Ling, B.W. A New IC Solder Joint Inspection Method for an Automatic Optical Inspection System Based on an Improved Visual Background Extraction Algorithm. *IEEE Trans. Compon. Packag. Manuf. Technol.* **2016**, *6*, 161–172. [CrossRef]
28. Wei, S.; Lai, S. Fast Template Matching Based on Normalized Cross Correlation with Adaptive Multilevel Winner Update. *IEEE Trans. Image Process.* **2008**, *17*, 2227–2235. [CrossRef]
29. Cheng, H.; Liu, Z.; Hou, L.; Yang, J. Sparsity-Induced Similarity Measure and Its Applications. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *26*, 613–626. [CrossRef]
30. Hassanin, A.I.M.; Abd El-Samie, F.E.; El Banby, G.M. A real-time approach for automatic defect detection from PCBs based on SURF features and morphological operations. *Multimed. Tools Appl.* **2019**, *78*, 34437–34457. [CrossRef]
31. Kang, D.; Lai, J.; Han, Y. Improving surface defect detection with context-guided asymmetric modulation networks and confidence-boosting loss. *Expert Syst. Appl.* **2023**, *225*, 120121. [CrossRef]
32. Luo, Z.; Lei, L.; Li, H. Rotation-Angle-Based Principal Feature Extraction and Optimization for PCB Defect Detection Under Uncertainties. *IEEE Trans. Ind. Inform.* **2025**, *21*, 932–939. [CrossRef]
33. Onshaunjit, J.; Srinonchat, J. Algorithmic Scheme for Concurrent Detection and Classification of Printed Circuit Board Defects. *Comput. Mater. Contin.* **2021**, *71*, 355–367. [CrossRef]
34. Jiao, L.; Zhang, F.; Liu, F.; Yang, S.; Li, L.; Feng, Z.; Qu, R. A Survey of Deep Learning-Based Object Detection. *IEEE Access* **2019**, *7*, 128837–128868. [CrossRef]
35. Zhong, Z.; Ma, Z. A Novel Defect Detection Algorithm for Flexible Integrated Circuit Package Substrates. *IEEE Trans. Ind. Electron.* **2022**, *69*, 2117–2126. [CrossRef]
36. Wang, Y.; Chen, Z.; Wang, J.; Shang, P.; Sowmya, A.; Sun, C. Printed Circuit Board Defect Detection Based on Lightweight Deep Learning Fusion Model. *Sensors* **2025**, *25*, 7403. [CrossRef] [PubMed]
37. Jun, H.; Jung, I.Y. Enhancement of Product-Inspection Accuracy Using Convolutional Neural Network and Laplacian Filter to Automate Industrial Manufacturing Processes. *Electronics* **2023**, *12*, 3795. [CrossRef]
38. Adibhatla, V.A.; Chih, H.; Hsu, C.; Cheng, J.; Abbod, M.F.; Shieh, J. Defect Detection in Printed Circuit Boards Using You-Only-Look-Once Convolutional Neural Networks. *Electronics* **2020**, *9*, 1547. [CrossRef]
39. Yang, Y.; Kang, H. An Enhanced Detection Method of PCB Defect Based on Improved YOLOv7. *Electronics* **2023**, *12*, 2120. [CrossRef]
40. Du, B.; Wan, F.; Lei, G.; Xu, L.; Xu, C.; Xiong, Y. YOLO-MBBi: PCB Surface Defect Detection Method Based on Enhanced YOLOv5. *Electronics* **2023**, *12*, 2821. [CrossRef]
41. Mo, C.; Hu, Z.; Wang, J.; Xiao, X. SGT-YOLO: A Lightweight Method for PCB Defect Detection. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 1–11. [CrossRef]
42. Wang, X.; Gao, J.-S.; Hou, B.-J.; Wang, Z.-S.; Ding, H.-W.; Wang, J. A Lightweight Modified YOLOX Network Using Coordinate Attention Mechanism for PCB Surface Defect Detection. *IEEE Sens. J.* **2022**, *22*, 20910–20920. [CrossRef]
43. Guo, Z.; Wang, C.; Yang, G.; Huang, Z.; Li, G. MSFT-YOLO: Improved YOLOv5 Based on Transformer for Detecting Defects of Steel Surface. *Sensors* **2022**, *22*, 3467. [CrossRef]
44. Wang, F.; Jiang, X.; Han, Y.; Wu, L. YOLO-LSDI: An Enhanced Algorithm for Steel Surface Defect Detection Using a YOLOv11 Network. *Electronics* **2025**, *14*, 2576. [CrossRef]
45. Yuan, Z.; Tang, X.; Ning, H.; Yang, Z. LW-YOLO: Lightweight Deep Learning Model for Fast and Precise Defect Detection in Printed Circuit Boards. *Symmetry* **2024**, *16*, 418. [CrossRef]
46. Li, C.; Qu, Z.; Wang, S.; Bao, K.; Wang, S. A Method of Defect Detection for Focal Hard Samples PCB Based on Extended FPN Model. *IEEE Trans. Compon. Packag. Manuf. Technol.* **2022**, *12*, 217–227. [CrossRef]
47. Chen, B.; Dang, Z. Fast PCB Defect Detection Method Based on FasterNet Backbone Network and CBAM Attention Mechanism Integrated With Feature Fusion Module in Improved YOLOv7. *IEEE Access* **2023**, *11*, 95092–95103. [CrossRef]
48. Yang, J.; Liu, Z.; Du, W.; Zhang, S. A PCB Defect Detector Based on Coordinate Feature Refinement. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–10. [CrossRef]
49. Hu, B.; Wang, J. Detection of PCB Surface Defects With Improved Faster-RCNN and Feature Pyramid Network. *IEEE Access* **2020**, *8*, 108335–108345. [CrossRef]

https://doi.org/10.3390/s26051541

50. Feng, B.; Cai, J. PCB Defect Detection via Local Detail and Global Dependency Information. *Sensors* **2023**, *23*, 7755. [CrossRef] [PubMed]
51. Jiang, W.; Li, T.; Zhang, S.; Chen, W.; Yang, J. PCB defects target detection combining multi-scale and attention mechanism. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106359. [CrossRef]
52. Wu, L.; Zhang, L.; Zhou, Q. Printed Circuit Board Quality Detection Method Integrating Lightweight Network and Dual Attention Mechanism. *IEEE Access* **2022**, *10*, 87617–87629. [CrossRef]
53. Chen, W.; Meng, S.; Wang, X. Local and Global Context-Enhanced Lightweight CenterNet for PCB Surface Defect Detection. *Sensors* **2024**, *24*, 4729. [CrossRef]
54. Liu, H.; Zhao, X.; Liu, Q.; Chen, W. An Optimization Method for PCB Surface Defect Detection Model Based on Measurement of Defect Characteristics and Backbone Network Feature Information. *Sensors* **2024**, *24*, 7373. [CrossRef]
55. Bhattacharya, A.; Cloutier, S.G. End-to-end deep learning framework for printed circuit board manufacturing defect classification. *Sci. Rep.* **2022**, *12*, 12559. [CrossRef]
56. Kim, J.; Ko, J.; Choi, H.; Kim, H. Printed Circuit Board Defect Detection Using Deep Learning via A Skip-Connected Convolutional Autoencoder. *Sensors* **2021**, *21*, 4968. [CrossRef]
57. Li, D.; Li, C.; Chen, C.; Zhao, Z. Semantic Segmentation of a Printed Circuit Board for Component Recognition Based on Depth Images. *Sensors* **2020**, *20*, 5318. [CrossRef]
58. Peng, J.; Fan, W.; Lan, S.; Wang, D. MDD-DETR: Lightweight Detection Algorithm for Printed Circuit Board Minor Defects. *Electronics* **2024**, *13*, 4453. [CrossRef]
59. Xie, Z.; Zou, X. MFAD-RTDETR: A Multi-Frequency Aggregate Diffusion Feature Flow Composite Model for Printed Circuit Board Defect Detection. *Electronics* **2024**, *13*, 3557. [CrossRef]
60. Wang, S. Domain adaptation using transformer models for automated detection of exterior cladding materials in street view images. *Sci. Rep.* **2025**, *16*, 2696. [CrossRef] [PubMed]
61. Deng, S.; Deng, L.; Sun, T.; Yu, S.; Wang, L.; Chen, B.; Hu, H.; Xie, Y.; Yin, H.; Xiao, J.; et al. EEDD: Edge-Guided Energy-Based PCB Defect Detection. *Electronics* **2023**, *12*, 2306. [CrossRef]
62. Lei, L.; Li, H.; Yang, H. Multiscale Convolution-Based Probabilistic Classification for Detecting Bare PCB Defects. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–8. [CrossRef]
63. Ding, R.; Dai, L.; Li, G.; Liu, H. TDD-net: A tiny defect detection network for printed circuit boards. *CAAI Trans. Intell. Technol.* **2019**, *4*, 110–116. [CrossRef]
64. Shao, X.; Liu, Z.; He, Q.; Yu, M.; Li, C. SM-FSOD: A Second-Order Meta-Learning Algorithm for Few-Shot PCB Defect Object Detection. *Electronics* **2025**, *14*, 3863. [CrossRef]
65. Lin, Q.; Takamasu, K.; Chen, M. An Optical Lens Defect Detection Method for Micro Vision Based on WGSO-YOLO. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–8. [CrossRef]
66. Shi, W.; Li, C.; Dai, J.; Niu, N. GCF-Net: Steel Surface Defect Detection Network Based on Global Attention Perception and Cross-Layer Interactive Fusion. *Electronics* **2025**, *14*, 1776. [CrossRef]
67. Annaby, M.H.; Fouda, Y.M.; Rushdi, M.A. Improved Normalized Cross-Correlation for Defect Detection in Printed-Circuit Boards. *IEEE Trans. Semicond. Manuf.* **2019**, *32*, 199–211. [CrossRef]
68. Zhang, Z.; Deriche, R.; Faugeras, O.; Luong, Q. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artif. Intell.* **1995**, *78*, 87–119. [CrossRef]
69. Ebayyeh, A.A.R.M.A.; Mousavi, A. A Review and Analysis of Automatic Optical Inspection and Quality Monitoring Methods in Electronics Industry. *IEEE Access* **2020**, *8*, 183192–183271. [CrossRef]
70. Zhou, Y.; Yuan, M.; Zhang, J.; Ding, G.; Qin, S. Review of vision-based defect detection research and its perspectives for printed circuit board. *J. Manuf. Syst.* **2023**, *70*, 557–578. [CrossRef]
71. 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]
72. Everingham, M.; Eslami, S.M.A.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes Challenge: A Retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [CrossRef]
73. Chiba, N.; Hashimoto, K. Ultra-Fast Multi-Scale Shape Estimation of Light Transport Matrix for Complex Light Reflection Objects. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21–25 May 2018; pp. 6147–6152. [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.

https://doi.org/10.3390/s26051541

sensors

MDPI

Article

Production System Monitoring Based on Petri Nets Enhanced with Multi-Source Information

Peng Liu [1], Xinze Li [1], Chenlong Zhang [2], Yanru Kang [1], Jun Qian [3,4,*] and Weizheng Chen [1,*]

[1] School of Mechanical and Aerospace Engineering, Jilin University, Changchun 130012, China; liupeng@jlu.edu.cn (P.L.); xzli9923@mails.jlu.edu.cn (X.L.); yanrukang_jlu@163.com (Y.K.)
[2] Jilin Zhongwei Protective Equipment Co., Ltd., Changchun 130507, China; zcl2988@163.com
[3] School of Mechanical Engineering, Zhejiang University, Hangzhou 310058, China
[4] Engineering Training Center, Zhejiang University, Hangzhou 310058, China
* Correspondence: qqjun@zju.edu.cn (J.Q.); wzchen@jlu.edu.cn (W.C.)

Abstract

As the manufacturing industry continues to advance its digital transformation, intelligent sensing technology has become a key support for achieving precise, efficient and automated quality control. However, current production line monitoring systems predominantly rely on fixed and costly monitoring equipment and sensors, lacking flexible and interactive first-person perspective perception approaches centered on on-site operators. Meanwhile, factory process monitoring often depends solely on visual expression rather than balancing the capabilities of the simulation model and visual state detection, leading to delayed responses to abnormal systems and hindering the adjustment strategy feedback. To address these limitations, this study provides wearable sensing for key workers, enriching the state perception capabilities in industrial scenarios. Furthermore, to achieve dynamic model and real-time visual representation of production line operations, a multi-source information-enhanced Petri nets model is proposed in terms of engineering and user-friendliness. With the solid mathematical basics of the Petri nets and the enriched human–machine data from the product line, this method provides an intuitive, dynamic and accurate reflection of the production system's real-time operational status, offering a scientific and reliable basis for operational decision-making. The proposed approach has been implemented in a real-world production system for reinforced concrete civil defense doors, and this engineering application can also be extended to many other scenarios.

Keywords: production line; condition monitoring; Petri nets; smart helmet

1. Introduction

As the manufacturing industry increasingly shifts toward multi-variety, small-batch, and customized production models, mixed-model assembly lines have emerged as a critical enabler for discrete manufacturing enterprises to improve market responsiveness and optimize resource utilization efficiency [1]. Comprehensive system monitoring is indispensable for ensuring operational reliability, process stability, and performance traceability in assembly lines and manufacturing facilities. Conventional existing condition monitoring platforms lack awareness of human–machine status, simulation-based planning capabilities, and intuitive human–computer interaction, which collectively impede data-informed decision-making. Consequently, there is an urgent need to establish a comprehensive production monitoring system capable of capturing material flow, equipment status, and

operator activities, while a real-time performance modeling evaluation method should be applied to enhance the analysis ability of monitoring [2–5].

Understanding production data and its flow is critical for system monitoring [6]. Nowadays, a human–machine monitoring approach is needed to enable comprehensive, real-time oversight of the entire production process [7–9]. In modern industrial settings, accurately identifying human–machine interaction states is key to automated monitoring and intelligent production management. Qadeer et al. achieved a trajectory prediction of mechanically thrown objects in modern industrial environments through multi-camera tracking and a Bi-LSTM deep neural network [10]. Research has made progress in recognizing human actions [11]. Yan et al. utilized a single 2D camera to estimate worker crowding density, achieving an error margin of approximately 0.6 m, albeit under the challenge of worker self-occlusion scenarios [12]. Wang et al. acknowledged the limitations of relying solely on single sensors for activity recognition in smart factories and thus adopted a multi-modal fusion approach by integrating IMU sensors with cameras to enhance accuracy [13]. Nonetheless, IMU sensors inherently possess drawbacks, including limited ranging distance and indirect position estimate, which result in relatively lower positional accuracy in industrial settings. Andronas et al. [14] combined visual, audio, and wearable sensor data to detect operator intent, boosting car assembly efficiency by 25%. EMG and video data together also improve detection of high-risk behaviors in dangerous tasks [15]. By combining multiple data sources, these methods expand the use of human–machine systems and support better behavior recognition in complex industrial environments.

After acquiring production system data, modeling methods are necessary for system analysis. There are three model categories: mathematical models (e.g., queuing theory, Markov chains), graphical models (e.g., flowcharts, Petri nets, GRAI nets), and simulation models (e.g., Plant Simulation, AnyLogic, FlexSim). Mathematical models require strong assumptions and perform poorly on systems with complex logic, high uncertainty, strong dynamics, or that are large-scale. Simulation models are tightly coupled to proprietary software, limiting reusability, portability, and modifiability. Graphical models, by contrast, offer both formal rigor and visual clarity, supporting system description, state monitoring, simulation setup and optimization design. Among them, Petri nets precisely model concurrency, synchronization, conflicts, and resource sharing, making them especially well-suited for dynamic behavior modeling [16]. For example, Cristian et al. [17] modeled the hydropower station's in-service turbine governor using Petri nets. Mehdi et al. [18] embedded Petri nets into a network-based digital twin to model industrial IoT systems. Du et al. [19] proposed a Petri nets-based robust deadlock prevention method for automated manufacturing systems. Owing to the inherent flexibility of Petri nets modeling, numerous researchers have extended or adapted Petri nets to address domain-specific requirements. For example, Yang et al. [20] proposed a double-layer mapping Petri nets (DMPN) that separates information feedback and decision-making into two dedicated subnets—one for real-time monitoring, the other for closed-loop control—resolving the virtual–real mapping challenge and capturing event-driven system dynamics. Oliveira Júnior et al. [21] embedded Nelson's rules and the Page–Hinkley algorithm into timed transitions, creating data-driven, AI-enhanced Petri nets for early detection of performance degradation and maintenance optimization in digital twins. Pla et al. [22] proposed Resource-Aware Petri Nets (RAPN), integrating time and resource constraints for intelligent workflows, enabling delay prediction and full-process monitoring in medical equipment maintenance. When integrated with anomaly recognition methods, Petri nets enable effective state monitoring and fault diagnosis in production systems. For example, Liu et al. [23] proposed self-modifying colored Petri nets (SMCPN) by combining colored Petri nets with the adaptive weight mechanism of self-modifying Petri nets, supporting systematic monitoring and di-

https://doi.org/10.3390/s26061785

agnosis across multiple operational modes. Chen et al. [24] introduced the Petri net method for anomaly recognition in cigarette rolling, designed specifically to monitor air-tightness indicators with high sensitivity and interpretability.

Research shows that Petri nets and their extended forms, by integrating artificial intelligence, the Internet of Things, and fault tree analysis, can effectively describe the concurrent, asynchronous, and uncertain behaviors of complex systems, providing strong modeling support for system state monitoring, fault diagnosis, and performance optimization. However, for large-scale or complicated production systems, Petri nets models tend to suffer from combinatorial explosion in size and structural complexity, leading to diminished comprehensibility and limited effectiveness in visual representation and analysis. Existing methods still have significant shortcomings when modeling discrete manufacturing systems: material flow and process logic are ambiguously represented, making it hard to distinguish batch-specific paths. Resource usage and dynamic flows—especially for equipment and personnel under large-scale production—are poorly captured, limiting accurate load tracking. Task triggers like order initiation and process transitions lack intuitive expression, slowing on-site comprehension. Worker movements across processes and regions are not effectively modeled. Moreover, Petri net modeling is complex and inflexible, hindering real-time adaptation. These limitations impair clear visual monitoring and prevent practical use in real-time production control, restricting applications to theoretical contexts [25].

To address the above engineering issues, this study proposes multi-source information-enhanced Petri nets (MSI-Petri nets) for produce line monitoring. Section 2 analyzes the target production workflow and details the multi-source video fusion strategy, combining panoramic views, wearable camera feeds, and UWB positioning to capture operator behavior and equipment states. Section 3 introduces the method of MSI-Petri nets with mechanisms to improve visual expressiveness. Section 4 applies the approach in a civil defense door manufacturing setting. The main contributions are: (1) Real-time data and human–machine behaviors are integrated into the Petri nets for dynamic monitoring. (2) MSI-Petri nets place explicitly model humans, machines, and materials; material consumption ratios intuitively show inventory surplus. (3) Task-oriented transitions simplify the Petri net structure for diverse processes, and a time threshold visually indicates remaining transaction time, enabling engineers to monitor each product unit's status at a glance.

2. Multi-Source Data Acquisition for Production Systems

2.1. Production System of Civil Defense Doors

Civil defense doors play an important role in urban infrastructure and in the disaster prevention and emergency response system. The product uses an integrated lattice structure, incorporating energy absorption, force mitigation, lightweighting, cost reduction, and ease of manufacturing and assembly into its design. This achieves both structural lightweighting and improved efficiency in production and assembly. This chapter presents a case analysis of the production process of civil defense doors at a manufacturing facility, focusing on data acquisition and operational status monitoring.

A comprehensive analysis of the factory's layout indicates that the current production system is equipped with key machinery, including a pipe cutting machine, plate cutting machine, bending machine, and welding robots, with operators strategically assigned in accordance with the specific process requirements of each manufacturing stage. Given the present production scale, this configuration of equipment and human resources has established a stable and highly efficient process flow system, as illustrated in Figure 1. The raw materials for civil defense doors are steel pipes and steel plates. Steel pipes are cut

https://doi.org/10.3390/s26061785

to a specified length of approximately 30 cm using a pipe cutting machine, while steel plates are processed via a plate cutting machine into plate parts, surface plate, and bottom plate. Plate parts are welded into a base structure, which is then assembled with pipe parts to form the door frame. Surface plates are shaped into door surface components using a bending machine. These and other components are subsequently integrated to form a semi-finished product (door-frame structure). The bottom plate is then welded to the finished assembly to complete the structural integration. Finally, the assembly undergoes surface treatment processes, including deburring, rust removal, and painting, resulting in a finished product that complies with established factory specifications.

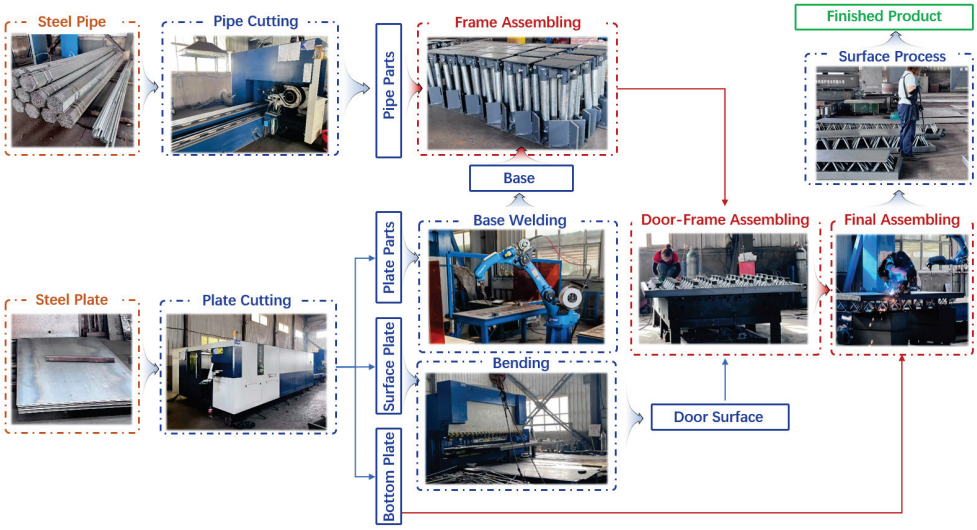


Figure 1. Process of the civil defense doors production system.

The status classifications of workers and machines were established based on eight core production processes of civil defense doors, namely, Pipe Cutting, Plate Cutting, Base Welding, Bending, Frame Assembling, Door-Frame Assembling, Final Assembling, and Surface Treatment, as detailed in Table 1. These states constitute the key state variables to be detected and identified by the data collection system proposed in this study. In Table 1, both human and machine states are uniformly represented using simplified coding: worker statuses are denoted by "HX" and equipment statuses by "MX".

Table 1. Classification of the status of different production units.

Production Unit	Status Classification
Pipe Cutting	**Human:** H0 (None), H1 (Idle), H2 (Loading materials), H3 (Operating), H4 (Unloading materials) **Machine:** M1 (Idle), M2 (Working)
Plate Cutting	**Human:** H0 (None), H1 (Idle), H2 (Loading materials), H3 (Operating), H4 (Unloading materials) **Machine:** M1 (Ideal), M2 (Plate Parts Cutting), M3 (Surface Plate Cutting), M4 (Bottom Plate)
Base Welding	**Human:** H0 (None), H1 (Idle), H2 (Loading materials), H3 (Operating), H4 (Unloading materials) **Machine:** M1 (Idle), M2 (Working)
Bending	**Human:** H0 (None), H1 (Idle), H2 (Loading materials), H3 (Operating), H4 (Unloading materials) **Machine:** M1 (Idle), M2 (Working)

https://doi.org/10.3390/s26061785

Table 1. *Cont.*

Production Unit	Status Classification
Frame Assembling	**Human:** H0 (None), H1 (Idle), H2 (Install support), H3 (Fasten), H4 (Transporting)
Door-Frame Assembling	**Human:** H0 (None), H1 (Idle), H2 (Replace), H3 (Calibration), H4 (Fasten), H5 (Welding), H6 (Transporting)
Final Assembling	**Human:** H0 (None), H1 (Idle), H2 (Replace), H3 (Welding), H4 (Transporting) **Machine:** M1 (Idle), M2 (Welding), M3 (Transporting)
Surface Treatment	**Human:** H0 (None), H1 (Idle), H2 (Deburring), H3 (Dedusting), H4 (Painting)

2.2. Design of Data Collection Method for the Produce System

This study presents a data acquisition system that integrates factory-wide monitoring with wearable devices for industrial workers. Helmets, as mandatory safety equipment continuously worn by personnel, offer inherent integration potential and consistent wearing stability, rendering them highly suitable as wearable sensing platforms. By embedding intelligent perception and feedback modules into helmets, the system enables efficient human–machine collaboration without imposing additional ergonomic load, while simultaneously capturing real-time production line status through the worker's first-person perspective—yielding high-value, operation-critical data. To address scenarios where helmet usage is impractical, smart badges are introduced as alternative sensing terminals, preserving core data acquisition and interaction capabilities and thereby ensuring the system's adaptability and operational flexibility across heterogeneous production environments.

In the domain of object recognition, YOLO (You Only Look Once) is a representative single-stage object detection framework that has been widely adopted across various computer vision applications owing to its high detection speed and accuracy. YOLOv8 further improves detection performance and inference efficiency through structural optimizations of the backbone network (CSPDarknet), the feature pyramid network (FPN), and the loss function. Nevertheless, conventional YOLO-based approaches exhibit limitations in modeling temporal dynamic information prevalent in industrial environments, particularly with regard to detection robustness under complex and cluttered background conditions. To overcome these challenges, this study proposes an enhanced YOLOv8 architecture integrated with multimodal temporal feature analysis, enabling more accurate and reliable recognition of dynamic interaction behaviors in complex industrial scenes.

The overall pipeline is illustrated in Figure 2. First, the system incorporates spatiotemporal localization data provided by Ultra-Wideband (UWB) devices to obtain the worker's spatial position and dynamic motion state. Based on this information, the corresponding region- or task-specific recognition models are triggered, enabling on-demand activation and resource optimization.

Subsequently, a first-person video stream is captured by the camera mounted on the helmet, with an original resolution of 1600×1200 pixels. To achieve an optimal balance between computational efficiency and detection accuracy, all frames are uniformly resized to a fixed dimension of 400×400 pixels and subjected to Gaussian filtering to suppress image noise. On this basis, the YOLOv8 model is applied to detect key elements in each frame, identifying equipment components and materials. The detection results are aggregated into an 80-dimensional environmental context feature vector, which includes: the occurrence frequency of each object category, confidence scores, and normalized minimum/maximum bounding box coordinates along with center coordinates.

https://doi.org/10.3390/s26061785

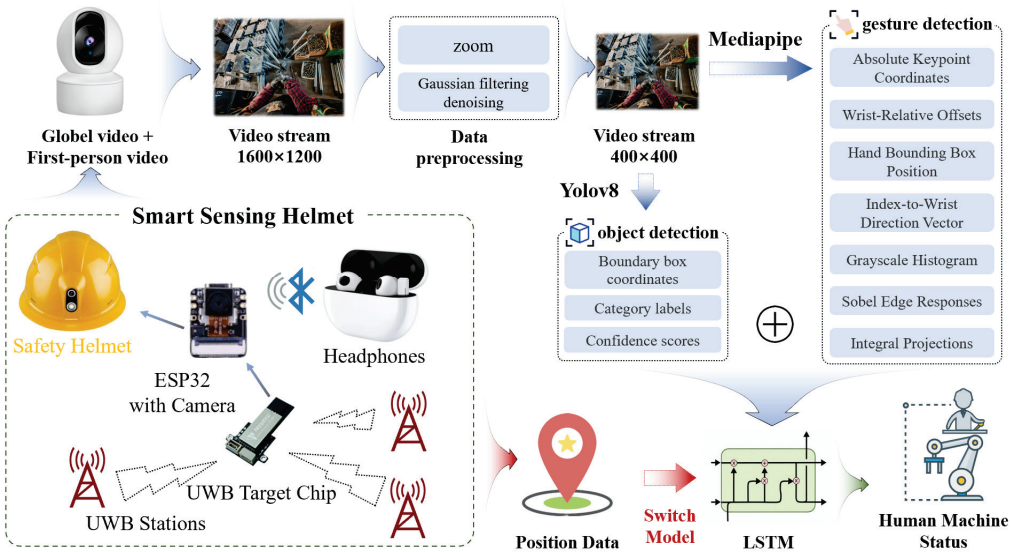


Figure 2. Schemes of data collection method.

On the other hand, MediaPipe (0.10.21) is employed for hand gesture recognition, extracting 21 hand key points and constructing a 200-dimensional hand-perception feature vector that comprises:

- The original absolute pixel coordinates of each key point in the image coordinate system;
- Offset vectors of all key points relative to the wrist joint (i.e., key point #0);
- Position information of the hand's minimum enclosing bounding box within the global image;
- A unit vector pointing from the tip of the index finger (keypoint #8) toward the wrist;
- Features extracted from the cropped hand region, including grayscale intensity histograms, horizontal and vertical edge responses computed via Sobel operators, and row/column-wise integral projections—collectively used to assist in inferring the worker's actions and states.

These hand features and environmental context features are then concatenated into a 280-dimensional joint representation and organized as a time sequence, providing structured input for subsequent temporal modeling.

Finally, to address the need for in-depth temporal feature modeling, the algorithm employs ReID (Re-Identification) to perform cross-frame association of the same target across consecutive frames. By linking detected instances over time, it constructs the target's motion trajectory and state transition sequence, thereby forming a comprehensive spatiotemporal state representation. To capture the long-term dependencies embedded within this sequence, an LSTM (Long Short-Term Memory network) is adopted to model and analyze the temporal dynamics. Leveraging its gated architecture, comprising the input gate, forget gate, and output gate, the LSTM effectively models long-range temporal correlations while mitigating the gradient vanishing problem inherent in conventional RNNs, thus enabling robust temporal understanding and classification of worker–device interaction states. Specifically, the model employs a two-layer LSTM architecture with an input dimension of 280 (combining 200-dimensional hand features and 80-dimensional object contextual features), a hidden size of 128, and a dropout rate of 0.3 to mitigate overfitting. During training, the Adam optimizer is used with an initial learning rate of 0.001, coupled with a StepLR scheduler (Learning Rate Scheduler) that halves the learning rate every 15 epochs; the model is trained for a total of 50 epochs with a batch size of 4. The model achieved a validation accuracy of 97.50%, with precision, recall, and F1-score all approaching 0.98. On a machine with a configuration of 9950X + 48G DDR5 + RTX

https://doi.org/10.3390/s26061785

5090D, the model inference speed is approximately 8.2 ms. These results indicate that the proposed approach performs excellently in both real-time capability and accuracy. In cases where automated recognition remains ambiguous or uncertain, the system further validates the production process state through real-time interaction with the wearer via Bluetooth headphones, ensuring accurate state confirmation through human-in-the-loop feedback.

In conclusion, by leveraging multiple workers equipped with first-person sensing devices, a comprehensive factory-wide data-monitoring network is established to enable continuous acquisition of real-time time-series data pertaining to materials, equipment, personnel, and other operational elements. Edge computing is employed to perform data preprocessing and feature extraction at the source, thereby constructing multi-dimensional data models that support critical applications such as quality traceability, energy efficiency analysis, and process simulation. This architecture provides a robust data infrastructure for achieving manufacturing transparency and facilitating lean, data-driven decision-making.

3. Information-Enhanced Petri Nets

This chapter begins with a review of the classical Petri nets model for production cells, upon which a visualization mechanism is integrated to enhance its state representation capability. Subsequently, by incorporating the real-time data acquisition features of production lines and the dynamic attributes of actual production tasks, an information-enhanced Petri net theoretical framework is established, aiming to enable more accurate and intuitive operational monitoring of the production process.

3.1. Visualized Petri Nets for Produce Unit

A classical Petri net is a 4-tuple $\boldsymbol{C} = (\boldsymbol{P}, \boldsymbol{T}, \boldsymbol{I}, \boldsymbol{O})$, where $\boldsymbol{P}$ denotes places: it is a finite non-empty set that characterizes resources, states, or buffer units in a production system; $\boldsymbol{P} = \{p_1, p_2, \ldots, p_n\}$, where n represents the total number of places. T denotes transitions: it is a finite non-empty set that characterizes production operations or process nodes; it satisfies $\boldsymbol{P} \cap \boldsymbol{T} = \Phi$ and $\boldsymbol{P} \cup \boldsymbol{T} = \Phi$; $\boldsymbol{T} = \{t_1, t_2, \ldots, t_m\}$, where m represents the total number of transitions. $\boldsymbol{I} \rightarrow 2^P$ is the mapping from transitions to their input places; a single transition can be associated with more than one input place. $\boldsymbol{O} \rightarrow 2^P$ is the mapping from transitions to their output places; a single transition can be associated with more than one output place.

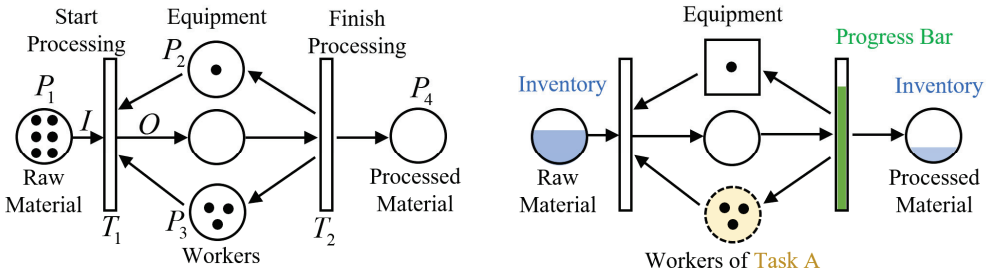
The marking of the Petri nets is represented by vector $\boldsymbol{\mu}$, where $\boldsymbol{\mu} = \{\mu_1, \mu_2, \ldots, \mu_n\}$ and μ_i denotes the number of tokens in the number i place. Therefore, marking can be used to characterize the transfer status of resources.

A production system is composed of the organic integration of multiple production units, and a specific production link within the system can be modeled using the basic Petri net structure illustrated in Figure 3a. In this framework, material flow serves as the central thread, while various production requirements and process methods are represented by transitions. The places are: raw material (P_1), equipment (P_2), workers (P_3), and processed material (P_4); the places at the top denote the machine and equipment resources utilized, reflecting the occupancy status of these resources, whereas the places at the bottom represent the human resources assigned to the corresponding process. The transitions are: start processing (T_1) and finish processing (T_2). The marking of the Petri nets is $\boldsymbol{\mu} = \{6, 1, 3, 0\}$.

In condition monitoring, accurately capturing the real-time operational status is of critical importance. To this end, this study implements the following information enhancement strategies for the basic production unit, as illustrated in Figure 3b. First, given the wide variety and large quantity of raw materials, the filling degree is adopted in the visualization to represent the proportion of each raw material type relative to the total inventory volume, rather than directly displaying token counts, thereby improving readability and

https://doi.org/10.3390/s26061785

information density. Second, to effectively differentiate among the three types of places in the production system, namely, human, machine, and material resources, distinct visual representations are employed: worker places are depicted as circles with dashed borders, with those corresponding to workers capable of performing identical tasks filled with the same color; machine places are represented by solid-line rectangles; and material places are shown as solid-line circles. Third, the completion status of transition processes is dynamically displayed using a progress bar. A time factor is considered to quantify the time required for transition completion, while color variations in the progress bar are used to reflect the degree of deviation between the current state and the expected normal state.



(a) Typical Petri Net for Production Unit (b) Visualized Petri Net for Production Unit

Figure 3. Petri Nets Model for production unit: (**a**) Typical Model, (**b**) Visualized Model.

3.2. Multi-Source Information-Enhanced Petri Nets

To address the requirements of production process modeling and monitoring for the entire civil defense doors production system, this paper proposes a multi-source information-enhanced Petri nets model (MSI-Petri nets). The core of this model lies in extending the elements and rules of traditional Petri nets, combined with further simplification and enhanced visualization, to better accommodate the key characteristics of complex production systems, such as resource state management, time constraints, handling of process interruptions, and personnel scheduling.

Definition 1. *Structure of MSI-Petri Nets.*

The MSI-Petri nets contain 7 tuples, namely, (P, T_{task}, F, M, θ, Γ, Σ), where each element is specified as follows:

Place set P: It is subdivided into three types of sub-places, namely, material places P_A, equipment places P_M, and worker places P_H. These three types have different shapes and are easy to distinguish.

The material place adopts a visualization form of filling degree, which is determined by the material consumption ratio (current material number divided by daily consumption volume). This design can directly reflect the utilization level of material inventory. The equipment place and the worker place retain the traditional Token to mark the resource occupation status.

Task-oriented Transition set T_{task}: It characterizes the start and end events of processing tasks. When equipment can switch between multiple processes, a task-oriented approach simplifies the system structure and supports condition monitoring.

Task switching mechanism: A mapping $T_{task} \rightarrow K$ (where K denotes the task set) is defined to bind production tasks to task-oriented transitions. During task switching, the ratio of the number of input arcs to output arcs, processing time, and visualization color of transitions all change dynamically, satisfying the following relationships:

https://doi.org/10.3390/s26061785

$$\begin{cases} \rho(t_{task}, k) = \frac{number\ of\ input\ arcs}{number\ of\ output\ arcs} (k \in K) \\ \tau(t_{task}, k) \in R^{+}(processing\ time\ for\ task\ k) \\ color(t_{task}, k) \in C(visual\ color\ for\ task\ k) \end{cases} \quad (1)$$

Progress bar-based state deviation visualization mechanism: The task time of transitions is bound to the color and filling degree of the progress bar. Based on the standard working hours/historical working hours, when the actual processing time deviates from the benchmark value beyond the threshold, the color of the progress bar changes to issue an early warning. For worker assembly tasks, the progress bar is divided into filling intervals to match different assembly stages.

Flow Relation *F***:** A set of directed arcs satisfying F $\subseteq$ (P $\times$ T) $\cup$ (T $\times$ P), which characterizes the material/resource flow between places and transitions. The flow velocity of arcs is visualized by the dynamic flickering frequency of the lines; when production is suspended, the arcs turn into dashed lines. The ratio of the number of input arcs to output arcs is determined by the material input-output ratio of the specific process.

Marking Function *M***:** A mapping M:P $\rightarrow$ N that denotes the number of tokens in places. For material places, the conventional token dots are replaced by "inventory container + filling progress bar"; for worker places and equipment places, traditional tokens are still used to indicate the availability status of resources.

Time Threshold Set θ: It is used to quantify the time boundaries for task execution and condition monitoring, with its mapping relationship defined as θ: $T_{task} \rightarrow (R^{+}, R^{+})$. That is, for each task-oriented transition $t_{task} \in T_{task}$, a pair of time thresholds (θ_{low}, θ_{up}) is assigned. Among them, θ_{low} is the minimum process time threshold, representing the shortest effective duration for task execution; if the actual processing time is lower than this threshold, the process execution is deemed insufficient, and an early warning is triggered. θ_{up} is the maximum process time threshold, whose value is based on standard working hours or historically optimal working hours and increased by a certain proportion; if the actual processing time exceeds this threshold, the process is judged abnormal, and the color of the progress bar triggers an overtime warning.

Resource Constraint Function Γ**:** A mapping Γ: T$\rightarrow 2^{P}$ that specifies the set of resource places required for transition firing.

Cycle Parameter Σ**:** A mapping Σ: F$\rightarrow R^{+}$ that assigns flow cycles to directed arcs, representing the transmission time of materials/resources on the arcs.

Definition 2. *Transition Firing Rules of MSI-Petri Nets.*

Combining the characteristics of Time Petri Nets and progress bar visualization, the transition firing rules of MSI-Petri Nets are as follows:

For an MSI-PN = (P, T_{task}, F, M, θ, Γ, Σ), a transition $t \in T$ satisfies the enabling condition (denoted as $M \rightarrow$) if and only if:

(1) Resource constraint is satisfied: For all $p \in \Gamma(t)$, $M(p) \geq 1$ (all required resource places contain at least one token).
(2) Input places are ready: For all $p \in \circ t$ (where $\circ t$ is input places), $M(p) \geq 1$ (raw materials or intermediate products are available).
(3) Cycle constraint is satisfied: The cycle parameter $\Sigma(f)$ corresponding to the flow relation F conforms to production requirements (the cycle value of the directed arc meets process specifications).
(4) Progress deviation is within the allowable range: The deviation between the filling degree of the progress bar and the standard working hours does not trigger the early warning threshold.

https://doi.org/10.3390/s26061785

When the transition t fires (denoted as $M \rightarrow M'$), the new marking M' is updated according to the following rule:

$$M'(p) = \begin{cases} M(p) - 1, p \in {}^{\circ}t \cup \Gamma(t) \\ M(p) + 1, p \in t^{\circ} \\ M(p), \text{Otherwise} \end{cases} \quad (2)$$

where t° represents the set of output places of transition t.

In terms of condition monitoring, when the production system is interrupted, the corresponding directed arcs turn into dashed lines, and a countdown for the buffer support duration is initiated. The duration time ($t_{duration}$) is determined by the current material number in the buffer (N_{buffer}) and the unit time consumption of the downstream process (T_{unit}):

$$t_{duration} = \frac{N_{buffer}}{T_{unit}} \quad (3)$$

Which reflects the maximum time that the system can maintain operation after interruption.

4. Case Study

This chapter presents a case study of a civil defense doors production system and constructs the corresponding MSI-Petri nets, providing an effective visualization framework for monitoring and analyzing the operational states of the production system.

All factory work areas are equipped with stable lighting infrastructure, ensuring consistent illumination conditions; consequently, the influence of lighting variations on the visual recognition system is negligible. However, electromagnetic interference may occur when workers operate within one meter of large mechanical equipment, leading to reduced positioning accuracy, system latency, and positional drift, with fluctuations reaching up to ± 0.8 m. Given the substantial physical separation between distinct operational zones, these localized disturbances have not significantly compromised the overall performance of the recognition system. Video data is compressed using the JPEG-based image format and transmitted via the TCP protocol, resulting in an average transmission delay of approximately 500 milliseconds. The data is wirelessly streamed in real time to a dedicated workstation capable of immediate processing.

The factory's specific dimensions and spatial layout are illustrated in Figure 4. There are 10 workers on duty, categorized by job type: 2 male workers (labeled 1 and 2) are assigned to the preparation process (encompassing pipe cutting, sheet cutting, and sheet bending); 5 male workers (labeled 3 to 7) are dedicated to heavy tasks (welding, assembly, transportation); and 3 female workers (labeled 8 to 10) are responsible for light assembly and post-processing operations. This workforce allocation strategy takes into account both the physical demands and technical complexity of each process, supporting the intensive and rational utilization of human resources. Workers within the same category can be dynamically reassigned across different tasks and stations in response to real-time production scheduling, thereby enhancing operational flexibility and resource efficiency. Leveraging UWB positioning chips embedded in smart helmets worn by workers and 3 UWB base stations deployed throughout the facility, the system enables real-time, centimeter-level accurate tracking of each worker's location, with their positions annotated numerically in Figure 4. Concurrently, visual data captured by helmet-mounted cameras or on-site cameras are displayed. By implementing the proposed multi-source perception algorithm, which integrates UWB-based localization with vision-based recognition, the system achieves precise, real-time awareness of "human-machine" states within each production unit, delivering reliable data support for dynamic process monitoring and optimization.

https://doi.org/10.3390/s26061785

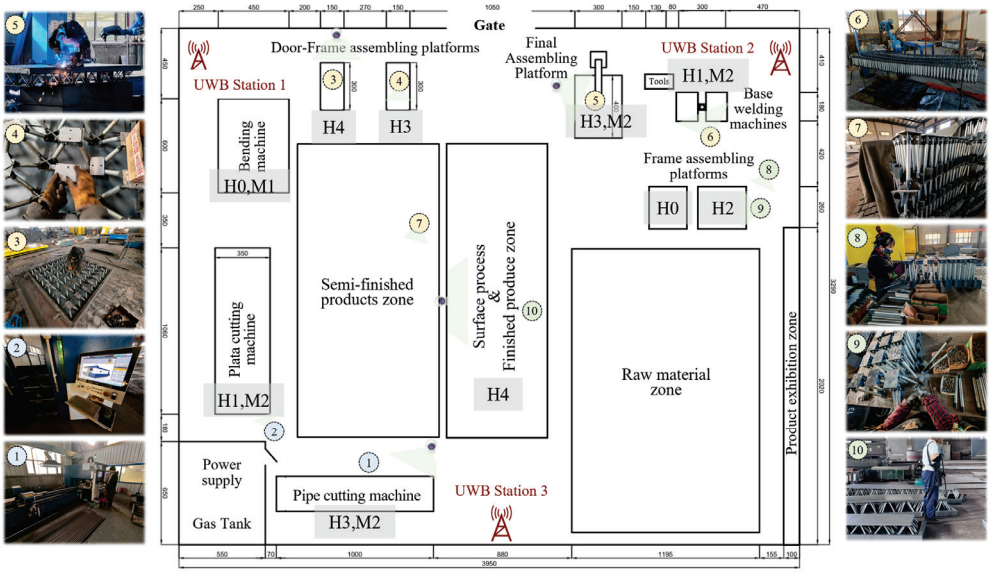


Figure 4. Human–machine status recognition of the production system.

Based on the state classification coding scheme defined in Table 1 (where "HX" denotes human states and "MX" denotes equipment states), all relevant zones in the figure are annotated with corresponding state codes to facilitate intuitive interpretation. According to the recognition results: the pipe cutting unit is in a state of active operation involving personnel and equipment working (H3, M2); the plate cutting unit exhibits idle personnel with machine conducting plate parts cutting task (H1, M2); the base welding unit shows idle personnel while the equipment is engaged in welding (H1, M2); the bending unit has no personnel present but the equipment is working (H0, M1); the left and right frame assembly units are respectively unoccupied (H0) and undergoing fastening operations (H2); the left and right door frame assembly units are in states of worker fastening (H4) and worker calibration (H3), respectively; the final assembly unit involves collaborative welding between personnel and equipment (H3, M2); and in the surface treatment unit, the worker is performing painting operations (H4).

As illustrated in Figure 5, the MSI-Petri nets model of the civil defense doors production system provides a comprehensive representation of system states. This model effectively captures the operational status of each production unit, particularly with respect to the utilization and flow dynamics of materials, human resources, and equipment resources. Transitions within the model serve as indicators of execution progress and enable the intuitive identification of abnormal conditions. The formal definitions of all places and transitions are provided in Tables 2 and 3, respectively. By applying Petri net-based simulation and analysis methods, the dynamic behavior of the production system can be systematically modeled. Through techniques such as state reachability analysis and performance evaluation, the model facilitates the identification of production bottlenecks and supports optimized resource allocation, thereby offering a quantitative foundation for process improvement. Furthermore, the model enables traceability analysis of abnormal state propagation paths, aiding in the development of targeted exception-handling strategies.

https://doi.org/10.3390/s26061785

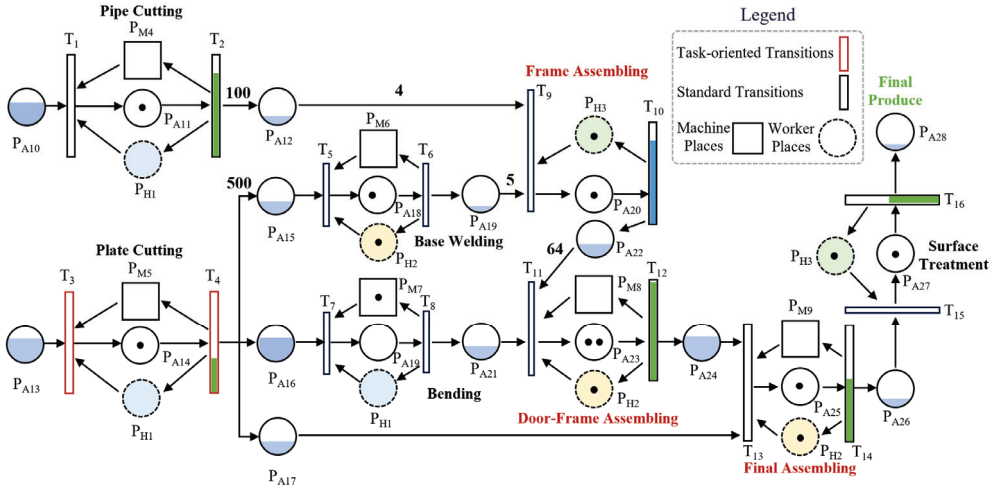


Figure 5. MSI-Petri net of the door production system.

Table 2. Definition of places in Petri nets.

Place	Name	Place	Name
P_{H1}	Workers of cutting, bending tasks	P_{H2}	Workers of welding task
P_{H3}	Workers of assembling task	P_{M4}	Pipe cutting machine available
P_{M5}	Plate cutting machine available	P_{M6}	Welding machine available
P_{M7}	Bending machine available	P_{M8}	Assembling platforms available
P_{M9}	Final assembling machine available	P_{A10}	Pipe raw material
P_{A11}	Pipe cutting in process	P_{A12}	Pipe parts
P_{A13}	Plate raw material	P_{A14}	Plate cutting in process
P_{A15}	Plate parts	P_{A16}	Surface plate
P_{A17}	Bottom plate	P_{A18}	Welding in process
P_{A19}	Base	P_{A20}	Bending in process
P_{A21}	Door surface	P_{A22}	Frame
P_{A23}	Door-Frame assembling in process	P_{A24}	Door-Frame
P_{A25}	Final assembling in process	P_{A26}	Complete door
P_{A27}	Surface treatment is in progress	P_{A28}	Final product

Table 3. Definition of transitions (T) in Petri nets.

T	Name	T	Name
T_1	Pipe cutting begins	T_2	Pipe cutting finished
T_3	Plate cutting begins	T_4	Plate cutting finished
T_5	Base welding begins	T_6	Base welding finished
T_7	Bending begins	T_8	Bending finished
T_9	Frame assembling begins	T_{10}	Frame assembling finished
T_{11}	Door-Frame assembling begins	T_{12}	Door-Frame assembling finished
T_{13}	Final assembling begins	T_{14}	Final assembling finished
T_{15}	Surface treatment begins	T_{16}	Surface treatment finished

In Figure 5, transitions T_3 and T_4 are designated as task-oriented, which means their output products vary depending on the specific task being executed, such as cutting plate parts, cutting surface plates, or bottom plates, and the processing time associated with each transition is accordingly task-dependent.

As illustrated in Figure 5, the Petri nets proposed in this study incorporate a progress bar-based state deviation visualization mechanism that dynamically links the task duration of each transition to the color and fill level of the corresponding progress bar. Taking T_2, the transition associated with steel pipe cutting as an example, its reference task time $Q1(T_2)$ represents the predefined standard cutting duration. The fill height of the progress bar is positively correlated with the actual execution time, while its color changes adaptively

https://doi.org/10.3390/s26061785

based on the degree of temporal deviation: when the actual duration equals $Q1(T_2)$, the progress bar displays as "baseline green," with the fill height precisely matching the expected duration; if the actual time exceeds $Q1(T_2)$ by 5%—for instance, due to increased processing time caused by equipment load fluctuations—the lower portion remains "baseline green" while the upper segment transitions to "warning yellow"; upon exceeding 10% or more, the majority of the progress bar shifts to "alarm red," providing a clear visual indication of cumulative task-level deviation. This mechanism overcomes the inherent limitation of traditional Petri nets, which rely solely on token counts to represent system states, thereby enabling dual-dimensional visualization of both temporal progression and deviation magnitude. As a result, anomaly detection in production processes is advanced from post-event statistical analysis to real-time process monitoring, significantly enhancing the real-time monitoring and control capabilities of complex assembly systems.

The marking function has been enhanced with an inventory-oriented visualization extension, representing an innovative adaptation to large-scale material-handling environments. The fill ratio of the progress bar quantitatively reflects the current inventory utilization level. Furthermore, interactive access to individual places allows users to retrieve the exact quantity of stored materials upon clicking, enabling production operators to rapidly assess raw material availability and warehouse capacity utilization. This enhancement significantly improves information readability and operational control efficiency in complex, high-volume production settings.

Almost all production systems encompass fundamental elements, including: human, machines, materials, and methods, the same as Petri Nets for the Produce Unit. As such, by reorganizing and modifying the method developed in this study, it can be applicable to the modeling, monitoring, and analysis of more complex systems. By incorporating a task-oriented mechanism, the proposed approach can be effectively deployed in mixed-model production lines, enabling dynamic switching of operational rules across different stages.

In the current stage, the efforts were primarily focused on enhancing the readability of the production system model through human factors engineering, visual design, and simplification of information flow. The overall direction is recognized as one that can enhance user experience and efficiency. Following a period of practical deployment, operator feedback, such as data gathered via structured interviews and standardized questionnaires, can be conducted to further provide a comprehensive analysis of user-friendliness.

5. Conclusions

The primary engineering challenge in monitoring complex production lines is the absence of multi-source heterogeneous, real-time synchronized data capturing worker–device collaboration. To address this, we propose a unified perception framework integrating UWB positioning, YOLOv8-based object detection, and MediaPipe-based pose estimation, coupled with an LSTM model for temporal recognition of human–machine interaction behaviors. To overcome the lack of embedded simulation capability in conventional monitoring systems, we adopt Petri nets as the formal modeling backbone and enhance them with interactive, semantics-aware visualization—yielding an integrated system that simultaneously supports real-time status monitoring and dynamic process simulation. This system serves as a deployable engineering solution for refined operational oversight and closed-loop management of complex production lines.

Focusing on the manufacturing system of civil defense doors as the application scenario, a multi-source perception and visualization framework is proposed and implemented by integrating intelligent helmets with information-enhanced Petri nets. On-site operational data are collected through a network of factory global cameras and intelligent helmets, which, when combined with dynamic modeling via Petri net structures,

https://doi.org/10.3390/s26061785

enable real-time visualization and representation of production states. This integration significantly improves the transparency and controllability of the production process. Case study results demonstrate that the proposed method enables recognition and monitoring of human–machine interactions, thereby providing reliable data support for production decision-making. With the foundation of the proposed MSI-Petri nets, future research can advance toward simulating the production system, enabling optimization decisions, and evaluating model performance based on real-world operational efficiency at the site. Furthermore, the method functions as a structured representation of a digital twin, facilitating not only real-time monitoring but also system-level analysis and process control, thereby offering robust decision-making support for expert systems.

The MSI-Petri nets model incorporates multiple information sources, including normalized resource identifiers, temporal attributes, arc state constraints, personnel scheduling logic, and equipment selection transitions, effectively overcoming the limited observability and interpretability inherent in traditional Petri nets within complex manufacturing environments. As a result, it facilitates fine-grained and visually interpretable modeling of production workflows. While this research successfully addresses monitoring requirements in a specific industrial context, its underlying technical framework exhibits strong potential for generalization to other manufacturing domains, offering practical insights for advancing industrial intelligence. Future work will focus on adaptive model optimization, cross-scenario generalization, and the application of the MSI-Petri nets to multi-objective optimization in production systems.

Author Contributions: Conceptualization, P.L. and W.C.; methodology, X.L.; investigation, C.Z.; data curation, C.Z.; writing—original draft preparation, X.L. and W.C.; writing—review and editing, P.L. and Y.K.; visualization, J.Q.; funding acquisition, W.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science and Technology Major Project of China, grant number 2025ZD1607300.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

Acknowledgments: DeepSeek (V3.2) is used to polish the language.

Conflicts of Interest: Author Chenlong Zhang was employed by the company Jilin Zhongwei Protective Equipment 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

1. Yang, H.Y. Balance of mixed flow assembly line based on industrial engineering mathematics and simulated annealing improved algorithm. *Results Eng.* **2024**, *22*, 102071. [CrossRef]
2. Yang, W.; Luo, G.; Li, H. GRU-based real-time scheduling method for production-logistics collaboration in digital twin workshop. *J. Manuf. Syst.* **2025**, *82*, 1269–1289. [CrossRef]
3. Zhang, Q.; Li, H.; Shen, S. Multi-objective optimization of the mixed-flow intelligent production line for automotive MEMS pressure sensors. *Appl. Intell.* **2025**, *55*, 72. [CrossRef]
4. Qian, W.; Guo, Y.; Zhang, H.; Huang, S.; Zhang, L.; Zhou, H.; Fang, W.; Zha, S. Digital twin driven production progress prediction for discrete manufacturing workshop. *Robot. Comput.-Integr. Manuf.* **2023**, *80*, 102456. [CrossRef]
5. Chen, J.; Zhang, Z.; Wang, L.; Tang, D.; Cai, Q.; Chen, K. Self-adaptive production scheduling for discrete manufacturing workshop using multi-agent cyber physical system. *Eng. Appl. Artif. Intell.* **2025**, *150*, 110638. [CrossRef]

https://doi.org/10.3390/s26061785

6. Bofill, J. Exploring Digital Twin-Based Fault Monitoring: Challenges and Opportunities. *Lat. Am. Symp. Depend. Secur. Comput.* **2023**, *23*, 7087. [CrossRef]
7. Yan, W.H. A Review of Real-Time Fault Diagnosis Methods for Industrial Smart Manufacturing. *Processes* **2023**, *11*, 369. [CrossRef]
8. Angleraud, A.; Ekrekli, A.; Samarawickrama, K.; Sharma, G.; Pieters, R. Sensor-based human–robot collaboration for industrial tasks. *Robot. Comput.-Integr. Manuf.* **2024**, *86*, 102663. [CrossRef]
9. Trabelsi, Z.; Alnajjar, F.; Parambil, M.M.A.; Gochoo, M.; Ali, L. Real-Time Attention Monitoring System for Classroom: A Deep Learning Approach for Student's Behavior Recognition. *Big Data Cogn. Comput.* **2023**, *7*, 48. [CrossRef]
10. Qadeer, N.; Shah, J.H.; Sharif, M.; Dahan, F.; Khokhar, F.A.; Ghazal, R. Multi-camera tracking of mechanically thrown objects for automated in-plant logistics by cognitive robots in Industry 4.0. *Vis. Comput.* **2024**, *40*, 9063–9082. [CrossRef]
11. Wang, Z.P. Multi-scale control and action recognition based human-robot collaboration framework facing new generation intelligent manufacturing. *Robot. Comput.-Integr. Manuf.* **2025**, *91*, 102847. [CrossRef]
12. Yan, X.Z.; Zhang, H.; Li, H. Estimating Worker-Centric 3D Spatial Crowdedness for Construction Safety Management Using a Single 2D Camera. *J. Comput. Civ. Eng.* **2019**, *33*, 04019030. [CrossRef]
13. Wang, Y.J.; Niu, X.; Yu, C. FFCI: A Camera and IMU Sensors-Based Multi-Modal Neural Network for Activity Recognition in Smart Factory. *IEEE Trans. Consum. Electron.* **2025**, *71*, 6528–6538. [CrossRef]
14. Andronas, D. Towards seamless collaboration of humans and high-payload robots: An automotive case study. *Robot. Comput.-Integr. Manuf.* **2023**, *83*, 102544. [CrossRef]
15. Yang, X.J. ATD-GCN: A human activity recognition approach for human-robot collaboration based on adaptive skeleton tree-decomposition. *Robot. Comput.-Integr. Manuf.* **2025**, *95*, 103019. [CrossRef]
16. Li, X.; Luo, J.; Li, J.; Yi, S.; Pan, C. Parallel Petri Nets Modeling Method of Manufacturing System Based on the improved PDDL. In Proceedings of the 2022 IEEE International Conference on Networking, Sensing and Control (ICNSC), Shanghai, China, 15–18 December 2022; pp. 1–6.
17. Valdez-Zumba, C.F.; Guerrero-Vásquez, L.F. Fault Management in Speed Control Systems of Hydroelectric Power Plants Through Petri Nets Modeling: Case Study of the Alazán Power Plant, Ecuador. *Energies* **2025**, *18*, 3176. [CrossRef]
18. Kherbache, M.; Ahmed, A.; Maimour, M.; Rondeau, E. Constructing a Network Digital Twin through formal modeling: Tackling the virtual–real mapping challenge in IIoT networks. *Internet Things* **2023**, *24*, 101000. [CrossRef]
19. Du, N.; Yang, Y. APetri net-based approach to robust deadlock prevention in automated manufacturing systems with unreliable resources. *Trans. Inst. Meas. Control* **2024**, *48*, 265–276. [CrossRef]
20. Yang, Y.H.; Liu, X.D.; Lu, W. A Cyber–Physical Systems-Based Double-Layer Mapping Petri Net Model for Factory Process Flow Control. *Appl. Sci.* **2023**, *13*, 8975. [CrossRef]
21. Oliveira Júnior, A.; Calvo-Rolle, J.-L.; Leitão, P. Artificial Intelligence Data-Driven Petri nets Approach for Virtualizing Digital Twins. In Proceedings of the 2023 IEEE International Conference on Industrial Technology (ICIT), Orlando, FL, USA, 4–6 April 2023; Volume 12, pp. 106249–106263.
22. Pla, A.; Gay, P.; Meléndez, J. Petri net-based process monitoring: A workflow management system for process modelling and monitoring. *J. Intell. Manuf.* **2014**, *25*, 539–554. [CrossRef]
23. Liu, H.; Qi, J.; Feng, H. Research on state monitoring and diagnosis models for multi-state systems based on Petri nets. *Results Eng.* **2025**, *27*, 106249. [CrossRef]
24. Chen, L.S.; He, Y.; Zhang, B.; Xia, C.; Qiu, Y.C.; Yu, Z.H. Petri Net-Based Information Enhancement for Quality Monitoring in the Cigarette Manufacturing Process. *Measurement* **2025**, *255*, 117999. [CrossRef]
25. Chen, Q.Y.; He, Y.H.; Yu, G.D.; Xu, C.; Liu, M.Y.; Li, Z.M. KBMP: Kubernetes-Orchestrated IoT Online Battery Monitoring Platform. *IEEE Internet Things J.* **2024**, *11*, 25358–25370. [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.

https://doi.org/10.3390/s26061785

sensors

MDPI

Article

Advancing Defect Detection in Laser Welding: A Machine Learning Approach Based on Spatter Feature Analysis

Gleb Solovev [1], Evgenii Klokov [1,2,*], Dmitrii Krasnov [1,2] and Mikhail Sokolov [1,2]

[1] AI Institute, ITMO University, Kronverksky Prospekt 49, Saint Petersburg 197101, Russia; gvsolovev@itmo.ru (G.S.); dmitrii.krasnov@itmo.ru (D.K.); mikhail.sokolov@itmo.ru (M.S.)
[2] Institute of Laser Technologies, ITMO University, Kronverksky Prospekt 49, Saint Petersburg 197101, Russia
* Correspondence: 336639@niuitmo.ru

Abstract

Full-penetration laser welding (FPLW) is increasingly adopted in manufacturing pipelines, yet its industrial scalability is constrained by in-process defect formation, particularly incomplete penetration. To address this, we propose a sensor-driven framework for non-destructive monitoring and automated defect detection that uses infrared (IR) thermography as the primary in situ sensing modality and applies deep learning to the acquired thermal signals. High-speed IR camera recordings were processed to track spatter and the weld zone, yielding a time series of physically interpretable spatiotemporal features (mean spatter area, mean spatter temperature, number of spatters, and mean welding zone temperature). Defect recognition is formulated as a multi-label classification problem targeting incomplete penetration, sagging, shrinkage groove, and linear misalignment, and multiple temporal models were evaluated on the same sensor-derived feature sequences. Experimental validation on 09G2S pipeline steel demonstrates that the proposed time series pipeline based on a hybrid CNN–transformer achieves a mean Average Precision (mAP) of 0.85 while preserving near-real-time inference on a CPU. The results indicate that IR thermography-based spatter dynamics provide actionable sensing signatures for automated defect prediction and can serve as a foundation for closed-loop quality control in industrial laser pipeline welding.

Keywords: machine learning; neural networks; quality assurance; signal processing; laser beam welding

1. Introduction

Laser Beam Welding (LBW) is widely employed in numerous industrial sectors due to its high precision, deep penetration capabilities, and overall process efficiency. Considerable research efforts have been devoted to further optimising this technique, as LBW is increasingly recognized as one of the most promising joining methods available. This is largely attributable to several distinctive advantages: (1) the non-contact nature of the process; (2) the ability to perform single-side joining; and (3) the exceptionally high energy density of the laser beam, which enables the formation of a weld within a fraction of a second. These characteristics contribute to a reduced heat-affected zone and, consequently, minimal deformation and residual stresses [1].

In addition to these advantages, LBW provides reduced heat input, minimal thermal distortion, and improved joint integrity compared with conventional welding techniques. These attributes make it particularly suitable for applications that demand high-quality and

highly reliable welds [2]. Among the various laser welding configurations, full-penetration laser welding is often considered a favourable option due to its capacity to produce single-pass welds, thereby increasing productivity and reducing operational costs [3].

In sectors such as oil and gas, energy, and water distribution, pipeline welding is critical to ensuring both the structural integrity and the long-term operational reliability of pipeline systems. Historically, pipeline fabrication has relied on arc-based techniques—including gas metal arc welding and submerged arc welding (SAW)—which typically require joint grooving, multiple welding passes, and extensive post-weld inspection procedures [4]. Although these methods have proven effective, they remain time-consuming and labour-intensive. By contrast, FPLW significantly reduces filler metal consumption—by up to 90% compared with SAW—and offers welding speeds that are three to five times higher [5].

Despite the well-documented advantages of laser welding, its widespread adoption in pipeline manufacturing and maintenance is still limited by several technical challenges. A major concern is the formation of weld defects—such as spattering, porosity, and keyhole instability—which may compromise joint strength and long-term performance [6]. Among these, spattering is considered a particularly critical phenomenon, as it can lead to contamination of the weld seam, material loss, and defects such as undercutting and inclusions [7].

Although a substantial body of research has examined the influence of welding parameters on spatter dynamics [8], there is still a gap in the development of real-time defect detection methods that will lead to enabling process control and continuous weld quality monitoring. Recent studies have investigated a range of factors affecting spatter generation, including keyhole behaviour, evaporation-driven recoil pressure, and melt pool dynamics. Correlations between spatter characteristics and key process parameters—such as laser power, as was shown by Cheng et al. [9] and Zoe et al. [10]; welding speed; and shielding gas conditions—have been well established in the laser welding literature, where variations in these parameters are shown to significantly influence spatter generation, morphology, trajectory, and thermal behaviour. Zhang et al. [11] analysed the influence of penetration mode and beam configuration on spatter network formation, and Wang et al. [12] clarified the dependence of oscillation parameters and heat input conditions on spatter behaviour. While these investigations have yielded valuable insights into the underlying physical mechanisms, there remains a need for advanced, data-driven techniques capable of extracting quantitative spatter features to support real-time prediction of defect formation.

Pipeline materials, particularly modern high-strength steels used in oil, gas, and hydrogen pipeline transportation, are explicitly designed with weldability as a critical engineering consideration. However, the problem of controllable laser penetration depth and other common laser welding defects remains a challenge [13].

In the FPLW of pipeline steels, defects such as porosity and cracking are relatively less prevalent due to the high weldability of these materials and their optimisation for joining processes. However, incomplete penetration or lack of fusion remain significant and comparatively common defects. These defects typically arise from suboptimal process conditions, including insufficient heat input, weld misalignment, or improper focal positioning, and can markedly compromise the structural integrity of the welded joint [14,15]. Addressing these issues requires a comprehensive understanding of the interplay between welding parameters and defect formation, as well as the development of advanced real-time monitoring and control systems to ensure consistent weld quality.

A variety of Weld Penetration Depth Control (WPDC) methods have been proposed in the literature, most of which rely on indirect process signals or require specific modifications to the welding setup. One of the earliest closed-loop WPDC approaches involved

https://doi.org/10.3390/s26061825

plasma analysis. Sibillano et al. [16] demonstrated that real-time correlation between electron temperature measured via spectrometry and the reference penetration depth in CO_2 laser welding could be used for control purposes. However, this method has a notable limitation—it requires precise calibration for each welding configuration, material, and desired penetration depth, thereby restricting its industrial flexibility. Harooni et al. [17] proposed a similar strategy based on correlating plasma ionisation temperature with penetration depth. Yet, as noted by You et al. [18], fluctuations in ionisation temperature can themselves indicate weld imperfections, which constrains the suitability of this approach for WPDC in industrial environments.

Other in-process monitoring and control techniques have leveraged photodiodes to collect signals such as thermal radiation, ultraviolet emissions, laser back-reflection, and visible light from the weld zone [19].

Acoustic signal processing offers significant potential for extracting meaningful information from the laser welding process. To better understand the relationship between optical emissions and airborne acoustic responses, Hoffman et al. [20] conducted a combined spectroscopic and acoustic analysis, demonstrating that characteristic audio frequencies depend on both the material type and its thickness. Further evidence was provided by Yusof et al. [21], who concluded that acoustic methods are suitable for online monitoring and control of weld penetration depth. However, the authors also highlighted that the inherent sensitivity of acoustic measurements to ambient noise presents a major limitation for their robust deployment in industrial environments. Complementary findings were reported by Fischer et al. [22], who showed, through experiments on glass and metal sheets, that high-frequency airborne ultrasound constitutes a potent technique for weld quality assessment and WPDC. This conclusion is supported by Authier et al. [23], who demonstrated that acoustic analysis can be further enhanced when combined synergistically with other sensing modalities.

As an alternative line of development, the transition from conventional LBW to hybrid welding processes expands the range of feasible monitoring and control strategies. For example, Bîrdeanu et al. [24] investigated WPDC in a hybrid welding configuration and demonstrated that the frequency of the TIG arc can be used as a reliable indicator of weld penetration depth. While relevant to the broader field of process control, a comprehensive review of hybrid or non-remote laser welding techniques lies outside the scope of the present study.

Despite their diversity, the methods described above share a common set of challenges, most notably the need for extensive databases of reference signals collected under tightly controlled conditions. Although these approaches remain promising—especially when viewed in light of recent advances in machine learning and artificial intelligence—the exclusive reliance on data-driven techniques is insufficient for industrial implementation. The inherent variability of real-world laser welding processes, including material composition, thickness, laser characteristics, joint configuration, and gap conditions, cannot be reliably captured by data alone. As noted in [25], effective WPDC solutions for industry require approaches that can accommodate these variations while maintaining robustness, adaptability, and real-time operability.

Analysing various methods for in-process quality control during laser welding, You et al. [18] suggested that the most reliable direct WPDC method remains X-ray imaging. However, the practical challenges associated with implementing X-ray systems in industrial environments—particularly concerning safety requirements, equipment complexity, and integration constraints—significantly restrict their applicability [26]. In their comprehensive investigation of weld penetration depth monitoring techniques, Bautze et al. [27] concluded

https://doi.org/10.3390/s26061825

that, aside from X-ray imaging, the only method capable of providing a direct, in-process measurement of keyhole depth is Optical Coherence Tomography.

Within the context of real-time defect detection in FPLW, the selection of sensing technology plays a decisive role in achieving accurate monitoring of process stability and weld quality. A wide range of sensors has been explored for weld diagnostics; however, infrared cameras offer advantages due to their ability to capture thermal emissions from the weld pool and from spatter dynamics. In contrast to visible-range imaging systems, which may be affected by process-related obstructions such as fumes, vapour plumes, or plasma, IR imaging provides a robust and contactless means of visualising temperature fluctuations and identifying thermal anomalies that are potentially associated with weld defects. Furthermore, IR cameras enable high-speed acquisition of spatter behaviour, allowing for the extraction of discriminative features such as spatter temperature, trajectory, and cooling rate—parameters that are examined in this study for their correlation with weld quality. By integrating IR camera-based data acquisition with machine learning-enabled analysis, the present work aims to advance defect prediction capabilities and ultimately contribute to improved process monitoring in FPLW.

Recent advancements in machine learning (ML) and computer vision have opened new possibilities for data-driven weld quality assessment. Various researchers have utilized automated vision systems to extract quantitative spatter metrics; for instance, Schweier et al. [28] developed a machine vision approach using Kalman filter-based tracking to identify spatter formation mechanisms during beam oscillation. Huang et al. [29] employed motion tracking to statistically analyze spatter ejection directions and velocities, linking them to vapor plume pressure. Furthermore, Hartung et al. [30] demonstrated the efficacy of deep learning-based semantic segmentation for real-time spatter detection and its use as a signal for defective seams. Given the stringent safety and durability requirements of pipeline infrastructure, integrating ML-based spatter analysis into defect detection frameworks could substantially enhance weld reliability while minimizing reliance on post-weld inspections.

A diverse range of ML techniques has been employed to establish relationships between sensor-derived features and defect occurrences during laser welding [31]. Classical ML methods, such as decision tree (DT), linear regression, random forest (RF), logistic regression (LR), gradient boosting (GB), and support vector machine (SVM), remain widely adopted in industrial applications due to their interpretability, relatively low computational cost, and robustness in structured data environments. For instance, Xiao et al. [32] utilized linear regression to predict welding depth using coaxial pyrometer data. Chianese et al. [33] applied DT, RF, and SVM to classify weld defects (e.g., overpenetration, lack of fusion) based on photodiode signal features. Zhao et al. [34] demonstrated the efficacy of RF models in monitoring weld penetration via laser keyhole imaging. Paulson et al. [35] employed LR and GB to link thermal history with keyhole porosity in laser powder bed fusion.

Beyond these classical approaches, modern deep learning (DL) methods offer advanced capabilities in handling complex, high-dimensional data. Multilayer perceptrons (MLPs) and convolutional neural networks (CNNs) have demonstrated promise in welding defect analysis [36]. Gao et al. [37] successfully achieved multi-class classification of five weld defect types using keyhole images and an MLP framework. Zhang et al. [38] implemented a CNN to monitor weld penetration states directly from raw pixel data captured by coaxial vision systems. These advancements underscore the potential of ML and DL techniques in enabling automated, high-precision weld quality assessment.

In addition to these domain-specific developments, it is instructive to consider methodologies established in broader fields such as video classification and time series analysis, as the task addressed in this study is fundamentally aligned with the classification of dynamic

https://doi.org/10.3390/s26061825

spatiotemporal patterns. Spatter behaviour, as captured by IR camera imaging, exhibits evolution in both space and time. As such, the analysis may leverage either (i) embeddings extracted directly from raw pixel sequences or (ii) engineered features organised into time series representations. These perspectives provide a conceptual bridge between welding-specific monitoring approaches and state-of-the-art ML architectures developed for sequence modelling and temporal pattern recognition, forming a foundation for the approach proposed in this work.

In the field of video classification, widely adopted baseline architectures typically rely on convolutional neural networks employing either 2D or 3D convolutional layers. For example, Hara et al. [39] introduced a three-dimensional extension of the ResNet architecture, demonstrating that 3D convolutions substantially improve video classification accuracy by explicitly capturing spatiotemporal dependencies. Nevertheless, studies such as [39,40] also acknowledge that 2D convolutional models remain capable of performing video classification tasks, albeit with reduced accuracy, despite their inherent limitation in modelling temporal evolution directly. These observations provide valuable context for selecting suitable architectures for analysing spatter dynamics in infrared video data, where both spatial structure and temporal progression are essential for reliable defect detection.

For time series analysis, the task involves processing sequences of one-dimensional signals. A classical approach is to construct a feature vector based on aggregated statistics (e.g., mean, standard deviation, skewness, or kurtosis) within a sliding window, followed by classification using basic ML or DL models. Currently, 1D CNNs [41], transformers [42], recurrent networks with Long Short-Term Memory (LSTM) cells [43], and hybrid architectures combining convolutional and transformer layers [44] are widely used for this purpose.

1. Characterize the relationship between spatter dynamics and weld defects in laser welding.
2. Develop a video processing framework for extracting spatiotemporal features from spatter dynamics.
3. Implement and evaluate a deep learning pipeline for classifying weld quality based on spatter behavior.
4. Validate the proposed methodology through experimental pipeline steel laser welding trials and assess its effectiveness in defect prediction and process optimization.

The innovation of this work lies in systematically defining a multi-dimensional spatiotemporal spatter feature set—including mean spatter area, mean spatter temperature, number of spatters, and mean welding zone temperature—specifically for 09G2S pipeline steel, and for the first time evaluating the correlation between these features and various defects to achieve a real-time classification model. Notably, the proposed method operates on a non-specialized thermal camera, providing a cost-effective and accessible alternative to expensive proprietary monitoring systems from vendors such as Precitec, Trumpf, or IPG.

2. Materials and Methods

2.1. Laser Welding Setup

The experiments were conducted using 09G2S structural pipeline steel samples of dimensions $3 \times 150 \times 450$ mm. The chemical composition and mechanical properties of the steel are given in Table 1.

Materials were welded by continuous wave multi-mode MAXPhotonics MFSC-3000 (Maxphotonics Co., Ltd., Shenzhen, China) in a butt joint welding setup. All experiments were performed with Argon shielding gas at a 20 L/min flowrate and without filler wire. The welding setup is shown in Figure 1 and equipment specification is shown in Table 2.

https://doi.org/10.3390/s26061825

The samples were mechanically cleaned before welding to remove rust and other possible weld contaminants.

Table 1. Chemical composition (wt., %) and mechanical properties of 09G2S grate steel according to GOST 19281-2014 [45].

C	Si	Mn	Ni	S	P	Cr	V	N	Cu	As
<0.12	0.5–0.8	1.3–1.7	<0.3	<0.035	<0.03	<0.3	<0.12	<0.008	<0.3	<0.08
Tensile strength (MPa)						**Elongation (%)**				
≥490						≥21				

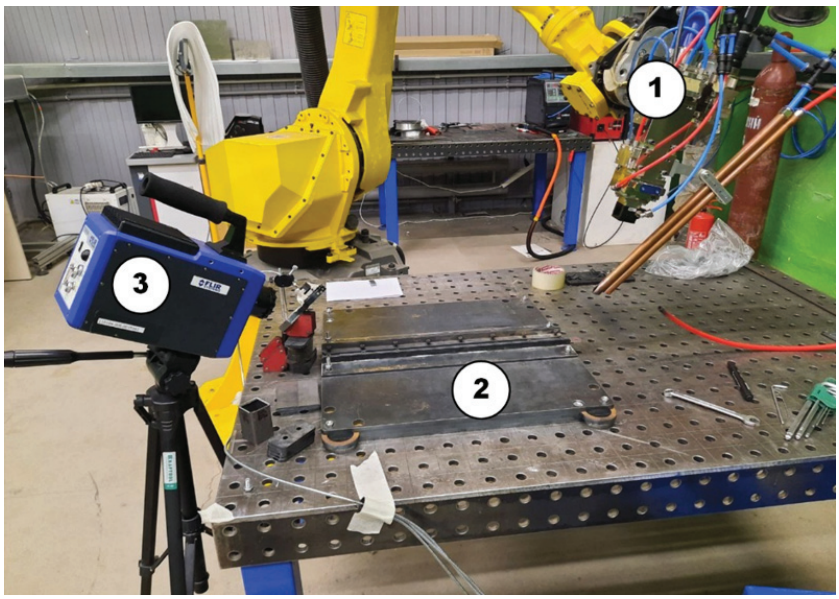


Figure 1. Experimental setup: (1) a WSX ND40 welding head with shielding gas pipes; (2) steel samples with clamping system; and (3) a FLIR Systems Titanium 520 M Camera (FLIR Systems, Inc., Wilsonville, OH, USA) installed at a distance of 50 cm from the weld zone.

Table 2. Specification of experimental setup.

MAXphotonics MFSC-3000	**Units**	**Levels**
Nominal output power	W	3000
Optical fibre diameter	µm	100
Spot diameter at focus	mm	0.14
Welding optics: WSX ND40		
Collimating length	mm	100
Focusing length	mm	250
Emission wavelength	nm	1080
FLIR Systems CEDIP TITANIUM 520 M Camera		
Resolution	pixel	320 × 240
Frequency	Hz	380
Spectral range	µm	1.5–5.0

After welding, the samples were cut at four places: 72, 152, 232, and 312 mm from the weld start, marked with letters (A, B, C, and D) as shown in Figure 2. Once the samples were extracted and polished, they were etched in a 4% Nital solution and the cross-sections were photographed using a LOMO MSP-1 microscope (LOMO (Leningrad Optical Mechanical

https://doi.org/10.3390/s26061825

Association), Saint Petersburg, Russia). The geometry of the weld and the weld defects were recognized on the cross-section photographs.

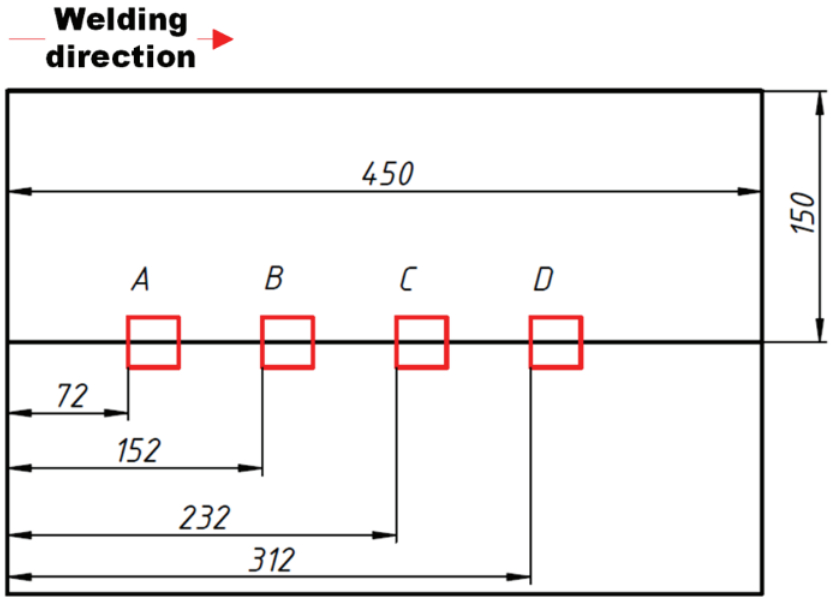


Figure 2. Sample cut locations.

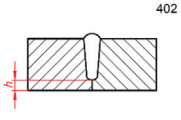
2.2. Design of Experiments

The experiments were focused on investigation of the ability to detect weld defects with an ML approach and analysis of the welding process parameters on the selected laser welding defects. A multi-level factorial design was employed, varying two key parameters—laser power (P_L) and welding speed (V_w)—at three levels each. Other parameters—the focal position (F_z) relative to the sample surface and the inclination angle (α) of the laser head—were maintained at fixed values. The experimental plan included a minimum of three replication points per condition, resulting in a total of 34 welding trials. From these trials, 136 macrosections were extracted and analysed. The experimental parameters, factors, and levels are summarized in Table 3. Analysis of these macrosections revealed several weld defects. The types, corresponding standards, frequencies, and visual examples of these defects are detailed in Table 4.

Table 3. Experimental parameters, factors, and levels.

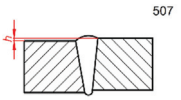
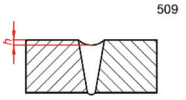
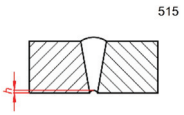
Constant Welding Parameter		**Value**	
F_z, mm		0	
α, °		15	
Factors and levels			
Factor	**Level 1**	**Level 2**	**Level 3**
P_L, W	1500	2000	2500
V_w, mm/s	30	35	40

Table 4. Investigated welding defects and their occurrence in the experimental set.

Defect Designation	**ISO 6520-1 [46]**	**ID**	**Occurr.**	**Illustration**
Incomplete penetration	402	h_i	17%	

https://doi.org/10.3390/s26061825

Table 4. *Cont.*

Defect Designation	ISO 6520-1 [46]	ID	Occurr.	Illustration
Linear misalignment	507	h_m	3%	
Sagging	509	h_s	5%	
Shrinkage groove	515	h_g	1%	

2.3. Feature Extraction from Spatter

Detecting spatters during laser welding is complicated by interfering radiation from the metal vapor plume and the background thermal emission of the workpiece. To reduce this interference, a background suppression filter is applied prior to spatter tracking. This filter enhances spatter-related radiation while attenuating the general background heat. The algorithm employs directional Laplacian-of-Gaussian (LoG) filtering combined with a minimum selection method as proposed by Kim [47]. Equation (1) show the overall process of generating directional (east, west, south, and north) LoG kernels, where (x, y) denotes a pixel position, σ is the standard deviation of the LoG kernel, and N is the kernel size ($N = 9$ and $\sigma = 1.8$ in our experiments). Equation (2) show the process of extracting directional feature maps with the generated LoG kernels via directional convolution. The final feature map was obtained as the minimum of the directional feature maps at each pixel position.

$$f_E(x) = \frac{1}{\sqrt{2\pi}\sigma^3}\left(1 - \frac{x^2}{\sigma^2}\right)\exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (1)$$

$$\text{where } x = 0, 1, 2, \ldots, N$$

$$f_W(x) = \frac{1}{\sqrt{2\pi}\sigma^3}\left(1 - \frac{x^2}{\sigma^2}\right)\exp\left(-\frac{x^2}{2\sigma^2}\right)$$

$$\text{where } x = 0, -1, -2, \ldots, -N$$

$$f_S(y) = \frac{1}{\sqrt{2\pi}\sigma^3}\left(1 - \frac{y^2}{\sigma^2}\right)\exp\left(-\frac{y^2}{2\sigma^2}\right)$$

$$\text{where } y = 0, 1, 2, \ldots, N$$

$$f_N(y) = \frac{1}{\sqrt{2\pi}\sigma^3}\left(1 - \frac{y^2}{\sigma^2}\right)\exp\left(-\frac{y^2}{2\sigma^2}\right)$$

$$\text{where } y = 0, -1, -2, \ldots, -N$$

$$I_E(x, y) = I(x, y) \otimes f_E(x) \quad (2)$$

$$I_W(x, y) = I(x, y) \otimes f_W(x)$$

$$I_S(x, y) = I(x, y) \otimes f_S(y)$$

$$I_N(x, y) = I(x, y) \otimes f_N(y)$$

This combination effectively removes false detections near the vapor plume edges and improves spatter visibility. Representative examples of the raw and filtered image frames are shown in Figure 3a and Figure 3b, respectively. Following preprocessing, the filtered

https://doi.org/10.3390/s26061825

frame is binarized using a constant threshold to produce a binary mask that highlights both detected spatters and the welding zone, as illustrated in Figure 3c. This binary mask is subsequently segmented into individual spatter regions, from which spatial coordinates and size parameters are extracted. Using the same coordinates, corresponding intensity values (which reflect temperature) are sampled from the original thermal data. For each frame, the following four spatiotemporal features are computed:

- Mean spatter area [spatter area]—the average area of spatter bounding boxes;
- Mean spatter temperature [spatter temp]—the average digital level from the FLIR camera pixels within spatter regions;
- Number of spatters per frame [number of spatters]—the total count of detected spatter regions;
- Mean welding zone temperature [welding zone temp]—the average digital level within the detected weld pool region.

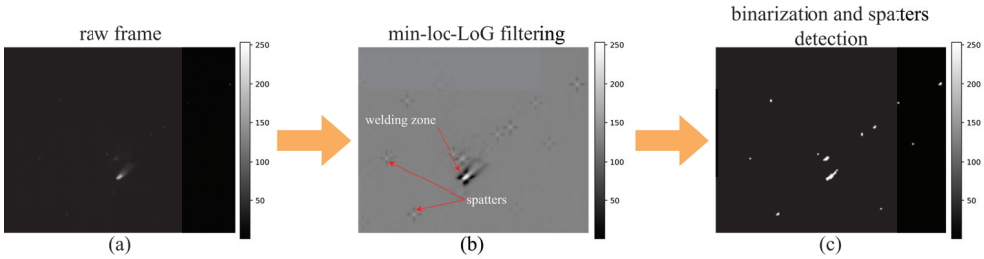


Figure 3. Image processing pipeline: (**a**) raw frame; (**b**) filtered feature map; and (**c**) binarized feature map.

These features were designed to encapsulate both the dynamic behaviour and the thermal characteristics of the welding process, thereby enabling the identification of patterns associated with weld quality and defect formation. The extracted features were computed for each frame, producing a continuous time series of spatiotemporal descriptors (Figure 4a).

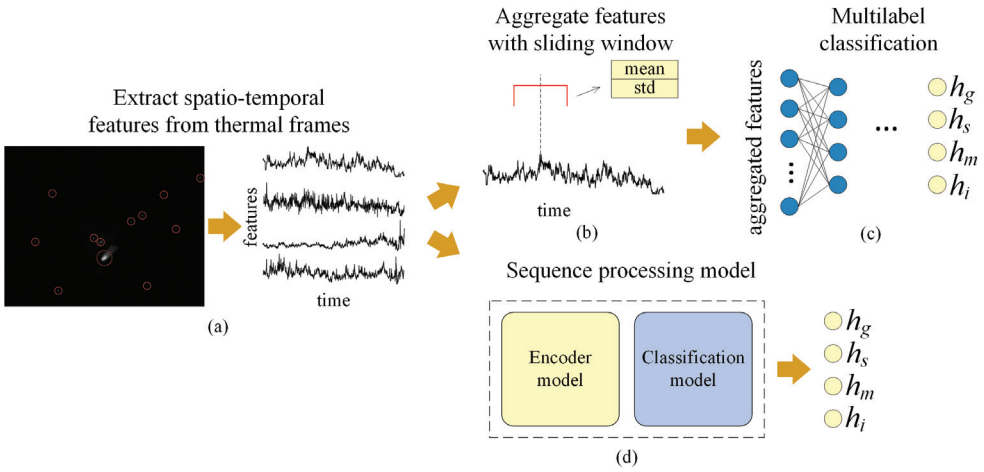


Figure 4. Presented defect prediction pipeline: (**a**) feature extraction; (**b**) feature aggregation; (**c**) classification with aggregated features; and (**d**) fixed-length sequence processing.

2.4. Deep Learning Approach

We formulated the defect classification task as a multilabel problem with four target classes: shrinkage groove (h_g), sagging (h_s), linear misalignment (h_m), and incomplete penetration (h_i). Multiple defects can co-occur within a single sample, and therefore we developed two distinct approaches to classify them based on extracted spatiotemporal

https://doi.org/10.3390/s26061825

features. These approaches correspond to different data preprocessing strategies, illustrated in Figure 4b,d.

The first approach aggregated per-frame feature time series within a sliding temporal window. For each window, the mean and standard deviation of the four extracted features were computed, yielding a fixed-dimensional representation (Figure 4b). Window size was optimized as a key hyperparameter to maximize classification performance. A custom MLP architecture with parallel class-specific heads classified these aggregated feature vectors (Figure 5a). The individual head outputs were concatenated and passed through a final shared linear layer to generate multi-label predictions (Figure 4c).

The second approach processed fixed-length sequences of per-frame features directly, without temporal aggregation (Figure 4d). A suite of sequence processing models was developed and evaluated, each comprising a temporal encoder for feature extraction followed by a linear classifier for final prediction.

Four encoder architectures were implemented: convolutional neural networks, long short-term memory networks, transformers, and a hybrid CNN–transformer architecture (Figure 5). Each encoder operated on identical input feature sequences, enabling direct comparison of temporal modelling capabilities. Key hyperparameters were systematically tuned across all models to optimize evaluation metrics, with the final network architectures illustrated in Figure 5.

Model performance was evaluated using stratified four-fold cross-validation with video-wise splits to prevent data leakage. Since each welding trial corresponds to a unique thermogram, partitioning was performed at the video level to ensure that all temporal segments derived from a single physical weld were assigned exclusively to the same fold. This procedure eliminates any correlation between training and validation sets at the weld level.

Stratification was performed based on the presence or absence of incomplete penetration (h_i) defects in each video in order to preserve class distribution across folds and mitigate class imbalance during training and validation. Notably, different defect types exhibit significant correlations; the presence of a given defect is often associated with the occurrence of other defect types. This correlation reinforces the need for stratification to ensure representative sampling of all defect combinations across folds. Alternative imbalance-handling techniques, including Focal Loss and resampling, were explored in preliminary experiments but did not yield measurable improvements, and were therefore not adopted. We attribute this to the fact that the engineered feature set already provides high class separability for the majority of samples, and the introduction of such techniques risked over-emphasizing rare noise patterns or outliers in the minority classes.

Temporal data were extracted with a 50% overlap, employing sliding windows in the first scheme and fixed-length sequences in the second. Preliminary experiments demonstrated that Z-score and Min–Max normalization degraded performance; consequently, raw feature values were used as direct model input. Normalization reduced classification accuracy, as it removed the absolute physical magnitudes that encode meaningful distinctions between defect types and welding conditions. Since the four descriptors are defined in physically meaningful units and already operate on a compact scale, global normalization eliminates these absolute differences, which are essential for reliable defect classification.

The risk of overfitting is mitigated by the compact architecture of the models, with the transformer and CNN–transformer comprising 1.65 M and 1.86 M parameters, respectively. By processing low-dimensional sequences of four engineered features instead of high-dimensional raw pixel data, the networks are constrained to learn essential spatiotemporal interactions between physical descriptors. This reduction in both parameter space and

https://doi.org/10.3390/s26061825

input dimensionality ensures robust generalization and stable convergence despite the modest dataset size.

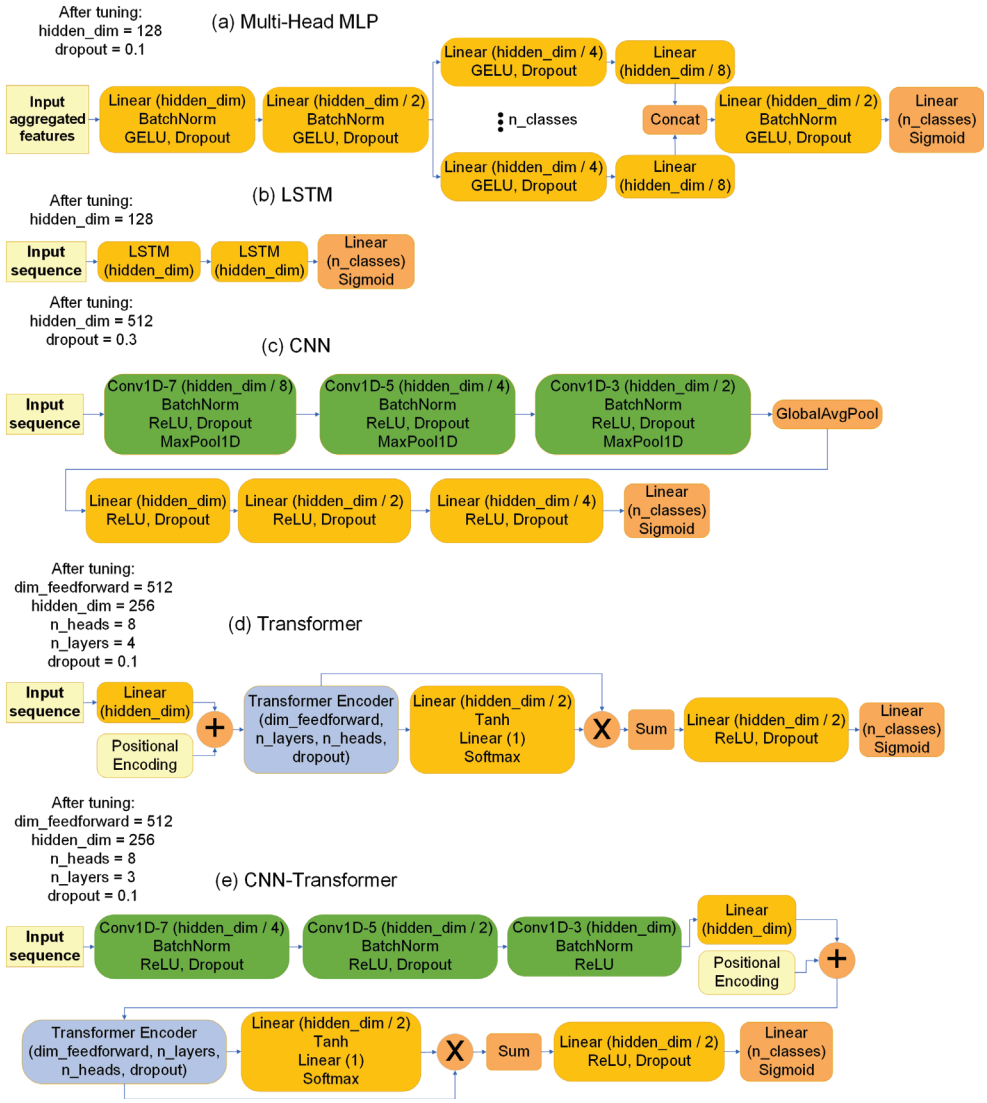


Figure 5. Architectures of the networks under comparison: (**a**) multi-head MLP; (**b**) LSTM; (**c**) CNN; (**d**) transformer; and (**e**) CNN–transformer.

Approaches were assessed using mAP and standard deviation across validation folds for window/sequence lengths of 8, 16, 32, 64, 96, and 128. Per-class performance was evaluated using maximum F1-score, precision, and recall at optimized decision thresholds. All experiments were implemented in Python 3.10 with the PyTorch 2.6 and Timm 1.0.24 frameworks. Training was performed on an NVIDIA GeForce RTX 2060 GPU (6 GB), while inference speed was benchmarked on an Intel Core i5-11600K CPU (3.90 GHz). Classifiers were trained using the AdamW optimizer (learning rate $= 1 \times 10^{-4}$, no decay), batch size 64, and binary cross-entropy loss.

3. Results

To assess the applicability of the proposed DL approach, we quantified the statistical relationships between the extracted spatter features and the observed weld defects. As the Kolmogorov–Smirnov test indicated that the feature distributions were non-normal, we employed Spearman's rank correlation analysis. Features were extracted and aggregated according to the method outlined in Section 2.3, utilizing a sliding window of 64 frames

https://doi.org/10.3390/s26061825

(Figure 4b). The resulting correlation coefficients are presented in Table 5, where coefficients with a p-value ≤ 0.05 are marked with an asterisk (*) to denote statistical significance.

Table 5. Spearman's feature–defect correlation coefficients and p-values.

Defect	Size Mean	Size Std	Number of Spatters Mean	Number of Spatters Std	Spatter Temp Mean	Spatter Temp Std	Welding Zone Temp Mean	Welding Zone Temp Std
h_g	0.08	−0.09	0.22 *	0.20 *	0.00	−0.27 *	−0.56 *	−0.39 *
h_s	−0.01	0.01	0.04	0.09	0.03	−0.03	−0.19 *	−0.07
h_m	0.16 *	0.06	0.00	−0.05	0.14 *	−0.14 *	−0.05	0.02
h_i	−0.35 *	0.10	−0.43 *	−0.40 *	−0.28 *	0.37 *	0.59 *	0.50 *

As shown in Table 5, the strongest correlations were observed for incomplete penetration (h_i) and shrinkage groove (h_g). For these defects, the most informative features were associated with spatter quantity and welding zone temperature, whereas features related to spatter area and spatter temperature showed weaker associations. However, the final classifiers achieved high predictive accuracy for sagging (h_s) and linear misalignment (h_m), suggesting that the full set of temporal features provides a richer detection signal than individual pairwise correlations alone. We performed additional ablation study to investigate the influence of each feature on the final classification precision.

Next, five DL pipelines (Methods; Figure 5) were trained and evaluated using multiple window/sequence lengths. The macro-averaged mAP and the corresponding parameter counts are summarized in Table 6.

Table 6. Mean Average Precision of Deep Learning Approaches.

Model	Sequence Size					
	8	**16**	**32**	**64**	**96**	**128**
Multi-Head MLP (0.43 M)	0.80 ± 0.06	0.82 ± 0.06	$\mathbf{0.83 \pm 0.05}$	0.83 ± 0.07	$\mathbf{0.83 \pm 0.07}$	0.83 ± 0.06
CNN (0.44 M)	0.79 ± 0.07	0.81 ± 0.08	0.81 ± 0.07	0.82 ± 0.07	0.78 ± 0.03	0.81 ± 0.08
LSTM (1.32 M)	0.80 ± 0.06	0.81 ± 0.06	0.81 ± 0.06	0.82 ± 0.05	0.83 ± 0.05	0.82 ± 0.09
Transformer (1.65 M)	0.79 ± 0.05	0.78 ± 0.05	0.80 ± 0.04	0.78 ± 0.05	0.78 ± 0.06	0.77 ± 0.05
CNN–Transformer (1.86 M)	$\mathbf{0.81 \pm 0.07}$	$\mathbf{0.82 \pm 0.07}$	0.82 ± 0.06	$\mathbf{0.85 \pm 0.04}$	0.82 ± 0.06	$\mathbf{0.83 \pm 0.07}$

The bold font highlights the best results in the table.

The results in Table 6 indicate that the hybrid CNN–transformer achieved the best overall performance, reaching an mAP of 0.85 at a sequence length of 64. This architecture also has the largest number of parameters among the tested models (1.86 M). In contrast, the simpler multi-head MLP, with substantially lower complexity (0.43 M parameters), delivered competitive performance (0.83 mAP), thereby offering a favorable efficiency–accuracy trade-off. Although the optimal sequence length differed across models, lengths between 32 and 96 frames generally provided the best balance between capturing sufficient temporal context and preserving an adequate number of training samples. Figure 6a shows the precision–recall curves of the best-performing CNN–transformer, aggregated across all validation folds and reported per defect class using micro-averaging. To select an operating point for deployment, we optimized the decision threshold by computing precision, recall, and the F1-score for thresholds from 0 to 1 in increments of 0.05. As shown in Figure 6b, the maximum F1-score occurred at a threshold of 0.55, which we therefore used as a general operating point. For specific industrial scenarios, this threshold can be adjusted to prioritize either higher precision or higher recall, depending on operational requirements.

https://doi.org/10.3390/s26061825

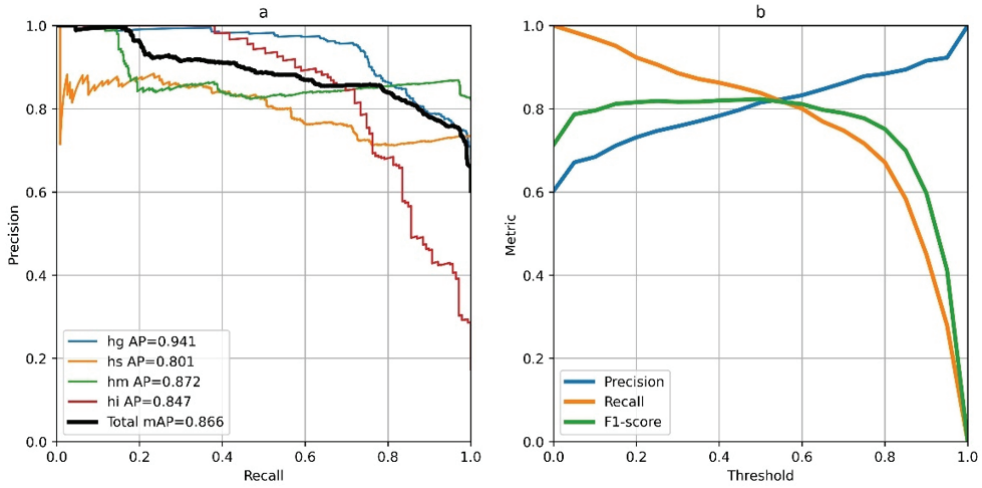


Figure 6. Evaluation curves of CNN–transformer: (**a**) precision–recall curves; (**b**) threshold–metric curves.

We performed additional ablation study on feature importance to investigate whether the full set of temporal sequences could provide a richer detection signal than individual pairwise correlation presented in Table 5. During these experiments we trained and evaluated the proposed CNN–transformer as described in Section 2.4 using all possible feature combinations with a sequence length equal to 64 (the best result in Table 6). The ablation results are presented in Table 7 as per-class fold-wise Average Precision (AP) and total mAP for each feature combination sorted by mAP. Fold-wise standard deviation was omitted to make the results less cumbersome.

Table 7. Ablation study on the feature importance.

Feature Combination	**AP**				**mAP**
	h_g	h_s	h_m	h_i	
spatter area + number of spatters + spatter temp + welding zone temp	0.94	0.85	0.87	0.72	0.85
number of spatters + spatter temp + welding zone temp	0.95	0.78	0.84	0.78	0.84
spatter area + spatter temp + welding zone temp	0.95	0.80	0.85	0.71	0.83
spatter temp + welding zone temp	0.94	0.83	0.86	0.68	0.83
spatter area + number of spatters + welding zone temp	0.94	0.78	0.88	0.71	0.83
number of spatters + welding zone temp	0.94	0.77	0.88	0.71	0.83
welding zone temp	0.95	0.76	0.82	0.76	0.82
spatter area + welding zone temp	0.94	0.75	0.86	0.73	0.82
number of spatters + spatter temp	0.83	0.74	0.87	0.56	0.75
spatter area + number of spatters + spatter temp	0.85	0.75	0.90	0.50	0.75
spatter area + number of spatters	0.84	0.71	0.89	0.55	0.75
number of spatters	0.81	0.73	0.88	0.56	0.75
spatter area + spatter temp	0.80	0.76	0.86	0.55	0.74
spatter temp	0.80	0.74	0.86	0.54	0.73
spatter area	0.75	0.74	0.88	0.54	0.73

These results indicate strong correlation between spatter features and defects which is enough to classify defects based even on a single feature despite the relatively small pairwise correlation presented in Table 5 (especially for sagging and linear misalignment). Although single features achieved good performance (0.73–0.82 mAP), their right combinations produced better results (0.85 mAP for spatter area + number of spatters + spatter temp + welding zone temp). Notably, the temperature of the welding zone was the most informative feature as it outperformed some combinations of other features, whereas the

https://doi.org/10.3390/s26061825

spatter area was less informative. The combination of all features was used to produce the final results.

For comparison, we implemented two baseline models—MobileNetV4 Small [48] backbone with a two-layer linear classifier (1280 and 640 neurons) and 3D Convolutional ResNet-18 [39]—to compare against the proposed feature-based CNN–transformer pipeline. To ensure a fair comparison, both baselines processed thermograms at their native resolution (320 × 240 pixels) using the maximum batch size permitted by the hardware (64 for MobileNetV4 and 16 for 3D ResNet-18). Both baselines were trained using the Adam optimizer with binary cross-entropy loss, where the learning rate was set to 7×10^{-5} after being selected from a range of 10^{-3} to 10^{-6} with a step of 10 and no decay. Table 8 reports the AP, precision, recall, and F1-score for each defect class at the 0.55 decision threshold, presented as mean ± standard deviation across validation folds for the proposed model and both baselines. Table 8 also includes the mean processing speed per thermogram frame measured from the preprocessing start (the division on 255 for MobileNetV4 and 3D ResNet18, the pipeline in Figure 4a,d for the CNN–transformer) to the end of model's inference over 1000 runs on the CPU.

The standard deviations across folds remain low for most defect classes, indicating overall model stability. However, the incomplete penetration (h_i) class exhibits significantly higher variability, suggesting that its detection performance is more sensitive to specific data splits.

Table 8. Per-class evaluation results.

Model	Metric	Classes				Macro Avg	Speed ms
		h_g	h_s	h_m	h_i		
CNN–Transformer (1.86 M)	AP	0.94 ± 0.02	0.86 ± 0.14	0.87 ± 0.09	0.72 ± 0.15	0.85	15.29
	Precision	0.87 ± 0.07	0.73 ± 0.20	0.87 ± 0.05	0.88 ± 0.13	0.84	
	Recall	0.81 ± 0.06	0.81 ± 0.14	0.97 ± 0.04	0.62 ± 0.30	0.80	
	F1-score	0.83 ± 0.01	0.74 ± 0.11	0.92 ± 0.02	0.67 ± 0.26	0.79	
MobileNetV4 (1.78 M)	AP	0.95 ± 0.03	0.84 ± 0.09	0.80 ± 0.09	0.75 ± 0.38	0.83	6.49
	Precision	0.90 ± 0.06	0.82 ± 0.12	0.84 ± 0.08	0.70 ± 0.34	0.81	
	Recall	0.87 ± 0.07	0.85 ± 0.10	0.83 ± 0.11	0.76 ± 0.17	0.83	
	F1-score	0.88 ± 0.06	0.83 ± 0.10	0.83 ± 0.08	0.70 ± 0.30	0.81	
3D ResNet18 (33.16 M)	AP	0.95 ± 0.03	0.93 ± 0.05	0.87 ± 0.05	0.73 ± 0.33	**0.87**	237.90
	Precision	0.89 ± 0.08	0.84 ± 0.10	0.85 ± 0.07	0.73 ± 0.28	0.83	
	Recall	0.88 ± 0.04	0.92 ± 0.06	0.89 ± 0.06	0.67 ± 0.30	0.84	
	F1-score	0.88 ± 0.04	0.87 ± 0.06	0.87 ± 0.06	0.69 ± 0.28	0.83	

Overall, the proposed time series pipeline based on the hybrid CNN–transformer provides a near-optimal balance between accuracy and computational efficiency, achieving 0.85 mAP (Table 8). It closely approaches the highest accuracy of the 3D ResNet-18 baseline (0.87 mAP) while delivering substantially faster inference (15.29 ms per frame), which supports real-time defect detection in pipeline welding. Moreover, the CNN–transformer attains the highest macro-averaged precision (0.84), indicating fewer false-positive predictions. By comparison, the lightweight MobileNetV4 baseline achieves the highest speed but yields the lowest overall accuracy (0.83 mAP). Consequently, the CNN–transformer is the most suitable architecture for the proposed pipeline, combining strong predictive performance with the responsiveness required for real-time monitoring.

4. Discussion

4.1. Interpretation of Key Findings

The findings of this study confirm the feasibility of detecting four defect types—incomplete penetration (h_i), shrinkage groove (h_g), sagging (h_s), and linear misalignment (h_m)—with the mAP of 0.85. This level of performance is achieved through the extraction of

https://doi.org/10.3390/s26061825

spatiotemporal features from thermogram sequences. While several defect types occur in fewer than 5% of cases, the use of macro-averaged metrics for model selection ensured that these rare classes contributed equally to the performance evaluation, preventing the model from simply ignoring them. The proposed pipeline integrates a thermogram processing algorithm (Figures 3 and 4a) for generating feature time series, followed by the classification of fixed-length segments using a CNN–transformer architecture (Figures 4d and 5e). Importantly, the low per-frame inference time of 15.3 ms on a CPU highlights the potential for real-time deployment in industrial environments, including implementation on edge devices such as Rockchip and NVIDIA Jetson. Such real-time feedback could support closed-loop control systems by enabling dynamic adjustment of laser parameters to reduce defect occurrence.

It should be acknowledged that the reported performance gap between the feature-based CNN–transformer and the end-to-end image-based baselines may be influenced by the depth of hyperparameter optimization. While more exhaustive tuning of architectures like MobileNetV4 could potentially close this gap, the primary focus of this study was to demonstrate that a compact set of physically informative features provides a robust and computationally efficient signal for defect detection. Unlike "black box" end-to-end models, our approach prioritizes physical interpretability and achieves high accuracy with significantly lower computational overhead for industrial real-time monitoring.

Additional insight comes from the per-class AP values in Table 8, reported as the mean and standard deviation across folds. The standard deviation serves as an indicator of model stability across different data splits. Notably, the incomplete penetration (h_i) defect exhibits a markedly higher standard deviation than the other classes. This variability is consistent across all tested architectures, suggesting it stems from characteristics of the dataset rather than a limitation of a specific model. In the current annotation strategy, a sample was labeled as defective when the h_i depth exceeded zero according to ISO 6520-1. Experimental observations indicate that a higher physical defect magnitude may be required before it produces a consistently discernible thermal signature. Consequently, some validation folds contained sequences with weak or ambiguous signals, lowering the AP for those specific folds.

4.2. Comparative Analysis

The findings of this study confirm the feasibility of using spatiotemporal features as a reliable indicator for weld quality monitoring in pipeline steel FPLW. Our results align with the broader research trend toward automated defect detection, yet provide distinct advantages in terms of feature interpretability and computational efficiency. A distinction of our approach lies in the input data representation. In contrast to the methodology proposed by Zhang et al. [38], which utilizes a convolutional neural network for real-time penetration state diagnosis based on high-dimensional raw pixel sequences, our pipeline leverages a compact set of four physically interpretable engineered features. While Zhang et al. achieved a high accuracy of 94.6% for specific penetration states, our framework addresses the more complex task of multi-label classification, targeting four distinct defect types (h_i, h_g, h_s, and h_m) simultaneously and achieving a competitive macro-averaged mAP of 0.85. Compared to the feature extraction method by Gao et al. [37], which relies on HU moments of the weld pool and keyhole morphology to achieve 91.29% accuracy for six defect types, our hybrid CNN–transformer architecture explicitly models the temporal evolution of spatter ejections. This allows for a more robust interpretation of process instabilities compared to static geometric descriptors. Our achieved inference speed of 15.3 ms compares favorably with recent Vision Transformer models, such as those presented by Din et al. [42], which report 97% accuracy for battery pole defects but require

https://doi.org/10.3390/s26061825

substantially larger datasets and more complex attention mechanisms to operate on raw images.

4.3. Limitations and Challenges

As demonstrated, statistically significant relationships between extracted features and weld defects were identified only for four defect types. The current set of extractable features is constrained by the spatial, temporal, and thermal properties of the spatters, which inherently limits the number of detectable welding anomalies.

A further challenge concerns the resolution of the thermal imaging sensor, which can impede the accurate detection of small spatters or introduce errors in estimating their size and temperature. Data availability remains a significant constraint—the dataset comprises 34 thermograms, yielding 800 samples with a window size of 64 and a 50% overlap. Furthermore, the industrial focus of this study was primarily on incomplete penetration (h_i), other defect types were included and reported to demonstrate the framework's potential for multi-label classification in broader welding despite the pronounced class imbalance. Moreover, the current study did not assess the impact of different materials, camera angles, or acquisition conditions, all of which may influence feature extraction and classifier performance.

4.4. Future Research Directions

Building on the findings of this study, future work should focus on the development of a substantially larger and more diverse dataset, incorporating diverse sets of materials, welding conditions, and thermogram acquisition configurations. Such a dataset would enable a more rigorous assessment of the generalizability and robustness of the proposed methodology. To enhance measurement precision, future studies should also explore the use of thermal imaging systems with higher spatial resolution and increased frame rates.

Future work will focus on enriching standard deep learning architectures with physics-informed inductive biases, including specialized attention mechanisms and tailored loss functions that reflect the underlying welding dynamics. Beyond the evaluated architectures, we plan to investigate hierarchical temporal models and multi-modal transformers to capture the evolution of ejections across multiple spatiotemporal scales. These developments will be complemented by the integration of domain-specific knowledge—such as temperature decay patterns and the transient nature of spatter ejections—into both architectural design and optimization objectives, thereby enhancing detection sensitivity and predictive accuracy. At the system level, further performance gains are expected through advanced classifier refinement and sensor fusion strategies that combine thermal imaging with visible and ultraviolet imaging, spectroscopy, or other auxiliary sensing modalities. Ultimately, such sensor fusion will allow for a more dependable way to detect various defects in real time, making industrial laser welding significantly more reliable.

Further development plans for this approach include its integration into a multi-agent system to enhance automation of engineering workflows and process control, while improving usability through natural-language interaction.

5. Conclusions

This work establishes a sensor-to-AI framework for non-destructive, in-process defect detection in full-penetration laser welding using infrared thermography as the primary monitoring modality. A dedicated thermogram-processing stage isolates spatter activity and converts high-speed IR video into compact, physically interpretable time series descriptors (spatter area, spatter temperature, number of spatters, and welding zone temperature), enabling data-driven quality assessment directly from sensor signals.

https://doi.org/10.3390/s26061825

A multi-label learning formulation was adopted to reflect real manufacturing conditions where multiple defects may co-occur, targeting incomplete penetration, sagging, shrinkage groove, and linear misalignment. Across the evaluated model families, the hybrid CNN–transformer that processes fixed-length feature sequences delivered the strongest overall performance, achieving 0.85 mAP at a sequence length of 64 frames, while maintaining near real-time inference (15.3 ms per frame on the CPU), which supports practical deployment without specialized compute.

This work establishes a robust real-time defect detection framework based on the first systematic evaluation of the relationships between multi-dimensional spatter dynamics and diverse weld defects in pipeline steel. Particularly strong results were achieved for incomplete penetration defects, where the correlation between spatter signatures and defect formation proved to be highly discriminative in real time.

By leveraging non-specialized thermal cameras rather than high-cost proprietary hardware from Precitec, Trumpf, or IPG, the proposed methodology provides a scalable and economically viable foundation for automated in-process quality assurance.

These findings demonstrate that for 09G2S pipeline steel, IR thermography coupled with temporal deep learning on engineered spatiotemporal features can provide actionable online diagnostics under laboratory conditions. While industrial deployment requires further testing across diverse setups, this work provides a preliminary foundation for the potential development of in-process parameter optimization systems in industrial pipeline laser welding.

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

Funding: This research was funded by the Ministry of Economic Development of the Russian Federation (IGK 000000C313925P4C0002) under agreement No. 139-15-2025-010.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and source code supporting the findings of this study can be downloaded from: https://github.com/ILT-ITMO/LaserWeldMonitor (accessed on 9 March 2026).

Conflicts of Interest: The authors state that they have no personal or financial conflicts of interest known to them that could have affected the research reported in this article.

Abbreviations

The following abbreviations are used in this manuscript:

FPLW	Full-Penetration Laser Welding
IR	Infrared
LBW	Laser Beam Welding
GMAW	Gas Metal Arc Welding
SAW	Submerged Arc Welding
WPDC	Weld Penetration Depth Control
ML	Machine Learning
DT	Decision Tree
RF	Random Forest
RL	Logistic Regression

https://doi.org/10.3390/s26061825

GB	Gradient Boosting
SVM	Support Vector Machine
DL	Deep Learning
MLP	Multilayer Perceptron
CNN	Convolutional Neural Networks
LSTM	Long Short-Term Memory
mAP	Mean Average Precision
LoG	Laplacian-of-Gaussian
h_g	Shrinkage Groove
h_s	Sagging
h_m	Linear Misalignment
h_i	Incomplete Penetration
AP	Average Precision

References

1. Klimpel, A. Review and Analysis of Modern Laser Beam Welding Processes. *Materials* **2024**, *17*, 4657. [CrossRef]
2. Kawahito, Y.; Mizutani, M.; Katayama, S. High quality welding of stainless steel with 10 kW high power fibre laser. *Sci. Technol. Weld. Join.* **2009**, *14*, 288–294. [CrossRef]
3. Li, Y.; Geng, S.; Shu, L.; Li, Y.; Jiang, P. Ultra-high-power laser welding of thick-section steel: Current research progress and future perspectives. *Opt. Laser Technol.* **2023**, *167*, 109663. [CrossRef]
4. Merchant, V.E. Laser welding in the pipeline industry. In *The Industrial Laser Handbook: 1992–1993 Edition*; Belforte, D., Levitt, M., Eds.; Springer: New York, NY, USA, 1992; pp. 81–88. [CrossRef]
5. Howse, D.; Scudamore, R.; Woloszyn, A.; Booth, G.; Howard, R. Development of the laser-MAG hybrid welding process for land pipeline construction. In Proceedings of the International Conference on the Application and Evaluation of High-Grade Linepipes in Hostile Environments, TWI and BP Exploration, Yokohama, Japan, 7–8 November 2002.
6. Hu, Y.; Zhang, X.; Liu, Z.; Yan, F.; Su, S.; Wang, C. Investigation on the formation mechanism of collapse defects, microstructural evolution and mechanical properties in full penetration laser welding of thick-section steel. *J. Manuf. Process.* **2024**, *118*, 341–356. [CrossRef]
7. Kaplan, A.; Powell, J. Spatter in laser welding. *J. Laser Appl.* **2011**, *23*, 032005. [CrossRef]
8. Liu, T.; Hu, R.; Chen, X.; Gong, S.; Pang, S. Localized boiling-induced spatters in the high-power laser welding of stainless steel: Three-dimensional visualization and physical understanding. *Appl. Phys. A* **2018**, *124*, 654. [CrossRef]
9. Cheng, H.; Zhou, L.; Li, Q.; Du, D.; Chang, B. Effect of welding parameters on spatter formation in full-penetration laser welding of titanium alloys. *J. Mater. Res. Technol.* **2021**, *15*, 5516–5525. [CrossRef]
10. Zou, J.; Zhu, B.; Zhang, G.; Guo, S.; Xiao, R. Power density effect on the laser beam-induced eruption of spatters in fiber laser keyhole welding. *Opt. Laser Technol.* **2022**, *147*, 107651. [CrossRef]
11. Zhang, Y.; Chen, J.; Zhang, W.; Li, C.; Qiu, C.; Ding, J.; Lu, H.; Zhang, K. Study of spatter net forming mechanism and penetration mode under flexible ring mode laser welding. *J. Mater. Res. Technol.* **2023**, *24*, 2213–2225. [CrossRef]
12. Wang, Z.; Shi, L.; Liao, W.; Liu, Y.; Zhang, S.; Gao, M. Numerical and experimental study on spatter in oscillating laser-arc hybrid welding of aluminum alloy. *J. Mater. Process. Technol.* **2024**, *332*, 118560. [CrossRef]
13. Sharma, S.; Maheshwari, S. A review on welding of high strength oil and gas pipeline steels. *J. Nat. Gas Sci. Eng.* **2017**, *38*, 203–2017. [CrossRef]
14. Sokolov, M.; Salminen, A. Improving Laser Beam Welding Efficiency. *Engineering* **2014**, *6*, 559–571. [CrossRef]
15. Zhang, M.; Chen, G.; Zhou, Y.; Liao, S. Optimization of deep penetration laser welding of thick stainless steel with a 10 kW fiber laser. *Mater. Des.* **2014**, *53*, 568–576. [CrossRef]
16. Sibillano, T.; Rizzi, D.; Mezzapesa, F.P.; Lugarà, P.M.; Konuk, A.R.; Aarts, R.; Veld, B.H.i.'t.; Ancona, A. Closed Loop Control of Penetration Depth during CO_2 Laser Lap Welding Processes. *Sensors* **2012**, *12*, 11077–11090. [CrossRef]
17. Harooni, M.; Carlson, B.; Kovacevic, R. Detection of defects in laser welding of AZ31B magnesium alloy in zero-gap lap joint configuration by a real-time spectroscopic analysis. *Opt. Lasers Eng.* **2014**, *56*, 54–66. [CrossRef]
18. You, D.; Gao, X.; Katayama, S. Detection of imperfection formation in disk laser welding using multiple on-line measurements. *J. Mater. Process. Technol.* **2015**, *219*, 209–220. [CrossRef]
19. Eriksson, I.; Norman, P.; Kaplan, A. Basic study of photodiode signals from laser welding emissions. In Proceedings of the Nordic Conference on Laser Processing of Materials, Turku, Finland, 24–26 August 2009.
20. Hoffman, J.; Szymanski, Z.; Jakubowski, J.; Kolasa, A. Analysis of acoustic and optical signals used as a basis for controlling laser-welding processes. *Weld. Int.* **2002**, *16*, 18–25. [CrossRef]

https://doi.org/10.3390/s26061825

21. Yusof, M.F.M.; Ishak, M.; Ghazali, M.F. Feasibility of using acoustic method in monitoring the penetration status during the Pulse Mode Laser Welding process. *IOP Conf. Ser. Mater. Sci. Eng.* **2017**, *238*, 012006. [CrossRef]
22. Fischer, B.; Rohringer, W.; Panzer, N.; Hecker, S. Acoustic Process Control for Laser Material Processing: Optical microphone as a novel "ear" for industrial manufacturing. *Laser Tech. J.* **2017**, *14*, 21–25. [CrossRef]
23. Authier, N.; Touzet, E.; Lücking, F.; Sommerhuber, R.; Bruyere, V.; Namy, P. Coupled membrane free optical microphone and optical coherence tomography keyhole measurements to setup welding laser parameters. In *High-Power Laser Materials Processing: Applications, Diagnostics, and Systems IX*; SPIE: Nuremberg, Germany, 2020; p. 8. [CrossRef]
24. Bîrdeanu, A.V.; Ciucă, C.; Puicea, A. Pulsed LASER-(micro) TIG hybrid welding: Process characteristics. *J. Mater. Process. Technol.* **2012**, *212*, 890–902. [CrossRef]
25. Sinha, S.; Glorieux, E.; Franciosa, P. 3D convolutional neural networks to estimate assembly process parameters using 3D point-clouds. In *Multimodal Sensing: Technologies and Applications*; SPIE: Nuremberg, Germany, 2019; p. 10. [CrossRef]
26. Stritt, P.; Boley, M.; Heider, A.; Fetzer, F.; Jarwitz, M.; Weller, D.; Weber, R.; Berger, P.; Graf, T. Comprehensive process monitoring for laser welding process optimization. In *High-Power Laser Materials Processing: Lasers, Beam Delivery, Diagnostics, and Applications V*; SPIE: Nuremberg, Germany, 2016; p. 97410Q. [CrossRef]
27. Bautze, T.; Moser, R.; Strebel, M.; Kogel-Hollacher, M. Use of inline coherent imaging for laser welding processes: Process control and beyond. In Proceedings of the Lasers in Manufacturing Conference, Munich, Germany, 22–25 June 2015; pp. 1–9.
28. Schweier, M.; Haubold, M.; Zaeh, M. Analysis of spatters in laser welding with beam oscillation: A machine vision approach. *CIRP J. Manuf. Sci. Technol.* **2016**, *14*, 35–42. [CrossRef]
29. Huang, Y.; Hua, X.; Li, F.; Shen, C.; Mou, G.; Tang, B. Spatter feature analysis in laser welding based on motion tracking method. *J. Manuf. Process.* **2020**, *55*, 220–229. [CrossRef]
30. Hartung, J.; Jahn, A.; Stambke, M.; Wehner, O.; Thieringer, R.; Heizmann, M. Camera-based spatter detection in laser welding with a deep learning approach. In *Forum Bildverarbeitung*; KIT Scientific Publishing: Karlsruhe, Germany, 2021. [CrossRef]
31. Lu, R.; Lou, M.; Xia, Y.J.; Huang, S.; Li, Z.; Lv, T.; Wu, Y.; Li, Y. Machine learning-based in-process monitoring for laser deep penetration welding: A survey. *Eng. Appl. Artif. Intell.* **2024**, *137*, 109059. [CrossRef]
32. Xiao, X.; Liu, X.; Cheng, M.; Song, L. Towards monitoring laser welding process via a coaxial pyrometer. *J. Mater. Process. Technol.* **2019**, *277*, 116409. [CrossRef]
33. Chianese, G.; Franciosa, P.; Sun, T.; Ceglarek, D.; Patalano, S. Using Photodiodes and Supervised Machine Learning for Automatic Classification of Weld Defects in Laser Welding of Thin Foils Copper-to-Steel Battery Tabs. *J. Laser Appl.* **2022**, *34*, 042040. [CrossRef]
34. Zhao, P.; Lu, R.; Wu, J.; Zhang, Y. Monitoring method of weld penetration in laser keyhole welding irradiated by laser auxiliary illuminant. *J. Laser Appl.* **2020**, *32*, 022037. [CrossRef]
35. Paulson, N.; Gould, B.; Wolff, S.; Stan, M.; Greco, A. Correlations between thermal history and keyhole porosity in laser powder bed fusion. *Addit. Manuf.* **2020**, *34*, 101213. [CrossRef]
36. Yu, R.; Cao, Y.; Chen, H.; Ye, Q.; Zhang, Y. Deep learning based real-time and in-situ monitoring of weld penetration: Where we are and what are needed revolutionary solutions? *J. Manuf. Process.* **2023**, *93*, 15–46. [CrossRef]
37. Gao, Y.; Zhong, P.; Tang, X.; Hu, H.; Xu, P. Feature Extraction of Laser Welding Pool Image and Application in Welding Quality Identification. *IEEE Access* **2021**, *9*, 120193–120202. [CrossRef]
38. Zhang, Z.; Li, B.; Zhang, W.; Lu, R.; Wada, S.; Zhang, Y. Real-time penetration state monitoring using convolutional neural network for laser welding of tailor rolled blanks. *J. Manuf. Syst.* **2020**, *54*, 348–360. [CrossRef]
39. Hara, K.; Kataoka, H.; Satoh, Y. Learning Spatio-Temporal Features with 3D Residual Networks for Action Recognition. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017.
40. Simonyan, K.; Zisserman, A. Two-Stream Convolutional Networks for Action Recognition in Videos. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 568–576.
41. Ozkat, E.C. Photodiode Signal Patterns: Unsupervised Learning for Laser Weld Defect Analysis. *Processes* **2025**, *13*, 121. [CrossRef]
42. Din, N.; Zhang, l.; Nawaz, M.S.; Yang, Y. Multi-model feature aggregation for classification of laser welding images with vision transformer. *J. King Saud Univ. Comput. Inf. Sci.* **2024**, *36*, 102049. [CrossRef]
43. Lindemann, B.; Müller, T.; Vietz, H.; Jazdi, N.; Weyrich, M. A survey on long short-term memory networks for time series prediction. *Procedia CIRP* **2021**, *99*, 650–655. [CrossRef]
44. Gui, X.; Du, J.; Song, L.; Yan, Z.; Guo, L. A Transformer-CNN Based Transferable Model for State-of-Health Prediction of Lithium-ion Batteries. In *2023 Global Reliability and Prognostics and Health Management Conference (PHM-Hangzhou)*; IEEE: New York, NY, USA, 2023; pp. 1–8.
45. Maksimova, E. Study of the structure of the 09G2S grade steel welded joints performed by welding at subzero temperatures. *Procedia Struct. Integr.* **2019**, *20*, 174–179. [CrossRef]

https://doi.org/10.3390/s26061825

46. *BS EN ISO 6520-1:2007*; Welding and Allied Processes—Classification of Geometric Imperfections in Metallic Materials—Part 1: Fusion Welding. BSI: London, UK, 2007.
47. Kim, S. Min-local-LoG filter for detecting small targets in cluttered background. *Electron. Lett.* **2011**, *47*, 105–106. [CrossRef]
48. Qin, D.; Leichner, C.; Delakis, M.; Fornoni, M.; Luo, S.; Yang, F.; Wang, W.; Banbury, C.; Ye, C.; Akin, B.; et al. *MobileNetV4: Universal Models for the Mobile Ecosystem*; Springer Nature: Cham, Switzerland, 2024; pp. 78–96. [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.

https://doi.org/10.3390/s26061825

sensors

MDPI

Article

Interval Prediction of Remaining Useful Life Based on Uncertainty Quantification with Bayesian Convolutional Neural Networks Featuring Dual-Output Units

Zhendong Qu [1,2], Jialong He [1,2], Yan Liu [1,2,*], Song Mao [1,2] and Xiaowu Han [1,2]

[1] Key Laboratory of CNC Equipment Reliability, Ministry of Education, Jilin University, Changchun 130022, China; quzd22@jlu.edu.cn (Z.Q.); hejl@jlu.edu.cn (J.H.); maosong23@jlu.edu.cn (S.M.); 18634874024@163.com (X.H.)
[2] School of Mechanical and Aerospace Engineering, Jilin University, Changchun 130022, China
* Correspondence: liuy_jlu@163.com; Tel.: +86-15754302451

Abstract

RUL prediction methods do not fully account for the uncertainties caused by data scarcity and inherent noise, and they also suffer from low reliability of RUL point estimates. To tackle these challenges, this paper proposes a Bayesian convolutional neural network with dual-output units for RUL interval predictions. The network employs the negative log-likelihood as the loss function. Thanks to its dual-output structure, it not only provides point estimates, but also quantifies the aleatoric uncertainty inherent in the data. During the training process, the CNN is reformulated using Bayesian principles, and the Bayes-by-backprop method is applied to train the network. This transformation converts model parameters from fixed values into random variables. As a result, epistemic uncertainty caused by model inaccuracies and limited data can be quantified. Experimental validation on the IEEE PHM Challenge 2012 dataset demonstrated that the proposed method achieved a higher prediction accuracy than state-of-the-art uncertainty-aware prediction approaches, demonstrating a better applicability in engineering practice.

Keywords: remaining useful life prediction; uncertainty quantification; Bayesian deep learning; convolutional neural network; interval prediction

1. Introduction

Prognostics and health management (PHM) is a decision-support system designed to ensure equipment safety and reduce maintenance costs by leveraging physical knowledge, operational information, and data from industrial equipment operation and maintenance [1–3]. The prediction of the remaining useful life (RUL) is a core component of PHM technology, serving as a prerequisite and foundation for system maintenance, and it represents a key research focus in the field of reliability engineering.

With the rapid development of sensor technology, increasingly integrated and miniaturized sensing devices are being widely deployed in mechanical systems to monitor equipment health conditions. In such a data-rich environment, deep learning techniques have demonstrated strong capabilities for RUL predictions. Various models, including convolutional neural networks (CNNs) [4], long short-term memory networks (LSTMs) [5], gated recurrent units (GRUs) [6], and graph convolutional networks (GCNs) [7], as well as their variants, have achieved a promising performance in bearing RUL prediction tasks.

However, most existing deep learning models focus on point estimation and fail to quantify the predictive uncertainty, which limits their reliability in practice [8]. In practice, data acquisition is often affected by noise and measurement errors, which are referred to as aleatoric uncertainty [9]. Homoscedastic uncertainty assumes a constant variance, whereas heteroscedastic uncertainty allows it to vary across inputs. Meanwhile, the uncertainty in model fitting caused by factors such as insufficient data or inappropriate model selection is referred to as epistemic uncertainty [10]. To quantify the uncertainties in practical engineering applications, Bayesian neural networks are widely employed in the field of remaining useful life predictions [11]. In Bayesian neural networks, all weights are represented by probability distributions of random variables, unlike traditional deep learning networks, where the weights are assigned fixed values [12]. During the training process, Bayesian neural networks quantify uncertainty through probability distributions, yielding results that are more reliable due to the provision of confidence intervals. This characteristic also facilitates a subsequent analysis and decision-making.

In addition to Bayesian approaches, recent studies have explored uncertainty-aware RUL predictions from different perspectives. Evidential deep learning [13] has been introduced to model predictive uncertainty through evidence theory without relying on sampling-based inference, while deep equilibrium models [14] have been proposed to enhance model representation and stability in RUL estimation. However, these methods do not explicitly provide a unified probabilistic framework for decomposing the predictive uncertainty into aleatoric and epistemic components.

Currently, the variational inference (VI) method is commonly used to train Bayesian neural network (BNN) models. Variational inference [15] employs an approximate distribution, parameterized by θ, to approximate the posterior distribution of network parameters, and iteratively optimizes this approximate distribution by solving an optimization problem. It is widely applicable and converges quickly. Two common forms of VI are the Monte Carlo (MC) dropout method and the Bayes-by-backprop method. Weiwen Peng et al. [16] integrated the MC dropout method with a bidirectional LSTM neural network and a multi-scale convolutional neural network, using dropout to achieve a Bayesian approximation for uncertainty quantification. Gaoyang Li et al. [10] proposed a novel Bayesian framework that combines the MC dropout method with a dual-output neural network. In this framework, the dual-output structure captures the aleatoric uncertainty, while dropout is used to implement Bayesian approximation in the neural network to capture epistemic uncertainty, thereby completing uncertainty quantification. Among the aforementioned methods, MC dropout is widely adopted by researchers for uncertainty quantification due to its relative ease of implementation. However, the MC dropout method can only capture epistemic uncertainty related to weights by randomly activating or deactivating neurons, while neglecting the aleatoric uncertainty caused by inherent noise in the data [17]. Additionally, the MC dropout method has been criticized for its inability to converge as the amount of data increases, a limitation that persists even in simple linear networks [18]. Kumar Shridhar et al. [19] utilized the Bayes-by-backprop method for uncertainty quantification in a BNN-CNN, employing a loss function composed of the sum of the negative log-likelihood and the Kullback–Leibler divergence. Lei Wang et al. [1] proposed a Bayesian large-kernel attention mechanism network for bearing remaining useful life predictions; it employs the Bayes-by-backprop method for network uncertainty quantification. Bayes-by-backprop implements the iterative optimization of the approximate distribution in the variational inference using the backpropagation algorithm [12]. Similar to the MC dropout method, this approach also overlooks the aleatoric uncertainty arising from inherent noise in the data. Yaguo Lei et al. [20] introduced a dual-output Res-ConvLSTM network, trained using a hybrid loss function that includes the mean squared error, log probability density, and

https://doi.org/10.3390/s26092592

confidence interval width definitions. While this method achieves quantification of the aleatoric uncertainty, it lacks a process for quantifying the epistemic uncertainty resulting from inaccuracies in the model's fitted distribution.

Based on the above analysis, the existing methods still face two major challenges in RUL predictions:

(1) The aleatoric uncertainty is often oversimplified or assumed to be constant; thus, the methods fail to capture its heteroscedastic nature [21].
(2) The epistemic uncertainty is frequently ignored in dual-output structures, resulting in incomplete uncertainty modeling [22].

Although BNNs have been widely used for uncertainty quantification, most existing approaches either focus primarily on epistemic uncertainty or model aleatoric uncertainty separately without a unified framework. As a result, they cannot simultaneously capture both types of uncertainty in a consistent probabilistic formulation.

To address the aforementioned limitations, this paper employs a convolutional neural network (CNN) as the backbone architecture, reformulates it within a Bayesian framework, and incorporates a dual-output structure. Bayesian inference is performed via the Bayes-by-backprop method for bearing RUL predictions and uncertainty quantification. The main contributions of this work are summarized as follows:

1. A unified probabilistic framework is proposed to jointly model aleatoric and epistemic uncertainties for RUL predictions. The dual-output structure enables the modeling of heteroscedastic aleatoric uncertainty, while the Bayesian formulation captures epistemic uncertainty through parameter distributions.
2. A Bayesian convolutional neural network (BCNN) with a dual-output structure is developed, in which model parameters are treated as random variables and optimized via Bayes-by-backprop, enabling consistent uncertainty quantification and an improved prediction reliability.

Compared with existing Bayesian neural network-based methods, the proposed approach provides a unified and principled framework for jointly modeling aleatoric and epistemic uncertainties, enabling more comprehensive and reliable uncertainty quantification for RUL predictions.

The structure of this paper is organized as follows: Section 2 introduces the theoretical foundations of Bayesian deep learning (BDL) networks and variational inference. Section 3 presents the proposed dual-output Bayesian convolutional neural network (BCNN) framework and elaborates on uncertainty quantification for remaining useful life predictions. Section 4 validates the effectiveness of the proposed method on a public bearing dataset and compares it with other uncertainty quantification approaches. Section 5 concludes the paper.

2. Methods and Theories

2.1. Bayesian Neural Networks

In a classical regression model, the remaining useful life prediction is often formulated as $y = f^w(x)$, where x and y represent the input and output of the system, respectively, and w denotes the parameters of the regression model. The essence of Bayesian deep learning (BDL) lies in preserving the topological structure of the deep learning network while treating its parameters as random variables, thereby enabling uncertainty quantification through probability distributions. This approach retains the superior feature extraction and analytical capabilities of deep learning models, while leveraging Bayesian inference to update and optimize the probability distributions of the model parameters.

https://doi.org/10.3390/s26092592

For a set of system run-to-failure training trajectories, the dataset is denoted as $D = \{(X_i, Y_i)\}_{i=0}^{N}$, where N is the total number of training trajectories. For the i-th trajectory, $X_i = \{x_{i,j}\}_{j=0}^{N_i}$ and $Y_i = \{y_{i,j}\}_{j=0}^{N_i}$, where $x_{i,j}$ represents the input sample at the j-th time step of the i-th trajectory, $y_{i,j}$ is the corresponding RUL label, and N_i denotes the number of samples in the i-th trajectory. The Bayesian deep learning (BDL) model is formulated as $y = f^w(x)$, with a prior distribution $p(w)$. The likelihood function after model regression is defined as $p(y_i|\, f^w(x_i))$. In bearing remaining useful life prediction tasks, this likelihood is often defined as a Gaussian distribution, i.e., $y_i \sim N(f^w(x_i), I)$, where I represents the standard deviation of the likelihood distribution. For newly observed data x^*, the predictive probability of the model is expressed as:

$$p(y^*|\, x^*, D) = \int p(y^*|\, f^w(x^*))p(w|\, D)dw \tag{1}$$

where $p(w|D)$ is referred to as the posterior distribution. According to Bayesian theory, the posterior distribution is computed as:

$$p(w|\, D) = \frac{p(w)p(D|\, w)}{p(D)} = \frac{p(w)p(D|\, w)}{\int p(D|\, w)p(w)dw} \tag{2}$$

It can be observed that the model predictive probability involves an integration problem, and the posterior distribution is generally intractable to compute directly, especially for complex network structures and high-dimensional data. This intractability poses a significant challenge for probability integration. Variational inference (VI) is a commonly used optimization method in machine learning for approximating intractable distributions [23].

2.2. Variational Inference Theory

The implementation of variational inference (VI) generally consists of two steps: First, a family of probability distributions $q_\phi(w)$ parameterized by ϕ is selected as the variational distribution family; second, the optimal variational distribution is found by minimizing the Kullback–Leibler (KL) divergence between $q_\phi(w)$ and the true posterior $p(w|\, D)$ [14]. The KL divergence is defined as follows:

$$\begin{aligned}
&KL\big[q_\phi(w)||\, p(w|\, D)\big] = \int q_\phi(w) \log \frac{q_\phi(w)}{p(w|\, D)} dw \\
&= \int q_\phi(w) \log \frac{q_\phi(w)p(D)}{p(D|\, w)p(w)} dw \\
&= \int q_\phi(w) \log \frac{q_\phi(w)}{p(w)} dw - \int q_\phi(w) \log p(D|\, w) dw + \int q_\phi(w) \log p(D) dw \\
&= \int q_\phi(w) \log \frac{q_\phi(w)}{p(w)} dw - \int q_\phi(w) \log p(D|\, w) dw + \log p(D) \\
&= KL\big[q_\phi(w)||\, p(w)\big] - E_{q_\phi(w)}[\log p(D|\, w)] + \log p(D)
\end{aligned} \tag{3}$$

In Formula (3), the posterior distribution is expanded for analysis. As a result, the KL divergence can be decomposed into three components. The first part represents the KL divergence values of the variational distribution family $q_\phi(w)$ and the prior $p(w)$; the second part is to convert the original integral term into the corresponding expected term, in order to facilitate further simplification and implementation using Monte Carlo (MC) sampling in the future; and the third part represents a constant that is independent of ϕ and

https://doi.org/10.3390/s26092592

can be ignored in minimization (3). If ϕ^* is used to represent the optimization problem, then after the above transformation, the expression of the optimization problem is as follows:

$$\begin{aligned} \phi^* &= \arg\min_{\phi} KL\left[q_\phi(w)||\, p(w|\, D)\right] \\ &= \arg\min_{\phi}\left\{KL\left[q_\phi(w)||\, p(w)\right] - E_{q_\phi(w)}[\log p(D|\, w)]\right\} \end{aligned} \tag{4}$$

The corresponding objective function takes the form of:

$$F(D,\phi) = KL\left[q_\phi(w)||\, p(w)\right] - E_{q_\phi(w)}[\log p(D|\, w)] \tag{5}$$

Most of the existing references are written in ELBO form, and its expression is as follows:

$$\begin{aligned} ELBO &= E_{q_\phi(w)}[\log p(D|\, w)] - KL\left[q_\phi(w)||\, p(w)\right] \\ &= E_{q_\phi(w)}[\log p(D|\, w)] - E_{q_\phi(w)}\left[\log q_\phi(w)\right] + E_{q_\phi(w)}[\log p(w)] \end{aligned} \tag{6}$$

In Formula (6), the minimization optimization problem is transformed into the maximization ELBO problem. The $KL\left[q_\phi(w)||\, p(w)\right]$ term is called the complexity cost, which represents the physical distance between the variational distribution family $q_\phi(w)$ and the model prior $p(w)$, and it is used to constrain $q_\phi(w)$ to approximate the prior $p(w)$. The $E_{q_\phi(w)}[\log p(D|\, w)]$ term is called the likelihood cost, which constrains the model to fit the training data. In the previous definition, let the prior $p(w)$ follow a Gaussian distribution with a mean of 0 and a variance of σ_p^2; the variational distribution $q_\phi(w)$ is also assumed to follow the Gaussian distribution $w \sim N(\mu, \sigma^2)$, where μ and σ^2 represent the mean and variance of the Gaussian distribution, respectively. In the second equation of Formula (6), the $KL\left[q_\phi(w)||\, p(w)\right]$ term is written as the logarithmic expected form of the variational distribution $q_\phi(w)$ and the prior $p(w)$, respectively. The purpose of this is to estimate the expected value through MC sampling.

2.3. Parameter Resampling

The MC sampling method estimates the expected value through multiple sampling processes, and the corresponding expression can be written as:

$$Loss = -\left\{\sum_{s=1}^{N_{train}^s}\left[\log p\left(D\middle|w^{(s)}\right) - \log q_\phi\left(w^{(s)}\right) + \log p\left(w^{(s)}\right)\right]\right\} \tag{7}$$

In the process of sampling w, we need to sample w from $N(\mu, \sigma^2)$. If we directly sample w, because the sampling action is discrete, the result is not differentiable. Therefore, in reality, there is no gradient information when used for backpropagation, and it is impossible to update the parameter gradient. To solve such problems, reparameterization techniques are often used to transform the sampling process into obtaining ε from the standard normal distribution $N(0,1)$, where $w = \mu + \varepsilon \cdot \sigma$, represents an element multiplication calculation. This can solve the problem of gradient non-transitivity. At the same time, in order to ensure that the value of the variational distribution parameter σ is greater than 0, the parameter σ is also resampled:

$$\sigma = \log(1 + e^{\rho}) \tag{8}$$

In this way, the distribution parameters of $q_\phi(w)$ become μ and ρ, which are different from the originally defined distribution parameters μ and σ. The uncertainty introduced by this sampling process can be regarded as global uncertainty. When performing Bayesian inference, it is necessary to sample all parameters globally. However, Diederik P. Kingma et al. [24] found that, if all parameters are independent Gaussian distributions, the results

https://doi.org/10.3390/s26092592

obtained through forward propagation will also be independent Gaussian distributions. This conclusion proves that we can directly calculate the forward-propagated values of the $q_\phi(w)$ distribution parameters for sampling, and then backpropagate them to the original distribution parameters for updating. The forward propagation values are calculated as follows:

$$F_\mu = XW_\mu + B_\mu \tag{9}$$

$$F_\sigma = XW_\sigma + B_\sigma \tag{10}$$

$$F = F_\mu + \varepsilon \cdot F_\sigma \tag{11}$$

Among them, X represents the input data matrix of the model, F_μ and F_σ correspond to the values of the $q_\phi(w)$ distribution parameters after the forward-propagation calculation, F represents the final return value of forward propagation, W_μ and B_μ represent the matrices corresponding to the mean parameter μ of the model weight and bias, and W_σ and B_σ represent the matrices corresponding to the standard deviation parameter $\log(1 + e^\rho)$ of the model weight and bias after resampling. This method is called the local reparameterization technique (LRT) [24] because it does not require sampling of global parameters, and is often used in Bayes-by-backprop methods to reduce the gradient variance during training and prevent overfitting.

The objective function in Formula (4) and the approximation in Formula (5) are performed by the model on the entire dataset. However, in the actual implementation process, the small-batch gradient descent method is often used to divide the dataset into batches, so correspondingly, the complexity cost $KL\left[q_\phi(w)||\ p(w)\right]$ needs to be scaled. Assuming that the entire dataset is divided into M batches, the simplest scaling method is to average the complexity cost of each batch, and the corresponding objective function can be written as:

$$F(D, \phi) = \frac{1}{M} KL\left[q_\phi(w)||\ p(w)\right] - E_{q_\phi(w)}[\log p(D|\ w)] \tag{12}$$

On this basis, Blundell et al. [12] proposed a new scaling method by modifying the proportion of the complexity cost in each batch, theoretically achieving the idea of focusing on fitting priors in the early stage of training and fitting data in the later stage. The corresponding objective function is written as follows:

$$F(D, \phi) = \pi_i KL\left[q_\phi(w)||\ p(w)\right] - E_{q_\phi(w)}[\log p(D|\ w)] \tag{13}$$

with $\pi_i \in [0, 1]^{\mathrm{M}}$ in Formula (13), and $\sum_{i=1}^{M} \pi_i = 1$, $\pi_i = \frac{2^{M-i}}{2^M - 1}$.

This paper applies the local reparameterization technique mentioned above to the Bayes-by-backprop method for Bayesian inference, and scales the complexity cost calculation involved in it, as shown in (13). The sampling Adams optimizer minimizes the objective function to optimize the network parameters.

3. BCNN Framework and RUL Uncertainty Quantification

The Bayesian framework is applied to convolutional neural networks (CNNs) to achieve RUL uncertainty quantization, and the method proposed in this paper is shown in Figure 1. There were three main modifications: firstly, the model parameters in the convolutional and fully connected layers of the CNN were converted into random variables, and probability distributions were used to represent the epistemic uncertainty of the model. Secondly, a dual-output unit was set up behind the fully connected layer of the CNN to

https://doi.org/10.3390/s26092592

predict the point estimate of the RUL and the inherent variance in the data themselves. Finally, the negative logarithmic probability density function was used as part of the loss function to optimize the dual-output unit results, and the variance in the data themselves was obtained as aleatoric uncertainty. At the same time, the sum of the KL divergence was used as another part of the loss function to optimize the variational distribution parameters, approximately replacing the posterior distribution of parameters in order to obtain the predicted probability to quantify the epistemic uncertainty.

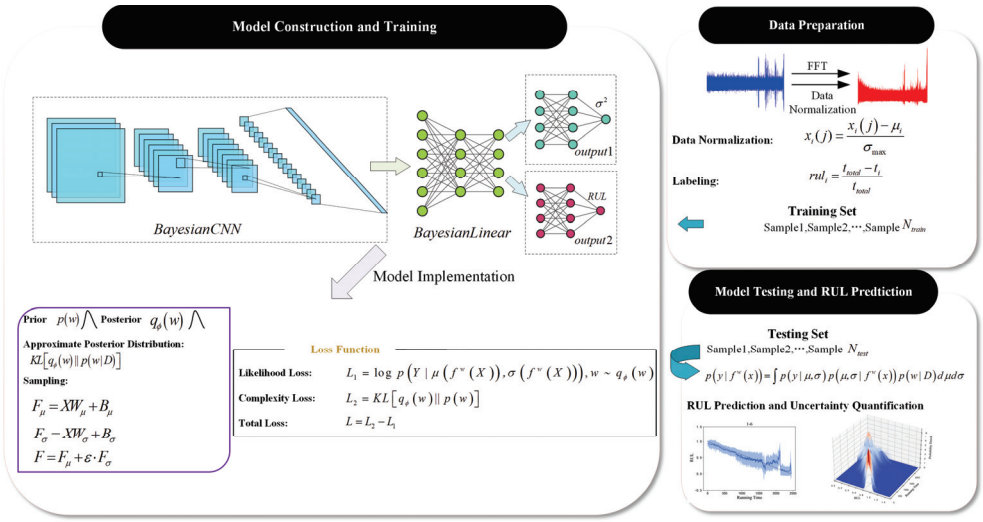


Figure 1. Method framework.

In the proposed BCNN model, the probability distributions in the convolutional and fully connected layers represent epistemic uncertainty, while the dual-output structure models heteroscedastic aleatoric uncertainty. When considering the life distribution parameterized by μ and σ, the likelihood function of the BCNN model in Formula (1) can be written as:

$$p(y|\, f^w(x)) = \int p(y|\, \mu,\sigma)p(\mu,\sigma|\, f^w(x))p(w|\, D)d\,\mu d\sigma \tag{14}$$

Among them, w represents the random variable of the model, corresponding to the posterior distribution of Formula (2), written as:

$$p(w|\, D) = \frac{p(w)\int p(y|\, \mu,\sigma)p(\mu,\sigma|\, f^w(x))d\,\mu d\sigma}{\int p(y|\, \mu,\sigma)p(\mu,\sigma|\, f^w(x))p(w)d\,\mu d\sigma dw} \tag{15}$$

The model prediction probability corresponding to Formula (1) is:

$$p(y^*|\, x^*, D) = \int p(y^*|\, \mu,\sigma)p(\mu,\sigma|\, f^w(x^*))p(w|\, D)d\,\mu d\sigma dw \tag{16}$$

From the previous description, it can be seen that the posterior distribution $p(w|\, D)$ usually cannot calculate an analytical solution, and variational inference theory needs to be used to approximate the posterior distribution $p(w|\, D)$ using the variational distribution $q(w)$. By substituting $q(w)$ into Formula (16), we can obtain:

$$p(y^*|\, x^*) = \int p(y^*|\, \mu,\sigma)p(\mu,\sigma|\, f^w(x^*))q(w)d\,\mu d\sigma dw \tag{17}$$

According to the Bayes-by-backprop method, minimizing the KL divergence of variational and posterior distributions is equivalent to maximizing the lower bound of evidence, and its optimization function can be expressed as

https://doi.org/10.3390/s26092592

$$\begin{aligned}\mathcal{L}_{ELBO} &= \int q(w)\log(\int p(Y \mid \mu,\sigma)p(\mu,\sigma \mid f^w(X))d\,\mu d\sigma)dw \\ &\quad -KL[q(w)||\,p(w)]\end{aligned} \tag{18}$$

In the explanation of Part 2, it is known that the optimization function can obtain its unbiased estimate using the Monte Carlo sampling method, and the equation is:

$$\begin{aligned}\mathcal{L}_{MC} &= E_{q_\phi(w)}[\log(\int p(Y \mid \mu,\sigma)p(\mu,\sigma \mid f^w(X))d\,\mu d\sigma)] \\ &\quad -E_{q_\phi(w)}\left[\log q_\phi(w)\right] + E_{q_\phi(w)}[\log p(w)]\end{aligned} \tag{19}$$

In Formula (19), parameters μ and σ are generated by model weights, so $p(Y \mid \mu,\sigma)$ $p(\mu,\sigma \mid f^w(X))$ can be simplified to $p(Y \mid \mu(f^w(X)),\sigma(f^w(X)))$. Substituting it into Formula (19) yields:

$$\begin{aligned}\mathcal{L}_{MC} &= E_{q_\phi(w)}[\log p(Y \mid \mu(f^w(X)),\sigma(f^w(X)))] \\ &\quad -E_{q_\phi(w)}\left[\log q_\phi(w)\right] + E_{q_\phi(w)}[\log p(w)]\end{aligned} \tag{20}$$

Among them, $w \sim q_\phi(w)$. By maximizing the lower bound of evidence as described above, the optimal variational distribution can be obtained to approximate the posterior distribution of the model. After the model training is completed, it is necessary to quantify the uncertainty of the RUL prediction results. According to the theory of variance accumulation [10], let the predicted variance $Var(y^* \mid x^*, w)$ represent the sum of aleatoric uncertainty and epistemic uncertainty. Reconstructing the total predicted variance yields the following equation:

$$\begin{aligned}Var(y^*|x^*,D) &= E_{p(w|D)}\left[(y^*)^2\right] - \left[E_{p(w|D)}(y^*)\right]^2 \\ &= E_{p(w|D)}\left[E\left((y^*)^2\middle|x^*,w\right)\right] - \left(E_{p(w|D)}[E(y^* \mid x^*,w)]\right)^2 \\ &= E_{p(w|D)}\left[Var(y^*|x^*,w) + [E(y^* \mid x^*,w)]^2\right] - \left(E_{p(w|D)}[E(y^* \mid x^*,w)]\right)^2 \\ &= E_{p(w|D)}\left[(E(y^* \mid x^*,w))^2\right] - \left(E_{p(w|D)}[E(y^* \mid x^*,w)]\right)^2 + E_{p(w|D)}[Var(y^*|x^*,w)] \\ &= Var_{p(w|D)}[E(y^*|x^*,w)] + E_{p(w|D)}[Var(y^*|x^*,w)]\end{aligned} \tag{21}$$

where $Var_{p(w|D)}[\cdot]$ and $E_{p(w|D)}[\cdot]$ denote the expectation and variance with respect to the posterior distribution of model parameters w, respectively.

The predicted total variance $Var(y^*|x^*,D)$ in Formula (21) is decomposed into two parts, where $Var_{p(w|D)}[E(y^*|x^*,w)]$ represents the variance in the expected $E(y^*|x^*,w)$, which can be used to represent the epistemic uncertainty, where the calculation of the expected value is influenced by the model parameters and data size. And $E_{p(w|D)}[Var(y^*|x^*,w)]$ eliminates the variability in the model parameters by calculating the expected value, which reflects the inherent uncertainty of the data itself, that is, the aleatoric uncertainty.

For a test sample x^*, the dual-output BCNN produces two outputs: the predictive mean $\mu(x^*,w)$ and the predictive variance $\sigma^2(x^*,w)$. Under the Gaussian likelihood assumption, the conditional predictive distribution of the RUL is given by $y^*\big|x^*,w \sim N\big(\mu(x^*,w),\sigma^2(x^*,w)\big)$. Accordingly, for a given model parameter sample w, the point prediction of the RUL is $\hat{y}(x^*,w) = E[y^*|x^*,w] = \mu(x^*,w)$, and the corresponding conditional variance is $Var(y^*|x^*,w) = \sigma^2(x^*,w)$, which characterizes the aleatoric uncertainty caused by inherent noise in the data.

Since the model parameters w follow the posterior distribution $p(w|\,D)$, the overall predictive distribution is obtained by marginalizing over w: $p(y^*|\;x^*,D) = \int p(y^*|\;x^*,w)p(w|\;D)dw$. The predictive mean of the RUL is therefore given by $\overline{y}^* = E_{p(w|D)}[\mu(x^*,w)]$, and the predictive variance can be decomposed according to the law of total variance as $Var(y^*|\;x^*,D) = E_{p(w|\;D)}[\sigma^2(x^*,w)] + Var_{p(w|\;D)}[\mu(x^*,w)]$,

https://doi.org/10.3390/s26092592

where the first term represents the aleatoric uncertainty and the second term represents the epistemic uncertainty.

Based on the predictive mean and total predictive variance, the RUL prediction interval at confidence level $(1-\alpha)$ is constructed as

$$\left[\overline{y}^* - z_{1-\alpha/2}\sqrt{Var(y^*|\, x^*, D)}, \overline{y}^* + z_{1-\alpha/2}\sqrt{Var(y^*|\, x^*, D)}\right] \tag{22}$$

where $z_{1-\alpha/2}$ is the standard normal quantile.

After training the BCNN model, its total prediction uncertainty can be obtained through Monte Carlo sampling, and its calculation formula is as follows:

$$\begin{aligned} Mean(y^*) &= \frac{1}{M}\frac{1}{N}\sum_{m=1}^{M}\sum_{n=1}^{N}\hat{y}_n^m \\ Var(y^*) &= \frac{1}{M}\frac{1}{N}\sum_{m=1}^{M}\sum_{n=1}^{N}(\hat{y}_n^m - Mean(y^*))^2 \end{aligned} \tag{23}$$

In Formula (22), calculations are performed by multiple sampling of $w_m \sim q_\phi(w)$ and $\hat{y}_n^m \sim p(y \mid \mu_m, \sigma_m)$, where the ranges of m and n are $[1, M]$ and $[1, N]$, respectively. This form of double sampling is determined by the applied BCNN dual-output unit structure network. In Formula (21), the epistemic uncertainty decomposed from the total prediction uncertainty needs to be solved for the variance of the mean, while aleatoric uncertainty only needs to calculate the mean of the variance. Relatively speaking, for the convenience of calculation, the aleatoric uncertainty in the predicted results can be solved by calculating the average variance of each model weight, which is given by the formula:

$$\begin{aligned} Mean_m(y^*) &= \frac{1}{N}\sum_{m=1}^{M}\sum_{n=1}^{N}\hat{y}_n^m \\ Var_{aleatoric}(y^*) &= \frac{1}{M}\frac{1}{N}\sum_{m=1}^{M}\sum_{n=1}^{N}(\hat{y}_n^m - Mean_m(y^*))^2 \end{aligned} \tag{24}$$

By substituting Formula (23) into Formula (21), the epistemic uncertainty can be calculated as follows:

$$Var_{epistemic}(y^*) = Var(y^*) - Var_{aleatoric}(y^*) \tag{25}$$

The overall training, sampling, and variance decomposition procedure of the proposed dual-output BCNN is summarized in Algorithm 1.

Algorithm 1. Dual-output BCNN variance decomposition.

Input: Training samples: $D_{train} = \{x_i, y_i\}_{i=1}^{N_{train}}$; testing samples: $D_{test} = \{x_i\}_{i=1}^{N_{test}}$;
Iterations: M and N; network parameters: ω;
Output: Variational distribution $q_\phi(\omega)$; predicted mean μ; predicted variance σ;

1. For m from 1 to M, do
2. Draw (μ_m, σ_m) with model parameters $\omega_m \sim q_\phi(\omega)$
3. For n from 1 to N, do
4. Draw sample $\hat{y}_n^m$ from $p(y \mid \mu_m, \sigma_m)$
5. End for
6. End for
7. Compute the $Mean(y^*)$ for $\hat{y}_n^m$ using (22)
8. Decompose the $Var(y^*)$ using (23) and (24)

https://doi.org/10.3390/s26092592

4. Case Description

In order to verify the effectiveness and superiority of the Bayesian convolutional neural network (BCNN) framework proposed in this paper for measuring the uncertainty in RUL predictions, this section uses the IEEE PHM 2012 dataset for experimental verification. The algorithm in this paper was trained on the Windows 10 system and an NVIDIA GeForce RTX 4060 GPU.

4.1. Overview of the Dataset

The IEEE PHM 2012 dataset was provided by the FEMTO-ST Institute and experiments were conducted on the experimental platform (PRONOSTIA). This experimental platform can meet the requirements of applying different working conditions to the test bearings and collect real-time bearing health monitoring data, as shown in Figure 2.

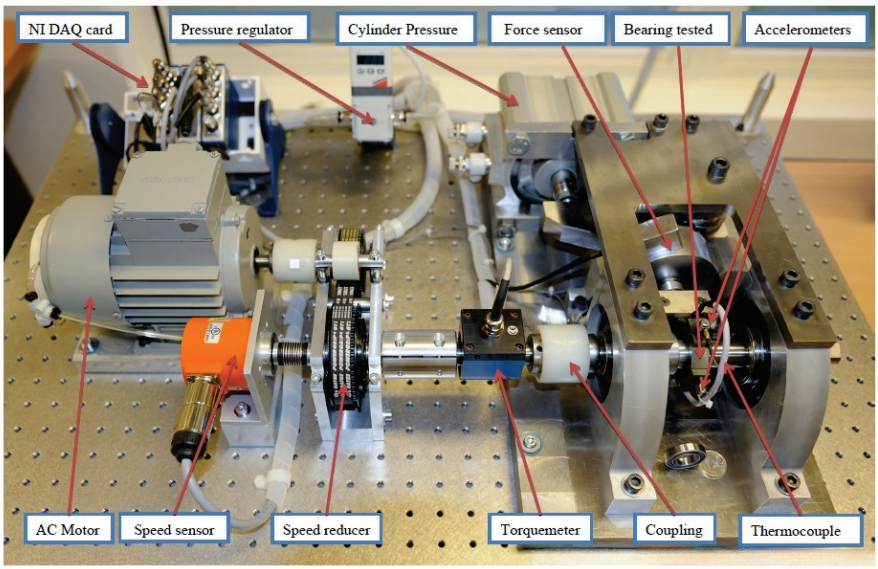


Figure 2. The PRONOSTIA testbed of rolling element bearings.

The PRNONSTIA platform consists of three parts: the rotating part, the load part, and the testing part. The rotating part is powered by a 250 W AC motor to apply torque to the shaft, ensuring that the second shaft reaches a speed of 2000 rpm. The load part applies a radial dynamic load of 4000 N to the tested bearing through a pneumatic jack, accelerating the degradation process of the bearing. The testing section is used to collect complete bearing degradation data, including collecting bearing vibration data and temperature data. For this purpose, vibration sensors are placed radially and axially on the outer ring of the bearing to measure data in both directions, with a sampling frequency of 25.6 kHz. Meanwhile, a resistance temperature detector is placed in the hole near the outer bearing ring to collect temperature data, with a sampling frequency of 0.1 Hz.

When conducting an RUL prediction analysis on bearings, the main data used are the vibration data of the bearing degradation process. This dataset contains horizontal and vertical vibration signals under three working conditions. The sensor collects signals every 10 s with a recording length of 0.1 s, and each collected vibration signal contains 2560 data points. During the entire bearing data collection process, for safety reasons, testing is stopped when the amplitude of the vibration signal exceeds 20 g. The statistics of the three operating conditions and the tested bearings are shown in Table 1.

https://doi.org/10.3390/s26092592

Table 1. The statistics of the operating conditions and the tested bearings.

Working Conditions	Load: 4000 Speed: 1800 rpm	Load: 4200 Speed: 1650 rpm	Load: 5000 Speed: 1500 rpm
Learning set	Bearing 1-1 Bearing 1-2	Bearing 2-1 Bearing 2-2	Bearing 3-1 Bearing 3-2
Test set	Bearing 1-3 Bearing 1-4 Bearing 1-5 Bearing 1-6 Bearing 1-7	Bearing 2-3 Bearing 2-4 Bearing 2-5 Bearing 2-6 Bearing 2-7	Bearing 3-3

4.2. Data Processing and Evaluation Indicators

To augment the training dataset, bearings 1-6, 1-7, 2-6, and 2-7 from operating conditions 1 and 2 were allocated to the test set, while the remaining bearings constituted the training set. Due to the unclear degradation features contained in the original vibration signal, in order to improve the training effect, it is necessary to perform a spectral analysis and z-score standardization on the dataset, and use the standardized data as the input for neural network training and testing. A comparison between the original vibration signal of the data and the standardized signal of the spectral analysis is shown in Figure 3.

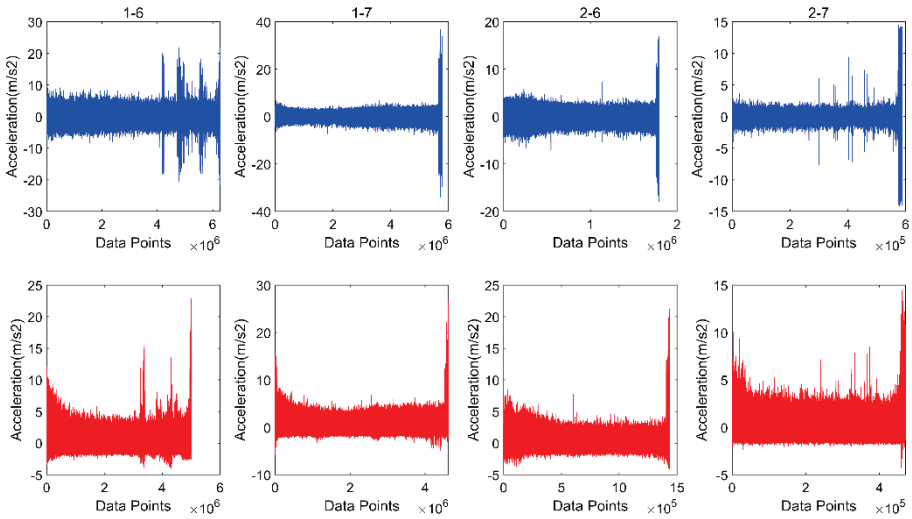


Figure 3. Time-domain and frequency-domain display of vibration data.

When assigning RUL labels to the dataset, linearly decreasing the RUL over time, and considering the differences in the service life of different bearings, setting the RUL range to 0–1 can achieve better training results.

To verify the effectiveness of the proposed method, the accuracy [25] and root mean square error (RMSE) were used as evaluation metrics for the RUL point estimation accuracy. Among them, accuracy represents the accuracy of the RUL prediction, and the larger the accuracy value, the higher the prediction accuracy; the RMSE reflects the average deviation between the predicted values and true values, with lower values indicating a better prediction performance. The calculation formulas are as follows:

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N} e^{\frac{-|m_i|}{RUL\_True}} \tag{26}$$

https://doi.org/10.3390/s26092592

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} m_i^2} \tag{27}$$

Among them, m_i is the prediction error value at the i-th time point, calculated as m_i = predicted remaining life − true remaining life; N is the number of data collection points.

In addition, in order to evaluate the performance of model uncertainty quantification, two indicators were used: the prediction interval coverage probability, PCIP, and the average prediction interval width, MPIW. Among them, PCIP represents the probability that the predicted interval contains the target value, and the larger its value, the higher the probability that the predicted target value is in the predicted interval. For uncertainty quantification goals, it is required that the PCIP value be large while the MPIW value be as small as possible, indicating a higher prediction accuracy. The calculation formula is as follows:

$$PCIP = \frac{1}{N}\sum_{i=1}^{n} rul_i \tag{28}$$

$$MPIW = \frac{1}{n}\sum_{i=1}^{n}(H_i - L_i) \tag{29}$$

Among them, if, and only if, $L_i \leq RUL_i \leq H_i$, $rul_i = 1$; in other cases, $rul_i = 0$. L_i and H_i represent the lower and upper limits of the prediction interval, respectively.

4.3. Experimental Analysis

The BCNN model transforms the training parameters in traditional CNNs into random variables without using the dropout method for uncertainty substitution. The network parameter graph is shown in Table 2.

Table 2. Network architecture diagram.

Block	Layer	Filter	Feature Size
1	Input		1 × 2048
	BConv1d ReLU()	In = 3, out = 8, k = 8	3 × 2041
	MaxPool1d	Size = 3	3 × 680
2	Same as Block 1	In = 3, out = 6	6 × 224
3	Same as Block 1	In = 6, out = 8	8 × 72
4	Same as Block 1	In = 8, out = 16	16 × 21
5	Flatten	/	336
6	BL	336 × 200	200
7	BL	200 × 200	200
8	BL Exp()	200 × 1	1
9	BL Sigmoid()	200 × 1	1

The BConv1d and BL layers used in BCNNs are Bayesian convolutional layers and Bayesian linear layers, respectively. They are the Conv1d and linear layers rewritten using Bayesian principles. The hyperparameters of the neural network were tuned using the Bayesian optimization algorithm Hyperopt, and the resulting hyperparameter values are shown in Table 3. The model training loss curve obtained using the following network configuration is shown in Figure 4. The negative loss values observed in Figure 4 were mainly due to the negative log-likelihood (NLL) term. For continuous distributions such as the Gaussian likelihood, the probability density may exceed 1, leading to positive log-

https://doi.org/10.3390/s26092592

likelihood values, and thus, negative NLL values. Although the KL divergence term is non-negative, the NLL term may dominate during training, resulting in an overall negative loss. This behavior is expected and does not indicate numerical instability. It was observed that the loss curve gradually approached a stable value at epoch 75, indicating the rationality of the training.

Table 3. The hyperparameters in the dataset.

Hyperparameters	IEEE PHM 2012
Train epoch	75
Prior distribution	N (0, 0.1)
Learning rate	0.01
Batch size	512

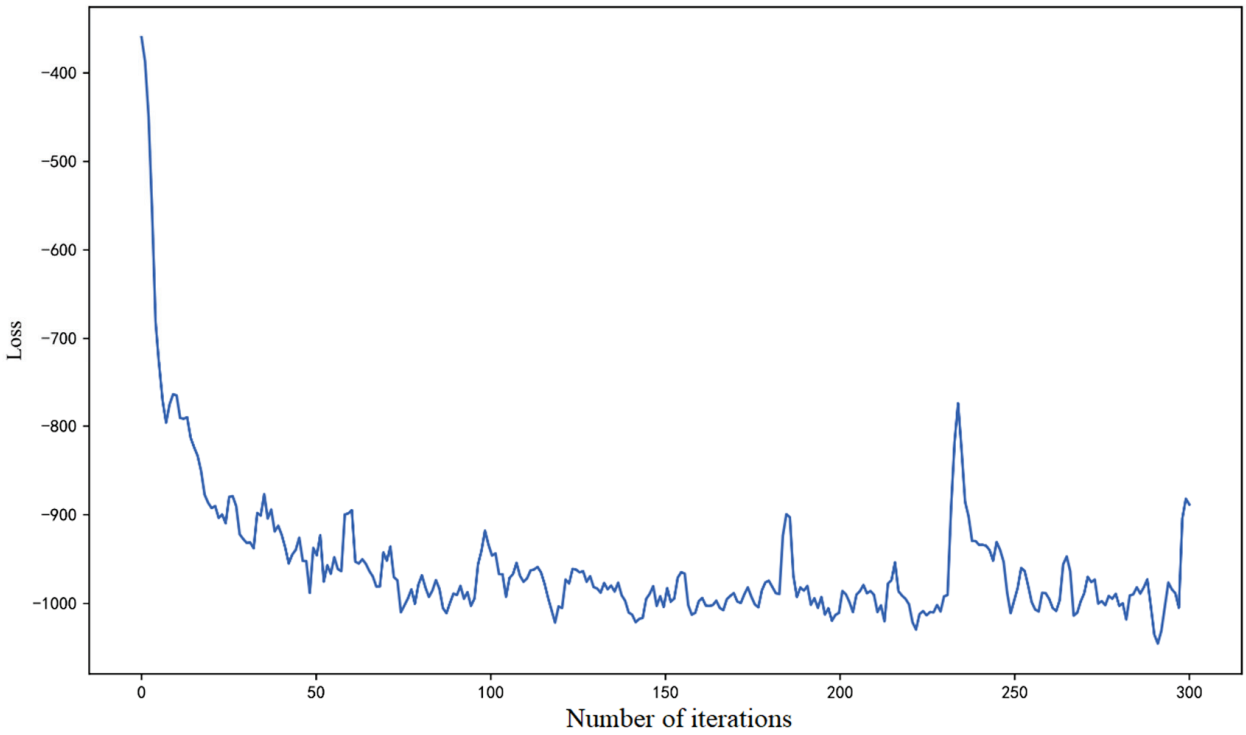


Figure 4. Training loss diagram of network model.

Using the training set described in Section 4.2 to train the proposed BCNN dual-output network, the original prediction performance of the test set is shown in Figure 5. In Figure 5, the original signal curve is close to the target value, indicating a relatively ideal prediction effect. However, there is a lot of signal stacking, and the overall trend observation is not obvious. In order to better observe and analyze the prediction performance, it is necessary to smooth the predicted data. It should be emphasized that sliding-window smoothing is used solely for visualization and does not influence the quantitative evaluation. Considering the significant differences in the number of data points in different test sets, the sliding window size for test sets 1-6 and 1-7 was selected as 10, while the sliding window size for test sets 2-6 and 2-7 was selected as 5. The resulting image is shown in Figure 6.

https://doi.org/10.3390/s26092592

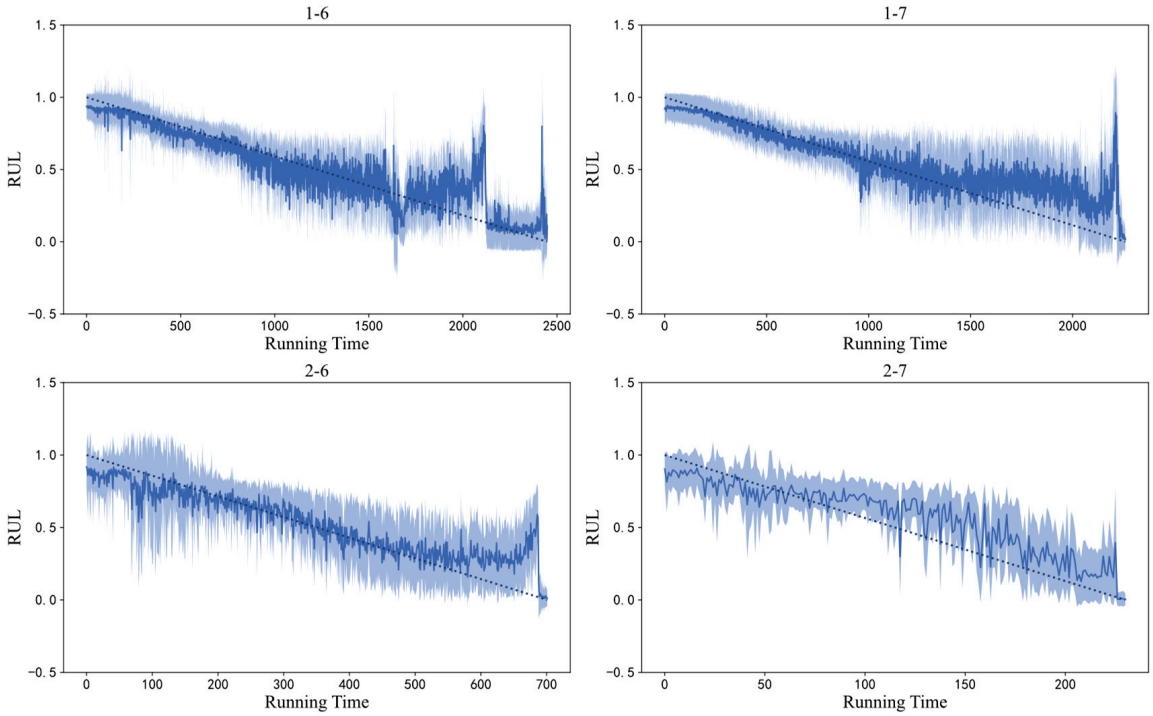


Figure 5. Unsmoothed image of the test set.

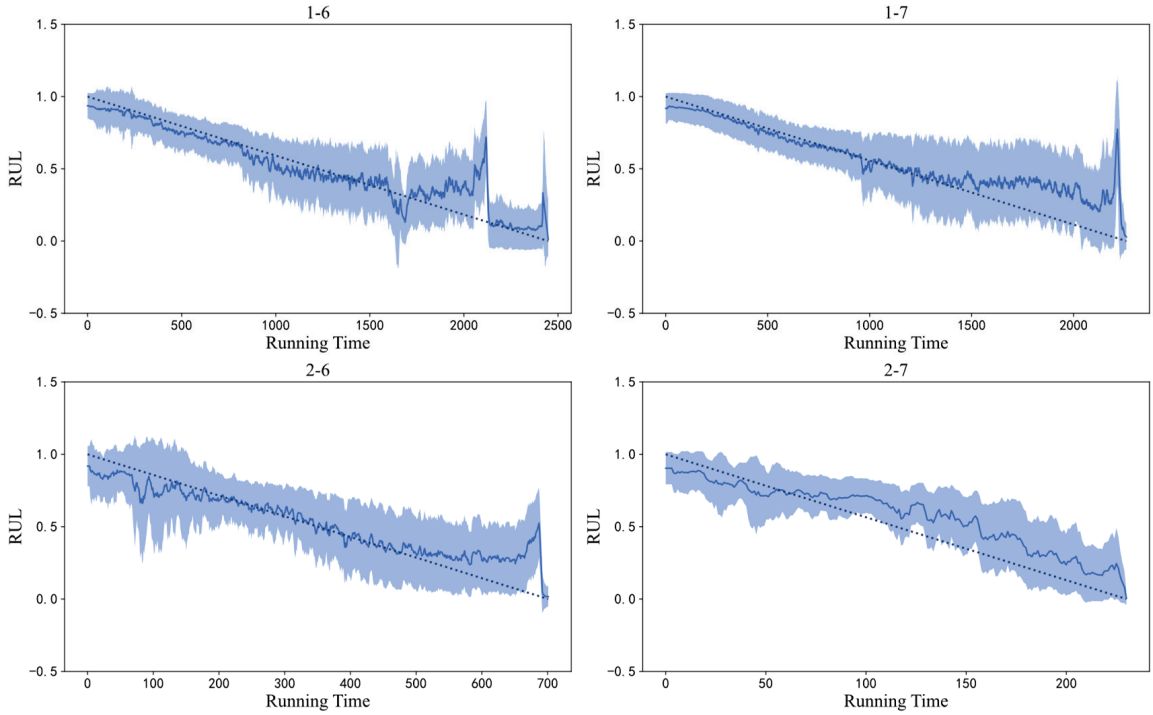


Figure 6. Smooth processing of test set images.

Based on Figures 5 and 6, it can be seen that the BCNN dual-output network had a good performance in measuring the uncertainty in bearing predictions. For testing bearings 1-6 and 1-6, in the early stage of the RUL, the degradation information contained in their vibration data was not obvious, and the data distribution was roughly the same as that of the training set, so the data itself had a low uncertainty. In the mid-term stage of the RUL, considering that individual bearings have different assembly and stress conditions, the noise signals generated during the rotation process varied from individual to individual. Compared with the training set, the uncertainty of the data itself increased, resulting in

https://doi.org/10.3390/s26092592

significant uncertainty in the predicted curve at this stage. As the bearing approached the degradation stage, due to the rapid degradation of the bearing, the degradation information that was contained increased. This situation also existed for all bearings in the training set, so their data distributions represented more degradation information, resulting in a lower uncertainty of the test data itself and a lower uncertainty of the prediction curve. This observation is important for equipment health monitoring, as an accurate prediction in the later stage is crucial for monitoring. The uncertainty mentioned in the above analysis mainly refers to aleatoric uncertainty. For epistemic uncertainty, due to the fact that the amount of data input in the training set is much larger than that in the test set, and the amount of data in the training set remains unchanged, the proportion of epistemic uncertainty is relatively small compared to aleatoric uncertainty. The presented image effect is shown in Figure 7.

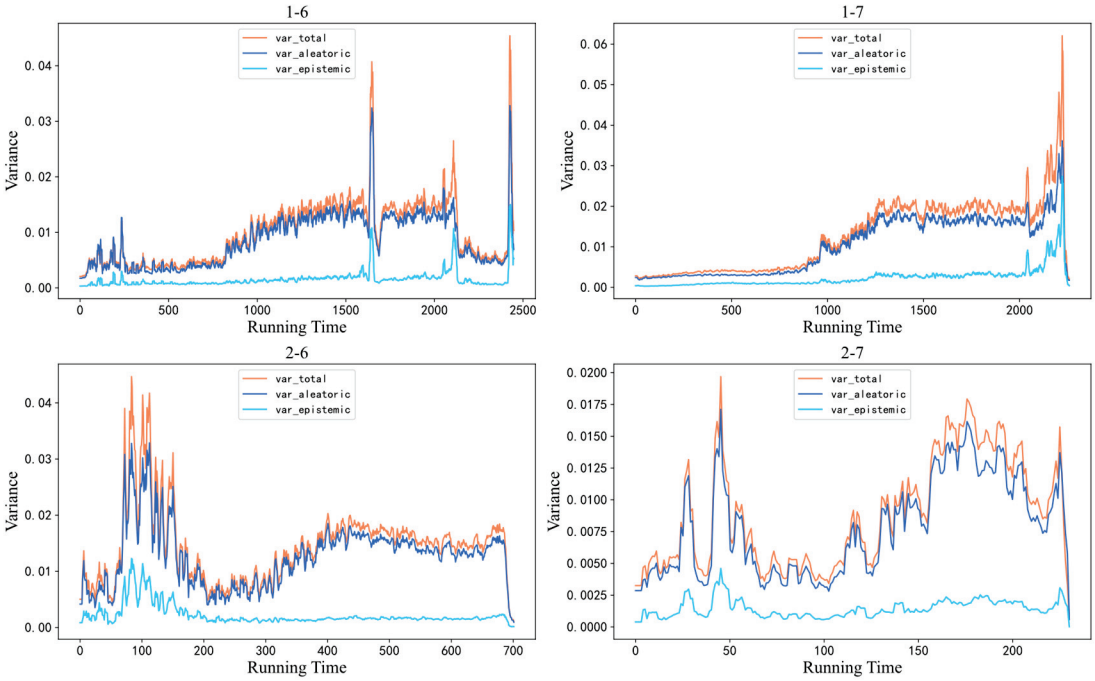


Figure 7. Test bearing uncertainty measurement diagram.

For testing bearings 2-6 and 2-7, there was significant uncertainty in the early stages of the RUL, possibly due to the individual bearing being more prone to minor failures during the early break in the stage, including obvious degradation information, resulting in a shorter overall lifespan. This phenomenon was caused by the unique assembly or stress conditions of the individual bearing, and it belongs to a special case. However, in the mid- to late stages, testing bearings 2-6 and 2-7 met the conditions analyzed above. As the bearings approached the degradation stage, their uncertainty decreased, which played a good role in bearing monitoring.

As shown in Figures 5 and 8, it can be seen that the RUL prediction performance of the method in this chapter was relatively accurate, and the confidence interval of the tested bearings included the vast majority of the target values. All four tested bearings had good predictive values near the degradation stage, almost 100% predicting the degradation failure point of the bearings. There was a certain gap between the predicted and ideal values of bearings 2-7 in the mid-term RUL stage, but their overall degradation trend remained consistent with the target value and the method still had a good prediction performance. In summary, the method proposed in this paper can effectively achieve RLU predictions and an uncertainty analysis of bearings.

https://doi.org/10.3390/s26092592

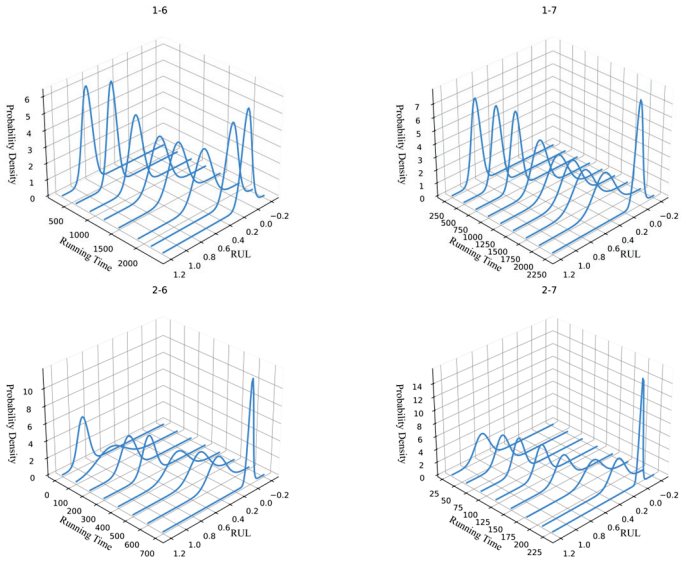


Figure 8. Test bearing PDF diagram.

4.4. Contrast Test

To evaluate the effectiveness of the proposed method, comparative experiments were conducted with several representative models for RUL predictions and uncertainty quantification, including a dual-output CNN, a single-output BCNN (Bayes-by-backprop), an MC dropout BCNN, and additional state-of-the-art architectures such as CNN-LSTM and transformer-BiLSTM variants. All the models were trained under the same experimental conditions to ensure a fair comparison. For clarity, the comparison focused on the test bearing 1-6, and the evaluation metrics defined in Section 4.2 were adopted. The quantitative comparison results are summarized in Table 4, including the prediction performance (RMSE and accuracy), uncertainty evaluation metrics (PCIP and MPIW), and computational complexity indicators (FLOPs and parameter size).

Table 4. Comparison results of model complexity and performance indicators.

Method	FLOPs	Params	RMSE	ACCURACY	PCIP	MPIW
Dual-output CNN	5.135 M	109.677 K	0.1253	0.7255	0.7304	0.2701
Single-output BCNN	10.175 M	218.8 K	0.0981	0.8009	0.7484	0.1977
MC dropout BCNN	5.135 M	109.677 K	0.0901	0.7987	0.7496	0.1686
Dual-output BCNN (proposed method)	10.176 M	219.2 K	0.0894	0.8061	0.9628	0.3740
Single-output CNN-LSTM	30.868 M	68.801 K	0.1297	0.7560	*	*
Dual-output CNN-LSTM	40.304 M	101.890 K	0.1180	0.7667	0.6299	0.1836
Single-output transformer-BiLSTM	46.277 M	67.169 K	0.1109	0.7675	*	*
Dual-output transformer-BiLSTM	46.310 M	67.234 K	0.1283	0.7557	0.7259	0.2026

* Indicates that this indicator is not applicable to the model.

From the results, it can be observed that the proposed dual-output BCNN achieved the best overall performance in uncertainty quantification. Specifically, it attained the highest PCIP value of 0.9628, and significantly outperformed all other methods. This indicates that the prediction intervals generated by the proposed model can effectively cover the true RUL values, providing more reliable decision support for practical applications. Although the MPIW of the proposed method was relatively larger, this is a reasonable trade-off, as a higher coverage probability is more critical in safety-related engineering scenarios.

Compared with the dual-output CNN, which introduces a dual-output structure to capture heteroscedastic uncertainty, the proposed method further incorporates Bayesian inference to model the epistemic uncertainty. As a result, it significantly improves both

https://doi.org/10.3390/s26092592

the prediction accuracy and interval reliability. The relatively low PCIP value of the dual-output CNN indicates that modeling only the aleatoric uncertainty is insufficient for reliable RUL predictions.

For the single-output BCNN and MC dropout BCNN, both methods attempt to quantify the epistemic uncertainty through Bayesian approximation. However, due to the absence of a dual-output structure, they fail to explicitly model the aleatoric uncertainty. Consequently, although they achieve relatively low MPIW values, their PCIP values remain limited, indicating that their prediction intervals cannot adequately capture the true RUL values. This further demonstrates the necessity of jointly modeling both types of uncertainty.

In addition, to enhance the comprehensiveness of the comparison, advanced sequence modeling architectures, including CNN-LSTMs and Transformer-BiLSTMs, were incorporated. These models leverage temporal feature extraction capabilities and show a competitive performance in their point prediction accuracy. For example, the single-output transformer-BiLSTM achieved a relatively low RMSE compared with conventional CNN-based models.

However, these models either lack uncertainty quantification mechanisms or exhibit an unsatisfactory uncertainty performance, as reflected by missing or relatively low PCIP values. Even for their dual-output variants, the PCIP values remained significantly lower than that of the proposed method, indicating a limited reliability in interval prediction. This suggests that simply increasing the model complexity or introducing temporal modeling does not guarantee effective uncertainty quantification.

From the perspective of computational complexity, the proposed dual-output BCNN introduces additional overhead due to Bayesian inference and variational parameterization. As shown in Table 4, its FLOPs and parameter size are approximately twice those of the standard CNN-based models. However, compared with more complex architectures such as CNN-LSTMs and Transformer-BiLSTMs, the proposed method maintains a moderate computational cost. More importantly, this increase in the computational complexity leads to substantial improvements in the uncertainty quantification performance, particularly in PCIP. Therefore, the proposed method achieves a favorable balance between computational efficiency and predictive reliability.

Overall, the experimental results demonstrate that the proposed dual-output BCNN not only achieves a competitive performance in point predictions compared with state-of-the-art models, but it also significantly outperforms them in uncertainty quantification. This makes it more suitable for practical PHM applications where both the prediction accuracy and reliability are essential.

5. Summary

This article proposes a dual-output unit Bayesian CNN that uses the Bayes-by-backprop method for variational inference for a Bayesian analysis. Meanwhile, the traditional single-output unit is replaced with a dual-output unit to simulate the distribution of the output data. Compared to other uncertainty quantification methods, the method proposed in this paper can capture and quantify epistemic uncertainty caused by data scarcity or model parameters, while utilizing the advantages of dual-output units to quantify any heteroscedastic uncertainty.

To verify the effectiveness of the proposed method, this paper conducted experimental validation using the PHM2012 dataset. The experimental results show that the method proposed in this paper can effectively complete the prediction and uncertainty quantification process of the RUL. In addition, by quantifying two types of uncertainties, the change process of the RUL prediction confidence interval was analyzed, providing a more valuable

https://doi.org/10.3390/s26092592

reference for equipment monitoring and maintenance. By comparing with other traditional uncertainty quantification methods, it was found that uncertainty quantification using our method almost completely includes the target value in the confidence interval, and has an average improvement of 28% on the PCIP index. It can be seen that the method proposed in this paper is superior to traditional uncertainty quantification methods in terms of uncertainty quantification. In addition, the ability to quantify two types of uncertainty provides more interpretable recommendations for monitoring and maintenance.

Author Contributions: Writing—original draft, Z.Q. and J.H.; Writing—review & editing, Z.Q.; Conceptualization, J.H.; Methodology, J.H. and Y.L.; Project administration, Y.L.; Supervision, S.M.; Visualization, X.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request from the authors.

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

References

1. Wang, L.; Cao, H.; Ye, Z.; Xu, H. Bayesian large-kernel attention network for bearing remaining useful life prediction and uncertainty quantification. *Reliab. Eng. Syst. Saf.* **2023**, *238*, 109421. [CrossRef]
2. Mishra, R.K.; Choudhary, A.; Fatima, S.; Mohanty, A.R.; Panigrahi, B.K. A self-adaptive multiple-fault diagnosis system for rolling element bearings. *Meas. Sci. Technol.* **2022**, *33*, 125018. [CrossRef]
3. Zio, E. Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice. *Reliab. Eng. Syst. Saf.* **2022**, *218*, 108119. [CrossRef]
4. Yang, B.; Liu, R.; Zio, E. Remaining Useful Life Prediction Based on a Double-Convolutional Neural Network Architecture. *IEEE Trans. Ind. Electron.* **2019**, *66*, 9521–9530. [CrossRef]
5. Shi, Z.; Chehade, A. A dual-LSTM framework combining change point detection and remaining useful life prediction. *Reliab. Eng. Syst. Saf.* **2021**, *205*, 107257. [CrossRef]
6. Ni, Q.; Ji, J.C.; Feng, K. Data-Driven Prognostic Scheme for Bearings Based on a Novel Health Indicator and Gated Recurrent Unit Network. *IEEE Trans. Ind. Inform.* **2023**, *19*, 1301–1311. [CrossRef]
7. Wang, L.; Cao, H.R.; Xu, H.; Liu, H.C. A gated graph convolutional network with multi-sensor signals for remaining useful life prediction. *Knowl.-Based Syst.* **2022**, *252*, 109340. [CrossRef]
8. Chen, K.; Liu, J.; Guo, W.; Wang, X. A two-stage approach based on Bayesian deep learning for predicting remaining useful life of rolling element bearings. *Comput. Electr. Eng.* **2023**, *109*, 108745. [CrossRef]
9. Sankararaman, S. Significance, interpretation, and quantification of uncertainty in prognostics and remaining useful life prediction. *Mech. Syst. Signal Process.* **2015**, *52*, 228–247. [CrossRef]
10. Li, G.; Yang, L.; Lee, C.-G.; Wang, X.; Rong, M. A Bayesian Deep Learning RUL Framework Integrating Epistemic and Aleatoric Uncertainties. *IEEE Trans. Ind. Electron.* **2021**, *68*, 8829–8841. [CrossRef]
11. Levin, E.; Tishby, N. A statistical approach to learning and generalization in layered neural networks. *Proc. IEEE* **2002**, *78*, 1568–1574. [CrossRef]
12. Blundell, C.; Cornebise, J.; Kavukcuoglu, K.; Wierstra, D. (Eds.) Weight Uncertainty in Neural Networks. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015.
13. Ayed, S.B.; Broujeny, R.S.; Hamza, R.T. Remaining Useful Life Prediction with Uncertainty Quantification Using Evidential Deep Learning. *J. Artif. Intell. Soft Comput. Res.* **2024**, *15*, 37–55. [CrossRef]
14. Plakias, S.; Boutalis, Y.S. A Deep Equilibrium Model for Remaining Useful Life Estimation of Aircraft Engines. *Electronics* **2025**, *14*, 2355. [CrossRef]
15. Benker, M.; Furtner, L.; Semm, T.; Zaeh, M.F. Utilizing uncertainty information in remaining useful life estimation via Bayesian neural networks and Hamiltonian Monte Carlo. *J. Manuf. Syst.* **2021**, *61*, 799–807. [CrossRef]
16. Peng, W.; Ye, Z.-S.; Chen, N. Bayesian Deep-Learning-Based Health Prognostics Toward Prognostics Uncertainty. *IEEE Trans. Ind. Electron.* **2020**, *67*, 2283–2293. [CrossRef]
17. Caceres, J.; Gonzalez, D.; Zhou, T.; Droguett, E.L. A probabilistic Bayesian recurrent neural network for remaining useful life prognostics considering epistemic and aleatory uncertainties. *Struct. Control Health Monit.* **2021**, *28*, e2811. [CrossRef]

https://doi.org/10.3390/s26092592

18. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Hassabis, D. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef]
19. Shridhar, K.; Laumann, F.; Liwicki, M. A Comprehensive guide to Bayesian Convolutional Neural Network with Variational Inference. *arXiv* **2019**, arXiv:1901.02731. [CrossRef]
20. Lei, Y.; Wang, W.; Yan, T.; Li, N.; Nandi, A. Residual Convolution LSTM Network for Machines Remaining Useful Life Prediction and Uncertainty Quantification. *J. Dyn. Monit. Diagn.* **2021**, *1*, 2–8.
21. Sun, M.; Zhang, T.; Wang, Y.; Strbac, G.; Kang, C. Using Bayesian Deep Learning to Capture Uncertainty for Residential Net Load Forecasting. *IEEE Trans. Power Syst.* **2020**, *35*, 201–211. [CrossRef]
22. Zhao, Z.; Wu, J.; Wong, D.; Sun, C.; Yan, R. Probabilistic Remaining Useful Life Prediction Based on Deep Convolutional Neural Network. In Proceedings of the 9th International Conference on Through-Life Engineering Service, Cranfield, UK, 3–4 November 2020.
23. Blei, D.M.; Kucukelbir, A.; McAuliffe, J.D. Variational Inference: A Review for Statisticians. *J. Am. Stat. Assoc.* **2017**, *112*, 859–877. [CrossRef]
24. Kingma, D.P.; Salimans, T.; Welling, M. Variational Dropout and the Local Reparameterization Trick. In Proceedings of the 32nd International Conference on Machine Learning (ICML), Lille, France, 7–9 July 2015.
25. An, Q.; Tao, Z.; Xu, X.; El Mansori, M.; Chen, M. A data-driven model for milling tool remaining useful life prediction with convolutional and stacked LSTM network. *Measurement* **2020**, *154*, 107461. [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.

https://doi.org/10.3390/s26092592

sensors

MDPI

Article

A Two-Stage Classification Method for Improved Fault Detection in Wind Turbines Based on SCADA Data

Jiazhi Dai [1,2], Mario Rotea [2,3] and Nasser Kehtarnavaz [1,2,*]

[1] Department of Electrical and Computer Engineering, University of Texas at Dallas, Richardson, TX 75080, USA; jiazhi.dai@utdallas.edu
[2] Center for Wind Energy, University of Texas at Dallas, Richardson, TX 75080, USA; rotea@utdallas.edu
[3] Department of Mechanical Engineering, University of Texas at Dallas, Richardson, TX 75080, USA
* Correspondence: kehtar@utdallas.edu

Abstract

Fault detection is essential for the reliable operation of wind turbines. Traditional supervised methods for fault detection based on SCADA data face highly imbalanced datasets of normal and fault samples. This paper presents a two-stage detection method to address this limitation by integrating unsupervised anomaly detection or classification with supervised classification. In the first stage, the unsupervised classifier of OCSVM, together with two complementary anomaly scores, is used to flag deviations from normal operation or separate abnormal data samples from normal data samples. In the second stage, the supervised classifier of CNN is applied to the detected abnormal data samples to identify fault samples among only these samples, thus enhancing the discrimination capability between normal and abnormal conditions. Experiments on real-world SCADA data show that the introduced two-stage detection method noticeably improves fault detection compared to supervised methods, both in terms of accuracy and missed fault rates.

Keywords: wind turbine fault detection; two-stage unsupervised and supervised fault detection; fault detection based on SCADA data

1. Introduction

Wind power constitutes a major source of renewable energy, contributing approximately 7.8% to electricity generation globally [1]. Reliable operation of wind turbines is of critical importance. Turbine faults can lead to downtime, increased maintenance costs, and loss of energy production. Accurate fault detection enables reliable wind turbine operation.

Supervisory Control and Data Acquisition (SCADA) sensors are widely deployed in wind turbines, providing continuous measurements of turbine conditions such as wind speed, power output, temperatures, vibration, and electrical signals [2]. Data from SCADA sensors, called SCADA data, allow addressing fault detection by machine learning in a data-driven manner. Existing machine learning approaches to fault detection involve supervised learning methods. As mentioned in [3,4], these methods require large amounts of labeled data consisting of both normal and fault conditions. In practice, fault-labeled datasets are often scarce, incomplete, or inconsistent. Furthermore, wind turbine data are often highly imbalanced, with normal-operation data samples dominating, making it difficult for supervised models to effectively distinguish fault from normal samples.

To improve fault detection beyond the existing supervised methods, this paper introduces a two-stage classification method by combining unsupervised classification with

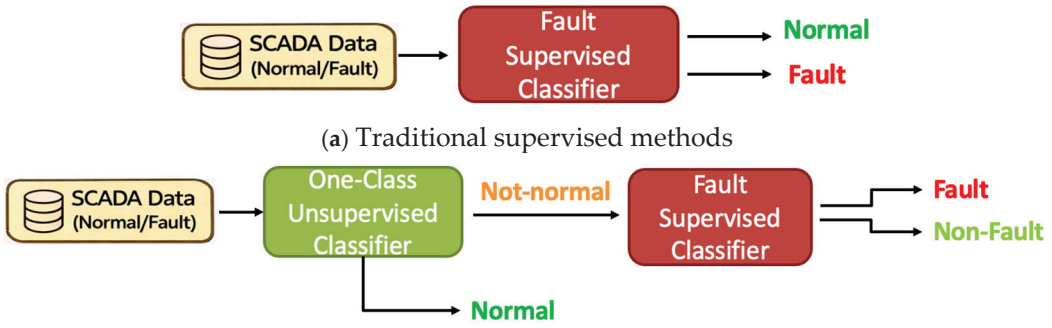
supervised classification. This two-stage method first identifies SCADA data samples that deviate from normal behavior using only normal data (a single class classification), and then applies a supervised model to those samples that are flagged as not-normal or abnormal samples to identify fault samples among them. By filtering out a large amount of normal data in the first stage, the introduced method reduces the search space for the supervised model to achieve a more accurate fault detection compared to just using supervised classification.

The main contributions in this paper are as follows: (i) Two complementary anomaly scores are introduced for carrying out the unsupervised classification: a feature-based anomaly score capturing deviations in individual SCADA data features, and a correlation-based anomaly score represented as a heatmap characterizing changes in the relationships among SCADA features. For each anomaly score, a one-class learning model is employed to detect abnormal samples. (ii) A two-stage fault detection framework is developed by integrating unsupervised anomaly detection with supervised fault detection to address the severe class imbalance commonly observed in SCADA datasets.

The remainder of the paper is organized as follows. Section 2 reviews related work on wind turbine fault detection. Section 3 presents the introduced methodology of two-stage classification. Section 4 describes the dataset and experimentations performed. The results obtained and their discussion are stated in Section 5, followed by the conclusion in Section 6.

2. Related Works

Wind turbine fault detection based on SCADA data has been extensively studied in the literature due to its importance in the reliable operation of wind turbines. Existing machine learning approaches are primarily supervised methods, while in this work, unsupervised learning is used as the frontend to supervised methods. The difference between our two-stage detection method and supervised methods is illustrated in Figure 1.



(**a**) Traditional supervised methods

(**b**) Our two-stage method

Figure 1. Difference between (**a**) traditional supervised detection methods and (**b**) the developed two-stage detection method. The label "Non-Fault" refers to not-normal samples identified by the first-stage unsupervised classifier that get classified by the second-stage classifier not to be fault samples.

Early studies on wind turbine condition monitoring mainly relied on statistical approaches. These methods include trend analysis, threshold-based alarms, clustering, and the Normal Behavior Model (NBM) [5]. Trending and threshold-based methods monitor deviations of key features or variables (e.g., temperature, power output) from their expected ranges. While simple and interpretable, as pointed out in [6], these methods often fail to capture complex relationships among SCADA features or variables, making them more susceptible to false alarms and performance degradation. NBM is one of the most widely used methods in monitoring SCADA data. In this approach, a model is trained

https://doi.org/10.3390/s26123865

using normal data to predict expected behavior, and deviations between predicted and observed values get flagged as anomalies. Several variations of NBM have been considered, including regression models, neural networks, and statistical models. These methods offer the advantage of using SCADA data that are readily available and do not require the installation of specialized sensors on a turbine.

The above traditional methods are often limited by their reliance on handcrafted features, which may not generalize well to different fault types. With advancements in machine learning, supervised models have been designed by treating wind turbine fault detection as a classification problem based on labeled data. Conventional machine learning models such as Support Vector Machine (SVM), Random Forest, Xgboost, and Ensemble methods, e.g., [7–9], have been used to classify different fault types based on SCADA features. More recently, deep learning models, including Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM), have been utilized for fault detection, e.g., [10–12]. These models learn complex patterns in data and have been shown to provide better performance compared to earlier methods.

The bottleneck in deep learning models is the need for large datasets of labeled normal and fault samples. In practice, as noted in [13], such datasets are scarce and highly imbalanced. Noting that fault events occur infrequently compared to normal operation, it becomes challenging to train supervised models. To address this shortcoming of supervised models, one-class unsupervised methods, which use only readily available normal data, can be considered to lower the search space of supervised models. In [14–18], unsupervised classifiers, including One-Class SVM (OCSVM), AutoEncoder (AE), and LSTM-AE, have been utilized. Despite their advantage of not requiring labeled data, unsupervised methods in general generate more false positives. Essentially, supervised methods are simple but struggle when normal data dominates a dataset, while unsupervised methods ease this shortcoming but are subject to less precision.

In this work, we combine unsupervised and supervised classification to take advantage of the strengths of each of these classifications in order to reach an improved detection outcome as compared to commonly used supervised methods.

3. Two-Stage Methodology

The two-stage method presented in this paper aims to automatically detect deviations from the normal state of a wind turbine using the entire SCADA data or measurements. The deviation from normal operation is carried out based on only normal data of a wind turbine that are readily available. As shown in Figure 2, the unsupervised classification is made up of three main steps: (1) data preprocessing, (2) one-class detection, and (3) anomaly score computation. The input consists of multidimensional SCADA data samples collected over time, and the output consists of SCADA data samples identified as deviations from the normal operation state.

Two complementary anomaly scores are computed to quantify deviations from normal behavior. Finally, two one-class learning models are trained on normal data to identify abnormal samples. These samples can be directly used for fault detection or further processed with a supervised classifier to identify fault type.

3.1. Data Preprocessing

3.1.1. Feature Merging

A raw SCADA dataset contains many measurements that are physically related and often provide overlapping information. Examples include three-phase electrical variables, multiple temperatures associated with the same component, and vibration measurements collected from closely located sensors. Using all the features increases dimensionality,

https://doi.org/10.3390/s26123865

which in turn makes the anomaly detection process more sensitive to noise. Therefore, a feature merging step is first performed to bring down the dimensionality.

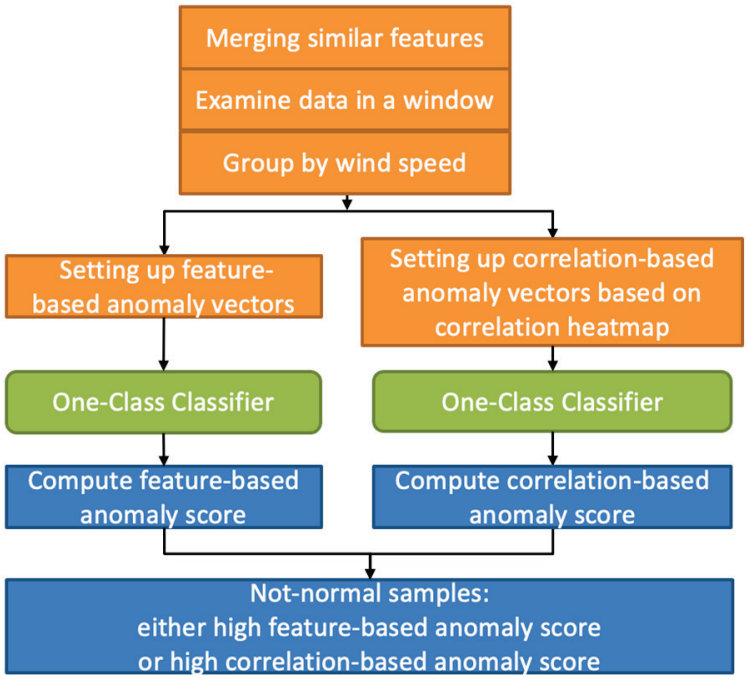


Figure 2. Flowchart of the developed two-stage fault detection method.

Feature merging is achieved here by the aggregation of very similar physical features. In other words, SCADA features that exhibit both physical consistency and similarity are merged. Physical consistency means that the features correspond to the same component, such as temperatures of different generator phases. Similarity means that these features exhibit similarity over time under normal operating conditions.

To decide which features to merge, the correlation among them is examined by computing the Pearson correlation coefficient. For a pair of SCADA features x_i and x_j, the Pearson correlation coefficient is computed as follows:

$$p_{ij} = \frac{cov(x_i, x_j)}{\sigma_i \sigma_j} \tag{1}$$

where $cov(x_i, x_j)$ denotes the covariance between the two features, σ_i and σ_j their standard deviations. As mentioned in [19], features are considered highly correlated when their correlation coefficients are close to 1 under normal operating conditions. In such cases, the features are interpreted as carrying the same information and are therefore suitable for merging. Figure 3 illustrates an example of using the Pearson correlation heatmap to group the SCADA features into speed, electric, and temperature groups. When the features within the same group exhibit strong correlations, they are merged. For example, a subset of features within the temperature group, such as the three transformer temperature measurements, shows strong dependencies or carries the same information, and thus they are merged into a single feature. Similar merging is performed for other feature groups or families. In other words, related features are merged when the following conditions are satisfied: (i) they represent the same physical quantity or the same component from different phases, and (ii) they exhibit very high correlation during normal operation.

Merged features are considered to be the average of the original features. Averaging is adopted here because it preserves the common trend in the same feature group while reducing random fluctuations from individual sensors. In situations where a different

https://doi.org/10.3390/s26123865

merging strategy is more physically meaningful (for example, for vibration signals), root-mean-square values are used.

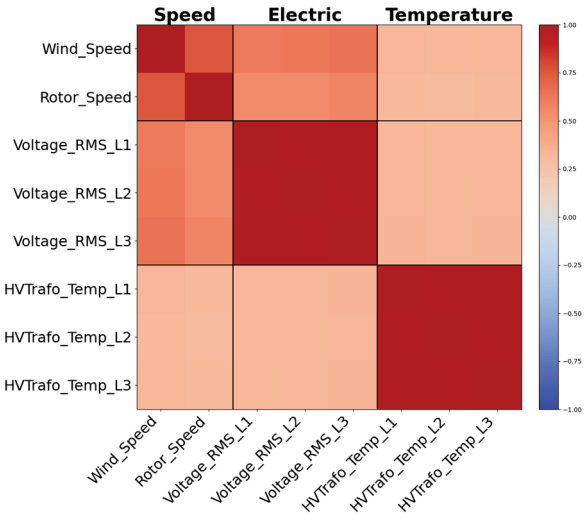


Figure 3. Example illustrating the use of Pearson correlation heatmap to conduct feature merging.

3.1.2. Window Size Selection

To capture temporal changes, the data are examined in a sliding window manner as illustrated in Figure 4. Given a time-series signal $x(t)$, each sample is constructed as a sequence of measurements over a fixed window of length T:

$$X_t = [x(t), x(t+1), \ldots, x(t+T-1)] \tag{2}$$

where T represents the window size. Many abnormal conditions in wind turbines do not appear as a single outlier at one time instant, but rather as a gradual deviation or a persistent abnormal pattern over a time period. If the window is too short, the resulting sample may only reflect instantaneous sensor noise, making it difficult to distinguish abnormal behavior from normal behavior. If the window is too long, abnormal behavior may get diluted by a large amount of normal data within the same window. Therefore, the window size should strike a balance between temporal continuity and preserving anomalous contrast. The selection of window size can be carried out by going from too short to too long windows and then seeing for what window size range more or less the same high-performance outcome is obtained.

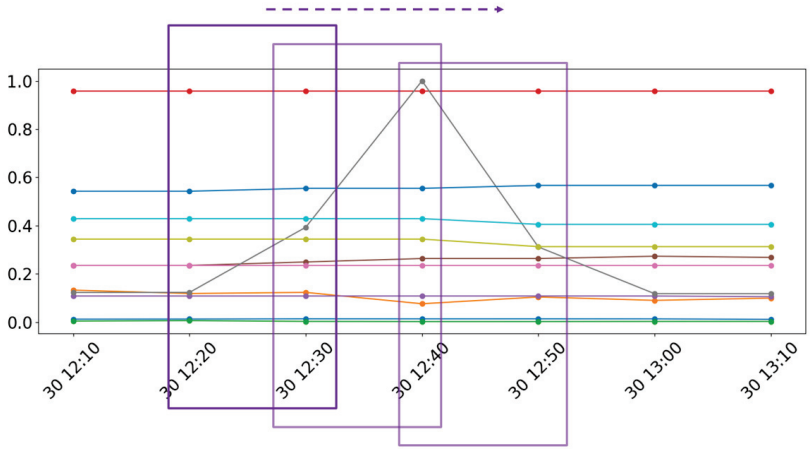


Figure 4. Illustration of sliding window.

https://doi.org/10.3390/s26123865

3.1.3. Grouping by Wind Speed

Under different wind speeds, the relationship between SCADA data features, such as power output, temperature, and vibration, changes considerably. Figure 5 displays a typical power curve of a wind turbine, indicating its power output vs. wind speed.

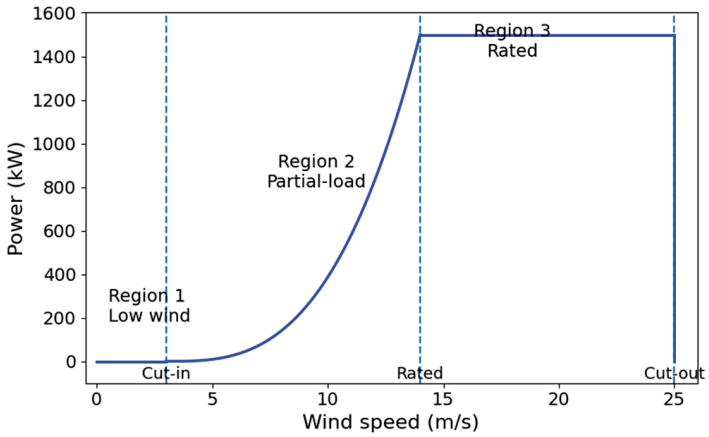


Figure 5. A typical wind turbine power curve.

To address this issue, the dataset is divided into three groups corresponding to the three regions of the power curve. At low wind speeds, typically below or slightly above the cut-in threshold, a turbine generates little or no power. In this region, SCADA data features such as power output, generator temperature, and rotational speed remain relatively low. In the mid wind speed region, the turbine power follows a cubic curve. At high wind speeds, the turbine reaches its rated power output, and this power is regulated to remain constant. In this region, the turbine control system adjusts blade pitch and other components to limit the mechanical load, ensuring a safe operation. Grouping by wind speed helps reduce false alarms caused by variations in turbine operation under different wind speeds. In the subsequent modules of the unsupervised classifier, each wind speed data group is treated independently.

3.2. One-Class Detection or Classification Using OCSVM

To detect deviations from normal turbine operation, a one-class classifier is utilized. This classifier is trained exclusively based on only normal data, allowing it to learn the underlying distribution of normal turbine data. Here, OCSVM is used as the one-class classifier as it has the ability to identify outliers without requiring labeled anomaly or abnormal samples.

A sliding window representation $X_t \in R^{m \times T}$ is first constructed, where m is the number of features and T is the window size. For each window, a set of anomaly scores is computed to characterize different types of deviations. These scores are aggregated into a feature vector $z_t = \varphi(X_t)$, serving as the input to the OCSVM. Here, z_t refers to either ${z_t}^F$ or ${z_t}^C$ depending on whether the feature-based or correlation-based OCSVM is used. Transformation for different anomalies is explained in Section 3.3. Given a set of normal samples $\{z_1, z_2, \ldots z_n\}$, the OCSVM learns a decision boundary that encloses the great majority of these samples by solving the following optimization problem

$$\min_{w,\xi,\rho} \frac{1}{2}\|w\|^2 + \frac{1}{\nu N}\sum_{t=1}^{N}\xi_t - \rho \tag{3}$$

subject to the constraint

$$(w \cdot \phi(z_t)) \geq \rho - \xi_t, \xi_t \geq 0 \tag{4}$$

https://doi.org/10.3390/s26123865

where the decision function is defined as:

$$f(z_t) = sgn((w \cdot \phi(z_t)) - \rho) \tag{5}$$

with $\phi(\cdot)$ being an RBF (Radial Basis Function) kernel; the parameter $\nu \in (0, 1]$ specifies the maximum allowable fraction of outliers; N denotes the number of samples, w the normal vector of the decision boundary, and ρ its location in the feature space. The slack variable ξ_t is added to allow for minor deviations due to noise. A data sample gets classified as follows:

$$f(z_t) \geq 0 \rightarrow \ Normal; f(z_t) < 0 \rightarrow \ Abnormal$$

In this work, the OCSVM is trained using only normal SCADA data within each wind speed group. The parameter ν is set to 0.05 based on the assumption that abnormal operating conditions constitute only a small fraction of SCADA data. An RBF kernel is employed where its scale parameter is automatically computed from the variance of training data. During inference, each SCADA window is evaluated using the learned decision function, and data samples that fall outside the normal region are identified as not-normal or abnormal samples.

3.3. Anomaly Score Computation

The outcome of the one-class classification largely depends on how the SCADA window X_t is represented. Different representations reflect different aspects of turbine behavior, and thus different anomalies. Here, two complementary anomaly representations or scores are considered: (i) feature-based anomaly, which captures deviations in individual features, (ii) correlation-based anomaly, which captures changes in relationships among the features. These two representations are used to compute anomaly vectors, which are then fed to the OCSVM model to obtain corresponding anomaly scores.

3.3.1. Feature-Based Anomaly Score

The feature-based anomaly score measures how much individual SCADA features deviate from their normal behavior. For each feature i, its mean μ_i and standard deviation σ_i are estimated from normal data, i.e., for a given window $X_{i,t}$, the deviation is computed as follows:

$$z_{i,t}{}^F = \varphi(X_{i,t}) = \frac{1}{T}\sum_{\tau=1}^{T}\frac{|x_i(t+\tau) - \mu_i|}{\sigma_i} \tag{6}$$

The vector $z_t{}^F = \left\{z_{1,t}{}^F,\ z_{2,t}{}^F, ..., z_{m,t}{}^F\right\}$ indicates the average normalized deviation of all the features over the window. A higher $z_{i,t}{}^F$ value indicates that the feature signals deviate more from their normal levels. This vector is then used as the input to the feature-based OCSVM to obtain the feature-based anomaly score. Since anomaly scores are concentrated within a narrow range under normal conditions, an exponential transformation is applied to amplify abnormal behavior. Meanwhile, very small scores are regarded as noise and set to zero.

3.3.2. Correlation-Based Anomaly Score

While a feature-based anomaly score captures individual feature deviations, it does not account for the relationships among features. However, under normal operation, SCADA features exhibit stable correlations. To capture changes in these relationships, a correlation-based anomaly score is considered. First, a baseline correlation matrix $R_{i,j}^{(0)} \in R_{m\times m}$ is

https://doi.org/10.3390/s26123865

computed using normal data in each wind speed region. Specifically, the data are divided into sliding windows, and for each window k, a correlation matrix is computed as follows:

$$R_{i,j}^{(k)} = p_{ij} = \frac{cov(x_i, x_j)}{\sigma_i \sigma_j} \tag{7}$$

The baseline correlation matrix is then obtained by averaging over N windows in the same wind speed region:

$$R_{i,j}^{(0)} = \frac{1}{N} \sum_{k=1}^{N} R_{i,j}^{(k)} \tag{8}$$

For each sliding window X_t, a correlation matrix $R^{(t)}$ is computed in the same manner. The difference between the current correlation matrix and the baseline correlation matrix is computed. The upper-triangular elements of the difference matrix are extracted and arranged into the following correlation-based anomaly vector:

$$z_t{}^C = \varphi^C(X_t) = vec_u\left(R^{(t)} - R^{(0)}\right) \tag{9}$$

where $vec_u(\cdot)$ denotes the upper-triangular entries. This vector is then used as the input to the correlation-based OCSVM to obtain the correlation-based anomaly score. This score captures relational anomalies; that is to say, even when individual features remain within their normal ranges, abnormal behavior can still appear via changes in their relationships.

3.3.3. Combination or Union of Two Anomaly Scores

After inputting the feature-based and correlation-based anomaly vectors separately into two one-class classifiers, the two corresponding anomaly scores are computed to capture complementary aspects of abnormal turbine behavior. The feature-based anomaly score focuses on individual feature deviations, identifying situations where one or more SCADA data features exceed their normal operating levels. This type of anomaly is typically associated with observable changes in magnitude, such as a sudden increase in bearing temperature or an abnormal drop in power output. In contrast, the correlation-based anomaly score focuses on capturing changes in the relationships among the features. Under normal operation, SCADA features exhibit stable correlations due to the inherent coupling of turbine components. When these relationships are altered, it may indicate abnormal operating conditions even if individual feature values remain within their normal limits. It is to be emphasized that the two anomaly scores are different. Feature-based anomaly scores reflect local deviations, while correlation-based anomaly scores reflect system-level changes. As a result, relying on only one type of anomaly score may not capture all possible abnormalities.

In practice, abnormal events are relatively rare and typically account for only a small fraction of SCADA data. For example, as reported in [20], the percentage of anomalous samples is less than 5% of the entire SCADA dataset. A percentile-based selection strategy is thus adopted here. For each anomaly score, a top percentage of samples with the highest scores are selected as abnormal data samples. The final set of abnormal data samples is obtained by taking the union of these two sets. This process is illustrated in Figure 6. In this figure, yellow bars represent abnormal SCADA samples across the time axis, and red bars represent fault samples across the time axis. By using both of the anomaly scores, ideally, abnormal samples should capture all the fault samples; i.e., all the fault samples should be marked as abnormal, and none of the fault samples should be marked as normal.

It is worth pointing out that the unsupervised classifier in Stage 1 is not intended to serve as the final fault detector. Instead, its primary role is to identify samples deviating from normal-operation samples and reduce the severe class imbalance commonly

https://doi.org/10.3390/s26123865

encountered in SCADA datasets. These samples are then used to train and evaluate the Stage-2 supervised classifier. In other words, the unsupervised classifier can be viewed as an anomaly-screening component for conducting supervised classification. Also, the unsupervised classification is applied independently to training and testing, and only the identified abnormal samples are passed to the supervised classifier.

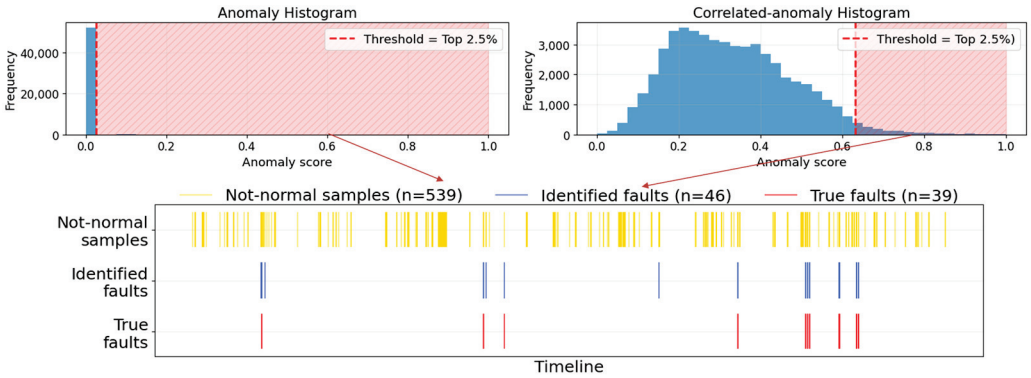


Figure 6. Histograms of anomaly scores; union of feature-based anomaly score and correlation-based anomaly score represented as yellow bars (abnormal samples), blue bars (identified fault samples), and red bars (true fault samples) across the time axis.

4. Experimental Studies

4.1. Dataset

Our experimental studies are reported in this section by using the public domain Zenodo Wind Farm B dataset provided in [21]. This dataset contains SCADA data or signals collected from three wind turbines (#27, #34, #77) that exhibit faults. The data is sampled at a 10-min interval and includes a variety of features such as wind speed, power output, and multiple sensor measurements. The dataset consists of normal and fault conditions. It was randomly divided into 70% training, 10% validation, and 20% testing subsets. The validation subset was used for anomaly-threshold sensitivity analysis and window-size selection, whereas the testing subset was reserved exclusively for performance evaluation. Since fault occurrences were highly imbalanced and unevenly distributed throughout the dataset, a fixed-ratio random split was adopted to ensure sufficient fault samples in each subset. To avoid temporal leakage, the training, validation, and testing windows were generated separately, and overlapping windows were used within each subset, but no windows were shared across different subsets. The number of training/validation/testing samples is provided in Table 1.

Table 1. Number of samples.

Turbine Number	# Training Samples (Normal/Fault)	# Validation Samples (Normal/Fault)	# Testing Samples (Normal/Fault)
27	36,633/134	5233/19	10,467/39
34	36,568/199	5224/28	10,449/57
77	36,669/98	5238/14	10,478/28

For the unsupervised model, only normal data was used during training to learn the distribution of normal data. Fault data was exclusively used for evaluation. For the supervised model, both normal and fault data were used for training and testing under the same data split regime. The preprocessing steps, including normalization and feature merging, as described in Section 3, are stated next.

https://doi.org/10.3390/s26123865

4.2. Preprocessing

4.2.1. Input Features

The SCADA features in this dataset reflect the electrical, thermal, and mechanical conditions of the wind turbines. They include voltage and current measurements, generator and transformer temperatures, bearing temperatures, vibration signals, and gearbox-related temperatures, as summarized in Table 1. Since many SCADA variables are collected from multiple phases or closely related components (e.g., L1, L2, L3 phases), strong correlations are observed among these signals. Feature merging was thus performed by aggregating highly correlated features into representative features. Features such as wind speed and wind direction were not merged, while variables like grid frequency, which remain constant over time, were excluded to reduce the dimensionality. All the merged features of the dataset are listed in Table 2.

Table 2. Original features and their corresponding merged features.

Original Features	Merged Features
RMS voltage L1, L2, L3	Voltage_RMS
RMS current L1, L2, L3	Current_RMS
Generator stator winding temperature phase 1, 2, 3	Gen_Phase_Temp
HV transformer temperature L1, L2, L3	HVTrafo_Phase_Temp
Generator bearing 1, Generator bearing 2	Gen_Bear_Temp
Drive train vibration Z, Tower vibration X, Y	Tower_Train_Vibration
Gearbox oil inlet temperature, oil tank temperature	Gearbox_Temp

4.2.2. Candidate Threshold Analysis

To determine an anomaly-selection threshold for the first stage, the degree of class imbalance commonly reported in SCADA datasets was considered. A previous study in [20] reported fault-to-normal ratios of approximately 1:360, 1:490, and 1:2930, corresponding to fault proportions below 0.5% of the total data. A similar fault proportion (approximately 0.5%) was thus considered in our study.

Noting that the objective of the first-stage unsupervised detector is not to directly identify fault samples but rather to identify suspicious samples for further examination by the second-stage supervised classifier, the number of retained candidate samples needs to be higher than the true faults. A sensitivity analysis was conducted using the validation subset. The 2-anomaly-score-based OCSVM was employed, and the retained percentage of not-normal samples for each anomaly score was varied from 0.5% to 5%. To avoid introducing assumptions about the final window size during the threshold selection, a temporary 12-h window was considered. The effect of window size was investigated in the subsequent experiment. For each threshold, both fault coverage and the resulting class imbalance after Stage 1 were examined. The results obtained are shown in Table 3.

Table 3. Sensitivity Analysis of Anomaly-Selection Threshold.

Threshold	Fault Coverage
0.5%	5%
1%	16%
1.5%	26%
2%	58%
2.5%	94%
3%	94%
4%	94%
5%	94%

https://doi.org/10.3390/s26123865

The results in Table 3 indicate that increasing the retained percentage improved fault coverage. Although a 5% threshold still achieved the highest fault coverage, it also retained a substantially larger number of normal samples, resulting in a post-filtering normal-to-fault ratio of approximately 19:1. In contrast, a target candidate ratio of approximately 2.5% achieved a substantially lower imbalance ratio while maintaining high fault coverage. Therefore, a target candidate ratio of 2.5% was selected.

4.2.3. Window Size Selection

As discussed in [22], for the SCADA data at a sampling rate of 10 min, the window length needs to be several hours to distinguish between normal and fault samples. Figure 7 presents the fault coverage rates obtained using window lengths ranging from 1 to 12 h.

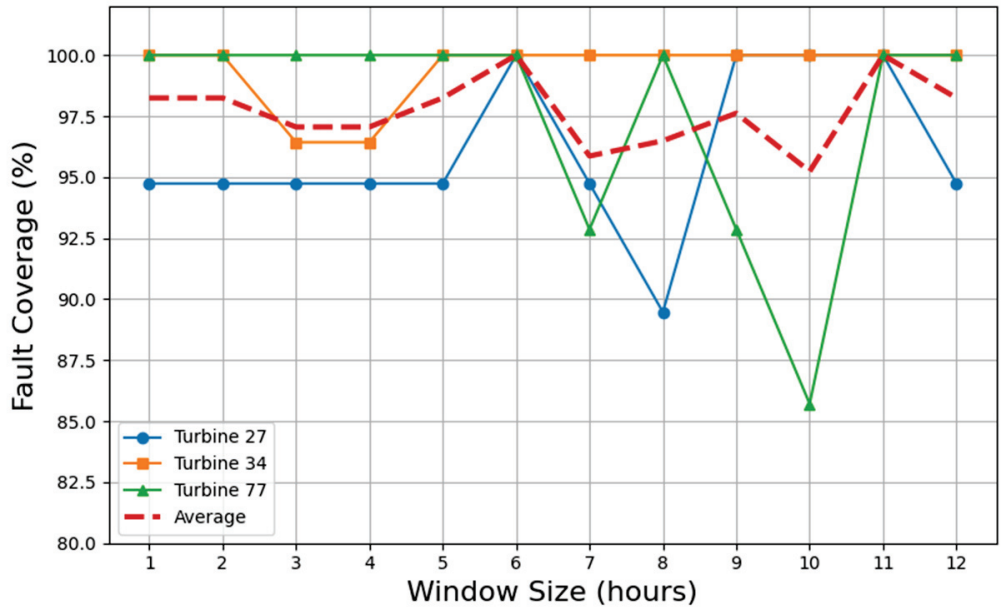


Figure 7. Fault coverage rate for each window size across different turbines.

High fault coverage was achieved across a broad range of window sizes, indicating that the developed framework is relatively robust to the window-length selection. Among the evaluated configurations, both the 6-h and 11-h windows achieved 100 fault coverage. The 6-h window size was selected since it was the shortest window length achieving complete fault coverage, was less susceptible to the influence of missing data and required lower computational cost than longer windows. This window length corresponds to 36 SCADA samples. The slight reduction in coverage for some longer windows is likely due to the dilution of fault-related patterns by a larger proportion of normal operating data within the window.

4.2.4. Grouping by Wind Speed

To account for different wind speed conditions, the data were grouped based on wind speed. Specifically, the entire dataset was divided into three wind speed intervals or regions (low, medium, and high wind speed regions). The power curves of the three turbines in the dataset are shown in Figure 8. It can be seen that the curves are similar, having the three wind speed regions of (0–2.5 m/s), (2.5–11.5 m/s), and (>11.5 m/s).

https://doi.org/10.3390/s26123865

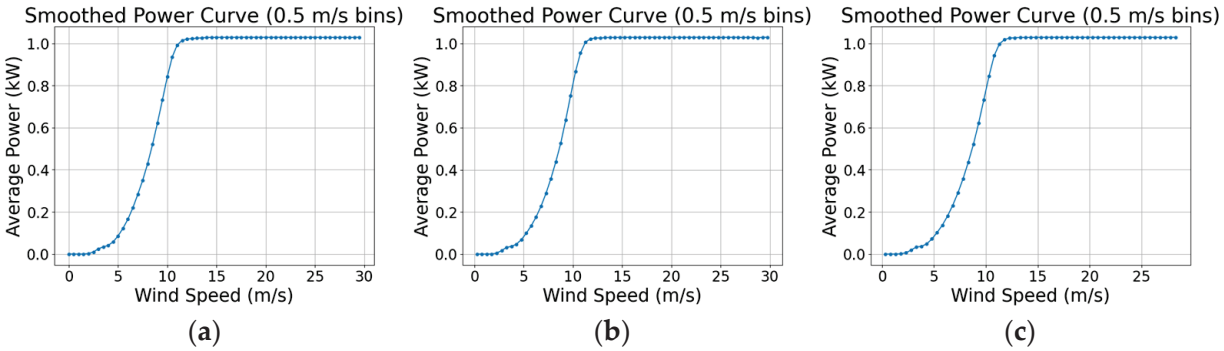


Figure 8. Power curve of the three turbines in the dataset: (**a**) turbine #27, (**b**) turbine #34, (**c**) turbine #77.

4.2.5. Selection of Unsupervised and Supervised Classifiers

For unsupervised classification, several models were evaluated for their ability to distinguish between normal and abnormal conditions. Among the unsupervised models examined (see Table 4), the One-Class Support Vector Machine (OCSVM) was selected because it provided greater discrimination between normal and fault samples. Similar to unsupervised models, NBM only uses normal data for analysis. For a typical NBM, a Linear Regression model in [23] was considered here as a baseline. This model was trained using only normal operating data to obtain the turbine power output, and the absolute prediction residual was used as the anomaly score. The top-ranked residual samples were then selected as abnormal candidates for fault coverage evaluation. For the neural-network models (Autoencoder, LSTM-AE, CNN, and the unsupervised graph representation learning model from [24]), the number of training epochs was fixed at 50, and the learning rate was set to 0.001 for consistency. OCSVM was configured with a radial basis function (RBF) kernel ($\nu = 0.05$) and did not require training epochs or a learning rate. All the models were tested with their best window size. These comparisons show that, except for OCSVM, the neural-network-based unsupervised models did not achieve better fault coverage of the dataset under examination here. This outcome suggests that increasing model complexity does not necessarily improve the Stage-1 fault screening performance.

Table 4. Performance of unsupervised models and normal behavior models on fault coverage rate.

Unsupervised Model or NBM	T27 Fault Coverage Rate (%)	T34 Fault Coverage Rate (%)	T77 Fault Coverage Rate (%)
OCSVM	100	100	100
LSTM-AE	73.7	35.7	35.7
Autoencoder	84.2	21.4	28.6
Linear Regression	55.7	37.7	70.0
Unsupervised graph learning	10.5	0	42.9

For supervised classification, several models were evaluated for their ability to distinguish filtered data between non-fault and fault conditions. As shown in Table 5, the Convolutional Neural Network (CNN) provided the fewest missed faults, and it was thus selected for the subsequent experimentations. All models were tested using their respective optimal window sizes. Recent approaches such as KNN and ConvLSTM-Transformer [25,26], which have shown promising performance in related applications, were also evaluated. However, they did not achieve superior performance. XGBoost missed fewer positive samples comparable to those of CNN, but CNN consistently missed fewer fault samples across the three

https://doi.org/10.3390/s26123865

turbines. These results suggest that CNN provided the best balance between fault detection capability and overall classification performance for the SCADA dataset examined.

Table 5. Performance of supervised models for fault detection.

Supervised Model	Missed Normal Samples (FP)			Missed Faults (FN)		
	T27	T34	T77	T27	T34	T77
CNN	1	4	5	0	1	0
XGboost	1	1	1	1	2	0
Random forest	0	0	0	0	2	2
LinearSVC	4	15	2	4	5	0
KNN	1	7	4	2	2	0
ConvLSTM + Transformer	2	2	6	0	3	0

Our two-stage detection thus consists of OCSVM as the unsupervised classifier in stage 1 and CNN as the supervised classifier in stage 2. This combination leverages the strength of unsupervised learning in identifying candidate anomalies and supervised learning in refining fault detection. Figure 9 shows the architecture of the OCSVM+CNN model.

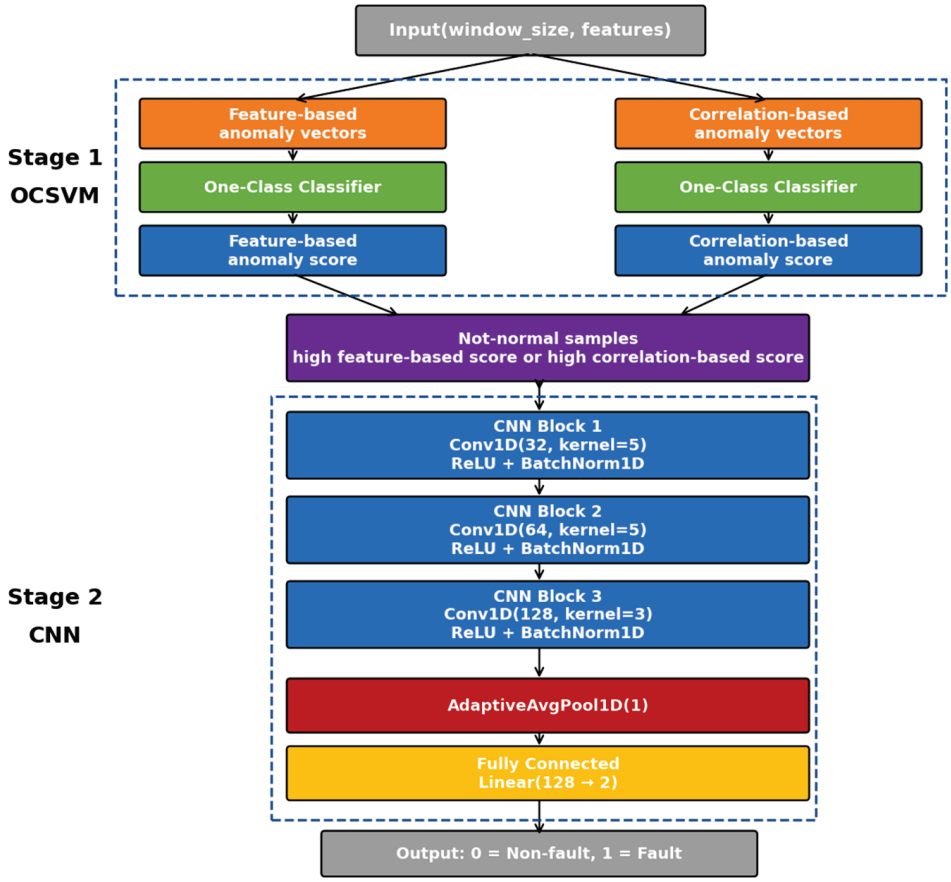


Figure 9. Architecture of the OCSVM+CNN model.

https://doi.org/10.3390/s26123865

4.3. Performance Metrics

4.3.1. Metric for Unsupervised Learning

For the unsupervised classification stage, performance is evaluated using the fault coverage rate, defined as the percentage overlap between detected abnormal samples and true fault samples.

$$Fault\ coverage\ rate = \frac{overlap\ between\ abnormal\ and\ fault\ samples}{total\ fault\ samples} \tag{10}$$

This metric measures the ability of the unsupervised anomaly detector or classifier in capturing fault events, regardless of fault type. A higher coverage rate indicates better detection of fault-including anomalies.

4.3.2. Metrics for Supervised Learning

For the supervised classification stage, the following metrics are considered:

$$Missed\ Normal\ Rate\ (MNR) = \frac{Missed\ normal\ samples}{Total\ normal\ samples} \tag{11}$$

This metric measures the proportion of normal samples incorrectly classified as faults (false positives).

$$Missed\ Fault\ Rate\ (MFR) = \frac{Missed\ fault\ samples}{Total\ fault\ samples} \tag{12}$$

This metric evaluates the proportion of fault samples that are not detected (false negatives), which is a critical measure for reliable wind turbine operation. In addition, the commonly used accuracy measure is considered. Accuracy is defined as

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{13}$$

where TP, FP, TN, and FN denote true positives, false positives, true negatives, and false negatives, respectively.

5. Results and Discussion

5.1. Combined Anomaly Score vs. Feature-Based Anomaly Score

This study compared the effectiveness of the feature-based anomaly score versus the combined anomaly score. Table 6 shows the outcome on the test data when using the combined anomaly scores introduced in this paper and the feature-based anomaly scores used in [27,28]. As shown in this table, the combined anomaly scores consistently achieved a fault coverage rate of 100%.

Table 6. Comparison of fault coverage between feature-based and combined anomaly scores.

Turbine	Method	Fault Coverage Rate
27	Feature-based anomaly score	100%
	Combined anomaly scores	100%
34	Feature-based anomaly score	98%
	Combined anomaly scores	100%
77	Feature-based anomaly score	91%
	Combined anomaly scores	100%

As shown in Figure 10, the yellow lines represent abnormal samples detected by the unsupervised classifier, the blue lines represent the faults identified by the supervised

https://doi.org/10.3390/s26123865

classifier, and the red lines indicate the fault samples. The fault samples completely overlap with the detected anomaly samples for all three turbines, indicating that the unsupervised classifier captured all the fault-related deviations. After processing these candidate samples, the CNN in Stage 2 identifies fault events that closely align with the ground-truth faults.

These results demonstrate that the combined anomaly score provides a more comprehensive representation of abnormal behavior and improves the robustness of fault detection. In addition, it indicates that the developed two-stage framework effectively suppresses false alarms while preserving fault-related information, enabling accurate fault detection with a significantly reduced search space.

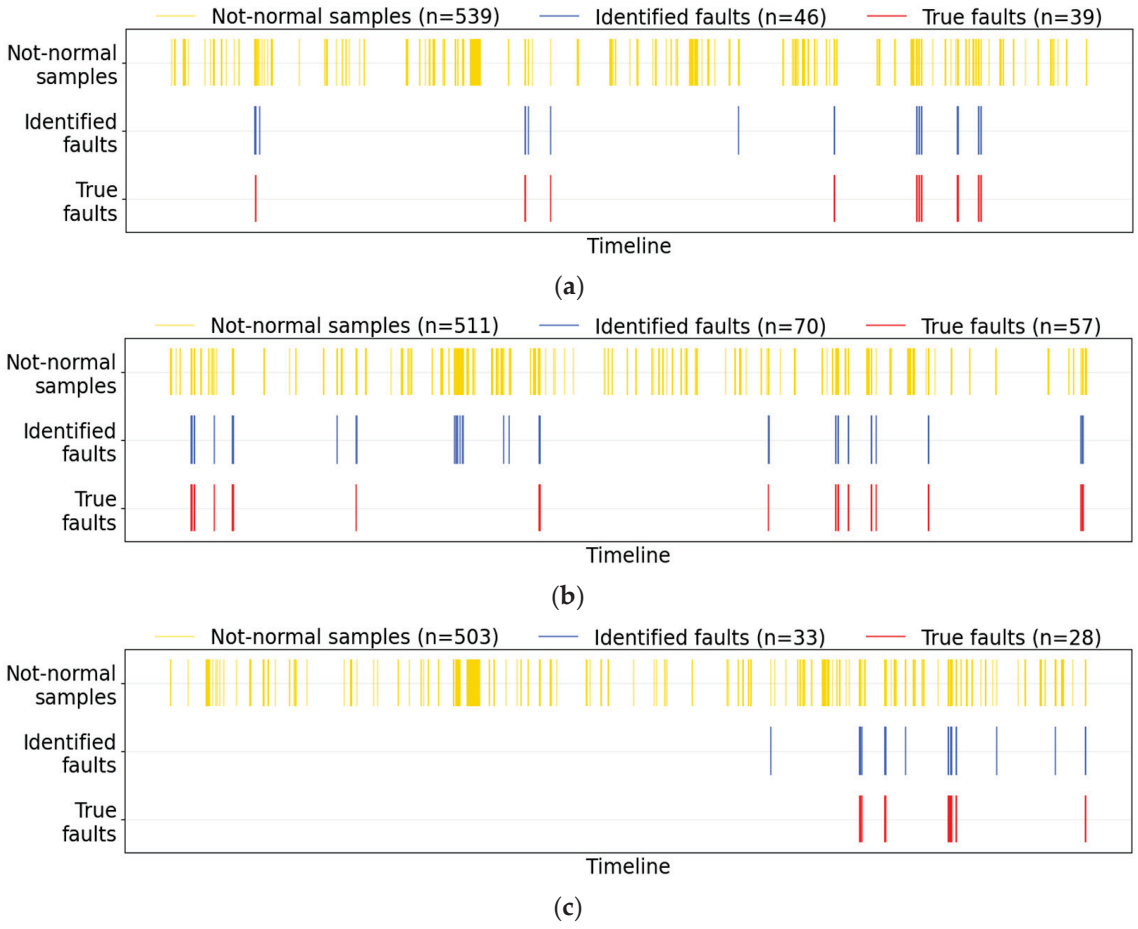


Figure 10. Temporal visualization of detected anomaly samples and fault samples for the three turbines of the dataset: (**a**) #27, (**b**) #34, and (**c**) #77.

5.2. Two-Stage Method vs. Supervised Method

This study evaluates the main contribution of the developed solution by comparing the two-stage method (unsupervised + supervised) with the supervised-only method. Due to severe class imbalance in SCADA data, the supervised model exhibits bias toward normal samples, leading to low fault-detection performance. In contrast, the two-stage method first applies the unsupervised OCSVM model to identify candidate anomaly samples, and then trains the CNN model only on these samples. As a result, the dominance of the normal samples is reduced, and the CNN model is trained to focus on abnormal samples of interest. It is noted that the Stage-1 OCSVM achieved 100% fault coverage on the testing data, and all the missed faults reported in Table 5 originate from the Stage-2 CNN classifier rather than the unsupervised screening stage.

Experimental results in Table 7 indicate that the supervised-only method achieved lower accuracy and suffered from higher missed normal/fault rates. The two-stage method

https://doi.org/10.3390/s26123865

significantly improved the fault detection performance, achieving lower missed fault rates and missed normal rate, and higher accuracy.

Table 7. Fault detection performance using the supervised method vs. the two-stage method.

Turbine	Method	Missed Normal/ Total Normal	Missed Faults/ Total Faults	Overall Accuracy
27	CNN	70/10,467	4/39	99.3%
	CNN + OCSVM	9/10,467	2/39	99.9%
34	CNN	193/10,449	7/57	98.1%
	CNN + OCSVM	13/10,449	0/57	99.9%
77	CNN	64/10,478	2/28	99.4%
	CNN + OCSVM	7/10,478	2/28	99.9%

After training, the inference of the combined OCSVM-CNN is less than 100 ms per sample on a typical CPU, which is much shorter compared with the 10-min sampling interval of typical SCADA datasets. Even for SCADA datasets that have a 1-min sampling rate, the additional computational cost of 100 ms does not impact the real-time deployment of the framework. In other words, the two-stage framework provides a practical online condition monitoring in utility-scale wind farms. Overall, this study highlights that incorporating an unsupervised stage before a supervised stage allows handling class imbalance and improves the effectiveness of supervised learning in fault detection.

5.3. Transferability Across Datasets from Similar Turbines

In this study, the transferability of the two-stage method across different turbines within the same wind farm (same make/model) was examined. Specifically, fault data from one turbine was considered with normal data from another turbine, and the above experiments were repeated. Since the turbines are of the same make/model and operate under similar conditions, their normal behavior is expected to produce similar normal SCADA data as mentioned in [29]. To evaluate the transferability of the proposed framework, fault samples from one turbine were combined with normal-operation samples from different turbines of the same make and model. For example, fault windows from turbine 27 were paired with normal windows from turbines 27, 34, and 77. This setup enables the investigation of whether fault characteristics learned from one turbine can be generalized to other turbines operating under similar conditions. Figure 11 provides an illustration of this transferability. The results of this study are shown in Table 8.

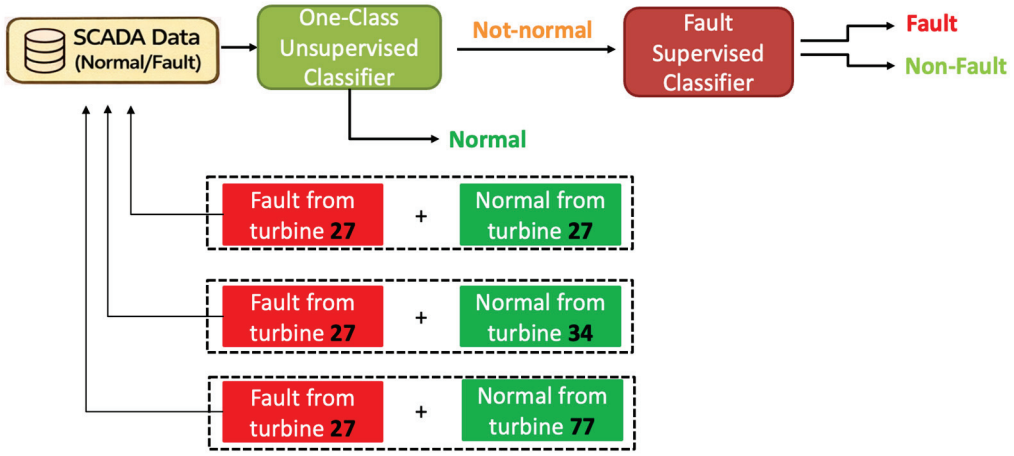


Figure 11. Illustration of cross-turbine transferability of normal data (faults from turbine #27 and normal data from the other turbines).

https://doi.org/10.3390/s26123865

Table 8. Performance comparison using normal data from other turbines.

Turbine	Normal Data Source	Missed Normal/ Total Normal	Missed Faults/ Total Faults	Overall Accuracy
27	27	9/10,467	2/39	99.9%
	34	16/10,467	5/39	99.8%
	77	16/10,467	3/39	99.8%
34	27	8/10,449	6/57	99.9%
	34	13/10,449	0/57	99.9%
	77	12/10,449	4/57	99.8%
77	27	32/10,478	5/28	99.6%
	34	15/10,478	6/28	99.8%
	77	7/10,478	2/28	99.9%

Although using normal data from other turbines leads to a slight reduction in accuracy compared to using data from the same turbine, the fault-detection capability remains high. In other words, the two-stage method remained effective under this cross-turbine setting, indicating the transferability of the fault detection framework when trained using normal data from similar make/model turbines. This finding has several important practical implications. First, if normal data from a specific turbine is unavailable, normal data from other turbines can be used without significantly affecting performance. Second, by using fault data from multiple turbines, it is possible to build a more generalizable fault detection classifier, reducing the need for turbine-specific training and lowering computational cost. Third, the unsupervised classifier trained solely on normal data can generate reliable abnormal alarm signals that serve as useful references even when fault labels are incomplete or unavailable.

Although the developed fault detection framework has demonstrated promising results, it is worth stating here its limitations. First, the anomaly selection threshold study and the transferability study were conducted using turbines of the same make and model. The optimal threshold and transferability may vary across different turbines and operating environments. In addition, while the fault detection framework effectively reduces the impact of class imbalance, fault types with limited samples would still remain a challenging task for supervised classifiers.

6. Conclusions

This paper has presented a two-stage method for detecting faults based on SCADA data. This method combines unsupervised anomaly detection or classification and supervised classification to address the problems of a lack of labeled data and severe class imbalance. In the first stage, the unsupervised classifier of OCSVM, together with the two feature-based and correlation-based anomaly scores, is used to identify abnormalities based on only normal SCADA data. In the second stage, the supervised classifier of CNN is applied to abnormal samples to identify fault samples among them. Experimental results show that the introduced two-stage detection method outperforms supervised-only methods, achieving higher accuracy and lower missed fault rates. Cross-turbine experiments further demonstrate that normal behavior learned from one turbine can be transferred to other turbines. Overall, the developed two-stage method provides improved wind turbine fault detection based on SCADA data.

https://doi.org/10.3390/s26123865

Author Contributions: Conceptualization, J.D., M.R. and N.K.; methodology, J.D. and N.K.; software, J.D.; investigation, J.D. and N.K.; resources, M.R. and N.K.; data curation, J.D., M.R. and N.K.; writing, J.D., M.R. and N.K.; supervision, M.R. and N.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Science Foundation under Award Nos. 1916715 and 1916776 (I/UCRC for Wind Energy, Science, Technology, and Research) and by the members of WindSTAR I/UCRC. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the sponsors.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used in this study is publicly available through the Zenodo repository stated in Reference [21].

Acknowledgments: The authors would like to acknowledge the industrial advisory members of the WindSTAR, NSF, and IUCRC for their feedback.

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

References

1. Ember. Available online: https://ember-climate.org/topics/wind/ (accessed on 16 June 2022).
2. Dai, J.; Rotea, M.; Kehtarnavaz, N. Obtaining Rotational Stiffness of Wind Turbine Foundation from Acceleration and Wind Speed SCADA Data. *Sensors* **2025**, *25*, 4756. [CrossRef] [PubMed]
3. Singh, U.; Rizwan, M. Analysis of Wind Turbine Dataset and Machine Learning Based Forecasting in SCADA-System. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 8035–8044. [CrossRef]
4. Singh, U.; Rizwan, M. SCADA System Dataset Exploration and Machine Learning Based Forecast for Wind Turbines. *Results Eng.* **2022**, *16*, 100640. [CrossRef]
5. Schlechtingen, M.; Santos, I.F.; Achiche, S. Wind Turbine Condition Monitoring Based on SCADA Data Using Normal Behavior Models. Part 1: System Description. *Appl. Soft Comput.* **2013**, *13*, 259–270. [CrossRef]
6. Meyer, A. Multi-Target Normal Behaviour Models for Wind Farm Condition Monitoring. *Appl. Energy* **2021**, *300*, 117342. [CrossRef]
7. Vidal, Y.; Pozo, F.; Tutivén, C. Wind Turbine Multi-Fault Detection and Classification Based on SCADA Data. *Energies* **2018**, *11*, 3018. [CrossRef]
8. Li, G.; Wang, C.; Zhang, D.; Yang, G. An Improved Feature Selection Method Based on Random Forest Algorithm for Wind Turbine Condition Monitoring. *Sensors* **2021**, *21*, 5654. [CrossRef] [PubMed]
9. Trizoglou, P.; Liu, X.; Lin, Z. Fault Detection by an Ensemble Framework of Extreme Gradient Boosting (XGBoost) in the Operation of Offshore Wind Turbines. *Renew. Energy* **2021**, *179*, 945–962. [CrossRef]
10. Xiang, L.; Wang, P.; Yang, X.; Hu, A.; Su, H. Fault Detection of Wind Turbine Based on SCADA Data Analysis Using CNN and LSTM with Attention Mechanism. *Measurement* **2021**, *175*, 109094. [CrossRef]
11. Rama, V.S.B.; Hur, S.-H.; Yang, J.-M. Short-Term Fault Prediction of Wind Turbines Based on Integrated RNN-LSTM. *IEEE Access* **2024**, *12*, 22465–22478. [CrossRef]
12. Delgado, I.; Fahim, M. Wind Turbine Data Analysis and LSTM-Based Prediction in SCADA System. *Energies* **2021**, *14*, 125. [CrossRef]
13. Liu, X.; Du, J.; Ye, Z.-S. A Condition Monitoring and Fault Isolation System for Wind Turbine Based on SCADA Data. *IEEE Trans. Ind. Inform.* **2022**, *18*, 986–995. [CrossRef]
14. Jin, X.; Pan, H.; Ying, C.; Kong, Z.; Xu, Z.; Zhang, B. Condition Monitoring of Wind Turbine Generator Based on Transfer Learning and One-Class Classifier. *IEEE Sens. J.* **2022**, *22*, 24130–24139. [CrossRef]
15. Tutivén, C.; Vidal, Y.; Insuasty, A.; Campoverde-Vilela, L.; Achicanoy, W. Early Fault Diagnosis Strategy for WT Main Bearings Based on SCADA Data and One-Class SVM. *Energies* **2022**, *15*, 4381. [CrossRef]
16. McKinnon, C.; Carroll, J.; McDonald, A.; Koukoura, S.; Infield, D.; Soraghan, C. Comparison of New Anomaly Detection Technique for Wind Turbine Condition Monitoring Using Gearbox SCADA Data. *Energies* **2020**, *13*, 5152. [CrossRef]
17. Ferreira, V.Z.; Brito, E.M.; Da Cunha, G.M.F.; Perez, F.A.R.; Dos Reis, E.D.W.; Pereira, E.R.; Louzada, F. Unsupervised Autoencoder-Based Anomaly Detection Under Limited Failure Data for Oil Industry. *IEEE Open J. Instrum. Meas.* **2026**, *5*, 2500112. [CrossRef]

https://doi.org/10.3390/s26123865

18. Chen, H.; Liu, H.; Chu, X.; Liu, Q.; Xue, D. Anomaly Detection and Critical SCADA Parameters Identification for Wind Turbines Based on LSTM-AE Neural Network. *Renew. Energy* **2021**, *172*, 829–840. [CrossRef]
19. Seifert, J.K.; Kraft, M.; Kühn, M.; Lukassen, L.J. Correlations of Power Output Fluctuations in an Offshore Wind Farm Using High-Resolution SCADA Data. *Wind. Energy Sci.* **2021**, *6*, 997–1014. [CrossRef]
20. Maldonado-Correa, J.; Valdiviezo-Condolo, M.; Artigao, E.; Martín-Martínez, S.; Gómez-Lázaro, E. Classification of Highly Imbalanced Supervisory Control and Data Acquisition Data for Fault Detection of Wind Turbine Generators. *Energies* **2024**, *17*, 1590. [CrossRef]
21. NIAID Data Discovery Portal. Wind Turbine SCADA Data for Early Fault Detection. Available online: https://data.niaid.nih.gov/resources?id=zenodo_10958774 (accessed on 15 March 2026).
22. Zhang, F.; Wen, Z.; Liu, D.; Jiao, J.; Wan, H.; Zeng, B. Calculation and Analysis of Wind Turbine Health Monitoring Indicators Based on the Relationships with SCADA Data. *Appl. Sci.* **2020**, *10*, 410. [CrossRef]
23. Dao, P.B. Condition Monitoring and Fault Diagnosis of Wind Turbines Based on Structural Break Detection in SCADA Data. *Renew. Energy* **2022**, *185*, 641–654. [CrossRef]
24. Zhu, Y.; Xie, B.; Wang, A.; Qian, Z. Wind Turbine Fault Detection and Identification via Self-Attention-Based Dynamic Graph Representation Learning and Variable-Level Normalizing Flow. *Reliab. Eng. Syst. Saf.* **2025**, *253*, 110554. [CrossRef]
25. Fazli, A.; Poshtan, J. Wind Turbine Fault Detection and Isolation Robust against Data Imbalance Using KNN. *Energy Sci. Eng.* **2024**, *12*, 1174–1186. [CrossRef]
26. Guo, J.; Song, X.; Tang, S.; Zhang, Y.; Wu, J.; Li, Y.; Jia, Y.; Cai, C.; Li, Q. Fault Diagnosis of Wind Turbine Blade Icing Based on Feature Engineering and the PSO-ConvLSTM-Transformer. *Ocean. Eng.* **2024**, *302*, 117726. [CrossRef]
27. Preciado-Grijalva, A.; Iza-Teran, V.R. Anomaly Detection of Wind Turbine Time Series Using Variational Recurrent Autoencoders. *arXiv* **2021**. [CrossRef]
28. Vásquez-Rodríguez, G.; Maldonado-Correa, J. Anomaly-Based Fault Detection in Wind Turbines Using Unsupervised Learning: A Comparative Study. *IOP Conf. Ser. Earth Environ. Sci.* **2024**, *1370*, 012005. [CrossRef]
29. Ulmer, M.; Jarlskog, E.; Pizza, G.; Huber, L.G. Cross-Turbine Training of Convolutional Neural Networks for SCADA-Based Fault Detection in Wind Turbines. *Annu. Conf. PHM Soc.* **2020**, *12*, 10. [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.

https://doi.org/10.3390/s26123865

sensors

MDPI

Article

Multi-Component Joint Maintenance Decision for Electro-Hydraulic Servo Fatigue Testing Machine Based on Multi-Head Deep Reinforcement Learning

Peng Liu †, Guotai Huang †, Jialu Xi and Jiaqi Wu *

School of Mechanical and Aerospace Engineering, Jilin University, Changchun 130022, China
* Correspondence: wujq@jlu.edu.cn
† These authors contributed equally to this work.

Abstract

To address the challenge of maintenance decision-making for critical components in electro-hydraulic servo material fatigue testing machine, characterized by weak state observability and difficulty in degradation prediction, a multi-component joint maintenance decision-making method based on multi-head deep reinforcement learning is proposed. Considering the heterogeneity of the degradation mechanisms and observation methods for the four components—bearing beam, fixture, main machine sensors, and hydraulic oil tank—a continuous-discrete hybrid state Markov decision process (HS-MDP) is constructed. To account for differences in maintenance strategies across components, a differentiated discrete action space for each component is designed, and engineering feasibility constraints are explicitly integrated into the policy through action masking. A data-quality loss term, determined by the degradation level of the sensors, is introduced into the reward function to align the optimization objective with the metrological properties of the fatigue testing machine. Based on the Branching Dueling DQN framework, a Q-network structure is constructed, incorporating a shared encoder, an inter-component attention mechanism, and multi-head branched outputs. Taking a 100 kN electro-hydraulic servo fatigue testing machine as a case study, comparisons with baseline strategies such as periodic maintenance, threshold-based condition-based maintenance (CBM), independent DQN, and PPO indicate that the proposed method reduces the average annual total cost by 60.3% compared to periodic maintenance and by 42.6% compared to threshold-based CBM. The number of failures decreases from 9.8 times/year to 1.4 times/year, while data efficiency increases from 82.1% to 96.2%. Ablation experiments and robustness tests further verify the critical contributions of three key design elements: action masking, inter-component attention, and data-quality loss.

Keywords: electro-hydraulic servo fatigue testing machine; multi-component joint maintenance; deep reinforcement learning; hybrid-state Markov decision process; opportunistic maintenance

1. Introduction

Intelligent operation and maintenance of electro-hydraulic servo fatigue testing equipment is pivotal for ensuring the continuity of material performance evaluation tasks. The coexistence of heterogeneous degradation characteristics across the internal load-bearing structures, actuators, and measurement-and-control units makes traditional single-component maintenance models struggle to balance high reliability with economic efficiency. Coordinated maintenance of multi-component systems involves not only the

independent physical degradation processes of each component, but it is also strongly influenced by the "opportunistic maintenance" effect induced by economic interdependencies among components.

This section first reviews the evolution of multi-component degradation modeling theory, examining the use of continuous stochastic processes and discrete state-transition models to describe heterogeneous failure mechanisms in complex mechanical systems. It then focuses on recent advances in deep reinforcement learning (DRL) for handling large-scale action spaces, embedding engineering logic constraints, and optimizing multi-objective cost functions. Finally, drawing on architectures such as branching networks and attention mechanisms, this section provides the theoretical underpinning for the joint maintenance decision framework proposed in this study, which is based on a Hybrid-State MDP (HS-MDP) and a multi-head branching Q-network (BDQ).

1. Degradation modeling of complex systems and multi-component maintenance logic.
   Degradation modeling of multi-component systems is the physical foundation for maintenance decision-making. In the maintenance management domain, de Jonge et al. conducted a systematic review of maintenance optimization models, emphasizing the impact of modeling accuracy on the effectiveness of decisions [1]. Nicolai et al. further elucidated the importance of economic interdependencies in multi-component systems, providing an early theoretical framework for "opportunistic maintenance" [2]. For structural components in fatigue testing equipment—such as bearing beams and fixture—that exhibit continuous cumulative damage, Thomas noted that maintenance of multi-unit systems should account for stochastic evolution processes [3]. Saha et al. systematically surveyed the application of Wiener processes in degradation prediction for highly reliable products [4], while Pan et al. explored multi-performance-index reliability modeling based on Gamma processes [5].
   However, the states of sensors and the hydraulic oil system in testing machines often exhibit discrete jumps as revealed by periodic calibration or laboratory analysis. Keizer et al. effectively modeled discrete degradation states using a partially observable Markov decision process (POMDP) [6]. In dealing with such complex systems, Si et al. reviewed data-driven techniques for remaining useful life (RUL) prediction [7]. Grall et al. pointed out that optimal maintenance policies must balance the benefit of a single repair against the fixed cost of a system-wide shutdown [8]. Castanier et al. further explored combining continuous degradation dynamics with discrete decision logic, laying the mathematical groundwork for continuous–discrete hybrid-state modeling [9].
2. Deep reinforcement learning for maintenance decision optimization.
   In the face of the exponential growth of the state space in multi-component systems, deep reinforcement learning (DRL) offers a model-free pathway for decision optimization in high-dimensional environments [10]. Zhang and Si were among the first to introduce DRL into condition-based maintenance (CBM) planning for multi-component systems, achieving an end-to-end mapping from multi-source monitoring signals to optimal maintenance actions [11]. To capture complex inter-component interactions, Do et al. and Chen et al. explored structure-dependent multi-component maintenance policies, demonstrating DRL's strong capability in handling economic interdependencies [12,13]. However, as precision metrological equipment, fatigue testing machines face a unique challenge: sensor degradation can render test data out of tolerance, which are thus scrapped. This distinctive "data-quality loss" is rarely represented in general maintenance studies, making algorithms difficult to directly adapt to the metrological attributes of fatigue testing equipment.

https://doi.org/10.3390/s26134087

3. Action-space decoupling, engineering masks, and coordination mechanisms.
   When maintenance decisions involve multiple heterogeneous components, the size of the joint action space grows exponentially. To address this challenge, the Dueling DQN architecture proposed by Wang et al. decouples the state value and action advantage [14]; building on this, the branching multi-head architecture (Branching BDQ) proposed by Tavakoli et al. greatly reduces sample complexity in high-dimensional discrete-action environments by decoupling the output heads [15]. Meanwhile, real maintenance operations must adhere to strict engineering constraints. De Giacomo et al. articulated the key role of action masking in ensuring logical compliance in reinforcement learning [16], and Huang et al. demonstrated that explicitly injecting state masks can significantly shrink the search space and accelerate policy convergence [17].
   In addition, modeling inter-component interactions has become crucial for improving collaborative maintenance efficiency. With the widespread adoption of attention mechanisms in feature engineering, researchers have introduced them into the PHM field to capture cross-component degradation synergies [18]. For example, Bai et al. combined attention mechanisms with DRL and improved maintenance decision accuracy by adaptively fusing multi-scale information [19]. Studies by Vu et al. and Olde Keizer et al. also show that modeling economic synergies among components can significantly enhance overall system availability [20,21]. These advanced algorithmic architectures provide the core tools in this study for injecting action masks and enabling cross-component coordinated decision-making in complex maintenance scenarios of fatigue testing equipment.

In summary, substantial progress has been made in multi-component maintenance modeling and reinforcement learning algorithm optimization. Nevertheless, existing research mostly focuses on homogeneous-component systems or simplified single-degradation models and lacks in-depth analysis of fatigue testing equipment, which integrates continuous degradation of precision mechanical structures with discrete deterioration of measurement-and-control systems—i.e., a continuous–discrete hybrid-state system. Consequently, for test machines exhibiting multi-source heterogeneous degradation, how to achieve intelligent decision-making that also accounts for metrological data quality under the dual constraints of an exploding action space and engineering feasibility remains largely unexplored. There is an urgent need to develop a multi-head branching deep reinforcement learning architecture that can explicitly handle action-masking constraints and inter-component attention dependencies.

Contributions: It should be stated at the outset that the algorithmic primitives employed in this work—the branching value decomposition, the dueling architecture, the scaled dot-product attention, and logit-level action masking—are adopted from the literature rather than newly proposed. The novelty of this study lies instead in the problem it formalizes, in one domain-specific modeling element, and in an integration whose components are shown to be individually necessary:

(1) A maintenance-decision formulation for a previously underexplored equipment class. Unlike production machinery, an electro-hydraulic servo fatigue testing machine is itself a metrological instrument: its load-bearing structures degrade continuously, whereas its measurement-and-control units degrade in discrete, calibration-revealed jumps. We cast the joint maintenance of these heterogeneous subsystems as a continuous–discrete hybrid-state MDP (HS-MDP), in which Wiener processes and operating-condition-modulated Markov chains are coupled through a shared condition variable. To the best of our knowledge, the coupling of continuous precision-

https://doi.org/10.3390/s26134087

mechanical degradation with discrete measurement-system degradation has not previously been formalized as a single maintenance-decision problem.

(2) A data-quality loss term that aligns the policy with the metrological function of the equipment. Because the deliverable of a testing machine is valid measurement data, sensor degradation silently drives test results out of tolerance and causes data to be scrapped. We introduce a data-quality cost $c_{\text{qual}}\left(z_t^{(3)}\right)$ that increases stepwise with the sensor degradation level—an objective dimension largely absent from general maintenance studies. Its effect on the learned policy is isolated in the ablation study, where it is shown to reshape sensor-maintenance behavior rather than merely re-label an existing cost.

(3) A purpose-built integration whose elements are empirically necessary. The hard/soft engineering action masks, the inter-component attention for opportunistic coordination, and the multi-head dueling heads are combined so that each component decision is simultaneously feasibility-constrained, condition-aware, and aware of the other components' states. The contribution of each element, and the physical interpretability of the learned attention weights, are examined in the ablation and attention analyses.

2. Problem Modeling and Hybrid-State Markov Decision Process

2.1. Maintenance Decision Scenario

During the long-term operation of an electro-hydraulic servo material fatigue testing machine, four critical components—the bearing beam, fixture, main force/displacement sensor, and hydraulic oil tank—undergo degradation processes driven by different mechanisms: The bearing beam experiences a gradual deviation in horizontality under prolonged eccentric and high-frequency alternating loads, exhibiting a continuous accumulation characteristic. The fixture undergoes contact wear due to repeated clamping and loading, with the wear amount similarly exhibiting continuous accumulation. The zero drift and sensitivity of the main sensors degrade over time and manifest as discrete levels observable through periodic calibration. The contamination level of hydraulic oil, following standards such as NAS 1638, intrinsically falls into discrete levels. At each decision time step t, the maintenance decision agent outputs a joint action $a_t = \left(a_t{}^1, a_t{}^2, a_t{}^3, a_t{}^4\right)$, where a_t^i represents the action chosen for the i-th component. After executing the action, the system transitions to a new state s_{t+1} and incurs an immediate cost c_t. The objective of the decision-making process is to find a policy $\pi : S \to A$ that minimizes the expected long-term discounted cumulative cost:

$$\pi^* = \arg\min_{\pi} \mathbb{E}_{\pi}\left[\sum_{t=0}^{T} \gamma^t c_t\right] \tag{1}$$

where $\gamma \in (0, 1)$ is the discount factor, and T represents the decision time horizon.

2.2. Hybrid-State Markov Decision Process

Considering the heterogeneity in the degradation mechanisms and observation methods of the four components, this paper adopts the Hybrid-State Markov Decision Process (HS-MDP) for modeling. HS-MDP is defined as a quintuple (S, A, P, R, γ), where the state space is the Cartesian product of the continuous subspace and the discrete subspace:

$$\mathcal{S} = \mathcal{X} \times \mathcal{Z} \subseteq \mathbb{R}^{d_c} \times \prod_j \{1, \ldots, K_j\} \tag{2}$$

The subspace $\mathcal{X}$ accommodates the degradation variables of the bearing beam and fixture, while the discrete subspace $\mathcal{Z}$ includes the degradation levels of the sensors and hydraulic

https://doi.org/10.3390/s26134087

oil tank. The action space A is the Cartesian product of the action subspaces of the four components. The transition function $P(s_{t+1} \mid s_t, a_t)$ propagates the continuous components using stochastic differential equations and the discrete components based on a state transition matrix, with both coupled through the shared operating condition variable u_t. The reward function $R(s_t, a_t)$ integrates maintenance costs, downtime costs, failure costs, and data-quality loss. This modeling approach not only ensures mathematical rigor but also considers the engineering relevance of the physical degradation characteristics of different components. The overall technical framework is illustrated in Figure 1.

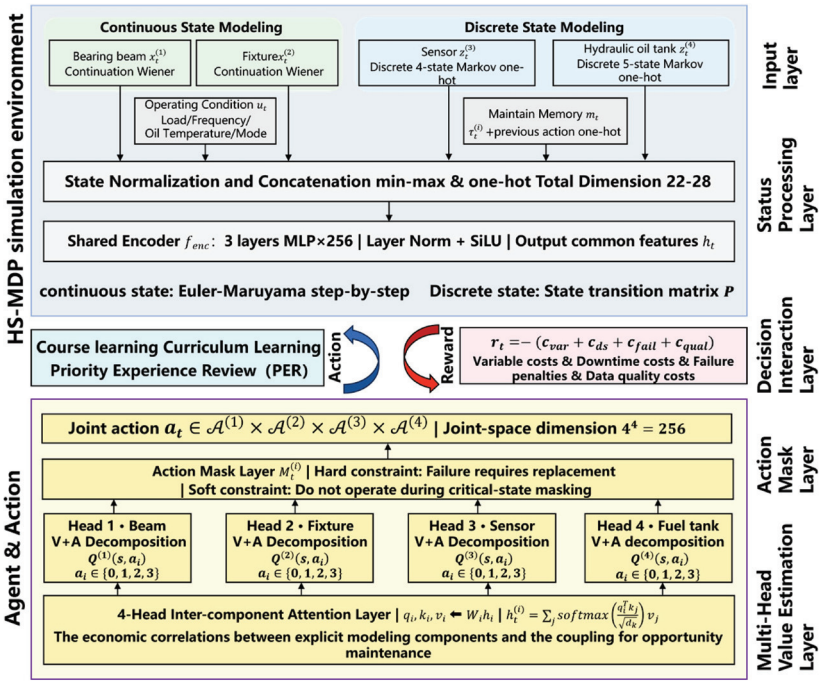


Figure 1. Maintenance Decision-Making Framework Based on Multi-Head Deep Reinforcement Learning.

3. Design of Hybrid State Space

3.1. Overall Structure of the State Vector

Define the system state vector at time t as:

$$s_t = \left[\mathrm{x}_t^{(1)}, \mathrm{x}_t^{(2)}, \mathrm{z}_t^{(3)}, \mathrm{z}_t^{(4)}, \mathrm{u}_t, \mathrm{m}_t\right] \tag{3}$$

where x_t^1 and x_t^2 are the continuous state sub-vectors for the bearing beam and fixture, respectively; z_t^3 and z_t^4 are the discrete state one-hot encodings for the main machine sensors and hydraulic oil tank, respectively; u_t is the global operating condition vector; and m_t is the maintenance memory vector.

3.2. Continuous State Modeling: Bearing Beam and Fixture

For the bearing beam, the core degradation variable is the horizontality deviation h_t. Define the composite horizontality deviation as $h_t = \sqrt{(h_t^x)^2 + \left(h_t^y\right)^2}$, and model it using a drift–diffusion Wiener process:

$$dh_t = \mu_h(\mathrm{u}_t, e_t)dt + \sigma_h dB_t \tag{4}$$

https://doi.org/10.3390/s26134087

where B_t is a standard Brownian motion and σ_h is the diffusion coefficient. The drift term adopts a multiplicative structure to reflect the coupling between operating conditions and degradation:

$$\mu_h(u_t, e_t) = \mu_{h,0} \cdot \left(\frac{L_t}{L_0}\right)^{p_1} \cdot \left(\frac{f_t}{f_0}\right)^{p_2} \cdot (1 + \lambda e_t) \tag{5}$$

where L_t and f_t are the current load amplitude and loading frequency, e_t is the eccentric loading coefficient, and p_1, p_2 and λ are coupling parameters. This structure enables the policy to learn the decision logic of preemptive leveling under heavy-load, high-frequency, and eccentric loading conditions. The continuous state sub-vector for the bearing beam is defined as $x_t^1 = (h_t, \dot{h}_t, D_t^1, e_t) \in \mathbb{R}^4$, where $\dot{h}_t$ is the drift-rate estimate over a recent window and $D_t^1 = \int_0^t (L\tau/L_0)^{p_1} d\tau$ is the cumulative load equivalent. The failure criterion is $h_t \geq h\_crit$.

For the fixture, the degradation variable is the contact wear w_t, which is likewise modeled using a Wiener process:

$$dw_t = \mu_w(u_t)dt + \sigma_w dB_t \tag{6}$$

where B_t is a standard Brownian motion, the drift coefficient $\mu_w(u_t)$ is related to the load and the number of clamping operations, and σ_w is the diffusion coefficient. The continuous state sub-vector is defined as $x_t^2 = (W_t, \dot{w}_t, n_t^2) \in \mathbb{R}^3$, where n_t^2 is the cumulative number of clamping operations (log-normalized). The failure criterion is $w_t \geq w_{crit}$.

The drift–diffusion Wiener process is adopted for the bearing beam and the fixture for three reasons. First, both the horizontality deviation and the contact wear are cumulative quantities whose one-step increment is governed by the current load condition rather than the full degradation history, so an independent-increment (Markov) model is appropriate. Second, the additive diffusion term represents the measurement noise of the horizontality/wear sensing together with the short-term stochasticity of the load profile, while the operating-condition-dependent drift (Equation (5)) lets the mean degradation rate respond to load amplitude, loading frequency, and eccentricity. Third, the Wiener process admits a closed-form first-passage-time (inverse-Gaussian) distribution to the failure threshold, which is convenient for reliability evaluation. Over the service horizon considered here, the drift dominates the diffusion, so the simulated trajectories are effectively monotone. Strictly monotone alternatives such as the Gamma or inverse-Gaussian process could be substituted without altering the HS-MDP or the decision algorithm.

In the simulation, the two continuous states are advanced in discrete time using the Euler–Maruyama scheme.

3.3. Discrete State Modeling: Sensors and Hydraulic Oil Tank

A discrete-state Markov chain is used for the sensors and the oil tank because both degradation variables are intrinsically ordinal and are observed only at discrete inspection epochs: sensor health is graded into levels by periodic calibration of zero drift and sensitivity, and oil cleanliness is graded into NAS 1638 contamination classes. A continuous state would therefore be neither observable nor physically meaningful for these two subsystems. The memoryless assumption—that the probability of advancing to the next level depends only on the current level and the prevailing operating condition—is consistent with the level-to-level structure of the calibration and oil-analysis record. Degradation is taken to be monotone in the absence of maintenance, which yields the upper-triangular transition matrices with an absorbing failure state; the dependence on operating conditions is captured through a stress-acceleration factor on loading frequency for the sensors (Equation (7)) and an Arrhenius-type factor on oil temperature for the oil tank (Equation (8)).

https://doi.org/10.3390/s26134087

The degradation of the main machine sensors primarily manifests as zero drift and sensitivity loss, which are detected through periodic calibration. A four-level discrete state is adopted for the sensors, with $K_3 = 4$ and ${z_t}^3 \in \{1, 2, 3, 4\}$. The hydraulic oil quality in the oil tank is classified according to NAS 1638 contamination levels, and a five-level discrete degradation state is used, with $K_4 = 5$ and $z_t^{(4)} \in \{1, 2, 3, 4, 5\}$. The meanings of each level are provided in Table 1.

Table 1. Definition of Discrete Degradation Levels for Sensors and the Hydraulic Oil Tank.

Component	Level	Meaning	Physical Correspondence
Sensor	1	Healthy	zero drift < 0.1% FS; sensitivity deviation < 0.2%
	2	Mild degradation	zero drift 0.1–0.3% FS
	3	Moderate degradation	zero drift 0.3–0.5% FS; affects data quality
	4	Failure	zero drift > 0.5% FS or sensitivity out of tolerance
Hydraulic Oil Tank	1	Excellent	≤NAS 6
	2	Normal	NAS 7–8
	3	Caution	NAS 9–10
	4	Severe contamination	NAS 11–12
	5	Replacement required	>NAS 12

Sensor degradation follows a Markov chain, with the state transition matrix $P^3(u_t) \in \mathbb{R}(4 \times 4)$ taking an upper-triangular form (degradation is monotonic and irreversible in the absence of maintenance), and state 4 is absorbing. The dependence on operating conditions is reflected by making the transition probabilities depend on the loading frequency:

$$p_{i,i+1}^{(3)}(\mathrm{u}_t) = p_{i,i+1}^{(3),0} \cdot \exp(\beta \cdot f_t) \tag{7}$$

The hydraulic oil tank state transition matrix $P^4(\theta_t)$ likewise has an upper-triangular structure and depends on the oil temperature θ_t:

$$p_{i,i+1}^{(4)}(\theta_t) = p_{i,i+1}^{(4),0} \cdot \exp(\alpha(\theta_t - \theta_0)) \tag{8}$$

The higher the oil temperature, the faster the oil oxidizes, and the transition probabilities increase. The discrete states are one-hot encoded when fed into the network: $onehot({z_t}^3) \in \{0,1\}^4$, $onehot({z_t}^4) \in \{0,1\}^5$.

3.4. Global Operating Conditions and Maintenance Memory

Operating conditions have a significant impact on the degradation rates of each component and must be included in the state: $\mathrm{u}_t = [L_t, f_t, \theta_t^{\mathrm{env}}, \mathrm{Mode}_t]$, where L_t is the normalized load amplitude level, θ_t^{env} is the ambient temperature, and $\mathrm{Mode_t}$ is a one-hot encoding of the test mode (e.g., tensile fatigue, bending fatigue, load-holding). Maintenance memory vector $\mathrm{m}_t = \left[\tau_t^{(1)}, \tau_t^{(2)}, \tau_t^{(3)}, \tau_t^{(4)}, \mathrm{LastAct}_t\right]$, where τ_t^{i} denotes the operating time since the last maintenance of component i, used to distinguish between "a newly replaced component" and "an old component whose degradation happens to be similar." After min–max normalization of continuous components and one-hot encoding of discrete components, they are concatenated to form the complete state vector.

4. Multi-Head Joint Action Space and Reward Function

4.1. Component-Differentiated Action Space

An exclusive action set is designed for each component based on its physical maintenance measures, uniformly covering a four-tier maintenance hierarchy: no action—

https://doi.org/10.3390/s26134087

inspection/prevention—minor repair—replacement. At time t, the agent outputs a joint action $a_t = (a_t^1, a_t^2, a_t^3, a_t^4) \in A_1 \times A_2 \times A_3 \times A_4$, and the joint action space has a cardinality of $4^4 = 256$. Through multi-head decomposition, each output head makes decisions only within a size-4 action space, significantly reducing sample complexity. The action designs for each component are shown in Table 2.

Table 2. Component-Differentiated Action Space Design.

Component	Action 0	Action 1	Action 2	Action 3
Bearing Beam	No action (natural degradation)	Recheck and tighten (reduce subsequent drift rate)	Leveling adjustment: $h_{t+1} = (1-\rho_1)h_t$	Full replacement: $h_{t+1} = 0$
Fixture	No action	Clean and lubricate (slow subsequent wear)	Grinding repair: $w_{t+1} = (1-\rho_2)w_t$	Full replacement: $w_{t+1} = 0$
Sensors	No action	Zero-point calibration: $z^3 \rightarrow max(z^3 - 1, 1)$	Full calibration (reset to 1 if $z^3 \leq 3$)	Replacement : $z^3 = 1$
Hydraulic Oil Tank	No action	Online filtration: $z^4 \rightarrow max(z^4 - 1, 1)$	Partial oil change: $z^4 \rightarrow max(z^4 - 2, 1)$	Full oil change : $z^4 = 1$

For the bearing beam and fixture, moderate maintenance is modeled as imperfect repair, with restoration ratios ρ_1 and $\rho_2 \sim Beta(\alpha, \beta)$, and an expected restoration rate $E[\rho] \in [0.7, 0.9]$. This modeling reflects the diminishing returns under the "leveling → re-degradation → re-leveling" sequence, forcing the policy to balance over the long term between repeated low-cost leveling and a one-time high-cost replacement—a dilemma typical near the end of the bearing beam's service life in practice. The differentiated rollback magnitudes in the oil tank actions compel the policy to learn hierarchical decision-making rather than uniformly selecting the strongest maintenance action.

4.2. Action Feasibility Constraints and Action Masking

Not all actions are meaningful in every state. This study explicitly injects engineering priors into the policy via an action mask $M_t^{(i)} \in \{0,1\}^{|\mathcal{A}^{(i)}|}$. Under hard rules, a component is forced to be replaced upon failure (e.g., $h_t \geq h_{\text{crit}} \Rightarrow M_t^{(1)} = [0,0,0,1]$). Under soft constraints, when the horizontality deviation approaches the tolerance limit, the "no action" option is masked ($h_t \geq 0.8h_{\text{crit}} \Rightarrow M_t^{(1)} = [0,1,1,1]$) to avoid missing critical maintenance windows. Invalid actions (e.g., when $z_t^3 = 1$, calibration is meaningless) can likewise be masked. The mask is implemented at the network output layer by setting the corresponding logits to $-\infty$:

$$\widetilde{Q}\left(s_t, a^{(i)}\right) = Q\left(s_t, a^{(i)}\right) + \log M_t^{(i)}\left(a^{(i)}\right) \tag{9}$$

This hard-constraint approach is more stable and more sample-efficient than reward-penalty shaping. Ablation experiments indicate that it is the design element contributing the most to the final performance.

4.3. Joint Execution and Multi-Objective Reward Function

To appropriately account for downtime costs, this study defines a joint execution rule: if any $a_t^i > 0$, a downtime event is triggered. The fixed downtime cost is charged once per event, $c_{\text{ds}}(a_t) = C_{\text{setup}} \cdot 1\left[\exists i : a_t^{(i)} > 0\right]$—this is the core manifestation of the opportunistic maintenance effect, i.e., performing maintenance actions on multiple components during a single shutdown does not increase the fixed downtime cost. The downtime duration follows parallel execution and is taken as the longest action duration, $T_{\text{down}}(a_t) = \max_i \tau^{(i)}\left(a_t^{(i)}\right)$.

https://doi.org/10.3390/s26134087

The total variable downtime cost equals the sum of materials and labor costs of the actions, plus the production loss rate per unit downtime multiplied by the downtime duration.

The reward function converts the maintenance objectives into a scalar signal:

$$r_t = -\left[c_{\text{var}}\left(a_t\right) + c_{\text{ds}}\left(a_t\right) + c_{\text{fail}}\left(s_t\right) + c_{\text{qual}}\left(z_t^{(3)}\right)\right] \tag{10}$$

where $c_{\text{var}}\left(a_t\right)$ is the sum of variable costs for all proactive maintenance operations, and $c_{\text{ds}}\left(a_t\right)$ is the fixed downtime cost. The failure penalty is $c_{\text{fail}}\left(s_t\right) = \sum_i c_{\text{f}}^{(i)} 1\left[s_t^{(i)} \in \mathcal{F}^{(i)}\right]$, where $\mathcal{F}^{(i)}$ is the failure-state set of component i, and $c_{\text{f}}^{(i)}$ is set far higher than the preventive replacement cost to reflect the severe consequences of failures (e.g., specimen damage, secondary damage). The term $c_{\text{qual}}\left(z_t^{(3)}\right)$ denotes the test-data-quality loss and is a distinctive element of the reward function in this study; its value increases stepwise with the sensor degradation level: $c_{\text{qual}}\left(1\right) = 0$ when ${z_{\text{t}}}^3 = 1$, and for ${z_{\text{t}}}^3 = 2, 3, 4$ it takes $\kappa_1 \cdot N_{\text{test}}(t)$, $\kappa_2 \cdot N_{\text{test}}(t)$, $\kappa_3 \cdot N_{\text{test}}(t)$, respectively, where N_{test} is the number of tests completed at the current time step and $\kappa_1 < \kappa_2 \ll \kappa_3$. This term aligns the policy with the metrological characteristics of the fatigue testing machine—an often overlooked yet crucial cost dimension in general maintenance decision studies.

5. Multi-Head Deep Reinforcement Learning Method

5.1. Algorithm Selection and Motivation

For the multi-component collaborative maintenance decision model of the electro-hydraulic servo fatigue testing machine constructed in this study, one of the core challenges lies in the high-dimensional, multi-dimensional discrete action space. The system comprises four key components, each associated with four mutually exclusive maintenance actions (no action, inspection/prevention, intermediate/partial repair, complete replacement), forming a complex joint action space of size 4^4 = 256. When selecting the base deep reinforcement learning (DRL) algorithm, it is essential to carefully balance the structural characteristics of the action space, sample efficiency, and the difficulty of embedding engineering feasibility constraints.

State-of-the-art policy-gradient methods (e.g., TD3 and SAC) perform well in industrial control but are natively designed for continuous action spaces. Forcibly applying them to the multi-dimensional discrete maintenance setting considered in this study typically requires continuous relaxations (e.g., Gumbel-Softmax reparameterization) or hard discretization of continuous outputs. Such transformations introduce substantial gradient-estimation noise and quantization errors, markedly increasing optimization difficulty and making convergence challenging in real engineering contexts where work-order data and physical samples are costly. Even with discrete-space variants (e.g., SAC-Discrete), these methods must construct and maintain a global probability distribution over the entire joint action space. As the number of maintained components grows, the joint action space explodes exponentially (M^N), forcing traditional architectures to output and optimize a policy vector as large as 256 dimensions in this case, which leads to severe parameter redundancy, poorer sample efficiency, and degraded training stability. More critically, continuous-control algorithms or traditional global-discrete policies do not natively and cleanly support per-component action masking. Under hard engineering constraints such as "a failed component must be replaced," imposing constraints on a 256-dimensional global joint distribution or on continuous outputs is highly complex and cannot guarantee that, during early exploration, the agent will avoid safety-violating invalid actions with 100% certainty.

https://doi.org/10.3390/s26134087

Based on the above considerations, this study adopts the Branching Dueling DQN (BDQ) as the core base algorithm. Through its distinctive branching structure and dueling decomposition, the BDQ architecture exhibits the following three notable advantages tailored to multi-component collaborative maintenance scenarios:

(1) Action-space decoupling and linearized parameterization: BDQ is specifically designed for multi-discrete dimensions with independent Q-value output branches, assigning a dedicated output head to each component. This reduces the parameter complexity at the output layer of policy network from the exponential growth of the joint space ($M \hat{} N = 256$) to linear growth ($\sum_{i=1}^{N} M_i = 4 \times 4 = 16$), effectively eliminating action-space explosion and greatly improving sample efficiency in multi-dimensional discrete control.

(2) Native and seamless support for action masking: Because the action spaces of individual maintenance components are fully decoupled at the output, the agent can conveniently apply component-specific state masks directly on each branch. By setting the logits of actions that do not meet engineering feasibility or logical constraints to $-\infty$ at the network output layer, the agent is guaranteed not to trigger dangerous or invalid maintenance combinations throughout the entire training process. Compared with traditional reward shaping, this hard-constraint mechanism delivers higher training stability and faster convergence.

(3) Capturing state–action advantages: BDQ integrates a dueling architecture within each branch, precisely decomposing the Q-value of each component into the global state value $V(s)$ and the component action advantage $A(s, a)$. In the operation of the fatigue testing machine, states such as severe degradation or critical failures occur infrequently; the dueling architecture enables the agent to accurately assess the overall safety value of system without frequently exploring every specific maintenance combination, thereby effectively mitigating learning lag under sparse major maintenance events.

5.2. Multi-Head Q-Network Architecture with Inter-Component Attention

The proposed network architecture consists of three parts: a shared encoder, an inter-component attention layer, and branching heads.

The shared encoder maps the state vector s_t to a common feature representation: $h_t = f_{\text{enc}}(s_t; \theta_{\text{enc}})$. The function f_{enc} is a three-layer multilayer perceptron with 256 neurons per layer, using LayerNorm and SiLU activation. Sharing the low-level representation improves the sample efficiency of the component-wise decisions.

The inter-component attention layer is introduced to explicitly model economic correlations among components and correlations induced by opportunistic maintenance. The shared feature h_t is linearly projected into each component's query, key, and value vectors $\mathrm{q}_i, \mathrm{k}_i, \mathrm{v}_i$, and then the cross-component attention-weighted features are computed:

$$\tilde{h}_t^{(i)} = \sum_{j=1}^{4} \mathrm{softmax}\left(\frac{\mathrm{q}_i^{\top} \mathrm{k}_j}{\sqrt{d_k}}\right) \mathrm{v}_j \tag{11}$$

$\tilde{h}_t^{(i)}$ is the context feature for decision-making on the i-th component, integrating the state information of all four components. The engineering intuition behind this design is that when the oil tank is close to the oil-change threshold, the sensors should tend toward opportunistic calibration even if they are only mildly degraded—this cross-component dependency is precisely what the attention mechanism models.

https://doi.org/10.3390/s26134087

The branching heads (with a dueling architecture) are independently configured for each component, using the dueling decomposition to split the Q-value into the state value and the advantage function:

$$Q^{(i)}\left(s_t, a^{(i)}\right) = V^{(i)}(s_t) + A^{(i)}\left(s_t, a^{(i)}\right) - \frac{1}{\left|\mathcal{A}^{(i)}\right|}\sum_{a'} A^{(i)}\left(s_t, a'\right) \tag{12}$$

The Q-value of the joint action is obtained by averaging the component-wise Q-values: $Q(s_t, a_t) = \frac{1}{N}\sum_{i=1}^{N} Q^{(i)}\left(s_t, a_t^{(i)}\right)$.

5.3. Training Algorithm

The algorithm adopts Double DQN target construction and an experience replay mechanism. The replay buffer stores transition tuples (s_t, a_t, r_t, s_{t+1}, M_{t+1}) (where M_{t+1} is the action mask of the next state). The training objective is to minimize the temporal-difference loss:

$$\mathcal{L}(\theta) = \mathbb{E}_{(s_t, a_t, r_t, s_{t+1})\sim\mathcal{D}}\left[\left(y_t - Q(s_t, a_t; \theta)\right)^2\right] \tag{13}$$

$$y_t = r_t + \gamma \cdot \frac{1}{N}\sum_{i=1}^{N} Q^{(i)}\left(s_{t+1}, \arg\max_{a'} \tilde{Q}^{(i)}\left(s_{t+1}, a'; \theta\right); \theta^-\right) \tag{14}$$

where θ^- denotes the target network parameters, and a soft update is used: $\theta^- \leftarrow \tau\theta + (1-\tau)\theta^-$.

This study employs three training techniques: (1) reward normalization, applying online running mean/variance normalization to each cost term to prevent large terms from dominating the gradients; (2) curriculum learning, constraining the simulation horizon and reducing the stochasticity of degradation rates in early training, then gradually restoring the target difficulty once the policy has largely formed; (3) prioritized experience replay, increasing the sampling probability of transitions with larger TD errors (typically corresponding to failures or major maintenance events). Key hyperparameters are: discount factor $\gamma = 0.99$, learning rate 3×10^{-4}, soft-update coefficient $\tau = 0.005$, replay buffer capacity 10^5, batch size 256, encoder hidden dimension 256, and number of attention heads 4.

6. Experimental Validation

6.1. Case Study and Simulation Environment

This study uses a 100 kN electro-hydraulic servo material fatigue testing machine of a certain model as the case object. The machine performs high-cycle and low-cycle fatigue tests on metal specimens, operates about 16 h per day, and runs 260 days per year. Based on the HS-MDP formalization in Sections 2–5, a simulation environment is constructed with a time step $\Delta t = 1$ day and an episode length T = 260 steps. Degradation and cost parameters are determined jointly from the manufacturer's maintenance manual, corporate financial records from the past 3 years, and site visits to three user organizations. The baseline horizontality drift rate of the bearing beam is $\mu_{h,0} = 1.2 \times 10^{-3}$ mm/day, the critical horizontality is $h\_crit = 0.50$ mm, and the baseline wear rate of the fixture is $\mu_{w,0} = 4.5 \times 10^{-3}$ mm/day. The discrete transition probabilities for the sensors and the oil tank are obtained via maximum likelihood estimation from dwell times at each level in work-order data. Operating conditions u_t are generated according to typical fatigue test schedules: load amplitude L_t is uniformly sampled in [0.6, 1.4], and loading frequency f_t is sampled from a normal distribution within 10–30 Hz. The simulation environment is implemented using the OpenAI Gymnasium standard interface, encapsulated as a four-component joint MDP, with the software platform comprising Python 3.10, NumPy 1.26, and PyTorch 2.2.1.

https://doi.org/10.3390/s26134087

The base transition probabilities of the two discrete chains are obtained by maximum-likelihood estimation from the level dwell times recorded in 3 years of calibration and maintenance work orders. For a one-step (pure-birth) chain, the dwell time in level i is geometrically distributed with parameter $p_{i,i+1}$, so the maximum-likelihood estimate is

$$\hat{p}^0_{i,i+1} = \frac{1}{\bar{d}_i}, \bar{d}_i = \frac{1}{N_i}\sum_{k=1}^{N_i} d_i^{(k)} \tag{15}$$

where $d_i^{(k)}$ is the dwell time (in time steps) of the k-th sojourn in level i and N_i is the number of such sojourns. These baseline estimates are referenced to the nominal operating condition (f_0, θ_0); the acceleration coefficients β and α in Equations (7) and (8) are then estimated by log-linear regression of the per-condition transition rates on loading frequency and oil temperature, respectively.

It should be emphasized that the manufacturer manuals, 3-year calibration and maintenance records, financial records, and site investigations were used to identify and calibrate the parameters of the HS-MDP simulator. However, all policy training, baseline comparisons, ablation studies, robustness tests, and sensitivity analyses reported in this section were conducted offline in this data-calibrated simulation environment. No learned policy was deployed online on the physical fatigue testing machine, and no intervention decisions were executed on the real equipment during this study. Therefore, the following results should be interpreted as simulation-based evidence of decision performance rather than as direct field validation.

6.2. Baselines and Evaluation Metrics

To systematically evaluate the relative advantages of the proposed method (denoted BDQ-Att-Mask), five representative baselines are selected for comparison:

- S1 Periodic: Periodic maintenance policy that performs medium-intensity maintenance and scheduled replacements at fixed intervals prescribed by the manufacturer's maintenance manual.
- S2 Threshold-CBM: Threshold-based condition-based maintenance strategy that triggers the corresponding maintenance when a component's state exceeds a preset threshold.
- S3 Independent-DQN: Four independent DQNs (one per component) with no awareness of other components' states, used to isolate the contribution of collaborative modeling.
- S4 PPO-MultiDisc.: PPO for multi-discrete actions, used to verify BDQ's advantage in sample efficiency.
- S5 Vanilla-BDQ: Standard BDQ without inter-component attention or action masking, serving as the ablation baseline.

To ensure a fair comparison, Independent-DQN, PPO-MultiDiscrete, Vanilla-BDQ, and BDQ-Att-Mask were trained under the same maximum episode horizon of 260 steps and the same final environment-interaction budget of [780,000] steps. The number of executed environment steps was counted directly from the training logs rather than inferred from the number of episodes, because curriculum learning and early termination may produce episodes of different lengths. All methods used the same set of training seeds. Each seed jointly controlled network initialization, environment stochasticity, action sampling, and minibatch sampling. After training, the network parameters were frozen, and every trained policy was evaluated using the same set of 30 previously unseen evaluation seeds. The evaluation seeds were not used for hyperparameter selection.

Where algorithmically applicable, the methods shared the same state preprocessing, reward normalization, curriculum schedule, network width, discount factor, and total environment-interaction budget. Independent-DQN, Vanilla-BDQ, and BDQ-Att-Mask

https://doi.org/10.3390/s26134087

additionally used the same replay-buffer capacity, batch size, Double-DQN target construction, target-network update rule, exploration schedule, and prioritized-replay settings. PPO used its standard on-policy rollout and update procedure and therefore did not use an experience replay buffer.

Hyperparameters were selected using a predefined search protocol. For each DRL method, same candidate configurations were evaluated using the same tuning interaction budget and the same set of validation seeds. The configuration with the lowest mean annual total cost on the validation seeds was selected. The final 30-seed test results were obtained only after hyperparameter selection had been completed. Thus, the proposed method and all learned baselines received the same data budget and comparable hyperparameter-tuning effort.

The evaluation metrics are constructed along three dimensions: economic benefits (average annual total cost and four subcategories), reliability (average annual number of component failures), and operational availability (annual cumulative downtime and data efficiency). All metrics are reported as mean $\pm$ standard deviation over 30 independent random seeds for yearly simulations. To ensure comparability, the DRL policies' network weights are frozen after training, and evaluation is conducted using the same seeds.

The five baselines are chosen to represent the relevant families rather than to enumerate individual papers: S1 the prevailing industrial practice (fixed schedule), S2 classical condition-based maintenance, S3 the decentralized multi-agent family (one independent learner per component), S4 the modern on-policy actor-critic family (multi-discrete PPO), and S5 the value-decomposition backbone without our additions. Recent maintenance-DRL methods—dependent-competing-risk CBM [11], structure-dependent and multi-agent formulations [12,13], and attention-augmented RUL-driven policies [19]—are formulated for homogeneous or purely discrete-state systems and do not natively support a continuous–discrete hybrid state, engineering action masking, or a data-quality objective. A faithful re-implementation in the present HS-MDP would therefore require substantial modification; the closest faithful representatives of their algorithmic cores (decentralized $\rightarrow$ S3, on-policy actor-critic $\rightarrow$ S4, value decomposition $\rightarrow$ S5) are already included.

6.3. Overall Performance Comparison

All four DRL methods were trained under the fairness protocol described in section above. Each method received the same final budget of [780,000] environment interactions and used the same set of independent training seeds.

In the simulation environment, the proposed method was trained for 3000 episodes, totaling approximately 7.8×10^5 interaction steps. The average annual total cost and its subcategories across 30 random seeds for each strategy are summarized in Table 3. The results show that, from the engineering baseline perspective, periodic maintenance has the highest costs in all categories, while threshold-based CBM leverages state feedback to reduce costs by about 30.8% relative to periodic maintenance. From the DRL baseline perspective, Independent DQN further reduces costs by about 13.0% compared with threshold-based CBM, and PPO performs similarly to Vanilla-BDQ. The proposed method achieves a 60.3% cost reduction relative to periodic maintenance, a 42.6% reduction relative to threshold-based CBM, and a 34.2% reduction relative to Independent DQN. The largest contributors to the cost reduction are failure costs and data-quality loss, which decrease by 88.6% and 74.9%, respectively, reflecting a marked improvement in the avoidance of major risks.

https://doi.org/10.3390/s26134087

Table 3. Parameter-provenance.

Quantity (Symbol)	Value	Source/Identification
Beam baseline drift $\mu_{h,0}$	1.2×10^{-3} mm/day	Fitted to 3 yr horizontality logs; cross-checked with maintenance manual
Beam failure threshold h_{crit}	0.50 mm	Alignment/horizontality tolerance for valid axial loading (cite governing standard, ISO 23788/ASTM E1012) [22]
Beam diffusion σ_h; coupling p_1, p_2, λ	$\sigma_h = 5 \times 10^{-4}, p_1 = 2.0$ $p_2 = 1.2,\ \lambda = 2.0$	Residual variance/regression on load- and frequency-stratified degradation rates
Fixture wear rate $\mu_{w,0}$	4.5×10^{-3} mm/day	Fitted to clamping-wear logs
Fixture failure threshold w_{crit} ; diffusion σ_w	$w_{\text{crit}} = 0.8$ mm $\sigma_w = 1 \times 10^{-3}$	Clamping-slip tolerance from acceptance spec
Sensor level bands (zero drift 0.1/0.3/0.5% FS)	Table 1	Force/displacement calibration accuracy classes (ISO 7500-1 [23])
Oil contamination levels	NAS 1638 classes	NAS 1638 [24]
Base transition probs $p^{(3),0}, p^{(4),0}$	$p^{(3),0} = 0.004\ , 0.006,\ 0.008$ $p^{(4),0} = 0.010,\ 0.015,\ 0.020,\ 0.025$	Dwell-time MLE from work orders (formula above)
Acceleration coeffs β, α	$\beta = 0.05, \alpha = 0.07$	Log-linear regression on frequency/oil temperature
Load amplitude L_t	$\sim U$ [0.6, 1.4]	Usage/test-schedule logs
Loading frequency f_t	N within 10–30 Hz	Usage/test-schedule logs
Fixed downtime cost C_{setup}; variable costs; failure penalty $c_f^{(i)}$	C_{setup} = 200 $c_f^{(i)}$ = 5000, 2000, 1500, 2000	3-yr corporate financial records (labor, materials, production-loss rate); scrapped-specimen cost
Data-quality coeffs $\kappa_1 < \kappa_2 \ll \kappa_3$	$\kappa_1, \kappa_2, \kappa_3$ = 20, 150, 1200	Cost of out-of-tolerance/discarded test data

In terms of reliability, the method proposed in this study yields a total of 1.4 failures per year across the four components, significantly lower than Periodic (9.8/year), Threshold-CBM (5.6/year), and Independent-DQN (3.7/year). In terms of operational availability, the method achieves 10.6 shutdowns per year, an annual cumulative downtime of 68.5 h (about 47% of Periodic), and a data efficiency of 96.2% (Periodic: 82.1%). The engineering significance lies in the markedly improved continuity of testing tasks—assuming the machine operates 16 h per day, each avoided unplanned failure prevents approximately 8–24 h of emergency downtime; compared with periodic maintenance, the method proposed in this study can recover about 52–156 h of effective operating time per year. Training curves and performance comparisons are shown in Figure 2.

The 60.3% reduction relative to periodic maintenance should be read together with its decomposition. The two categories that the fixed manufacturer schedule cannot react to—failure cost and data-quality loss—account for most of the gap, falling by 88.6% and 74.9%, respectively, whereas the maintenance and downtime costs, which periodic maintenance already incurs deliberately, decrease by a more moderate ~40%. The large headline figure therefore reflects the avoidance of high-consequence failures and out-of-tolerance data rather than a uniform across-the-board cost saving. Against the strongest learned baselines, the margin is correspondingly more modest but still consistent: the proposed method reduces total cost by 24.5% relative to Vanilla-BDQ and by 29.2% relative to PPO-MultiDiscrete, isolating the contribution of the inter-component attention and action masking from that of the BDQ backbone itself. All the experiment results has been listed in Table 4.

https://doi.org/10.3390/s26134087

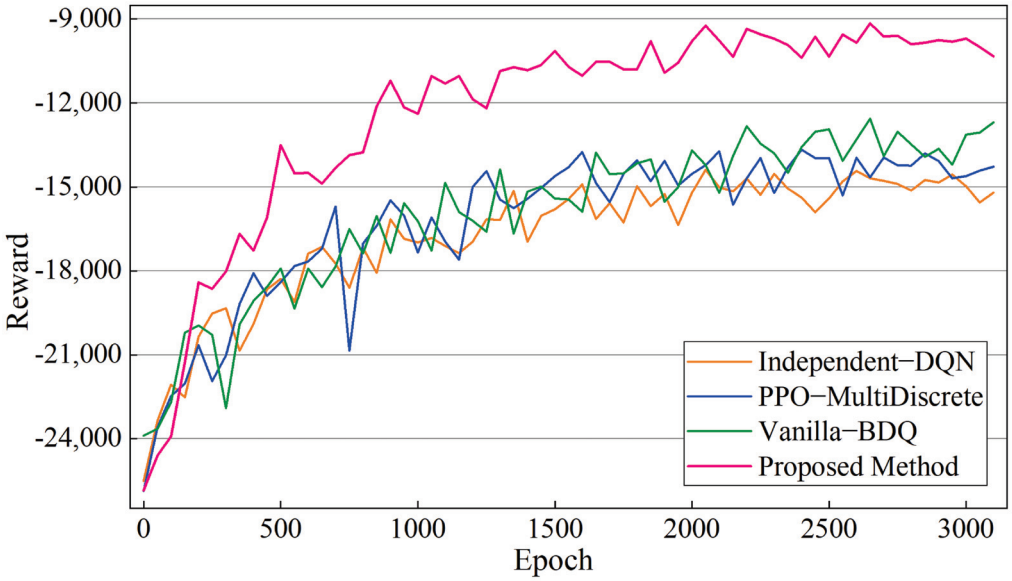


Figure 2. Comparison of episode reward curves during training.

Table 4. Comparison of average annual cost metrics for each strategy (unit: CNY).

Strategy	Maintenance Cost	Downtime Cost	Failure Cost	Data Quality Loss	Total Cost	Relative Periodic
S1 Periodic	8420 ± 210	5260 ± 180	6850 ± 920	4180 ± 310	24,710 ± 1220	—
S2 Threshold-CBM	6420 ± 228	4150 ± 172	3640 ± 712	2890 ± 245	17,100 ± 950	−30.8%
S3 Independent-DQN	5680 ± 305	4820 ± 256	2280 ± 520	2150 ± 228	14,930 ± 1085	−39.6%
S4 PPO-MultiDisc.	5950 ± 340	4150 ± 288	1950 ± 485	1820 ± 260	13,870 ± 1218	−43.9%
S5 Vanilla-BDQ	5520 ± 298	3980 ± 242	1820 ± 465	1690 ± 235	13,010 ± 1132	−47.4%
S6 BDQ-Att-Mask (Proposed Method)	4870 ± 185	3120 ± 162	780 ± 245	1050 ± 152	9820 ± 702	−60.3%

6.4. Opportunistic Maintenance and Cross-Component Attention Analysis

Define the opportunistic maintenance ratio ρ_opp as the proportion of "downtime events in which maintenance on multiple components is executed" to "the total number of downtime events." The proposed method achieves $\rho\_opp = 62.4\%$, far higher than Threshold-CBM (21.5%) and Independent-DQN (28.2%). This proportion directly indicates that the proposed method actively captures the economic effect of "a single shutdown enabling parallel maintenance of multiple components": among the 10.6 annual shutdowns, about 6.6 involve two or more components, and each opportunistic maintenance event involves an average of 2.5 components, as shown in Figure 3. Each part of this stacked bar chart represents maintenance downtime for a single component, two components, three components, and four components, respectively, and the red line represents the dividing line between single-component maintenance downtime and multi-component maintenance downtime.

Furthermore, the inter-component attention weights of the trained policy network are visualized on the test set (Figure 4). The attention matrix exhibits clear off-diagonal elements: (1) Oil tank → sensors ($\alpha = 0.34$): periods of elevated oil temperature and oil degradation typically coincide with high-load, high-frequency operating conditions, accelerating sensor degradation; (2) Fixture → bearing beam ($\alpha = 0.28$): wear of the fixture leads to clamping misalignment, which imposes eccentric loading on the beam and accelerates horizontality degradation; (3) Sensors → oil tank ($\alpha = 0.22$): sensor degradation can impair overload protection, indirectly causing oil overheating. These coupling patterns

https://doi.org/10.3390/s26134087

are highly consistent with the physical causal chain of the mechano-hydraulic system, demonstrating good interpretability of the inter-component attention mechanism.

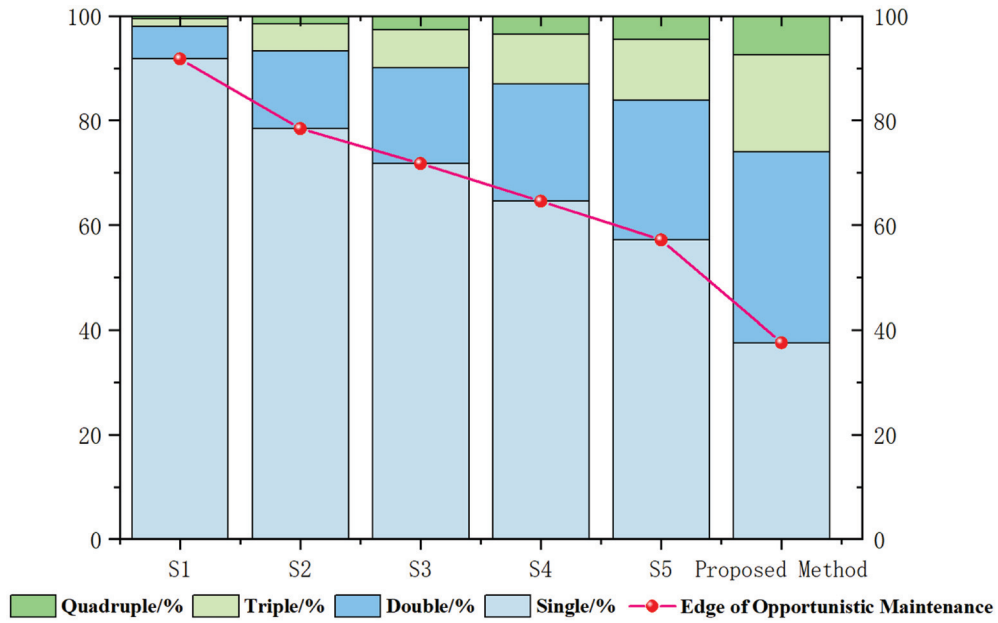


Figure 3. Comparison of opportunistic maintenance.

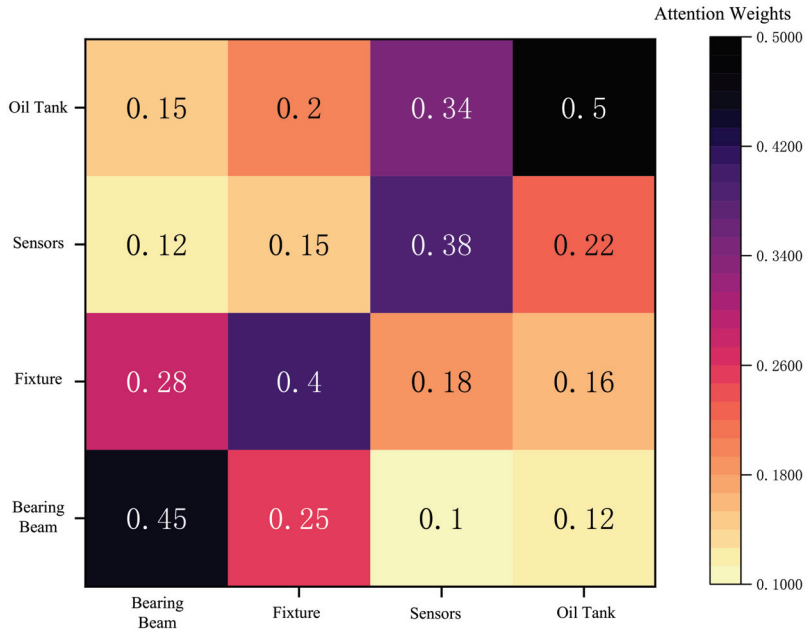


Figure 4. Heatmap of inter-component attention weights.

6.5. Ablation Experiments

To quantitatively assess the independent contribution of each core component of the method, this study designs the following five ablation variants. All variants are retrained and evaluated in the same environment and with the same set of 30 random seeds, and the results are summarized in Table 5: (A) Full method (BDQ-Att-Mask); (B) −Attention: remove inter-component attention, with branching heads deciding directly based on shared encoder features; (C) −Mask: remove action masking, suppressing invalid actions via a reward penalty (−1000); (D) −Quality: remove the data-quality loss term from the reward function; (E) −Dueling: replace dueling heads with standard Q-heads.

https://doi.org/10.3390/s26134087

Table 5. Performance comparison of ablation experiments (mean over 30 seeds).

Variant	Total Cost(CNY)	Number of Failures	Number of Shutdowns	Relative Loss
A. BDQ-Att-Mask	9820 ± 702	1.4	10.6	—
B. −Attention	$11{,}340 \pm 905$	2.2	11.2	+15.5%
C. −Mask	$12{,}860 \pm 1185$	3.6	13.8	+30.9%
D. −Quality	$11{,}580 \pm 945$	1.5	10.4	+18.0%
E. −Dueling	$10{,}520 \pm 820$	1.7	11.0	+7.1%

Although Table 5 reports only the aggregate cost together with the annual failure and shutdown counts, these two counts already act as a decomposition lens: the failure count is a proxy for the failure-cost contribution and the shutdown count for the downtime contribution, so the source of each variant's degradation can be inferred without additional instrumentation.

Removing action masking (C) is the most damaging variant (+30.9%) and degrades both proxies at once—failures rise from 1.4 to 3.6 per year and shutdowns from 10.6 to 13.8. Without the mask, feasibility and safety are enforced only through the -1000 reward penalty, which the agent must learn by repeatedly sampling infeasible actions; more importantly, the soft mask that forbids "no action" near the tolerance limit is lost, so critical maintenance windows are missed and the failure-driven cost dominates the increase.

Removing inter-component attention (B) raises the cost by 15.5% but adds only 0.6 shutdowns ($10.6 \rightarrow 11.2$). The increase is therefore not explained by more frequent downtime but by poorer coordination within each downtime: the opportunistic–maintenance ratio falls from its full-method value of 62.4%, so maintenance that the complete model bundles into a single shutdown is now executed across separate visits, and the moderate rise in failures (to 2.2) reflects the loss of cross-component anticipation such as the oil-tank $\rightarrow$ sensor coupling.

Removing the data-quality term (D) is the most informative case. Its failure count (1.5) and shutdown count (10.4) are statistically indistinguishable from—indeed marginally lower than—those of the full method, yet the total cost rises by 18.0%. Because neither reliability nor availability has changed, the entire increase must originate in a cost dimension that these two columns do not capture, namely the silent loss of out-of-tolerance test data: once the term is removed, the agent's incentive to maintain the sensors collapses, the frequency of the sensor reaching level 3 increases roughly 5.2-fold, and the policy pays through discarded data rather than through additional shutdowns. This is direct evidence that the data-quality term is not a redundant relabelling of existing costs but a distinct objective that reshapes the sensor-maintenance behaviour.

Removing the dueling decomposition (E) yields the smallest degradation (+7.1%), with failures and shutdowns close to the full method but a larger seed-to-seed standard deviation than the full method (820 vs. 702). Its contribution is therefore concentrated in optimization stability rather than in the structure of the converged policy.

6.6. Simulation-Based Robustness and Parameter Sensitivity Analysis

The following robustness and sensitivity analyses were performed entirely in the calibrated HS-MDP simulation environment. They were designed to examine the response of the learned policy to controlled model-parameter and cost-accounting perturbations. Unless otherwise stated, the policy trained under the nominal simulation parameters was frozen and evaluated without retraining.

For the degradation-model robustness test, the degradation-rate parameters in Table 3 were simultaneously perturbed by $\pm 10\%$, $\pm 20\%$, and $\pm 30\%$ in the evaluation environment. For each perturbation level, all compared policies were evaluated using the same

https://doi.org/10.3390/s26134087

30 test seeds. This setting represents model misspecification between the nominal training simulator and the perturbed evaluation simulator.

(1) Degradation-Parameter Perturbation Robustness

In engineering practice, there is unavoidable uncertainty in estimating degradation parameters; therefore, it is necessary to assess the robustness of policy to parameter perturbations. We took three levels of relative perturbations (±10%, ±20%, ±30%) are simultaneously applied during evaluation. For each perturbation level, 30 random test seeds are generated to evaluate the average annual total cost of the method presented in this chapter (directly using the policy trained under baseline conditions, without retraining) and three baseline strategies. The results are shown in Figure 5.

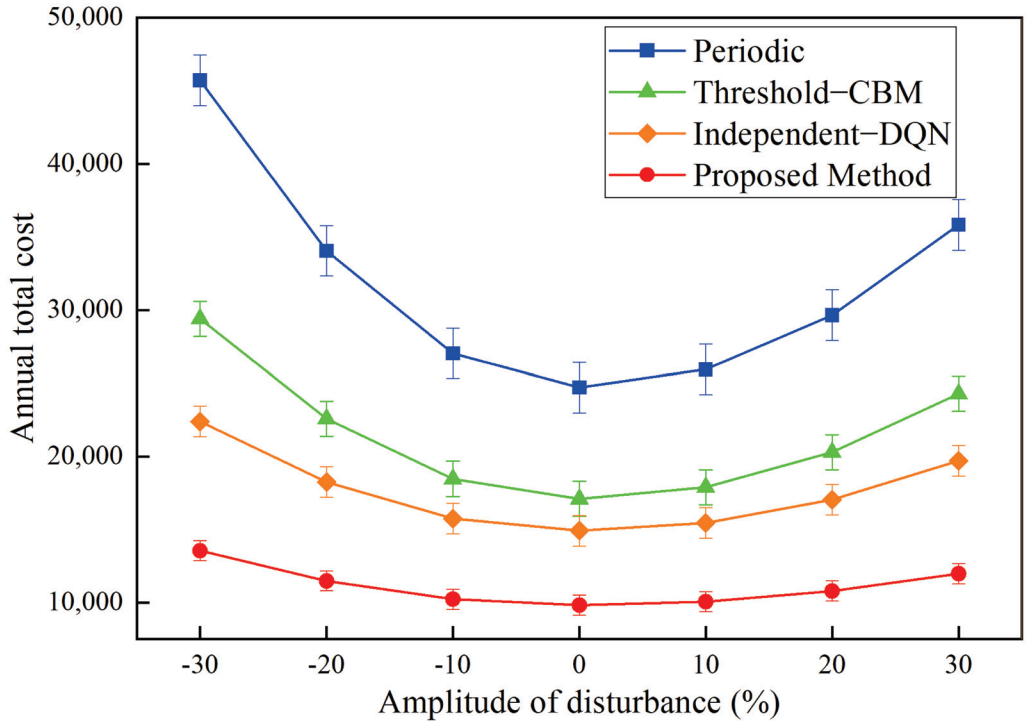


Figure 5. Robustness test.

The results indicate that when the degradation parameters are underestimated by 30%, the number of failures rises markedly for all strategies; the cost of the method presented in this chapter increases by about 38.2%, while Threshold-CBM increases by about 72.6%. When the degradation parameters are overestimated by 30%, all strategies show a stronger tendency toward over-maintenance; the cost of the method presented in this study increases by about 15.4%, and Periodic by about 21.8%. The method presented in this study exhibits the smallest performance degradation under both perturbation extremes, with robustness clearly surpassing all baselines. These results suggest that the DRL-learned policy does not "overfit" the degradation parameters used during training, but it instead has learned more generalizable state-based rules for multi-component coordinated decision-making.

(2) Cost-Parameter Sensitivity Analysis

Maintenance cost structures may vary significantly across enterprises and time periods. This section evaluates the transferability of the policy by perturbing cost parameters. Sensitivity is examined for three categories of cost parameters: the failure penalty coefficient (×0.5, ×1, ×2, ×4), the fixed downtime cost (×0.5, ×1, ×2, ×4), and the data-quality coefficient (×0.5, ×1, ×2, ×4). Under each perturbation level, the method presented in this

https://doi.org/10.3390/s26134087

chapter (using the policy trained under baseline conditions) is re-evaluated, and the results are summarized in Table 6.

Table 6. Cost-Parameter Sensitivity Analysis (relative cost changes with respect to the method proposed in this study).

Parameter	×0.5	×1.0 (Baseline)	×2.0	×4.0
Failure penalty coefficient	−18.5%	0	+8.2%	+15.6%
Fixed downtime cost	−12.4%	0	+10.5%	+22.8%
Data-quality coefficient	−8.6%	0	+5.2%	+12.4%

The results show that the failure penalty is the most sensitive factor; however, even when amplified by 4×, the cost of the method proposed in this study increases by only +15.6%—because failures are already rare under the baseline policy (1.4 per year), so even when the per-failure cost is multiplied severalfold, its total contribution remains limited. For perturbations to the fixed downtime cost, the cost change is relatively larger (+22.8% at ×4) because this term directly couples with the number of shutdowns; nevertheless, the policy has learned to reduce shutdowns via opportunistic maintenance, limiting the amplification effect. The data-quality coefficient exhibits the lowest sensitivity, indicating that the sensor-maintenance policy has been trained into a robust behavior that is largely insensitive to this term. Taken together, these results indicate that the method proposed in this study maintains good transferability across different cost structures.

6.7. Discussion and Limitations

While the proposed BDQ-Att-Mask framework demonstrates substantial improvements in multi-component joint maintenance, it is fundamentally constrained by the assumption of full state observability. Currently, the HS-MDP relies on the accurate and continuous acquisition of structural degradation and discrete sensor states. Consequently, if sensor failures, missing data, or abnormal measurements occur in practical industrial environments, the decision-making agent will receive corrupted state inputs. This perturbation can lead to sub-optimal policies, such as missing critical opportunistic maintenance windows or executing redundant repairs. Furthermore, as the application extends to larger-scale industrial systems, the interactions between mechanical structures and hydraulic systems become significantly more complex. In such scenarios, relying solely on high-level macro-maintenance scheduling is insufficient to guarantee structural safety under severe state constraints or dynamic environmental disturbances. A highly promising future extension to address these intertwined limitations is transitioning towards a "Control-Maintenance Co-design" paradigm, which integrates deep reinforcement learning with advanced bottom-level robust control. Specifically, to mitigate the vulnerabilities associated with missing data and sensor anomalies, extended-state-observer (ESO) based adaptive control methods can be introduced as a state-estimation front end [25]. By utilizing ESOs to dynamically reconstruct the true system states under full-state constraints, the DRL agent can be supplied with filtered, robust state representations even when physical sensors fail. Additionally, incorporating optimized-based fault-tolerant control with disturbance rejection can ensure that the electro-hydraulic system maintains safe operational stability and rejects external disturbances while operating in a degraded state prior to the execution of scheduled maintenance [26].

7. Conclusions

Addressing the joint maintenance decision-making problem for multiple components in electro-hydraulic servo material fatigue testing equipment, this study proposes a method

https://doi.org/10.3390/s26134087

based on multi-head deep reinforcement learning. The main contributions and conclusions are as follows:

1. A continuous–discrete hybrid-state MDP modeling framework is constructed. The continuous degradation trajectories of the bearing beam and fixture are characterized by Wiener processes, while the main sensors and the hydraulic oil tank are modeled by discrete Markov chains to align with periodic calibration and oil-grade inspections in practice. A component-differentiated four-tier action space is designed, and engineering feasibility constraints are injected via action masking.
2. A multi-head Branching Dueling DQN network with an inter-component attention mechanism is proposed, reducing the parametric complexity of the joint action from $\prod i|Ai|$ to $\Sigma i|Ai|$. Through a hierarchical design of shared encoder-attention layer-branching heads, component-wise decisions become aware of each other's states and autonomously learn opportunistic maintenance patterns.
3. The reward function introduces a data-quality loss term determined by the sensor degradation level, aligning the optimization objective with the metrological attributes of the fatigue testing machine. Ablation experiments show that although this term appears to be merely one cost component, its removal leads to an overall relaxation of sensor maintenance and an 18.0% increase in implicit data-discarding costs.
4. The case study indicates that the proposed method reduces the average annual total cost by 60.3% relative to periodic maintenance and by 42.6% relative to threshold-based CBM; the number of failures drops from 9.8/year to 1.4/year, and data efficiency rises from 82.1% to 96.2%. The opportunistic maintenance ratio reaches 62.4%, far exceeding other baselines. Robustness tests and attention visualizations further validate the method's transferability and interpretability.

Future work will focus on three directions: (1) transferring the policy trained in simulation to physical equipment and studying policy correction methods in Sim-to-Real settings; (2) extending to multi-machine collaborative maintenance scenarios, considering cross-machine spare parts sharing and the coupling with personnel scheduling; (3) exploring online adaptive estimation of degradation model parameters to further enhance the policy's generalization to unknown operating conditions.

Author Contributions: Conceptualization, P.L.; Methodology, G.H.; Validation, G.H.; Data curation, J.W.; Writing—original draft, G.H. and J.W.; Writing—review & editing, G.H. and J.W.; Visualization, G.H. and J.X.; Supervision, P.L.; Funding acquisition, P.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (NSFC), General Program, grant number 52475559.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available upon reasonable request.

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

References

1. De Jonge, B.; Scarf, P.A. A review on maintenance optimization. *Eur. J. Oper. Res.* **2020**, *285*, 805–824. [CrossRef]
2. Nicolai, R.P.; Dekker, R. Optimal Maintenance of Multi-component Systems: A Review. In *Complex System Maintenance Handbook*; Springer: London, UK, 2008; pp. 263–286. [CrossRef] [PubMed]
3. Thomas, L.C. A survey of maintenance and replacement models for maintainability and reliability of multi-item systems. *Reliab. Eng.* **1986**, *16*, 297–309. [CrossRef]

https://doi.org/10.3390/s26134087

4. Saha, A.; Kar, S.; Maiti, M. Multi-item fuzzy-stochastic supply chain models for long-term contracts with a profit sharing scheme. *Appl. Math. Model.* **2015**, *39*, 2815–2828. [CrossRef]
5. Pan, Z.; Balakrishnan, N. Reliability modeling of degradation of products with multiple performance characteristics based on gamma processes. *Reliab. Eng. Syst. Saf.* **2011**, *96*, 949–957. [CrossRef]
6. Ang, F.; Kerstens, P.J. Decomposing the Luenberger–Hicks–Moorsteen Total Factor Productivity indicator: An application to U.S. agriculture. *Eur. J. Oper. Res.* **2017**, *260*, 359–375. [CrossRef]
7. Si, X.S.; Wang, W.; Hu, C.H.; Zhou, D.H. Remaining useful life estimation—A review on the statistical data driven approaches. *Eur. J. Oper. Res.* **2011**, *213*, 1–14. [CrossRef]
8. Grall, A.; Bérenguer, C.; Dieulle, L. A condition-based maintenance policy for stochastically deteriorating systems. *Reliab. Eng. Syst. Saf.* **2002**, *76*, 167–180. [CrossRef]
9. Castanier, B.; Grall, A.; Bérenguer, C. A condition-based maintenance policy with non-periodic inspections for a two-unit series system. *Reliab. Eng. Syst. Saf.* **2005**, *87*, 109–120. [CrossRef]
10. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef] [PubMed]
11. Zhang, N.; Si, W. Deep reinforcement learning for condition-based maintenance planning of multi-component systems under dependent competing risks. *Reliab. Eng. Syst. Saf.* **2020**, *203*, 107094. [CrossRef]
12. Chen, J.; Wang, Y. A deep reinforcement learning approach for maintenance planning of multi-component systems with complex structure. *Neural Comput. Applic.* **2023**, *35*, 15549–15562. [CrossRef]
13. Do, P.; Nguyen, V.T.; Voisin, A.; Iung, B.; Neto, W.A.F. Multi-agent deep reinforcement learning-based maintenance optimization for multi-dependent component systems. *Expert Syst. Appl.* **2024**, *245*, 123144. [CrossRef]
14. Wang, Z.; Schaul, T.; Hessel, M.; Van Hasselt, H.; Lanctot, M.; De Freitas, N. Dueling network architectures for deep reinforcement learning. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning*; JMLR: Norfolk, MA, USA, 2016; Volume 48, pp. 1995–2003.
15. Tavakoli, A.; Pardo, F.; Kormushev, P. Action Branching Architectures for Deep Reinforcement Learning. *Proc. AAAI Conf. Artif. Intell.* **2018**, *32*, 4131–4138. [CrossRef]
16. Varricchione, G.; Alechina, N.; Dastani, M.; De Giacomo, G.; Logan, B.; Perelli, G. Pure-Past Action Masking. *Proc. AAAI Conf. Artif. Intell.* **2024**, *38*, 21646–21655. [CrossRef]
17. Wang, Z.; Li, X.; Sun, L.; Zhang, H.; Liu, H.; Wang, J. Learning State-Specific Action Masks for Reinforcement Learning. *Algorithms* **2024**, *17*, 60. [CrossRef]
18. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*; Curran Associates Inc.: New York, NY, USA, 2017; pp. 6000–6010.
19. Bai, Y.; Du, J.; Li, H.; Bao, X.; Li, L.; Zhang, C.; Yan, J.; Wang, R.; Xu, Y. An Adaptive Framework for Remaining Useful Life Prediction Integrating Attention Mechanism and Deep Reinforcement Learning. *Sensors* **2025**, *25*, 6354. [CrossRef] [PubMed]
20. Vu, H.C.; Do, P.; Fouladirad, M.; Grall, A. Dynamic opportunistic maintenance planning for multi-component redundant systems with various types of opportunities. *Reliab. Eng. Syst. Saf.* **2020**, *198*, 106854. [CrossRef]
21. Olde Keizer, M.C.A.; Teunter, R.H.; Veldman, J.; Babai, M.Z. Condition-based maintenance for systems with economic dependence and load sharing. *Int. J. Prod. Econ.* **2018**, *195*, 319–327. [CrossRef]
22. *ISO 23788:2012*; Information and Documentation: International Standard Identifier for Libraries and Related Organizations (ISIL). International Organization for Standardization: Geneva, Switzerland, 2012.
23. *ISO 7500-1:2018*; Metallic Materials—Calibration and Verification of Static Uniaxial Testing Machines—Part 1: Tension/Compression Testing Machines—Calibration and Verification of the Force-Measuring System. International Organization for Standardization: Geneva, Switzerland, 2018.
24. *NAS 1638*; Cleanliness Requirements of Parts Used in Hydraulic Systems. Aerospace Industries Association of America: Washington, DC, USA, 1964.
25. Phan, V.D.; Ahn, K.K. Optimized-Based Fault-Tolerant Control of an Electro-Hydraulic System with Disturbance Rejection. *Appl. Sci.* **2022**, *12*, 9197. [CrossRef]
26. Xu, Z.; Deng, W.; Shen, H.; Yao, J. Extended-State-Observer-Based Adaptive Prescribed Performance Control for Hydraulic Systems With Full-State Constraints. *IEEE/ASME Trans. Mechatron.* **2022**, *27*, 5615–5625. [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.

https://doi.org/10.3390/s26134087

sensors

MDPI

Article

A Span-Prior-Guided Explainable Multimodal Neural Network Method for Final-State Quality Inspection of Hairpin Windings

Xiaopeng Chang [1], Bangcheng Zhang [1,2,*], Zhi Gao [1], Siyu Chen [1] and Jingru Liu [3]

[1] School of Mechatronic Engineering, Changchun University of Technology, Changchun 130012, China; 1202201011@stu.ccut.edu.cn (X.C.); gaozhi@ccut.edu.cn (Z.G.); 1202301013@stu.ccut.edu.cn (S.C.)
[2] School of Mechanical and Electrical Engineering, Changchun Institute of Technology, Changchun 130103, China
[3] FAW Tool & Die Manufacturing Co., Ltd., No. 1999, Jieda Road, Changchun 130011, China; liuqr_td@faw.com.cn
* Correspondence: zhangbangcheng@ccut.edu.cn

Abstract

For final-state quality inspection of three-dimensional stamped hairpin windings, existing studies still lack multimodal methods that integrate mechanical geometric constraints, prior-guided fusion, and decision interpretability. This study proposes a span-prior-guided explainable multimodal neural network method and develops SPIMA-Net. The final-state images were acquired at a fixed inspection station with a fixed camera position and imaging angle under a CCD vision light source. Final-state images are used as visual inputs, while geometric priors are constructed from span measurements and model-type information. A visual branch and a span branch extract image and prior features, and a span-prior-assisted gating mechanism modulates visual features to enable collaborative fusion. Experimental results show that SPIMA-Net achieves an accuracy of 98.14%, an F1-score of 96.55%, and an AUC of 0.9983 on the test set. Its nonconforming-class F1-score is improved by 10.44, 3.04, 6.27, 1.38, and 0.66 percentage points over the image-only, span-only, direct-fusion, SE-fusion, and CBAM-fusion models, respectively, while the total number of misclassifications decreases to five. Interpretability analysis shows that the model mainly focuses on span openings, end profiles, and local abnormal regions. Relative and absolute span deviations are identified as the main mechanical geometric factors affecting final-state quality classification. The proposed method provides a neuro-mechanical fusion approach that demonstrates high discriminative performance and engineering interpretability for hairpin winding quality inspection on the investigated industrial dataset.

Keywords: hairpin windings; final-state quality inspection; span prior; explainable multimodal neural network; multimodal sensing fusion

1. Introduction

With the development of intelligent manufacturing and industrial quality inspection, quality assessment of complex manufactured components is shifting from experience-based inspection to intelligent inspection that integrates visual images, geometric dimensions, and structured information. For mechanically formed components, final-state quality depends not only on external morphology but also on whether key geometric dimensions satisfy design requirements. In such inspection tasks, effective use of image information requires digital image representation, visual feature extraction, and convolutional representation learning to support morphology recognition and discriminative feature extraction [1–3].

Therefore, combining deep visual representation with mechanical geometric priors is important for final-state quality inspection with strong discriminative capability, explicit geometric constraints, and engineering interpretability.

Hairpin windings for flat-wire motors have advantages such as high slot fill factor, high power density, compact end-winding structure, and favorable heat dissipation, making them important for traction motor stators in new energy vehicles. Zou et al. reported that winding design factors such as slot–pole combinations, conductor transposition, end connections, and parallel branches affect the manufacturability and performance stability of hairpin windings [4]. Fleischer et al. further pointed out that electric drive manufacturing involves complex process chains, where manufacturing quality is critical for large-scale electric vehicle production [5]. After three-dimensional stamping, the final opening morphology, end profile, and span dimension jointly determine the assembly compatibility and quality state of hairpin windings. Existing studies have mainly addressed forming control, welding quality monitoring, visual inspection, and process quality assurance [6–10], but the joint evaluation of final-state images and key span dimensions after stamping remains insufficiently studied.

Deep visual neural networks have been widely used in industrial defect recognition, anomaly detection, and assembly quality assessment. Bono et al. further emphasized that quality control in automated production lines remains challenging under highly inconsistent operating conditions [11]. Related studies have explored industrial visual anomaly detection, synthetic data-based assembly inspection, few-shot defect detection, electronic board inspection, multiview inspection, and AI-enabled product defect detection [12–17]. These studies show that visual models are effective in extracting contour, edge, and local morphological features. However, for final-state inspection tasks with explicit geometric constraints, image information alone may be affected by visually similar appearances and may not fully identify quality abnormalities caused by key geometric deviations.

For three-dimensional stamped hairpin windings, final-state images provide information on end profiles, opening regions, and local morphological abnormalities, while span-related geometric dimensions provide mechanical constraints directly associated with quality criteria. These two types of information are complementary: image-only models may ignore explicit span deviation information, whereas span-only models cannot fully capture local contour abnormalities, end bending, or asymmetric deformation. Therefore, final-state images and span-related geometric information should be jointly modeled for quality classification.

Multimodal learning provides a useful basis for such joint modeling. Baltrušaitis et al. summarized representation, alignment, fusion, and co-learning issues in multimodal machine learning [18], while Zhao et al. reviewed deep multimodal data fusion methods and highlighted the value of multisource information fusion for improving model representation and robustness [19]. Jiao et al. further categorized multimodal fusion strategies into early fusion, deep fusion, late fusion, and hybrid fusion [20]. In industrial scenarios, multimodal modeling of images, sensors, process parameters, and structured data has been applied to diagnosis, process monitoring, quality assessment, and intelligent manufacturing systems [21–24]. However, many existing multimodal models still rely on feature concatenation or conventional fusion, and span information is rarely used to actively guide the visual branch toward opening regions, end profiles, and local abnormal locations.

Attention mechanisms, gating mechanisms, and feature recalibration methods have been widely used to help neural networks focus on task-relevant regions and discriminative features. Schlemper et al. introduced attention gating to guide networks toward important regions [25], and Hu et al. proposed channel-wise feature recalibration to enhance discriminative representation [26]. These studies provide a methodological basis for prior-guided

https://doi.org/10.3390/s26154857

neural fusion. However, in final-state quality inspection of three-dimensional stamped hairpin windings, how to transform measured final span, theoretical span, span deviations, and model-type information into mechanical priors for visual feature modulation remains insufficiently investigated. In addition, industrial intelligent inspection models require not only classification accuracy but also engineering interpretability. Selvaraju et al. proposed Grad-CAM to visualize key image regions involved in convolutional neural network decisions [27]. Related studies in process quality management, visual quality control, industrial fault diagnosis, quality state monitoring, and zero-defect manufacturing have shown that explainable artificial intelligence can improve the transparency and practical value of industrial AI systems [28–34]. Therefore, final-state quality inspection of hairpin windings should not only output conforming/nonconforming decisions but also explain the image regions and geometric factors influencing the decisions.

In summary, previous studies on hairpin windings have mainly focused on forming control, welding monitoring, and visual inspection, while final-state quality inspection after three-dimensional stamping remains insufficiently studied. In this study, the final-state images were acquired at a fixed inspection station with a fixed camera position and imaging angle under a CCD vision light source. This task requires the joint evaluation of final-state morphology and span dimensions. Image-only methods cannot explicitly use span deviation information, whereas span-only methods cannot capture local contour abnormalities. Conventional multimodal fusion usually relies on feature concatenation and lacks active guidance from span priors to visual features. Therefore, a span-prior-guided and interpretable multimodal method is needed for final-state quality inspection of three-dimensional stamped hairpin windings.

To address these issues, this study proposes a span-prior-guided explainable multimodal neural network method for final-state quality inspection of hairpin windings and constructs the Span-Prior-Informed Multimodal Assessment Network (SPIMA-Net) as its core implementation model. The proposed method integrates final-state image representations with span-related geometric priors. Through a visual branch, a span branch, and a prior-assisted gating mechanism, it enables collaborative fusion between neural visual features and mechanical geometric constraints, and is applied to final-state quality classification and risk scoring of three-dimensional stamped hairpin windings for flat-wire motors.

The main contributions of this study are summarized as follows:

(1) A multisource collaborative representation method is proposed for final-state quality inspection of hairpin windings.

Final-state image information and mechanical span geometric priors are jointly modeled, enabling the model to simultaneously exploit visual morphological features and span deviation information for final-state quality classification and risk quantification.

(2) A span-prior-guided multimodal neural network model, SPIMA-Net, is constructed.

Mechanical geometric priors are encoded through a span branch, and a span-prior-assisted gating mechanism is designed to modulate visual features, thereby enhancing the model's discriminative ability for span opening regions, end profiles, and local abnormal regions.

(3) A dual-sided explainable analysis method is proposed for image regions and mechanical span factors.

By combining Grad-CAM with permutation feature importance analysis, the key abnormal regions attended to by the model are identified from the image side, while the contributions of major geometric factors to quality classification are quantified from

https://doi.org/10.3390/s26154857

the span side. This result supports the interpretability and practical applicability of the proposed method for final-state quality inspection of hairpin windings.

2. Methods

2.1. Problem Definition and Span Prior Construction

The final-state quality of three-dimensional stamped hairpin windings is jointly affected by end morphology, span dimensions, and differences among model types. Figure 1A presents the simplified structure of the stamped hairpin winding and the definition of the theoretical span. The stamped hairpin winding is formed from a flat copper conductor through bending and three-dimensional stamping, where the theoretical span s_i^{ref} denotes the designed distance between the two end regions. Figure 1B shows the practical assembly scene of the hairpin winding, in which the final-state span directly affects its insertion accuracy and assembly compatibility. When the span deviation is excessive, assembly difficulty or local interference may occur, as shown in Figure 1C,D. Therefore, final-state quality inspection should consider both visual morphology information and span-related geometric information.

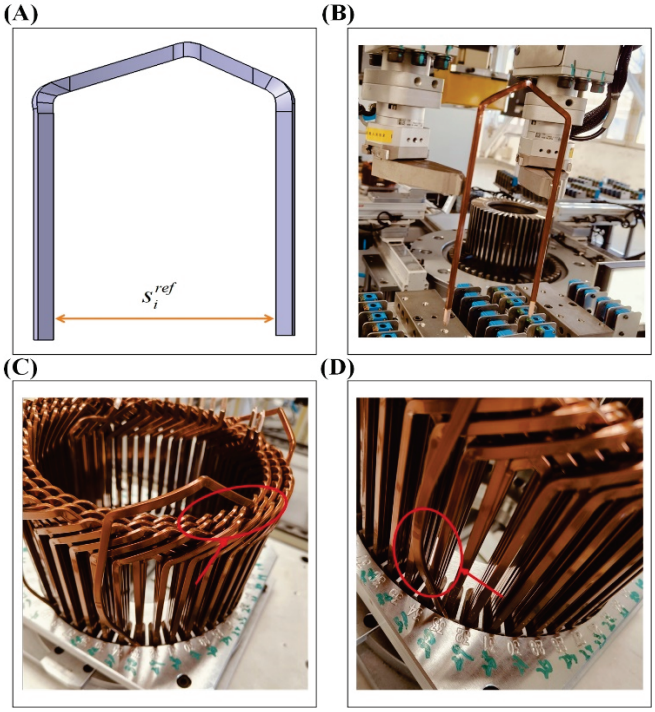


Figure 1. Structure and assembly of the stamped hairpin winding: (**A**) simplified structure and definition of the theoretical span; (**B**) practical assembly scene; (**C**,**D**) examples of local interference caused by excessive span deviation.

Based on the stamped hairpin winding structure and span definition, to enable joint modeling of final-state image information and mechanical geometric priors, the i-th sample is defined as:

$$x_i = \left\{ I_i, s_i, s_i^{ref}, m_i, y_i \right\} \tag{1}$$

where I_i denotes the final-state image of the i-th sample, s_i denotes the measured final span, s_i^{ref} denotes the theoretical span corresponding to the model type, m_i denotes the model-type information, and $y_i \in \{0,\ 1\}$ denotes the quality label, where 1 indicates conforming and 0 indicates nonconforming.

https://doi.org/10.3390/s26154857

Because the theoretical span varies across different hairpin winding model types, using only the measured final span s_i cannot accurately reflect its deviation from the target dimension. Therefore, the absolute span deviation is defined as:

$$e_i = \left| s_i - s_i^{ref} \right| \tag{2}$$

and the relative span deviation is further defined as:

$$d_i = \frac{\left| s_i - s_i^{ref} \right|}{s_i^{ref} + \varepsilon} \tag{3}$$

where ε is a small constant introduced to avoid division by zero. In this study, ε was fixed as 1×10^{-8}. This value was used only for numerical stability and was not tuned according to different model types. The absolute span deviation e_i represents the actual dimensional deviation, while the relative span deviation d_i is used to reduce the influence of scale differences in theoretical span among different model types on quality classification.

Based on these definitions, the span prior vector is constructed as:

$$g_i = \left[s_i, s_i^{ref}, e_i, d_i, m_i \right] \tag{4}$$

where the model-type information m_i can be encoded using one-hot encoding. This vector provides mechanical geometric information, including the measured dimension, target dimension, deviation magnitude, and model-type context. The variables s_i, s_i^{ref}, e_i, and d_i in the vector are not independent raw observations but correlated geometric descriptors constructed for final-state span quality judgment. Specifically, e_i and d_i characterize the deviation of s_i from s_i^{ref} from the perspective of absolute span deviation and relative span deviation, respectively. Therefore, this interdependence does not introduce an additional modeling problem but helps explicitly represent the engineering judgment information of span deviation and provides structured geometric priors for subsequent multimodal feature fusion.

The objective of this study is to learn the following mapping:

$$f : \{ I_i, g_i \} \rightarrow \{ \hat{y}_i, \hat{r}_i \} \tag{5}$$

where $\hat{y}_i$ denotes the predicted quality class, and $\hat{r}_i$ denotes the risk score. Since this study focuses on identifying nonconforming samples, the predicted probability of the nonconforming class is defined as the risk score. A higher risk score indicates a greater risk of final-state quality abnormality.

2.2. SPIMA-Net Network Architecture

As shown in Figure 2, SPIMA-Net is constructed as a span-prior-guided multimodal classification framework. The model takes the final-state image I_i and the span prior vector g_i as two complementary inputs. The visual branch maps I_i into image feature representation z_i^v, while the span branch encodes g_i into span prior feature representation z_i^s. Unlike direct feature concatenation, SPIMA-Net uses the span prior representation to generate gating weights, which are then applied to the visual features for adaptive feature modulation. This design allows the geometric prior information to participate in the visual feature learning process before multimodal fusion. The modulated visual representation and span prior representation are subsequently concatenated and passed into the classification head to obtain the conforming/nonconforming prediction and the nonconforming risk score.

In the visual branch, ResNet18 is adopted as the visual backbone for extracting final-state image features [35]. ResNet18 is a residual convolutional network composed of stacked residual blocks with shortcut connections, which helps stable extraction of contour,

https://doi.org/10.3390/s26154857

edge, and local morphology features from final-state images and reduces overfitting risk under limited industrial samples:

$$z_i^v = f_v(I_i; \theta_v) \tag{6}$$

where z_i^v denotes the visual feature vector, θ_v represents the parameters of the visual branch, and $f_v(\bullet)$ denotes the visual feature extraction function.

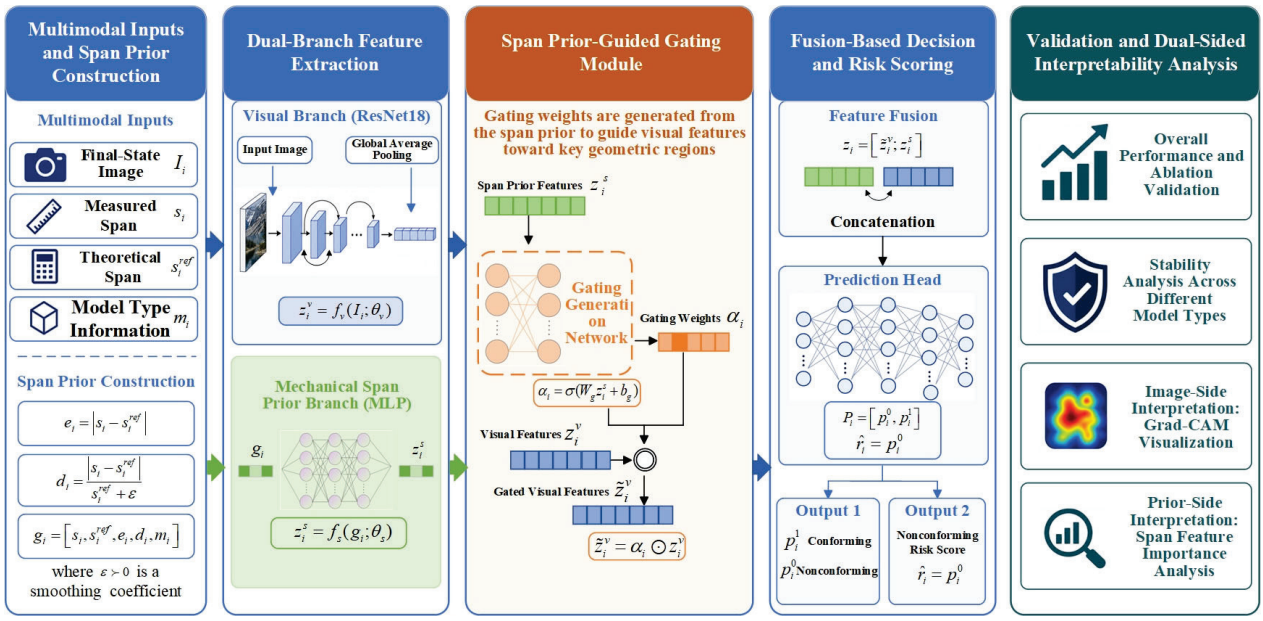


Figure 2. SPIMA-Net span-prior-guided multimodal final-state quality inspection framework.

In the span branch, the span prior vector g_i constructed in Section 2.1 is fed into a multilayer perceptron (MLP) encoder to obtain the span prior feature:

$$z_i^s = f_s(g_i; \theta_s) \tag{7}$$

where z_i^s denotes the span prior feature vector, while $f_s(\bullet)$ and θ_s represent the span branch encoding function and its parameters, respectively. The span branch adopts a two-layer fully connected structure, with ReLU activation and Dropout introduced to suppress overfitting.

To avoid passive concatenation of image features and span information only at the final stage, this study designs a span-prior-guided gating mechanism. This module uses span prior features to generate gating weights with the same dimensionality as the visual features:

$$\alpha_i = \sigma\left(W_g z_i^s + b_g\right) \tag{8}$$

where α_i denotes the gating weight generated from the span prior, $\sigma(\bullet)$ is the sigmoid function, and W_g and b_g are the weight matrix and bias term of the gating mapping, respectively. In this study, the visual feature vector extracted by the ResNet18 branch is $z_i^v \in \mathbb{R}^{512}$, and the span prior feature vector is $z_i^s \in \mathbb{R}^{32}$. The linear gating layer maps z_i^s to $\alpha_i \in \mathbb{R}^{512}$, which has the same dimension as z_i^v. Therefore, the visual features are modulated by element-wise multiplication as follows:

$$\tilde{z}_i^v = \alpha_i \odot z_i^v \tag{9}$$

where $\odot$ denotes element-wise multiplication, and $\tilde{z}_i^v$ denotes the visual feature representation modulated by the span prior. This mechanism enables span prior information to participate in visual feature modulation and strengthens the discriminative representation of the model.

https://doi.org/10.3390/s26154857

The modulated visual features are then fused with the span prior features:

$$z_i = [\tilde{z}_i^v; z_i^s] \tag{10}$$

where z_i denotes the fused multimodal feature vector, and $[\tilde{z}_i^v; z_i^s]$ denotes the concatenation of the span-prior-modulated visual feature and the span prior feature.

The class probabilities are then produced by a fully connected classification head:

$$P_i = \text{soft max}(W_2\phi(W_1 z_i + b_1) + b_2) \tag{11}$$

where W_1, W_2, b_1 and b_2 are the parameters of the classification head, and $\phi(\bullet)$ denotes a nonlinear activation function. The fully connected layers in the span-related branch and the fusion classification head use ReLU activation. The gating module adopts a sigmoid activation function to generate element-wise modulation weights for visual features, and the final classification logits are converted into class probabilities using the softmax function. The Dropout rate is set to 0.2 in the span-related branch and 0.3 in the fusion classification head.

$$P_i = \left[p_i^0, p_i^1\right] \tag{12}$$

where P_i^0 and P_i^1 denote the softmax outputs corresponding to the nonconforming and conforming classes, respectively. The risk score is defined as:

$$\hat{r}_i = p_i^0 \tag{13}$$

That is, a higher softmax output for the nonconforming class indicates a higher final-state instability risk. In this study, $\hat{r}_i$ is used to represent the relative risk degree that the model assigns a sample to the nonconforming class. Considering that softmax outputs may be affected by data distribution bias and class imbalance, it is not interpreted as an absolute probability estimate after additional probability calibration.

In summary, SPIMA-Net does not simply concatenate image information and span information. Instead, it employs a span-prior-guided gating mechanism that enables mechanical geometric priors to participate in visual feature modulation, thereby achieving joint output of final-state quality classification and risk scoring.

2.3. Comparative Models and Ablation Experiment Design

To validate the effectiveness of the proposed SPIMA-Net and analyze the roles of image information, span priors, multimodal fusion, and the gating mechanism in final-state quality inspection, six comparative models are constructed: the image-only model, the span-only model, the direct-fusion model, the SE-fusion model, the CBAM-fusion model, and SPIMA-Net. The structural components and input information of these six models are summarized in Table 1.

Table 1. Structural components and input information of different models.

Model Name	Image Branch	Span Branch	Fusion Strategy	Output
Image-only	ResNet18	No	No	Class probability
Span-only	No	MLP	No	Class probability
Direct-fusion	ResNet18	MLP	Direct concatenation	Class probability
SE-fusion	ResNet18	MLP	SE-based fusion	Class probability
CBAM-fusion	ResNet18	MLP	CBAM-based fusion	Class probability
SPIMA-Net	ResNet18	MLP	Prior-assisted gating + concatenation	Class probability + risk score

Among them, the image-only model takes final-state images as input and uses ResNet18 for quality classification, thereby evaluating the independent discriminative capability of visual morphological information. The span-only model takes the span prior vector g_i as input and adopts an MLP for classification, thereby assessing the contribution

https://doi.org/10.3390/s26154857

of mechanical geometric priors to quality decision-making. The direct-fusion model simultaneously introduces image features and span features, but performs multimodal fusion only through feature concatenation, so as to evaluate the effect of conventional multimodal fusion. The SE-fusion and CBAM-fusion models introduce attention-based fusion modules to evaluate whether general attention-based mechanisms can improve multimodal feature interaction compared with direct feature concatenation. SPIMA-Net further incorporates a span-prior-guided gating mechanism on the basis of direct fusion, thereby examining the effectiveness of mechanical geometric priors in modulating visual features. Therefore, this set of experiments is used to analyze the effects of image information, span priors, direct fusion, attention-based fusion, and span-prior-guided gating on final-state quality classification performance.

2.4. Joint Loss and Dual-Sided Interpretability Analysis

To jointly ensure class discriminative and span-consistency constraints for risk scoring, this study adopts a joint optimization scheme that combines classification loss with a span auxiliary loss.

The classification loss is defined using weighted cross-entropy:

$$\mathcal{L}_{cls} = -\frac{1}{N}\sum_{i=1}^{N}\left[\alpha_1 y_i \log p_i^1 + \alpha_0 (1-y_i)\log p_i^0\right] \tag{14}$$

where N denotes the number of samples, while α_1 and α_0 are the class weights for the conforming and nonconforming classes, respectively, which are used to mitigate the influence of class imbalance during training.

To make the risk score reflect the degree of final-state span instability, a span auxiliary loss is further introduced. Considering that the risk score $\hat{r}_i$ lies within [0, 1], the relative span deviation d_i is normalized by the maximum value in the training set as follows:

$$\widetilde{d}_i = \frac{d_i}{d_{\max}^{train}} \tag{15}$$

and clipped to the interval [0, 1]. Subsequently, $\widetilde{d}_i$ is used as the risk constraint target, and the span auxiliary loss is defined as:

$$\mathcal{L}_{span} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{r}_i - \widetilde{d}_i\right)^2 \tag{16}$$

Therefore, the total loss function is defined as:

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda \mathcal{L}_{span} \tag{17}$$

where λ is a balancing coefficient, which was set to 0.20 in this study. The class weights in the weighted cross-entropy loss were calculated according to the inverse class frequency in the training set. Since the nonconforming class was labeled as 0 and the conforming class as 1, the class weights were set to α_0 = 1.8646 for the nonconforming class and α_1 = 0.6832 for the conforming class. This joint loss enables the model to perform conforming/nonconforming classification while producing risk scores that are consistent with the degree of span deviation.

For interpretability analysis, this study reveals the model decision basis from both the image side and the span side. On the image side, Grad-CAM is used to generate class activation heatmaps for locating the key abnormal regions attended to by the model. On the span side, permutation feature importance analysis is performed for the measured final span, theoretical span, absolute span deviation, relative span deviation, and model-type information. Each factor is perturbed separately. The changes in F1-score and AUC are then used to quantify its contribution to quality classification.

https://doi.org/10.3390/s26154857

In summary, the proposed framework combines joint-loss optimization for class decision and span-risk constraint with dual-sided interpretation from image regions and span-related factors.

3. Results

3.1. Dataset and Experimental Settings

To validate the effectiveness of the proposed SPIMA-Net for final-state quality inspection of hairpin windings, this study uses field-collected samples of flat-wire motor hairpin windings after the three-dimensional stamping process. The samples were collected from one stator production batch, which contained seven different hairpin winding geometries/model types. Image acquisition was performed at a fixed inspection station using one industrial camera, GV-CRM120-10, under a CCD vision light source. During acquisition, the camera position, imaging angle, and lighting condition were kept fixed.

The dataset consists of final-state images, quality labels, measured final spans, theoretical spans, span deviation features, and model-type information. The dataset was divided into training and test sets using a stratified 80%/20% random split according to the quality labels and model types. The same image or the same physical sample was not duplicated in both the training and test sets. Therefore, the experimental results were obtained under a sample-level random partition within the investigated stator production batch. The dataset composition and partitioning are summarized in Table 2.

Table 2. Dataset composition and partition statistics.

Dataset	Number of Conforming Samples	Number of Nonconforming Samples	Total Number of Samples
Overall Samples	982	361	1343
Training Set (80%)	786	288	1074
Test Set (20%)	196	73	269

In the experiments, nonconforming samples are treated as the positive class, with emphasis placed on evaluating the model's ability to identify final-state quality abnormalities. The evaluation metrics include accuracy, precision, recall, F1-score, and AUC. Accuracy measures the overall classification correctness; precision and recall reflect the correctness and completeness of nonconforming sample identification, respectively; F1-score provides a comprehensive evaluation of nonconforming-class recognition performance; and AUC assesses the overall separability between conforming and nonconforming samples.

Following the comparative design in Section 2.3, the compared models were trained and evaluated under the same training–test split. The network input image size was set to 224×224. During training, light image augmentation was applied, including brightness and contrast perturbation with a factor of 0.05, while no augmentation was used during testing. Images were normalized using the ImageNet mean and standard deviation.

Span-related variables were standardized using the mean and standard deviation calculated from the training set. For the span auxiliary loss, the relative span deviation was normalized by the maximum relative span deviation in the training set and clipped to [0, 1]. ResNet18 was trained from scratch. AdamW was used with a learning rate of 1×10^{-4}, a weight decay of 1×10^{-4}, a batch size of 16, and 20 training epochs. The balancing coefficient in the total loss was set to $\lambda = 0.20$, and the weighted cross-entropy class weights were set to $\alpha_0 = 1.8646$ for the nonconforming class and $\alpha_1 = 0.6832$ for the conforming class. The main random seed was 42. No additional learning-rate scheduler or early stopping strategy was used; all models were trained under this fixed 20-epoch setting. The experiments were implemented in Python 3.10.7 using PyTorch 2.13.0 and run in the

https://doi.org/10.3390/s26154857

PyCharm 2023.1 environment on a computer equipped with an NVIDIA GeForce RTX 2060 (6 GB) GPU. On this platform, SPIMA-Net has 11.27 M parameters and achieved an average inference time of 166.24 ms/sample, corresponding to 6.02 FPS. To verify the sufficiency of the 20-epoch training schedule, the training convergence curves of SPIMA-Net were further analyzed, as shown in Figure 3A–D. The classification loss and total loss gradually decreased and stabilized within 20 epochs, while the span auxiliary loss remained at a low and stable level. Meanwhile, the F1-score increased and became stable in the later epochs, indicating that the model had basically converged under the adopted training setting. The best model was selected mainly according to the F1-score of the nonconforming class. Overfitting was controlled using light augmentation, Dropout, weight decay, and class-weighted cross-entropy loss.

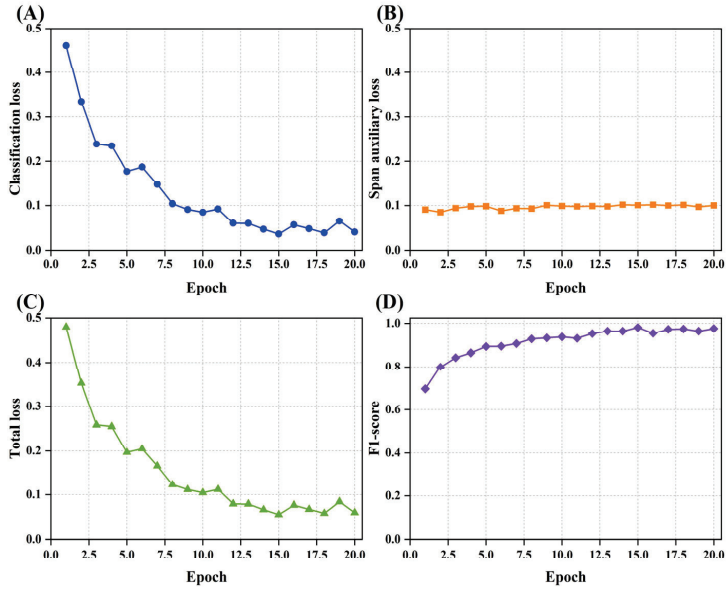


Figure 3. Training convergence curves of SPIMA-Net over 20 epochs. (**A**) Classification loss; (**B**) span auxiliary loss; (**C**) total loss; (**D**) F1-score.

3.2. Comparative Model and Ablation Analysis

To validate the effectiveness of the proposed SPIMA-Net and analyze the roles of image information, span priors, and the prior-guided gating mechanism in final-state quality classification, six models are compared: the image-only model, the span-only model, the direct-fusion model, the SE-fusion model, the CBAM-fusion model, and SPIMA-Net. Their overall performance on the test set is reported in Table 3.

Table 3. Overall performance comparison of different models on the test set.

Model Name	Accuracy%	Precision%	Recall%	F1%	AUC
Image-only	92.57	87.32	84.93	86.11	0.9586
Span-only	96.28	88.89	98.63	93.51	0.9986
Direct-fusion	94.80	91.55	89.04	90.28	0.9903
SE-fusion	97.40	95.83	94.52	95.17	0.9965
CBAM-fusion	97.77	95.89	95.89	95.89	0.9978
SPIMA-Net	98.14	97.22	95.89	96.55	0.9983

In this study, nonconforming samples are treated as the positive class. Accordingly, TP denotes the number of truly nonconforming samples correctly predicted as nonconforming, FN denotes the number of truly nonconforming samples incorrectly predicted as conforming, FP denotes the number of truly conforming samples incorrectly predicted

https://doi.org/10.3390/s26154857

as nonconforming, and TN denotes the number of truly conforming samples correctly predicted as conforming. The evaluation metrics are calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{21}$$

AUC is used to evaluate the overall discriminative ability of the model under different decision thresholds.

As shown in Table 3, SPIMA-Net achieves the best accuracy, precision, and F1-score, reaching 98.14%, 97.22%, and 96.55%, respectively, with an AUC of 0.9983. Compared with the image-only model, the accuracy and F1-score of SPIMA-Net are approximately 5.57 and 10.44 percentage points higher, respectively, indicating that final-state images alone are insufficient to fully identify quality abnormalities caused by span deviations. Compared with the span-only model, SPIMA-Net improves accuracy, precision, and F1-score by approximately 1.86, 8.33, and 3.04 percentage points, respectively, suggesting that span priors are important for identifying nonconforming samples, but geometric information alone cannot fully capture local contour and morphological abnormalities. Compared with the direct-fusion model, SPIMA-Net improves accuracy and F1-score by approximately 3.34 and 6.27 percentage points, respectively, demonstrating that the span-prior-guided gating mechanism provides stronger fusion-based discriminative capability than simple feature concatenation. Compared with SE-fusion and CBAM-fusion, SPIMA-Net further improves accuracy by 0.74 and 0.37 percentage points and F1-score by 1.38 and 0.66 percentage points, respectively. In particular, SPIMA-Net achieves the same recall as CBAM-fusion while obtaining higher precision and fewer misclassifications, indicating a more balanced trade-off between nonconforming-sample recognition and false-alarm control. This result also indicates that simply adding image information to span information does not necessarily improve the classification performance. Since span deviation is a strong discriminative cue in this task, naïve feature concatenation may introduce irrelevant visual variations, such as local appearance changes, imaging noise, or background interference, thereby weakening the dominant role of span prior information. This further supports the necessity of the proposed span-prior-guided gating mechanism for adaptive multimodal fusion.

To further evaluate the stability of model performance under different data partitions, repeated stratified random experiments were conducted. As shown in Table 4, SPIMA-Net maintained stable performance across different random splits, indicating that the reported results were not dependent on a single train–test partition.

Table 4. Repeated random experiment results under different data partitions.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC	ECE	Brier Score
Image-only	94.55 ± 1.20	89.96 ± 2.04	89.95 ± 2.76	89.95 ± 2.25	0.9767 ± 0.0132	0.0397 ± 0.0080	0.0421 ± 0.0077
Span-only	97.65 ± 1.41	92.92 ± 4.95	99.09 ± 0.79	95.85 ± 2.43	0.9991 ± 0.0008	0.0462 ± 0.0068	0.0175 ± 0.0068
Direct-fusion	96.78 ± 0.86	92.16 ± 2.17	96.35 ± 0.79	94.20 ± 1.51	0.9952 ± 0.0035	0.0239 ± 0.0067	0.0247 ± 0.0090
SE-fusion	97.89 ± 0.86	96.42 ± 3.30	95.89 ± 1.37	96.13 ± 1.54	0.9972 ± 0.0021	0.0214 ± 0.0026	0.0181 ± 0.0051
CBAM-fusion	97.40 ± 1.34	94.30 ± 3.99	96.35 ± 0.79	95.29 ± 2.29	0.9960 ± 0.0021	0.0278 ± 0.0108	0.0220 ± 0.0114
SPIMA-Net	98.02 ± 1.31	94.37 ± 3.16	98.63 ± 1.37	96.44 ± 2.31	0.9989 ± 0.0006	0.0200 ± 0.0114	0.0147 ± 0.0090

https://doi.org/10.3390/s26154857

To further evaluate the statistical reliability of the observed performance differences, a statistical significance analysis was conducted based on three repeated stratified random experiments with different random seeds. The nonconforming-class F1-score and AUC were used as evaluation metrics. SPIMA-Net was compared with each compared model, and the mean difference, Bootstrap 95% confidence interval, paired *t*-test *p*-value, and Wilcoxon signed-rank test *p*-value were calculated. The mean difference was defined as the performance of SPIMA-Net minus that of the compared model.

As shown in Table 5, SPIMA-Net achieved positive mean differences over most compared models. For the two fusion baselines, SPIMA-Net improved the nonconforming-class F1-score and AUC by 0.0032 and 0.0017 over SE-fusion, respectively. Compared with CBAM-fusion, SPIMA-Net improved the nonconforming-class F1-score by 0.0116, with a Bootstrap 95% confidence interval of [0.0075, 0.0139] and a paired *t*-test *p*-value of 0.0291. The AUC was improved by 0.0029, with a Bootstrap 95% confidence interval of [0.0013, 0.0046]. These results provide additional statistical evidence for the performance improvement of SPIMA-Net over the compared fusion baselines.

Table 5. Statistical significance analysis between SPIMA-Net and compared models.

Compared Model	Metric	Mean Difference	Bootstrap 95% CI	Paired *t*-Test *p*-Value	Wilcoxon *p*-Value
Image-only model	F1-score	0.0650	[0.0515, 0.0821]	0.0187	0.2500
Span-only model	F1-score	0.0059	[−0.0077, 0.0253]	0.6108	0.7500
Direct-fusion model	F1-score	0.0224	[0.0071, 0.0331]	0.1044	0.2500
SE-fusion model	F1-score	0.0032	[−0.0126, 0.0147]	0.7343	0.7500
CBAM-fusion model	F1-score	0.0116	[0.0075, 0.0139]	0.0291	0.2500
Image-only model	AUC	0.0221	[0.0073, 0.0303]	0.0962	0.2500
Span-only model	AUC	−0.0002	[−0.0013, 0.0009]	0.7697	0.7500
Direct-fusion model	AUC	0.0037	[0.0007, 0.0064]	0.1580	0.2500
SE-fusion model	AUC	0.0017	[−0.0002, 0.0028]	0.2171	0.5000
CBAM-fusion model	AUC	0.0029	[0.0013, 0.0046]	0.0936	0.2500

To further analyze the misclassification characteristics of different models, confusion matrices were constructed by treating nonconforming samples as the positive class. The confusion matrix statistics of the six compared models are presented in Table 6 and Figure 4.

Table 6. Confusion matrix statistics of the six compared models.

Model Name	TP	FN	FP	TN	Number Predicted as Nonconforming	Number Predicted as Conforming	Total Number of Misclassifications
Image-only	62	11	9	187	71	198	20
Span-only	72	1	9	187	81	188	10
Direct-fusion	65	8	6	190	71	198	14
SE-fusion	69	4	3	193	72	197	7
CBAM-fusion	70	3	3	193	73	196	6
SPIMA-Net	70	3	2	194	72	197	5

As shown in Figure 4A, the image-only model produces 20 misclassifications, including FN = 11 and FP = 9, indicating that final-state images alone remain insufficient for identifying some span-related abnormal samples. In Figure 4B, the span-only model reduces the total number of misclassifications to 10, with FN = 1 and FP = 9. This suggests that span priors are highly sensitive to nonconforming samples, but geometric information alone tends to generate false alarms for boundary conforming samples. As shown in Figure 4C, the direct-fusion model yields 14 misclassifications, including FN = 8 and FP = 6.

https://doi.org/10.3390/s26154857

Although it improves upon the image-only model, simple feature concatenation does not fully exploit the complementarity between image information and span priors.

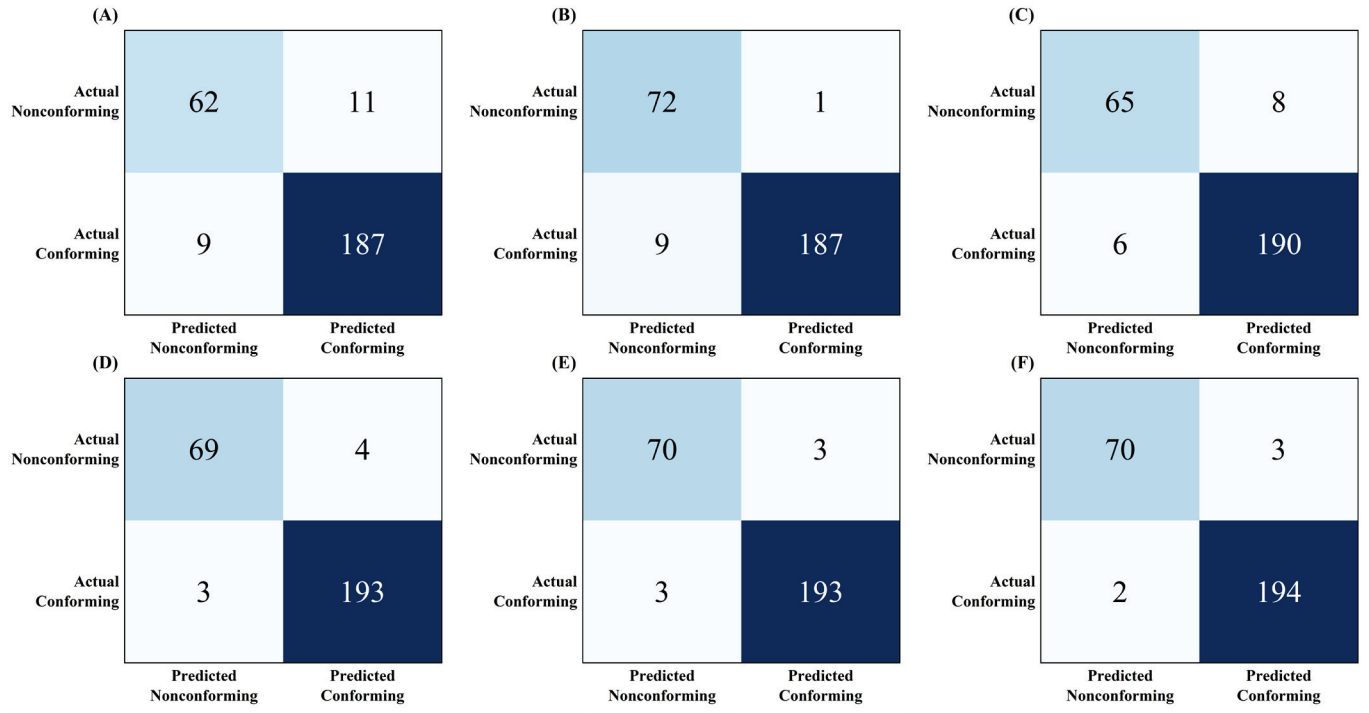


Figure 4. Confusion matrices of the six compared models: (**A**) Image-only; (**B**) Span-only; (**C**) Direct-fusion; (**D**) SE-fusion; (**E**) CBAM-fusion; (**F**) SPIMA-Net.

After introducing attention-based fusion mechanisms, the classification balance is further improved. As shown in Figure 4D, SE-fusion produces seven misclassifications, with FN = 4 and FP = 3. In Figure 4E, CBAM-fusion further reduces the total number of misclassifications to six, with FN = 3 and FP = 3, showing competitive performance close to SPIMA-Net. By comparison, Figure 4F shows that SPIMA-Net achieves the lowest number of misclassifications, with FN = 3 and FP = 2. More specifically, SPIMA-Net maintains a low missed-detection rate for nonconforming samples while further reducing false alarms for conforming samples compared with CBAM-fusion. This indicates that the span-prior-guided gating mechanism can more effectively use mechanical geometric information to modulate visual features, thereby strengthening the discriminative capability of final-state quality inspection for hairpin windings.

Taken together, Tables 3–6 and Figure 4 indicate that final-state images and span priors are clearly complementary. On this basis, the introduction of the span-prior-guided gating mechanism further improves overall performance, nonconforming-sample recognition, and misclassification control, thereby validating the effectiveness of the SPIMA-Net architecture.

3.3. Stability Analysis Across Different Model Types

In this study, model-type labels such as 1-2-6, 2-3-6, and 6-6-7 denote internal product identifiers for different stamped hairpin winding specifications on the production line. To further evaluate the stability of the proposed method across different model types, the F1-scores for the nonconforming class are calculated for six models on the test samples from seven model types, as shown in Figure 5.

Overall, SPIMA-Net maintains high recognition performance for most model types, with an overall test-set F1-score of 0.9655. Specifically, the F1-scores for the nonconforming class reach 1.0000 for the 2-3-6, 4-5-6, 6-6-5, and 6-6-7 model types, while those for the 1-2-6

https://doi.org/10.3390/s26154857

and 3-4-6 model types are 0.9000 and 0.9787, respectively. The SE-fusion model achieves identical F1-scores to SPIMA-Net on five model types but drops markedly on the 5-6-6 (0.7273) and 6-6-7 (0.9231) model types. The CBAM-fusion model reaches 1.0000 on the 1-2-6, 2-3-6, and 6-6-5 model types, and achieves 0.9231 on the 5-6-6 and 6-6-7 model types, yet it exhibits lower F1-scores on the 3-4-6 (0.9545) and 4-5-6 (0.8571) model types. These results indicate that the attention-based fusion models generally improve model-type-level stability, while SPIMA-Net achieves a more balanced performance across different model types.

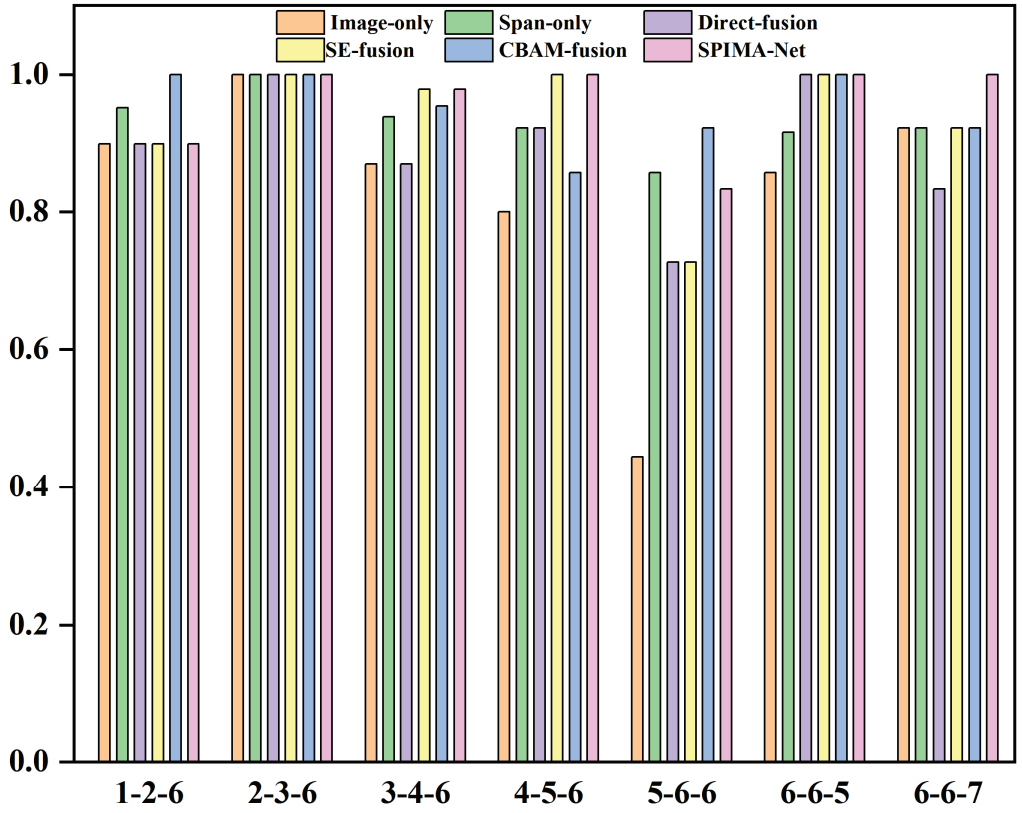


Figure 5. Comparison of nonconforming-class F1-scores among the six models across different model types.

For the 5-6-6 model type, SPIMA-Net achieves an F1-score of 0.8333, which is substantially higher than that of the image-only model (0.4444), the direct-fusion model (0.7273), and the SE-fusion model (0.7273), and close to that of the span-only model (0.8571). The CBAM-fusion model achieves a higher F1-score of 0.9231 on this model type. This result indicates that SPIMA-Net is not superior for every individual model type, but it maintains a more balanced overall performance across different specifications. For samples with more complex local morphological differences and closer geometric boundaries, image information alone is more susceptible to interference from visually similar appearances, whereas span priors can provide effective mechanical geometric constraints for quality classification. Figure 5 shows that the span-prior-guided mechanism helps maintain classification stability across different model types.

3.4. Risk Scoring and Interpretability Analysis

To further analyze the discriminative capability of model output probabilities, the representational effectiveness of risk scores, and the basis for model decisions, the proposed model and the compared baselines are evaluated from four aspects: ROC curves, risk-score distributions, Grad-CAM visualization, and span-branch feature importance.

https://doi.org/10.3390/s26154857

Figure 6 presents the ROC curves of the six compared models on the test set. All six models exhibit discriminative capability to different degrees. SPIMA-Net achieves an AUC of 0.9983, indicating strong capability to distinguish conforming and nonconforming samples. Although the span-only model obtains a slightly higher AUC of 0.9986, the confusion matrix results reported above show that it produces more false positives under the practical decision threshold. Compared with CBAM-fusion, SPIMA-Net achieves a slightly higher AUC and fewer misclassifications. These results indicate that SPIMA-Net maintains strong probability-level discrimination while achieving better misclassification control under the practical decision threshold.

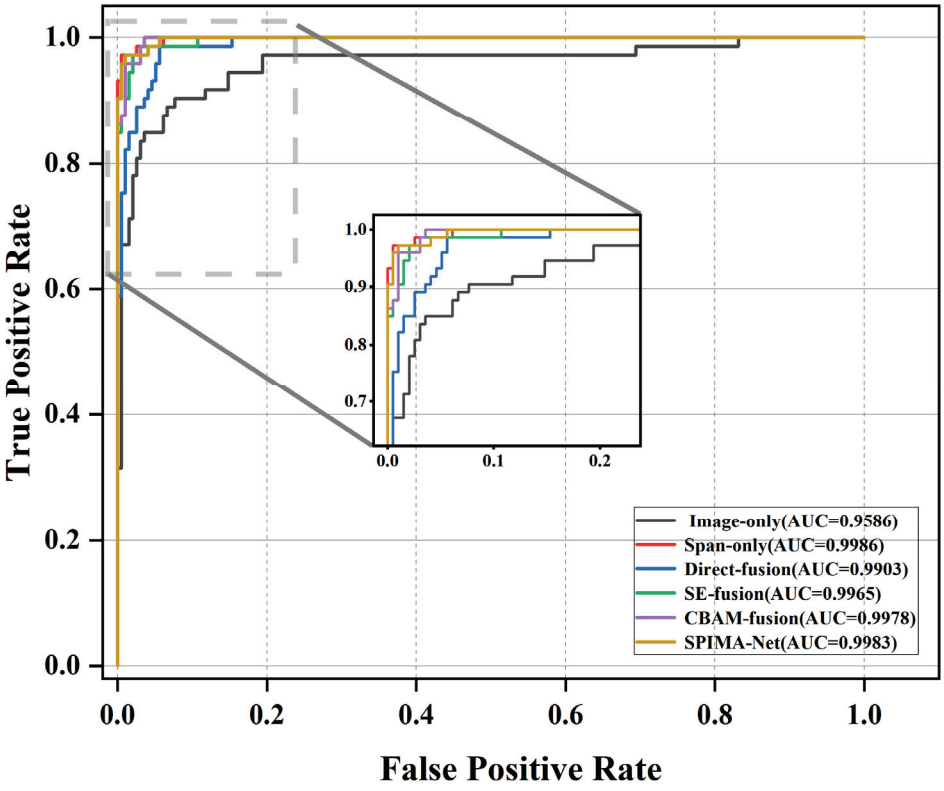


Figure 6. Comparison of ROC curves among the six compared models.

Figure 7 shows the distributions of risk scores for the six compared models on the test set, with 0.5 used as the risk decision threshold. As shown in Figure 7A, for the image-only model, 187 out of 196 conforming samples fall into the low-risk region, and 62 out of 73 nonconforming samples fall into the high-risk region, although some overlap still exists. Figure 7B shows that the span-only model assigns 72 nonconforming samples to the high-risk region, but at the same time classifies nine conforming samples as high-risk, indicating that although span information alone is highly sensitive to nonconforming samples, it still tends to assign excessively high risk scores to some boundary samples. In Figure 7C, the direct-fusion model shows an improved risk distribution compared with the image-only model, with 190 conforming samples located in the low-risk region and 65 nonconforming samples located in the high-risk region, but the two classes are still not fully separated. The attention-based fusion variants also exhibit a more compact and better separated risk-score distribution. As shown in Figure 7D, SE-fusion assigns 69 nonconforming samples to the high-risk region while keeping 193 conforming samples in the low-risk region. Similarly, Figure 7E shows that CBAM-fusion assigns 70 nonconforming samples to the high-risk region and retains 193 conforming samples in the low-risk region. These results indicate that SE-fusion and CBAM-fusion improve the risk separation between conforming and

https://doi.org/10.3390/s26154857

nonconforming samples, although a small number of boundary samples still remain near or across the decision threshold. By comparison, the distribution of SPIMA-Net in Figure 7F is slightly more concentrated, with 194 conforming samples located in the low-risk region and 70 nonconforming samples located in the high-risk region, with only a few samples crossing the threshold. This indicates that the span-prior-guided gating mechanism can further improve the discrimination capability and stability of risk scoring.

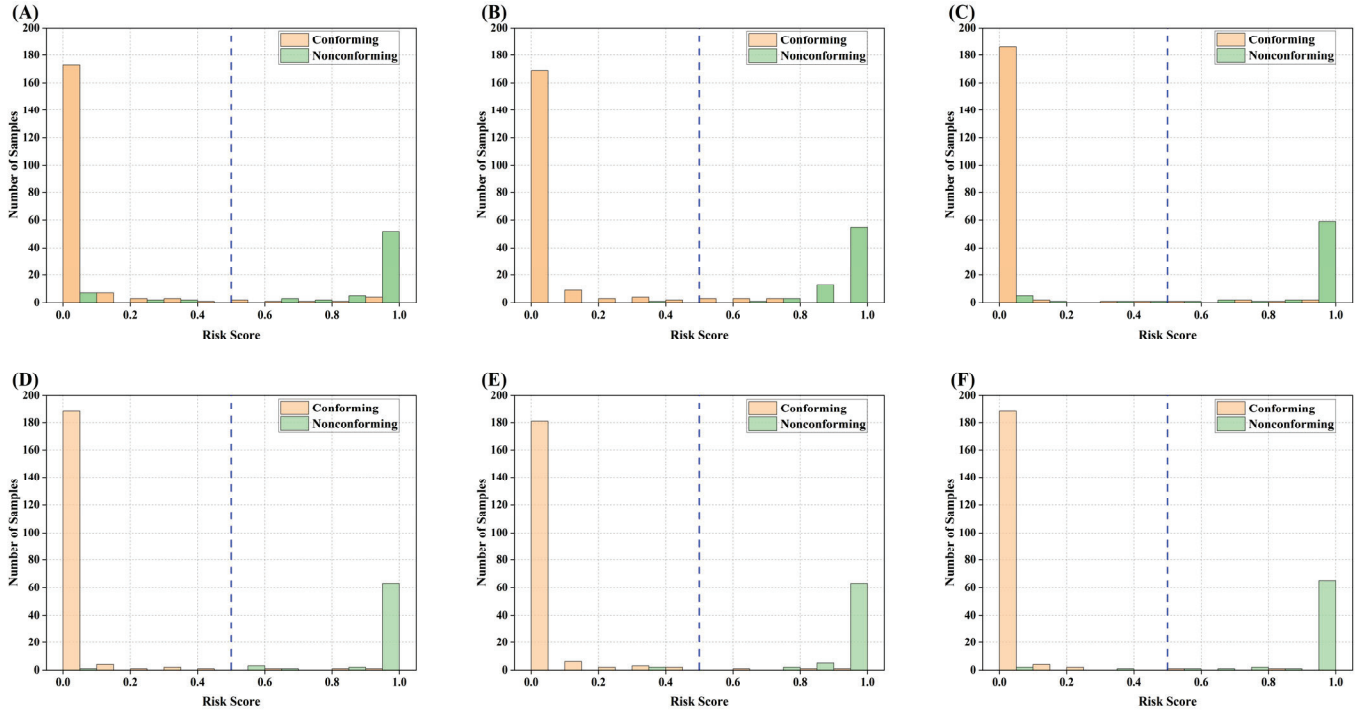


Figure 7. Risk-score distributions of the six compared models: (**A**) Image-only; (**B**) Span-only; (**C**) Direct-fusion; (**D**) SE-fusion; (**E**) CBAM-fusion; (**F**) SPIMA-Net.

Based on the risk-score distributions of the compared models shown in Figure 7, SPIMA-Net exhibits clear risk discrimination between conforming and nonconforming samples. On this basis, the calibration behavior of its predicted risk scores was further evaluated using the ECE and Brier score results in Table 4 together with the reliability diagram in Figure 8. As reported in Table 4, SPIMA-Net achieved the lowest ECE of 0.0200 ± 0.0114 and the lowest Brier score of 0.0147 ± 0.0090 among the compared models, indicating favorable calibration performance. Figure 8 further shows a clear calibration tendency of SPIMA-Net, particularly in the low-risk and high-risk regions. The local fluctuations in the observed nonconforming rate within the middle-probability intervals may be influenced by the uneven sample distribution across these bins. Therefore, the predicted risk score of SPIMA-Net can provide useful relative risk information for final-state quality inspection, while it should not be interpreted as a fully calibrated absolute probability under all deployment conditions.

Figure 9 presents representative Grad-CAM visualization results of SPIMA-Net. Specifically, Figures 9A and 9B show the Grad-CAM results of a nonconforming and a conforming sample of model type 3-4-6, respectively. Figures 9C and 9D show the Grad-CAM results of a nonconforming and a conforming sample of model type 6-6-5, respectively. Figure 9E shows the Grad-CAM result of a conforming sample of model type 6-6-5, and Figure 9F further provides the comparison between this sample and the terminal span-based quality assessment ROI. Warmer colors indicate stronger Grad-CAM responses, whereas cooler colors indicate weaker responses. Background color differences outside the activated re-

https://doi.org/10.3390/s26154857

gions are used only for visual distinction and are not involved in the evaluation. The red dashed boxes indicate the predefined ROI regions related to terminal span-based quality assessment. It can be observed that the high-response regions of SPIMA-Net are mainly concentrated around the span opening and the terminal pins on both sides, rather than being distributed over irrelevant background regions. This result indicates that, when distinguishing between conforming and nonconforming samples, the model can effectively focus on local morphological features associated with final-state span quality. The highlighted regions are therefore consistent with the key geometric regions considered in final-state quality inspection of hairpin windings, further supporting the interpretability of the proposed method.

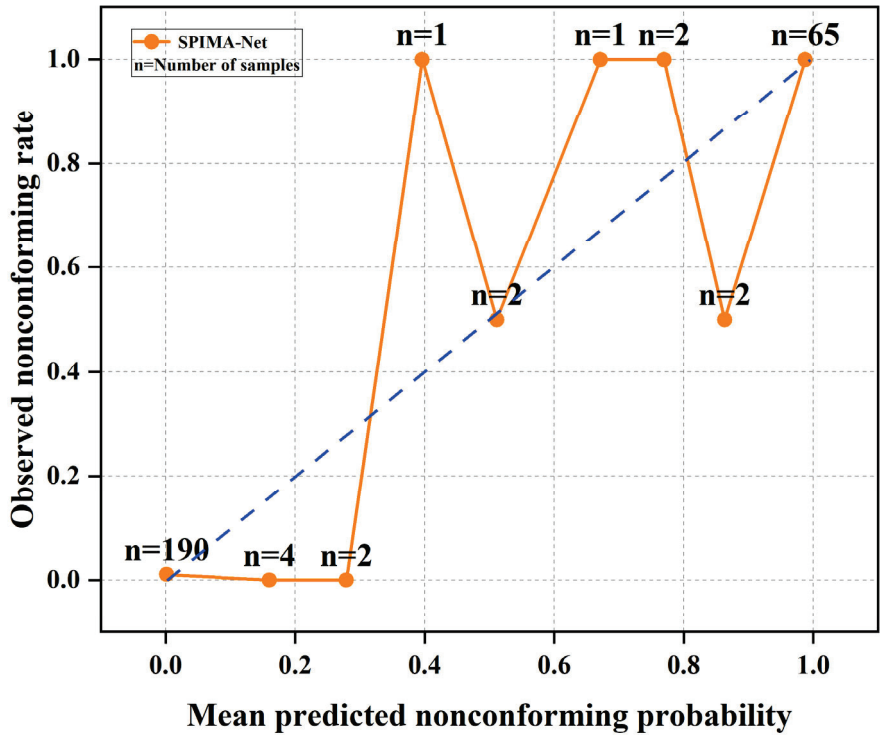


Figure 8. Reliability diagram of SPIMA-Net.

Based on the Grad-CAM visualization results in Figure 9, an ROI-based activation analysis was performed to quantify the image-side interpretability of SPIMA-Net. The terminal-span quality-assessment regions were defined as ROIs. The ROI activation ratio was calculated as the Grad-CAM activation sum within the ROI divided by the total Grad-CAM activation sum of the whole image, and the ROI enrichment factor was calculated as the ROI activation ratio divided by the ROI area ratio. Therefore, the ROI activation ratio reflects the proportion of model response located in the ROI, while the ROI enrichment factor reflects whether the response is concentrated in the ROI relative to its area.

As shown in Table 7, the representative nonconforming samples show substantially higher ROI activation ratios and ROI enrichment factors than the conforming samples. Based on the sample results in Table 7, the mean ROI activation ratio was 0.3550 for nonconforming samples and 0.0033 for conforming samples, while the mean ROI enrichment factor was 3.0091 for nonconforming samples and 0.0279 for conforming samples. These results quantitatively indicate that SPIMA-Net assigns stronger Grad-CAM responses to terminal-span-related ROI regions when identifying nonconforming final states.

The feature-group importance results of the span branch are presented in Table 8 and Figure 10. To provide a unified evaluation of the contribution of different span prior variables to model discrimination, a permutation-based feature-group importance analysis

https://doi.org/10.3390/s26154857

was conducted on the test set. Specifically, each span prior feature group was permuted, and the decreases in F1-score and AUC before and after perturbation were calculated. The mean F1 decrease and mean AUC decrease were then combined to obtain the Overall Importance Score. Since these span prior variables jointly describe the geometric deviation between the measured final span and the theoretical target, their importance scores are analyzed as feature-group-level sensitivity rather than as isolated single-variable effects. This treatment is consistent with the purpose of the analysis, which is to identify which type of span-related information has a greater influence on the classification performance of SPIMA-Net.

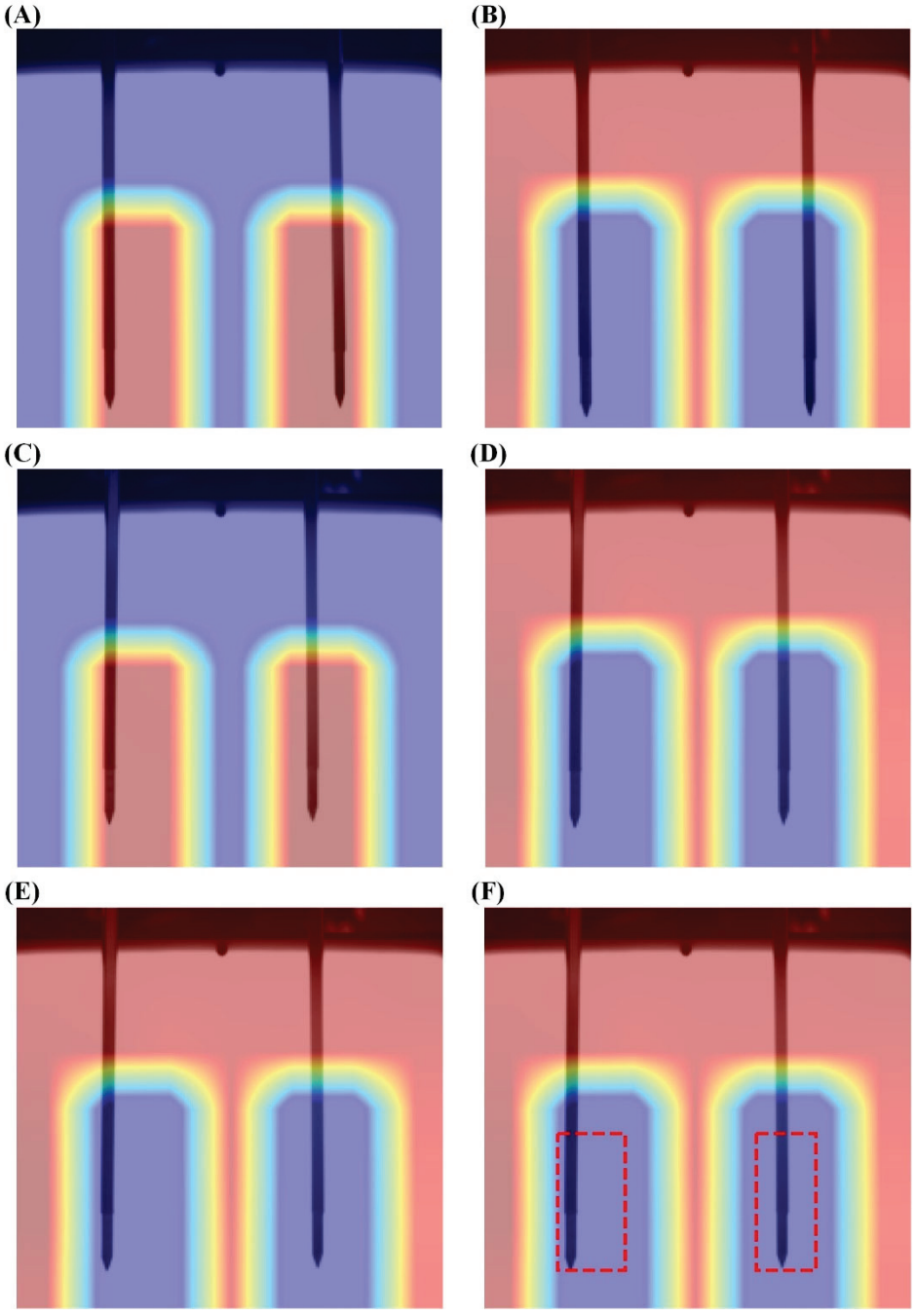


Figure 9. Grad-CAM visualization and terminal ROI comparison of SPIMA-Net: (**A**) nonconforming sample of model type 3-4-6; (**B**) conforming sample of model type 3-4-6; (**C**) nonconforming sample of model type 6-6-5; (**D**) conforming sample of model type 6-6-5; (**E**) Grad-CAM visualization of a conforming sample of model type 6-6-5; (**F**) terminal ROI comparison of the sample in (**E**).

https://doi.org/10.3390/s26154857

Table 7. Quantitative Grad-CAM ROI activation analysis of SPIMA-Net.

Hairpin Type	Sample Type	Distance	ROI Activation Ratio	ROI Enrichment Factor
6-6-5	Nonconforming	57.34	0.3551	3.0095
6-6-5	Nonconforming	57.34	0.3551	3.0095
3-4-6	Nonconforming	65.23	0.3550	3.0086
3-4-6	Nonconforming	65.32	0.3550	3.0086
6-6-5	Conforming	58.96	0.0033	0.0279
6-6-5	Conforming	58.99	0.0033	0.0279
3-4-6	Conforming	63.93	0.0033	0.0279
3-4-6	Conforming	64.41	0.0033	0.0279

Table 8. Ranking of feature importance in the span branch of SPIMA-Net.

Rank	Feature Name	Mean F1 Decrease	Mean AUC Decrease	Overall Importance Score
1	Relative Span Deviation	0.157395	0.020816	0.178211
2	Absolute Span Deviation	0.141929	0.015667	0.157596
3	Model-Type Information	0.018320	0.002411	0.020732
4	Measured Final Span	0.000840	0.000198	0.001038
5	Theoretical Span	−0.001337	0.000262	0.000262

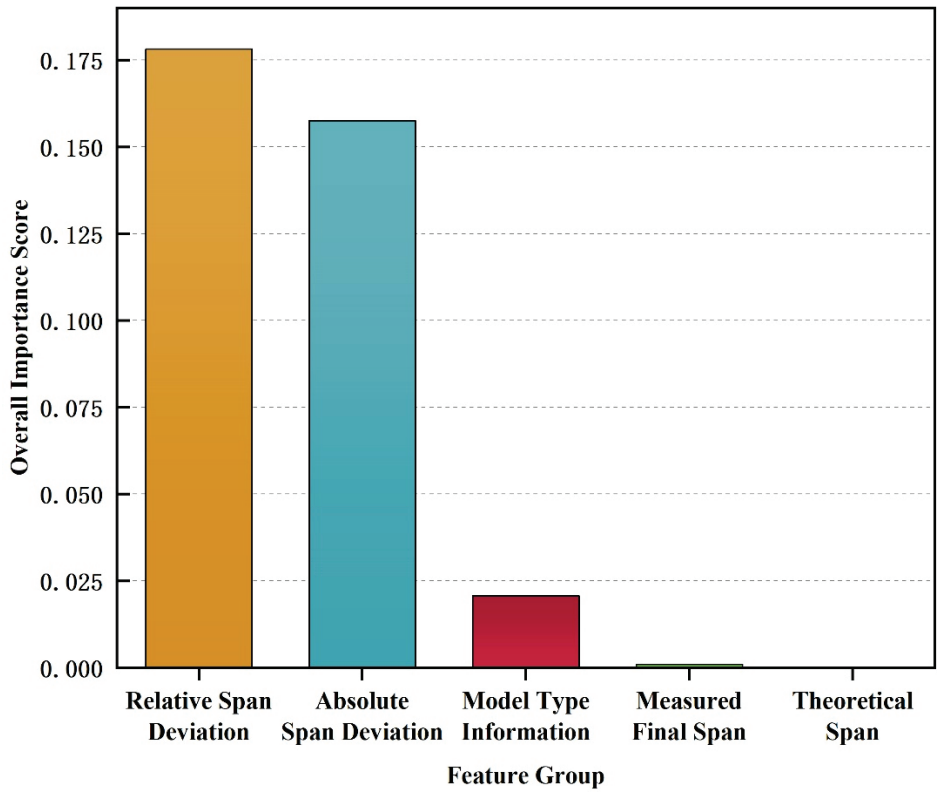


Figure 10. Feature importance of the span branch in SPIMA-Net.

As shown in Table 8 and Figure 10, relative span deviation and absolute span deviation achieve the highest Overall Importance Scores, reaching 0.1782 and 0.1576, respectively, which are substantially higher than those of the other features. This indicates that span-deviation-related information is the main mechanical geometric factor affecting nonconforming classification. The Overall Importance Score of model-type information is 0.0207, suggesting that differences in geometric boundaries among model types also

https://doi.org/10.3390/s26154857

influence model decisions. By contrast, the importance scores of the measured final span and theoretical span are relatively low, indicating that the model pays more attention to the deviation of the measured dimension from the theoretical target than to a single raw dimensional value.

Taken together, the above results show that the advantage of SPIMA-Net does not simply arise from the addition of input information but from the effective guidance of visual features by span priors. The proposed method achieves a balanced performance in classification accuracy, risk discrimination capability, and engineering interpretability, demonstrating its effectiveness for final-state quality inspection of hairpin windings.

4. Discussion

Existing studies on industrial visual inspection and multimodal fusion have shown that image representation and multisource information integration can improve complex quality inspection performance. However, for final-state quality inspection of hairpin windings with explicit geometric decision criteria, relying only on visual information or using conventional feature concatenation remains insufficient to capture the relationships among final-state morphology, span deviation, and quality status. The results of this study indicate that using span geometric priors as guiding signals for visual feature modulation is more effective for this type of mechanically formed quality inspection task.

This finding verifies the core hypothesis of this study, namely that span prior guidance can enhance the consistency between visual representation and quality classification. Span information provides evidence of key dimensional deviations, while final-state images supplement discriminative cues such as end profiles, opening regions, and local abnormal morphologies. The advantage of SPIMA-Net does not arise from simple stacking of input modalities. Instead, it relies on a collaborative decision mechanism. In this mechanism, geometric priors guide the decision direction, while visual features provide local evidence. In addition, Grad-CAM visualization and span feature importance analysis show that the model's attention regions and major decision factors are consistent with engineering inspection knowledge. This indicates that the proposed method not only improves inspection performance but also enhances the interpretability of the decision process. The proposed idea may provide a useful reference for other mechanically formed quality inspection tasks with explicit dimensional constraints, and its applicability can be further extended by incorporating richer three-dimensional geometric features.

In this study, the classification performance and interpretability of SPIMA-Net were evaluated using an industrial dataset covering seven hairpin winding model types with different geometric characteristics. All samples were collected from the same manufacturing environment, production line, production batch, and fixed imaging setup. Variations in independent production batches, different production lines, illumination conditions, camera positions, and backgrounds may lead to distribution shifts in both image features and span prior features, thereby affecting model performance, risk-score reliability, and deployment adaptability. Therefore, model evaluation under cross-batch, cross-line, and cross-imaging conditions will be an important direction of future work to further improve the robustness and industrial applicability of the proposed method.

5. Conclusions

This study proposes a span-prior-guided explainable multimodal neural network method and develops SPIMA-Net as its core model. The proposed method combines final-state images with span-related geometric priors and employs a span-prior-guided gating mechanism to modulate visual features, thereby achieving collaborative fusion

https://doi.org/10.3390/s26154857

between image representations and mechanical geometric information for final-state quality inspection and risk scoring of three-dimensional stamped hairpin windings.

Experimental results show that SPIMA-Net achieves an accuracy of 98.14%, a precision of 97.22%, a recall of 95.89%, an F1-score of 96.55%, and an AUC of 0.9983 on the test set. Compared with the image-only, span-only, direct-fusion, SE-fusion, and CBAM-fusion models, the F1-score for the nonconforming class is improved by 10.44, 3.04, 6.27, 1.38, and 0.66 percentage points, respectively, while the total number of misclassifications decreases to five. Although CBAM-fusion shows competitive performance close to SPIMA-Net in some evaluations, SPIMA-Net achieves more balanced overall performance in terms of classification accuracy, nonconforming-sample recognition, and misclassification control. Further analysis shows that SPIMA-Net maintains stable classification performance across different model types, with an overall test-set F1-score of 0.9655. Risk-score distributions, reliability analysis, Grad-CAM visualization, and span-feature importance analysis further verify the effectiveness of the proposed method in risk representation and engineering interpretability.

Although SPIMA-Net achieved high classification performance and interpretable decision evidence on the investigated industrial dataset covering seven hairpin winding model types, further model evaluation under independent production batches, different production lines, and different imaging conditions will be investigated in future work. For production deployment, an input-checking module can be introduced before SPIMA-Net inference to identify missing model-type information, abnormal span measurements, or corrupted images. When model-type or span-related information is unavailable, the system can switch to an image-only backup inference mode and flag the sample for manual review; when the image input is corrupted, image reacquisition should be triggered before quality judgment. Future work will combine three-dimensional geometric enhancement, multimodal representation learning, and cross-scenario adaptation mechanisms to further improve the robustness, extensibility, and long-term deployment potential of the proposed framework.

Author Contributions: Writing—original draft preparation, X.C.; conceptualization, B.Z.; funding acquisition, Z.G.; supervision, S.C.; project administration, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Jilin Provincial Scientific and Technological Development Program (No. 20250601069RC).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study are not publicly available due to confidentiality requirements of the industrial application scenario.

Conflicts of Interest: Author Jingru Liu was employed by the company FAW Tool & Die Manufacturing 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

1. Szeliski, R. *Computer Vision: Algorithms and Applications*, 2nd ed.; Springer Nature: Cham, Switzerland, 2022. [CrossRef]
2. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
3. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
4. Zou, T.J.; Gerada, D.; La Rocca, A.; Moslemin, M.; Cairns, A.; Cui, M.M. A comprehensive design guideline of hairpin windings for high power density electric vehicle traction motors. *IEEE Trans. Transp. Electrif.* **2022**, *8*, 3578–3593. [CrossRef]

https://doi.org/10.3390/s26154857

5. Fleischer, J.; Ceglarek, D.; Franke, J.; Herrmann, C. Production technologies and systems for electric mobility. *CIRP Ann.* **2025**, *74*, 1047–1072. [CrossRef]
6. Wirth, F.; Hausmann, L.; Fleischer, J. Model-based closed-loop process control for the manufacturing of hairpin coils. *Prod. Eng.* **2024**, *18*, 875–888. [CrossRef]
7. Yu, H.Y.; Yin, Y.S.; Ji, S.A.; Zhang, C.R.; Zhu, T.S.; Liu, Z.H. Laser welding method and quality analysis of hairpin windings based on visual recognition of gap and displacement matching process. *Opt. Laser Technol.* **2024**, *178*, 111224. [CrossRef]
8. Omlor, M.; Reith, J.; Steffen, F.; Schwarz, J.; Dilger, K. Inline quality assurance of hairpin welding using a machine learning approach based on photodiode signals and images. *J. Laser Appl.* **2025**, *37*, 032022. [CrossRef]
9. Omlor, M.; Seitz, N.; Butzmann, T.; Petrich, T.; Graf, R.; Hesse, A.C.; Dilger, K. Quality characteristics and analysis of input parameters on laser beam welding of hairpin windings in electric drives. *Weld. World* **2023**, *67*, 1491–1508. [CrossRef]
10. D'Arcangelo, S.; Caprio, L.; Chesi, D.; Nocciolini, D.; Corbinelli, R.; Previtali, B.; Demir, A.G. Comprehensive benchmarking of laser welding technologies including novel beam shapes and wavelengths for e-drive copper hairpins. *Opt. Laser Technol.* **2024**, *169*, 109964. [CrossRef]
11. Bono, F.M.; Radicioni, L.; Cinquemani, S. A novel approach for quality control of automated production lines working under highly inconsistent conditions. *Eng. Appl. Artif. Intell.* **2023**, *122*, 106149. [CrossRef]
12. Zhuo, L.; Yan, Y.H.; Wang, X.H.; Ge, Y.F.; Meng, L. A survey of deep learning for industrial visual anomaly detection. *Artif. Intell. Rev.* **2025**, *58*, 279. [CrossRef]
13. Zhu, X.M.; Mårtensson, P.; Hanson, L.; Björkman, M.; Maki, A. Automated assembly quality inspection by deep learning with 2D and 3D synthetic CAD data. *J. Intell. Manuf.* **2025**, *36*, 2567–2582. [CrossRef]
14. Zajec, P.; Rožanec, J.M.; Theodoropoulos, S.; Fontul, M.; Koehorst, E.; Fortuna, B.; Mladenić, D. Few-shot learning for defect detection in manufacturing. *Int. J. Prod. Res.* **2024**, *62*, 6979–6998. [CrossRef]
15. Lara-Chávez, J.A.; Avilés-Cruz, C.; Magos-Rivera, M. Deep learning model for automated visual inspection of electronic boards. *J. Intell. Manuf.* **2025**. [CrossRef]
16. Bhavanasi, G.; Neven, D.; Arteaga, M.; Ditzel, S.; Dehaeck, S.; Bey-Temsamani, A. Enhanced vision-based quality inspection: A multiview artificial intelligence framework for defect detection. *Sensors* **2025**, *25*, 1703. [CrossRef] [PubMed]
17. Nahar, L.; Awrangjeb, M.; Islam, M.S. AI-enabled defect detection in industrial products: A comprehensive survey, key insights and future research challenges. *Adv. Eng. Inform.* **2026**, *69*, 104067. [CrossRef]
18. Baltrušaitis, T.; Ahuja, C.; Morency, L.P. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 423–443. [CrossRef] [PubMed]
19. Zhao, F.; Zhang, C.C.; Geng, B.C. Deep multimodal data fusion. *ACM Comput. Surv.* **2024**, *56*, 216. [CrossRef]
20. Jiao, T.Z.; Guo, C.P.; Feng, X.Y.; Chen, Y.M.; Song, J. A comprehensive survey on deep learning multi-modal fusion: Methods, technologies and applications. *Comput. Mater. Contin.* **2024**, *80*, 1–35. [CrossRef]
21. Wang, Y.F.; Yu, W.Y.; Yu, C.Y.; Shi, H.; Li, W.L. A survey on multimodal learning for industrial diagnostics: A data dimensionality perspective. *Artif. Intell. Rev.* **2026**, *59*, 126. [CrossRef]
22. McKinney, M.; Garland, A.; Cillessen, D.; Adamczyk, J.; Bolintineanu, D.; Heiden, M.; Fowler, E.; Boyce, B.L. Unsupervised multimodal fusion of in-process sensor data for advanced manufacturing process monitoring. *J. Manuf. Syst.* **2025**, *78*, 271–282. [CrossRef]
23. Groiseheide, J.; Tao, Z.; Cetin, B.; Wolfschläger, D.; Dewulf, W.; Schmitt, R.H. Enabling multimodal sensor fusion in additive manufacturing using a voxel-based data model. *Prod. Eng.* **2025**, *19*, 1195–1216. [CrossRef]
24. Leng, J.W.; Wang, J.J.; Zhou, L.H.; Zhao, R.L.; Chen, C.; Zhang, D.; Zheng, S.A.; Liu, Q.; Shen, W.M. A review of multi-modal deep learning towards agentic smart manufacturing. *Adv. Eng. Inform.* **2026**, *70*, 104117. [CrossRef]
25. Schlemper, J.; Oktay, O.; Schaap, M.; Heinrich, M.; Kainz, B.; Glocker, B.; Rueckert, D. Attention gated networks: Learning to leverage salient regions in medical images. *Med. Image Anal.* **2019**, *53*, 197–207. [CrossRef] [PubMed]
26. Hu, J.; Shen, L.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [CrossRef] [PubMed]
27. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359. [CrossRef]
28. Senoner, J.; Netland, T.; Feuerriegel, S. Using explainable artificial intelligence to improve process quality: Evidence from semiconductor manufacturing. *Manag. Sci.* **2022**, *68*, 5704–5723. [CrossRef]
29. Müller, R.; Reindel, D.F.; Stadtfeld, Y.D. The benefits and costs of explainable artificial intelligence in visual quality control: Evidence from fault detection performance and eye movements. *Hum. Factors Ergon. Manuf. Serv. Ind.* **2024**, *34*, 396–416. [CrossRef]
30. Puthanveettil Madathil, A.; Luo, X.C.; Liu, Q.; Walker, C.; Madarkar, R.; Qin, Y. A review of explainable artificial intelligence in smart manufacturing. *Int. J. Prod. Res.* **2025**, *63*, 8654–8697. [CrossRef]

https://doi.org/10.3390/s26154857

31. Cação, J.; Santos, J.; Antunes, M. Explainable AI for industrial fault diagnosis: A systematic review. *J. Ind. Inf. Integr.* **2025**, *47*, 100905. [CrossRef]
32. Ahangar, M.N.; Farhat, Z.A.; Sivanathan, A.; Ketheesram, N.; Kaur, S. Explainable AI-driven quality and condition monitoring in smart manufacturing. *Sensors* **2026**, *26*, 911. [CrossRef] [PubMed]
33. Singh, S.A.; Desai, K.A. Integrating explainable artificial intelligence with vision-based surface defect detection systems for realizing zero-defect manufacturing. *Int. J. Prod. Res.* **2025**. [CrossRef]
34. Tzionis, G.; Mouratidis, P.; Kougka, G.; Gialampoukidis, I.; Vrochidis, S.; Kompatsiaris, I.; Vlachopoulou, M. A review of explainable AI methods and their application in manufacturing systems. *Discov. Appl. Sci.* **2025**, *8*, 52. [CrossRef]
35. 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, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [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.

https://doi.org/10.3390/s26154857

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Sensors Editorial Office
E-mail: sensors@mdpi.com
www.mdpi.com/journal/sensors



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest Editor. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editor and contributors and not of MDPI. MDPI disclaims 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 978-3-7258-9210-5