

sensors

Advanced Fault Diagnosis and Health Monitoring Techniques for Complex Engineering Systems

Edited by

Yongbo Li, Bing Li, Jinchun Ji and Hamed Kalhori

Printed Edition of the Special Issue Published in *Sensors*

Advanced Fault Diagnosis and Health Monitoring Techniques for Complex Engineering Systems

Advanced Fault Diagnosis and Health Monitoring Techniques for Complex Engineering Systems

Editors

Yongbo Li

Bing Li

Jinchen Ji

Hamed Kalhori

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin



Editors

Yongbo Li
School of Aeronautics,
Northwestern
Polytechnical University,
Xi'an, China

Bing Li
School of Aeronautics,
Northwestern
Polytechnical University,
Xi'an, China

Jinchen Ji
School of Mechanical and
Mechatronic Engineering,
Faculty of Engineering and
IT, University of Technology,
Sydney, Australia

Hamed Kalhori
School of Mechanical and
Mechatronic Engineering,
The University of Technology Sydney,
Sydney, Australia

Editorial Office

MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Sensors* (ISSN 1424-8220) (available at: https://www.mdpi.com/journal/sensors/special_issues/0U137VN8YS).

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

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

ISBN 978-3-0365-6462-3 (Hbk)

ISBN 978-3-0365-6463-0 (PDF)

Cover image courtesy of Unsplash.com at <https://unsplash.com/photos/jXd2FSvcRr8>.

© 2023 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license, which allows users to download, copy and build upon published articles, as long as the author and publisher are properly credited, which ensures maximum dissemination and a wider impact of our publications.

The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons license CC BY-NC-ND.

Contents

Yongbo Li, Bing Li, Jinchun Ji and Hamed Kalhori

Advanced Fault Diagnosis and Health Monitoring Techniques for Complex Engineering Systems

Reprinted from: *Sensors* **2022**, *22*, 10002, doi:10.3390/s222410002 1

Yoonjae Lee, Minho Jo, Gyoujin Cho, Changbeom Joo and Changwoo Lee

Impact of Sensor Data Characterization with Directional Nature of Fault and Statistical Feature Combination for Defect Detection on Roll-to-Roll Printed Electronics

Reprinted from: *Sensors* **2021**, *21*, 8454, doi:10.3390/s21248454 5

Jie Chen, Qingshan Xu, Yingchao Guo and Runfeng Chen

Aircraft Landing Gear Retraction/Extension System Fault Diagnosis with 1-D Dilated Convolutional Neural Network

Reprinted from: *Sensors* **2022**, *22*, 1367, doi:10.3390/s22041367 23

Qiang Pan, Tian He, Wendong Liu, Xiaofeng Liu and Haibing Chen

Study on a Fault Mitigation Scheme for Rub-Impact of an Aero-Engine Based on NiTi Wires

Reprinted from: *Sensors* **2022**, *22*, 1796, doi:10.3390/s22051796 43

Ghufran Aldawood and Hamzeh Bardaweel

Self-Powered Self-Contained Wireless Vibration Synchronous Sensor for Fault Detection

Reprinted from: *Sensors* **2022**, *22*, 2352, doi:10.3390/s22062352 59

Gang Mao, Zhongzheng Zhang, Sixiang Jia, Khandaker Noman and Yongbo Li

Partial Transfer Ensemble Learning Framework: A Method for Intelligent Diagnosis of Rotating Machinery Based on an Incomplete Source Domain

Reprinted from: *Sensors* **2022**, *22*, 2579, doi:10.3390/s22072579 83

Feng Duan, Shuai Zhang, Yinze Yan and Zhiqiang Cai

An Oversampling Method of Unbalanced Data for Mechanical Fault Diagnosis Based on MeanRadius-SMOTE

Reprinted from: *Sensors* **2022**, *22*, 5166, doi:10.3390/s22145166 103

Linlin Kou, Jiaxian Chen, Yong Qin and Wentao Mao

The Robust Multi-Scale Deep-SVDD Model for Anomaly Online Detection of Rolling Bearings

Reprinted from: *Sensors* **2022**, *22*, 5681, doi:10.3390/s22155681 119

Lei Hu, Ligui Wang, Yanlu Chen, Niaoqing Hu and Yu Jiang

Bearing Fault Diagnosis Using Piecewise Aggregate Approximation and Complete Ensemble Empirical Mode Decomposition with Adaptive Noise

Reprinted from: *Sensors* **2022**, *22*, 6599, doi:10.3390/s22176599 133

Samira Asadi, Mehrdad Moallem and G. Gary Wang

Simultaneous Sensor and Actuator Fault Reconstruction by Using a Sliding Mode Observer, Fuzzy Stability Analysis, and a Nonlinear Optimization Tool

Reprinted from: *Sensors* **2022**, *22*, 6866, doi:10.3390/s22186866 147

Lili Bykerk and Jaime Valls Miro

Vibro-Acoustic Distributed Sensing for Large-Scale Data-Driven Leak Detection on Urban Distribution Mains

Reprinted from: *Sensors* **2022**, *22*, 6897, doi:10.3390/s22186897 161

Jing Guo, Biao Ma, Tiangang Zou, Lin Gui and Yongbo Li
Composite Multiscale Transition Permutation Entropy-Based Fault Diagnosis of Bearings
Reprinted from: *Sensors* **2022**, *22*, 7809, doi:10.3390/s22207809 177

Editorial

Advanced Fault Diagnosis and Health Monitoring Techniques for Complex Engineering Systems

Yongbo Li ^{1,*}, Bing Li ¹, Jinchen Ji ² and Hamed Kalhori ³¹ School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China² School of Mechanical and Mechatronic Engineering, Faculty of Engineering and IT, University of Technology, Sydney, P.O. Box 123, Broadway, NSW 2007, Australia³ School of Mechanical and Mechatronic Engineering, The University of Technology Sydney, Sydney, NSW 2007, Australia

* Correspondence: yongbo@nwpu.edu.cn

1. Introduction

Fault diagnosis and health condition monitoring have always been critical issues in the engineering research community. Over the past decade, with the rapid development of artificially structured materials, advanced sensing and data-driven intelligence algorithms, fascinating technical possibilities have been reported in the area of fault diagnosis and in the health condition monitoring of complex engineering systems. However, with the development of highly efficient intelligent algorithms, recent fault diagnosis and health monitoring strategies have become highly automated and are encountering sophisticated problems in terms of data availability, computational complexity, accuracy, etc. Meanwhile, combined with advanced intelligent algorithms, flourishing developments such as new sensing techniques, diagnostic approaches and the design of new types of metamaterials have also enabled significant advances and emerging opportunities in the field of system health condition monitoring. These studies will doubtlessly promote the reliability, availability and robustness of systems for the fault diagnosis and health monitoring of complex engineering systems.

This Special Issue of *Sensors* aims to collect research works encompassing the whole area of fault diagnosis and health monitoring techniques for engineering systems. This collection contains a total of 11 papers representing the current status of the research related to different methods of monitoring the health and reliability of engineering systems.

Targeting the limitations of the original transition permutation entropy (TPE) method, Guo et al. [1] propose a multiscale transition permutation entropy (MTPE) method. Furthermore, considering the weaknesses of the proposed multiscale approach, the feature extraction ability of the MTPE method is further improved by proposing a composite multiscale transition permutation entropy (CMTPE) approach. Lastly, the researchers input the features extracted using the CMTPE method into an extreme learning machine (ELM) to perform the fault diagnosis of a bearing.

Bykerk et al. [2] used vibro-acoustic sensors for detecting leaks in the water distribution mains of an urban area. The real-time data collected from the extensive deployment of the vibro-acoustic sensors across a sprawling metropolitan city were used to monitor the presence and absence of pipe leaks using a convolutional neural network (CNN) after pre-processing via short-time Fourier transform (STFT). Different external factors, such as pipeline size, pipeline material and the soil condition around the pipe, are also taken into consideration.

Asadi et al. [3] designed a Takagi–Sugeno (TS) fuzzy-based sliding mode observer (SMO) to reconstruct the faults in actuators and sensors installed in a nonlinear system subjected to unknown external disturbance. A non-quadratic Lyapunov function (NQLF) and fmincon function were used to guarantee the stability of the proposed SMO as a matlab

Citation: Li, Y.; Li, B.; Ji, J.; Kalhori, H. Advanced Fault Diagnosis and Health Monitoring Techniques for Complex Engineering Systems. *Sensors* **2022**, *22*, 10002. <https://doi.org/10.3390/s222410002>

Received: 10 November 2022

Accepted: 11 November 2022

Published: 19 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

optimization tool. The influence of unknown disturbances and uncertainties are minimized by utilizing a performance criterion. The proposed method provides better accuracy, less conservative optimization conditions and improved generality in comparison to other existing state-of-the-art methods.

Hu et al. [4] combined piecewise aggregate approximation (PAA) with complete ensemble empirical mode decomposition (CEEMDAN) to alleviate the high memory requirements and low computational efficiency of the CEEMDAN method in bearing fault diagnosis. Vibration signals were used to study the efficacy of the proposed method. An enhanced bearing fault diagnosis performance was obtained using the proposed method.

Aiming to solve the problem of unavailable data in online fault detection in rolling element bearings, a multiscale deep support vector data description (Deep-SVDD) approach is proposed by Kou et al. [5] By utilizing data enhancement technology, training data were transformed into multiple subspaces. Then, a subsequent clustering algorithm was utilized to enhance the robustness of the features. Lastly, the proposed Deep-SVDD model was constructed to achieve the online monitoring of the health of rolling element bearings. The proposed method can be utilized to detect incipient faults in a bearing.

A new oversampling algorithm, namely, MeanRadius-SMOTE, is proposed by Duan et al. [6] for diagnosing mechanical faults regarding unbalanced data. The newly proposed method can effectively avoid the generation of useless and noisy samples and solve the multiclassification problem regarding different mechanical faults. A complete diagnosis of the faults in mechanical equipment can be achieved using the proposed method.

Mao et al. [7] addressed the challenges of an incomplete training dataset using a cross-domain intelligent fault diagnosis approach and proposed a novel deep learning approach called the partial transfer ensemble learning framework (PT-ELF). After substituting the missing health states with another dataset, the proposed method was able to address the variable data distribution challenge by training a weak global classifier and two partial domain adaption classifiers. Lastly, a specific ensemble strategy was used to combine these classifiers for fault diagnosis.

Aldawood et al. [8] developed a self-vibration-powered energy harvester sensor system to tackle the environmental threat posed by unused batteries in battery-powered sensors in wireless sensor networks (WSN). Dual moving magnets bordered by coil windings were used for power and signal generation in a harvester sensor unit. A radio frequency (RF) transmitter is operated using the power generated from the harvester, and the generated signal from this sensor is transmitted as the vibration signal. Lastly, a custom-made APP is utilized to detect faults in this system.

A 1D dilated convolutional neural network (1-DDCNN) is proposed by Chen et al. [9] for the fault diagnosis in an aircraft retraction/extension (R/E) system. Aiming to solve the limited feature information extraction and fault diagnosis ability of 1-DCNNs, multiple feature parameters have been used. Moreover, the main fault mode of the R/E system for aircraft landing gears has been studied, specifically exploring its working principal and the influence of convolutional kernel size on the classification accuracy.

Lee et al. [10] studied the optimal sensor selection criteria in a multi-sensor-based fault diagnosis of a roll-to-roll printed electronics system. Data are collected for four major defects of a Gravure roll-to-roll printed electronic system with three triaxial acceleration signals. Smart data were formed from the collected raw data obtained by a sensor data efficiency evaluation; a sensitivity evaluation for axis selection considering the directional nature of faults; and feature variable optimization using the feature combination matrix method. The progressive application of the aforementioned phases enhanced the fault diagnosis results in terms of accuracy, calculation time, predictive ability and data storage capacity.

Pan et al. [11] proposed a new method to investigate the mitigation of commonly occurring rotor–stator rub impact faults in aero-engines. A pre-strained, two-way shape memory alloy (SMA) wire was used in the design of a current-driven active control actuator

to mitigate the occurrence of rub impact faults. The feasibility of the proposed scheme is verified by different properties of the used NiTi wires. Finally, a prototype of the schemed actuator was designed and manufactured for testing under various conditions. The status of the rub impact fault was monitored using an acoustic emission sensor.

On behalf of all the editors of this Special Issue, we would like to extend our heartiest gratitude for the contributions from the authors to this project. We would also like to extend our sincere thanks to all the reviewers and members of the editorial board of *Sensors*.

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

References

1. Guo, J.; Ma, B.; Zou, T.; Gui, L.; Li, Y. Composite Multiscale Transition Permutation Entropy-Based Fault Diagnosis of Bearings. *Sensors* **2022**, *22*, 7809. [[CrossRef](#)] [[PubMed](#)]
2. Bykerk, L.; Valls Miro, J. Vibro-Acoustic Distributed Sensing for Large-Scale Data-Driven Leak Detection on Urban Distribution Mains. *Sensors* **2022**, *22*, 6897. [[CrossRef](#)] [[PubMed](#)]
3. Asadi, S.; Moallem, M.; Wang, G.G. Simultaneous Sensor and Actuator Fault Reconstruction by Using a Sliding Mode Observer, Fuzzy Stability Analysis, and a Nonlinear Optimization Tool. *Sensors* **2022**, *22*, 6866. [[CrossRef](#)] [[PubMed](#)]
4. Hu, L.; Wang, L.; Chen, Y.; Hu, N.; Jiang, Y. Bearing Fault Diagnosis Using Piecewise Aggregate Approximation and Complete Ensemble Empirical Mode Decomposition with Adaptive Noise. *Sensors* **2022**, *22*, 6599. [[CrossRef](#)] [[PubMed](#)]
5. Kou, L.; Chen, J.; Qin, Y.; Mao, W. The Robust Multi-Scale Deep-SVDD Model for Anomaly Online Detection of Rolling Bearings. *Sensors* **2022**, *22*, 5681. [[CrossRef](#)] [[PubMed](#)]
6. Duan, F.; Zhang, S.; Yan, Y.; Cai, Z. An Oversampling Method of Unbalanced Data for Mechanical Fault Diagnosis Based on MeanRadius-SMOTE. *Sensors* **2022**, *22*, 5166. [[CrossRef](#)] [[PubMed](#)]
7. Mao, G.; Zhang, Z.; Jia, S.; Noman, K.; Li, Y. Partial Transfer Ensemble Learning Framework: A Method for Intelligent Diagnosis of Rotating Machinery Based on an Incomplete Source Domain. *Sensors* **2022**, *22*, 2579. [[CrossRef](#)] [[PubMed](#)]
8. Aldawood, G.; Bardaweel, H. Self-Powered Self-Contained Wireless Vibration Synchronous Sensor for Fault Detection. *Sensors* **2022**, *22*, 2352. [[CrossRef](#)] [[PubMed](#)]
9. Chen, J.; Xu, Q.; Guo, Y.; Chen, R. Aircraft Landing Gear Retraction/Extension System Fault Diagnosis with 1-D Dilated Convolutional Neural Network. *Sensors* **2022**, *22*, 1367. [[CrossRef](#)] [[PubMed](#)]
10. Lee, Y.; Jo, M.; Cho, G.; Joo, C.; Lee, C. Impact of Sensor Data Characterization with Directional Nature of Fault and Statistical Feature Combination for Defect Detection on Roll-to-Roll Printed Electronics. *Sensors* **2021**, *21*, 8454. [[CrossRef](#)] [[PubMed](#)]
11. Pan, Q.; He, T.; Liu, W.; Liu, X.; Chen, H. Study on a Fault Mitigation Scheme for Rub-Impact of an Aero-Engine Based on NiTi Wires. *Sensors* **2022**, *22*, 1796. [[CrossRef](#)] [[PubMed](#)]

Article

Impact of Sensor Data Characterization with Directional Nature of Fault and Statistical Feature Combination for Defect Detection on Roll-to-Roll Printed Electronics

Yoonjae Lee ¹, Minhjo Jo ¹, Gyoujin Cho ², Changbeom Joo ³ and Changwoo Lee ^{4,*}

¹ Department of Mechanical Production Engineering and Design, Konkuk University, Seoul 05030, Korea; dldbswp913@konkuk.ac.kr (Y.L.); als8080@konkuk.ac.kr (M.J.)

² Research Engineering Center for R2R Printed Flexible Computer, Department of Biophysics, Institute of Quantum Biophysics, Sungkyunkwan University, Suwon-si 16419, Korea; gcho1004@skku.edu

³ Department of Mechanical Engineering, Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ 07030, USA; cjoo@stevens.edu

⁴ Department of Mechanical and Aerospace Engineering, Konkuk University, Seoul 05030, Korea

* Correspondence: leewoo1220@konkuk.ac.kr; Tel.: +82-2-450-3570; Fax: +82-2-454-0428

Citation: Lee, Y.; Jo, M.; Cho, G.; Joo, C.; Lee, C. Impact of Sensor Data Characterization with Directional Nature of Fault and Statistical Feature Combination for Defect Detection on Roll-to-Roll Printed Electronics. *Sensors* **2021**, *21*, 8454. <https://doi.org/10.3390/s21248454>

Academic Editors: Yongbo Li, Bing Li, Jinchun Ji and Hamed Kalhori

Received: 19 November 2021

Accepted: 14 December 2021

Published: 18 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Gravure printing, which is a roll-to-roll printed electronics system suitable for high-speed patterning of functional layers have advantages of being applied to flexible webs in large areas. As each of the printing procedure from inking to doctoring followed by ink transferring and setting influences the quality of the pattern geometry, it is necessary to detect and diagnose factors causing the printing defects beforehand. Data acquisition with three triaxial acceleration sensors for fault diagnosis of four major defects such as doctor blade tilting fault was obtained. To improve the diagnosis performances, optimal sensor selection with Sensor Data Efficiency Evaluation, sensitivity evaluation for axis selection with Directional Nature of Fault and feature variable optimization with Feature Combination Matrix method was applied on the raw data to form a Smart Data. Each phase carried out on the raw data progressively enhanced the diagnosis results in contents of accuracy, positive predictive value, diagnosis processing time, and data capacity. In the case of doctor blade tilting fault, the diagnosis accuracy increased from 48% to 97% with decreasing processing time of 3640 s to 16 s and the data capacity of 100 Mb to 5 Mb depending on the input data between raw data and Smart Data.

Keywords: defect detection; Directional Nature of Fault; gravure printing; fault diagnosis; roll-to-roll printed electronics; sensor data characterization

1. Introduction

Roll-to-roll processing is highly advantageous because it results in multiple functional layers of electronic circuitry printed on large flexible materials (i.e., web) [1–3]. Gravure printing is the desirable mode for fabricating these printed electronic devices, owing to its characteristic high-speed patterning of component layers [4–6]. Gravure printing can be classified into the following four phases: inking, doctoring, ink transfer, and ink setting [7,8]. Printing defects can be generated by undesired printing conditions and ink characteristics during each printing phase [9–12]. For example, during the doctoring phase, the misalignment of the doctor blade at either side can degrade the ink uniformity in the engraved patterns in the width direction (i.e., transverse direction (TD)). Moreover, non-uniform nip roll pressure can negatively affect the uniformity of the pattern thickness in the TD. To derive high-quality patterns with uniform thickness using the roll-to-roll gravure printing process, it is necessary to recognize and diagnose these.

In this study, a method of data characterization using sensor data efficiency evaluation (S_E), directional nature of fault (DNF), and feature combination matrix (FCM) is proposed to diagnose these major faults. The aim is to recognize defects in advance and improve the

diagnosis results by optimizing the training (input) data acquired from multiple sensors for the machine-learning fault diagnosis model. We find that the misalignment of the doctor blade, eccentricity of the nip and printing rolls, and non-uniform nip pressure can be indirectly measured via the vibration of the doctor blade, the nip roll, and the frames supporting the printing module. Through the acquisition of vibration data using multiple sensors, a vibration dataset (i.e., Raw data) is acquired. The smart data clearly show the characteristics of the vibration caused by the factors mentioned above, and they are selected from the raw dataset using the proposed methods in three phases to maximize performance efficiency. The evaluation criteria include diagnosis accuracy, positive predictive value (PPV), processing time for diagnosis, and data capacity. The performance of the machine-learning model developed using smart data was compared to that of the model just using the raw dataset.

With significant growth of industrial machines, recent studies have raised concerns regarding the maintenance of operating conditions. Profound interest in the fields of fault diagnosis based on data acquisition of sensors has been shown in recent research. Xia et al. presented convolutional neural network-based feature extraction approaches for fault diagnosis of rotating machines with multiple sensors [13]. Duan et al. have reviewed fields of fault diagnosis and condition monitoring based on multi-sensors for rolling bearings by presenting foundational knowledge [14]. Studies with multirate data and sensors for fault diagnosis by feature extracting deep learning models has been carried out by Zhao et al. and Huang et al. [15,16]. Research for fault diagnosis based on data optimization in recent studies has been shown by Bazan et al. and Wang et al. [17,18]. Lee et al. proposed quantification methods of fault features for rotary machine fault diagnosis. Most studies regarding fault diagnosis have shown methods of feature extraction to improve the results of machine learning from the data acquisition of sensors.

As shown in studies abovementioned, diagnosing the abnormal conditions with multiple sensors show promising results of fault diagnosis; however, the efficiency of diagnosis performance is without consideration. As studies focus on methods or strategies to conclude in diagnosis, this paper proposes methods to optimize multiple sensor data by selecting an optimal sensor. Furthermore, in comparison with Bazan et al., the performances of diagnosis results regarding accuracy, and data reduction stretch to positive predictive value and diagnosis processing time [17]. Related to Lee et al., this paper proposes strategies based on quantification methods to evaluate the efficiency of each phase [19].

2. Methodology of Data Characterization

2.1. Procedure of Data Characterization from Raw Data to Smart Data

Procedure of data characterization is led with data acquisition with three acceleration sensors which are attached to the doctor's blade and the frame of the gravure printing system. Each sensor is capable of acquiring data with three axes. Then on, experimentally acquired raw data is achieved in three phases, as shown in Figure 1. During Phase 1, the acquired sensor data are evaluated for efficiency (S_E), and the most efficient (optimal) sensor is chosen for DNF processing in Phase 2 to extract the most sensitive of three axes from the sensor. Then, a list of feature variables is tallied for training data using the FCM method in Phase 3. Finally, the processed smart data are used as input to the machine-learning fault diagnosis model to classify the printing process operating conditions during the major fault occurrences. Further description of smart data characterization through Phases 1–3 will be extensively illustrated in detail in Sections 2.2–2.4.

2.2. Sensor Data Efficiency Evaluation

The optimally efficient sensor is selected using an evaluation procedure based on Equation (1), which leverages three variables. α is the ratio of the data capacity between raw data and single-sensor data. β is the ratio of the data processing time, and γ is the ratio of the misclassification rate. Likewise, β and γ is a ratio between raw data and single-sensor data. Since the value of S_E in Equation (1) is dependent on the ratio of three variables

of two comparing data, the sensor rating the highest S_E is selected as the optimal single sensor. In other words, a sensor with the clearest distinction to the raw data in three aspects abovementioned is likely to score the highest S_E .

$$S_E = \frac{\alpha + \beta}{2\gamma} \tag{1}$$

In the case of this experiment, the diagnosis results from the raw data of three triaxial sensors were compared.

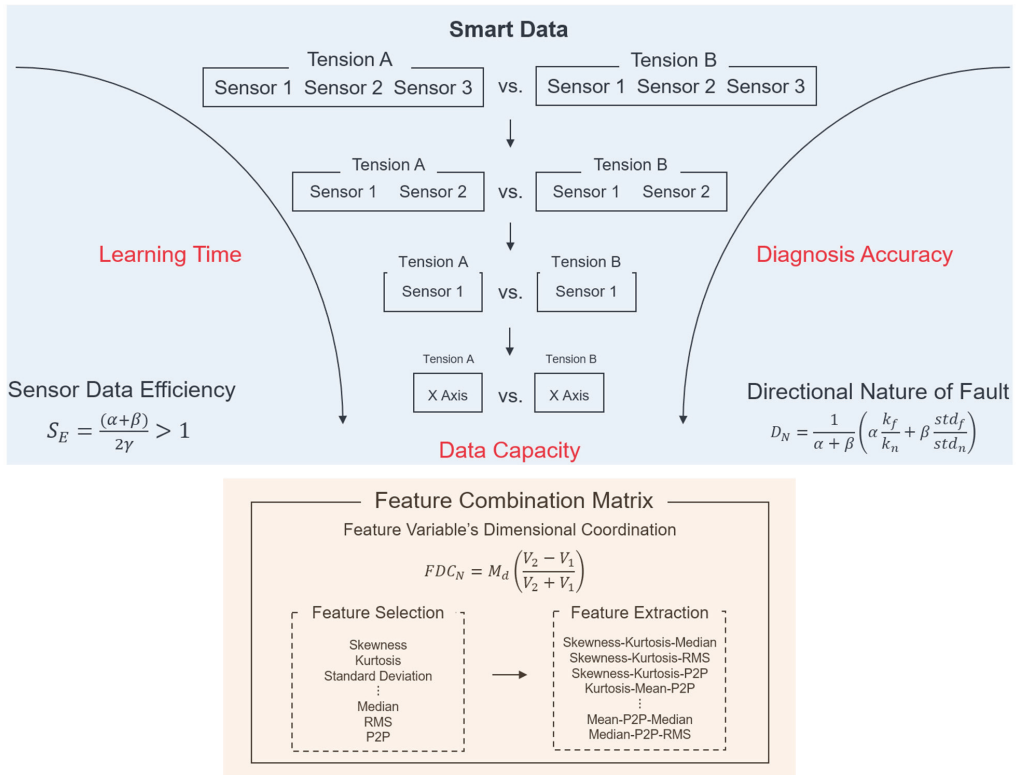


Figure 1. Smart data characterization procedure from raw data.

2.3. Directional Nature of Fault

The DNF method extracts valid data from raw data by evaluating the sensitivity of the axial information from a single sensor. After Phase 1, the DNF method evaluates axes X, Y, and Z to extract valid data for fault diagnosis. The DNF method is defined in Equation (2), where α and β are weight factors defining the relative ratio between kurtosis and standard deviation. k_f and k_n are the kurtosis of the fault and normal conditions, respectively. std_f and std_n are the standard deviation of the fault and normal conditions, respectively. Based on the probability distribution curve, the standard deviation of the abnormal condition data has a wide distribution of data points [20,21]. The kurtosis of an abnormal condition has an imbalanced distribution [22]. The DNF number based on

Equation (2) can thus be acquired from each axis. The axis with the highest DNF number defines the most sensitive and valid data for training.

$$D_N = \frac{1}{\alpha + \beta} \left(\alpha \frac{k_f}{k_n} + \beta \frac{std_f}{std_n} \right) \tag{2}$$

2.4. Feature Combination Matrix

The FCM method selects and extracts statistical feature variables. As shown in Figure 2, feature extraction is performed when the list of statistical feature variables is acquired from the dataset from Phase 2 [23,24]. The extracted features are then combined into the three features of a three-dimensional volume. As mentioned in Section 2.3, based on a normal distribution, the distribution of data points is likely to be imbalanced, broad, skewed, or irregular [25–27]. Comparing the volume acquired from the combination of the three features, the volume of the normal condition data is smaller than that of the abnormal condition. Hence, the combination producing the largest difference between the two volumes of different conditions reflects higher classification accuracy. The distance between the two datasets is also a factor that improves classification performance because it distinguishes between normal and abnormal conditions. The Mahalanobis distance is applied to evaluate the distance between two datasets in a multivariate space, including correlated points for multiple variables, considering the densities of the datasets [28–31]. Using the volumes of normal/abnormal feature combinations and the Mahalanobis distance feature variables, the Feature Variable’s Dimensional Coordination number (FDC_N) can be obtained. As shown in Equation (3), the FDC_N evaluates the combination of extracted features to ranks them according to efficiency. V_1 represents the volume of the normal condition feature combination, V_2 represents the volume of the abnormal condition feature combination, and M_d represents the Mahalanobis distance between V_1 and V_2 .

$$FDC_N = M_d \left(\frac{V_2 - V_1}{V_2 + V_1} \right) \tag{3}$$

The selected feature combination through evaluation of the FDC_N is then applied to be used as training data for developing a machine learning fault diagnosis model.

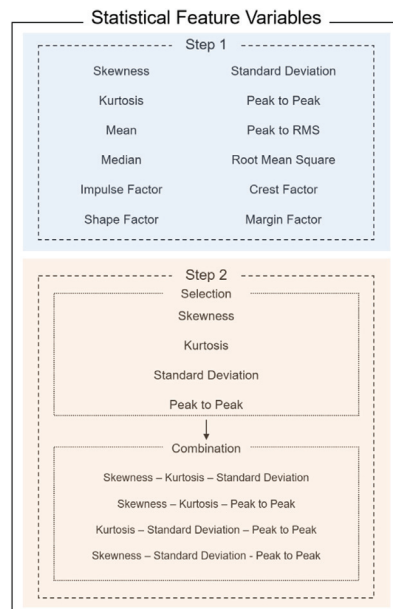


Figure 2. Feature engineering for feature combination matrix.

3. Experimental Data Acquisition

The experimental data acquisition for major fault diagnosis of the gravure printing system is shown in Figure 3. As shown, acceleration Sensors 1, 2, and 3 were installed on both sides of the doctor blade and the frame supporting the printing module. All sensor outputs were obtained using a data acquisition module (NI-9230, National Instruments). Table 1 lists the specifications of the acceleration sensor and the NI-9230 module. When the sensors obtained the vibration data, they were transferred to the LabVIEW software to monitor and save the acquired data.

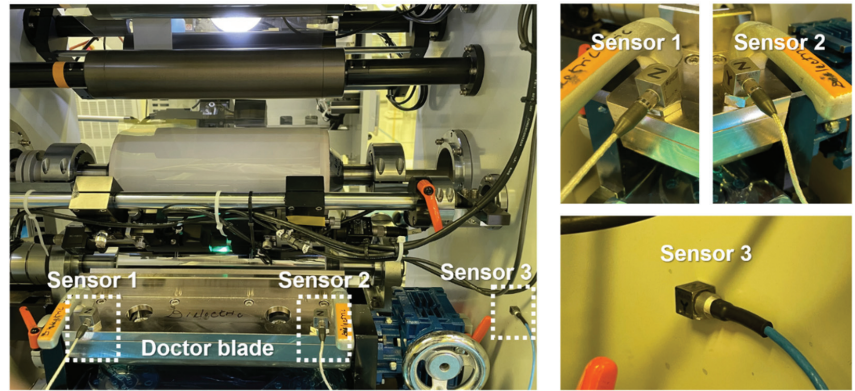


Figure 3. Experimental data acquisition by sensor position designation within the printing section.

Table 1. Specifications of acceleration sensor and NI-9230 module.

| Item | Parameter | Value |
|--------|---|-------------------------------|
| Sensor | Sensor Type | Share Accelerometer, Triaxial |
| | Sensitivity [$\text{mV}/(\text{m}/\text{s}^2)$] | 5.15 |
| | Measurement Range [m/s^2] | ± 1000 g peak |
| | Frequency Range [Hz] | 2–5000 Hz |
| | Resolution [m/s^2] | 0.003 (rms) |
| Module | Sampling Rate Range [Hz] | 0–1651.6 |
| | Sampling Time per Trial [s] | 36,000 |

The possible main faults during the printing process of the gravure printing system are shown in Figure 4. The four main faults of the experimental design include doctor blade tilting, printing roll eccentricity, nip roll eccentricity, and nip force non-uniformity. To detect the main faults for diagnosis, the experimental variables included the doctor blade, nip force, and tension. Cases with and without doctoring, and cases with and without nipping were tested under tensions of 2, 4, and 6 kgf. Regarding the nip force, the nipping cases were tested under 5 and 10 kgf, as shown in Table 2.

As shown in Table 3, each case was tested under different tension, nip force, and doctoring conditions. The data used for diagnosing the doctor blade tilting fault required Cases 1 and 2 at an operating tension of 2 kgf, Cases 7 and 8 at an operating tension of 4 kgf, and Cases 13 and 14 at an operating tension of 6 kgf. Cases 1, 7, and 13 had different operating tensions; however, they were tested without and without doctoring. Cases 2, 8, and 14 also had different operating tensions with and without doctoring. The data for the fault diagnosis of the doctor blade tilting fault were acquired from the comparison of each case at the same operating tension. The data for diagnosis printing roll eccentricity were acquired from Cases 1, 7, and 13, which lack nipping and doctoring. Case comparison for nip roll eccentricity required conditions without doctoring; hence, Cases 5, 11, and 17 with a nip force of 10 kgf were compared to cases 1, 7, and 13. Nip force non-uniformity cases

were selected using the nip force data shown in Figure 5. Cases 9 and 15 with uniform nip forces were compared to Cases 11 and 17.

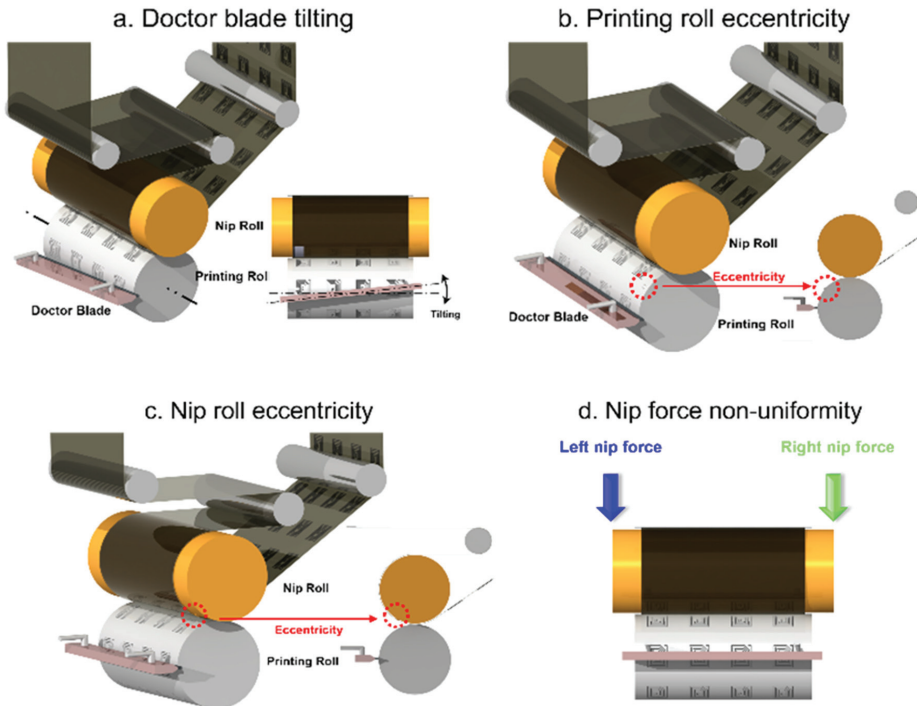


Figure 4. Possible main faults during gravure printing process: (a) Doctor blade tilting fault; (b) Printing roll eccentricity fault; (c) Nip roll eccentricity fault; and (d) Nip force non-uniformity fault.

Table 2. Experimental design of data acquisition with experimental variables of tension, nip force, and doctoring.

| Case No. | Tension [kgf] | Nip Force [kgf] | Doctoring |
|----------|---------------|-----------------|-------------------|
| 1 | 2 | Without Nipping | Without Doctoring |
| 2 | | Without Nipping | With Doctoring |
| 3 | | 5 | Without Doctoring |
| 4 | | 5 | With Doctoring |
| 5 | | 10 | Without Doctoring |
| 6 | | 10 | With Doctoring |
| 7 | 4 | Without Nipping | Without Doctoring |
| 8 | | Without Nipping | With Doctoring |
| 9 | | 5 | Without Doctoring |
| 10 | | 5 | With Doctoring |
| 11 | | 10 | Without Doctoring |
| 12 | | 10 | With Doctoring |
| 13 | 6 | Without Nipping | Without Doctoring |
| 14 | | Without Nipping | With Doctoring |
| 15 | | 5 | Without Doctoring |
| 16 | | 5 | With Doctoring |
| 17 | | 10 | Without Doctoring |
| 18 | | 10 | With Doctoring |

Table 3. Case comparison for fault diagnosis of possible main faults during printing process of gravure printing system.

| Case No. | Tension [2 kgf] | Tension [4 kgf] | Tension [6 kgf] |
|----------------------------|-------------------|--------------------|---------------------|
| Doctor Blade Tilting | Case 1 vs. Case 2 | Case 7 vs. Case 8 | Case 13 vs. Case 14 |
| Printing Roll Eccentricity | Case 1 | Case 7 | Case 13 |
| Nip Roll Eccentricity | Case 1 vs. Case 5 | Case 7 vs. Case 11 | Case 13 vs. Case 17 |
| Nip Force Non-Uniformity | - | Case 9 vs. Case 11 | Case 15 vs. Case 17 |

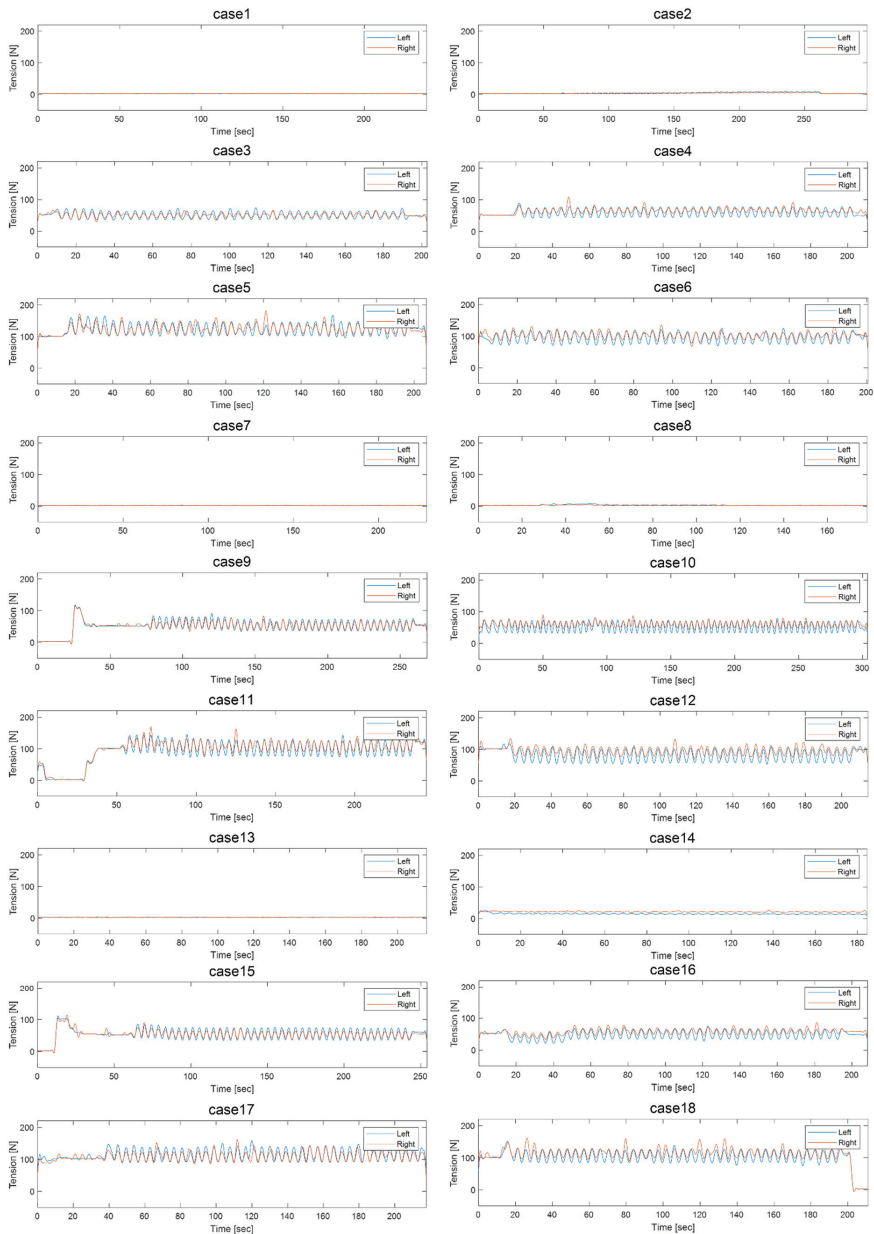


Figure 5. Nip force uniformity data of Cases 1–18.

4. Results

4.1. Doctor Blade Tilting Fault Diagnosis

4.1.1. Doctor Blade Tilting Fault Diagnosis Based on Raw Data

In this section, the fault diagnosis results of the doctor blade tilting fault based on the raw data are presented in Table 4. The raw data in this case include all data acquired from Sensors 1, 2, and 3. The diagnosis of a doctor blade tilting fault at an operating tension of 2 kgf showed 58.2 with a diagnosis accuracy of 1508.9 s and a processing capacity of 115 Mb. For a tension of 4 kgf, accuracy rates of 48.1% at 3640.4-s processing time required 100-Mb data capacity. At a tension of 6 kgf, the accuracy of fault diagnosis rates was 67.2%, which was the highest among tensions by 368.4 s with 113-Mb data size.

Table 4. Doctor blade tilting fault diagnosis based on raw data (i.e., Sensors 1, 2, and 3).

| [SVM] | 2 kgf | 4 kgf | 6 kgf |
|-------------------------------|--------|--------|-------|
| Accuracy [%] | 58.2 | 48.1 | 67.2 |
| Positive Predictive Value [%] | 51.4 | 36.7 | 64.4 |
| Processing Time [s] | 1508.9 | 3640.4 | 368.4 |
| Data Capacity [Mb] | 115 | 100 | 113 |

4.1.2. Optimal Sensor Selection Based on Sensor Efficiency Evaluation Method

The sensor data efficiency method described in Section 2.2 was applied to the raw data to select a single optimal sensor for performance improvement. Because the raw data comprised all sensor data, the sensor data efficiency method evaluates the sensors individually, as shown in Table S1. To evaluate the efficiency of S_E , the data capacity (α), processing time (β), and misclassification rate (γ) must be obtained from individual sensors. Sensors 1 and 2 from Figure 3 were evaluated because both were installed on the doctor blade in the same directions as the X, Y, and Z axes. Tables S1–S3 show the results of the sensor data efficiency evaluation, comparing the raw data to the data of Sensors 1 and 2. The results of the doctor blade tilting fault diagnosis for optimal sensor selection in Tables S1–S3 show that the highest S_E results for Sensor 1 are as listed in Table 5.

Table 5. Result of sensor data efficiency evaluation for optimal sensor selection of doctor blade tilting fault.

| S_E | Sensor 1 | Sensor 2 |
|-------|----------|----------|
| 2 kgf | 6 | 5.79 |
| 4 kgf | 6.69 | 5.70 |
| 6 kgf | 8.17 | 5 |

The result of the optimal sensor selection can be verified in Table S4 as compared with Table S5, based on the performance of the diagnosis results. It can also be seen that the diagnosis result of Sensor 1 was improved in accuracy, processing time, and data capacity compared with the result of raw data diagnosis shown in Table 4.

4.1.3. Optimal Axis Selection Based on the DNF Method

Sensor 1 from the raw data of the doctor blade tilting cases was selected as the optimal sensor, and the DNF method was used to evaluate axes X, Y, and Z from Sensor 1 to extract the most sensitive axis. As mentioned in Section 2.3, based on the kurtosis and standard deviation of normal and abnormal conditions, the DNF number was calculated. The axis having the highest number of DNFs resulted in the highest diagnostic performance. As shown in Table 6, the DNF number evaluation of the X, Y, and Z axes from Sensor 1 is shown. As shown in Table 6, the axis having the highest DNF number differed depending on the operating tension. For a tension of 2 kgf, the Y-axis resulted in the highest DNF number. Tensions of 4 and 6 kgf showed the highest DNF numbers on the X-axis. The

theory of achieving the highest diagnosis performance depending on the DNF number is verified in Tables S6–S8. Table S6 shows the highest accuracy of diagnosis for tensions of 4 and 6 kgf along the X-axis, and Table S7 illustrates the best result for a tension of 2 kgf. The proposed method evaluates the sensitivity of the axis using the DNF number, which resulted in a high rate of diagnosis accuracy and decreased processing time and data capacity requirements.

Table 6. DNF Number of axis X, Y, and Z from Sensor 1 of doctor blade tilting fault.

| D_N | X Axis | Y Axis | Z Axis |
|-------|--------|--------|--------|
| 2 kgf | 8.27 | 11.66 | 6.00 |
| 4 kgf | 16.76 | 15.53 | 8.64 |
| 6 kgf | 1.59 | 1.32 | 1.01 |

4.1.4. Feature Variable Optimization Based on FCM Method

As shown in Figure 2, 12 feature variables were extracted from the data acquired during Phases 1 and 2. From the 12 feature variables, four were selected to be coordinated into a feature combination. The four variables in this case were skewness, kurtosis, standard deviation, and peak-to-peak. The left and right sides of the statistical feature variables are generally symmetrical around the mean on a normal distribution. Hence, skewness and kurtosis are selected as indicators to determine how far the distribution shape of the data deviates from normal. Skewness measures the asymmetry of the distribution. The more symmetric the data, the closer the skewness to zero. Furthermore, because kurtosis is a measure of outliers present in the distribution, there are clear criteria for discriminating between normal and abnormal, such as finding a value of three in the Gaussian probability distribution. In the case of peak-to-peak, peak vibration can be observed on the distribution chart when an abnormality occurs. Hence, the FCM method was applied to skewness, kurtosis, standard deviation, and peak-to-peak. The coordination of three feature variables of the selected four forms a volume, as shown in Figure 6. The red volume represents the three-dimensional feature variables of the abnormal condition data. The blue volume represents normal condition data. A significant volume difference between normal and abnormal conditions is visible. After evaluating the coordination of feature combinations from the selected feature variables using the FDC number from Equation (3), the combination having the highest FDC number was used as input data to train the machine-learning fault diagnosis model. As shown in Table 7, the fault diagnosis results of the doctor blade tilting condition improved, owing to the data characterization process of Phases 1, 2, and 3. Compared with the results of the raw data-based diagnosis in Table 4, the smart data-based fault diagnosis resulted in an improved accuracy of 90.1% from 58.2% at a tension of 2 kgf. At 4 kgf, the accuracy improved from 48.1% to 86.2%, and 67.2% to 97.0% at a tension of 6 kgf. The processing time reduced from 1508.9 s to 33.9 s at a tension of 2 kgf. It reduced from 3640.4 s to 37.5 s at 4 kgf. It reduced from 368.4 s to 16.6 s at 6-kgf tension. The data capacity was also reduced from approximately 113 Mb to 4 Mb.

4.2. Printing Roll Eccentricity Fault Diagnosis

4.2.1. Printing Roll Eccentricity Fault Diagnosis Based on Raw Data

The fault diagnosis of printing roll eccentricity was conducted using the raw data of processes at tensions of 2, 4, and 6 kgf, as listed in Table 3. As shown in Table 8, the results based on the raw data showed a diagnosis accuracy of 69.7–76.9%. The processing time of the raw data diagnosis ranged from 208.0 s to 237.9 s.

4.2.2. Printing Roll Eccentricity Fault Diagnosis Based on Smart Data

The diagnosis of the printing roll eccentricity fault data was performed in the same order as the doctor blade tilting diagnosis procedure described in Section 4.1. Based on the raw data of Phase 2, the sensor data efficiency evaluation was applied to select a

single optimal sensor. As shown in Tables S9–S11, the data capacity, processing time, and misclassification rate of each case were computed to obtain S_E , as shown in Table 9. S_E results of Sensor 2 reflected the highest value for all tensions. The fault diagnosis results based on Sensors 1 and 2 are shown in Tables S12 and S13 as applied to the verification of the sensor data efficiency evaluation.

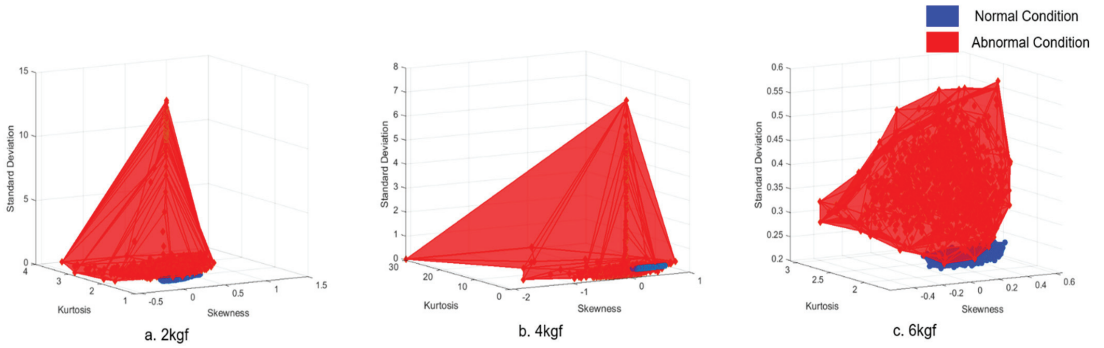


Figure 6. Volume comparison of normal and abnormal condition data: (a) Operating tension of 2 kgf; (b) Operating tension of 4 kgf; and (c) Operating tension of 6 kgf.

Table 7. Doctor blade tilting fault diagnosis based on smart data.

| [SVM] | 2 kgf [Y Axis] | 4 kgf [X Axis] | 6 kgf [X Axis] |
|-------------------------------|----------------|----------------|----------------|
| Accuracy [%] | 90.1 | 86.2 | 97.0 |
| Positive Predictive Value [%] | 89.8 | 85.9 | 97.0 |
| Processing Time [s] | 33.9 | 37.5 | 16.6 |
| Data Capacity [Mb] | 5 | 4 | 5 |

Table 8. Printing roll eccentricity diagnosis based on raw data (i.e., Sensors 1, 2, and 3).

| [SVM] | 2 and 4 kgf | 2 and 6 kgf | 4 and 6 kgf |
|-------------------------------|-------------|-------------|-------------|
| Accuracy [%] | 74.8 | 76.9 | 69.7 |
| Positive Predictive Value [%] | 70.2 | 73.5 | 54.2 |
| Processing Time [s] | 237.9 | 208.0 | 237.0 |
| Data Capacity [Mb] | 111 | 111 | 110 |

Table 9. Result of sensor data efficiency evaluation for optimal sensor selection of printing roll eccentricity fault.

| S_E | Sensor 1 | Sensor 2 |
|-------------|----------|----------|
| 2 and 4 kgf | 24.07 | 29.29 |
| 2 and 6 kgf | 20.29 | 25.15 |
| 4 and 6 kgf | 19.09 | 27.24 |

Based on the selected optimal Sensor 2 data, the DNF method was applied to extract the most sensitive axis information based on the DNF number. The results of the computation of the number of DNFs are listed in Table 10. The X-axis for tension 2 (4 kgf) resulted in the highest DNF number followed by the Z-axis for the remaining cases. Verification results of the selected axis depended on the cases based on the DNF number and are shown in Tables S14–S16. Compared with Table 10, the diagnostic performance of the selected axis having the highest DNF number provided the most efficient outcome.

Table 10. DNF Number of axis X, Y, and Z from Sensor 2 of printing roll eccentricity fault.

| D_N | X-Axis | Y-Axis | Z-Axis |
|-------|--------|--------|--------|
| 2 kgf | 1.12 | 0.91 | 1.06 |
| 4 kgf | 1.09 | 0.86 | 1.12 |
| 6 kgf | 0.95 | 1.01 | 1.06 |

As shown in Figure 7, the feature variables were extracted and combined into three feature combinations for evaluation. The selected and extracted feature variables were identical to those described in Section 4.1.4. The conditions of normal and abnormal data formed a volume measure for each feature variable, as shown in Figure 7. The two conditions were then computed using Equation (3) to select the training input data. Based on the results of the FCM method, it was then used as input data for printing roll eccentricity fault diagnosis. The results are listed in Table 11. Compared with Table 8, smart data increased the diagnosis accuracy up to 99.1% with a processing time of 3.7 s and a data capacity of 4 Mb. In summary, diagnosing the main fault printing roll eccentricity with smart data improved the diagnostic performance with less time consumption and fewer data requirements.

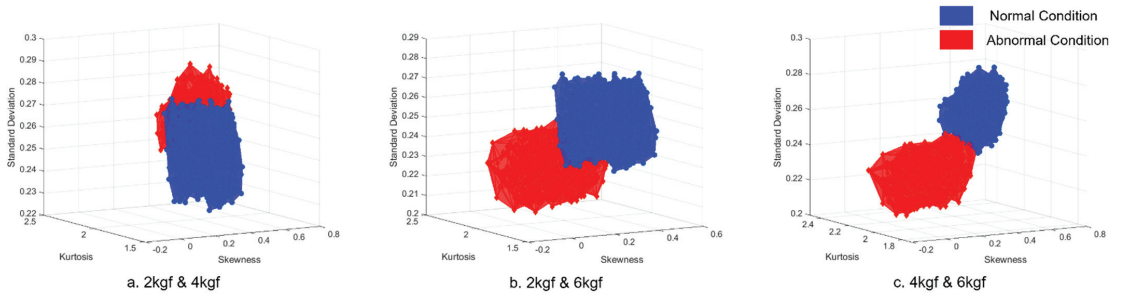


Figure 7. Volume comparison of normal and abnormal condition data: (a) Operating tensions of 2 and 4 kgf; (b) Operating tensions of 2 and 6 kgf; and (c) Operating tensions of 4 and 6 kgf.

Table 11. Printing roll eccentricity fault diagnosis based on smart data.

| [SVM] | 2 and 4 kgf [X Axis] | 2 and 6 kgf [Z Axis] | 4 and 6 kgf [Z Axis] |
|-------------------------------|----------------------|----------------------|----------------------|
| Accuracy [%] | 97.9 | 99.1 | 96.3 |
| Positive Predictive Value [%] | 93.4 | 94.9 | 92.0 |
| Processing Time [s] | 6.1 | 5.1 | 3.7 |
| Data Capacity [Mb] | 5 | 5 | 4 |

4.3. Nip Roll Eccentricity Fault Diagnosis

4.3.1. Nip Roll Eccentricity Fault Diagnosis Based on Raw Data

The fault diagnosis of the nip roll eccentricity based on raw data is shown in Table 12. The results for cases of tensions 2, 4, and 6 kgf rated 42.1% to 56.0% diagnosis accuracy with 425.4 s to 597.0 s of processing time. The data capacity of the raw data ranged from 111 Mb to 114 Mb, like the raw data capacity of doctor blade tilting and printing roll eccentricity faults.

4.3.2. Nip Roll Eccentricity Fault Diagnosis Based on Smart Data

The smart data transition from the raw data is presented in this section. The evaluation of the sensor data efficiency in Phase 1 used to select the optimal sensor is shown in Table 13. Sensor 1 was selected as the optimal sensor for the next phase of the DNF method. It can be seen that the S_E of each case at Sensor 1 was higher than that of Sensor 2. As shown

in Tables S17–S19, the data capacities of Sensors 1 and 2 maintained an average value of 43. As the capacity difference of both sensors merely influenced factor α , the major factor influencing the outcome of S_E was at factors β and γ . Tables S20 and S21 verify that the sensor having the highest S_E maintained the diagnosis result with higher accuracy.

Table 12. Nip roll eccentricity fault diagnosis based on raw data (i.e., Sensors 1, 2, and 3).

| [SVM] | 2 kgf | 4 kgf | 6 kgf |
|-------------------------------|-------|-------|-------|
| Accuracy [%] | 53.8 | 56.0 | 42.1 |
| Positive Predictive Value [%] | 46.7 | 47.7 | 33.9 |
| Processing Time [s] | 425.4 | 574.4 | 597.0 |
| Data Capacity [Mb] | 111 | 111 | 114 |

Table 13. Result of sensor data efficiency evaluation for optimal sensor selection of nip roll eccentricity fault.

| S_E | Sensor 1 | Sensor 2 |
|-------|----------|----------|
| 2 kgf | 24.16 | 19.6 |
| 4 kgf | 14.54 | 13.82 |
| 6 kgf | 19.5 | 16.88 |

The evaluation of the X, Y, and Z axes of Sensor 1 was carried out based on the DNF method and the DNF number. The results from the most sensitive axis for each case are listed in Table 14. For the case of the tension of 2 kgf, the Z-axis rate had the highest D_N , whereas tensions of 4 and 6 kgf rates were the highest in the X-axis. The diagnosis results for each case, based on the axis of Sensor 1, are shown in Tables S22–S24.

Table 14. DNF Number of axis X, Y, and Z from Sensor 1 of nip roll eccentricity fault.

| D_N | X Axis | Y Axis | Z Axis |
|-------|--------|--------|--------|
| 2 kgf | 0.89 | 0.86 | 1.09 |
| 4 kgf | 1.53 | 1.30 | 1.04 |
| 6 kgf | 1.67 | 1.15 | 0.92 |

The FCM method was carried out based on the results of Phase 2 in this section. The feature variables used for coordination of the combination were identical to the results of Sections 4.1 and 4.2 by skewness, kurtosis, standard deviation, and peak-to-peak. Kurtosis considers the effect of data at the end of the distribution on the probability curve. Based on the standard distribution, the kurtosis value increased with the weight of the outer values. Hence, kurtosis refers to the sharpness of the distribution, and if the degree of dispersion is large, the data are heterogeneous, and the height of the distribution is lowered. On the other hand, if the degree of dispersion is small, the data are homogeneous, and the height of the distribution increases.

The volume of normal and abnormal conditions based on the coordinated feature variables can be seen in Figure 8. Normal volume is shown in blue, and abnormal volumes are shown in red and yellow. The abnormal volumes differ depending on the nip force of the data. Table 15 shows the results of the nip roll eccentricity fault diagnosis based on the smart data. In the case of the tension of 2 kgf, the diagnostic accuracy rates were 100% with a data capacity of 4 Mb and a processing time of 4.63 s. Compared with the results of the raw data in Table 12, it can be seen that the fault diagnosis model performances improved in areas of accuracy, positive predictive value, processing time, and data capacity.

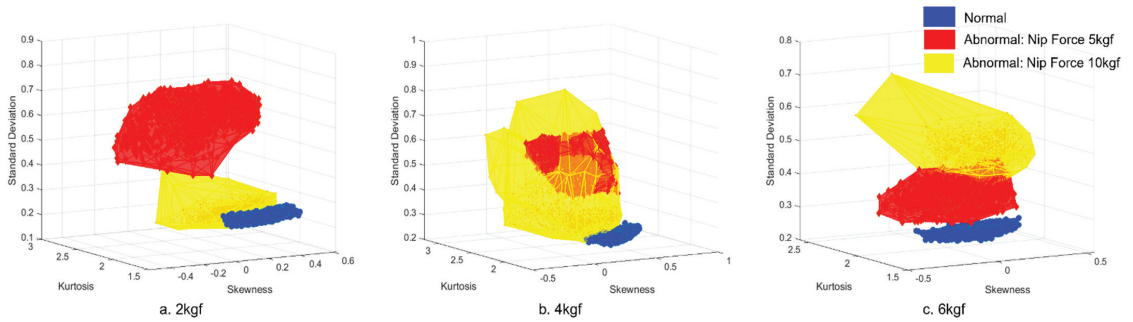


Figure 8. Volume comparison of normal and abnormal condition data: (a) Operating tension of 2 kgf; (b) Operating tension of 4 kgf; and (c) Operating tension of 6 kgf.

Table 15. Nip roll eccentricity fault diagnosis based on smart data.

| [SVM] | 2 kgf [Z Axis] | 4 kgf [X Axis] | 6 kgf [X Axis] |
|-------------------------------|----------------|----------------|----------------|
| Accuracy [%] | 100.0 | 98.4 | 99.5 |
| Positive Predictive Value [%] | 98.8 | 97.0 | 98.2 |
| Processing Time [s] | 4.63 | 4.38 | 4.40 |
| Data Capacity [Mb] | 4 | 4 | 4 |

4.4. Nip Force Non-Uniformity Fault Diagnosis

4.4.1. Nip Force Non-Uniformity Fault Diagnosis Based on Raw Data

Fault diagnosis based on raw data was performed to detect nip force non-uniformity. Figure 5 shows the data of the nip force for Cases 1–18. As Cases 11 and 17 in Figure 5 showed non-uniformity nip forces, the data of both cases were used as abnormal condition data for fault diagnosis. Table 16 shows the performance of the fault diagnosis at tensions of 4 and 6 kgf.

Table 16. Nip force non-uniformity fault diagnosis based on raw data (i.e., Sensors 1, 2, and 3).

| [SVM] | 4 kgf | 6 kgf |
|-------------------------------|-------|-------|
| Accuracy [%] | 65.5 | 65.4 |
| Positive Predictive Value [%] | 60.3 | 59.4 |
| Processing Time [s] | 281.7 | 515.4 |
| Data Capacity [Mb] | 115 | 116 |

4.4.2. Nip Force Non-Uniformity Fault Diagnosis Based on Smart Data

The sensor data efficiency evaluation results are shown in Table 17 based on the computation of Tables S25 and S26. It can be seen that Sensor 2 had the highest S_E among the raw data. Tables S27 and S28 can be used to verify the optimal sensor selection results of the sensor data efficiency evaluation.

The DNF method was used to evaluate the axis of Sensor 2 by X, Y, and Z for tension cases of 4 and 6 kgf. The DNF numbers for both cases are shown in Table 18, where the result of a tension of 4 kgf showed axis Y as the most valid, and X for the tension case of 6 kgf. The results of the fault diagnosis based on Sensor 2 for the triaxis are shown in Tables S29–S31.

With identical feature variables coordinated through the FCM method, the volumes of normal and abnormal conditions are shown in Figure 9. It can be seen from Figure 9a that the volume of the normal condition overlaps with the volume of the abnormal condition. Thus, the peak values and the distribution of data points for abnormal conditions were

broad, compared with the normal volume condition. Based on the results of the FCM, the nip force non-uniformity fault diagnosis results with smart data are shown in Table 19.

Table 17. Result of sensor data efficiency evaluation for optimal sensor selection of nip force non-uniformity fault.

| S_E | Sensor 1 | Sensor 2 |
|-------|----------|----------|
| 4 kgf | 11.29 | 11.83 |
| 6 kgf | 11.91 | 12.45 |

Table 18. DNF Number of X, Y, and Z axes from Sensor 2 of nip force non-uniformity fault.

| D_N | X Axis | Y Axis | Z Axis |
|-------|--------|--------|--------|
| 4 kgf | 0.97 | 1.12 | 1.10 |
| 6 kgf | 1.16 | 1.03 | 1.04 |

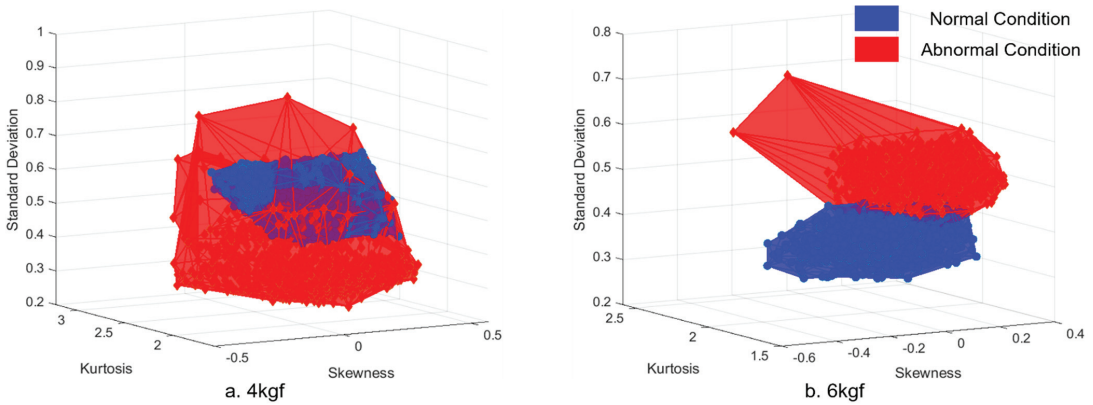


Figure 9. Volume comparison of normal and abnormal condition data: (a) Operating tension of 4 kgf; (b) Operating tension of 6 kgf.

Table 19. Nip force non-uniformity fault diagnosis based on smart data.

| [SVM] | 4 kgf [Y Axis] | 6 kgf [X Axis] |
|-------------------------------|----------------|----------------|
| Accuracy [%] | 97.9 | 95.2 |
| Positive Predictive Value [%] | 93.5 | 90.7 |
| Processing Time [s] | 25.4 | 28.4 |
| Data Capacity [Mb] | 6 | 5 |

4.5. Simultaneous Fault Diagnosis

In Sections 4.1–4.4, defects caused during the printing process of gravure printing system has been diagnosed independently. However, occasionally in real applications it is likely for the gravure printing system to malfunction with more than one single fault. In this section, characterized smart data has been applied under the assumption of multiple faults appearing simultaneously to present the effectiveness of the diagnosis model performance.

Cases 6, 12, and 18 from Table 2 has been selected for the multiple fault data since the experimental condition included with nipping and doctoring at tensions 2, 4, and 6 kgf. Diagnosis results of simultaneous multiple faults is shown in Table 20. The effectiveness of the smart data characterization is shown with comparison to the diagnosis result with raw data. As the raw data of simultaneous faults contain various disturbances with noticeable

peaks, it is less complex for the raw data-based diagnosis model to clarify the distinct conditions for classification. Hence, the average accuracy of raw data diagnosis is at 72.3% in which rates a higher value compared to single fault diagnosis results. Therefore, results based on smart data rates at an average of 99% on the grounds of abovementioned basis. In short, detecting simultaneous multiple faults based on smart data shows positive results as shown in Table 20.

Table 20. Simultaneous fault diagnosis result based on big data and smart data.

| [SVM]. | 2 kgf [X Axis] | 4 kgf [Y Axis] | 6 kgf [Y Axis] |
|-------------------------------|----------------|----------------|----------------|
| Accuracy [%] | 70→97 | 74→100 | 73→100 |
| Positive Predictive Value [%] | 69→95 | 72→99 | 73→99 |
| Processing Time [s] | 4501→52 | 4035→34 | 4722→49 |
| Data Capacity [Mb] | 110→6 | 112→8 | 114→7 |

4.6. Raw Data and Smart Data Comparison for Fault Diagnosis

The fault diagnosis of four possible major faults during the printing process of the gravure printing system based on raw and smart data is shown in Table 21. Table 21 summarizes the impact of data characterization methods for the diagnosis of the four suggested major faults and the simultaneous faults of the gravure printing system printing process. The diagnosis performance comparison results are shown based on raw and smart data. All diagnosis results based on raw data and smart data are processed through support vector machine algorithm. In Tables S32–S35, diagnosis results of the four major faults depending on the machine learning algorithm is shown. A total of eight different algorithms have been applied to each of the faults and consequently shows that the most efficient outcome of the performance regarding accuracy, positive predictive value, and processing time concludes with the use of a support vector machine algorithm to diagnose all faults of the printing process.

Table 21. Raw data and smart data diagnosis comparison.

| Main Faults | Accuracy [%] | PPV [%] | Processing Time [s] | Data Capacity [Mb] |
|----------------------------|--------------|---------|---------------------|--------------------|
| Doctor Blade Tilting | 48→97 | 36→97 | 3640→16 | 110→5 |
| Printing Roll Eccentricity | 69→99 | 54→94 | 237→5 | 110→5 |
| Nip Roll Eccentricity | 42→100 | 33→98 | 597→4 | 114→4 |
| Nip Force Non-Uniformity | 65→97 | 59→93 | 515→25 | 116→6 |
| Simultaneous Faults | 74→100 | 72→99 | 4035→34 | 112→8 |

Based on the results of Table 21, techniques to increase the accuracy of the classification has been applied to faults of doctor blade tilting, printing roll eccentricity, and nip force non-uniformity. As the abovementioned faults maintain an accuracy of 97% to 99%, it is possible to improve the final diagnosis results by adjusting the parameter of window size. As shown is Equation (4), the window size can be adjusted using the sampling rate and revolutions per minute. As x is the revolutions per minute, and α as the sampling rate (Hz), it is possible to obtain the value W_s . Once the value W_s is obtained for the three faults it is then applied to as a fixed parameter to be diagnosed based on the smart data. The results show in Table 22 that the contents of accuracy, PPV, and processing time have improved in comparison to the results of Table 21.

$$W_s = \alpha \left(\frac{x}{60} \right)^{-1} \quad (4)$$

Table 22. Smart data diagnosis improvement with window size adjustment.

| Main Faults | Accuracy [%] | PPV [%] | Processing Time [s] | Data Capacity [Mb] |
|----------------------------|--------------|---------|---------------------|--------------------|
| Doctor Blade Tilting | 48→100 | 36→99 | 3640→13 | 110→5 |
| Printing Roll Eccentricity | 69→100 | 54→99 | 237→8 | 110→5 |
| Nip Force Non-Uniformity | 65→100 | 59→98 | 515→17 | 116→6 |

5. Conclusions

Printing defects generated by the misalignment of the doctor blade, eccentricity of the nip and printing rolls, and non-uniform nip roll pressures can negatively affect the performance of printed electronic devices. To prevent printing defects and to obtain high-quality printed functional layers, it is necessary to recognize and diagnose factors that cause printing defects. In this study, a method for data characterization using sensor data efficiency evaluation (S_E), DNF, and FCM methods was proposed to diagnose the possible four major faults in the roll-to-roll gravure printing process, followed by experimental verification. The misalignment of the doctor blade, printing roll eccentricity, nip roll eccentricity, nip force non-uniformity, and simultaneous faults rated an average value of 56% accuracy with raw data. However, with smart data, the accuracy rated 100.0% on average. The positive predictive value increased when the learning time was reduced from 1247 s to 12 s on average. The data capacity was reduced from 112 Mb to 5 Mb, depending on the selection of the sensor and its axis with optimized feature variable coordination. It is known that, with the use of smart data through sensor data efficiency evaluation, the feature combination matrix, and DNF methods, machine learning fault diagnosis model performance improves for classifying normal and abnormal conditions of datasets. The proposed smart data process in this paper is the most novel and contributory aspect of this paper because it leads to the near-perfect performance of the machine learning fault detection model. It is also faster and less computer-memory intensive than the results found from raw sensor data. This poses a contribution to the field, and countless industries can benefit from the improved and most cost-efficient production of printed electronics. Further research regarding the methodologies proposed in this paper plans to expand the application for fault diagnosis despite the numerous numbers of sensors.

Supplementary Materials: The following are available online at <https://www.mdpi.com/1424-8220/21/24/8454/s1>, Table S1: Sensor data efficiency evaluation of doctor blade tilting fault under operating tension 2 kgf, Table S2: Sensor data efficiency evaluation of doctor blade tilting fault under operating tension 4 kgf, Table S3: Sensor data efficiency evaluation of doctor blade tilting fault under operating tension 6 kgf, Table S4: Doctor blade tilting fault diagnosis result based on sensor 1, Table S5: Doctor blade tilting fault diagnosis result based on sensor 2, Table S6: Doctor blade tilting fault diagnosis result based on X axis of sensor 1, Table S7: Doctor blade tilting fault diagnosis result based on Y axis of sensor 1, Table S8: Doctor blade tilting fault diagnosis result based on Z axis of sensor 1, Table S9: Sensor data efficiency evaluation of printing roll eccentricity fault under case 2 kgf and 4 kgf, Table S10: Sensor data efficiency evaluation of printing roll eccentricity fault under case 2 kgf and 6 kgf, Table S11: Sensor data efficiency evaluation of printing roll eccentricity fault under case 4 kgf and 6 kgf, Table S12: Printing roll eccentricity fault diagnosis result based on sensor 1, Table S13: Printing roll eccentricity fault diagnosis result based on sensor 2, Table S14: Printing roll eccentricity fault diagnosis result based on X axis of sensor 2, Table S15: Printing roll eccentricity fault diagnosis result based on Y axis of sensor 2, Table S16: Printing roll eccentricity fault diagnosis result based on Z axis of sensor 2, Table S17: Sensor data efficiency evaluation of nip roll eccentricity fault under case 2 kgf, Table S18: Sensor data efficiency evaluation of nip roll eccentricity fault under case 4 kgf, Table S19: Sensor data efficiency evaluation of nip roll eccentricity fault under case 6 kgf, Table S20: Nip roll eccentricity fault diagnosis result based on sensor 1, Table S21: Nip roll eccentricity fault diagnosis result based on sensor 2, Table S22: Nip roll eccentricity fault diagnosis result based on X axis of sensor 1, Table S23: Nip roll eccentricity fault diagnosis result based on Y axis of sensor 1, Table S24: Nip roll eccentricity fault diagnosis result based on Z axis of sensor 1, Table S25: Sensor data efficiency evaluation of nip force non-uniformity fault under case 4 kgf, Table S26: Sensor

data efficiency evaluation of nip force non-uniformity fault under case 6 kgf, Table S27. Nip force non-uniformity fault diagnosis result based on sensor 1, Table S28. Nip force non-uniformity fault diagnosis result based on sensor 2, Table S29. Nip force non-uniformity fault diagnosis result based on X axis of sensor 2, Table S30. Nip force non-uniformity fault diagnosis result based on Y axis of sensor 2, Table S31. Nip force non-uniformity fault diagnosis result based on Z axis of sensor 2, Table S32. Doctor blade tilting fault diagnosis with various machine learning algorithms, Table S33. Printing roll eccentricity fault diagnosis with various machine learning algorithms, Table S34. Nip roll eccentricity fault diagnosis with various machine learning algorithms, Table S35. Nip force non-uniformity fault diagnosis with various machine learning algorithms.

Author Contributions: Conceptualization, Y.L. and C.L.; methodology, Y.L. and M.J.; software, Y.L. and G.C.; formal analysis, Y.L., M.J. and G.C.; data curation, M.J. and G.C.; writing—original draft preparation, Y.L., M.J. and C.L.; writing—review and editing, M.J. and C.J.; visualization, G.C. and C.J.; supervision, C.L.; project administration, C.L.; funding acquisition, C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A5A1019649) and the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0012770).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Park, H.; Sun, J.; Jung, Y.; Park, J.; Maskey, B.B.; Shrestha, K.; Koirala, G.R.; Parajuli, S.; Shrestha, S.; Chung, A.; et al. The First Step towards a R2R Printing Foundry via a Complementary Design Rule in Physical Dimension for Fabricating Flexible 4-Bit Code Generator. *Adv. Electron. Mater.* **2020**, *6*, 2000770. [\[CrossRef\]](#)
2. Jo, M.; Lee, J.; Kim, S.; Cho, G.; Lee, T.M.; Lee, C. Resistance Control of an Additively Manufactured Conductive Layer in Roll-to-Roll Gravure Printing Systems. *Int. J. Precis. Eng. Manuf. Green Technol.* **2021**, *8*, 817–828. [\[CrossRef\]](#)
3. Nguyen, H.A.D.; Lee, C.; Shin, K.H. Approach to optimizing printed conductive lines in high-resolution roll-to-roll gravure printing. *Robot. Comput. Integr. Manuf.* **2017**, *46*, 122–129. [\[CrossRef\]](#)
4. Lee, S.H.; Lee, S. Fabrication of Comb-Structured Acceleration Sensors by Roll-to-Roll Gravure Printing. *Int. J. Precis. Eng. Manuf. Green Technol.* **2021**, 1–21. [\[CrossRef\]](#)
5. Lee, J.; Jo, M.; Lee, C. Advanced Tension Model for Highly Integrated Flexible Electronics in Roll-to-Roll Manufacturing. *IEEE/ASME Trans. Mechatron.* **2021**. [\[CrossRef\]](#)
6. Bisconti, F.; Giuri, A.; Suhonen, R.; Kraft, T.M.; Ylikunnari, M.; Holappa, V.; Po', R.; Biagini, P.; Savoini, A.; Marra, G.; et al. One-step polymer assisted roll-to-roll gravure-printed perovskite solar cells without using anti-solvent bathing. *Cell. Rep. Phys. Sci.* **2021**, *2*, 100639. [\[CrossRef\]](#)
7. Lee, J.; Byeon, J.; Lee, C. Fabrication of Thickness-Controllable Double Layer Electrolyte Using Roll-to-Roll Additive Manufacturing System. *Int. J. Precis. Eng. Manuf. Green Tech.* **2020**, *7*, 635–642. [\[CrossRef\]](#)
8. Lee, T.M.; Lee, S.H.; Noh, J.H.; Kim, D.S.; Chun, S. The effect of shear force on ink transfer in gravure offset printing. *J. Micromech. Microeng.* **2010**, *20*, 125026. [\[CrossRef\]](#)
9. Nguyen, H.A.D.; Shin, K.; Lee, C. Effect of nip force on ink transfer in high resolution roll-to-roll printing. *Int. J. Precis. Eng. Manuf.* **2015**, *16*, 517–523. [\[CrossRef\]](#)
10. Kang, H.; Lee, C. Optimal Design of Pneumatic Flotation for Roll-to-Roll Conveyance in the Production of Printed Circuits. *Appl. Sci.* **2020**, *10*, 5440. [\[CrossRef\]](#)
11. Nguyen, H.A.D.; Lee, C.; Shin, K.H. A mathematical model to predict surface roughness and pattern thickness in roll-to-roll gravure printed electronics. *Robot. Comput. Integr. Manuf.* **2013**, *29*, 26–32. [\[CrossRef\]](#)
12. Lee, J.; Byeon, J.; Lee, C. Theories and Control Technologies for Web Handling in the Roll-to-Roll Manufacturing Process. *Int. J. Precis. Eng. Manuf. Green Tech.* **2020**, *7*, 525–544. [\[CrossRef\]](#)
13. Xia, M.; Li, T.; Xu, L.; Liu, L.; De Silva, C.W. Fault Diagnosis for Rotating Machinery Using Multiple Sensors and Convolutional Neural Networks. *IEEE/ASME Trans. Mechatron.* **2018**, *23*, 101–110. [\[CrossRef\]](#)
14. Duan, Z.; Wu, T.; Guo, S.; Shao, T.; Malekian, R.; Li, Z. Development and trend of condition monitoring and fault diagnosis of multi-sensors information fusion for rolling bearings: A review. *Int. J. Adv. Manuf. Technol.* **2018**, *96*, 803–819. [\[CrossRef\]](#)
15. Zhao, S.; Li, J.; Cheng, X. Sensor Fault Diagnosis Based on Adaptive Arc Fuzzy DBN-Petri Net. *IEEE Access.* **2021**, *9*, 20305–20317. [\[CrossRef\]](#)

16. Huang, K.; Wu, S.; Li, F.; Yang, C.; Gui, W. Fault Diagnosis of Hydraulic Systems Based on Deep Learning Model With Multirate Data Samples. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–13. [[CrossRef](#)]
17. Bazan, G.H.; Goedel, A.; Duque-Perez, O.; Morinigo-Sotelo, D. Multi-fault diagnosis in three-phase induction motors using data optimization and machine learning techniques. *Electronics* **2021**, *10*, 1462. [[CrossRef](#)]
18. Wang, Z.; Zhao, W.; Du, W.; Li, N.; Wang, J. Data-driven fault diagnosis method based on the conversion of erosion operation signals into images and convolutional neural network. *Process. Saf. Environ. Prot.* **2021**, *149*, 591–601. [[CrossRef](#)]
19. Lee, J.; Park, B.; Lee, C. Fault diagnosis based on the quantification of the fault features in a rotary machine. *Appl. Soft Comput. J.* **2020**, *97*, 106726. [[CrossRef](#)]
20. Udmale, S.S.; Singh, S.K. Application of spectral kurtosis and improved extreme learning machine for bearing fault classification. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 4222–4233. [[CrossRef](#)]
21. Lebed, E.V. The accuracy of statistical computing of the standard deviation of a random variable. *IOP Conf. Ser. Earth Environ. Sci.* **2017**, *90*, 12150. [[CrossRef](#)]
22. Chaffin, W.W.; Rhiel, G.S. The effect of skewness and kurtosis on the one-sample t test and the impact of knowledge of the population standard deviation. *J. Stat. Comput. Simul.* **1993**, *46*, 79–90. [[CrossRef](#)]
23. Zebari, R.; Abdulazeez, A.; Zeebaree, D.; Zebari, D.; Saeed, J. A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. *J. Appl. Sci. Technol. Trends* **2020**, *1*, 56–70. [[CrossRef](#)]
24. Sisiaridis, D.; Markowitch, O. Reducing Data Complexity in Feature Extraction and Feature Selection for Big Data Security Analytics. In Proceedings of the 2018 1st International Conference on Data Intelligence and Security (ICDIS), South Padre Island, TX, USA, 8–10 April 2018; pp. 43–48. [[CrossRef](#)]
25. Kira, K.; Rendell, L.A. *A Practical Approach to Feature Selection*; Morgan Kaufmann Publishers, Inc.: Burlington, MA, USA, 1992.
26. Altinors, A.; Yol, F.; Yaman, O. A sound based method for fault detection with statistical feature extraction in UAV motors. *Appl. Acoust.* **2021**, *183*, 108325. [[CrossRef](#)]
27. Minhas, A.S.; Singh, S. A new bearing fault diagnosis approach combining sensitive statistical features with improved multiscale permutation entropy method. *Knowl.-Based Syst.* **2021**, *218*, 106883. [[CrossRef](#)]
28. Ji, H. Statistics Mahalanobis distance for incipient sensor fault detection and diagnosis. *Chem. Eng. Sci.* **2020**, *230*, 116233. [[CrossRef](#)]
29. Zhu, Q.; Zhang, N.; Xu, Y.; He, Y. Novel Mahalanobis Distance Based Fault Diagnosis Using Discrimination Neighborhood Preserving Embedding for Industrial Process. In Proceedings of the 2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS), Suzhou, China, 14–16 May 2021; pp. 18–22.
30. Chatterjee, S.; Maji, B. A Mahalanobis distance based algorithm for assigning rank to the predicted fault prone software modules. *Appl. Soft Comput. J.* **2018**, *70*, 764–772. [[CrossRef](#)]
31. Lee, B.R.; Yun, T.J.; Oh, W.B.; Lee, C.W.; Kim, H.H.; Jeong, Y.J.; Kim, I.S. A study on cutting quality using a mahalanobis distance. *J. Korean Soc. Precis. Eng.* **2021**, *38*, 253–260. [[CrossRef](#)]

Article

Aircraft Landing Gear Retraction/Extension System Fault Diagnosis with 1-D Dilated Convolutional Neural Network

Jie Chen ^{1,*}, Qingshan Xu ¹, Yingchao Guo ¹ and Runfeng Chen ²

¹ School of Civil Aviation, Northwestern Polytechnical University, Xi'an 710072, China; niguanfeixian1994@163.com (Q.X.); guo_ying_chao2020@163.com (Y.G.)

² China Academy of Space Technology (CAST), Beijing 100081, China; chenrunfeng2018@126.com

* Correspondence: shuimujie@mail.nwpu.edu.cn

Abstract: The faults of the landing gear retraction/extension(R/E) system can result in the deterioration of an aircraft's maneuvering conditions; how to identify the faults of the landing gear R/E system has become a key issue for ensuring aircraft take-off and landing safety. In this paper, we aim to solve this problem by proposing the 1-D dilated convolutional neural network (1-DDCNN). Aiming at developing the limited feature information extraction and inaccurate diagnosis of the traditional 1-DCNN with a single feature, the 1-DDCNN selects multiple feature parameters to realize feature integration. The performance of the 1-DDCNN in feature extraction is explored. Importantly, using padding dilated convolution to multiply the receptive field of the convolution kernel, the 1-DDCNN can completely retain the feature information in the original signal. Experimental results demonstrated that the proposed method has high accuracy and robustness, which provides a novel idea for feature extraction and fault diagnosis of the landing gear R/E system.

Keywords: landing gear retraction/extension(R/E) system; 1-D dilated convolutional neural network (1-DDCNN); fault diagnosis; feature integration

Citation: Chen, J.; Xu, Q.; Guo, Y.; Chen, R. Aircraft Landing Gear Retraction/Extension System Fault Diagnosis with 1-D Dilated Convolutional Neural Network. *Sensors* **2022**, *22*, 1367. <https://doi.org/10.3390/s22041367>

Academic Editor: Andrea Cataldo

Received: 4 January 2022

Accepted: 7 February 2022

Published: 10 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The landing gear R/E system is the significant subsystem for aircrafts, after long-term running under complex and variable conditions, with heavy loads and strong impact, the key parts in the landing gear R/E system will inevitably generate multifarious faults, which may affect take-off, landing, and flight safety.

Firstly, Hinton proposed a deep learning method in 2006, which set off a new wave of research on artificial intelligence and its applications [1]. In particular, deep learning models have shown significant success in image processing, speech recognition, target detection, information retrieval, natural language processing, and so on [2]. Moreover, as an important network structure, CNNs are widely applied in computer vision and natural language processing [3]. Machine learning methods have made great progress in the field of fault diagnosis. For example, Gligorijevic et al. proposed a method for rolling bearing fault diagnosis. Through the five-level wavelet decomposition of the vibration signals, the standard deviations of the wavelet coefficients from six sub-bands were extracted as representative features; feature dimensionality reduction was then performed, and the diagnosis accuracy reached 98.9% [4]. However, some scholars gradually introduced CNNs into the field of fault diagnosis. By converting 1-D timeseries vibration signals into 2-D input matrices, some experts and scholars constructed 2-D convolutional neural network models for fault diagnosis of rotating machinery. Janssens et al. performed a short-time Fourier transform on the vibration information of rotating machinery, then input the transformed coefficient map into a constructed CNN model for feature extraction to achieve CNN-based multi-fault identification [5]. Jing et al. proposed an adaptive multi-sensor data fusion method based on deep convolutional neural networks, for fault

diagnosis. The proposed method can learn features from raw data and adaptively optimize a combination of different fusion levels to effectively diagnose different faults of planetary gearboxes [6]. Wen et al. proposed a signal-to-image conversion method, using CNN techniques to extract features of the converted images, and achieved excellent test results for three famous datasets, comprising a motor bearing dataset, a self-priming centrifugal pump dataset, and an axial piston hydraulic pump dataset [7].

The main difference between the 1-DCNN and 2-DCNN lies in the dimension of input data and the sliding mode of the convolution kernel. Wu et al. proposed a 1-DCNN model for fault diagnosis of original vibration signals, for which test diagnosis accuracy reached 99.3% [8]. Huang et al. proposed a multi-scale cascade convolutional neural network (MC-CNN) model for fault diagnosis of bearings, achieving satisfactory results under non-stationary operating conditions [9]. He et al. combined a 1-DCNN and a LSTM (long short-term memory) network to construct a novel network model for fault diagnosis of bearings, with an average accuracy of over 99% [10].

For the traditional CNN, the convolution and pooling operations are carried out alternately, which reduces the feature maps' size while increasing the receptive field. However, for the pixel-level prediction problem of image segmentation, the final feature output size is required to be consistent with the original image size, which involves the down-sampling and up-sampling processes, image resolution reduction, and information loss. To solve these problems, dilated convolution came into being [11]. Zhuang et al. proposed a stacked residual dilated convolutional neural network (SRDCNN) for real-time fault diagnosis of bearings by combining dilated convolution, LSTM, and residual networks, and their experimental results show that the proposed model has improved denoising performance and adaptability [12]. Feng et al. used dilated convolution to replace conventional convolution and pooling structures, and introduced instance normalization (IN) to solve the issue of data style transfer. The proposed 1-D stacked dilated convolutional neural network (1D-SDCNN) model has an average accuracy of 96.8% for fault diagnosis of rolling bearings with variable loads [13]. Liang et al. combined LSTM, dilated convolution, and capsule networks to construct a new capsule network with gate-structure dilated convolutions (GDCCN). Their experimental results demonstrated that the proposed model has strong noise resistance and generalization in fault diagnosis of motors under variable load conditions [14]. As equipment has become more intelligent, complex, and integrated, it is difficult to accurately determine the characteristics of the failure status with a single feature. Moreover, using the 1-DCNN for fault diagnosis is highly dependent on fault datasets for roller bearing. The CNN fault diagnosis model can directly extract features from the original data to achieve end-to-end fault diagnosis without complicated data pre-processing. Dimensionality reduction methods, such as principal component analysis (PCA), cannot effectively preserve the time-dependence of timeseries data; moreover, information loss occurs in the process of dimensionality reduction.

In order to solve the above problems, this paper proposes a fault diagnosis method for the aircraft landing gear R/E system based on a 1-D dilated convolutional neural network (1-DDCNN). The main work of this paper is summarized as follows:

1. Analyze the aircraft landing gear R/E system's main fault mode with its working principle.
2. A conventional 1-DCNN was constructed to classify faults based on the actuator's displacement; experimental results show the average accuracy of the test set reached 91.80%.
3. The receptive field size affects the extraction of the original information by the model. There is a nonlinear relationship between the receptive field and the convolution kernel size for the expanded convolution. The convolution kernel size's influence on classification accuracy is explored.
4. There is a low classification result with a single feature; consequently, 1-DDCNN selects multi-feature parameters with the system pressure, the pressure at the right and left end of the actuator cylinder. Experimental results reveal that the average test accuracy of the model reaches 99.80%.

2. The Typical Aircraft Landing Gear R/E System Analysis

2.1. System Working Principle and Composition

The landing gear system of a typical aircraft mainly includes the following: front landing gear and cabin door, main undercarriage and cabin door, landing gear R/E system, wheels and brakes, turning control system, landing gear position indication system, etc. Among them, the R/E system mainly completes normal R/E and emergency R/E functions, and provides the landing gear position indication signal.

The front landing gear R/E process is similar to that of the main landing gear; therefore, only the front landing gear retraction process is described here. The working status of the front landing gear is shown in Figure 1. When the plane takes off, and the landing gear wheels are off the ground, the pilot sets the landing gear R/E control switch to the “UP” position, and the current flows to the landing gear R/E electromagnetic switch and the accumulator charging electromagnetic switch. The hydraulic fluid from the three-position four-way directional valve enters the front landing gear lower lock and actuating cylinder. As the accumulator charging solenoid switch is turned on, oil from the pump is supplied to the actuating cylinder, and the oil in the accumulator is also released to aid the landing gear retraction. When the aircraft is about to land, the pilot moves the control switch to the “DOWN” position, the three-position four-way directional valve is switched to the down circuit, and the fluid enters the front landing gear upper lock. Once the lock is opened, the oil enters the lowering chamber of the R/E actuator to lower the landing gear.

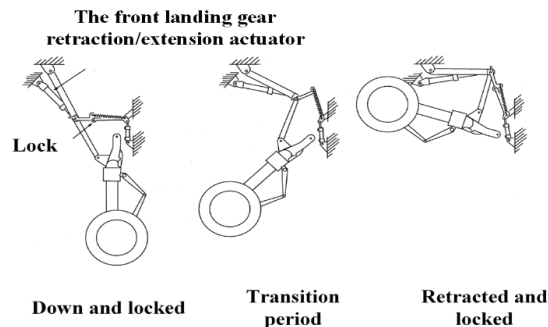


Figure 1. Front landing gear retracting process.

The landing gear R/E system mainly includes the following components: constant pressure variable pump, tank, hydraulic motor, filter, accumulator, actuator, press control, throttle valve, and three-position four-way directional control valve.

Through providing certain pressure and oil mass, the pump converts mechanical energy into hydraulic energy. The tank is used to store hydraulic oil. The filter is used to filter the hydraulic oil and remove its impurities. The accumulator not only supplies oil at both weak and heavy flows, it also compensates for leakage and maintains constant pressure. The actuator is a device that converts hydraulic energy into mechanical energy for linear reciprocating motion, which overcomes the load (including friction) and maintains the speed of motion using pressure-driven liquid flow. The relief valve is one of the common pressure valves used to regulate or limit the pressure in a hydraulic system. The throttle valve is a hydraulic component that regulates and controls the flow of oil in a hydraulic system. The function of the one-way throttle valve is to ensure that the oil flows in one direction, with no backflow, using the throttling effect. The solenoid directional valve is one of the frequently used hydraulic components in hydraulic systems, and is used to switch the direction of the hydraulic circuit. This article uses a three-position four-way solenoid directional valve. The “Position” refers to the working position of the spool. The “Way” marks the valve body of the oil port [15].

2.2. Failure Mode and Effect Analysis

The failure mode and effect analysis (FMEA), derived from typical civil aircraft design data in this subsection, can be used for parameter selection and fault injection into the simulation model, whereby fault datasets are obtained. The FMEA was carried out on the system's main components for subsequent fault diagnosis, the results of which are shown in Table 1 below. The failure analysis in this paper focuses on the component level of the landing gear R/E system and does not explore the specific internal failure of each component. For the excessive noise from the pump, the failure threshold can be obtained by changing the air content of the oil. If the clogging of the throttle valves at both ends of the actuator cylinder has different effects on the system, it is necessary to change the throttle valve's diameter at both ends of the actuator cylinder to obtain the failure threshold. Regarding the system failure caused by constant pressure variable pump leakage, a throttle valve should be connected to the constant pressure variable pump in parallel, and the throttle valve's diameter should be changed to simulate different degrees of leakage of constant pressure variable pump. Actuator cylinder leakage also affects the normal operation of the system, and the failure threshold can be obtained by changing the actuator leakage coefficient.

Table 1. FMEA for the landing gear R/E system.

| Components | Failure Mode | Failure Cause |
|---------------------------------|--|---|
| Constant pressure variable pump | Insufficient oil output and pressure of the pump | Severe pump leakage or low pump speed |
| | Excessive noise from the pump | The pump is severely worn or air mixed in the oil |
| Actuator cylinder | Higher temperature | Abnormal wear in the pump |
| | Piston rod stuck | The actuator cylinder is leaking seriously or press control is faulty |
| | The speed did not reach the standard value | The actuator cylinder is leaking seriously and it is subjected to excessive external load |
| Directional control valve | Creeping phenomenon | Air entered the actuator cylinder |
| | Insufficient flow after reversing spool | Insufficient valve opening |
| | Spool stands still | Spool stuck or electromagnet failed |
| Filter | Excessive pressure drops | Improper setting of some parameters |
| | Filter clogging | Oil is contaminated |
| Throttle valve | The range of flow regulation is limited | Clogged throttle hole |

3. 1-DDCNN Fault Diagnosis Model

3.1. 1-DCNN

The receptive field is defined as the area size mapped on the original image by each pixel on the feature map output from each layer in the CNN. The neuron's receptive field value decides the original range it can cover, meaning that it may contain more global features. Figure 2a shows the range of neuron receptive fields in the third layer of the 1-DCNN, with a convolutional kernel size of 3×1 and a step size of 1×1 . The marked blue neurons in the third layer are mapped from the blue regions in the first layer, that is, the receptive field size of the input sequence data corresponding to a neuron in the output feature map 2 is 5×1 .

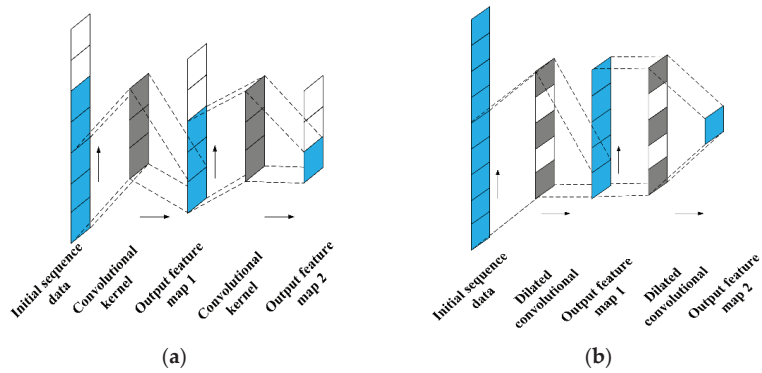


Figure 2. The receptive field of 1-DCNN. (a) 1-DCNN; (b) 1-DDCNN.

Before inputting 1-D time series signals into 2-DCNN, the common method is to rearrange and combine signal sampling points, using a simple procedure, and convert them into 2-D matrix form. The 1-DCNN has the advantage that 1-D time series signals can be input directly without the need for cumbersome conversions.

The output receptive field of the n -th layer is:

$$r_n = r_{n-1} + ((k_n - 1) \times \prod_{i=1}^{n-1} s_i), n \geq 2, \tag{1}$$

where r_n is the receptive field size of the n -th layer, k_n is the filter size of the n -th layer, and s_i is the movement step size of the i -th layer filter.

According to the receptive field's design principle, the size of the neuron receptive field in the last layer is close to the input signal's length, that is, it satisfies the condition $r_n = L$, where L is the length of the input signal. The convolution kernel size is k , and the sliding step size of the convolution kernel is s . Each convolutional layer is followed by the maximum pooling layer, where the step size of the maximum pooling layer is $k_{pool} = 2$, and the sliding window of the maximum pooling layer is $k_{pool} = 2$. When $n > 2$, the receptive field of the convolutional layer is:

$$r_n = r_{n-1} + (k - 1) \times \prod_{i=1}^{n-1} s_i, \tag{2}$$

and the receptive field of the pooling layer is:

$$r_{n+1} = r_n + (k_{pool} - 1) \times \prod_{i=1}^{n-1} s_i \times s, \tag{3}$$

when $n > 2$, the n -th network layer is a convolutional layer and n is odd, then the difference between the front and back receptive fields is 2^{n-1} ; thus, the expression for the receptive field r_n of the neurons in the last pooling layer at the input signal, when n is even number, is:

$$r_n = k + 1 + 2k + 4k + \dots + 2^{\frac{n}{2}-1}k = 1 + 2^{\frac{n}{2}}k - k \tag{4}$$

According to the condition $r_n = L$, the value of k is obtained from:

$$k \approx \frac{L - 1}{2^{\frac{n}{2}} - 1}, \tag{5}$$

3.2. 1-DDCNN

Dilated convolution is also called expanded convolution; it replaces the traditional CNN pooling operation by introducing the expansion ratio, which can completely retain the feature information in the original signal so that the convolution kernel of the same size can obtain a larger receptive field.

$$\begin{cases} r_1 = k \\ r_n = r_{n-1} + (k_n - 1)l_n, \quad n \geq 2' \end{cases} \quad (6)$$

where r_n is the receptive field size of the n -th layer network structure, k_n is the convolutional kernel size of the n -th layer network structure, and l_n is the expansion rate of the n -th layer network structure.

Figure 2b shows the receptive fields' range in the third layer of the 1-DDCNN (output feature map 2) for the first layer(input sequence data) and the second layer (output feature map 1). The convolutional kernel size is 3×1 ($k = 3$). The step size is 1×1 . The expansion rate is 2 ($l_1 = l_2 = 2$). The receptive field size in the second layer corresponding to output feature 2 is 5×1 , and the receptive field size in the first layer is 9×1 .

3.3. 1-DDCNN Fault Diagnosis Model Framework

The structural framework of the proposed fault diagnosis method based on the 1-DDCNN is shown in Figure 3. The model fault diagnosis process was as follows:

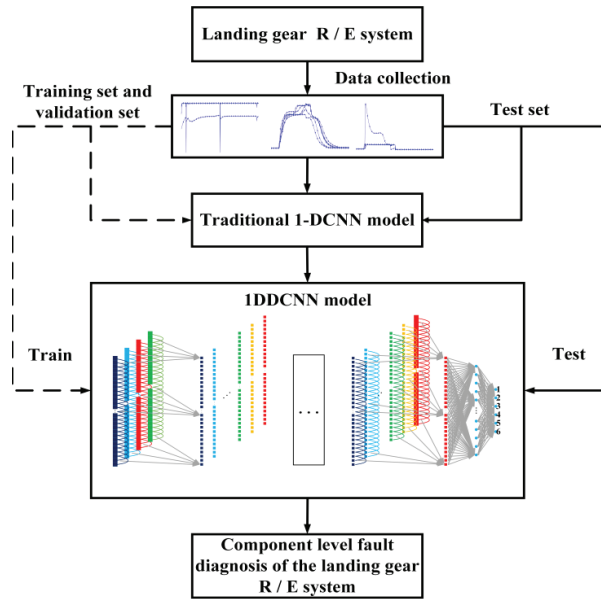


Figure 3. 1-DDCNN fault diagnosis process.

- Step 1: Datasets were divided into training set, validation set, and test set.
- Step 2: According to the structure and parameters of the traditional 1-DCNN model, the 1-DDCNN was preliminarily designed.
- Step 3: The diagnostic accuracy of the multi-feature 1-DDCNN model under different convolutional kernel sizes was investigated to determine the final model hyper-parameters.
- Step 4: The proposed model was trained and tested with a test set to obtain the fault diagnosis accuracy.

4. Experimental Implementations

4.1. Data Description and Operating Environment

Due to the fault data insufficiency regarding operation conditions, there is an incentive to use AMESim[®] to model the landing gear R/E system model and obtain fault datasets. According to the mutual logical relationship between components, the landing gear R/E system model was established, and is presented in Figure 4. The blue section represents the hydraulic subsystem, the green section signifies the mechanical subsystem, and the red section denotes the external load of the system. Component parameter settings in the model are shown in Table 2.

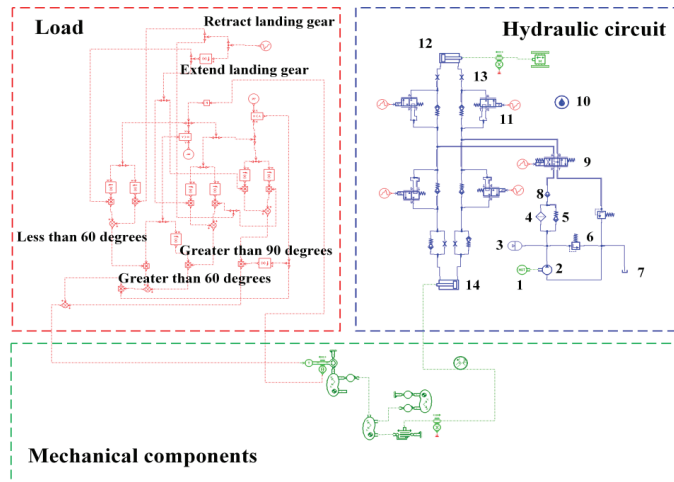


Figure 4. Simulation of landing gear R/E system. 1. Hydraulic motor; 2. Constant pressure variable pump; 3. Accumulator; 4. Filter; 5. Spring check valve; 6. Press control; 7. Oil tank; 8. Hydraulic check valve; 9. Three-position four-way directional control valve; 10. Hydraulic fluid; 11. Two-position three-way directional control valve; 12. Unlock actuator; 13. Flow control; 14. Actuator cylinder.

Table 2. Component parameter settings in the landing gear R/E system model.

| Sub-Model | Parameter | Value |
|---------------------------------|-----------------------------|---------------------------|
| Hydraulic fluid | Density | 850 kg/m ³ |
| | Bulk modulus | 1.7 × 10 ⁴ bar |
| | Temperature | 40 °C |
| | Air/gas content | 0.1% |
| Constant pressure variable pump | Pump displacement | 5 cc/rev |
| Hydraulic motor | Shaft speed | 5000 rev/min |
| | Equivalent orifice diameter | 6 mm |
| Filter | Critical flow number | 2320 |
| | Maximum flow coefficient | 0.7 |
| | Polytrophic index | 1.4 |
| Accumulator | Gas pre-charge pressure | 118 bar |
| | Initial pressure | 152 bar |
| | Accumulator volume | 2 L |
| Actuator cylinder | Piston diameter | 69.596 mm |
| | Rod diameter | 30.099 mm |
| | Length of stroke | 276.1 mm |
| | Leakage coefficient | 0 L/min/bar |
| Unlock actuator | Piston diameter | 25.12 mm |
| | Rod diameter | 11.049 mm |
| | Length of stroke | 70.84 mm |

Table 2. Cont.

| Sub-Model | Parameter | Value |
|---|--|--------------|
| Press control | Relief value cracking pressure | 206.843 bar |
| Throttle valve | Relief value flow rate pressure gradient | 20 L/min/bar |
| | Diameter | 2 mm |
| | Valve natural frequency | 80 Hz |
| Three-position four-way directional control valve | Valve damping ratio | 0.8 |
| | Flow rate | 39 L/min |
| | Pressure drop | 2.5 bar |
| | Valve rated current | 40 mA |

The specific parameters of the FMEA in Section 2.2 are shown in Table 3.

Table 3. Labels and failure threshold.

| Failure Status | Labels | Failure Threshold |
|--|--------|---|
| Normal | 1 | - |
| Excessive noise from the pump | 2 | Content of air in the hydraulic fluid is more than 5% |
| Throttle valve blocking at right end of actuating cylinder | 3 | The diameter of the throttle valve is less than 2.5 mm |
| Throttle valve blocking at left end of actuating cylinder | 4 | The diameter of the throttle valve is less than 2 mm |
| Constant pressure variable pump leakage | 5 | The diameter of the throttle valve is more than 3 mm |
| Actuator cylinder leakage | 6 | Leakage coefficient of actuator cylinder is greater than 0.01 |

The failure status:1 curve in the subgraphs a, b, c, and d of Figure 5 shows the main parameters' variation trends under normal conditions, and that the entire landing gear R/E process time is 32 s, during which the landing gear retraction time is 7.5 s and the extension time is 10.8 s. These times are similar to those specified in the manual, and the manual requires that the R/E time shall not exceed the specified time by 1 s, or it will be regarded as a fault [16].

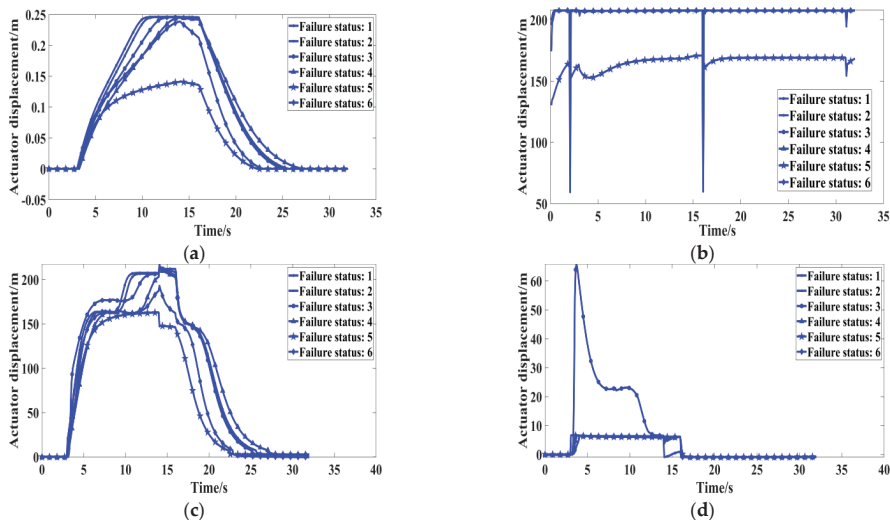


Figure 5. Comparison of the six fault types under different feature parameters. (a) The displacement of actuating cylinder; (b) system pressure; (c) the pressure at the right end of actuating cylinder; (d) the pressure at the left end of actuating cylinder.

According to the fault thresholds in Table 3, 300 simulations were conducted for each of the six fault states, and the four parameters (actuator cylinder displacement, system pressure, and the pressure at the right and left end of the actuating cylinder) were sampled. The sampling frequency was set as 0.01 to obtain 1800 samples. Training, validation, and test sets were divided as 8:1:1, respectively. The details of the single-feature and multi-feature datasets are shown in Tables 4 and 5, respectively, and the operation environment for the simulation is described in Table 6.

Table 4. Single-feature dataset A.

| Data Division | Failure Status | | | | | |
|----------------|----------------|-----|-----|-----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Training set | 240 | 240 | 240 | 240 | 240 | 240 |
| Validation set | 30 | 30 | 30 | 30 | 30 | 30 |
| Test set | 30 | 30 | 30 | 30 | 30 | 30 |

Table 5. Multi-feature dataset B.

| Data Division | Feature Parameters | Failure Status | | | | | |
|----------------|---|----------------|-----|-----|-----|-----|-----|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| Training set | System pressure | 240 | 240 | 240 | 240 | 240 | 240 |
| | The pressure at the right end of actuating cylinder | 240 | 240 | 240 | 240 | 240 | 240 |
| | The pressure at the left end of actuating cylinder | 240 | 240 | 240 | 240 | 240 | 240 |
| Validation set | System pressure | 30 | 30 | 30 | 30 | 30 | 30 |
| | The pressure at the right end of actuating cylinder | 30 | 30 | 30 | 30 | 30 | 30 |
| | The pressure at the left end of actuating cylinder | 30 | 30 | 30 | 30 | 30 | 30 |
| Test set | System pressure | 30 | 30 | 30 | 30 | 30 | 30 |
| | The pressure at the right end of actuating cylinder | 30 | 30 | 30 | 30 | 30 | 30 |
| | The pressure at the left end of actuating cylinder | 30 | 30 | 30 | 30 | 30 | 30 |

Table 6. Experimental operation environment.

| Operating Environment | Version |
|-------------------------|------------------------------------|
| CPU | Intel(R) Core(TM) i5-9300H 2.40GHz |
| System | Windows 10 Home Chinese Version |
| Interpreter | Anaconda2.0.3 |
| Compiler | PyCharm2019.3.4 |
| Deep learning framework | Tensorflow2.0 |

4.2. Experimental Model

Zhou [17] analyzed the following three important factors that have an impact on the performance of the CNN: network organization structure, network depth, and feature maps number. On the one hand, increasing the network depth can improve the recognition accuracy; on the other hand, increasing the feature maps number can also improve the recognition accuracy. Therefore, it is necessary to conduct a comparative study separately to determine the final model parameters. From Section 4.1, it is known that the sequence length of a single sample is 3201. The total number of convolutional and pooling layers is $n = 12$ (excluding the dropout layer). On the basis of Equation (5), the convolution kernel size of the first convolutional layer is 50. From the comparative test, the model with convolution kernel size 50, convolution number 4, and moving step size 1 at the first convolution layer, has the best diagnostic effect. The specific parameters of the traditional 1-DCNN model are shown in Table 7.

Table 7. 1-DCNN model details.

| No. | Layer Type | Kernel Size/Stride | Kernel Number | Output Size (Width × Depth) | Padding | Dropout Rate |
|-----|--------------|--------------------|---------------|-----------------------------|---------|--------------|
| 1 | Convolution1 | 50 × 1/1 × 1 | 4 | 3201 × 4 | Yes | - |
| 2 | MaxPooling1 | 2 × 1/2 × 1 | 4 | 1600 × 4 | No | - |
| 3 | Convolution2 | 3 × 1/1 × 1 | 16 | 1600 × 16 | Yes | - |
| 4 | MaxPooling2 | 2 × 1/2 × 1 | 16 | 800 × 16 | No | - |
| 5 | Dropout1 | - | - | 800 × 16 | - | 0.2 |
| 6 | Convolution3 | 3 × 1/1 × 1 | 16 | 800 × 16 | Yes | - |
| 7 | MaxPooling3 | 2 × 1/2 × 1 | 16 | 400 × 16 | No | - |
| 8 | Convolution4 | 3 × 1/1 × 1 | 32 | 400 × 32 | Yes | - |
| 9 | MaxPooling4 | 2 × 1/2 × 1 | 32 | 200 × 32 | No | - |
| 10 | Dropout2 | - | - | 200 × 32 | - | 0.2 |
| 11 | Convolution5 | 3 × 1/1 × 1 | 32 | 200 × 32 | Yes | - |
| 12 | MaxPooling5 | 2 × 1/2 × 1 | 32 | 100 × 32 | No | - |
| 13 | Dropout3 | - | - | 100 × 32 | - | 0.2 |
| 14 | Convolution6 | 3 × 1/1 × 1 | 64 | 100 × 64 | Yes | - |

Activation function of the convolution layer: ReLU
 Activation function of fully-connected layer: Softmax
 Optimization: Adam(1.0×10^{-5})
 Iteration: 100

Due to the limited feature information extraction and inaccurate classification of the traditional 1-DCNN model with a single-feature parameter (actuator cylinder displacement), the three features, e.g., system pressure and the pressure at the right and left end of actuating cylinder, are selected to jointly characterize six failure statuses.

Referring to the traditional 1-DCNN, in 1-DDCNN we initially set the convolution kernel size as 50, the step size as 1, and the expansion factor as $l_n = 2^{n-1}$, the calculation formula for the receptive field is:

$$r_n = 1 + (k - 1)(2^n - 1) \tag{7}$$

The network structure and initial settings are shown in Figure 6 and Table 8, respectively. The design principle of the model is that the output feature graph size of the last convolution layer is similar to, or exactly the same as, the size of the input data. The proposed 1-DDCNN has the following advantages: firstly, it constructs the convolutional kernel to obtain a larger receptive field and completely retain the feature information in the original signal; secondly, it can act as a dropout layer to prevent over-fitting.

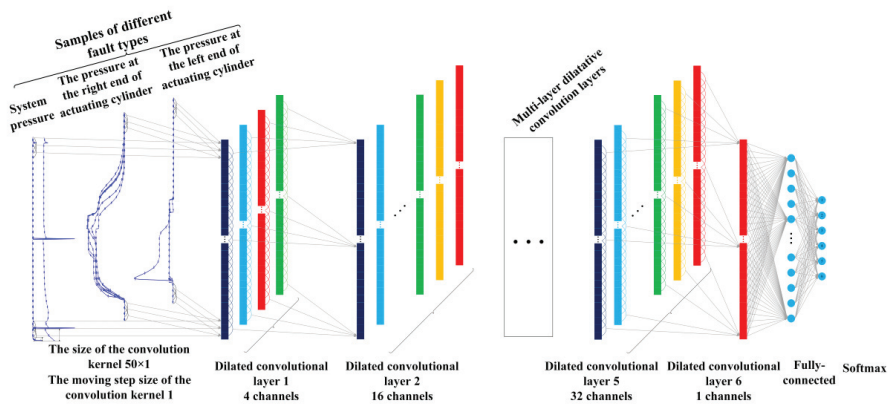


Figure 6. The structure of 1-DDCNN.

Table 8. 1-DDCNN model details.

| No. | Layer Type | Kernel Size/Stride | Kernel Number | Expansion Factor | Output Size (Width × Depth) | Padding | Receptive Field |
|-----|-----------------|--------------------------|---------------|------------------|-----------------------------|---------|-----------------|
| 1 | Convolution 1 | $50 \times 1/1 \times 1$ | 4 | 1 | 3201×4 | Yes | 50 |
| 2 | Convolution 2 | $50 \times 1/1 \times 1$ | 16 | 2 | 3201×16 | Yes | 148 |
| 3 | Convolution 3 | $50 \times 1/1 \times 1$ | 16 | 4 | 3201×16 | Yes | 344 |
| 4 | Convolution 4 | $50 \times 1/1 \times 1$ | 32 | 8 | 3201×32 | Yes | 736 |
| 5 | Convolution 5 | $50 \times 1/1 \times 1$ | 32 | 16 | 3201×32 | Yes | 1520 |
| 6 | Convolution 6 | $50 \times 1/1 \times 1$ | 1 | 32 | 3201×1 | Yes | 3088 |
| 7 | Fully-connected | 256 | 1 | - | 256×1 | - | - |
| 8 | Softmax | 6 | 1 | - | 6 | - | - |

4.3. Experimental Results and Analysis

4.3.1. Research on the Size of Convolution Kernel of 1-DDCNN Model

According to Equation (7), once the expansion factor is determined, the parameter that has a decisive influence on the receptive field size is the convolution kernel size. Since the output size of each dilated convolution layer in the 1-DDCNN is 3201×1 , the convolution kernel size does not affect the output features' size of the dilated convolution layer, but has a great influence on the feature extraction degree of the original data. Therefore, it is necessary to investigate the convolutional kernel size's effect on the classification accuracy. The convolution kernel size was set to 30, 40, 50, 60, and 70 in turn to investigate the diagnostic accuracy of the 1-DDCNN under different conditions, and to determine the final model hyper-parameters.

Table 9 and Figure 7 show the detailed diagnosis results for the effect of convolution kernel size on the test samples in each trial. As the convolution kernel size increases, the total training parameters rise, and the model running time expands accordingly. The average accuracy of the 1-DDCNN with different convolution kernel sizes reached more than 90%. In particular, when the convolution kernel size was 40, the highest average accuracy of 99.80% was achieved in five training sessions. When the convolution kernel size was 50, its standard deviation was at least 0.0000, which indicates that the model had the highest stability under this condition.

Table 9. Fault diagnosis results of 1-DDCNN under different sizes of convolution kernels.

| The Convolution Kernel Size | Average Accuracy (%) | Standard Deviation | Total Parameter | Average Running Time/s |
|-----------------------------|----------------------|--------------------|-----------------|------------------------|
| 30 | 99.00 | 0.0071 | 878355 | 187.2152 |
| 40 | 99.80 | 0.0045 | 897355 | 211.7023 |
| 50 | 99.00 | 0.0000 | 916355 | 231.7082 |
| 60 | 90.80 | 0.1355 | 935355 | 262.0429 |
| 70 | 96.20 | 0.0740 | 953355 | 280.4578 |

Figure 8 shows that accuracies were close to 100% at the 10th iteration with convolution kernel size ranging from 30×1 to 50×1 , and fluctuated around 94.5% from the 50th iteration onwards, with convolution kernel size ranging from 60×1 to 70×1 . In fact, when the convolution kernel size is 60×1 , the accuracy actually dropped. It can be seen from Figure 9 that the loss values within the convolution kernel size range of 30×1 to 50×1 approach 0 at the 10th iteration. The loss value at convolution kernel size 60×1 remained around 0.92 after 10th iteration, which indicates over-fitting. The loss value of the model corresponding to the convolution kernel size 70×1 remains around 0.14 after the 40th iteration. In particular, at the 65th iteration, the training and validation loss values at convolution kernel size 40×1 were both less than 1.0×10^{-5} .

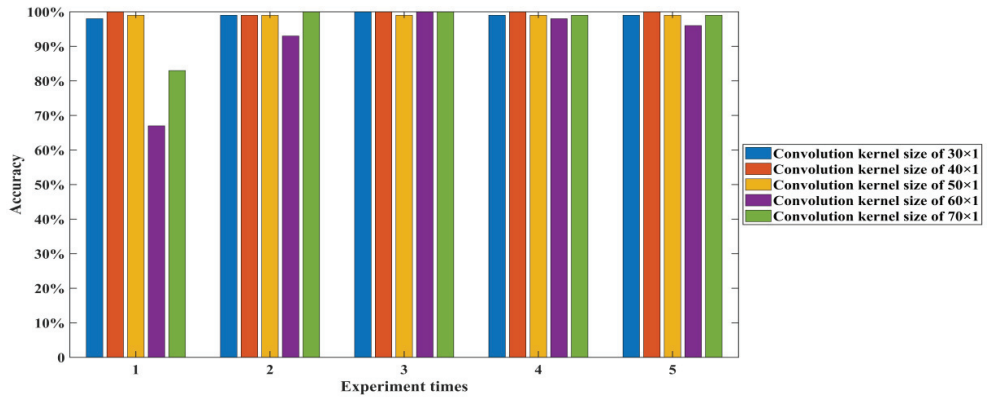


Figure 7. Test accuracy of 1-DDCNN under different sizes of convolution kernels.

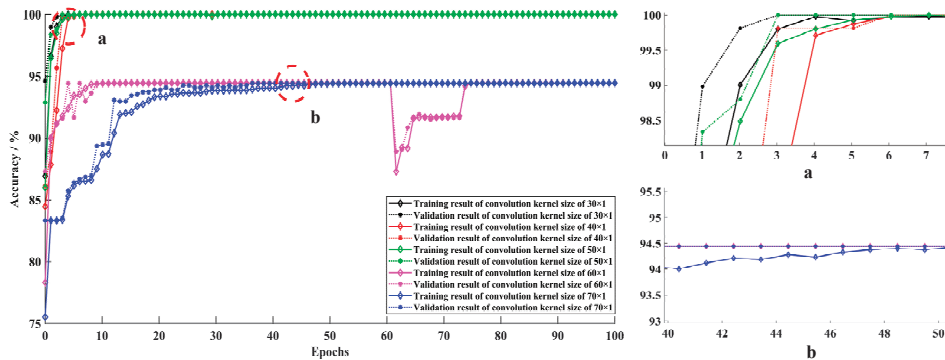


Figure 8. Training and validation accuracy at different convolution kernel sizes.

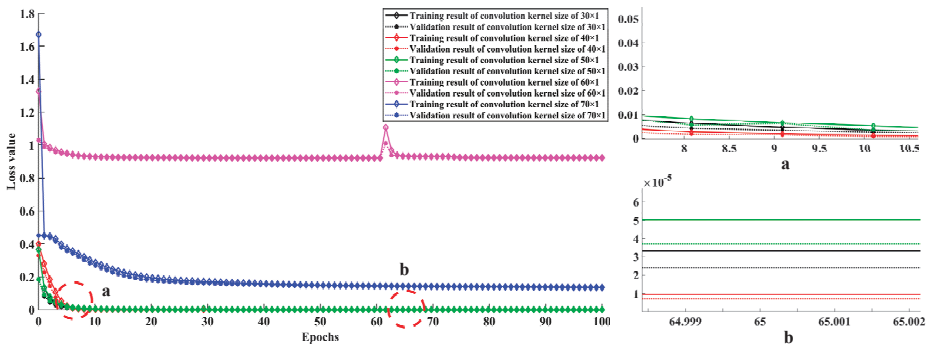


Figure 9. Training and validation loss values at different convolution kernel sizes.

Considering the accuracy, stability, and training cost comprehensively, results are optimal when convolution kernel size is set as 40×1 .

4.3.2. Comparative Experiment of Three Models under Different Datasets

To show the dilated convolution’s advantages in feature extraction and information loss prevention, the comparative experiments of three models, e.g., traditional 1-DCNN, 1-DDCNN, and 1-DDCNN II, with dataset A and dataset B, were conducted. Depending

on whether the same size is maintained between the feature map and the input data, two types of convolution operations exist: VALID (without padding) and SAME (with padding) convolution operations. Compared to the 1-DDCNN, the 1-DDCNN II's dilated convolution layer was VALID, and the convolution kernel size was uniformly set to 51×1 . Inputting the dataset into 1-DDCNN II, the output size of the flattening layer was 3264×1 after six dilated convolution layers, which is slightly larger than the sequence length of the input samples. The model structure is similar to the 1-DDCNN, and specific parameter settings are shown in Table 10.

Table 10. 1-DDCNN II details.

| No. | Layer Type | Kernel Size/Stride | Kernel Number | Expansion Factor | Output Size (Width×Depth) | Padding | Receptive Field |
|-----|-----------------|--------------------------|---------------|------------------|---------------------------|---------|-----------------|
| 1 | Convolution 1 | $50 \times 1/1 \times 1$ | 4 | 1 | 3151×4 | No | 51 |
| 2 | Convolution 2 | $50 \times 1/1 \times 1$ | 16 | 2 | 3051×16 | No | 151 |
| 3 | Convolution 3 | $50 \times 1/1 \times 1$ | 16 | 4 | 2851×16 | No | 351 |
| 4 | Convolution 4 | $50 \times 1/1 \times 1$ | 32 | 8 | 2451×32 | No | 751 |
| 5 | Convolution 5 | $50 \times 1/1 \times 1$ | 32 | 16 | 1651×32 | No | 1551 |
| 6 | Convolution 6 | $50 \times 1/1 \times 1$ | 64 | 32 | 51×64 | No | 3151 |
| 7 | Fully-connected | 256 | 1 | - | 256×1 | - | - |
| 8 | Softmax | 6 | 1 | - | 6 | - | - |

It can be seen from Table 11 and Figure 10 that, compared with dataset A, the diagnostic accuracies of the 1-DCNN, 1-DDCNN, and 1-DDCNN II with dataset B were higher, and the total training parameters and training time increased slightly. The average accuracies of the 1-DDCNN and 1-DDCNN II reached more than 99%, and the standard deviation of both was 0.0045, which indicates that both models are stable.

Table 11. Diagnosis results of different models.

| Models | Datasets | Average Accuracy (%) | Standard Deviation | Total Parameter | Average Running Time/s |
|------------|----------|----------------------|--------------------|-----------------|------------------------|
| 1-DCNN | A | 91.80 | 0.0399 | 833074 | 60.9052 |
| 1-DDCNN | | 75.60 | 0.3278 | 897035 | 191.7420 |
| 1-DDCNN II | | 93.60 | 0.0055 | 1036854 | 153.4254 |
| 1-DCNN | B | 97.40 | 0.0089 | 833474 | 76.4105 |
| 1-DDCNN | | 99.80 | 0.0045 | 897355 | 211.7023 |
| 1-DDCNN II | | 99.20 | 0.0045 | 1037262 | 171.7298 |

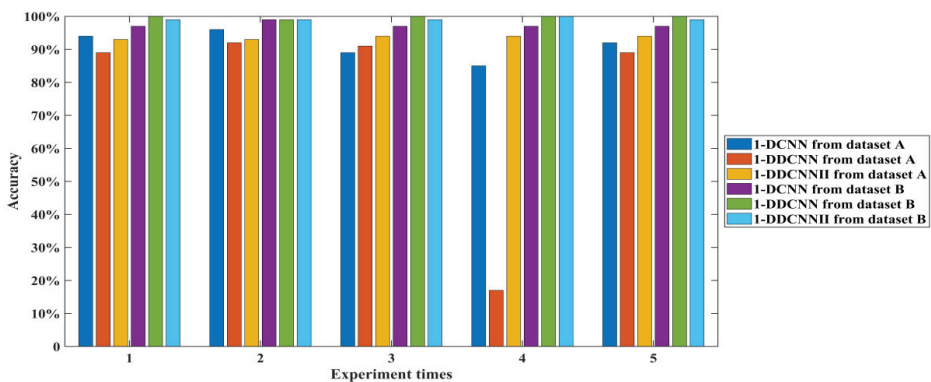


Figure 10. Test accuracy of different models from dataset A and dataset B.

The training processes and confusion matrices for the three models are shown in Figures 11–13.

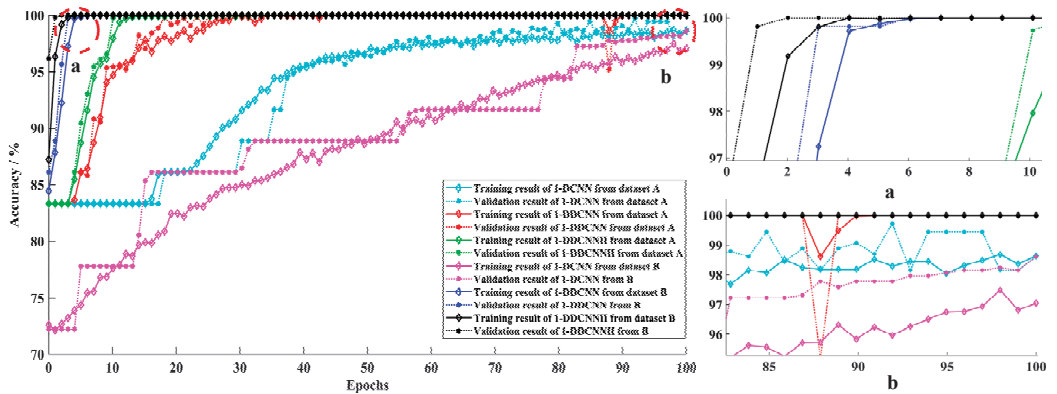


Figure 11. Training and validation accuracy of different models from dataset A and B.

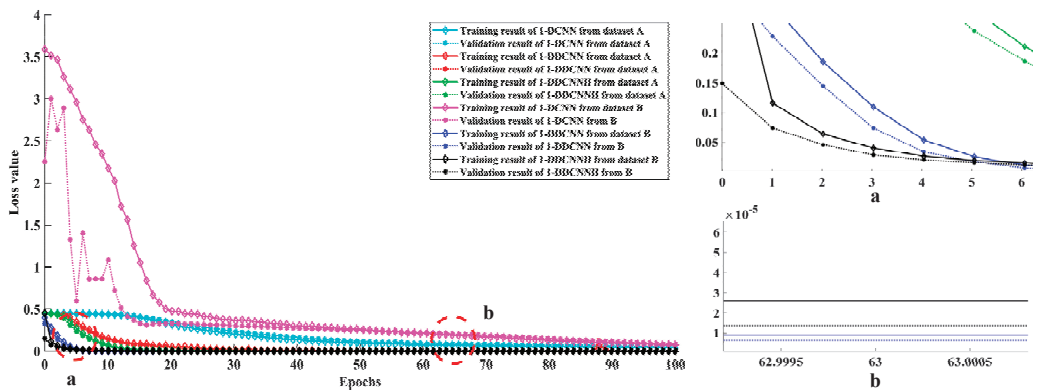


Figure 12. Training and validation loss value of different models from dataset A and B.



Figure 13. Confusion matrix of diagnosis results of different models based on dataset A and B. The following are the results of: (a) 1-DCNN from dataset A; (b) 1-DCNN from dataset B; (c) 1-DDCNN from dataset A; (d) 1-DDCNN from dataset B; (e) 1-DDCNN II from dataset A; (f) 1-DDCNN II from dataset B.

In Figure 13, the rows represent the predicted class (output class) and the columns represent the true class (target class). The diagonal cells represent the observations that were correctly classified. The off-diagonal cells represent incorrectly classified observations. Both the observation number and the percentage of the total observation numbers, are shown in each cell. The far-right column in the plot shows the percentages of all the predicted examples belonging to each class that were correctly and incorrectly classified. These metrics are often called the precision and false rate, respectively. The bottom row in the plot shows the percentages of all the examples belonging to each class that are correctly and incorrectly classified. These metrics are often called the recall (or true positive rate) and false negative rate, respectively. The cell in the bottom right of the plot shows the overall accuracy.

When the 1-DCNN model with dataset A was located at the 100th iteration, the training accuracy was 98.63%, and the training loss value was 0.0441. The validation accuracy was 98.61% and the validation loss value was 0.0609. The confusion matrix corresponding to the test set is shown in Figure 13a; four fault samples caused by “excessive noise from the pump” were incorrectly classified as “normal”, and this corresponded to 2.2% of all 180 samples in dataset A. Similarly, six fault samples caused by “throttle valve blocking at left end of actuating cylinder” were incorrectly classified as “excessive noise from the pump”, and this corresponded to 3.3% of all data. Overall, 93.9% of the classifications were correct and 6.1% were wrong. It is believed that the potential reason for the identification error was the high similarities between sample sequences of different failure status, as shown in Figure 5a.

It can be seen in Figure 11 that the accuracies of both 1-DDCNNs with dataset B were close to 100% at the 10th iteration, while the accuracy of the traditional 1-DCNN model with dataset A and B did not reach 100%, even after 100 epochs, which indicates that the traditional 1-DCNN model loses a large amount of information during the pooling process.

Figure 12 shows the loss values of the 1-DDCNN and 1-DDCNN II with dataset B are both close to 0 at the 10th iteration; the convergence trend of the loss value of the 1-DCNN with dataset B was slower than that of the other models. The training and validation loss value convergence curve of the 1-DCNN had a significant gap after the previous 20 iterations. In particular, at the 63th iteration, training and validation loss values of the 1-DDCNN with dataset B were both less than 1.0×10^{-5} . In terms of training cost, the iteration number should be set as 63 in subsequent model training.

It can be seen in Figure 13 that the test accuracies of subgraphs b, d, and f are higher than those of the subgraphs a, c, and e. In particular, the test accuracy of the 1-DDCNN (subgraph d) with dataset B reached 100%.

Figure 14 shows the output features visualization of Conv1 and Flatten layers in the 1-DDCNN with six failure statuses. The feature similarity of the four-channel output of Conv1 was relatively high for the six failure statuses. After convolution and flattening operations, the six failure statuses exhibited unique characteristics, which are conducive to distinguishing the failure status of the model.

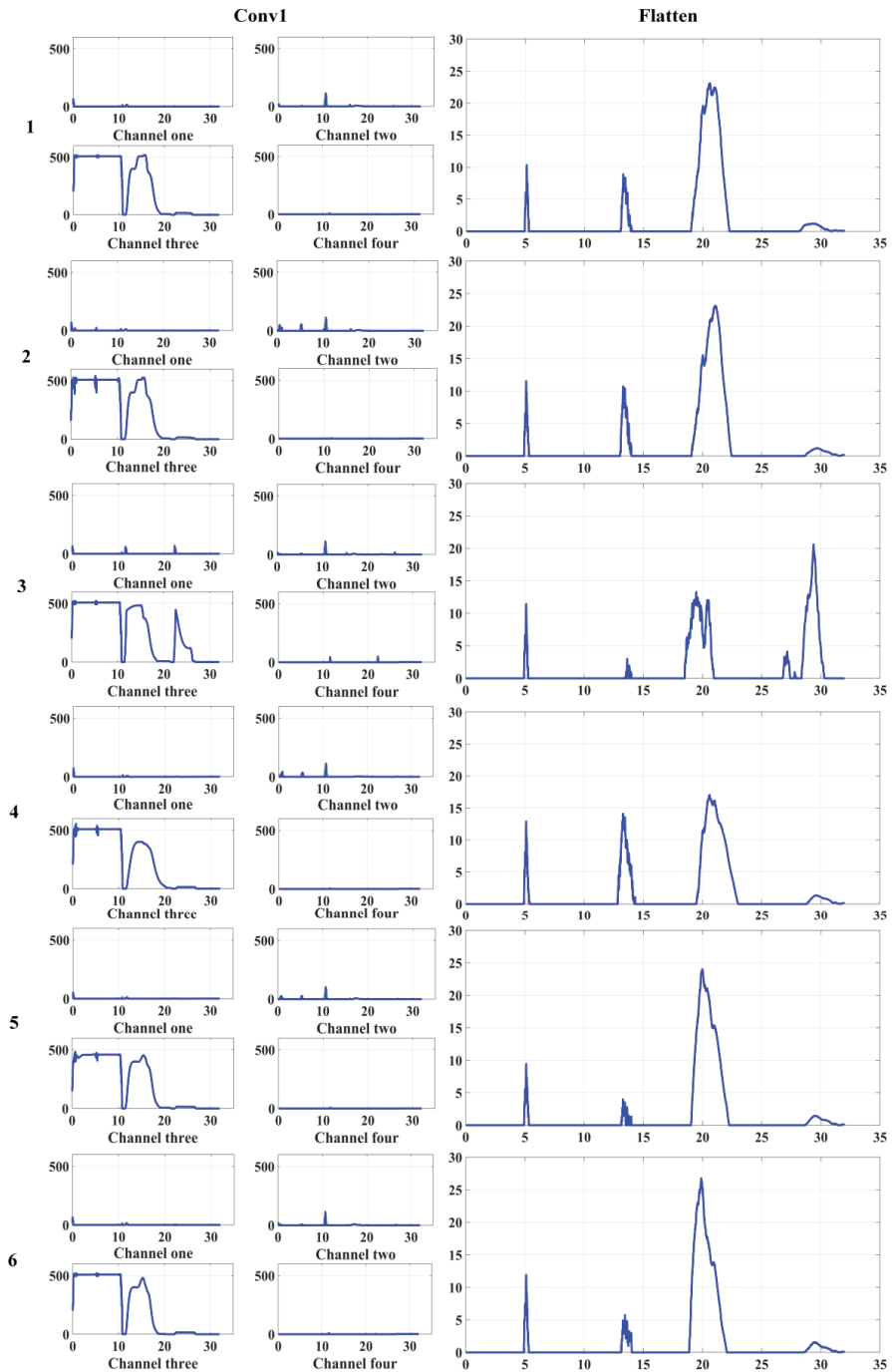


Figure 14. Visualization of output features of Conv1 and Flatten layers with the six failure statuses of 1-DDCNN. (1–6) correspond to the six failure statuses in Table 3.

5. Conclusions

Since the 2-DCNN cannot directly process one-dimensional time series data, which often requires complex pre-processing, a novel 1-DDCNN is proposed for landing gear R/E system fault diagnosis in this paper. Dilated convolution can exponentially increase the receptive field of the convolution kernel by adding the convolution layer, which could acquire more redundant information to alleviate the influence of randomness. The displacement of the actuator cylinder was selected as the feature parameter, and the diagnosis classification was carried out on the traditional 1-DCNN model, for which the average diagnosis accuracy reached 91.80%. Due to the limited feature information extraction and inaccurate diagnosis for a single feature in the traditional 1-DCNN, multiple feature parameters are selected to jointly represent the fault and to input into the proposed model for feature integration. The convolution kernel size's influence on classification accuracy is explored. When the convolution kernel size is 50, the model has the highest stability. The results show that the average diagnostic accuracy of the proposed model is 99.80%, compared with other models.

Future work will be carried out on the following two aspects. Firstly, the system has noise in the actual working environment, and it is necessary to verify the robustness of the proposed model on noisy data. Secondly, this paper only considers the influence of a single parameter, such as oil mixing into the air or actuator leakage, on the system operating state. In the future, the complex situation of the simultaneous failure of multiple internal components, and the consequent effects on the system operating state, should be studied.

Author Contributions: J.C.: Resources, Funding acquisition, Project administration, Writing—Review & editing. Q.X.: Methodology, Software, Writing—Original. Y.G.: Data acquisition, Investigation. R.C.: Validation, Project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Nature Science Foundation of China (Grant no.61873203), China Academy of Space Technology Innovation Foundation(CAST-2020-02-11), Aeronautical Science Foundation of China (Grant No.20200033053001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

References

1. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
2. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)] [[PubMed](#)]
3. LeCun, Y. Generalization and network design strategies. *Connect. Perspect.* **1989**, *19*, 143–155.
4. Gligorijevic, J.; Gajic, D.; Brkovic, A. Online Condition Monitoring of Bearings to Support Total Productive Maintenance in the Packaging Materials Industry. *Sensors* **2016**, *16*, 316. [[CrossRef](#)] [[PubMed](#)]
5. Janssens, O. Convolutional neural network based fault detection for rotating machinery. *J. Sound Vib.* **2016**, *377*, 331–345. [[CrossRef](#)]
6. Jing, L. An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox. *Sensors* **2017**, *17*, 414. [[CrossRef](#)] [[PubMed](#)]
7. Wen, L. A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Trans. Ind. Electron.* **2017**, *65*, 5990–5998. [[CrossRef](#)]
8. Wu, C.; Jiang, P.; Ding, C. Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network. *Comput. Ind.* **2019**, *108*, 53–61. [[CrossRef](#)]
9. Huang, W.; Cheng, J.; Yang, Y. An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis. *Neurocomputing* **2019**, *359*, 77–92. [[CrossRef](#)]
10. Pan, H. An Improved Bearing Fault Diagnosis Method using One-Dimensional CNN and LSTM. *J. Mech. Eng.* **2018**, *64*, 443–452.
11. Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. *arXiv* **2015**, arXiv:1511.07122.
12. Zhuang, Z. A deep learning method for bearing fault diagnosis through stacked residual dilated convolutions. *Appl. Sci.* **2019**, *9*, 1823. [[CrossRef](#)]
13. Zhang, C.; Feng, J.R. An Intelligent Fault Diagnosis Method of Rolling Bearing Under Variable Working Loads Using 1-D Stacked Dilated Convolutional Neural Network. *IEEE Access* **2020**, *8*, 63027–63042. [[CrossRef](#)]

14. Liang, Y.; Li, B.B.; Jiao, B. A deep learning method for motor fault diagnosis based on a capsule network with gate-structure dilated convolutions. *Neural Comput. Appl.* **2021**, *33*, 1401–1418. [[CrossRef](#)]
15. Wang, H.T. *Aircraft Hydraulic Components and Systems*; National Defense Industry Press: Beijing, China, 2012.
16. Shi, X.D.; Ying, F.J. Performance simulation and fault analysis of aircraft landing gear extension and retraction system. *Comput. Appl. Softw.* **2021**, *38*, 65–71.
17. Zhou, F.Y.; Jin, L.P.; Dong, J. Review of convolutional neural network. *Chin. J. Comput.* **2017**, *40*, 1229–1251.

Article

Study on a Fault Mitigation Scheme for Rub-Impact of an Aero-Engine Based on NiTi Wires

Qiang Pan ¹, Tian He ^{1,*}, Wendong Liu ², Xiaofeng Liu ¹ and Haibing Chen ¹

¹ School of Transportation Science and Engineering, Beihang University, Beijing 100191, China; panqiang@buaa.edu.cn (Q.P.); liuxf@buaa.edu.cn (X.L.); aw_chenhb@buaa.edu.cn (H.C.)

² AECC Shenyang Engine Research Institute, Shenyang 110000, China; lwdbuaa@163.com

* Correspondence: hetian@buaa.edu.cn; Tel.: +86-131-4128-1907

Abstract: The aim of this study was to solve the frequently occurring rotor-stator rub-impact fault in aero-engines without causing a significant reduction in efficiency. We proposed a fault mitigation scheme, using shape memory alloy (SMA) wire, whereby the tip clearance between the rotor and the stator is adjusted. In this scheme, an acoustic emission (AE) sensor is utilized to monitor the rub-impact fault. An active control actuator is designed with pre-strained two-way SMA wires, driven by an electric current via an Arduino control board, to mitigate the rub-impact fault once it occurs. In order to investigate the feasibility of the proposed scheme, a series of tests on the material properties of NiTi wires, including heating response rate, ultimate strain, free recovery rate, and restoring force, were carried out. A prototype of the actuator was designed, manufactured, and tested under various conditions. The experimental result verifies that the proposed scheme has the potential to mitigate or eliminate the rotor-stator rub-impact fault in aero-engines.

Keywords: rub-impact; tip clearance; shape memory alloy; aero-engine; fault mitigation

Citation: Pan, Q.; He, T.; Liu, W.; Liu, X.; Chen, H. Study on a Fault Mitigation Scheme for Rub-Impact of an Aero-Engine Based on NiTi Wires. *Sensors* **2022**, *22*, 1796. <https://doi.org/10.3390/s22051796>

Academic Editors: Bing Li, Jinchun Ji, Hamed Kalhori and Yongbo Li

Received: 21 January 2022

Accepted: 21 February 2022

Published: 24 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The efficiency of a rotating machine such as an aero-engine is strongly dependent on the tip clearance between the stationary and rotating parts [1]. In order to improve the efficiency of an aero-engine, the clearance should be designed as small as possible. It has been reported that a 0.0254 mm reduction in the tip clearance of a high-pressure turbine may lead to a decrease of 0.1 percent in specific fuel consumption [2,3]. However, minimizing the clearance is usually associated with undesired rub-impact phenomena occurring between the rotor and the casing due to mechanical, aerodynamic excitation, or thermal gradient during engine operation [4,5]. This leads to material or structural damage, e.g., plastic deformations, changes in the microstructure on the blade tips, crack initiation, and the break out of liner material at the rubbing zone [6], and, sometimes, catastrophic accidents [7].

Researchers have strived to understand the rub-impact fault, for example, exploring the fault mechanism [8–10], the fault feature extraction method [11–13], source localization [14,15], and intelligent fault diagnosis [16,17], etc. These studies contributed to the development of rub-impact fault diagnosis and to practical applications. However, most studies focused on how to monitor the occurrence of a fault and analyze its type or location; few were concerned with the mitigation or elimination of the rub-impact phenomenon. Unfortunately, rub-impact is a fault which may induce disastrous accidents, if it cannot be mitigated in time.

In order to protect the rotor from rub-impact from the casing of an aero-engine, a feasible solution is to adjust the tip clearance when it occurs. The most popular scheme to achieve this that is used in the design of aero-engines is based on the active clearance control (ACC) method [18]. For example, Bucaro et al. designed a new gas turbine engine thermal control device to improve the control efficiency in terms of the active thermal

control system [19]. Tillman et al. proposed a system and method of active thermal control that includes processing aircraft data when the aircraft is flying at altitude cruise conditions [20]. Decastro et al. used two actuators consisting of electrohydraulic servo valves and piezoelectric stacks to adjust the shroud in a high-pressure turbine section [21]. However, ACC is generally used to schedule the clearance to improve the performance of an operating aero-engine. The actuators based on ACC management do not execute any action spontaneously when a rub-impact occurs between the rotor and casing, though such events frequently occur [22].

Recently, researchers have explored ways of limiting the rub-impact in an aero engine. They have proposed adding damping in the rotating machinery to control the abnormal vibration and reduce the rubbing-induced stress. Ma et al. [23] proposed a multi-objective optimization method to analyze the vibration attenuation effects of the squeeze film damper parameters on the dynamic response of the system. It was found that in the misalignment-rubbing coupling fault, the amplitude of the fundamental frequency reduced by 7.4%, the amplitude of $2 \times$ the fundamental frequency dropped by 51.5%, and the amplitude of $3 \times$ the fundamental frequency reduced by 16.8%. Their study provided a theoretical reference for vibration control and the optimal design of rotating machinery. Xu et al. [24] suggested an impulsive control method to eliminate the rotor-stator rubbing, based on the phase characteristic. It utilized the vibration energy and the phase difference to trigger the control of the rotor-stator rubbing by impulse. The impulse was applied in directions x and y , several times, to avoid the rotor-stator rubbing. However, it is a theoretical study, applicable to simple rotors only, and it did not account for the implementation of impulse in practical rotating machinery. Shang et al. [25] investigated the influence of cross-coupling effects on the rubbing-related dynamics of rotor/stator systems. They proposed a control method by generating cross-coupling damping on the stator through the active auxiliary bearing, thereby suppressing the contact severity to avoid rubbing instability. This method was validated by numerical analysis on the Jeffcott rotor model, yet it lacks theoretical analysis and experimental data for complicated rotating systems. However, adding damping to the vibration system is beneficial with minor faults but it is not a valid approach once the rub impact is serious. The research into adding damping in the rotating machinery is not yet well developed.

It has been recognized that rub-impact originates from a very small tip clearance between the stationary and rotating parts in an operating aero-engine [26]. If the clearance can be systematically controlled, the rub-impact might be effectively mitigated or eliminated. Garg [27] suggested shape memory alloy (SMA) as an attractive actuation material in an aero-engine due to its high order-of-magnitude energy density and low energy consumption. DeCastro et al. [28] advanced the concept of a prototype actuator consisting of high-temperature shape memory alloys for ACC actuation in the high-pressure turbine section of a modern turbofan engine. Based on the intrinsic properties of SMA and on the development of tip clearance control technology management, this paper presents a method to monitor rub-impact faults and mitigate them via SMA. The proposed method is schematically shown in Figure 1.

This paper presents a promising solution for rub-impact fault detection and mitigation for an aero-engine. An active control scheme, based on two-way SMA wires actuation is proposed. The rub-impact fault is monitored by an acoustic emission (AE) sensor. An SMA wire-based actuator prototype operated by an Arduino control board is established. A series of tests to establish the material properties of NiTi wires, including heating response rate, ultimate strain, free recovery rate, and restoring force, were carried out. The mechanism and design of the actuator are described in detail in this paper. The feasibility of the proposed model for rub-impact fault is verified by our experimental research and our results show that the proposed active control actuator can effectively mitigate rub-impact fault when it occurs. Therefore, the main contribution of this paper is that it provides a potential way to mitigate or even eliminate the accidental rub-impact fault, without a significant reduction in the engine's efficiency.

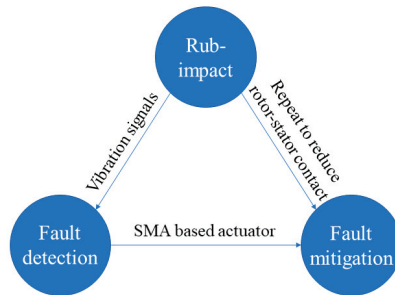


Figure 1. The basic idea of the proposed rub-impact fault mitigation method.

2. Mechanical Behaviors of SMA Wires

Typical SMA, such as NiTi, is capable of recovering its original shape after plastic deformation by heating above its characteristic transition temperature via the shape memory effect (SME) [29]. This unique mechanical behavior results from a phase transformation between high-temperature austenite and low-temperature martensite phases [30] and has led to many actuation applications [31,32]. In this study, NiTi wires with a two-way shape memory effect [30] were used in our design to drive the actuator through the application of an electric current. As the primary functional component of the actuator, the SME mechanical behaviors of NiTi wires were analyzed in a series of tests prior to design assembly. The material properties of the NiTi wires used in this study and their transition temperatures are listed in Tables 1 and 2, respectively.

Table 1. Material properties of NiTi wires.

| Name | Composition | Density (Kg·m ⁻³) | Melting Point (°C) |
|------|-------------|-------------------------------|--------------------|
| NiTi | 50.3 at%Ni | 6500 | 1295 |

Table 2. Phase transition temperatures of NiTi wires.

| M _f (°C) | M _s (°C) | A _s (°C) | A _f (°C) |
|---------------------|---------------------|---------------------|---------------------|
| 19.8 | 36.3 | 39.7 | 50.1 |

2.1. Heating Response Rate of NiTi Wires

The heating response rate of the NiTi wire is used as a measure of how fast the wire's SME function works. This function is represented by the phase transformation spending time, which is the time it takes the wire to reach its austenite finish temperature A_f from its starting state. The shorter the phase transformation spending time, the higher the heating response rate. In order to investigate the heating response rate of NiTi wires, a series of tests under various temperatures were conducted. The experimental setup is schematically shown in Figure 2. The test rig consisted of a DC power supply, a tension test platform, NiTi wire, a thermocouple, and a data acquisition system.

The length and diameter of the NiTi wire were 100 mm and 0.8 mm, respectively. During the tests, the wire was clamped on a tension test platform and one thermocouple was attached to the wire to measure its surface temperature. At room temperature $T_0 = 21$ °C, the wires were heated by 2, 3, 4, and 5 A electric currents provided by the DC power supply. The temperature variations of the wires are shown in Figure 3. The temperature of the NiTi wire rose slowly initially, becoming almost invariable, with an increased electric current. This means that the rising temperature rate of the NiTi wire gradually decreases, and tends to become stable when subjected to a continuous electric current. Figure 3 also indicates that the rising temperature rate of the wire is dependent on the amplitude of the electric current. A large current amplitude is associated with a high rising temperature rate. In

order to evaluate the response of the NiTi wires to different electric currents, the phase transformation spending time, which is the time that it takes for the wire to reach A_f from its starting state, was measured, using various currents. The comparison results are shown in Table 3. The heating response rate rose with an increased electric current, but an increase in the response rate was not obvious when the electric current increased beyond 6 A.

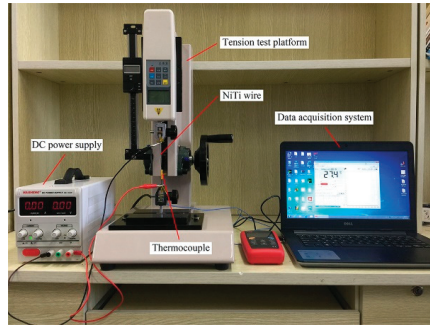


Figure 2. Experimental setup of the heating response rate tests.

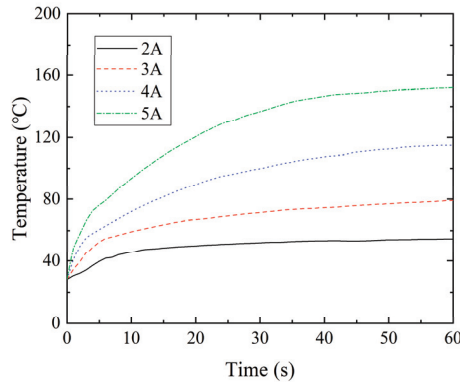


Figure 3. Temperature variation of NiTi SMA wire subjected to different currents.

Table 3. The phase transformation spending time to reach A_f using different currents.

| I (A) | 2 | 2.5 | 3 | 3.5 | 4 | 5 | 6 | 7 | 8 |
|-------|------|-----|-----|-----|-----|-----|-----|-----|-----|
| t (s) | 14.9 | 6.3 | 3.2 | 1.9 | 1.5 | 1.1 | 0.7 | 0.4 | 0.3 |

2.2. Ultimate Strain of NiTi Wires

A tension test was conducted using an Instron 5565 Universal Testing Machine. The length of the wire specimen used was 750 mm. Because the wire is very fine, it is difficult to accurately measure the ultimate strain directly during its deformation. As such, a load-displacement curve was plotted from the tension test results, as shown in Figure 4. The wire underwent elastic deformation from 0 to point A, plastic deformation from point A to B, a hardening stage from point B to C, and it fractures at point C. The elongations at points A, B, and C are 0.68%, 4.89%, and 13.33% of the original length of the wire, respectively. We found that the NiTi wire behaved well in terms of plasticity and its pre-strain was not allowed to exceed 13.33% due to its ultimate strain.

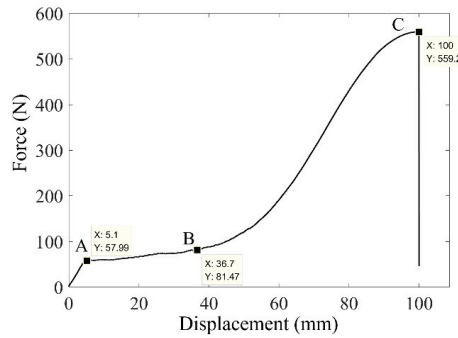


Figure 4. A load-displacement curve under tension test.

2.3. Free Recovery Rate of NiTi Wires

The free recovery rate of NiTi wires under different pre-strain was investigated using the same experimental setup as was used in the heating response rate tests. The initial length of the wire specimens was 100 mm and they were trained to have the characteristic of two-way SME. The wires were stretched with pre-strains of 2%, 3%, 4%, 5%, 6%, and 7%, and were heated to recover their deformations, then cooled to room temperature. As shown in Figure 5, L_0 , L_1 , L_2 , and L_3 represent the original length of the wire, the length of the wire after being stretched and unloaded, the length of the wire after being heated, and the length of the wire after being cooled to room temperature, respectively. The elongations of the wires were measured to calculate the two-way free recovery rate η , which is defined by:

$$\eta = \frac{L_3 - L_2}{L_1 - L_0} \tag{1}$$

Based on Equation (1), the variation in the free recovery rate with respect to the cycle number at different pre-strain levels is plotted in Figure 6. The two-way free recovery rate initially increased with increased training cycles and then tended to stabilize after 60 cycles. Additionally, the large pre-tension strain resulted in a low free recovery rate. However, the largest free recovery rate appeared when the pre-strain was 4%. Therefore, a median deformation of 4% is preferred in order to obtain the maximum recovery rate of the NiTi wire used in a two-way SME.

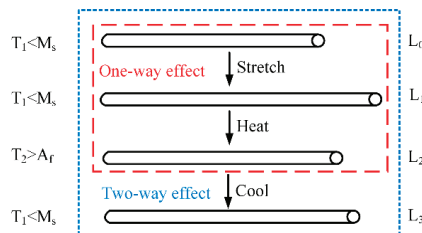


Figure 5. Two-way shape memory effects.

2.4. Restoring Force of NiTi Wires

The restoring forces of the NiTi wires, with various pre-strains and which were subjected to 3, 4, 5, and 6 A of electric current, were tested and measured. Figure 7 shows that the restoring force of each wire, at a given electric current, begins to rise sharply and then tends to stabilize. In addition, the maximum restoring force of the NiTi wire with a given pre-strain is dependent on the amplitude of the electric current applied. The restoring force rose with increased current. To compare the heating efficiency of the wires subjected

to different electric currents, the response time required to reach the maximum restoring force of each wire was measured and is listed in Table 4. The results indicate that the heating response time to reach the maximum force under a large current does not change significantly. Moreover, the wire with a 4% pre-strain presents the largest restoring force at 6 A. Thus, the designed actuator, installed with a NiTi wire with a 4% pre-strain, may obtain the largest restoring force in an acceptable response time under such conditions.

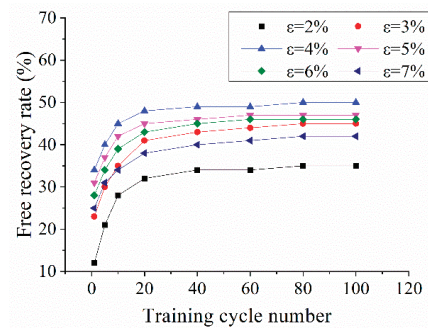


Figure 6. Free recovery rate curve of NiTi wires.

Table 4. Maximum restoring force and required time at different electric currents.

| Strain (%) | Electric Current (A) | Maximum Restoring Force (N) | Time to Reach the Maximum Restoring Force (s) |
|------------|----------------------|-----------------------------|---|
| 2 | 3 | 71.3 | 31.2 |
| | 4 | 81.8 | 16.3 |
| | 5 | 92.7 | 4.2 |
| | 6 | 95.8 | 3.1 |
| 3 | 3 | 90.6 | 31.0 |
| | 4 | 103.5 | 11.1 |
| | 5 | 116.4 | 5.6 |
| | 6 | 123.3 | 3.1 |
| 4 | 3 | 142.4 | 32.8 |
| | 4 | 156.7 | 16.5 |
| | 5 | 181.0 | 7.1 |
| | 6 | 180.1 | 5.0 |
| 5 | 3 | 123.7 | 30.7 |
| | 4 | 138.0 | 15.9 |
| | 5 | 158.4 | 4.9 |
| | 6 | 165.9 | 3.3 |
| 6 | 3 | 124.6 | 33.2 |
| | 4 | 154.7 | 13.7 |
| | 5 | 181.0 | 7.6 |
| | 6 | 181.8 | 4.4 |
| 7 | 3 | 109.6 | 29.8 |
| | 4 | 148.0 | 13.6 |
| | 5 | 159.5 | 6.9 |
| | 6 | 159.3 | 4.7 |

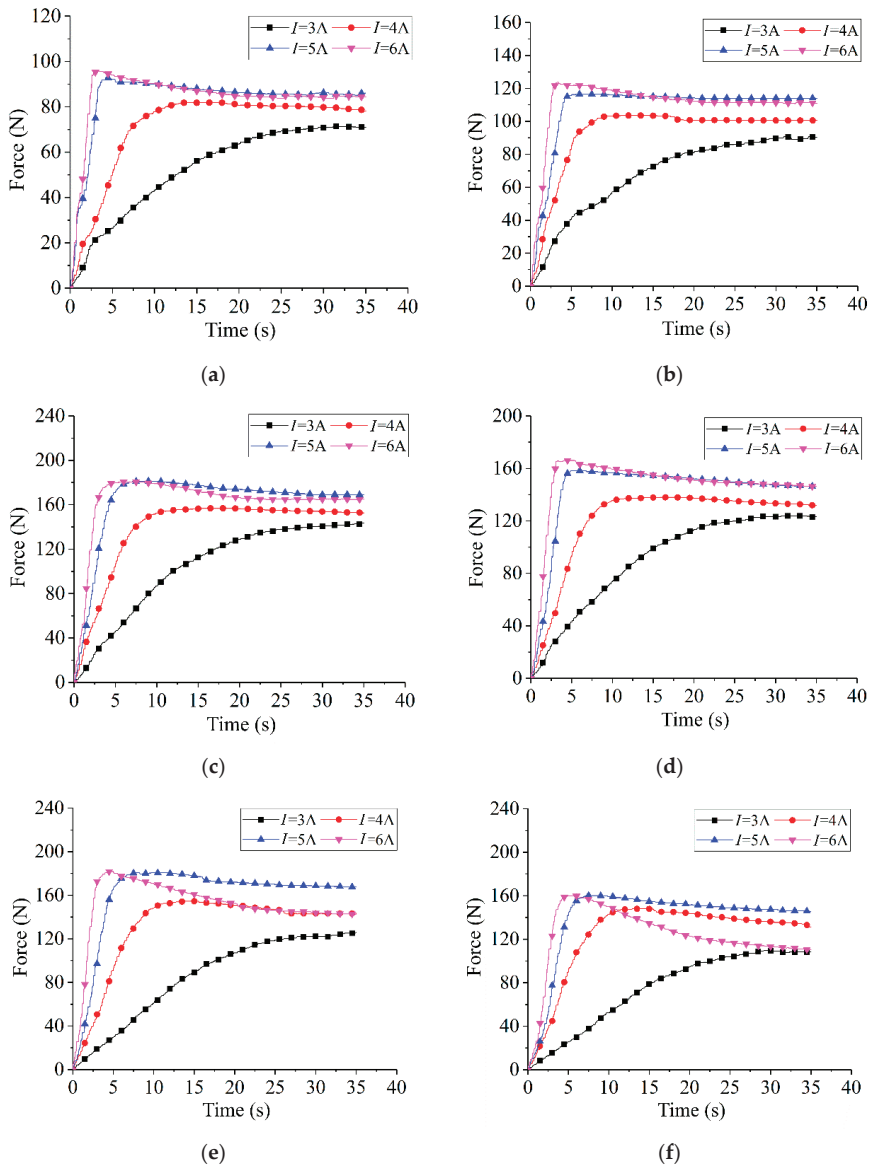


Figure 7. The restoring force of NiTi wires subjected to different electric currents: (a) $\epsilon = 2\%$; (b) $\epsilon = 3\%$; (c) $\epsilon = 4\%$; (d) $\epsilon = 5\%$; (e) $\epsilon = 6\%$; and (f) $\epsilon = 7\%$.

3. Design of the Active Clearance Control Actuator

3.1. Prototype of the Actuator

The proposed actuator was designed with the aim of mitigating or eliminating the rub-impact fault when it occurs between the rotor and casing during the operation of an aero-engine. A flowchart of the proposed fault mitigation scheme is shown in Figure 8. The design of the proposed actuator consists of a package, an electrode plate, an insulation layer, pre-stretched NiTi wires, a driving lever, a limit roller, a baffle, a spring, and external and internal casings, as is shown schematically in Figure 9.

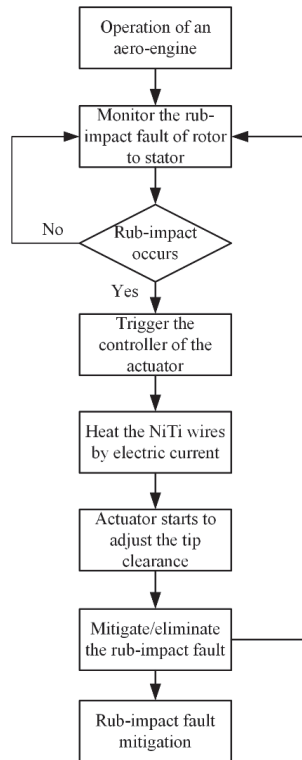


Figure 8. Flowchart of the proposed fault mitigation scheme.

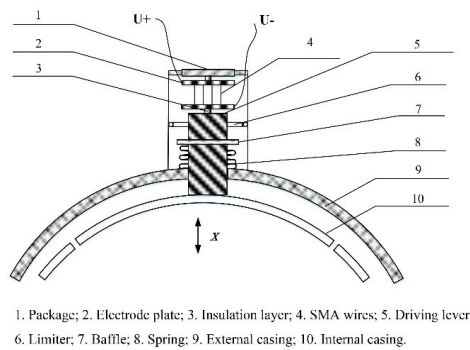


Figure 9. Design of the proposed actuator.

The actuator works by using a driving lever that passes through the external casing. The two ends of the lever are fixed on the internal casing and the bottom electrode plate, respectively, which forces the internal casing to deform when the lever moves. Two insulation layers are placed between the electrode plates and the other parts of the actuator for the insulation and protection of the entire structure. The two ends of NiTi wires are bolted at each end to two electrode plates and are heated by an electric current during the operation of the actuator. A baffle is welded onto the driving lever and a spring is installed between the baffle and the external casing to generate a restoring force. In addition, one limiter with three rollers, circumferentially threaded and installed on the driving lever, act

to improve the accuracy of the motion. The prototype of the actuator is shown in Figure 10. To allow us to observe the behavior of the actuator, the package was not completely sealed, as shown in Figure 10d.

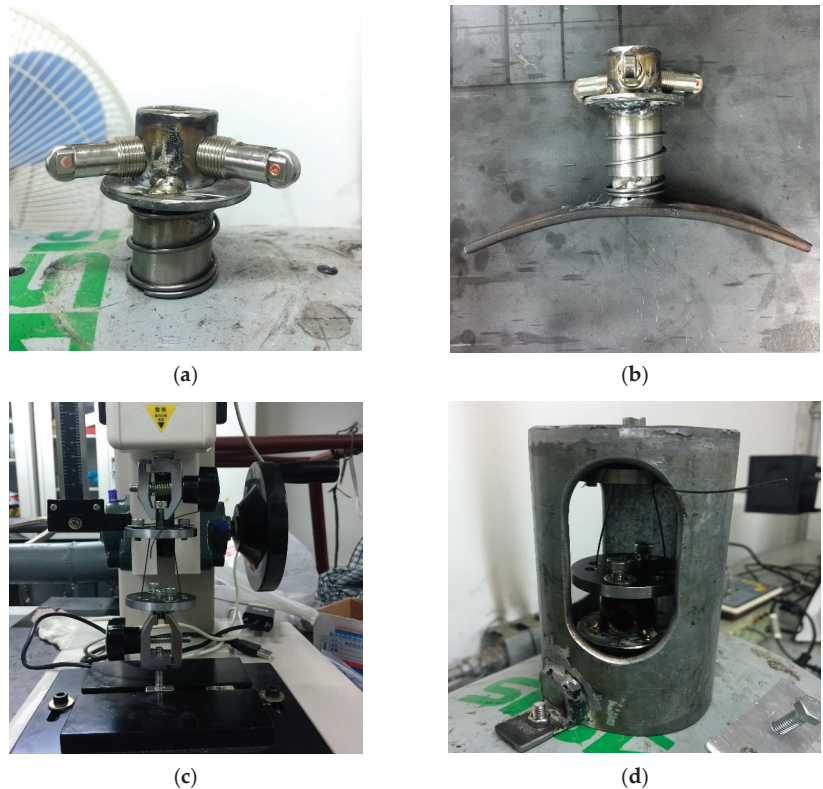


Figure 10. Prototype of the SMA-based actuator: (a) Limiter; (b) driving lever and internal casing; (c) electrode plates; and (d) package.

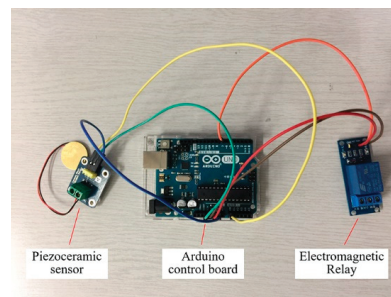
Once a rub-impact fault is detected, the external direct current (DC) power supply system begins to heat the SMA wires. When the temperature is beyond the austenite start temperature of the wires A_s , the wire undergoes a martensite phase transformation and is compressed due to the unique property of SME. This compression forces the internal casing to move upwards via the driving lever. Due to the motion of the internal casing, the tip clearance between the rotor and the casing is enlarged and, as a result, the rub-impact phenomenon is mitigated or possibly eliminated. The aim of using two-way NiTi wires is to mitigate the rubbing fault, whilst guaranteeing engine efficiency, which is achieved when the clearance is reduced once the NiTi wires cool to the martensite phase. To verify the feasibility of the proposed model, a prototype of an SMA-based actuator was designed and its geometric parameters are shown in Table 5.

Table 5. Design of the proposed actuator prototype.

| Parts | Shape | Dimensions |
|-----------------|--|--|
| Package | Hollow cylinder | Internal diameter: 70 mm External diameter: 76 mm Height: 150 mm |
| Electrode plate | Disc | Diameter: 40 mm Thickness: 3 mm |
| NiTi wire | Wire | Length: 50 mm Diameter: 0.8 mm Pre-strain: 4% |
| Driving lever | Rod | Diameter: 26 mm Height: 70 mm |
| Spring | Compression spring | Internal diameter: 28 mm Free height: 25 mm Wire diameter: 1.5 mm |
| Limiter | Hollow cylinder with three rollers on the side of the cylinder | Internal diameter of cylinder: 70 mm External diameter of cylinder: 76 mm Height of cylinder: 150 mm External diameter of roller: 6 mm Thickness of roller: 2 mm |

3.2. Control Scheme of the Actuator

In order to realize the self-healing of rub-impact faults, a control system based on an Arduino control board was investigated for the proposed actuator. This system consists of an Arduino Uno R3 control board, an RB-02S082A piezoelectric sensor, and an SRD-05VDC-SL-C 5V electromagnetic relay, as shown in Figure 11. When the system is running, the AE signals are acquired by the piezoelectric sensor and are then transferred to the Arduino Uno R3 control board, which is capable of identifying whether the acquired signals are fault signals. The electric circuit switching function is implemented by compiled coding. In normal conditions, the current circuit is shut down and the actuator does not work. Once a rub-impact fault is identified by the board, an order is sent to the electromagnetic relay to switch on the current circuit to heat the NiTi wires. A flowchart depiction of the algorithm of the control scheme is shown in Figure 12.

**Figure 11.** Configuration of the control system.

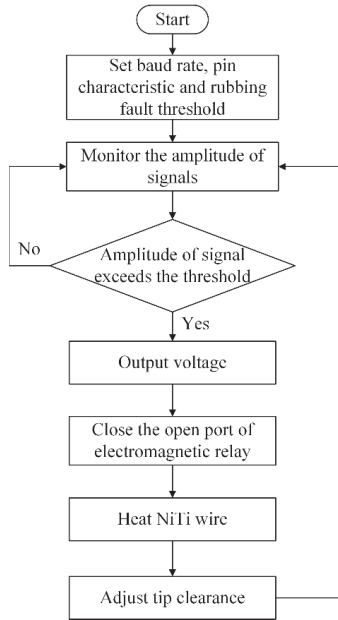


Figure 12. Flowchart of the control scheme.

4. Experimental Verification of the Clearance Control Mechanism

In order to evaluate the effectiveness of the clearance control actuator, the mechanism was verified by an experimental study. Figure 13 shows the experimental setup, which consisted of a rotor test rig, a rotor power supply, an accelerator, an electromagnetic relay, an Arduino control board, a piezoelectric sensor, an AE sensor, a preamplifier, a signal acquisition instrument, a DC power supply, and a PC and data analysis system.

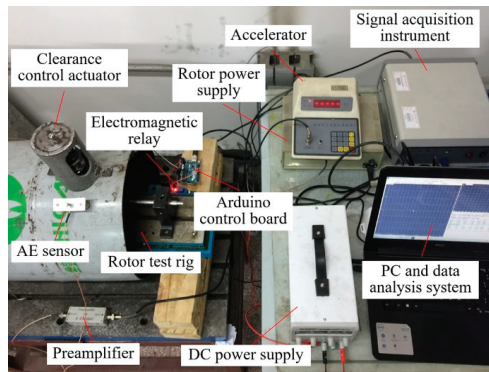


Figure 13. Equipment of the clearance control mechanism.

The NiTi wires with 4% pre-strain were installed on the clearance control actuator, and the actuator was fixed on the external casing of the aero-engine. The power of the rotor test rig was provided by a rotor power supply and the rotational speed of the rotor was controlled by an accelerator. The piezoelectric sensor was attached to the upper surface of the internal casing and was connected to the Arduino control board to identify the rubbing and control the electromagnetic relay operation. An AE sensor was attached to the surface

of the external casing to acquire the AE signals. These AE signals were processed by a signal preamplifier and then transmitted to the PC.

4.1. Setting of the Threshold and Sampling Frequency

The accuracy of the proposed control scheme is dependent on the selection of the signaling threshold and sampling frequency. The operation of an aero-engine is always accompanied by vibrations, whether or not rub-impact occurs. Therefore, the primary goal is to ascertain the appropriate threshold at which rub-impact faults occur, by reference to the amplitude of the vibration signal. Figure 14 shows the signals acquired by the piezoelectric sensor at rotational speeds of 1500, 2000, 2500, and 3000 rpm. The amplitudes of all vibration signals without the rub-impact fault are lower than 20. Therefore, the threshold may be set to 20 as the threshold for the occurrence of the rub-impact fault. Regarding the sampling, in our experience, it is generally better for the set to be greater than 5 times the highest frequency of the vibration signals, to ensure the reliability of the acquired data. Because the rotational frequency of the rotor at 3000 rpm was measured at 50 Hz in our tests, the sampling frequency in this study was set to 500 Hz, which was 10 times the highest frequency.

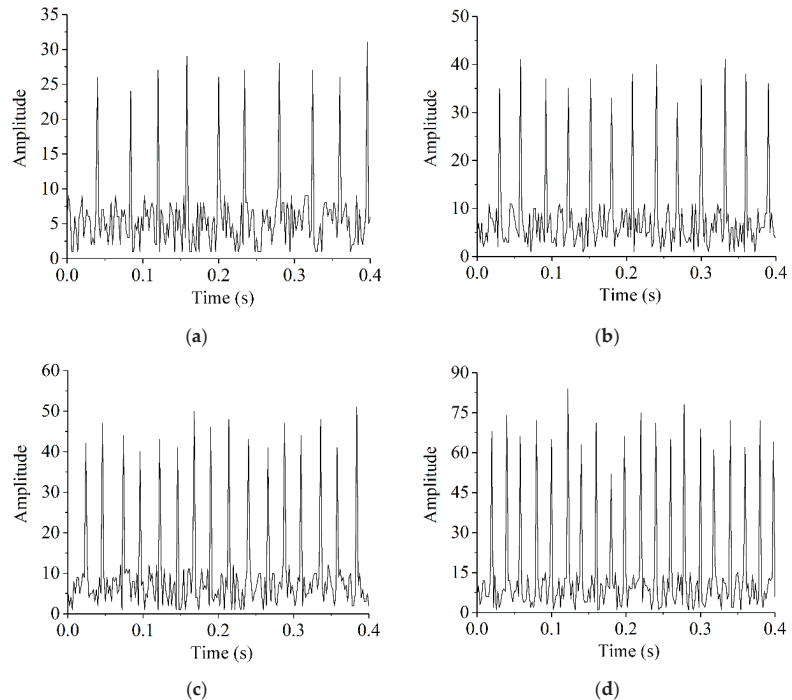


Figure 14. Vibration signal at various rotational speeds: (a) $n = 1500$ rpm; (b) $n = 2000$ rpm; (c) $n = 2500$ rpm; and (d) $n = 3000$ rpm.

4.2. Verification of the Clearance Control Actuator

(1) Effect of rotational speed

The effectiveness of the proposed clearance control actuator was verified at rotational speeds of 1500, 2000, 2500, and 3000 rpm, with a 6 A electric current, in our tests. Assuming that $t_0, t_1, t_2, t_3, t_4, t_5$ represent the points at which a rub-impact fault occurs, the amplitude of the AE signal reaches its maximum, the fault is eliminated, the second rub-impact fault

begins, the amplitude of the AE signal reaches its maximum again, and the second fault is eliminated, consecutively. The specific durations are defined by:

$$T_1 = t_1 - t_0 \tag{2}$$

$$T_2 = t_2 - t_1 \tag{3}$$

$$T_3 = t_3 - t_2 \tag{4}$$

$$T_4 = t_4 - t_3 \tag{5}$$

$$T_5 = t_5 - t_4 \tag{6}$$

where T_1 is the duration from the start of rub-impact to its most serious state, T_2 denotes the rub-impact fault elimination time spent by the proposed actuator, T_3 represents the time interval between the rub-impact fault and the start of the following rub-impact fault event, T_4 is the duration from the start of the next fault to its most serious state, and T_5 denotes the time spent to eliminate the fault.

Table 6 gives the spending times measured at various rotational speeds in the rotation test, at a 6 A electric current. The results show that the control times $T_1, T_2, T_3, T_4,$ and T_5 are almost the same at different rotational speeds. This means the control time of the clearance control mechanism based on two-way NiTi wires is independent of the rotational speed of the rotor.

Table 6. Control time of the clearance control actuator at different rotational speeds.

| <i>I</i> (A) | <i>n</i> (rpm) | T_1 (s) | T_2 (s) | T_3 (s) | T_4 (s) | T_5 (s) |
|--------------|----------------|-----------|-----------|-----------|-----------|-----------|
| 6 | 1500 | 2.1 | 1.0 | 14.3 | 1.3 | 1.0 |
| 6 | 2000 | 1.8 | 0.9 | 13.5 | 1.2 | 0.8 |
| 6 | 2500 | 1.8 | 1.1 | 14.0 | 1.4 | 0.7 |
| 6 | 3000 | 2.1 | 1.0 | 13.7 | 1.3 | 1.1 |

A series of rotational rubbing tests were conducted under the condition of a 6 A current and a 1500 rpm rotational speed, and the AE signal during the operational process was measured. As shown in Figure 15, the amplitude of the AE signal continuously rises with increased rotor speed during the start-up process until it reaches the maximum at $t_1 = 2.143$ s, i.e., the rubbing becomes serious and a fault at this moment. At $t_2 = 3.177$ s, this rubbing fault is eliminated. The next rubbing fault occurs at $t_3 = 17.49$ s, becoming severe at $t_4 = 18.77$ s. At $t_5 = 19.78$ s, the fault is eliminated.

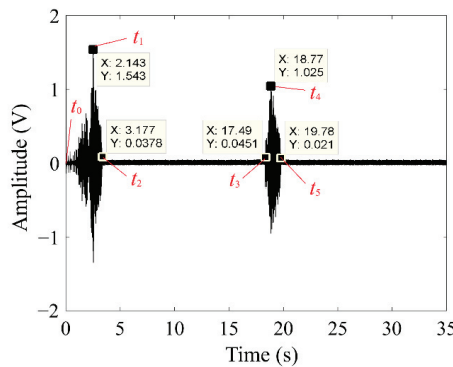


Figure 15. AE signals acquired during the clearance control process based on two-way NiTi SMA wires.

(2) Effect of the electric current

A series of rotational rubbing tests at 4, 5, and 6 A electric currents were conducted at rotational speeds of 1500 rpm. The rubbing fault eliminating time spent during the control process is recorded in Table 7. It was found that T_2 , T_3 , and T_5 changed significantly with different electric currents. The durations of T_2 and T_5 gradually descended with increased current. Therefore, a large electric current contributes to a rapid control response time. In addition, the interval between one rub-impact fault and the next fault, T_3 , becomes larger at a higher current compared to lower currents, resulting from the fact that the NiTi wires need more time to cool. The proposed actuator is affected significantly by the electric current.

Table 7. Control time of the NiTi wire based actuator at various electric currents.

| I (A) | n (rpm) | T_1 (s) | T_2 (s) | T_3 (s) | T_4 (s) | T_5 (s) |
|---------|-----------|-----------|-----------|-----------|-----------|-----------|
| 4 | 1500 | 2.3 | 13.6 | 7.2 | 1.1 | 13.1 |
| 5 | 1500 | 2.0 | 4.7 | 10.4 | 1.1 | 4.4 |
| 6 | 1500 | 2.1 | 1.0 | 14.3 | 1.3 | 1.0 |

5. Conclusions

A fault mitigation scheme based on a two-way SMA wire system to mitigate and possibly eliminate the rotor-stator rub-impact of an aero-engine, by controlling the tip clearance between the rotor and stator, was proposed in this study. Based on the inherent characteristics of SMA, a prototype of an SMA-based actuator was designed, manufactured, and tested. Through our experimental study, the proposed scheme was verified and the following conclusions were drawn:

- (1) The mechanical properties of NiTi wire are highly dependent upon the heating electric current and on its pre-strain level. It is recommended to apply a 6 A electric current to heat the wires and to use wires with a 4% pre-strain in the design of an SMA wire-based actuator.
- (2) The time spent in the fault mitigation/elimination process is not significantly dependent on the rotational speed of the rotor.
- (3) The proposed SMA-based actuator is a promising model that mitigates the rub-impact fault that occurs between the rotor and stator in aero-engines, whilst guaranteeing engine efficiency.

Author Contributions: Conceptualization, Q.P. and T.H.; methodology, Q.P. and X.L. and T.H.; software, X.L. and H.C.; validation, W.L. and H.C.; formal analysis, X.L.; investigation, Q.P. and W.L.; resources, T.H.; writing—original draft preparation, W.L. and Q.P.; writing—review and editing, Q.P. and T.H.; visualization, W.L.; supervision, T.H.; project administration, X.L. and H.C.; funding acquisition, X.L. 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 (2017-I-0007-0008).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|-----|--------------------------|
| SMA | Shape memory alloy |
| AE | Acoustic emission |
| ACC | Active clearance control |
| SME | Shape memory effect |
| DC | Direct current |

References

1. Ahmad, S. Rotor casing contact phenomenon in rotor dynamics—Literature survey. *J. Vib. Control* **2010**, *16*, 1369–1377. [[CrossRef](#)]
2. Wiseman, M.W.; Guo, T.H. An investigation of life extending control techniques for gas turbine engines. In Proceedings of the American Control Conference, Arlington, VA, USA, 25–27 June 2001.
3. Lattime, S.B.; Steinetz, B.M. High-pressure-turbine clearance control systems: Current practices and future directions. *J. Propuls. Power* **2004**, *20*, 302–311. [[CrossRef](#)]
4. Sato, W.; Ota, S.; Yamagata, A. A study on unsteady aerodynamic excitation forces on radial turbine blade due to rotor-stator interaction. *Turbomachinery* **2012**, *40*, 605–611.
5. Jacquet-Richardet, G.; Torkhani, M.; Cartraud, P.; Thouverez, F.; Nouri Baranger, T.; Herran, M.; Gibert, C.; Baguet, S.; Almeida, P.; Peletan, L. Rotor to stator contacts in turbomachines. Review and application. *Mech. Syst. Signal Process.* **2013**, *40*, 401–420. [[CrossRef](#)]
6. Nitschke, S.; Wollmann, T.; Ebert, C.; Behnisch, T.; Langkamp, A.; Lang, T.; Johann, E.; Gude, M. An advanced experimental method and test rig concept for investigating the dynamic blade-tip/casing interactions under engine-like mechanical conditions. *Wear* **2019**, *422*, 161–166. [[CrossRef](#)]
7. Hong, J.; Yu, P.C.; Zhang, D.Y.; Ma, Y.H. Nonlinear dynamic analysis using the complex nonlinear modes for a rotor system with an additional constraint due to rub-impact. *Mech. Syst. Signal Process.* **2019**, *116*, 443–461. [[CrossRef](#)]
8. Ma, H.; Yin, F.; Wu, Z.; Tai, X.; Wen, B. Non-linear vibration response analysis of a rotor-blade system with blade-tip rubbing. *Nonlinear Dyn.* **2016**, *84*, 1225–1258. [[CrossRef](#)]
9. Zeng, J.; Zhao, C.G.; Ma, H.; Yu, K.; Wen, B.C. Rubbing dynamic characteristics of the blisk-casing system with elastic supports. *Aerosp. Sci. Technol.* **2019**, *95*, 105481. [[CrossRef](#)]
10. Hong, J.; Yu, P.C.; Zhang, D.Y.; Liang, Z.C. Modal characteristics analysis for a flexible rotor with non-smooth constraint due to intermittent rub-impact. *Chin. J. Aeronaut.* **2018**, *31*, 498–513. [[CrossRef](#)]
11. Braut, S.; Zigulic, R.; Skoblar, A.; Roncovic, G.S. Partial rub detection based on instantaneous angular speed measurement and variational mode decomposition. *J. Vib. Eng. Technol.* **2020**, *8*, 351–364. [[CrossRef](#)]
12. He, T.; Ye, W.; Pan, Q.; Liu, X.D. An automatic abrupt information extraction method based on singular value decomposition and higher-order statistics. *Meas. Sci. Technol.* **2016**, *27*, 025007. [[CrossRef](#)]
13. Prosvirin, A.E.; Islam, M.; Kim, J.Y.; Kim, J.M. Rub-impact fault diagnosis using an effective IMF selection technique in ensemble empirical mode decomposition and hybrid feature models. *Sensors* **2018**, *18*, 2040. [[CrossRef](#)]
14. Yu, M.Y.; Feng, Z.G.; Huang, J.J.; Zhu, L.L. Aero-engine rotor-stator rubbing position identification based on casing velocity signal. *J. Vibroeng.* **2016**, *18*, 2123–2134. [[CrossRef](#)]
15. Pan, Q.; Zhou, R.; Su, J.Y.; He, T.; Zhang, Z.B. Automatic localization of the rotor-stator rubbing fault based on acoustic emission method and higher-order statistics. *J. Mech. Sci. Technol.* **2019**, *33*, 513–524. [[CrossRef](#)]
16. Wu, X.Y.; Peng, Z.K.; Ren, J.S.; Cheng, C.M.; Zhang, W.M.; Wang, D. Rub-impact fault diagnosis of rotating machinery based on 1-d convolutional neural networks. *IEEE Sens. J.* **2020**, *20*, 8349–8363. [[CrossRef](#)]
17. Yang, W.; Li, J.; Peng, W.; Deng, A. A rub-impact recognition method based on improved convolutional neural network. *CMC-Comput. Mat. Contin.* **2020**, *63*, 283–299. [[CrossRef](#)]
18. Zhang, H.; Cai, Y. Research on active clearance control system of HPT. In Proceedings of the 4th International Symposium on Jet Propulsion and Power Engineering, Xi’an, China, 10–12 September 2012.
19. Bucaro, M.T.; Ruiz, R.J.; Albers, R.J.; Estridge, S.A.; Wartner, R.F. Thermal Control of Gas Turbine Engine Rings for Active Clearance Control. Patent US7597537, 6 October 2009.
20. Tillman, D.; Gage, R.; Barrett, R. Gas Turbine Engine Turbine Blade Tip Active Clearance Control System and Method. Patent US9266618, 23 February 2016.
21. DeCastro, J.; Melcher, K. A study on the requirements for fast active turbine tip clearance control systems. In Proceedings of the 40th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, Fort Lauderdale, FL, USA, 11–14 July 2004.
22. Yu, P.C.; Zhang, D.Y.; Ma, Y.H.; Hong, J. Dynamic modeling and vibration characteristics analysis of the aero-engine dual-rotor system with Fan blade out. *Mech. Syst. Signal Process.* **2018**, *106*, 158–175. [[CrossRef](#)]
23. Ma, L.; Wang, J.; Zhang, G.C. The dynamic characteristic of a faulted rotor system with multi-objective optimization designed SFD. *Appl. Sci.* **2019**, *9*, 3628. [[CrossRef](#)]
24. Xu, J.; Li, Q.; Wang, S. Impulsive control of the rotor-stator rub based on phase characteristic. *Abstr. Appl. Anal.* **2014**, *2014*, 495747. [[CrossRef](#)]
25. Shang, Z.; Zhang, H.; Shaymurat, T. Control methods using cross-coupling effects for suppression of rotor/stator rubbing system. In Proceedings of the 7th International Conference on Mechanical and Electronics Engineering, Dalian, China, 26–27 September 2015.
26. Gao, T.; Cao, S.Q.; Hou, L.L.; Hou, Y.H. An experimental study on the nonlinear vibration phenomenon of a rotor system subjected to barrel roll flight and coupled rub-impact faults. *Measurement* **2020**, *153*, 107406. [[CrossRef](#)]
27. Garg, S. *NASA Glenn Research in Controls and Diagnostics for Intelligent Aerospace Propulsion Systems*; NASA/TM Report, 2005-214036; NASA: Washington, DC, USA, 2007.
28. DeCastro, J.A.; Melcher, K.J.; Noebe, R.D. System-level design of a shape memory alloy actuator for active clearance control in the high pressure turbine. In Proceedings of the 41st Joint Propulsion Conference and Exhibit, Tucson, AZ, USA, 10–13 July 2005.

29. Humbeeck, J.V. Shape memory alloys: A material and a technology. *Adv. Eng. Mater.* **2001**, *3*, 837–850. [[CrossRef](#)]
30. Jani, J.M.; Leary, M.; Subic, A.; Gibson, M.A. A review of shape memory alloy research, applications and opportunities. *Mater. Design.* **2014**, *56*, 1078–1113. [[CrossRef](#)]
31. Duerig, T.; Pelton, A.; Stockel, D. An overview of nitinol medical applications. *Mater. Sci. Eng. A* **1999**, *273*, 149–160. [[CrossRef](#)]
32. Van Humbeeck, J. Non-medical applications of shape memory alloys. *Mater. Sci. Eng. A* **1999**, *273*, 134–148. [[CrossRef](#)]

Article

Self-Powered Self-Contained Wireless Vibration Synchronous Sensor for Fault Detection

Ghufran Aldawood¹ and Hamzeh Bardaweel^{1,2,3,*}

¹ Institute for Micromanufacturing, College of Engineering and Science, Louisiana Tech University, Ruston, LA 71272, USA; gja005@latech.edu

² Department of Mechanical Engineering, College of Engineering and Science, Louisiana Tech University, Ruston, LA 71272, USA

³ Department of Nanosystems Engineering, College of Engineering and Science, Louisiana Tech University, Ruston, LA 71272, USA

* Correspondence: hamzehb@latech.edu; Tel.: +1-318-257-5134

Abstract: Failure in dynamic structures poses a pressing need for fault detection systems. Inter-connected sensor nodes of wireless sensor networks (WSN) offer a solution by communicating information about their surroundings. Nonetheless, these battery-powered sensors have an immense labor cost and require periodical battery maintenance and replacement. Batteries pose a significant environmental threat that is expected to cause irreversible damage to the ecosystem. We introduce a fully integrated vibration-powered energy harvester sensor system that is interfaced with a custom-developed fault detection app. Vibrations are used to power a radio frequency (RF) transmitter that is integrated with the vibration sensor subunit. The harvester-sensor unit is comprised of dual moving magnets that are bordered by coil windings for power and signal generation. The power generated from the harvester is used to operate the transmitter while the signal generated from the sensor is transmitted as a vibration signal. Transmitted values are streamed into a high precision fault detection app capable of detecting the frequency of vibrations with an error of 1%. The app employs an FFT algorithm on the transmitted data and notifies the user when a threshold vibration level is reached. The total energy consumed by the transmitter is 0.894 μJ at a 3 V operation. The operable acceleration of the system is 0.7 g [m/s^2] at 5–10.6 Hz.

Keywords: vibration energy harvesting; vibration sensor; self-powered sensor; clean technology; wireless vibration sensor; IoT support technology

Citation: Aldawood, G.; Bardaweel, H. Self-Powered Self-Contained Wireless Vibration Synchronous Sensor for Fault Detection. *Sensors* **2022**, *22*, 2352. <https://doi.org/10.3390/s22062352>

Academic Editors: Hamed Kalhori, Yongbo Li, Bing Li and Jinchun Ji

Received: 17 February 2022

Accepted: 15 March 2022

Published: 18 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Internet of Things (IoT) technologies are blooming and are expected to reach a 1567 billion USD market value by 2025 [1]. Currently, there are a little over 7 billion sensor nodes worldwide [1]. These sensors represent the backbone of IoT systems since they are responsible for detecting essential information about the surrounding environment and monitoring the health conditions of structures such as compressors, pumps, bridges, tunnels, railroads, and other dynamic structures [2]. Monitoring the health conditions of structures using these sensors helps in preventing catastrophic failures and loss of lives [2].

Presently, the majority of these sensors are powered using traditional batteries [3]. The use of these traditional batteries limits the scope and value of these IoT sensors [4,5]. This is due to the fact that the use of conventional batteries as a power source for IoT sensors results in several challenges. Not only do these batteries require continuous and frequent replacement and maintenance [6], but also they have a limited lifespan and pose an environmental threat [7]. Given the projected worldwide spread of these IoT sensors, this environmental concern has become a pressing issue to deal with [8]. Moreover, in harsh environments and remote locations, including wildlife, gas and oil fields, and drilling

in mining fields, a self-powered solution becomes the only viable option for deploying long-lasting, stand-alone, eco-friendly, and sustainable sensors [9].

Consequently, there has been a growing interest in developing autonomous, self-powered, environment-friendly sensor technologies as a necessary and integral part of the flourishing IoT systems and technologies [10]. Tremendous recent studies have investigated the issue of developing autonomous sensors powered by free sources of energies surrounding these sensors including solar and vibration energies. Vibration-powered sensors are an attractive option since they are not constrained by the availability of sunlight and can operate indoors and underground [3]. Moreover, vibration energy consists of a broadband vibration spectrum rich in low frequencies with a power density as high as $500 \mu\text{W}/\text{cm}^3$ [11]. Vibration energy also represents a wealthy form of freely-available energy in transportation [12] and industry sectors [13].

To this end, Xin Li et al. developed a vibration-powered sensor node [3]. The system used a piezoelectric transducer to harvest ambient vibrations. The fully integrated self-powered sensing and transmitting system consisted of a few units including energy generation, energy transduction, energy-boosting, energy management and circuitry, and demonstration unit (mobile interface). The system was successfully demonstrated under various vibrations conditions. In another effort, two wireless sensor nodes were powered using an electromagnetic vibration energy harvesting system [1]. A custom-built power conditioning system was integrated into the energy harvesting system and then used to power a sensor node with a duty cycle of 30 s. The self-powered system was shown to produce enough power to receive and transmit information at intervals of less than 60 s. Moreover, in Ref [14] a piezoelectric energy harvester was used to power a wireless platform which consisted of a vibration sensor, a microcontroller, a power management circuitry, and a custom-built low power radio transmitter. The fully integrated system was operated at acceleration level and frequency of $0.25 \text{ g [m/s}^2\text{]}$ and 100 Hz, respectively. The system was able to transmit the sensor data every 10 s for a duty cycle of 0.2%. Moreover, in their study, Lu Wang et al. [15] built a wireless temperature sensor node powered by a piezoelectric bimorph cantilever vibration energy harvesting system. A big proof mass was attached to the harvester to lower its resonant frequency to approximately 22 Hz. A power management circuitry was built for rectifying the output from the harvester which was then used to power the temperature sensor. In Ref [16], the authors utilized a commercial piezoelectric cantilever as an energy harvester for powering a wireless temperature sensor node. The study focused on investigating the design methodology for the power management circuitry used in their work. A demonstration of the self-powered temperature sensor was performed. Additionally, Lu Wang et al. [17] built a hybrid piezoelectric-triboelectric unit and used it to construct an autonomous wireless sensor node where the piezoelectric generator served as the energy source and the triboelectric worked as an accelerometer (i.e., sensing unit). The piezoelectric harvester produced approximately 6.5 [mW] at $1 \text{ g [m/s}^2\text{]}$ and 25 [Hz] and the triboelectric accelerometer showed a sensitivity of 15 V/g for acceleration range 0–1.5 g [m/s²]. In another effort, a piezoelectric energy harvesting-sensor unit was developed and implemented in monitoring airflow from an HVAC outlet [18]. To avoid signal distortion from the sensor the proposed system used two separate piezoelectric devices (i.e., one for energy harvesting and one for sensing purposes).

One of the major constraints in wireless sensor technologies is associated with the limitations in the supplied energy to the wireless sensor nodes [19–21]. Overcoming this challenge can only be achieved through minimizing energy consumption by means of ultra-low-power techniques. Several techniques to reduce energy consumption have emerged in recent years, such as duty cycling [22]. Other techniques include topology control [23], which deals with the distribution of the wireless sensor nodes in order to reduce energy consumption while eliminating interference at the lowest cost possible. Additionally, data transmission network protocol selection is based on application, where different protocols have variable bandwidths with varying energy consumptions [24]. Cyber security of those

networks is also amongst the necessary features that when added would require more processing power from the sensor node [25].

The work presented in this article is focused on developing a novel, self-powered, self-contained, environment-friendly, and wireless vibration sensor. One of the major issues that the vibration-powered wireless sensor node designers are faced with is the unstable power source causing much noise to the transmitted data [14]. This causes a need for stringent supply regulation using linear or switching regulators that amounts to power losses due to heat as a result of the regulation. In this work, power is conserved from not using any regulators, and instead, the noise corrupted data is retrieved from post-processing through a custom-developed dynamic fast Fourier transform (FFT) app. A given dynamic structure has a vibration signature where a frequency shift can be detected by implementing an FFT algorithm to the time series data [26]. Furthermore, there is an ample amount of noise introduced into the signal due to interference from the use of radiofrequency modules [14] and the use of FFT in post-processing of the data lessens the impact of noise on the signal [27]. The custom developed app in this work can fetch the sensor bit-stream, buffer the data, and use it to plot the vibration signal amplitude and its frequency in real-time.

Moreover, unlike in the relatively high voltages produced in piezoelectric transducers that would set a need for voltage regulation [3], in this work, electromagnetic transduction is used and the output voltages are within the electronics acceptable supplied voltage range. Also, unlike the aforementioned studies and state-of-the-art developments, in this work, the self-powered sensor uses vibrations to synchronously perform two functions. First, these free vibrations are converted into useful electric power through the presented energy harvester system. Second, these vibrations are detected as electric signals (voltage) by the presented vibration sensor, and are then transmitted wirelessly to the workstation (laptop). Thus, the harvester-sensor hardware is self-contained and self-powered. That is, the mechanical power required to operate the sensor is obtained from the energy harvester. A charge pump circuit, also known as a voltage multiplier, is used to rectify the AC output of the energy harvester. The output DC is then stored in a supercapacitor that provides the energy to the microcontroller and sensor transmitter circuit. The voltage multiplier circuit allows immediate circuit startup due to the transmitter circuit having sufficient voltage and electric current to operate. The signal from the sensor is sampled using a 10-bit analog to digital converter (ADC) and is transmitted over an RF amplitude modulated (AM) carrier. When monitoring the health conditions of a structure, a shift in its vibrations signature may indicate a malfunction in the structure which could lead to impending failure [2,28]. In this article, collected sensor data are analyzed through a custom-developed dynamic displacement monitoring software which helps in mitigating damage to vibrating structures.

The structure and organization of the article are outlined next. The design and structure of the self-powered self-contained wireless sensor are presented in Section 2. The manufacturing and fabrication of the system components are detailed in Section 3. Experimental methods and testing techniques are detailed in Section 4. Results and findings from this work and system operation and demonstration are discussed in Section 5. Finally, Section 6 presents the major conclusions and summarizes the results from this work.

2. Design Concept and System Configuration

The concept and overall structure of the proposed self-powered self-contained fully integrated system are shown in Figure 1. The overall structure of the self-powered sensor consists of a few main sub-systems, namely a vibration energy harvester-sensor unit, a transmitter circuit, a receiver circuit, and a custom-developed fault detection app. The main elements and components of these sub-systems are shown in Figure 2.

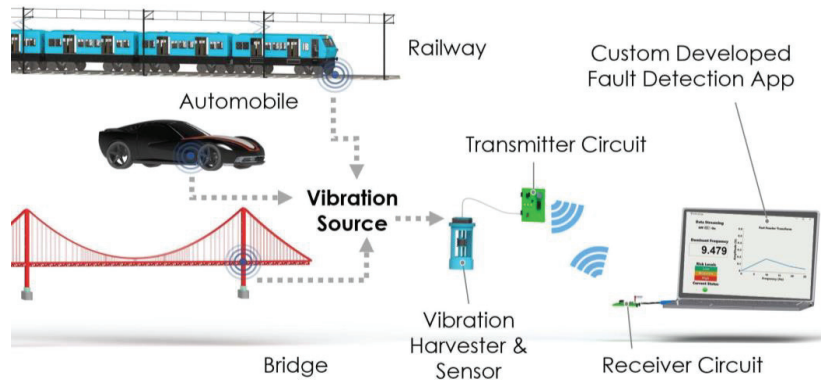


Figure 1. Design and concept of the wireless self-powered, self-contained, and eco-friendly vibration sensor system and its main components.

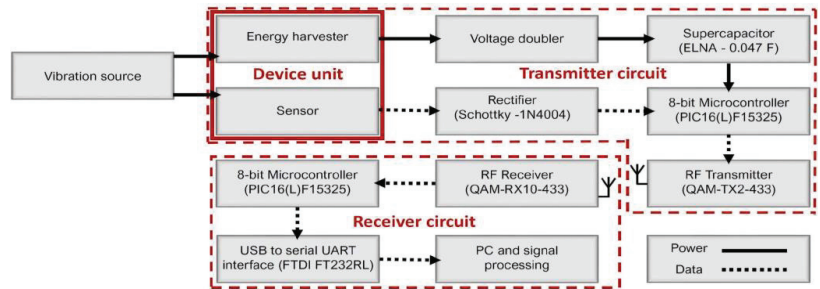


Figure 2. Block diagram showing the layout of the proposed wireless self-powered, self-contained, eco-friendly vibration sensor system.

2.1. Vibration Energy Harvester-Sensor Subsystem

The vibration harvester-sensor unit consists of two major components: a vibration energy harvester and a vibration sensor (as shown in Figure 3). The harvester consists of two (top and bottom) magnets with a third magnet that is levitated between them. The magnets are arranged in a repulsive configuration with alike poles facing each other and, therefore, the levitated magnet is floating between the top and bottom magnets. The bottom magnet is fixed while the top magnet is glued to, and guided by, a mechanical FR4 sensor diaphragm. A 40 AWG stationary copper coil is wound around the levitated magnet for electric power generation. The vibration sensor consists of the mechanical diaphragm and its guided top magnet, and copper coil windings are positioned around the top magnet as shown in Figure 3. The 3D printed guiding rail of the levitated magnet is designed to provide a restricted travel pathway for the levitated magnet between the top guided magnet and the bottom fixed magnet. The energy harvester coil windings are wound around fixed cylindrical support that is positioned in alignment with the center of the levitated magnet. The fixed magnet support is used to hold the bottom magnet to the guiding rail frame. The 3D printed holder casing is designed to hold the guiding rail of the levitated magnet to the rest of the components as shown in Figure 3.

When subject to external vibrations, first, the levitated magnet moves inside the harvester’s coil windings, thus converting the kinetic energy from these oscillations into electric power that is used to operate the system shown in Figure 1. Consequently, dynamic displacement is induced in the sensor diaphragm and top magnet as a result of these excitations. In turn, induced vibrations in the sensor’s diaphragm and top magnet result in induced voltage in the top coil surrounding the top guided magnet. The voltage signal

from the top coil is then sampled by the microcontroller in the transmitter circuit. This is discussed in further detail next.

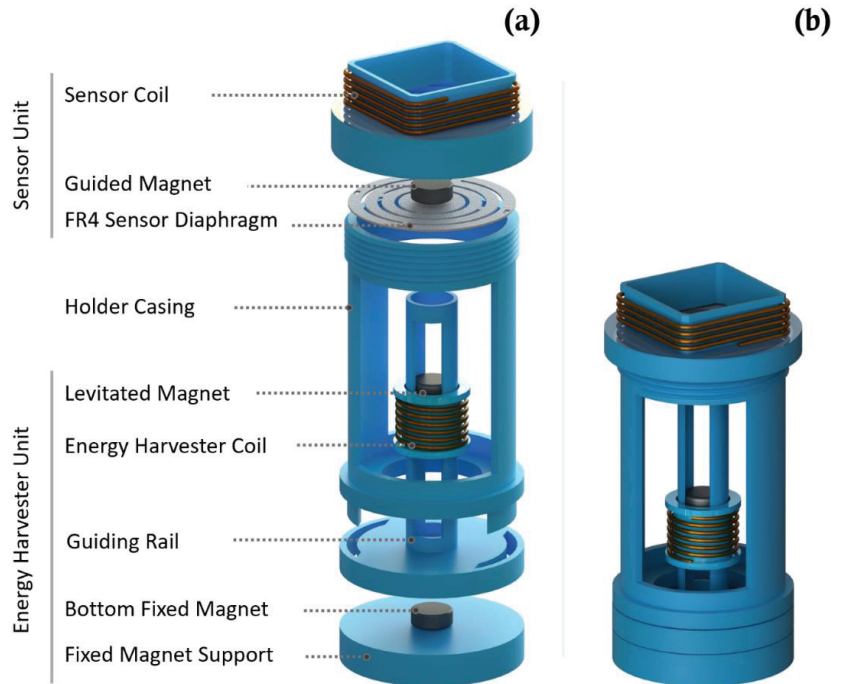


Figure 3. Design and 3D view of the structure of the vibration energy harvester-sensor unit presented in this work: (a) exploded view and (b) collapsed view.

2.2. Transmitter Subsystem

The main components of the transmitter sub-system are shown in Figure 2. The circuitry encapsulates a Microchip Technology PIC microcontroller that is enabled when the energy harvester has sufficient energy to power the circuit load. The input AC voltage from the energy harvester is rectified by a two-stage voltage multiplier, known as ‘voltage doubler’. This is shown in Figure 4. In the voltage doubler circuit, the diode in the first stage of the doubler is forward biased during the negative half cycle of the input sinusoidal waveform. This allows charging up of both capacitors. Meanwhile, during the positive half cycle of the input, the diode in the first stage of the multiplier is reverse biased and is blocking the discharging of the capacitor in the first stage. This allows for the capacitor on the second stage to charge up to approximately twice the voltage of the input source voltage. The output DC voltage from the voltage doubler is stored in the supercapacitor to power the load.

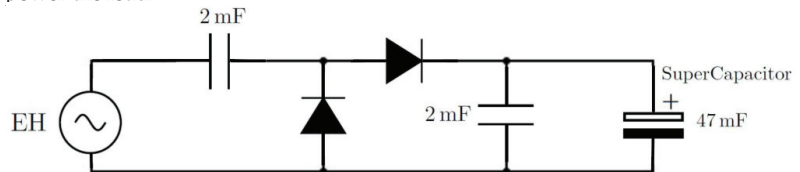


Figure 4. Circuit diagram of the voltage doubler used to rectify the input AC output voltage from the energy harvester.

The energy buffer for the system load is chosen to be a supercapacitor, due to being maintenance-free as well as for its long lifetime and superior power density over chemical batteries [29]. The microcontroller draws the stored energy in the supercapacitor and uses it to power the system while allowing the supercapacitor to recharge in between the transmission cycles. An RF solutions radio frequency transmitter/receiver module with a high operating voltage range is used for the vibration sensor packetized data ahead of post-processing by the custom-developed app. To allow for minimum energy consumption and constant power savings, when vibration energy falls below a threshold value, the microcontroller is set to go into idle mode. Similar to sleep mode, in idle mode, the CPU clock is turned off. However, in idle mode, the microcontroller peripheral clock stays on.

2.3. Receiver Subsystem

As shown in Figure 2, the receiver subsystem is composed of the RF receiver, an 8-bit microcontroller, USB to serial UART interface, and the custom-developed dynamic FFT app. The AM radio frequency receiver can receive the transmitted data at a range of 50 m. The microcontroller receives the bit-stream through a universal serial bus (USB) to transistor-transistor logic (TTL) interface at a rate of 300 baud. The received data packet is composed of 1 start bit, 8 data frame bits, and 1 stop bit. The data sampled from vibrations is then run through the in-house custom-made dynamic spectrum analysis FFT app. In the app, the user is prompted to turn data streaming on or off. Once the streaming is turned on, the app stores both voltage and time data into an equal size window buffer. FFT algorithm is carried out on the buffered vibration data, and the output vibration and amplitude of the signal are plotted in real-time on the graphical user interface (GUI) axes. A frequency tracking numeric field feature is also integrated into the app to offer a more distinguishable frequency monitoring.

3. Manufacturing and Prototyping

This section describes the details of manufacturing, integration, and assembly of the hardware of the self-powered sensor presented in this work. Additionally, details of signal acquisition and data post-processing via the custom-developed fault detection MATLAB app are discussed in this section.

3.1. Fabrication and Integration of System Hardware

The energy harvester-sensor unit was designed using a 3D CAD designing software (SolidWorks). The casing of the energy harvester was printed using a 3D printer and polylactic acid (PLA) filament. The FR4 sensor diaphragm customized patterns were fabricated using a Kern laser cutter (KER4824-Ti100 micro). The 130 W laser cutter was set at 80% of full power and a cutting rate of 20 mm/s. Three permanent solid magnets were used in the device assembly. The bottom magnet was fixed to the bottom support, the top magnet was fixed to the FR4 sensor diaphragm, and the levitated magnet was left to float. The levitated magnet was guided by the walls of the guiding rail tube as shown in Figure 3b. A manual winding machine (MXBAOHENG NZ-1) was used to wind the enameled copper coils around the harvester and sensor. Details and dimensions of the designed and fabricated structures are shown in Figures 5 and 6. A view of the final assembled energy harvester-sensor unit is shown in Figure 7. A list of the design specification and materials used to fabricate the energy harvester-sensor unit is shown in Table 1.

The transmitter circuit electronic components, shown in Figure 8, were affixed on an insulating board using soldering. The PIC microcontroller that was chosen (PIC16LF15325/45) had ultra-low-power features where it typically consumed only 8 μ A at 32 kHz oscillator frequency and 50 nA at 1.8 V in sleep mode. The PIC also included a windowed watchdog timer feature that was able to issue a reset to the microcontroller in the event of software failure. Furthermore, the microcontroller peripheral module disable (PMD) feature was used to disable all of the unused peripherals to minimize power consumption.

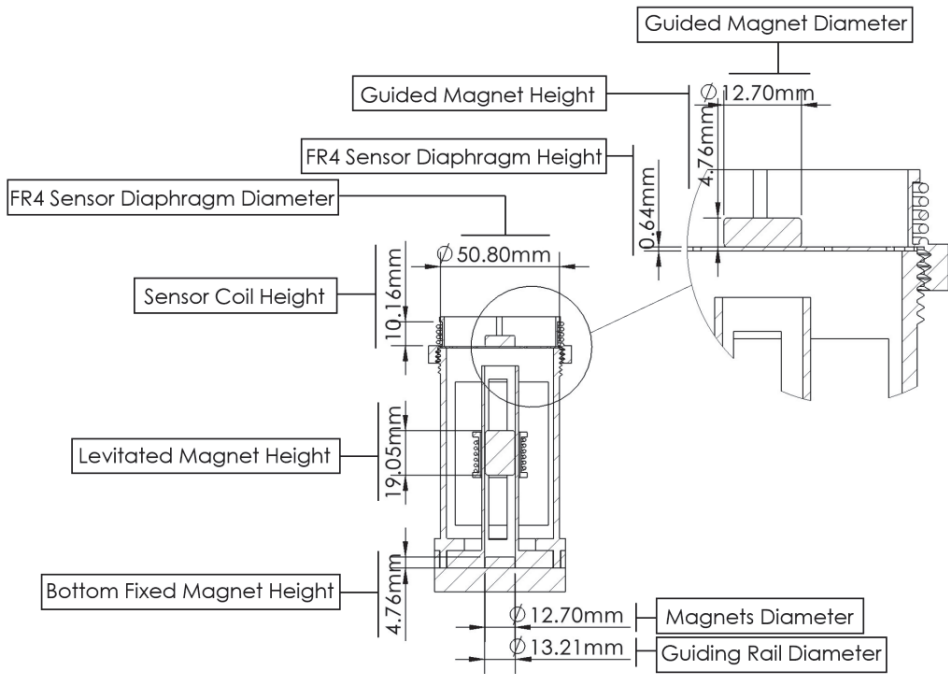


Figure 5. Cross-section view of the vibration energy harvester-sensor unit along with a blow-up detailed section view.

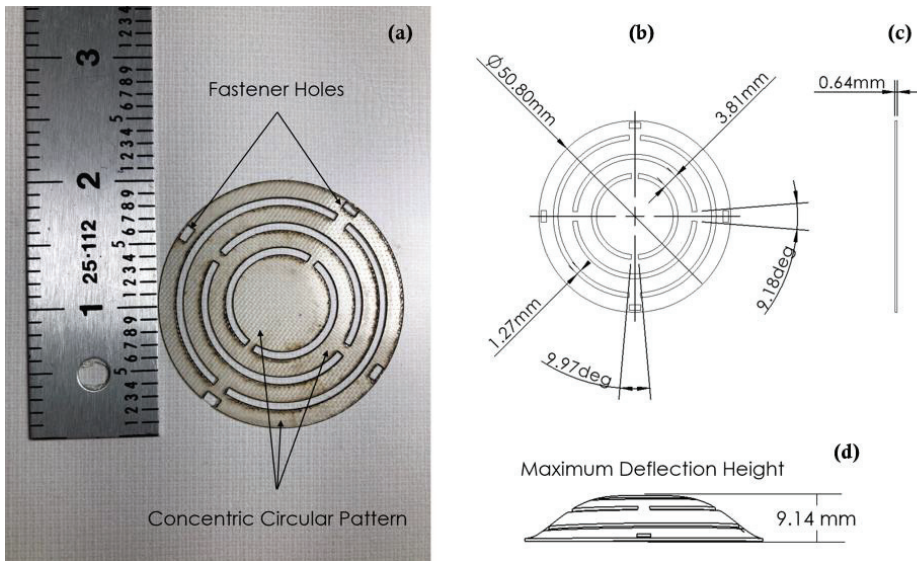


Figure 6. Design and fabrication of the FR4 Sensor diaphragm: (a) Fabricated sensor diaphragm next to a scale, (b) CAD model top view and dimensions of the sensor diaphragm, (c) CAD model side view of the sensor diaphragm and, and (d) simulated CAD view of the deflected sensor diaphragm.

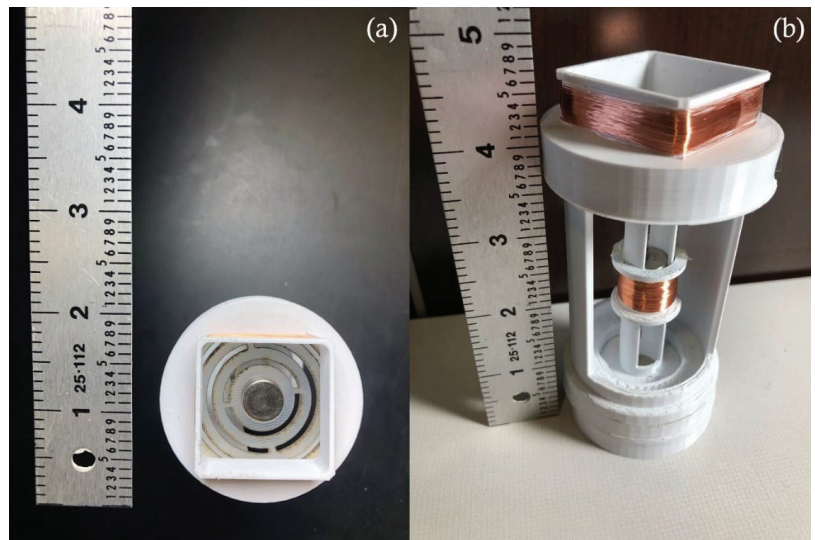


Figure 7. Fully assembled energy harvester-sensor unit next to a scale: (a) top view of the unit and (b) side view of the unit.

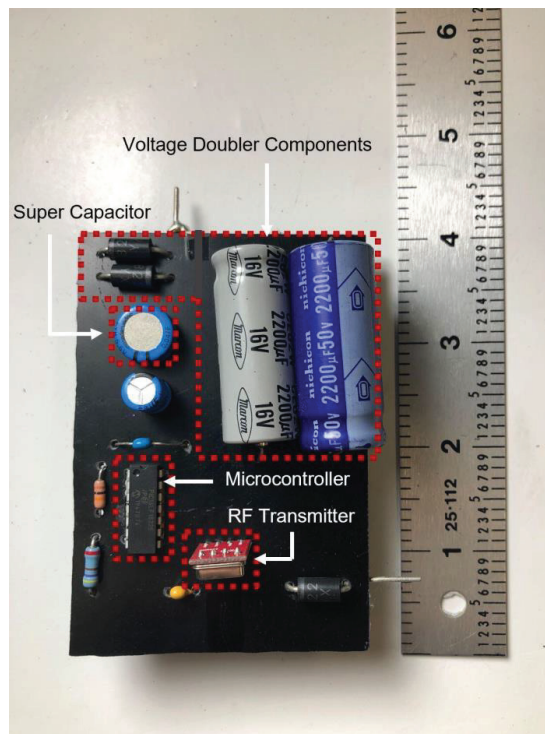


Figure 8. Circuit and components of the transmitter subsystem as part of the manufacturing process of the overall wireless sensor system presented in this work.

Table 1. Specification and details of the fabricated energy harvester-sensor unit.

| Specification | Value |
|-------------------------------|---|
| Energy Harvester | |
| Coil turns number | 450 |
| Coil resistance | 93 Ω |
| Minimum operable acceleration | 0.7 g m/s ² |
| Excitation frequency range | 5–10.6 Hz |
| Casing material | 3D printed polylactic acid (PLA) |
| Coil type | 40 AWG enameled copper |
| Magnet’s material | NdFeB type N42 |
| Sensor | |
| Coil turns number | 1500 |
| Coil resistance | 890 Ω |
| Spring diaphragm material | FR4 glass-reinforced epoxy resin laminate |
| Coil type | 40 AWG enameled copper |
| Magnet’s material | NdFeB type N42 |

Other components on the circuit board included a 433 MHz RF transmitter module, voltage doubler components incorporating two 16 V, 2200 μ F electrolytic capacitors, and two Schottky diodes. The circuit also holds a supercapacitor energy buffer size 47 mF with a 5.5 V voltage rating and a low equivalent series resistance (ESR) of 25 Ω . A circuit diagram and interconnections of the transmitter circuit board are shown in Figure 9.

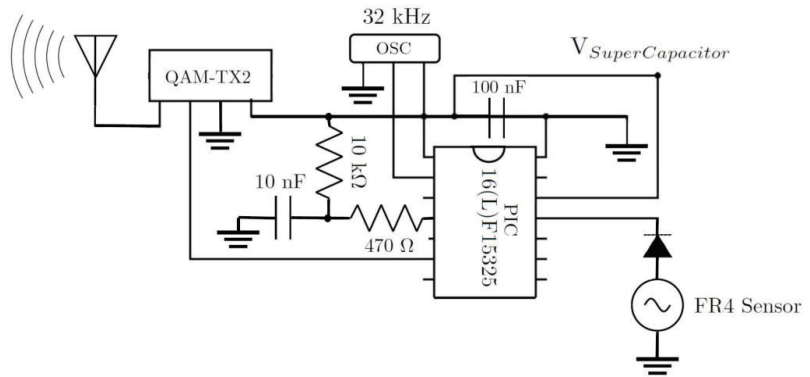


Figure 9. Circuit diagram and interconnections of the microcontroller transmitter subsystem components used in the wireless vibration sensor system presented in this work.

3.2. Signal Acquisition

Both microcontrollers used in the transmitter and receiver subsystems were programmed through the PIC embedded applications development freeware (MPLAB X IDE) and a PICKit 3 in-circuit debugger. The vibration signal acquisition process described in Figure 10 starts when the microcontroller exits the idle (sleep) mode. The microcontroller is programmed to stay in sleep mode until it is prompted to ‘wake up’ when triggered by the voltage level held by the supercapacitor. The microcontroller waking up process is set up to initiate when the sampled signal from the supercapacitor reaches (2 V). The supercapacitor voltage signal sampling takes place at one of the 10-bit low power successive approximation ADC channels of the microcontroller (ADC-CH1). The ADC allows conversion of the analog signal into a 10-bit binary form of that input signal. The ADC channel input voltage level can vary from 0 V up to a maximum voltage that needs to be set as the reference voltage of the channel. A reference voltage is needed by the ADC to create a range of voltages that are mapped into specific length binary values. The supercapacitor sampled input voltage is then compared to those binary values from the reference. Typically,

the reference voltage used for ADC channels is the supply voltage to the microcontroller. This poses a challenge since the supply voltage and the sampled voltage would both be of the supercapacitor voltage. This was overcome by using the microcontroller internal fixed voltage reference (FVR) feature that is independent of the microcontroller supply voltage. A programmable independent buffer gain amplifier is used at the output of the FVR and is set to amplify the voltage reference to a desired selectable voltage level. During sampling, the supercapacitor voltage is compared to the FVR and when the value returned matches the selected voltage level, power is delivered to the rest of the system after waking the microcontroller up. During the microcontroller system initialization process, voltage rails are stabilized, and the CPU starts fetching code instructions and data to operate the necessary control registers.

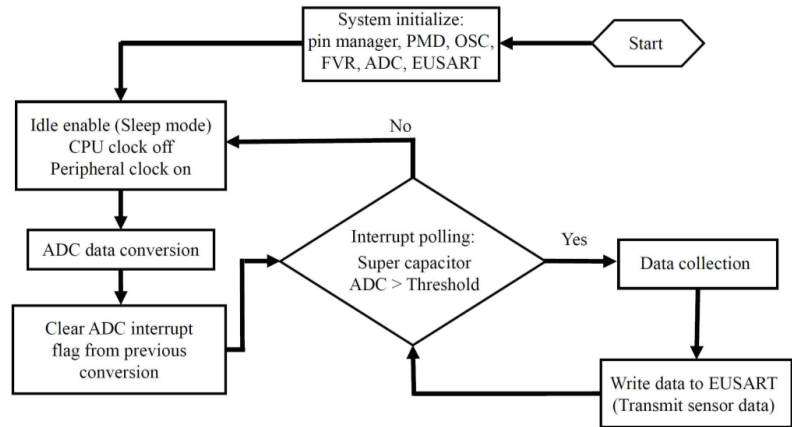


Figure 10. A signal acquisition process flowchart of the wireless vibration sensor presented in this work showing detailed description of the program starting at the initialization process of the microcontroller and ending at the transmission of data.

Embedded systems interrupts are hardware features that preempt normal program code operation in order to execute a command that requires CPU attention. As shown in Figure 10, when an interrupt is set, an interrupt service routine firmware (ISR) determines the source of the interrupt by the process of polling. The ISR polling protocol is an active process of monitoring interrupt flag bits from the interrupt flag register. The peripheral interrupt from the supercapacitor voltage level value reaching the set threshold allows the CPU to service that interrupt and wake the microcontroller from sleep. As indicated in Figure 10, unless the supercapacitor voltage level is above the preset threshold, the ISR will continue polling and remain in sleep mode. A command to clear the interrupt flag from a previous interrupt is necessary to execute following the servicing of an interrupt. If the interrupt flag is not cleared, and if the supercapacitor voltage level is above the preset threshold, interrupts will occur repeatedly overriding necessary CPU functions.

In idle mode the microcontroller CPU core and memory operations are halted while the internal peripheral clocks such as ADC channels clock continue to operate. Once the microcontroller is woken up, the sensor data are sampled through a second ADC channel (ADC-CH2). In Figure 10, the ADC logs the sensor data and writes it to the enhanced universal synchronous and asynchronous receiver transmitter (EUSART) register upon waking up from idle mode. The data are then wirelessly transmitted via the 433 MHz AM RF transmitter module. The signal is then received by a compatible receiver module (QAM-RX10) that is connected to a second microcontroller in order to receive the data through the serial port of a PC by utilizing a USB to serial converter. The antenna-equipped receiver provides a two-way communication that transforms the electromagnetic waves

into electrical signals. Modulation of the baseband data onto the carrier is accomplished by amplitude shift keying (ASK) of the signal.

3.3. Custom Developed Fault Detection MATLAB App

In this work, the monitoring app is developed using MATLAB-GUI to extract useful information from the collected vibration signal including maximum displacement amplitude and frequency. The extracted information is then compared against preset threshold values to assess the risk level associated with the operation performed.

Inside the MATLAB app development environment, the GUI components including the data streaming switch, frequency tracking numeric field, status indicator, and real time FFT plot are identified as dynamic objects. The objects are chosen from a MATLAB supported components library as seen in Figure 11. The dynamic objects are configured as public access properties that allow data exposure to the user through the GUI. The corresponding values to the communication port and baud rate of the receiver board are then specified. This allows for initiation of data streaming through the serial port when the 'on' dynamic object under the data streaming label is selected by the user. Evenly sized sectioned data buffers are set up for both voltage and time elements to allocate for real time data plotting. Parameters of the FFT measurements including the length of the signal, sampling frequency, and Nyquist frequency are identified to convert the windowed time domain data into frequency domain. A peak finder function is then used to detect the dominant frequency from the converted data for display in the GUI numeric field region. The amplitude of the resultant peak is then compared against fixed threshold values to vary the color of the status lamp indicator on the GUI. The variable colors of the lamp give the user a risk severity measure of the performed operation. Therefore, the indicator lamp switches color from green when the energy harvester sensor unit is subject to low amplitude vibrations to red once it experiences higher vibration amplitudes. Further details about the custom developed fault detection MATLAB app are discussed later in this article (i.e., Section 5).

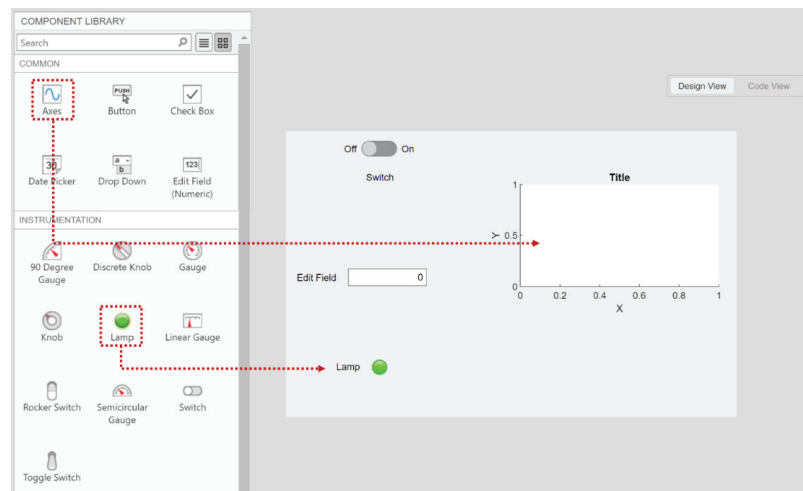


Figure 11. The design view browser inside MATLAB app development environment. The dynamic objects are chosen from a component library and are populated into the GUI canvas.

4. Experimental Methods and Characterization Techniques

The experimental testing setup shown in Figure 12a was used to measure the dynamic frequency response of the wireless sensor system presented in this work. An illustration of the signal and power flow of the experimental testing equipment setup is shown in

Figure 12b. In the closed loop vibration testing system, as shown in Figure 12a, a vibration controller (S81B-P02, SENTEK DYNAMICS) is directed by a PC software. Initially, the test is conducted by setting up frequency, acceleration as well as time elements through the software’s control settings. The settings are managed through Crystal Instrument’s engineering and data management (EDM) vibration control system (VCS) software. The preset commands of the test are then transferred into a vibration controller. The controller sends a drive signal to a power amplifier in order to drive a rectilinear shaker table. The shaker table (VT-500 by SENTEK DYNAMICS) transforms the drive signal into mechanical vibrations that are transferred into the energy harvester-sensor unit attached to the shaker armature.

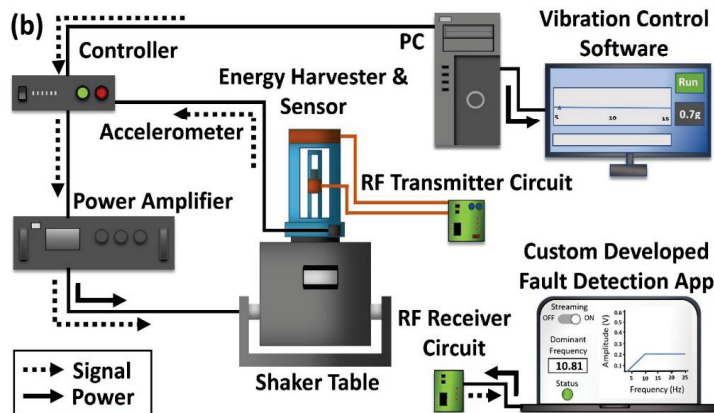
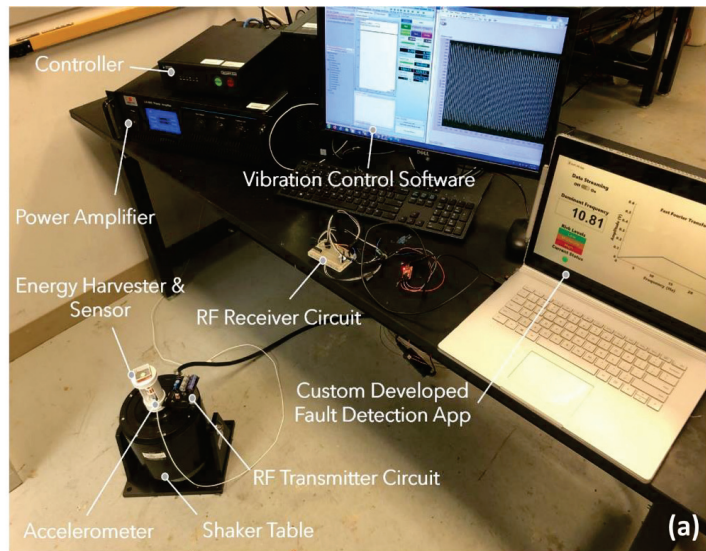


Figure 12. Experimental methods used in this work: (a) experimental apparatus used for characterization of the wireless self-powered vibration sensor system presented in this work; (b) cartoon schematic of the characterization setup showing signal and power flow in the equipment.

When performing the experiments, the lower end of the energy harvester-sensor unit is secured on top of the shaker table as shown in Figure 12. The vibration response from the energy harvester is measured by an accelerometer (PCB333B30 by PCB Piezotronics).

The RF transmitter board situated on the static outer rim of the shaker table is connected to the energy harvester-sensor unit as shown in Figure 12a. The RF receiver circuit board is connected to the laptop via the USB to serial TTL level FTDI cable for live frequency response analysis as shown in Figure 12a,b.

5. Results and Discussion

Using the experimental apparatus shown in Figure 12, dynamic characterization of the fabricated energy harvester-sensor unit was performed. The resulting voltage frequency responses of both the energy harvester and sensor are shown in Figures 13 and 14. The energy harvester-sensor unit was subject to fixed input acceleration values ranging from 0.1 g up to 0.7 g [m/s^2] while the frequency was swept at a rate of 0.0833 Hz/s.

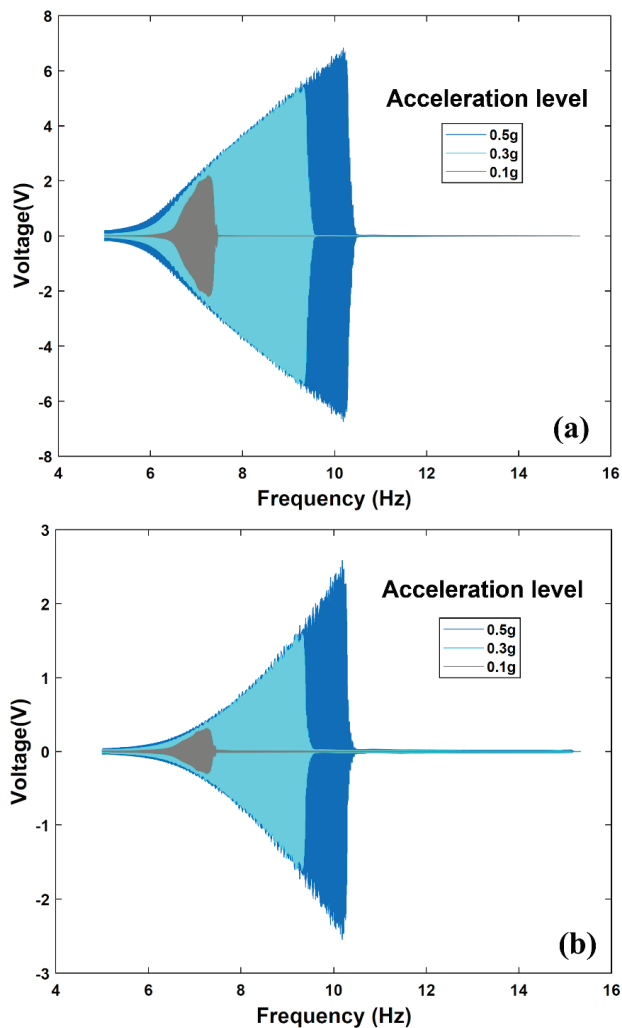


Figure 13. Energy harvester-sensor unit open circuit frequency response at a range of input accelerations. (a) Open circuit voltage of the energy harvester. (b) Open circuit voltage from the FR4 sensor diaphragm.

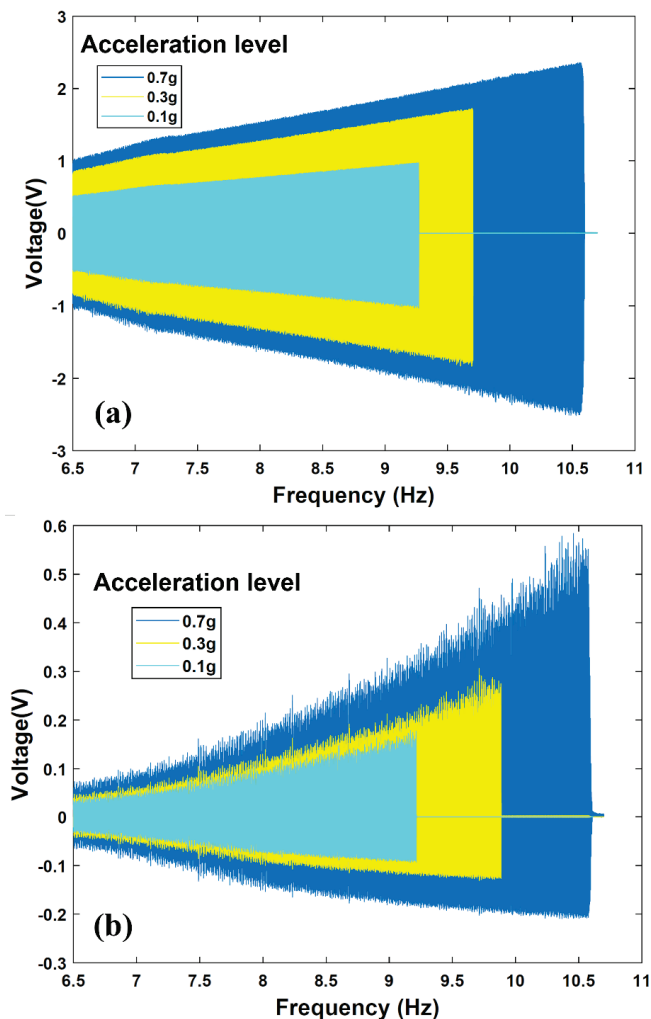


Figure 14. Energy harvester-sensor unit closed circuit frequency response at a range of input accelerations. (a) Energy harvester closed circuit voltage values measured across the microcontroller subsystem. (b) FR4 sensor diaphragm closed circuit voltage values measured across the input to the microcontroller.

The nonlinear magnetic spring stiffness nature of the energy harvester is evident in the voltage frequency response as shown in both Figures 13 and 14. That is, the repulsive magnetic forces experienced by the levitated magnet can be described as a nonlinear spring force [7,8]. This results in a hardening effect that is evident when comparing the trend in output voltage peaks shown in Figures 13 and 14. One can notice that these peaks are shifting to higher frequencies as the input acceleration is increased. The nonlinear behavior of magnetic levitation-based energy harvesting system was studied extensively in our prior work [5,7,8,29]. Furthermore, during the frequency sweep, an abrupt sharp decline of the energy harvester's frequency response is evident. This drop, known as the frequency jump phenomenon [7,8], is attributed to the coexistence of multiple energy states at the frequency

branch. The discontinuity in the response is a characteristic of magnetic levitation-based energy harvesting systems [5,29].

The supercapacitor charging and discharging cycles were measured using a data acquisition device (NI myDAQ) and the graphical programming environment LabVIEW software. The supercapacitor charging history is shown in Figure 15 where the maximum charge is held at 3.7 V after 122 s. During the charging cycle and after approximately 39 s, a slight shift in voltage level that lasts for 4 s occurs during the microcontroller waking up and voltage rails stabilization stage. The rate of the voltage held by the supercapacitor is seen to decrease from 44 mV/s to 22 mV/s after the microcontroller startup period. This is likely due to the fact that during this stage the system is draining the supercapacitor in order to operate. After approximately 122 s of charging, while simultaneously the energy harvester-sensor unit crosses the resonant frequency point during the frequency sweep, the supercapacitor discharging cycle starts to take effect at a voltage decline rate of 15 mV/s. The supercapacitor charging period after the microcontroller startup takes 79 s while it takes 108 s to discharge before the microcontroller enters the sleep mode as shown in Figure 15. At 230 s and 2 V, the microcontroller enters the sleep mode for a period that lasts approximately 50 s before the supercapacitor voltage level drops below 1.8 V which is no longer a sufficient amount of power for the microcontroller to operate. Consequently, the supercapacitor starts a self-discharge process as shown in Figure 15.

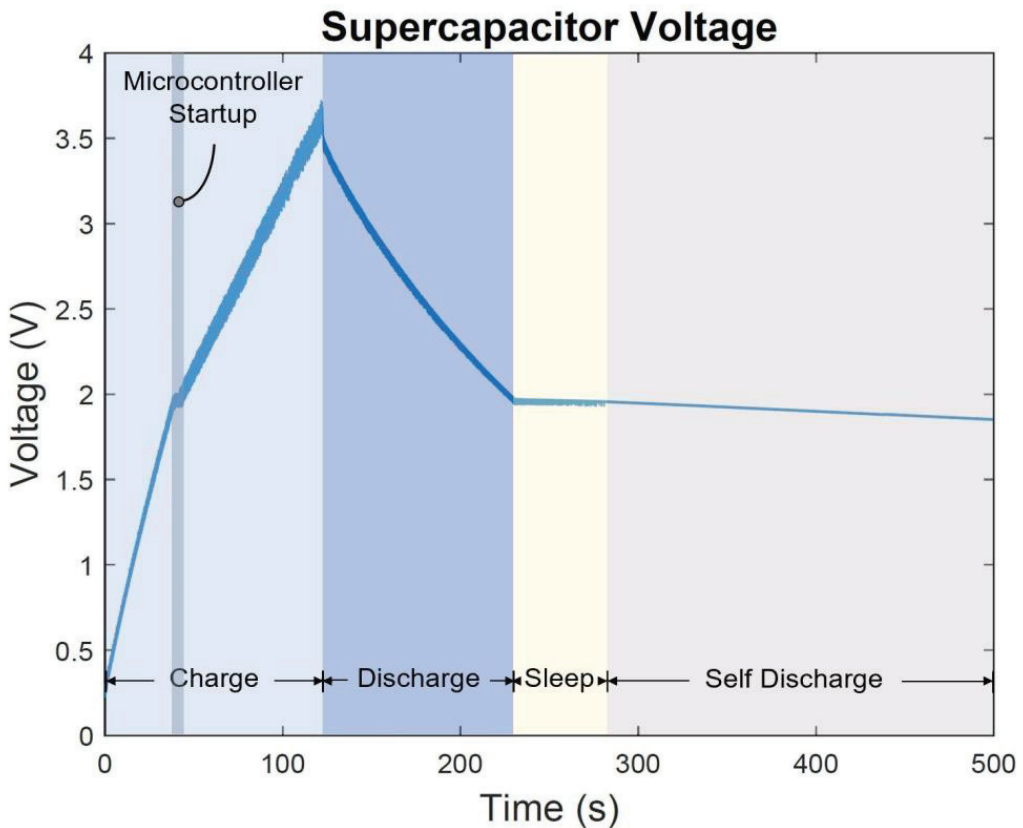


Figure 15. The supercapacitor voltage charge and discharge cycle during frequency sweep 6.5–15 Hz at 1 g [m/s²] acceleration level.

Next, the fully integrated, self-powered, self-contained sensor system was subject to harmonic oscillations and was tested under lab-controlled settings using the experimental apparatus shown in Figure 12. The resulting energy consumptions at different system states are shown in Table 2. The microcontroller internal oscillator was set to the lowest clock frequency of 32 kHz to minimize energy consumption. The ADC peripheral was configured to use the internal system clock oscillator. The microcontroller takes 10 measurements from the sensor signal during analog to digital bit conversion at a sampling rate of 51 kHz. After the measurements conversion is completed in 19.55 μ s, the data are stored in the 16-bit ADC results register. The data are then retrieved by the EUSART serial communication peripheral, and the 10-bit data frame is transmitted asynchronously with 1 start bit and 1 stop bit added to each 8-bit sensor data packet and no parity bit.

Table 2. Transmitter circuit subsystem energy consumption at different program states.

| Component | Power | Duration | Energy |
|--|----------|---------------|---------------|
| Microcontroller startup (32 kHz) | 0.879 mW | 22.5 μ s | 19.7 nJ |
| Microcontroller operation (32 kHz) | 0.756 mW | 17.6 μ s | 13.32 nJ |
| Radio transmission (8 bits, 76.8 kbps) | 8.58 mW | 0.1 ms | 0.858 μ J |
| Sensor reading | 0.144 mW | 19.55 μ s | 2.8 nJ |
| Total | | | 0.894 μ J |

Figure 16 shows the power consumed by the transmitter subsystem as well as the power generated at an unloaded voltage doubler circuit. The power generated from the unloaded voltage doubler at resonant frequency is 12.6 mW and it exceeds the power consumed by the transmitter subsystem load of 5.1 mW. In the power consumed curve, the decline in power is more gradual compared to the abrupt decline seen in the power generated curve due to having the supercapacitor in the transmitter subsystem circuitry. In both power curves, the profiles show maximum power generation and consumption at resonant frequency of 9.9 Hz.

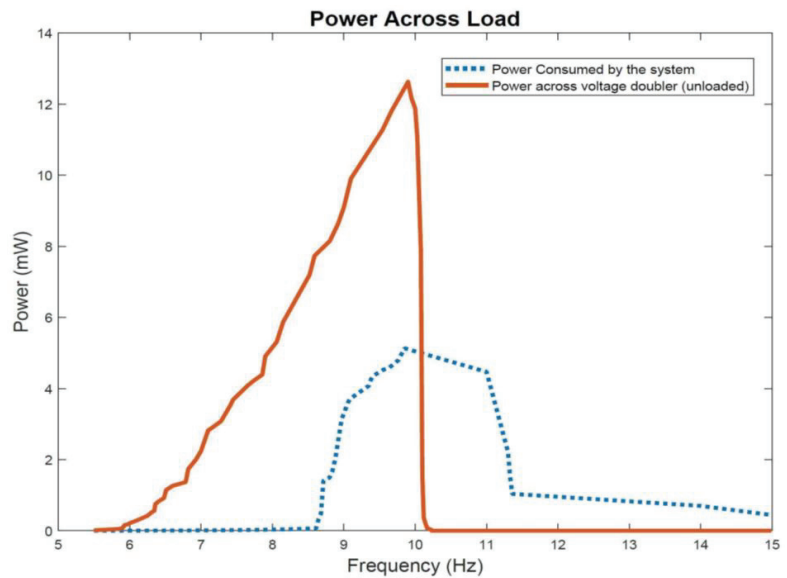


Figure 16. Power generated from the energy harvester at an unloaded voltage doubler output compared to power consumed by the microcontroller-transmitter subsystem load.

The operation and realization of the fully integrated self-powered fault detection system are examined next. A video recording of the self-powered, self-contained, environment-friendly, and wireless vibration sensor during data acquisition and transmission process is included in the Supplementary Materials. The video recording includes segments from the custom developed fault detection app along with the EDM shaker table vibration control software and the energy harvester and sensor voltage waveforms as well as the experimental setup showing the fabricated energy harvester-sensor unit mounted on the shaker table.

Figures 17–20 show selected timeframes showing the operation of the fully integrated self-powered wireless fault detection system at different stages during the system's operation. Figures 17–20 demonstrate the ability to self-power the sensor and the ability of the frequency tracker in the custom developed app to detect the vibration frequency from the self-powered sensor. For example, Figure 17a shows the vibration frequency from the sensor, detected by the frequency tracker in the custom developed app, at 9.479 Hz. The sensor vibration frequency value can be confirmed by the EDM software shaker table preset frequency, as shown in Figure 17b. Here, the preset frequency was approximately 9.007 Hz. Thus, there is a very slight discrepancy between the frequency detected by the frequency tracker in the custom developed app and the preset frequency from the shaker table. The error is estimated at approximately 5%. This discrepancy may be attributed to a few factors including the high noise in the AM transmitter/receiver module that is not filtered out during the FFT filtering process. Other factors include the use of the internal microcontroller oscillator in this work as opposed to using an external oscillator which would have resulted in a more accurate frequency stability. A bitrate estimated error of approximately 1.24% in the transmitted data is also a contributing factor to the overall error in the received and filtered data. Moreover, Figure 17c demonstrates the output signal from the fabricated self-powered, self-contained sensor as the top magnet moves inside the top coil in response to the detected vibrations. That is, the dynamic displacement induced in the sensor causes induced voltage in the top coil surrounding the top guided magnet as shown in Figure 17c. Here, the power generated by the energy harvester is used to operate the RF transmitter circuit. The output voltage from the sensor is then sampled by the microcontroller on the transmitter circuit. The app, shown in Figure 17a, also demonstrates successful status indicator monitoring capabilities where it shows a green light, indicating a low level of acceleration experienced by the sensor (i.e., low risk operation). Similarly, Figures 18–20 show subsequent timeframes from the demonstration experiment. In those timeframes, the sensor is self-powered, and the app detects a transition from a lower risk operation to a higher risk operation as seen signaled by the status indicator turning red in Figure 18. In Figure 18, the preset frequency from the vibration source (i.e., the shaker table) was approximately 9.652 Hz whereas the app detected tracker frequency was at 9.748 Hz. This results in a small error of approximately 1%. Likewise, the next timeframe in Figure 19 shows the preset frequency was at 10.98 Hz with the app detecting a 10.86 Hz also corresponding to a 1% error in the detected frequency value. The last presented timeframe from the demonstration experiment is shown in Figure 20 with the app detecting a frequency of 11.05 Hz compared to the preset frequency of 11.15 Hz resulting in only a 0.9% error. The mean absolute error value for all frequencies under investigation is found to be 0.8 while the mean percent error value is 2.6%.

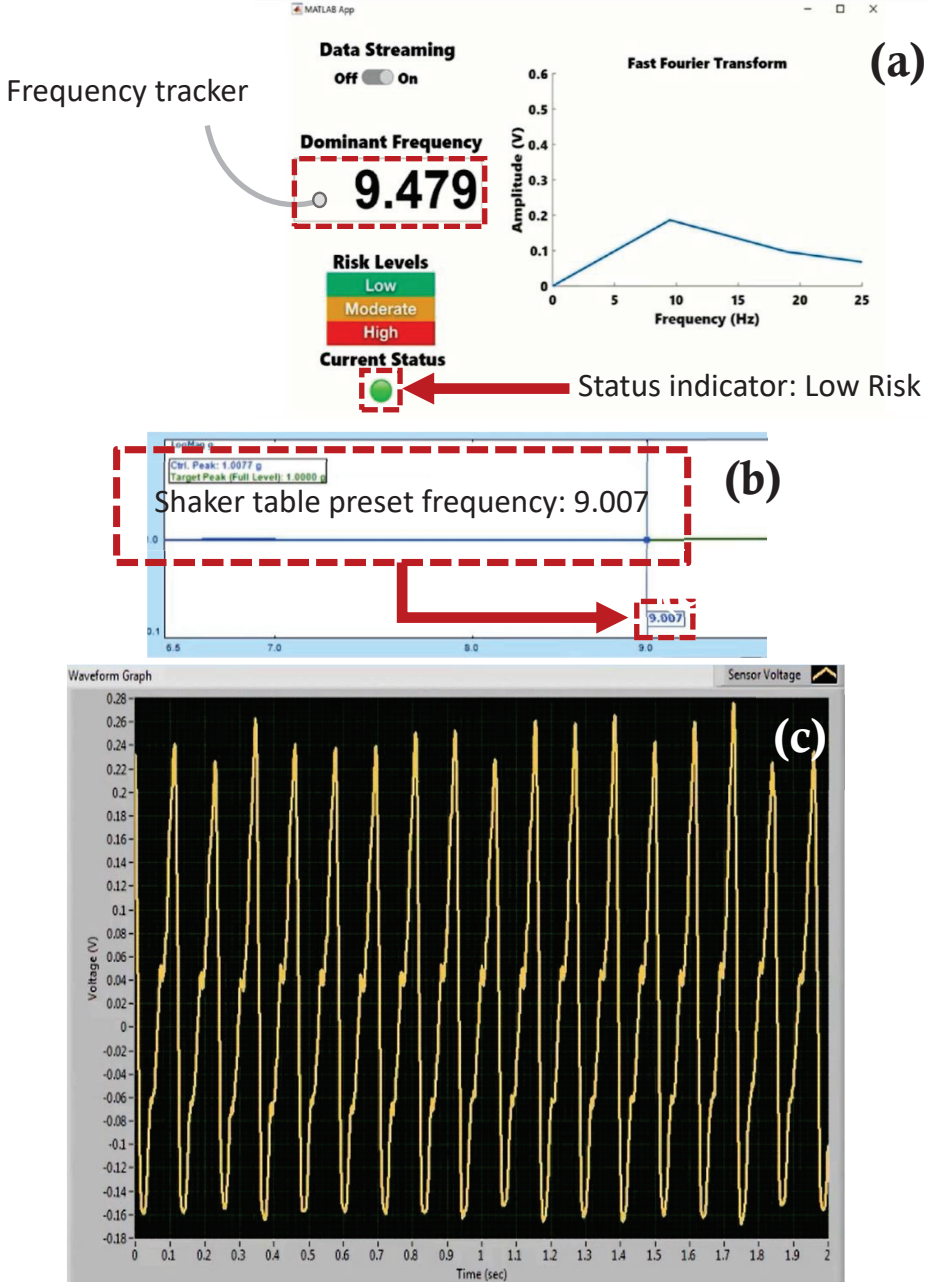


Figure 17. Screen captures of the self –powered, self –contained, vibration sensor system taken from the demonstration experiment: (a) custom developed fault detection app with a frequency tracker feature and status indicator for risk monitoring, (b) EDM shaker table vibration control software with preset frequency monitor, and (c) sensor voltage waveform monitor at preset frequency 9.007 Hz.

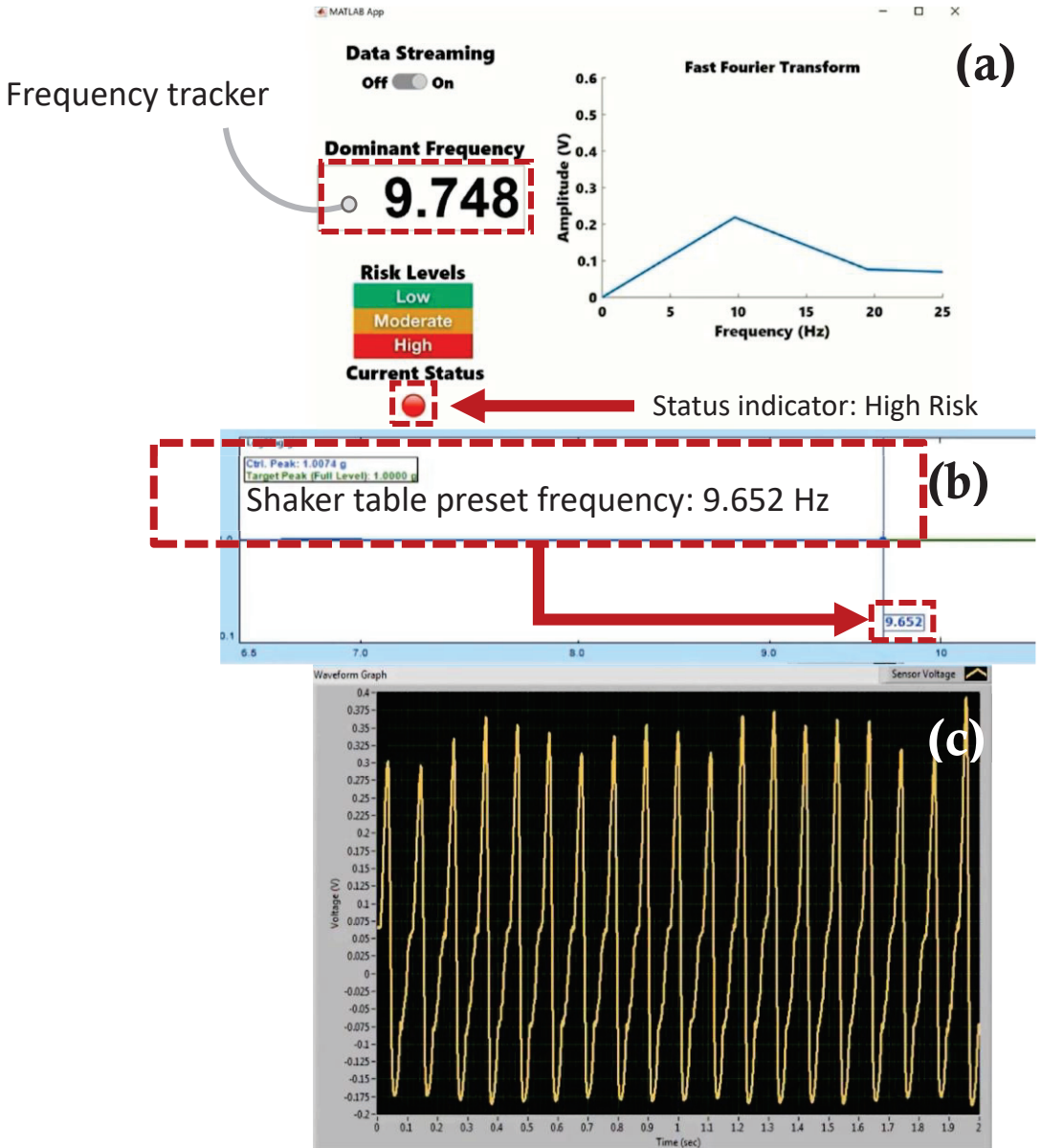


Figure 18. Screen captures of the self –powered, self –contained, vibration sensor system taken from the demonstration experiment: (a) custom developed fault detection app with a frequency tracker feature and status indicator for risk monitoring, (b) EDM shaker table vibration control software with preset frequency monitor, and (c) sensor voltage waveform monitor at preset frequency 9.652 Hz.

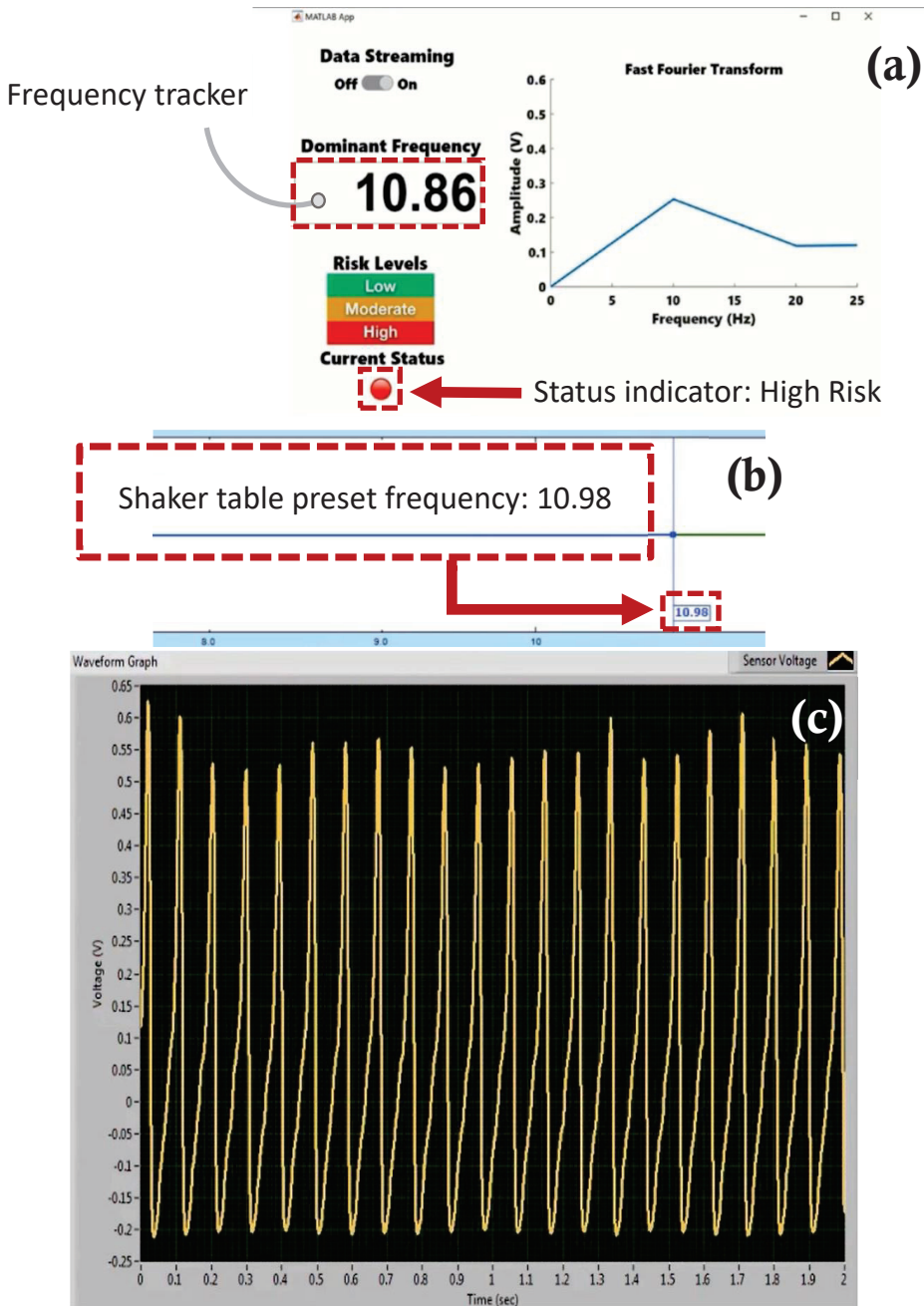


Figure 19. Screen captures of the self –powered, self –contained, vibration sensor system taken from the demonstration experiment: (a) custom developed fault detection app with a frequency tracker feature and status indicator for risk monitoring, (b) EDM shaker table vibration control software with preset frequency monitor, and (c) sensor voltage waveform monitor at preset frequency 10.98 Hz.

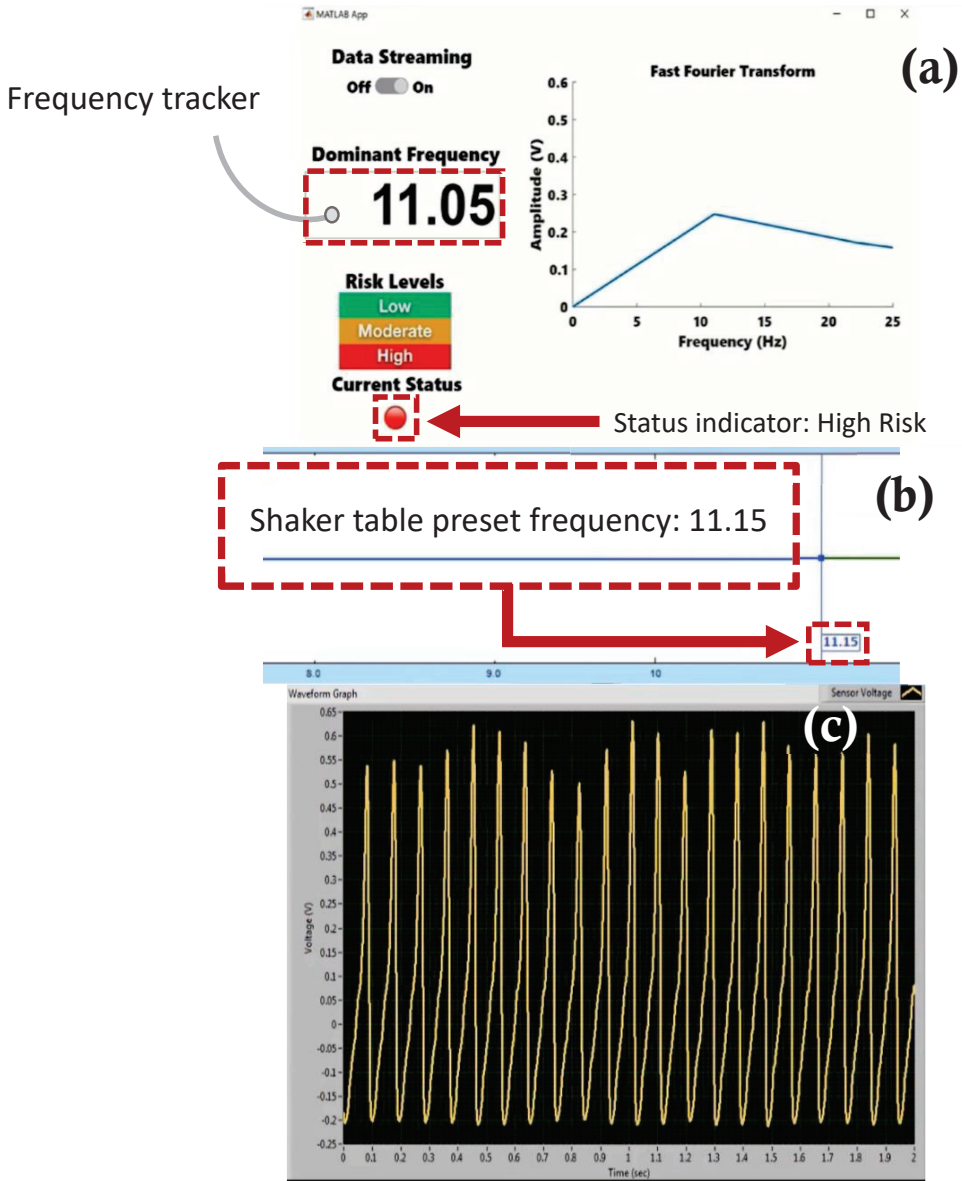


Figure 20. Screen captures of the self –powered, self –contained, vibration sensor system taken from the demonstration experiment: (a) custom developed fault detection app with a frequency tracker feature and status indicator for risk monitoring, (b) EDM shaker table vibration control software with preset frequency monitor, (c) sensor voltage waveform monitor at preset frequency 11.15 Hz.

6. Conclusions

In this work, we have introduced a novel self-powered self-contained wireless vibration sensor for fault detection in dynamic structures. The energy harvester-sensor unit is based on dual mass moving magnets. The voltages are extracted from the moving magnets by the coil surrounding the casing around the magnets. The power produced by the energy

harvester subunit is used to operate an RF based transmitter subsystem. The transmitter subsystem sends mechanical vibration levels through a sensor subunit to a custom developed fault detection app wirelessly. The app notifies the user of the degree of risk associated with the operation by applying an FFT algorithm to the transmitted vibration data. The app can identify the frequency of the vibration with a low error of approximately 1% in most of the transmitted values. Unlike commonly studied self-powered vibration based WSN transmitter subsystems, this work utilizes the active power of the sensor subunit during the transmission process as opposed to requiring to power a passive vibration sensor. The transmitter subsystem operates at ultra-low power where the total consumption of energy to transmit a sensor value is approximately 0.894 μJ at 3 V. The transmitter subsystem can transmit the data from the sensor at a minimum operable acceleration of 0.7 g [m/s^2] and an excitation range of 5–10.6 Hz. The significance in this work also lies in the low frequency required to operate the energy harvester-sensor unit that is widely available in many surrounding environments. Future work will consider employing a more energy-aware circuit with maximum power point (MPPT) tracking capability. This enhancement will allow for a self-sufficient stand-alone field operation of the energy harvester sensor unit.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/s22062352/s1>. Video S1: Demonstration experiment.

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

Funding: This work was partially supported by the Louisiana Department of Transportation-Transportation Research Center [Contract DOTLT1000369], Louisiana Board of Regents-National Science Foundation (BoR/NSF) Grant and Cooperative Agreement Number [80NSSC20M0110], and Louisiana Board of Regents Support Fund contract number [LEQSF(2020-24)-LaSPACE]. Dr. Bardaweel is W.W. Chew II Endowed Professor. W.W. Chew II Professorship is made available through the State of Louisiana Board of Regents Support Fund. The APC of this article was funded by W.W. Chew II Professorship. The views expressed in this article are those of the authors and do not reflect the official policy or position of the funding agencies.

Data Availability Statement: Data are contained within the article or Supplementary Material. The data presented in this study are available in [this manuscript].

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

References

- Rodriguez, J.; Nico, V.; Punch, J. Powering Wireless Sensor Nodes for Industrial IoT Applications using Vibration Energy Harvesting. In Proceedings of the 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), Limerick, Ireland, 15–18 April 2019; pp. 392–397.
- Avci, O.; Abdeljaber, O.; Kiranyaz, S.; Hussein, M.; Gabbouj, M.; Inman, D. A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. *Mech. Syst. Signal Process.* **2021**, *147*, 107077. [[CrossRef](#)]
- Li, X.; Teng, L.; Tang, H.; Chen, J.; Wang, H.; Liu, Y.; Fu, M.; Liang, J. VIPSN: A vibration-powered iot platform. *IEEE Internet Things J.* **2020**, *8*, 1728–1739. [[CrossRef](#)]
- Nguyen, H.; Bardaweel, H. Design guidelines and parametric study of nonlinear magnetic springs for vibration systems. In *Energy Harvesting and Storage: Materials, Devices, and Applications X*; International Society for Optics and Photonics: Bellingham, WA, USA, 2020; Volume 11387, pp. 12–21.
- Nguyen, H.; Genov, D.; Bardaweel, H. Mono-stable and bi-stable magnetic spring based vibration energy harvesting systems subject to harmonic excitation: Dynamic modeling and experimental verification. *Mech. Syst. Signal Process.* **2019**, *134*, 106361. [[CrossRef](#)]
- Kim, J.; Salauddin, M.; Cho, H.; Rasel, M.; Park, J. Electromagnetic energy harvester based on a finger trigger rotational gear module and an array of disc Halbach magnets. *Appl. Energy* **2019**, *250*, 776–785. [[CrossRef](#)]
- Nguyen, H.; Genov, D.; Bardaweel, H. Vibration energy harvesting using magnetic spring based nonlinear oscillators: Design strategies and insights. *Appl. Energy* **2020**, *269*, 115102. [[CrossRef](#)]

8. Aldawood, G.; Nguyen, H.; Bardaweel, H. High power density spring-assisted nonlinear electromagnetic vibration energy harvester for low base-accelerations. *Appl. Energy* **2019**, *253*, 113546. [\[CrossRef\]](#)
9. Reza, A.; Talevski, A.; Carlsen, S.; Petersen, S. Applications of wireless sensor networks in the oil, gas and resources industries. In Proceedings of the 2010 24th IEEE International Conference on Advanced Information Networking and Applications, Perth, Australia, 20–23 April 2010; pp. 941–948.
10. Carneiro, P.; dos Santos, M.P.; Rodrigues, A.; Ferreira, J.A.; Simões, J.A.; Marques, A.T.; Kholkin, A.L. Electromagnetic energy harvesting using magnetic levitation architectures: A review. *Appl. Energy* **2020**, *260*, 114191. [\[CrossRef\]](#)
11. Roundy, S.; Wright, P.; Rabaey, J. A study of low level vibrations as a power source for wireless sensor nodes. *Comput. Commun.* **2003**, *26*, 1131–1144. [\[CrossRef\]](#)
12. Xiong, H.; Wang, L. Piezoelectric energy harvester for public roadway: On-site installation and evaluation. *Appl. Energy* **2016**, *174*, 101–107. [\[CrossRef\]](#)
13. Miller, L.; Halvorsen, E.; Dong, T.; Wright, P. Modeling and experimental verification of low-frequency (MEMS) energy harvesting from ambient vibrations. *J. Micromech. Microeng.* **2011**, *21*, 45029. [\[CrossRef\]](#)
14. Reilly, E.; Burghardt, F.; Fain, R.; Wright, P. Powering a wireless sensor node with a vibration-driven piezoelectric energy harvester. *Smart Mater. Struct.* **2011**, *20*, 125006. [\[CrossRef\]](#)
15. Wang, L.; Zhao, L.; Luo, G.; Zhao, Y.; Yang, P.; Jiang, Z.; Maeda, R. System level design of wireless sensor node powered by piezoelectric vibration energy harvesting. *Sens. Actuators A Phys.* **2020**, *310*, 112039. [\[CrossRef\]](#)
16. Huet, F.; Boitier, V. Design Strategy of Conventional Electronic for Wireless Sensor Node Powered by Vibration Energy Harvester. In Proceedings of the International Conference on Renewable Energies and Power Quality (ICREPQ'19), Tenerife, Spain, 10–12 April 2019.
17. Wang, L.; He, T.; Zhang, Z.; Zhao, L.; Lee, C.; Luo, G.; Mao, Q.; Yang, P.; Lin, Q.; Li, X.; et al. Self-sustained autonomous wireless sensing based on a hybridized TENG and PEG vibration mechanism. *Nano Energy* **2021**, *80*, 105555. [\[CrossRef\]](#)
18. Xiang, T.; Chi, Z.; Li, F.; Luo, J.; Tang, L.; Zhao, L.; Yang, Y. Powering indoor sensing with airflows: A trinity of energy harvesting, synchronous duty-cycling. In Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, Roma, Italy, 11–15 November 2013.
19. Singh, J.; Kaur, R.; Singh, D. Energy harvesting in wireless sensor networks: A taxonomic survey. *Int. J. Energy Res.* **2021**, *45*, 118–140. [\[CrossRef\]](#)
20. Carrano, R.; Passos, D.; Magalhaes, L.; Albuquerque, C. Survey and Taxonomy of Duty Cycling Mechanisms in Wireless Sensor Networks. *IEEE Commun. Surv.* **2014**, *16*, 181–194. [\[CrossRef\]](#)
21. Mazunga, F.; Nechibvute, A. Ultra-low power techniques in energy harvesting wireless sensor networks: Recent advances and issues. *Sci. Afr.* **2021**, *11*, e00720. [\[CrossRef\]](#)
22. Bahbahani, M.S.; Alsusa, E. A Cooperative Clustering Protocol With Duty Cycling for Energy Harvesting Enabled Wireless Sensor Networks. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 101–111. [\[CrossRef\]](#)
23. Wang, Y. Topology control for wireless sensor networks. In *Wireless Sensor Networks and Applications*; Springer: Boston, MA, USA, 2008; pp. 113–147.
24. Jawhar, I.; Mohamed, N.; Al-Jaroodi, J. Networking architectures and protocols for smart city systems. *J. Internet Serv. Appl.* **2018**, *9*, 26. [\[CrossRef\]](#)
25. Kumar, V.; Jain, A.; Barwal, P.N. Wireless sensor networks: Security issues, challenges and solutions. *Int. J. Inf. Comput. Technol.* **2014**, *4*, 859–868.
26. Amirtharajah, R.; Collier, J.; Siebert, J.; Zhou, B.; Chandrakasan, A. DSPs for energy harvesting sensors: Applications and architectures. *IEEE Perv. Comput.* **2005**, *4*, 72–79. [\[CrossRef\]](#)
27. Nguyen, T.Q. A Data-Driven Approach to Structural Health Monitoring of Bridge Structures Based on the Discrete Model and FFT-Deep Learning. *J. Vib. Eng. Technol.* **2021**, 1–23. [\[CrossRef\]](#)
28. He, K.; Zhu, W. Structural damage detection using changes in natural frequencies: Theory and applications. *J. Phys. Conf. Ser.* **2011**, *305*, 012054. [\[CrossRef\]](#)
29. Nammari, A.; Caskey, L.; Negrete, J.; Bardaweel, H. Fabrication and characterization of non-resonant magneto-mechanical low-frequency vibration energy harvester. *Mech. Syst. Signal Process.* **2018**, *102*, 298–311. [\[CrossRef\]](#)

Article

Partial Transfer Ensemble Learning Framework: A Method for Intelligent Diagnosis of Rotating Machinery Based on an Incomplete Source Domain

Gang Mao, Zhongzheng Zhang, Sixiang Jia, Khandaker Noman and Yongbo Li *

MIT Key Laboratory of Dynamics and Control of Complex System, School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China; mg0207@yeah.net (G.M.); zhangzhongzheng@mail.nwpu.edu.cn (Z.Z.); sixiang_j@163.com (S.J.); khandakernoman93@nwpu.edu.cn (K.N.)
* Correspondence: yongbo@nwpu.edu.cn

Abstract: Most cross-domain intelligent diagnosis approaches presume that the health states in training datasets are consistent with those in testing. However, it is usually difficult and expensive to collect samples under all failure states during the training stage in actual engineering; this causes the training dataset to be incomplete. These existing methods may not be favorably implemented with an incomplete training dataset. To address this problem, a novel deep-learning-based model called partial transfer ensemble learning framework (PT-ELF) is proposed in this paper. The major procedures of this study consist of three steps. First, the missing health states in the training dataset are supplemented by another dataset. Second, since the training dataset is drawn from two different distributions, a partial transfer mechanism is explored to train a weak global classifier and two partial domain adaptation classifiers. Third, a particular ensemble strategy combines these classifiers with different classification ranges and capabilities to obtain the final diagnosis result. Two case studies are used to validate our method. Results indicate that our method can provide robust diagnosis results based on an incomplete source domain under variable working conditions.

Keywords: partial transfer learning; ensemble strategy; fault diagnosis; deep adversarial convolutional neural network

Citation: Mao, G.; Zhang, Z.; Jia, S.; Noman, K.; Li, Y. Partial Transfer Ensemble Learning Framework: A Method for Intelligent Diagnosis of Rotating Machinery Based on an Incomplete Source Domain. *Sensors* **2022**, *22*, 2579. <https://doi.org/10.3390/s22072579>

Academic Editor: Mohammad Noori

Received: 8 March 2022

Accepted: 25 March 2022

Published: 28 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rotating components play a significant role in system performance and are widely applied in engineering machinery such as aerobots, engines, and gearbox systems [1,2]. The failure of rotating components may cause unexpected downtime and economic losses. Therefore, it is crucial to precisely identify and detect the fault states of rotating machinery [3]. Recently, intelligent fault diagnosis has become a hotspot because it can analyze vast amounts of measured data and provide intuitionistic diagnosis results [4].

Intelligent fault diagnosis has received a lot of attention in recent years from both industrial engineers and academic researchers and has accomplished remarkable achievements [5]. For example, shallow machine learning techniques such as support vector machine (SVM) [6] and random forest (RF) [7] have been studied. Deep learning methods have been researched that can adaptively extract the fault features hidden in a collected signal, such as recurrent neural network (RNN) [8], convolutional neural network (CNN) [9], and stack autoencoder (SAE) [10]. In addition, some variant models are being studied, such as dilated CNN [11], CNN with capsule network [12], and multiscale CNN [13]. However, the existing methods are developed based on statistics, which assume that adequate labeled samples are obtainable to train the models. In addition, these methods require the data distribution of training and testing to be identical [14]. In actual industry settings, obtaining a large amount of labeled data is unrealistic. Even if the labeled data can be acquired, the aforementioned methods may fail to recognize the unlabeled data collected

from another machine or under different working conditions due to the inconsistent data distribution [15].

The proposal of transfer learning aims to solve this problem by promoting models trained by labeled data from a relevant domain to the target fields [16]. The implementation of transfer learning for machine fault diagnosis mainly includes two scenarios: (1) A few target-domain-labeled data are available but are insufficient to support the model training. Qian et al. [17] implemented bearing fault diagnosis under diverse working conditions by transferring the parameters of SAE. Chen et al. [18] studied the use of transferable CNN to recognize the fault states of rotary machinery by pre-training a 1D-CNN using the source data and fine-tuning it with the limited labeled samples in the target domain. (2) There are no available labeled target data to participate in the model training process. One solution is to add a domain adaptation term to the loss function, such as the Maximum Mean Discrepancy (MMD) [4,19,20], Wasserstein distance [21]. Another solution is to implement the transfer learning by use of an adversarial network, in which case a feature extractor aims to extract domain-insensitive features from the target and source domains by adversarial training [22–24].

The existing cross-domain fault diagnosis methods can obtain superior results in the target domain, but the precondition lies in the assumption that the health states in the target domain are identifiable with the source domain. However, given the variation of operations and unpredictability of the fault states, it is difficult to guarantee that the current or future fault states have all been learned in the training phase. Therefore, the source training dataset is usually incomplete, and there are some additional failure states in the target domain. This causes negative transfer and misclassification in the testing stage. These private failure state data can be collected from another component, but the working conditions, such as speed, load, and frequency, are completely different from the source domain and target test data. Figure 1 shows an example of such a situation. Dataset *A* is collected from bearing 1 and contains five health states. However, during the test, more fault states appeared due to the change in working conditions, resulting in seven health states. The data for the two missing health states can be supplemented from dataset *B*. Dataset *B* is collected from bearing 2 and includes four health states total. So, the data source domain discrepancy between *A* and *B* also needs to be taken into consideration; this creates some difficulties for the implementation of transfer learning diagnostic methods.

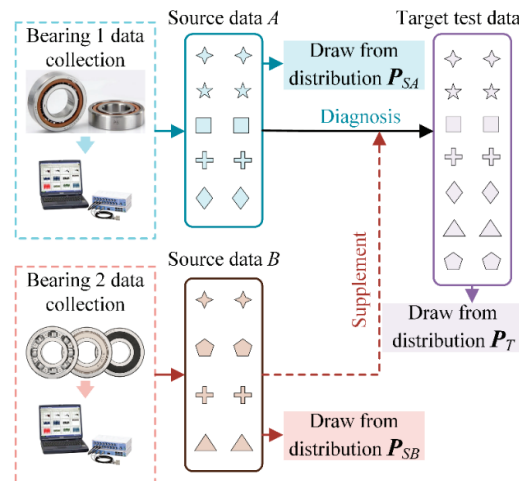


Figure 1. Example of the situation of fault diagnosis with new health states.

This research studies a partial transfer ensemble learning framework (PT-ELF) to solve the above problem. First, two incomplete source domain datasets collected from

different components or under different working conditions are defined. Note that neither of them contains all the health states present in the target domain data. They are used to form a complete dataset in which all the health states are included. Then, a weak global classifier based on the complete dataset and two partially strong classifiers based on the deep adversarial network are established. Finally, since the classification ability and classification range of classifiers differ, a particular ensemble strategy is designed to combine these two strong partial classifiers and the weak global classifier, resulting in the final diagnostic results. The main contributions of this research are summarized as follows:

- (1) A partial transfer ensemble learning framework is designed to diagnose the fault with incomplete training datasets under various conditions;
- (2) To incorporate the classification ability of multiple classifiers into the PT-ELF model, a particular ensemble strategy is designed to combine a weak global classifier and two partial domain adaptation classifiers;
- (3) Two case studies using rotor bearing test bench data and motor bearing data are performed to validate and demonstrate the superiority of the proposed method.

The rest of this article is arranged as follows: Section 2 presents the basic theories. The details of the proposed PT-ELF are given in Section 3. Section 4 validates the proposed method and analyzes the results. Finally, the conclusion in Section 5 brings the study to a close.

2. Basic Theory

2.1. Convolutional Neural Network

A standard CNN usually includes convolution, pooling, fully connected, and output layers. In addition, batch normalization operation is usually used in CNN [25]. A convolution layer is combined with a pooling layer to form a convolution block, and a deep architecture is built from several such blocks. A Softmax Regression layer usually serves as the last layer and performs regression or classification [26]. In a convolutional layer, the local receptive is adopted, in which only part of the input sample points connect to each node. This operation rapidly decreases the number of parameters and the model complexity. To identify the local features throughout the input sample, weights and biases are shared between the hidden neurons in one convolutional layer [27]. The process in the convolutional layer can be expressed as:

$$\mathbf{z}_n^l = \sum_k \mathbf{x}_k^{l-1} * \mathbf{w}_n^l + \mathbf{b}_n^l \quad (1)$$

where \mathbf{x}_k^{l-1} is the k -th node in $l - 1$ layer. $*$ represents the convolution operation. \mathbf{w}_n^l and \mathbf{b}_n^l represent the weight and the corresponding bias. Additionally, the activation function $\varphi(\bullet)$ is given to transform the convolution layers nonlinearly, which can be denoted as:

$$\mathbf{c}_n^l = \varphi(\mathbf{z}_n^l) \quad (2)$$

where \mathbf{c}_n^l represents the k -th nonlinear feature value in $l - 1$ layer. Sigmoid and ReLU activation functions are commonly used in CNN. Sigmoid can normalize the input data to between 0 and 1. ReLU can enhance the efficiency of the model training and decrease the risk of gradient disappearance [28].

In a pooling layer, the down-sampling operation can decrease the dimension of the features and enhance their robustness. Mathematically, a maximum pooling operation is defined as:

$$p_{oj} = \max_{i \in m_j} \{ \mathbf{c}_j(i) \} \quad (3)$$

where c_j represents the j -th location, and the po_j is the output of the pooling. For classification tasks, after several convolution blocks and fully connected layers, the Softmax function is usually utilized to predict categories. The loss objective function can be expressed as:

$$H(r, p) = -\sum_i r_i \log(p_i) \quad (4)$$

where p represents the output probability, and r corresponds to the actual labels.

2.2. Deep Adversarial Convolutional Neural Network

Generally, a deep adversarial convolutional neural network (DACNN) consists of a feature extractor G_f , a domain discriminator G_d , and a classifier G_y [29–31]. The feature extractor, namely several convolution blocks, serves as a contestant in the DACNN. It can be expressed as $G_f = G_f(x, \theta_f)$, which indicates that the features are extracted from the input sample x with parameters θ_f . In addition, a discriminator (binary classifier) is treated as the opponent, which is expressed as $G_d = G_d(G_f(x), \theta_d)$. Input the source and target samples into the feature extractor, and the output features are further distinguished by the discriminator G_d . The binary cross-entropy loss is taken as an objective function, which is described as:

$$L(G_d(G_f(x_i)), d_i) = d_i \log \frac{1}{G_d(G_f(x_i))} + (1 - d_i) \times \log \frac{1}{1 - G_d(G_f(x_i))} \quad (5)$$

where d_i denotes the binary variable for x_i . Through the adversarial training between two parts, the feature extractor G_f tends to extract the common features from the two types of data and makes it hard to differentiate 0 or 1 as the discriminator. Hence, the model can perform well on both the source and target datasets. The loss function is expressed as:

$$E(\theta_f, \theta_d) = -\left(\frac{1}{n} \sum_{i=1}^n L_d^i(\theta_f, \theta_d) + \frac{1}{N-n} \sum_{i=n+1}^N L_d^i(\theta_f, \theta_d)\right) \quad (6)$$

where n and $N - n$ represent the sample number of the source and target domain.

Additionally, all of the labeled samples should be supervised during training to ensure the accuracy of the diagnosis in the adversarial procedure. Thus, a classifier is established and is expressed as $G_y = G_y(G_f(x), \theta_y) : R^D \rightarrow R^L$ with parameters θ_y , in which L is the number of classes. The cross-entropy loss is applied in the Softmax function and is described as:

$$L(G_y(G_f(x_i)), y_i) = \log \frac{1}{G_y(G_f(x_i))_{y_i}} \quad (7)$$

Adding Equation (7) to the objective function (6), the optimization objective can be expressed as:

$$E(\theta_f, \theta_y, \theta_d) = \frac{1}{n} \sum_{i=1}^n L_y^i(\theta_f, \theta_y) - \lambda \left(\frac{1}{n} \sum_{i=1}^n L_d^i(\theta_f, \theta_d) + \frac{1}{N-n} \sum_{i=n+1}^N L_d^i(\theta_f, \theta_d) \right) \quad (8)$$

where $L_y^i(\theta_f, \theta_y) = L(G_y(G_f(x_i)), y_i)$ and λ is a non-negative hyper-parameter trade-off for the losses of the discriminator. In the whole training procedure of the DACNN, the optimization parameters $\theta_f, \theta_y, \theta_d$ can be obtained by:

$$(\hat{\theta}_f, \hat{\theta}_y) = \operatorname{argmax}_{\theta_f, \theta_y} E(\theta_f, \theta_y, \hat{\theta}_d) \quad (9)$$

$$\hat{\theta}_d = \operatorname{argmax}_{\theta_d} E(\hat{\theta}_f, \hat{\theta}_y, \theta_d) \quad (10)$$

The flowchart of the DACNN is displayed in Figure 2. By optimizing Equations (9) and (10), the DACNN tends to train a feature extractor G_f that can extract suitable representations from input samples that can be classified accurately by the classifier G_y but weakens the ability of the discriminator G_d to differentiate which domain this representation is from. In the phases of testing, the domain-insensitive features are extracted by the feature extractor G_f and fed into the health state classifier G_y to identify the states immediately.

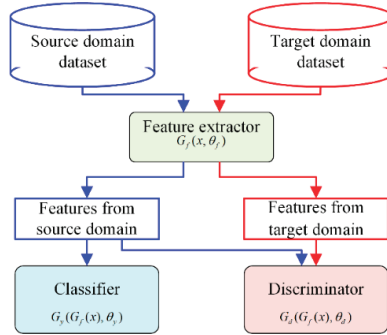


Figure 2. The schematic of the DACNN.

3. The Proposed Method

This section describes the proposed method in detail. It mainly includes problem formulation, the training of the three classifiers, and the classifiers’ ensemble.

3.1. Problem Formulation

Before implementing the proposed method, two incomplete source domain datasets A and B are defined as shown in Figure 3. The source dataset $A = \{(x_i^A, y_i^A)\}_{i=1}^{n_A}$ of n_A labels instances associated with $|D_A|$ classes and is drawn from distribution P_{SA} . The source dataset $B = \{(x_i^B, y_i^B)\}_{i=1}^{n_B}$ of n_B labels instances associated with $|D_B|$ classes collected from another same-type component and is drawn from distribution P_{SB} . The class label spaces of A and B are denoted as D_A and D_B , respectively. The collection of different components results in variations in the operating conditions (such as load, speed, etc.) in a real industrial environment; this means that $P_{SA} \neq P_{SB}$. In addition, there must be some shared health states contained in both source dataset A and source dataset B , which are denoted as $D = D_A \cap D_B$ and shown in Figure 3. $\hat{D}_A = D_A \setminus D_B$ denotes the private label space of the A and $\hat{D}_B = D_B \setminus D_A$ denotes the private label sets of B .

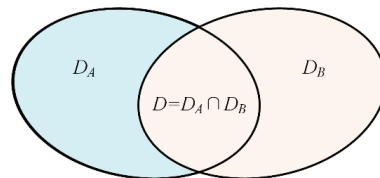


Figure 3. Two different source domain datasets.

However, in the testing stage of the actual machine fault diagnosis scenario, all possible health states may appear. Therefore, the target domain dataset includes all health states; it can be expressed as $T = \{(x_i^T)\}_{i=1}^{n_T}$ of n_T unlabeled instances associated with $|D_T|$ classes drawn from distribution P_T . The D_T represents the label sets of the target domain and $D_T = D_A \cup D_B$. In addition, the target domain distribution P_T is different in source domain distributions P_{SA} and P_{SB} .

This paper aims to establish a fault diagnosis model to realize fault diagnosis based on incomplete source training data under different operating conditions.

3.2. Classifier Training

This section describes the training procedure for the three classifiers (weak classifier C_W , classifier C_A , and classifier C_B) concretely.

First, a complete dataset C that contains all of the classes can be formed based on the incomplete source datasets A and B , as shown in Figure 4. In the complete dataset C , the sample in label space \hat{D}_A is from source dataset A , and the sample in label space \hat{D}_B is from source dataset B . For the samples in shared label space D , a portion of them come from A , and the rest come from B . Thus, the label space of dataset C is the same as T , and it includes $|D_T|$ health states. Second, a standard CNN classifier C_W is trained using the complete dataset C . However, since the source domain datasets A and B are collected under various work conditions, the samples in the dataset C are drawn from two types of distributions. In addition, the data distribution in the testing set P_T is different in P_{SA} and in P_{SB} . Therefore, the classifier C_W has poor classification ability for the target domain data. However, the classifier C_W has the ability to classify all health states.

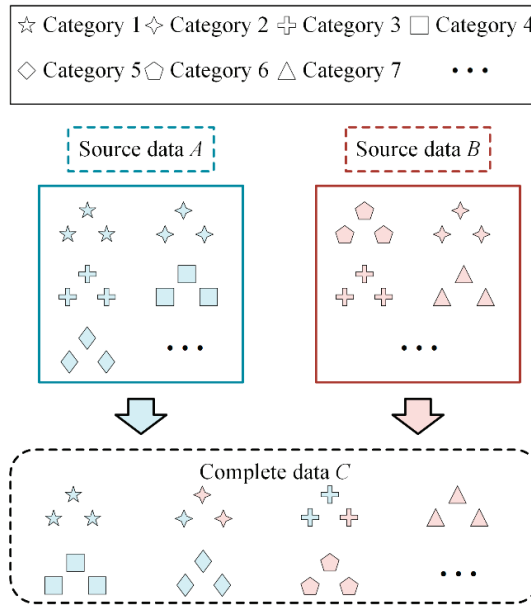


Figure 4. The process of forming a complete dataset C .

After the weak classifier C_W is obtained, the test samples from the target domain $T = \{(x_i^T)\}_{i=1}^{n_T}$ of n_T unlabeled instances associated with $|D_T|$ classes are classified, and the result is served as a pseudo-label to participate in the subsequent training. Target domain samples whose pseudo-label is in D_A are obtained to construct the target domain training set A_T . The samples whose pseudo-label is in D_B are obtained to construct the target domain training set B_T . Thus, the datasets A and A_T have the same label space D_A , and the datasets B and B_T have the same label space D_B .

Dataset A and A_T have the same health states but draw from different distributions. So, a DACNN model can be trained using the datasets A and A_T . A feature extractor and a classifier in this DACNN are combined to form a block, which is taken as classifier C_A . The classifier C_A is constructed by a DACNN using domain adaptation techniques, so that it has a strong classification ability for the unlabeled target domain dataset. However, the classification range of strong classifier C_A is limited to $|D_A|$ classes. After the training of classifier C_A is completed, classifier C_B is trained in the same way. Similarly, the classification range of C_B is limited to $|D_B|$ classes.

In the implementation process of the DACNN, the SELU activation function is used in convolutional layers; its mathematical expression is expressed as Equation (11):

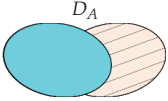
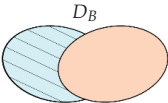
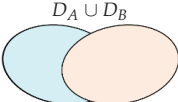
$$\text{SELU}(x) = \lambda \begin{cases} \alpha e^x - \alpha & (x \leq 0) \\ x & (x > 0) \end{cases} \tag{11}$$

where the value of α is 1.6732, and the value of λ is 1.0507. The SELU activation function can automatically normalize the sample distribution to 0 mean value and unit variance to avoid the gradient exploding or disappearing. The activation function used in the fully connected layer in the state classifier and domain discriminator is ReLU, and it is expressed as Equation (12):

$$\text{ReLU}(x) = \begin{cases} 0 & (x \leq 0) \\ x & (x > 0) \end{cases} \tag{12}$$

In this way, three well-trained classifiers are achieved, including one weak global classifier C_W , one strong partial classifier C_A , and one strong partial classifier C_B . The details of the three classifiers are listed in Table 1.

Table 1. Classification range and ability of the three classifiers.

| Classifiers | Range of Classification | Ability of Classification |
|-------------|--|---------------------------|
| C_A |  | Strong |
| C_B |  | Strong |
| C_W |  | Weak |

3.3. Classifiers' Ensemble

After the three classifiers are obtained, this section designs a particular ensemble strategy to combine their results. The procedure for the ensemble strategy is presented in Figure 5.

After inputting a testing sample x into the three classifiers, the classification result y_W , y_A , and y_B can be output from the three classifiers, which can be expressed as:

$$\begin{cases} y_W = C_W(x) \\ y_A = C_A(x) \\ y_B = C_B(x) \end{cases} \tag{13}$$

If $y_W = y_A \parallel y_W = y_B \parallel y_A = y_B$ is satisfied, the final result y can be obtained by a majority voting strategy immediately. Otherwise, it means that the results of the three classifiers are different from each other. In such cases, because the classifier C_W is a global classifier, y_W is served as the reference standard. If $y_W \in D_A$ is satisfied, that means that the actual label of x may be in D_A . In this range, the classifier C_A has perfect classification ability, and thus y_A is served as the final result. Similarly, if $y_W \in D_B$ is satisfied, y_B is served as the final result. However, if $y_W \in D$ is satisfied, both the classifiers C_A and C_B

have good classification ability in this shared range. In this case, y is determined according to the output probability p in the Softmax layer of classifiers, and it can be expressed as:

$$\begin{cases} y = y_A & \text{if } p_A = \max(p_A, p_B, p_W) \\ y = y_B & \text{if } p_B = \max(p_A, p_B, p_W) \\ y = y_W & \text{if } p_W = \max(p_A, p_B, p_W) \end{cases} \quad (14)$$

where the p_A , p_B , and p_W represent the Softmax output probability of classifiers C_A , C_B , and C_W ; $\max(\cdot)$ is the maximum function.

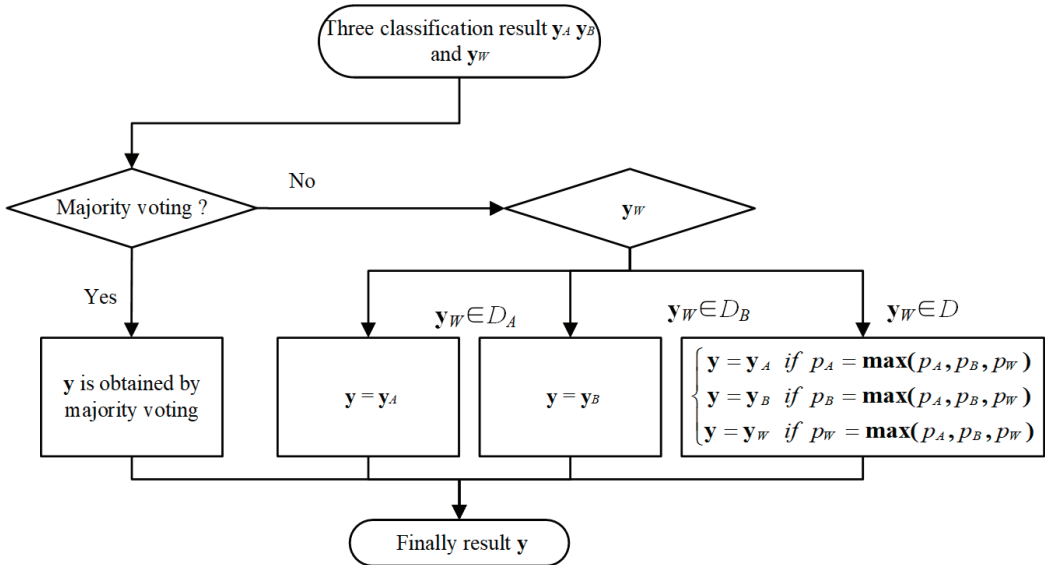


Figure 5. The flowchart of the classifiers' ensemble.

3.4. Architecture of the Proposed Method

The architecture of our method for fault diagnosis is presented in Figure 6, and the process is summarized below.

- (1) Collect original vibration signals from different components or under different working conditions, and convert them into frequency domain signals for subsequent model training;
- (2) Construct a complete dataset by combing these incomplete datasets, and train a weak global classifier CNN;
- (3) Classify the target domain data using the weak classifier to obtain the two target domain training sets;
- (4) Train two DACNN models using two source datasets and target domain training sets to construct two strong partial classifiers;
- (5) Design a particular ensemble strategy to combine the three classifiers and obtain the final classification results.

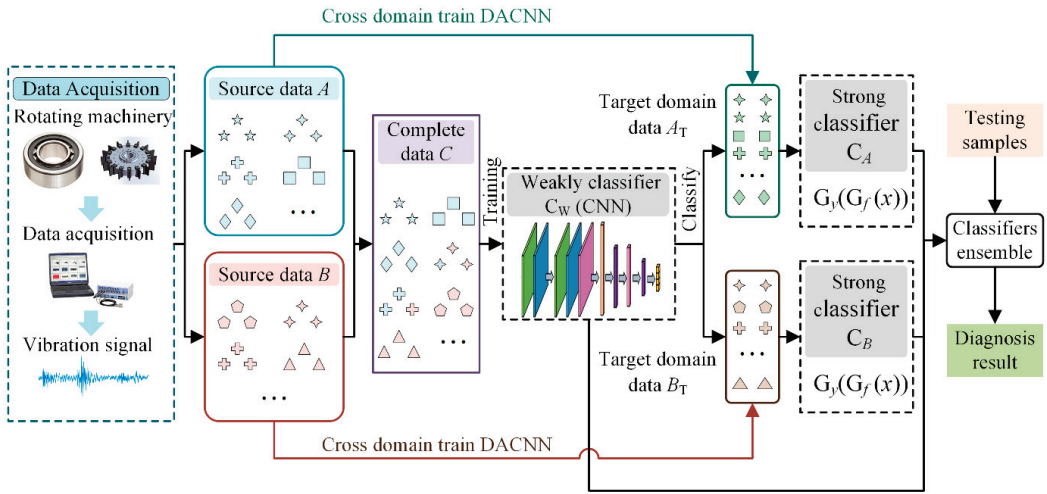


Figure 6. The overall procedures of the proposed method.

4. Experimental Verification

To validate the effectiveness of the proposed PT-ELF method, rotor and rolling bearing experiments are designed. Note that the code for the proposed method is written in Pytorch 1.2 and runs with 16G RAM and a Core i5 10400F CPU.

4.1. Case 1

4.1.1. Rotor Experiment

Case 1 adopts the rotor dataset from Northwestern Polytechnical University. As shown in Figure 7a, the experimental system is composed of a three-phase variable frequency motor, single-span rotor shafting, torque speed sensor, rolling bearing seat, shafting load plate, rubbing mounting bracket, platform bottom plate, radial loading device, coupling, system control cabinet, and fault suite. A displacement sensor is mounted on the rotor test bench to collect vertical vibration signals under a health state and six different fault states as shown in Figure 8, and the sample frequency is 10,240 Hz. Figure 7b depicts the sensor and single-span rotor shaft layout. The structural components are listed in Table 2.

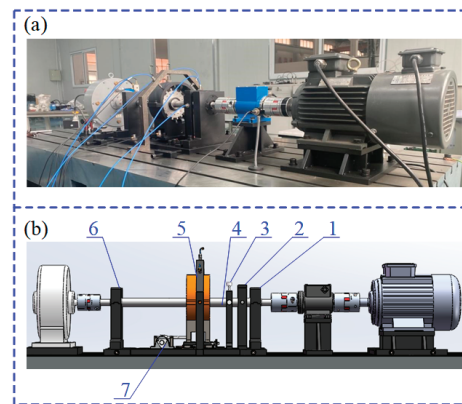


Figure 7. The rolling bearing experiment system: (a) the experimental test rig; (b) the layout of the test rig.

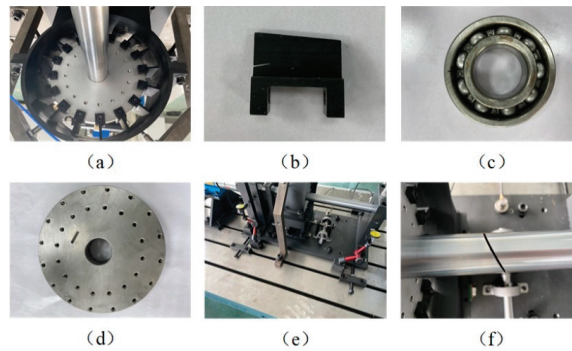


Figure 8. Six different fault states: (a) full annular rub; (b) blade crack; (c) bearing fault; (d) blisk crack; (e) Shaft coupling fault; (f) Shaft crack.

Table 2. The structural components of the single-span rotor shafting.

| No | Component |
|----|--|
| 1 | Support bearing pedestal |
| 2 | Displacement sensor bracket |
| 3 | Friction assembly and bracket |
| 4 | Shaft |
| 5 | Casing friction support and blade disc |
| 6 | Test bearing pedestal |
| 7 | Worm gear and worm |

The rotor vibration data are collected under three working load conditions of 0%, 20%, and 40%. As detailed in Table 3, for each load, data from seven health states (including a health state and six fault states) are used. The data in each state are divided into 300 samples, with 80 randomly selected as tests and the remaining 220 used to train. Each sample, each consisting of 800 data points, is used to verify the method proposed in this paper. Figure 9 shows the waveform of the original displacement signal and the spectral distributions of each health state under 0% load. The left shows the spectral signal, and the right shows the corresponding spectrum. The signals have a large amplitude of around 10–30 Hz, showing relatively similar characteristics, which makes it hard to recognize the health states.

Table 3. Seven health states of the rotor.

| Label | Health States | The Number of Training/Testing Samples |
|-------|-------------------------------|--|
| 0 | Health | 220/80 |
| 1 | Full annular rub | 220/80 |
| 2 | Blade crack and bearing fault | 220/80 |
| 3 | Blade crack | 220/80 |
| 4 | Blisk crack | 220/80 |
| 5 | Shaft coupling fault | 220/80 |
| 6 | Shaft crack | 220/80 |

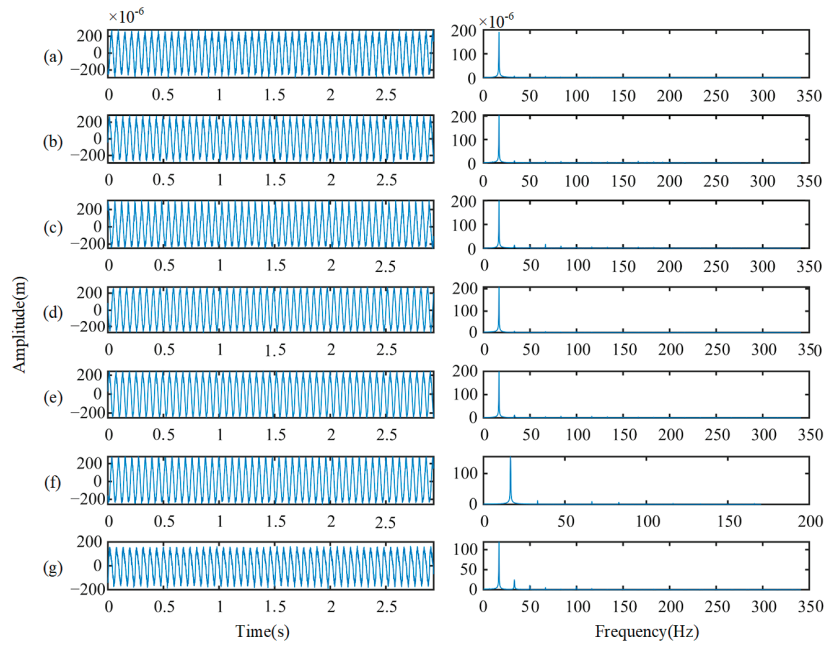


Figure 9. Original displacement signals and spectral distributions: (a) health; (b) full annular rub; (c) blade crack and bearing fault; (d) blade crack; (e) blisk crack; (f) shaft coupling fault; (g) shaft crack.

4.1.2. Results and Discussion

In this case study, two incomplete source datasets are constructed, as shown in Table 4. The source dataset *A* contains five kinds of health states (states 1–5), and the source dataset *B* contains four kinds of health states (states 4–7).

Table 4. Distribution of health states in two source domains and one target domain.

| States | Source Domain Dataset A Data | Source Domain Dataset A Labels | Source Domain Dataset B Data | Source Domain Dataset B Labels | Target Domain Data Data | Target Domain Data Labels |
|--------|---------------------------------|-----------------------------------|---------------------------------|-----------------------------------|----------------------------|------------------------------|
| 1 | ✓ | ✓ | | | ✓ | |
| 2 | ✓ | ✓ | | | ✓ | |
| 3 | ✓ | ✓ | | | ✓ | |
| 4 | ✓ | ✓ | ✓ | ✓ | ✓ | |
| 5 | ✓ | ✓ | ✓ | ✓ | ✓ | |
| 6 | | | ✓ | ✓ | ✓ | |
| 7 | | | ✓ | ✓ | ✓ | |

First, the source domain datasets *A* and *B* are mixed to form a training set that contains all health states, which is used to train a weak classifier C_W . The classifier C_W has a classification ability for all of the health states (seven kinds of health states). Second, according to the classification results (the pseudo-label) of the weak classifier C_W on the target domain samples, two transfer models based on a DACNN are trained. They are transferred from source domain dataset *A* and source domain dataset *B* to the target domain. Thus, two strong classifiers C_A and C_B are trained. Finally, after classifying a test sample by the classifiers C_A , C_B , and C_W , three results are obtained and fused by the proposed ensemble strategy described in Section 3.3.

To demonstrate that our method is applicable to various operating conditions, five test scenarios (test scenarios A1–E1) are designed to test the proposed method. As listed in Table 5, the source domain *A*, source domain *B*, and target domain are served by the collected dataset under different loads. In source dataset *A*, only five kinds of labeled samples in states 1–5 are available. Similar to source domain *A*, in source dataset *B*, only four types of labeled samples in states 4–7 are available. The test data in the target domain contain all seven kinds of unlabeled samples in states 1–7.

Table 5. Five different test scenarios.

| Test Scenarios | Source Dataset A | Source Dataset B | Target Data |
|----------------|-----------------------|-----------------------|-----------------------|
| A1 | Load 0% (states 1–5) | Load 20% (states 4–7) | Load 40% (states 1–7) |
| B1 | Load 0% (states 1–5) | Load 40% (states 4–7) | Load 20% (states 1–7) |
| C1 | Load 40% (states 1–5) | Load 20% (states 4–7) | Load 0% (states 1–7) |
| D1 | Load 20% (states 1–5) | Load 0% (states 4–7) | Load 40% (states 1–7) |
| E1 | Load 40% (states 1–5) | Load 0% (states 4–7) | Load 20% (states 1–7) |

The accuracies of the three classifiers (two strong partial classifiers and a weak global classifier) and the proposed PT-ELF method in the five test scenarios are listed in Table 6, and a bar diagram is shown in Figure 10a. Note that the accuracy of C_A is tested by states 1–5, and the accuracy of C_B is tested using states 4–7. The result of the weak classifier C_W and the ensemble result are tested using target domain test data that contain all of the health states (states 1–7).

Table 6. Results of different classifiers.

| Test Scenarios | Strong Classifier C_A | Strong Classifier C_B | Weak Classifier C_W | Proposed Method |
|----------------|-------------------------|-------------------------|-----------------------|-----------------|
| A1 | 92.14% | 98.58% | 85.89% | 91.08% |
| B1 | 95.15% | 98.28% | 92.14% | 95.41% |
| C1 | 81.50% | 99.68% | 78.03% | 83.75% |
| D1 | 99.50% | 91.07% | 89.07% | 92.89% |
| E1 | 98.14% | 96.56% | 87.50% | 90.48% |
| Average | 93.29% | 96.83% | 86.52% | 90.73% |

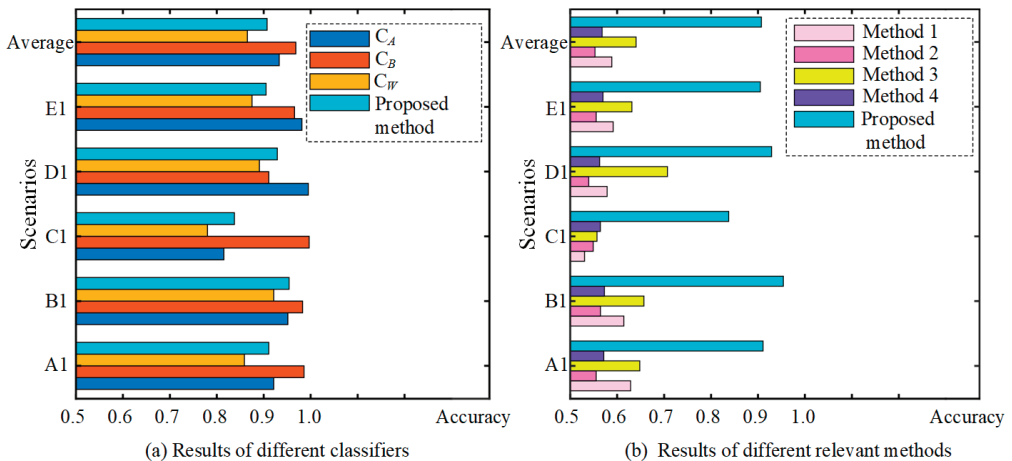


Figure 10. The result diagram for different classifiers.

It can be seen from Table 6 that the two strong classifiers C_A and C_B have high accuracy in the corresponding classification range, with averages of 93.29% and 96.83%. On the one hand, this is because the two strong classifiers are trained by a domain adversarial network DACNN, which can extract domain-insensitive features to classify. On the other hand, they are just tested by partial health states. The result of the weak classifier C_W is relatively poor, with an average accuracy of 86.52%. This is because the data of the target domain and two source domains are not uniformly distributed, leading to the decrease in classification performance.

Out of five test scenarios, the result in scenario B1 is the highest at 95.41%; scenario C1 has the lowest accuracy at 83.75%, and the average is 90.73%. This is significantly higher than the weak classifier C_W , and maintains a high classification accuracy. This is because the proposed ensemble strategy can cause the test sample to be classified by the corresponding strong classifier as far as possible. It indicates that our method can still achieve good results even under incomplete training data.

In addition, to prove the superiority of our method, relevant methods for a CNN and a DACNN, trained by source dataset A and source dataset B , respectively, are used as comparison methods (Method 1–4). The result is listed in Table 7, and a bar diagram of the various methods is shown in Figure 10b. It can be observed that the average accuracies of the CNN trained by source domains A and B are 58.87% and 55.27%, respectively. The average accuracies of the DACNN trained by source domains A and B are 64.02% and 56.79%, respectively, which are significantly higher than the accuracy of the CNN. This is because the DACNN can extract domain-insensitive features using adversarial training; this restrains the model's performance decrease caused by a distribution discrepancy and further improves the accuracy of the model in the target domain. However, since the source domain A is incomplete, a model (CNN or DACNN) trained by source dataset A is unable to classify the testing samples whose actual label is in \hat{D}_B (states 6–7). Similarly, a model (CNN or DACNN) trained by source dataset B is unable to classify the testing samples whose actual label is in \hat{D}_A (states 1–3); therefore, the results of methods 1–4 are poor compared to our method. The average accuracy of our method is as high as 90.73%, which indicates that the proposed method has good classification ability for all health states presented in the testing dataset in the target domains.

Table 7. Results of different methods.

| Test Scenarios | Method 1 (CNN Trained by Source A) | Method 2 (CNN Trained by Source B) | Method 3 (DACNN Trained by Source A) | Method 4 (DACNN Trained by Source B) | The Proposed Method |
|----------------|------------------------------------|------------------------------------|--------------------------------------|--------------------------------------|---------------------|
| A1 | 62.86% | 55.54% | 64.82% | 57.14% | 91.08% |
| B1 | 61.43% | 56.43% | 65.71% | 57.28% | 95.41% |
| C1 | 53.04% | 54.89% | 55.71% | 56.42% | 83.75% |
| D1 | 57.86% | 53.93% | 70.71% | 56.25% | 92.89% |
| E1 | 59.14% | 55.54% | 63.14% | 56.96% | 90.48% |
| Average | 58.87% | 55.27% | 64.02% | 56.79% | 90.73% |

4.2. Case 2

4.2.1. Rolling Bearing Experiment

The rolling bearing vibration data utilized in case 2 are from Case Western Reserve University [32]. As shown in Figure 11, the setup mainly consists of a loading motor, an induction motor, and testing bearings. The vibration signals used in this case are collected by an accelerometer installed near the drive end. As listed in Table 8, the vibration signals were collected under four different loads (Load 1–Load 4). Each fault was artificially implanted into the bearings with different severity levels from 0.007 to 0.028 inches in diameter (1 inch = 25.4 mm). The details of the test bearing are listed in Table 9.

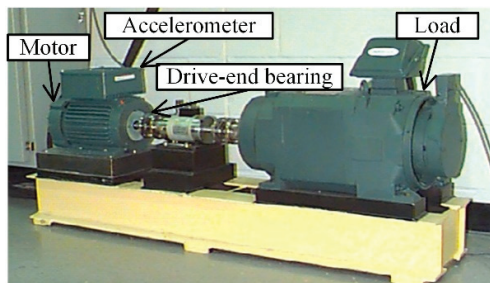


Figure 11. The experiment setup of rolling bearing.

Table 8. Four different loads.

| Loads | Values |
|--------|----------------|
| Load 1 | 1797 rpm, 0 hp |
| Load 2 | 1772 rpm, 1 hp |
| Load 3 | 1750 rpm, 2 hp |
| Load 4 | 1750 rpm, 3 hp |

Table 9. Details of the test bearing.

| Parameters | Values |
|---------------------|-------------------------|
| Type | 6205-2RS JEM SKF |
| The number of balls | 9 |
| Pitch diameter | 1.537 inches |
| Ball diameter | 0.3126 inches |
| Sampling frequency | 12 (kHz) |
| Motor speed | 1797/1772/1750/1730 rpm |

The vibration data collected under four different loads are used to test the proposed method. Each of them includes 12 health states, which include different failure locations (shown in Figure 12), different failure orientations, and different failure severities. As detailed in Table 10, each health state contains 300 samples, which consist of 400 continuous data points. At random, 200 samples are selected to train, and the remaining 100 are used to test. The raw vibration is under 1797 rpm (0 hp) (in the left column), and the corresponding spectral distributions (in the right column) are shown in Figure 13. In terms of raw vibration signals, the health state vibration amplitude is relatively small (Figure 13a). The fault signals (Figure 13b–i) have an obvious impact. The spectral distribution contains the fault frequency and the bearing natural frequency. In addition to the health signals, the other fault vibration signals have a higher amplitude of around 3–4 kHz. It is still very unrealizable to accurately distinguish the fault location, dimension, and orientation across different working conditions with new fault states.

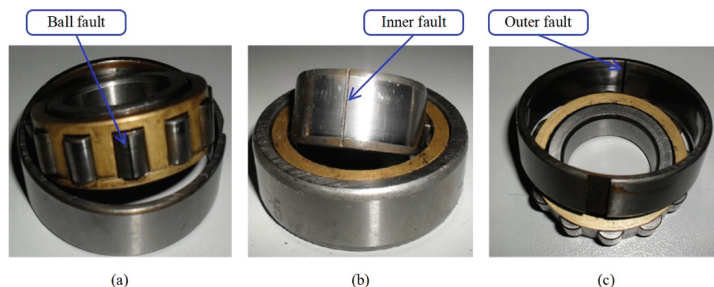


Figure 12. The faults of bearing in three locations: (a) ball fault; (b) inner fault; (c) outer fault.

Table 10. The details of the 12 operating states.

| Labels | Failure Location | Failure Orientation | Failure Severities (Inches) | The Number of Testing/Training Samples |
|--------|------------------|---------------------|-----------------------------|--|
| 0 | Health | - | 0 | 100/200 |
| 1 | Rolling element | - | 0.007 | 100/200 |
| 2 | Rolling element | - | 0.014 | 100/200 |
| 3 | Rolling element | - | 0.021 | 100/200 |
| 4 | Inner race | - | 0.007 | 100/200 |
| 5 | Inner race | - | 0.021 | 100/200 |
| 6 | Inner race | - | 0.028 | 100/200 |
| 7 | Outer race | Center | 0.007 | 100/200 |
| 8 | Outer race | Vertical | 0.007 | 100/200 |
| 9 | Outer race | Center | 0.014 | 100/200 |
| 10 | Outer race | Center | 0.021 | 100/200 |
| 11 | Outer race | Vertical | 0.021 | 100/200 |

The proposed method mainly studies the case in which only partial health state labeled data are available in the source domain. To verify our method, we assume that source domain dataset *A* only contains eight kinds of fault state labeled data, while source domain dataset *B* contains seven kinds of labeled data. Among them, three categories overlap, as shown in Table 11. In addition, all target domain data are unlabeled; these data contain 12 kinds of health states.

Table 11. Distribution of health states in source and target data.

| States | Source Domain Dataset <i>A</i> | | Source Domain Dataset <i>B</i> | | Target Domain Data | |
|--------|--------------------------------|--------|--------------------------------|--------|--------------------|--------|
| | Data | Labels | Data | Labels | Data | Labels |
| 1 | ✓ | ✓ | | | ✓ | |
| 2 | ✓ | ✓ | | | ✓ | |
| 3 | ✓ | ✓ | | | ✓ | |
| 4 | ✓ | ✓ | | | ✓ | |
| 5 | ✓ | ✓ | | | ✓ | |
| 6 | ✓ | ✓ | ✓ | ✓ | ✓ | |
| 7 | ✓ | ✓ | ✓ | ✓ | ✓ | |
| 8 | ✓ | ✓ | ✓ | ✓ | ✓ | |
| 9 | | | ✓ | ✓ | ✓ | |
| 10 | | | ✓ | ✓ | ✓ | |
| 11 | | | ✓ | ✓ | ✓ | |
| 12 | | | ✓ | ✓ | ✓ | |

4.2.2. Results and Discussion

Similar to Case 1, the source datasets *A* and *B* are first mixed to form a training set containing all health states, and it is used to train the weak classifier C_W . Thus, C_W has a classification ability for all of the health states, but the classification ability is weak.

In the following step, two DACNN models are trained based on source domain datasets *A* and *B* to adapt target domain data. Then, two strong classifiers C_A and C_B can be obtained. In each DACNN, the feature extractor G_f contains two convolution blocks. Meanwhile, the classifier G_y contains a fully connected layer and output by a Softmax function. The $G_y(G_f(x))$ in the DACNN is taken as the classifier. Finally, three well-trained classifiers C_A , C_B , and C_W with different classification capabilities and classification ranges are integrated using the ensemble strategy introduced in Section 3.3 to obtain the final diagnosis result.

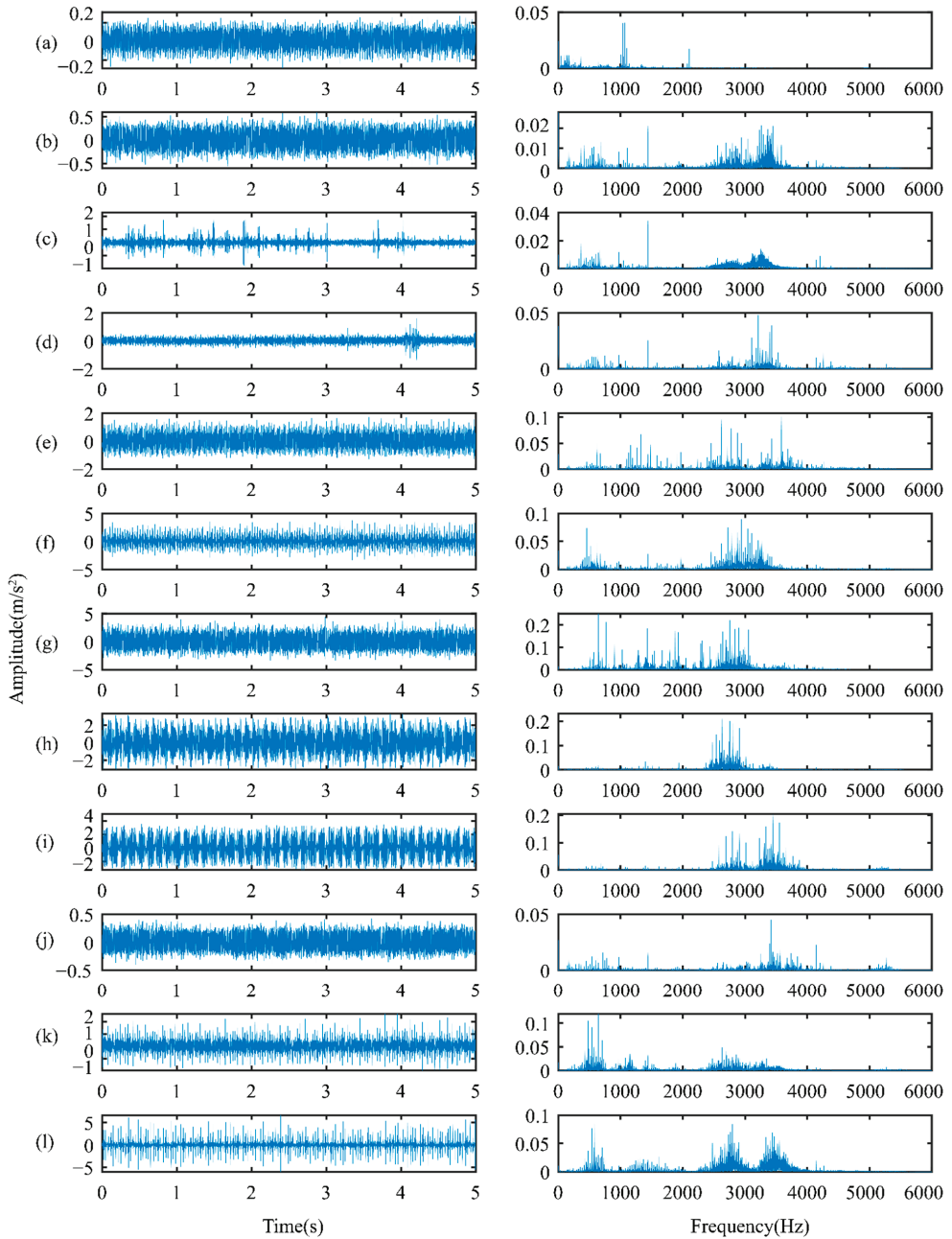


Figure 13. Waveform of raw signals and spectral distributions of the rolling bearing: (a) health; (b) rolling element failure (0.007); (c) rolling element failure (0.014); (d) rolling element failure (0.021); (e) inner race failure (0.007); (f) inner race failure (0.021); (g) inner race failure (0.028); (h) outer race failure (0.007 Center); (i) outer race failure (0.007 Vertical); (j) outer race failure (0.014 Center); (k) outer race failure (0.021 Center); (l) outer race failure (0.021 Vertical).

To demonstrate that our method is applicable to different working conditions, five test scenarios (test scenarios A2–E2) with incomplete data are used to test the proposed method, as shown in Table 12. In source dataset *A*, eight kinds of labeled samples in states 1–8 are available, and in source dataset *B*, seven kinds of labeled samples in states 6–12 are available. The target data, which contains 12 kinds of unlabeled samples in states 1–12, is used to test. In the five test scenarios, source domain datasets *A* and *B* and the target domain dataset are served by data collected under different loads. To indicate the superiority of our method, two conventional deep learning methods based on CNN (method 1 and method 2) and two transfer learning methods based on DACNN (method 3 and method 4) are used for comparison in five test scenarios; the results are listed in Table 13. Method 1 and method 3 are trained using source dataset *A*, and method 2 and method 4 are trained using source dataset *B*. In order to show the comparison results visually, the results bar diagram for different methods is shown in Figure 14.

Table 12. Five different test scenarios.

| Test Scenarios | Source Dataset A | Source Dataset B | Target Data |
|----------------|---------------------|----------------------|----------------------|
| A2 | Load 1 (states 1–8) | Load 2 (states 6–12) | Load 3 (states 1–12) |
| B2 | Load 3 (states 1–8) | Load 4 (states 6–12) | Load 1 (states 1–12) |
| C2 | Load 2 (states 1–8) | Load 3 (states 6–12) | Load 4 (states 1–12) |
| D2 | Load 1 (states 1–8) | Load 2 (states 6–12) | Load 4 (states 1–12) |
| E2 | Load 2 (states 1–8) | Load 3 (states 6–12) | Load 1 (states 1–12) |

Table 13. Results of different methods.

| Test Scenarios | Method 1 (CNN Trained Using Source Dataset A) | Method 2 (CNN Trained Using Source Dataset B) | Method 3 (DACNN Trained Using Source Dataset A) | Method 4 (DACNN Trained Using Source Dataset B) | The Proposed Method |
|----------------|--|--|--|--|------------------------|
| A2 | 63.17% | 57.13% | 65.75% | 58.33% | 98.08% |
| B2 | 60.50% | 58.25% | 65.83% | 58.08% | 95.41% |
| C2 | 66.50% | 58.08% | 66.67% | 58.33% | 99.66% |
| D2 | 66.08% | 58.14% | 66.58% | 58.33% | 99.25% |
| E2 | 65.08% | 56.08% | 66.25% | 57.17% | 95.83% |
| Average | 64.27% | 57.53% | 66.22% | 58.05% | 97.65% |

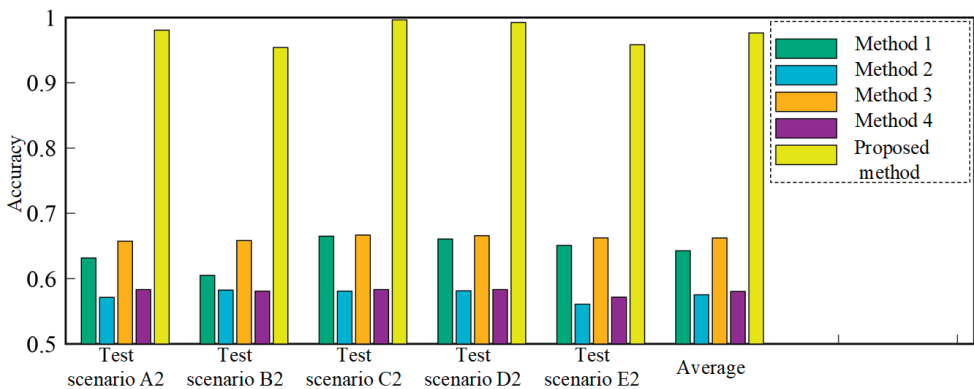


Figure 14. The results diagram for different methods.

As shown in Table 13 and Figure 14, the average accuracies of methods 1 and 2 are 64.27% and 57.53%, respectively. The average accuracies of method 3 and method 4, based on transfer learning, are 66.22% and 58.05%, respectively. This is because the

DACNN can solve the problem of cross-domain fault diagnosis well and enhances the recognition accuracy in the target domain. However, since the source datasets *A* and *B* are incomplete, neither of them contains all the health states presented in the testing data; the fault classification accuracy is still relatively low even if the transfer strategy is used. The accuracy of the method proposed can achieve 98.08%, 95.41%, 99.66%, 99.25%, and 95.83% in five test scenarios, respectively. Accuracy is the lowest in test scenario B2, but it can still remain at 95.41%. In test scenario C2, the classification accuracy is the highest at 99.66%. The comparison results demonstrate that the proposed PT-ELF method exhibits satisfactory cross-domain diagnostic ability with new health states.

5. Conclusions

This paper proposes a rotating machinery fault diagnosis method based on partial transfer learning and ensemble learning. Unlike other existing cross-domain diagnostic methods with the assumption of the same health states in the source and target domains, the proposed method can provide a reliable diagnosis result in the target domain even when the source domain is incomplete and only contains partial health states. As the core of the proposed method, partial transfer learning can avoid the problem induced by incomplete training data and train two classifiers with strong classification capabilities for partial categories. Then, a particular ensemble strategy is designed to combine the output of the three classifiers (a weak global classifier and two strong partial classifiers). The effectiveness of the proposed method is validated on a rotor experiment and a bearing experiment. After comparing with four related methods, results indicate that the proposed method can achieve superior performance and provide a reliable diagnosis result based on incomplete source domain under various working conditions.

In this preliminary study, the proposed method lies in the assumption that the missing health states in the source domain training set can be obtained from another dataset or another component. The unseen health states will be considered in our future research.

Author Contributions: Data curation, Z.Z.; Formal analysis, G.M.; Funding acquisition, Y.L.; Investigation, S.J.; Methodology, G.M.; Resources, Z.Z.; Software, S.J.; Validation, G.M. and K.N.; Writing—original draft, G.M.; Writing—review & editing, K.N. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research is supported by the National Natural Science Foundation of China under Grant 51805434 and 12172290 and Key Laboratory of Equipment Research Foundation under Grant 6142003190208.

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.

Conflicts of Interest: The authors declare that they have no known competing financial interest or personal relationship that relate to the work reported in this paper.

Abbreviations

| | |
|--------|---|
| CNN | convolutional neural network |
| DACNN | deep adversarial convolutional neural network |
| MMD | maximum mean discrepancy |
| PT-ELF | partial transfer ensemble learning framework |
| RF | random forest |
| RNN | recurrent neural network |
| SAE | stack autoencoder |
| SVM | support vector machine |

References

1. Yongbo, L.; Xiaoqiang, D.; Fangyi, W.; Xianzhi, W.; Huangchao, Y.J. Rotating machinery fault diagnosis based on convolutional neural network and infrared thermal imaging. *Chin. J. Aeronaut.* **2020**, *33*, 427–438.
2. Shao, H.; Jiang, H.; Zhang, X.; Niu, M. Rolling bearing fault diagnosis using an optimization deep belief network. *Meas. Sci. Technol.* **2015**, *26*, 115002. [[CrossRef](#)]
3. Haidong, S.; Hongkai, J.; Xingqiu, L.; Shuaipeng, W. Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine. *Knowl.-Based Syst.* **2018**, *140*, 1–14. [[CrossRef](#)]
4. Yang, B.; Lei, Y.; Jia, F.; Xing, S. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. *Mech. Syst. Signal Process.* **2019**, *122*, 692–706. [[CrossRef](#)]
5. 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](#)]
6. Zhang, X.; Chen, W.; Wang, B.; Chen, X.J. Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization. *Neurocomputing* **2015**, *167*, 260–279. [[CrossRef](#)]
7. Ma, K.; Ben-Arie, J. Compound exemplar based object detection by incremental random forest. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 2407–2412.
8. Liu, H.; Zhou, J.; Zheng, Y.; Jiang, W.; Zhang, Y.J. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Trans.* **2018**, *77*, 167–178. [[CrossRef](#)]
9. Zhang, W.; Li, C.; Peng, G.; Chen, Y.; Zhang, Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Signal Process.* **2018**, *100*, 439–453. [[CrossRef](#)]
10. Sun, W.; Shao, S.; Zhao, R.; Yan, R.; Zhang, X.; Chen, X.J. A sparse auto-encoder-based deep neural network approach for induction motor faults classification. *Measurement* **2016**, *89*, 171–178. [[CrossRef](#)]
11. Khan, M.A.; Kim, Y.-H.; Choo, J. Intelligent fault detection via dilated convolutional neural networks. In Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing, Shanghai, China, 15–17 January 2018; pp. 729–731.
12. Zhu, Z.; Peng, G.; Chen, Y.; Gao, H. A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis. *Neurocomputing* **2019**, *323*, 62–75. [[CrossRef](#)]
13. Zhu, J.; Chen, N.; Peng, W. Estimation of bearing remaining useful life based on multiscale convolutional neural network. *IEEE Trans. Ind. Electron.* **2019**, *66*, 3208–3216. [[CrossRef](#)]
14. Guo, L.; Lei, Y.; Xing, S.; Yan, T.; Li, N. Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines With Unlabeled Data. *IEEE Trans. Ind. Electron.* **2019**, *66*, 7316–7325. [[CrossRef](#)]
15. Han, T.; Liu, C.; Yang, W.; Jiang, D. A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults. *Knowl.-Based Syst.* **2019**, *165*, 474–487. [[CrossRef](#)]
16. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q.J. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2020**, *109*, 43–76. [[CrossRef](#)]
17. Qian, W.; Li, S.; Wang, J.; Xin, Y.; Ma, H. A New Deep Transfer Learning Network for Fault Diagnosis of Rotating Machine Under Variable Working Conditions. In Proceedings of the 2018 Prognostics and System Health Management Conference (PHM-Chongqing), Chongqing, China, 26–28 October 2018; pp. 1010–1016.
18. Chen, Z.; Gryllias, K.; Li, W. Intelligent Fault Diagnosis for Rotary Machinery Using Transferable Convolutional Neural Network. *IEEE Trans. Ind. Inform.* **2019**, *16*, 339–349. [[CrossRef](#)]
19. Wen, L.; Gao, L.; Li, X. A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Trans. Syst. Man Cybern. -Syst.* **2017**, *49*, 136–144. [[CrossRef](#)]
20. Xu, K.; Li, S.; Wang, J.; An, Z.; Qian, W.; Ma, H.J. A novel convolutional transfer feature discrimination network for imbalanced fault diagnosis under variable rotational speed. *Meas. Sci. Technol.* **2019**, *30*, 105107. [[CrossRef](#)]
21. Zhang, M.; Wang, D.; Lu, W.; Yang, J.; Li, Z.; Liang, B. A Deep Transfer Model With Wasserstein Distance Guided Multi-Adversarial Networks for Bearing Fault Diagnosis Under Different Working Conditions. *IEEE Access* **2019**, *7*, 65303–65318. [[CrossRef](#)]
22. Zhang, Z.; Li, X.; Wen, L.; Gao, L.; Gao, Y. Fault Diagnosis Using Unsupervised Transfer Learning Based on Adversarial Network. In Proceedings of the 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE), Vancouver, BC, Canada, 22–26 August 2019; pp. 305–310.
23. Zhang, B.; Li, W.; Hao, J.; Li, X.-L.; Zhang, M.J. Adversarial adaptive 1-D convolutional neural networks for bearing fault diagnosis under varying working condition. *arXiv* **2018**, arXiv:1805.00778.
24. Wang, B.; Shen, C.; Yu, C.; Yang, Y. Data Fused Motor Fault Identification Based on Adversarial Auto-Encoder. In Proceedings of the 2019 IEEE 10th International Symposium on Power Electronics for Distributed Generation Systems (PEDG), Xi'an, China, 3–6 June 2019; pp. 299–305.
25. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 7–9 July 2015.
26. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.J. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [[CrossRef](#)]
27. Shao, S.; McAleer, S.; Yan, R.; Baldi, P.J. Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Trans. Ind. Inform.* **2018**, *15*, 2446–2455. [[CrossRef](#)]

28. Jiao, J.; Zhao, M.; Lin, J.; Zhao, J.J. A multivariate encoder information based convolutional neural network for intelligent fault diagnosis of planetary gearboxes. *Knowl. Based Syst.* **2018**, *160*, 237–250. [[CrossRef](#)]
29. Wang, J.; Li, S.; Han, B.; An, Z.; Bao, H.; Ji, S. Generalization of Deep Neural Networks for Imbalanced Fault Classification of Machinery Using Generative Adversarial Networks. *IEEE Access* **2019**, *7*, 111168–111180. [[CrossRef](#)]
30. Jia, F.; Lei, Y.; Lu, N.; Xing, S.J. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. *Mech. Syst. Signal Process.* **2018**, *110*, 349–367. [[CrossRef](#)]
31. Arjovsky, M.; Bottou, L.J. Towards principled methods for training generative adversarial networks. *arXiv* **2017**, arXiv:1701.04862.
32. Available online: <https://csegroups.case.edu/bearingdatacenter/home> (accessed on 7 March 2022).

Article

An Oversampling Method of Unbalanced Data for Mechanical Fault Diagnosis Based on MeanRadius-SMOTE

Feng Duan, Shuai Zhang *, Yinze Yan and Zhiqiang Cai

School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an 710072, China; fengduan@mail.nwpu.edu.cn (F.D.); yanyinze@mail.nwpu.edu.cn (Y.Y.); caizhiqiang@nwpu.edu.cn (Z.C.)
* Correspondence: zhangshuai5000@nwpu.edu.cn

Abstract: With the development of machine learning, data-driven mechanical fault diagnosis methods have been widely used in the field of PHM. Due to the limitation of the amount of fault data, it is a difficult problem for fault diagnosis to solve the problem of unbalanced data sets. Under unbalanced data sets, faults with little historical data are always difficult to diagnose and lead to economic losses. In order to improve the prediction accuracy under unbalanced data sets, this paper proposes MeanRadius-SMOTE based on the traditional SMOTE oversampling algorithm, which effectively avoids the generation of useless samples and noise samples. This paper validates the effectiveness of the algorithm on three linear unbalanced data sets and four step unbalanced data sets. Experimental results show that MeanRadius-SMOTE outperforms SMOTE and LR-SMOTE in various evaluation indicators, as well as has better robustness against different imbalance rates. In addition, MeanRadius-SMOTE can take into account the prediction accuracy of the overall and minority class, which is of great significance for engineering applications.

Keywords: mechanical fault diagnosis; unbalanced data set; MeanRadius-SMOTE; minority class

Citation: Duan, F.; Zhang, S.; Yan, Y.; Cai, Z. An Oversampling Method of Unbalanced Data for Mechanical Fault Diagnosis Based on MeanRadius-SMOTE. *Sensors* **2022**, *22*, 5166. <https://doi.org/10.3390/s22145166>

Academic Editor: Jongmyon Kim

Received: 13 June 2022

Accepted: 8 July 2022

Published: 10 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the continuous innovation of technology, industrial equipment has developed rapidly in the direction of large-scale, automated, integrated, and intelligent, such as aircraft engines, steam turbines, wind turbines, centrifuges, etc. In order to meet the requirements of mechanical equipment reliability and precision in the industrial field, PHM (Prognostics and Health Management) was initiated to ensure the stable operation of mechanical equipment and reduce maintenance costs [1–3].

With the development of big data in the industrial field, data-driven mechanical fault diagnosis research has received more and more attention [4–6]. Mechanical fault diagnosis generally starts by extracting vibration signals from the operation of the equipment, because vibration signals can provide sufficient fault features to reflect the fault status and serve as the input of the prediction model [7,8]. However, due to the low frequency of some faults, the vibration signals of such faults are too small, and the classifier cannot predict them accurately, which is the problem of unbalanced data sets in fault diagnosis. In the multi-classification mechanical fault diagnosis problem, the machine learning classifier emphasizes the accuracy of the overall prediction, which leads to sacrificing the prediction accuracy of the minority class to ensure the prediction of the majority class samples [9]. However, there are infrequent failures in some mechanical equipment, which will lead to huge economic losses once they occur. Therefore, it is necessary to research the problem of unbalanced data sets in mechanical fault diagnosis.

At present, the research on the problem of unbalanced data sets is relatively mature, but this research in the mechanical fault diagnosis field has just begun [10]. Many fault diagnosis techniques rely on reliable and complete data sets, such as multi-sensing fusion techniques [11]. However, since machinery usually operates under normal conditions, it is

difficult to collect enough failure data, so that the actual data set lacks completeness [12,13]. The lack of samples with specific labels can lead to data imbalance problems. In recent years, many scholars have begun to pay attention to this problem and have given their own methods [14,15]. Generally, the solution to the problem of unbalanced data sets is mainly divided into data and algorithm aspects, and sometimes they are combined [16].

For the data aspect, scholars mainly use resampling technology to copy, synthesize, delete, and perform other operations on original samples, to adjust the number of samples to reduce the impact of unbalanced data sets. Resampling techniques are divided into oversampling for minority class samples and undersampling for majority class samples. The main idea of oversampling is to increase the number of minority class samples to achieve class balance. The main methods are divided into replicating samples and generating new samples. ROS (Random Oversampling) is to randomly replicate original samples to expand the number of minority class samples, but it may cause the replication of noise samples to affect the quality of the data set [17]. The method of generating new samples derives new samples from one or more original samples, and the new samples can indirectly reflect the features of the minority class. The most classic oversampling is the SMOTE algorithm [18]. The SMOTE algorithm selects the line connecting the two original samples as the range of the new sample and determines a point on the line as the new sample. However, SMOTE still does not avoid the generation of noise samples, and the new samples are easily affected by the distribution of the original samples, which may cause the new samples to deviate from the actual distribution. Later scholars improved SMOTE in terms of noise reduction and generation algorithms, such as Borderline-SMOTE [19], Adasyn [20], LR-SMOTE [21], etc. Undersampling achieves class balance by reducing the number of majority class samples, such as undersampling based on the clustering algorithm and ENN (Edited Nearest Neighbor) [22]. In fact, most of the unbalanced data sets are caused by too few samples in the minority class, so oversampling is the key research in this field [23].

For the algorithm aspect, with the rapid development of machine learning, many classifiers have responded to the problem of unbalanced data sets. On the premise that each sample is equal, the number of samples determines which class the classifier prefers, so setting the weight of the sample, the threshold of the decision boundary, or the objective function of the classifier can strengthen the ability of the classifier to combat unbalanced data sets [24,25]. Adjusting these can make the classifier's decision boundary less sensitive to the sample size [26]. Moreover, adding a proper regularization term to the objective function can reduce the impact of the imbalance rate on the classifier [27].

There is no universal solution to the problem of unbalanced data sets in mechanical fault diagnosis; although, scholars have tried in various directions. From the perspective of features, extracting more abundant features from vibration signals is beneficial to solving the problem, because the failure can be reflected in the energy of the vibration of the equipment [28]. In addition to features in the time and frequency domains, there are features based on wavelet packet energy and entropy values [29,30], and the fault features are also extracted using a bag-of-visual-word approach from the infrared thermography images [31]. However, the increase of features will undoubtedly increase the workload of feature screening. From the perspective of resampling, scholars use various existing resampling methods to conduct experiments on mechanical equipment [32]. Once there are more failure types or concurrent failures, existing oversampling algorithms may fail. Therefore, analyzing the commonality of mechanical faults and proposing a new oversampling algorithm is the key to solving this problem in the mechanical field [33,34]. From the perspective of the classifier, scholars mainly set the cost matrix, and change the loss function or network structure to make the classifier aware of this imbalance [35]. These classifiers are often only suitable for identifying faults in stationary parts, such as gears or bearings [36].

Although new oversampling algorithms are emerging, there are still the following problems: (1) The solutions are generally only aimed at the prediction of bearing failures or gear failures, so the methods cannot comprehensively diagnose the running state of

complete mechanical equipment. (2) Most of the solutions are aimed at the two-category problem, which is obviously not practical. For a simple secondary planetary gear, there are already as many as eight failure types. (3) The new samples are not effective enough that the existing oversampling methods generate. Although the number has reached a balance, it is far from enough in terms of the amount of fault-type information contained in the sample.

In view of the existing problems, this paper improves SMOTE and proposes an oversampling algorithm called MeanRadius-SMOTE, which is specially used to solve the multi-classification problems in mechanical fault diagnosis. MeanRadius-SMOTE can reduce the production of noise samples and add more samples with the ability to affect the decision boundary, and it is easier to inherit the feature information from the original samples. The complexity of the MeanRadius-SMOTE algorithm is not high compared to SMOTE.

The main contributions of this paper are as follows: To solve the problem of multi-classification unbalanced data sets in mechanical fault diagnosis, a new oversampling algorithm, MeanRadius-SMOTE, is proposed. The algorithm takes into account the performance of prediction of overall and minority class, and especially in the minority class, prediction accuracy is greatly improved. In this paper, a large number of comparative experiments are carried out on data sets with various specifications and imbalance rates, and the effectiveness, stability, and robustness of the algorithm are verified.

The rest of this paper is divided into five parts. In Section 2, the SMOTE algorithm and the improved LR-SMOTE algorithm based on SMOTE are introduced. In Section 3, the specific process of the MeanRadius-SMOTE algorithm is introduced in detail. In Section 4, we mainly introduce the source and processing of the data set, as well as the selection of classifiers and evaluation indicators in the experiment. In Section 5, we introduce the experimental process and experimental results. In the following sections, we discuss and summarize the MeanRadius-SMOTE algorithm based on experiments, and we propose future research directions.

2. Related Works

Since the machine learning algorithm is greedy in the face of multi-classification problems, the classifier will give priority to ensuring the highest overall accuracy, resulting in an inaccurate prediction of some minority class samples. In the real industrial field, in the face of some faults with low probability but high maintenance cost, operators hope that the model can accurately predict these faults. Therefore, this section introduces the commonly used methods to deal with unbalanced data sets, namely, the traditional SMOTE method and the improved LR-SMOTE method.

2.1. SMOTE

The SMOTE algorithm was proposed by Chaw La et al. in 2002 [18], and the algorithm is an improved method based on ROS. In the SMOTE algorithm, new samples are generated based on the original samples, which has a greater probability of obtaining effective features than random oversampling of new samples. The steps of the SMOTE algorithm are as follows:

- (1) For each sample x in the training set, calculate their Euclidean distance to each minority class sample x_i , and obtain the k nearest neighbors of each minority class sample.
- (2) According to the sample imbalance rate, set the sampling ratio N . For x_i , randomly select N samples from its k nearest neighbors, denoted as x_h .
- (3) According to Equation (1), build new samples based on x_i and x_h until the classes are balanced, denoted as x_{new} .

$$x_{\text{new}} = x_i + \text{rand}(0, 1) * (x_h - x_i) \quad (1)$$

Although the SMOTE algorithm overcomes the overfitting problem of the ROS algorithm, SMOTE still has some problems with noise samples and useless samples. Many scholars have improved SMOTE. For example, Han proposed the Borderline-SMOTE algorithm [19]. The algorithm first classifies the original samples into safe, dangerous, and noise, then uses the dangerous samples to generate new samples. It not only reduces the interference of noise points but also enables new samples to better reflect the features of the data set. However, how to accurately divide the three labels is a more difficult problem for different data sets.

2.2. LR-SMOTE

Based on the SMOTE algorithm, Wang proposed the LR-SMOTE algorithm [21]. The algorithm first uses SVM (Support Vector Machine) and K-means to remove the noise samples in the original data set, then changes the generation rules of new samples and considers the center of the samples to generate new samples. The specific steps of the LR-SMOTE algorithm are as follows:

- (1) Use SVM to classify the data set, and then for the wrongly classified minority samples use the K-means method to judge and remove the noise samples.
- (2) Use K-means to find the center x_c of the minority class sample, calculate the distance d_i from each minority class sample to the center x_c , and calculate the average distance d_m .
- (3) For each minority class sample x_i , calculate the ratio M_i of the average distance d_m and the distance d_i .
- (4) According to the number of the same samples in the neighbor samples, set the weight of each minority class sample, and then randomly select a minority class sample x_i and build new samples x_{new} according to Equation (2).

$$x_{new} = x_i + rand(0, M_i) * (x_c - x_i) \quad (2)$$

- (5) Repeat steps 3 and 4 until the number of samples of the majority class and minority class is balanced.

In the LR-SMOTE algorithm, the new samples are generated based on the functional relationship between the sample center and each sample, rather than any two minority samples. Therefore, the new samples will not deviate from the range of the minority samples, and the features are closer to the original sample. LR-SMOTE provides a good direction for generating rules so that the algorithm determines the distribution of samples according to the sample center. This paper also proposes a new algorithm along this way to solve the unbalanced data sets in the mechanical field. We use the MeanRadius-SMOTE algorithm to experiment on a variety of mechanical failure data sets, and the experimental results show that the MeanRadius-SMOTE algorithm is suitable for solving the problem of unbalanced data sets in the mechanical field.

3. Proposed Method

In an oversampling algorithm, new samples at different geometric locations have different improvements in classifier training. In general, the more new samples near the decision boundary, the greater the impact on the classifier. This paper proposes the MeanRadius-SMOTE (MR-SMOTE) algorithm considering the sample center and radius. When using machine learning to predict mechanical failures, we deal with noise samples in advance, so noise reduction is performed in feature preprocessing. Noise reduction is not involved in the MeanRadius-SMOTE, and the noise reduction algorithm will be introduced in the next section.

The MeanRadius-SMOTE algorithm mainly changes the generation rules of the SMOTE algorithm, so that the new samples are more likely to be distributed near the average radius of the minority class samples, and the new samples have a stronger ability to affect the decision boundary of the classifier. In the MeanRadius-SMOTE algorithm, the new sample is determined by k vectors of the sample center to the samples, and the

distance between the new sample and the sample center follows a normal distribution. The steps of the MeanRadius-SMOTE algorithm are as follows:

- (1) According to each minority class sample, calculate the geometric center, denoted as the sample center x_c of the minority class sample.
- (2) Calculate the Euclidean distance from each minority class sample to the sample center, and then obtain the average distance, denoted as the sample radius d_m of the minority class.
- (3) Randomly select k minority class samples, and then obtain k vectors v_i from the sample center x_c to the samples. Compute the resultant vector of k vectors.
- (4) Use a normal distribution with mean d_m and variance $\frac{d_m}{\theta}$ to determine the distance between the new sample and the sample center. According to Equation (3), build new samples.

$$x_{new} = x_c + r * \sum_{i=0}^k v_i \quad r \sim N(d_m, \frac{d_m}{\theta}) \tag{3}$$

- (5) Repeat steps 3 and 4 until the number of samples of the majority class and minority class is balanced.

In order to show the flow of the algorithm more conveniently, we draw the flow chart of the MeanRadius-SMOTE algorithm, as shown in Figure 1.

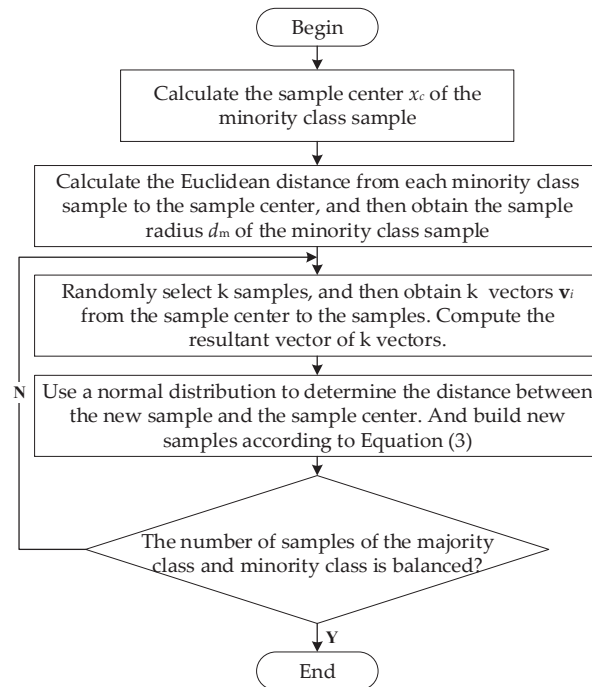


Figure 1. The flow chart of the MeanRadius-SMOTE algorithm.

In the MeanRadius-SMOTE algorithm, k and θ are hyperparameters of the algorithm, which are determined according to the number of minority class samples and the imbalance rate. If k is too large, the direction of the new sample relative to the sample center will become meaningless, and θ directly affects the distribution of the new sample. As shown in Figure 2, new samples under different θ are likely to be distributed in colored areas. When θ is too small, the new sample may be far from the sample center. When θ is too large, the new sample is too conservative and cannot balance the number of positive and

negative samples near the decision boundary. Therefore, in general, the selection range of parameters k is 2 to 5 and the selection range of parameters θ is 4 to 10.

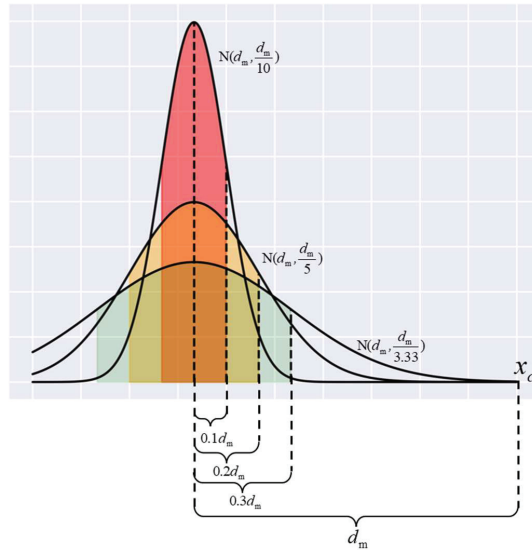


Figure 2. New samples distribution under different θ .

For mechanical equipment, some concurrent faults and the original fault have similar vibration states, and the two types of samples often overlap in distribution. Whether the classifier can find an excellent decision boundary is the key to determining the accuracy of the model. In the MeanRadius-SMOTE algorithm, most of the new samples are concentrated around the sample radius to ensure the validity of the new samples. The new sample is determined by k samples and is related to the sample center, so that the new sample can better inherit the features of the minority class. The geometric positions of new samples generated by different oversampling algorithms have their own characteristics, so we plot the examples of SMOTE, LR-SMOTE, and MeanRadius-SMOTE on two-dimension feature samples, as shown in Figure 3. The information of the two-dimension feature samples is shown in Table 1.

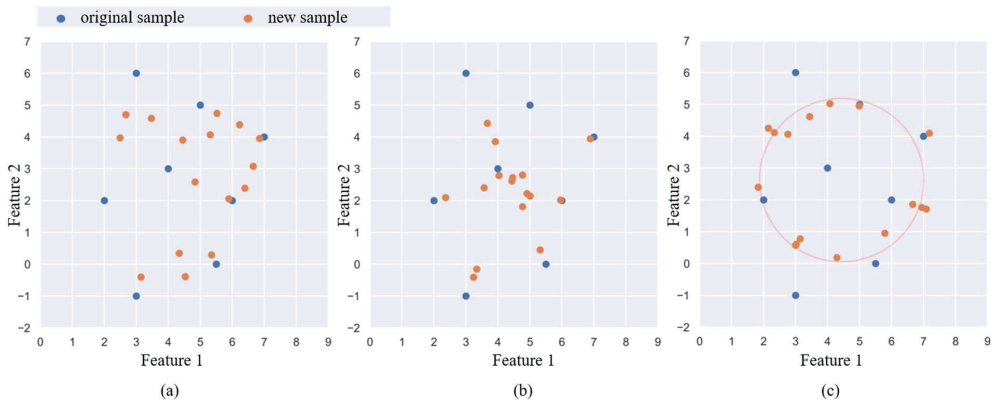


Figure 3. New samples on oversampling algorithms: (a) SMOTE, (b) LR-SMOTE, (c) MeanRadius-SMOTE.

Table 1. The information of the two-dimension feature samples.

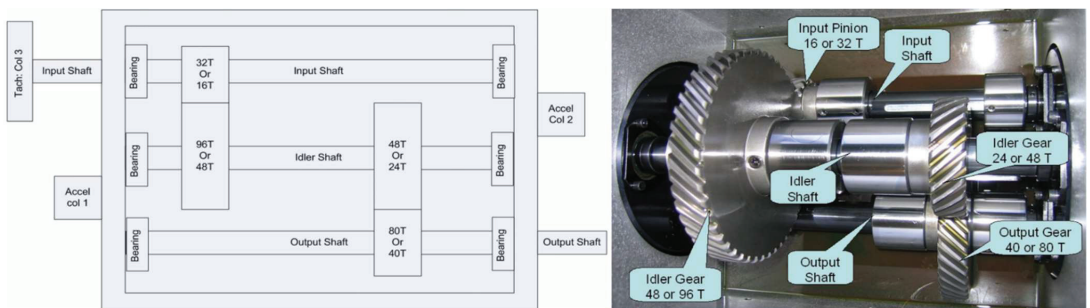
| | Sample 1 | Sample 2 | Sample 3 | Sample 4 | Sample 5 | Sample 6 | Sample 7 | Sample 8 |
|-----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Feature 1 | 3 | 4 | 6 | 7 | 5 | 2 | 3 | 5.5 |
| Feature 2 | 6 | 3 | 2 | 4 | 5 | 2 | −1 | 0 |

The new samples of SMOTE are more inclined to be generated in locations with a high density of the original samples. Since LR-SMOTE randomly chooses a sample to determine the orientation of the new sample, the new sample is more clustered and radial. In MeanRadius-SMOTE, the orientation of new samples is relatively random, and the new samples are generated around the sample radius.

4. Experimental Preparation

4.1. Data Set

Our experimental data set is the 2009 PHM data challenge of gearbox [37]. The data set is a typical industrial gearbox data set, which contains 3 shafts, 4 gears, and 6 bearings, and its experimental bench is shown in Figure 4. The data set tests two sets of gears: spur gear and helical gear. The spur gear data set contains 8 health states, and the helical gear data set contains 6 health states. The data set consists of two channels of accelerometer signals and one channel of tachometer signals. The sampling frequency is 66.67 kHz, and the tachometer signals are collected at 10 pulses per revolution. There are five types of shaft speeds: 30 Hz, 35 Hz, 40 Hz, 45 Hz, and 50 Hz, with high and low loads. In the experiment, we chose the low load spur gear operating data at 30 Hz, and we used the vibration data of the two acceleration channels for feature extraction. The 8 health states of spur gears are as follows in Table 2.

**Figure 4.** Gearbox used in PHM 2009 challenge data.**Table 2.** A brief description of the faults.

| Label | Description |
|---------|---|
| Label 1 | Good |
| Label 2 | Gear chipped and eccentric |
| Label 3 | Gear eccentric |
| Label 4 | Gear eccentric and broken, bearing ball fault |
| Label 5 | Gear chipped and eccentric and broken, bearing inner and ball and outer fault |
| Label 6 | Gear broken, bearing inner and ball and outer fault, shaft imbalance |
| Label 7 | Bearing inner fault, shaft keyway sheared |
| Label 8 | Bearing ball and outer fault, shaft imbalance |

Mechanical equipment frequently fails in the harsh environment of high temperature and high pressure due to concurrent failures composed of multiple single failures [38]. In

the PHM dataset, there are many types of concurrent failures, such as labels 4 to 8. They are all combinations of different types of failures in gears and bearings.

For the vibration signal, we sampled the data set using a sliding window with a stride of 100 and a width of 1000. Then we extracted time–frequency domain features for each vibration signal sample and add labels [39]. The formula of 23 features is shown in Table 3.

Table 3. The time–frequency domain features.

| Time-Domain Feature | | Frequency-Domain Feature | |
|---|--|--|---|
| $F_1 = \frac{\sum_{n=1}^N x(n)}{N}$ | $F_7 = \frac{\sum_{n=1}^N (x(n)-F_1)^4}{(N-1)F_2^4}$ | $F_{12} = \frac{\sum_{k=1}^K s(k)}{K}$ | $F_{18} = \sqrt{\frac{\sum_{k=1}^K f_k^2 s(k)}{\sum_{k=1}^K s(k)}}$ |
| $F_2 = \sqrt{\frac{\sum_{n=1}^N (x(n)-F_1)^2}{N-1}}$ | $F_8 = \frac{F_5}{F_4}$ | $F_{13} = \frac{\sum_{k=1}^K (s(k)-F_{12})^2}{K-1}$ | $F_{19} = \sqrt{\frac{\sum_{k=1}^K f_k^4 s(k)}{\sum_{k=1}^K f_k^2 s(k)}}$ |
| $F_3 = \left(\frac{\sum_{n=1}^N \sqrt{x(n)}}{N}\right)^2$ | $F_9 = \frac{F_5}{F_3}$ | $F_{14} = \frac{\sum_{k=1}^K (s(k)-F_{12})^3}{K(\sqrt{F_{13}})^3}$ | $F_{20} = \frac{\sum_{k=1}^K f_k^2 s(k)}{\sqrt{\sum_{k=1}^K s(k) \sum_{k=1}^K f_k^2 s(k)}}$ |
| $F_4 = \sqrt{\frac{\sum_{n=1}^N (x(n))^2}{N}}$ | $F_{10} = \frac{F_4}{\frac{1}{N} \sum_{n=1}^N x(n) }$ | $F_{15} = \frac{\sum_{k=1}^K (s(k)-F_{12})^4}{K F_{13}^2}$ | $F_{21} = \frac{F_{12}}{F_{16}}$ |
| $F_5 = \max x(n) $ | $F_{11} = \frac{F_5}{\frac{1}{N} \sum_{n=1}^N x(n) }$ | $F_{16} = \frac{\sum_{k=1}^K f_k s(k)}{\sum_{k=1}^K s(k)}$ | $F_{22} = \frac{\sum_{k=1}^K (s(k)-F_{16})^3 s(k)}{K F_{17}^3}$ |
| $F_6 = \frac{\sum_{n=1}^N (x(n)-F_1)^3}{(N-1)F_2^3}$ | | $F_{17} = \frac{\sum_{k=1}^K (f_k - F_{16})^2 s(k)}{K}$ | $F_{23} = \frac{\sum_{k=1}^K (s(k)-F_{16})^{0.5} s(k)}{K \sqrt{F_{17}}}$ |
| where $x(n)$ is a signal series for $n = 1 - N$, and N is the number of data points. | | where $s(k)$ is a signal series for $k = 1 - K$, and K is the number of spectrum lines; f_k is the frequency value of the k th spectrum line. | |

In the experiment, we used the K-nearest neighbor algorithm to denoise the data set. If the five nearest samples around a sample are not of this class, we consider it to be a noise sample and delete it. After the above preprocessing, we obtained 2656 samples per label, of which 1000 samples per label were taken as the test set. Additional samples were used to construct unbalanced data sets.

4.2. Classifiers

In order to comprehensively evaluate the oversampling algorithm, we chose different classifiers to build the experimental model, which excludes the influence of the classifier and verifies the generality of the oversampling algorithm. Through experiments in a large number of mechanical fault diagnoses, the SVM classifier generally has a good training effect, so we chose SVM to establish a classification model. With the continuous development of the decision tree algorithm, the ensemble learning model is also favored by scholars because of its excellent generalization ability. Therefore, we chose RF (Random Forest) representing bagging ensemble mode, and GBDT (Gradient Boosting Decision Tree) representing boosting ensemble mode for experiments.

4.3. Evaluation Indicators

Traditional evaluation indicators can well evaluate the performance of the model in the two-category problem. However, in the multi-classification problem, due to the partiality of the classifier, these indicators cannot comprehensively evaluate the model on unbalanced data sets. The expectation of the oversampling algorithm in this paper is to improve the prediction performance of the minority class without losing the overall prediction accuracy of the classifier. Therefore, we will use the traditional evaluation indicators and the prediction indicator of the minority class to evaluate the prediction model. For class i samples, we define the prediction results as follows, as shown in Table 4:

Table 4. Predicting results for class i samples.

| | Positive Prediction | Negative Prediction |
|----------------|---------------------|---------------------|
| Positive class | TP_i | FN_i |
| Negative class | FP_i | TN_i |

We choose the following four evaluation indicators:

- (1) Accuracy (Acc): The Acc value is the ratio of the number of correctly predicted samples to the total number of samples. The calculation method is as shown in Equation (4):

$$\text{Acc} = \frac{\sum_n TP_i + FN_i}{\sum_n FP_i + TN_i + TP_i + FN_i} \quad (4)$$

The Acc value evaluates the overall prediction, but in the case of unbalanced data sets, it is not a good indicator to measure the results.

- (2) Macro-Precision (Mac-P): The calculation method of the Precision value for class i samples is as shown in Equation (5):

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \quad (5)$$

In the multi-classification problem, the Precision value is divided into Macro and Micro methods. Micro-Precision focuses more on types of samples with a large number of samples, so it is more susceptible to the majority class. However, Mac-P will treat each type of sample equally, so it can better describe the model's ability to deal with unbalanced data sets. The calculation method is as shown in Equation (6):

$$\text{Mac - P} = \frac{\sum_n \text{Precision}_i}{n} \quad (6)$$

- (3) Macro-F1 (Mac-F1): It is contradictory to improve the Precision value and Recall value at the same time. The F1 value is a balance point with high Precision value and high Recall value, and its calculation method is as shown in Equation (7):

$$F1_i = \frac{2 * \text{Precision}_i * \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (7)$$

In the multi-classification problem, The F1 value also has Macro and Micro methods such as the Precision value. This paper selects Mac-F1, which can better take into account the minority class. The calculation method is as shown in Equation (8):

$$\text{Mac - F1} = \frac{\sum_n F1_i}{n} \quad (8)$$

- (4) Precision-Minority (Pre_{small}): In order to pay more attention to the prediction effect of the model on the minority class samples after oversampling algorithms, we will calculate the Precision value of the minority class as an indicator, and its calculation method is as shown in Equation (9):

$$\text{Pre}_{small} = \frac{TP_{small}}{FP_{small} + TP_{small}} \quad (9)$$

5. Experimental Design and Results

In this paper, we will design unbalanced data sets of various sizes for experiments. According to the distribution of sample data volume within each class, unbalanced data

sets can be divided into two forms, linear imbalance and step imbalance. The distribution of sample data volume for the two forms is as shown in Figure 5.

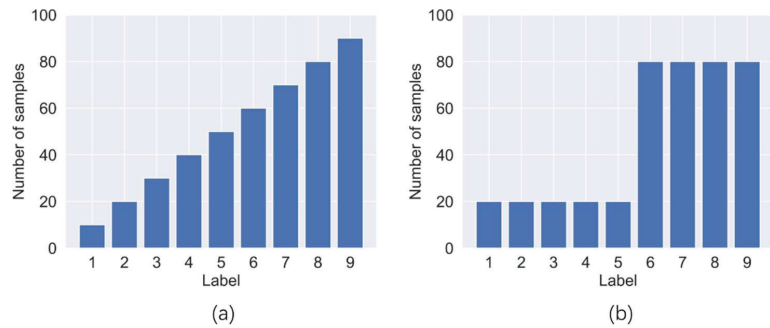


Figure 5. Two imbalance forms: (a) linear imbalance, (b) step imbalance.

In this paper, we design three linear unbalanced data sets and four step unbalanced data sets. In order to reduce the interference of the class on the Pre_{small} in different experiments, we set the number of samples for labels 4 to 50 as the smallest minority class. We set the normal label as the large sample class, and the imbalance rate is designed to be 30, 20, and 15, through which the number of other labels can be determined. The details of the seven unbalanced data sets are shown in Table 5. For line-1 to 3, their imbalance rates are not the same. Moreover, the label linear order is shuffled. For stage-1 to 4, there are differences in the imbalance rate and the ratio of minority class and majority class labels.

Table 5. Unbalanced data sets description.

| Imbalance Forms | Name | Number of Samples | | | | | | | | Imbalance Rate |
|-----------------|---------|-------------------|---------|---------|---------|---------|---------|---------|---------|----------------|
| | | Label 1 | Label 2 | Label 3 | Label 4 | Label 5 | Label 6 | Label 7 | Label 8 | |
| linear | line-1 | 1500 | 465 | 258 | 50 | 672 | 879 | 1293 | 1086 | 30 |
| | line-2 | 1000 | 864 | 592 | 50 | 728 | 321 | 185 | 457 | 20 |
| | line-3 | 750 | 550 | 450 | 50 | 150 | 350 | 650 | 250 | 15 |
| step | stage-1 | 1500 | 50 | 1500 | 50 | 1500 | 1500 | 1500 | 50 | 30 |
| | stage-2 | 750 | 50 | 750 | 50 | 750 | 750 | 750 | 50 | 15 |
| | stage-3 | 1500 | 50 | 1500 | 50 | 50 | 50 | 50 | 1500 | 30 |
| | stage-4 | 750 | 50 | 750 | 50 | 50 | 50 | 50 | 750 | 15 |

In the experiment, we will use the SMOTE, LR-SMOTE, and MeanRadius-SMOTE to oversample the seven unbalanced data sets, so that each class label becomes balanced. Then, we conduct experiments on the original data set and the three processed data sets on SVM, RF, and GBDT classifiers. In order to eliminate the training bias caused by random data, all experiments were performed with 5-fold cross-validation and repeated 10 times to obtain the average number of indicators.

The experimental results of Acc, Mac-P, and Mac-F1 on the linear unbalanced data sets and step unbalanced data sets are shown in Tables 6 and 7, where the values with bold mean the highest value in four compared models.

From Table 6, it can be found that the oversampling algorithm can effectively improve Acc, Mac-P, and Mac-F1, and MeanRadius-SMOTE is the best in most cases. In some experiments, SMOTE performs better than MeanRadius-SMOTE, but the gap between them is very small. However, in the SVM experiment, MeanRadius-SMOTE improves the three indicators much better than SMOTE and LR-SMOTE.

From Table 7, since there are more minority classes in the step unbalanced data sets, the three indicators are all lower in the experiments without oversampling, and are more affected by the imbalance rate. The SVM classifier combined with any oversampling algorithm is better than the ensemble learning classifier, and there are obvious gaps in the three indicators.

On the step unbalanced data sets, MeanRadius-SMOTE outperforms SMOTE and LR-SMOTE in all cases, and the gap is especially significant on the SVM classifier.

Table 6. Experimental results of the linear unbalanced data set.

| Data Set | Methods | SVM | | | RF | | | GBDT | | |
|----------|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | | Acc | Mac-P | Mac-F1 | Acc | Mac-P | Mac-F1 | Acc | Mac-P | Mac-F1 |
| line-1 | None | 0.8675 | 0.8896 | 0.8484 | 0.7726 | 0.8122 | 0.7256 | 0.8126 | 0.8243 | 0.7842 |
| | SMOTE | 0.9045 | 0.9148 | 0.8997 | 0.8555 | 0.8685 | 0.8471 | 0.8774 | 0.8849 | 0.8731 |
| | LR-SMOTE | 0.9065 | 0.9161 | 0.9012 | 0.8339 | 0.8528 | 0.8182 | 0.8662 | 0.8746 | 0.8591 |
| | MR-SMOTE | 0.9206 | 0.9233 | 0.9186 | 0.8678 | 0.8730 | 0.8643 | 0.8739 | 0.8773 | 0.8698 |
| line-2 | None | 0.8733 | 0.8945 | 0.8607 | 0.7668 | 0.8075 | 0.7243 | 0.8209 | 0.8412 | 0.8011 |
| | SMOTE | 0.8891 | 0.9062 | 0.8836 | 0.8548 | 0.8629 | 0.8501 | 0.8626 | 0.8713 | 0.8566 |
| | LR-SMOTE | 0.8923 | 0.9059 | 0.8865 | 0.8354 | 0.8497 | 0.8271 | 0.8588 | 0.8685 | 0.852 |
| | MR-SMOTE | 0.9139 | 0.9160 | 0.9131 | 0.8675 | 0.8702 | 0.8657 | 0.8733 | 0.8780 | 0.8698 |
| line-3 | None | 0.8754 | 0.8890 | 0.8644 | 0.792 | 0.8261 | 0.7583 | 0.8344 | 0.8496 | 0.818 |
| | SMOTE | 0.8995 | 0.9069 | 0.8954 | 0.8646 | 0.8726 | 0.8618 | 0.8748 | 0.8782 | 0.8716 |
| | LR-SMOTE | 0.8988 | 0.9058 | 0.8947 | 0.8464 | 0.8575 | 0.8415 | 0.8683 | 0.8730 | 0.8639 |
| | MR-SMOTE | 0.9175 | 0.9183 | 0.9168 | 0.8691 | 0.8720 | 0.8679 | 0.8803 | 0.8807 | 0.8784 |

Table 7. Experimental results of the step unbalanced data set.

| Data Set | Methods | SVM | | | RF | | | GBDT | | |
|----------|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | | Acc | Mac-P | Mac-F1 | Acc | Mac-P | Mac-F1 | Acc | Mac-P | Mac-F1 |
| Stage-1 | None | 0.7403 | 0.8207 | 0.7066 | 0.6144 | 0.7610 | 0.5051 | 0.6793 | 0.7553 | 0.6194 |
| | SMOTE | 0.8418 | 0.8685 | 0.8332 | 0.7614 | 0.8166 | 0.7447 | 0.8250 | 0.8487 | 0.8182 |
| | LR-SMOTE | 0.8566 | 0.8789 | 0.8512 | 0.7103 | 0.7950 | 0.6729 | 0.8021 | 0.8403 | 0.7915 |
| | MR-SMOTE | 0.9039 | 0.9062 | 0.9023 | 0.844 | 0.8592 | 0.8408 | 0.8596 | 0.8706 | 0.8561 |
| Stage-2 | None | 0.7746 | 0.8365 | 0.7528 | 0.6398 | 0.7694 | 0.5538 | 0.7193 | 0.7936 | 0.6838 |
| | SMOTE | 0.8575 | 0.8760 | 0.8525 | 0.7790 | 0.8242 | 0.7682 | 0.8368 | 0.8551 | 0.8330 |
| | LR-SMOTE | 0.8649 | 0.8833 | 0.8602 | 0.7429 | 0.8073 | 0.7202 | 0.8205 | 0.8481 | 0.8142 |
| | MR-SMOTE | 0.9064 | 0.9078 | 0.9051 | 0.838 | 0.8529 | 0.8357 | 0.8621 | 0.8723 | 0.8601 |
| Stage-3 | None | 0.6465 | 0.8034 | 0.6369 | 0.4534 | 0.7669 | 0.3607 | 0.5651 | 0.6999 | 0.5259 |
| | SMOTE | 0.7828 | 0.8481 | 0.7847 | 0.7390 | 0.8181 | 0.7336 | 0.8048 | 0.8297 | 0.8022 |
| | LR-SMOTE | 0.8118 | 0.8546 | 0.8116 | 0.6766 | 0.8001 | 0.6654 | 0.7641 | 0.8044 | 0.7583 |
| | MR-SMOTE | 0.8771 | 0.8826 | 0.8759 | 0.8163 | 0.8366 | 0.8151 | 0.8351 | 0.8431 | 0.8327 |
| Stage-4 | None | 0.6823 | 0.8124 | 0.6767 | 0.5186 | 0.7697 | 0.4671 | 0.6491 | 0.7391 | 0.6334 |
| | SMOTE | 0.8221 | 0.8606 | 0.8223 | 0.767 | 0.8119 | 0.7634 | 0.8098 | 0.8315 | 0.8082 |
| | LR-SMOTE | 0.8440 | 0.8700 | 0.8434 | 0.7186 | 0.7957 | 0.7131 | 0.7871 | 0.8205 | 0.7848 |
| | MR-SMOTE | 0.8766 | 0.8829 | 0.8762 | 0.8135 | 0.8278 | 0.8119 | 0.8436 | 0.8513 | 0.8425 |

By analyzing Acc, Mac-P, and Mac-F1, all oversampling algorithms can effectively improve the overall prediction performance of the classifier on both forms of unbalanced data sets, and the MeanRadius-SMOTE algorithm proposed in this paper has the most obvious effect. We still need to focus on the prediction performance of the algorithm on the minority class; the experimental results of Pre_{small} are shown in Table 8, where the values with bold mean the largest value in four compared models.

From Table 8, Pre_{small} does not even exceed five in the None experiments. SMOTE and LR-SMOTE only improved Pre_{small} by around five in most experiments. However, MeanRadius-SMOTE can help the classifier to more accurately predict the minority class, improving Pre_{small} by around six or seven. In addition, MeanRadius-SMOTE is more stable in experiments with different imbalance rates, and does not fluctuate greatly like SMOTE and LR-SMOTE.

Table 8. Pre_{small} on the data sets.

| Data Set | Methods | SVM Pre_{small} | RF Pre_{small} | GBDT Pre_{small} | Data Set | Methods | SVM Pre_{small} | RF Pre_{small} | GBDT Pre_{small} |
|----------|----------|-------------------|------------------|--------------------|----------|----------|-------------------|------------------|--------------------|
| line-1 | None | 0.277 | 0.048 | 0.184 | stage-1 | None | 0.329 | 0.039 | 0.154 |
| | SMOTE | 0.563 | 0.472 | 0.588 | | SMOTE | 0.555 | 0.484 | 0.545 |
| | LR-SMOTE | 0.553 | 0.338 | 0.508 | | LR-SMOTE | 0.584 | 0.322 | 0.506 |
| | MR-SMOTE | 0.703 | 0.625 | 0.603 | | MR-SMOTE | 0.781 | 0.615 | 0.619 |
| line-2 | None | 0.358 | 0.052 | 0.261 | stage-2 | None | 0.403 | 0.062 | 0.259 |
| | SMOTE | 0.427 | 0.554 | 0.519 | | SMOTE | 0.616 | 0.506 | 0.637 |
| | LR-SMOTE | 0.501 | 0.449 | 0.433 | | LR-SMOTE | 0.606 | 0.36 | 0.555 |
| | MR-SMOTE | 0.768 | 0.681 | 0.612 | | MR-SMOTE | 0.791 | 0.695 | 0.690 |
| line-3 | None | 0.403 | 0.110 | 0.308 | stage-3 | None | 0.445 | 0.042 | 0.368 |
| | SMOTE | 0.583 | 0.607 | 0.618 | | SMOTE | 0.657 | 0.738 | 0.697 |
| | LR-SMOTE | 0.579 | 0.525 | 0.565 | | LR-SMOTE | 0.662 | 0.605 | 0.616 |
| | MR-SMOTE | 0.791 | 0.701 | 0.681 | | MR-SMOTE | 0.768 | 0.738 | 0.708 |
| stage-4 | None | 0.476 | 0.164 | 0.42 | stage-4 | None | 0.476 | 0.164 | 0.42 |
| | SMOTE | 0.732 | 0.745 | 0.697 | | SMOTE | 0.732 | 0.745 | 0.697 |
| | LR-SMOTE | 0.754 | 0.648 | 0.705 | | LR-SMOTE | 0.754 | 0.648 | 0.705 |
| | MR-SMOTE | 0.783 | 0.760 | 0.778 | | MR-SMOTE | 0.783 | 0.760 | 0.778 |

To better compare the effects of SMOTE, LR-SMOTE, and MeanRadius-SMOTE, we draw the line charts of Mac-P, Mac-F1, and Pre_{small} , as shown in Figure 6. Since the data of Acc and Mac-F1 are close and their trend is basically the same, we only choose Mac-F1 to draw the line chart.

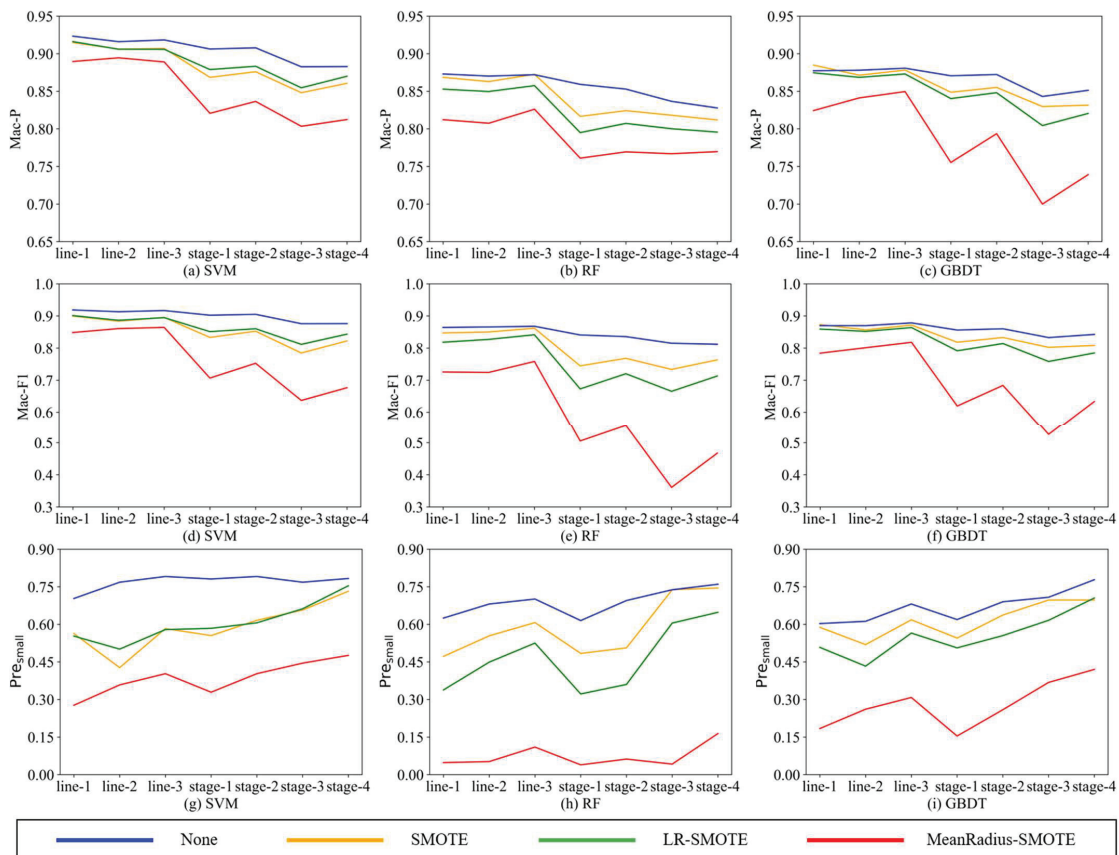


Figure 6. The line charts of Mac-P, Mac-F1, and Pre_{small} .

According to Figure 6, the following conclusions can be drawn:

- (1) Since these seven unbalanced data sets are homologous, the better the oversampling algorithm, the closer the indicators should be. Comparing the nine charts, all indicators are relatively stable in the MeanRadius-SMOTE experiment, which is less affected by the imbalance rate and data set form, and this stabilization is more obvious in the SVM classifier. This shows that MeanRadius-SMOTE has good robustness.
- (2) Analyzing the three charts—Figure 6a,d,g, in the seven data sets, MeanRadius-SMOTE on the SVM classifier can not only ensure that the overall prediction indicators reach about 0.9 but also ensure that Pre_{small} is relatively high, about 0.75.
- (3) Comparing the three charts—Figure 6g–i, the SVM experiment can achieve a higher Pre_{small} , and in most experiments, Pre_{small} is greatly affected by the data sets, especially in the RF experiments. However, only in the model composed of MeanRadius-SMOTE and SVM do we obtain a very flat line, which shows that this model has good robustness and accuracy in predicting the minority class.
- (4) Comparing the three charts—Figure 6a–c, for SMOTE and LR-SMOTE, LR-SMOTE performs better than SMOTE on SVM, while it is the opposite on RF and GBDT. In addition, SMOTE even outperforms MeanRadius-SMOTE in some GBDT experiments. LR-SMOTE is also an oversampling algorithm for binary classification problems, which is more suitable for a classifier that is essentially a binary classification algorithm-SVM. Therefore, it can be inferred that MeanRadius-SMOTE is also more suitable for SVM classifiers.

In summary, MeanRadius-SMOTE shows excellent performance in all experiments, which can take into account the prediction performance of the overall and minority class. In individual experiments, SMOTE is slightly higher than MeanRadius-SMOTE in Acc, Mac-P, and Mac-F1, but lower than MeanRadius-SMOTE in Pre_{small} . We can think that this is the result of sacrificing the prediction performance of the minority class. Therefore, it can still be considered that MeanRadius-SMOTE is better than SMOTE and LR-SMOTE. Furthermore, the model composed of MeanRadius-SMOTE and SVM can improve prediction accuracy and stability.

6. Conclusions and Outlook

Mechanical fault diagnosis has always been a key issue in the PHM. Since the development of machine learning, although mechanical fault diagnosis has been solved by many effective methods, fault diagnosis under unbalanced data sets has always been a stubborn problem. The oversampling algorithm is currently recognized as an effective means to solve the problem of unbalanced data sets. The traditional oversampling algorithm is not only affected by the sample distribution, but also easily generates noise samples, which makes the decision boundary blurred. These drawbacks are not conducive to the classifier making predictions.

Based on the SMOTE, this paper proposes the new algorithm, MeanRadius-SMOTE, combining the sample center and radius. MeanRadius-SMOTE effectively avoids useless samples and noise samples in the process of generating new samples. In this paper, we conduct comparative experiments for SMOTE, LR-SMOTE, and MeanRadius-SMOTE algorithms and use SVM, RF, and GBDT classifiers on three linear unbalanced data sets and four step unbalanced data sets. Experimental results show that the MeanRadius-SMOTE algorithm can effectively balance data classes and improve the prediction performance of machine learning classifiers. From the perspective of various indicators, the MeanRadius-SMOTE algorithm is better than SMOTE and LR-SMOTE, and has better robustness. In the problem of unbalanced data sets, MeanRadius-SMOTE can more accurately predict the minority class without sacrificing the prediction performance of other classes, which is of great significance for mechanical fault diagnosis, and the combined model of MeanRadius-SMOTE and SVM is proved to be much better than other models.

Although this paper proves on PHM09 challenge data that MeanRadius-SMOTE has a good ability to deal with unbalanced data sets, considering the actual situation, future research can be carried out from the following aspects:

- (1) In this paper, in order to ensure that the experiment is carried out under a variety of unbalanced data sets, we use artificial unbalanced data sets in experiments. In future research, we will collect the failure unbalanced data sets of actual mechanical equipment to continue the verification experiment.
- (2) When constructing the data set in this paper, we only extracted the time–frequency domain features from the vibration signal. Currently, there are more methods to extract features from vibration signals, such as convolutional neural networks, wavelet packet decomposition, etc. Training sets composed of different types of features may have an impact on the performance of MeanRadius-SMOTE.

Author Contributions: Data curation, F.D. and S.Z.; validation, F.D.; investigation, S.Z. and Y.Y.; writing—original draft preparation, F.D. and Y.Y.; writing—review and editing, F.D., S.Z., and Z.C.; supervision, S.Z. and Z.C. All authors have read and agreed to the published version of the manuscript.

Funding: This article was supported by the National Natural Science Foundation of China (71871181); the Basic Research Project of Natural Science of Shaanxi Province (2022JM-433); the Key R&D Program of Shaanxi Province (2022GY-207, 2021ZDLGY10-06).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data in this article are the 2009 PHM data challenge of the gearbox.

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

References

1. Yongbo, L.; Wang, S.; Deng, Z. Intelligent fault identification of rotary machinery using refined composite multi-scale Lempel–Ziv complexity. *J. Manuf. Syst.* **2021**, *61*, 725–735.
2. Jiang, W.; Xie, C.; Zhuang, M.; Shou, Y.; Tang, Y. Sensor data fusion with Z-numbers and its application in fault diagnosis. *Sensors* **2016**, *16*, 1509. [[CrossRef](#)] [[PubMed](#)]
3. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal. Process.* **2018**, *108*, 33–47. [[CrossRef](#)]
4. Lu, S.; He, Q.; Yuan, T.; Kong, F. Online fault diagnosis of motor bearing via stochastic-resonance-based adaptive filter in an embedded. *IEEE Trans. Syst. Man Cybern A* **2018**, *47*, 1111–1122. [[CrossRef](#)]
5. Cai, B.; Zhao, Y.; Liu, H.; Xie, M. A data-driven fault diagnosis methodology in Three-Phase inverters for PMSM drive systems. *IEEE Trans. Power Electron.* **2017**, *32*, 5590–5600. [[CrossRef](#)]
6. Cofre-Martel, S.; Drogue, E.; Modarres, M. Big Machinery Data Preprocessing Methodology for Data-Driven Models in Prognostics and Health Management. *Sensors* **2021**, *21*, 6841. [[CrossRef](#)]
7. Akilu, Y.; Sinha, J.; Nembhard, A. A novel fault diagnosis technique for enhancing maintenance and reliability of rotating machines. *Struct. Health Monit.* **2015**, *14*, 604–621.
8. Yongbo, L.; Wang, X.; Si, S.; Huang, S. Entropy Based Fault Classification Using the Case Western Reserve University Data: A Benchmark Study. *IEEE Trans. Reliab.* **2020**, *69*, 754–767.
9. Fernandez, A.; Garcia, S.; Luengo, J.; Bernado-Mansilla, E.; Herrera, F. Genetics-based machine learning for rule induction: State of the art, Taxonomy, and Comparative Study. *IEEE Trans. Evol. Comput.* **2010**, *14*, 913–941. [[CrossRef](#)]
10. Swana, E.; Doorsamy, W.; Bokoro, P. Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset. *Sensors* **2022**, *22*, 3246. [[CrossRef](#)]
11. He, Z.; Shao, H.; Jing, L.; Cheng, J.; Yang, Y. Transfer fault diagnosis of bearing installed in different machines using enhanced deep auto-encoder. *Measurement* **2020**, *152*, 107393.
12. Zhang, T.; Chen, J.; Li, F.; Zhang, K.; Lv, H.; He, S.; Xu, E. Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions. *ISA Trans.* **2022**, *119*, 152–171. [[PubMed](#)]
13. Shao, H.; Lin, J.; Zhang, L.; Galar, D.; Kumar, U. A novel approach of multisensory fusion to collaborative fault diagnosis in maintenance. *Inf. Fusion* **2021**, *74*, 65–76. [[CrossRef](#)]
14. Martin-Diaz, I.; Morinigo-Sotelo, D.; Duque-Perez, O.; De, R. Early fault detection in induction motors using adaboost with imbalanced small data and optimized sampling. *IEEE Trans. Ind. Electron.* **2022**, *53*, 3066–3075. [[CrossRef](#)]

15. Zhang, T.; Chen, J.; Li, F.; Pan, T.; He, S. A small sample focused intelligent fault diagnosis scheme of machines via multi-modules learning with gradient penalized generative adversarial networks. *IEEE Trans. Ind. Electron.* **2021**, *68*, 10130–10141. [[CrossRef](#)]
16. Branco, P.; Torgo, L.; Ribeiro, R. A survey of predictive modeling on imbalanced domains. *ACM Comput. Surv.* **2017**, *45*, 351–400. [[CrossRef](#)]
17. Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. Learning from class-imbalanced data: Review of methods and applications. *Expert Syst. Appl.* **2017**, *73*, 220–239. [[CrossRef](#)]
18. Chawla, N.; Hall, L.; Bowyer, K.; Kegelmeyer, W. SMOTE: Synthetic minority over-sampling technique. *Artif. Intell.* **2002**, *16*, 321–357. [[CrossRef](#)]
19. Han, H.; Wang, W.; Mao, B. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In Proceedings of the Conference on Aristophanes Upstairs and Downstairs, Magdalen Coll, Oxford, UK, 16–18 September 2004.
20. Haibo, H.; Yang, B.; Garcia, E.; Shutao, L. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In Proceedings of the International Joint Conference on Neural Networks, Hong Kong, China, 1–8 June 2008.
21. Liang, X.; Jiang, A.; Li, T.; Xue, Y.; Wang, G. LR-SMOTE—An improved unbalanced data set oversampling based on K-means and SVM. *Knowl.-Based Syst.* **2020**, *196*, 105846. [[CrossRef](#)]
22. Yen, S.; Lee, Y. Under-sampling approaches for improving prediction of the minority class in an imbalanced dataset. In Proceedings of the International Conference on Intelligent Computing, Kunming, China, 16–19 August 2006.
23. Laurikkala, J. Improving identification of difficult small classes by balancing class distribution. In Proceedings of the 8th Conference on Artificial Intelligence in Medicine in Europe, Cascais, Portugal, 1–4 July 2001.
24. Zhang, C.; Gao, W.; Song, J.; Jiang, J. An imbalanced data classification algorithm of improved autoencoder neural network. In Proceedings of the 8th International Conference on Advanced Computational Intelligence, Chiang Mai, Thailand, 14–16 February 2016.
25. Cheng, F.; Zhang, J.; Wen, C. Cost-sensitive large margin distribution machine for classification of imbalanced data. *Pattern Recognit. Lett.* **2016**, *80*, 107–112. [[CrossRef](#)]
26. Xu, Q.; Lu, S.; Jia, W.; Jiang, C. Imbalanced fault diagnosis of rotating machinery via multi-domain feature extraction and cost-sensitive learning. *J. Manuf. Syst.* **2020**, *31*, 1467–1481. [[CrossRef](#)]
27. Zhao, X.; Jia, M.; Lin, M. Deep laplacian auto-encoder and its application into imbalanced fault diagnosis of rotating machinery. *Measurement* **2020**, *31*, 1467–1481. [[CrossRef](#)]
28. Yongbo, L.; Du, X.; Wang, X.; Si, S. Industrial gearbox fault diagnosis based on multi-scale convolutional neural networks and thermal imaging. *ISA Trans.* **2022**, *in press*.
29. Yongbo, L.; Wang, S.; Yang, Y.; Deng, Z. Multiscale symbolic fuzzy entropy: An entropy denoising method for weak feature extraction of rotating machinery. *Mech. Syst. Signal. Process.* **2022**, *162*, 108052.
30. Zhang, C.; Chen, J.; Guo, X. A gear fault diagnosis method based on EMD energy entropy and SVM. *Shock Vib.* **2010**, *29*, 216–220.
31. Yongbo, L.; Wang, X.; Si, S.; Du, X. A New Intelligent Fault Diagnosis Method of Rotating Machinery under Varying-Speed Conditions Using Infrared Thermography. *Complexity* **2019**, *2019*, 2619252.
32. Farajzadeh-Zanjani, M.; Razavi-Far, R.; Saif, M. Efficient sampling techniques for ensemble learning and diagnosing bearing defects under class imbalanced condition. In Proceedings of the IEEE Symposium Series on Computational Intelligence, Athens, Greece, 6–9 December 2016.
33. 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](#)]
34. Razavi-Far, R.; Farajzadeh-Zanjani, M.; Saif, M. An integrated class-imbalanced learning scheme for diagnosing bearing defects in induction motors. *IEEE Trans. Ind. Inform.* **2017**, *13*, 2758–2769. [[CrossRef](#)]
35. Jia, F.; Lei, Y.; Lu, N.; Xing, S. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. *Mech. Syst. Signal. Process.* **2018**, *110*, 349–367. [[CrossRef](#)]
36. Lan, Y.; Han, X.; Xiong, X.; Huang, J.; Zong, W.; Ding, X.; Ma, B. Two-step fault diagnosis framework for rolling element bearings with imbalanced data based on GSA-WELM and GSA-ELM. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2018**, *232*, 2937–2947. [[CrossRef](#)]
37. Jing, L.; Zhao, M.; Li, P.; Xu, X. A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement* **2017**, *111*, 1–10. [[CrossRef](#)]
38. Zhang, K.; Zhou, H.; Yi, C. Review of multiple fault diagnosis methods. *IET Control Theory Appl.* **2015**, *32*, 1143–1157.
39. Lei, Y.; He, Z.; Zi, Y.; Hu, Q. Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs. *Mech. Syst. Signal. Process.* **2007**, *21*, 2280–2294. [[CrossRef](#)]

Article

The Robust Multi-Scale Deep-SVDD Model for Anomaly Online Detection of Rolling Bearings

Linlin Kou ^{1,2}, Jiaxian Chen ³, Yong Qin ¹ and Wentao Mao ^{3,*}

¹ State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China; koulinlin5829@bjsubway.com (L.K.); koull01@163.com (Y.Q.)

² Beijing Mass Transit Railway Operation Corp. Ltd., Beijing 100044, China

³ College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China; 12120823@bjtu.edu.cn

* Correspondence: maowt@htu.edu.cn

Abstract: Aiming at the online detection problem of rolling bearings, the limited amount of target bearing data leads to insufficient model in training and feature representation. It is difficult for the online detection model to construct an accurate decision boundary. To solve the problem, a multi-scale robust anomaly detection method based on data enhancement technology is proposed in this paper. Firstly, the training data are transformed into multiple subspaces through the data enhancement technology. Then, a prototype clustering method is introduced to enhance the robustness of features representation under the framework of the robust deep auto-encoding algorithm. Finally, the robust multi-scale Deep-SVDD hyper sphere model is constructed to achieve online detection of abnormal state data. Experiments are conducted on the IEEE PHM Challenge 2012 bearing data set and XJTU-TU data set. The proposed method shows much greater susceptibility to incipient faults, and it has fewer false alarms. The robust multi-scale Deep-SVDD hyper sphere model significantly improves the performance of incipient fault detection for rolling bearings.

Keywords: incipient fault detection; robustness; reinforcement learning; anomaly detection

Citation: Kou, L.; Chen, J.; Qin, Y.; Mao, W. The Robust Multi-Scale Deep-SVDD Model for Anomaly Online Detection of Rolling Bearings. *Sensors* **2022**, *22*, 5681. <https://doi.org/10.3390/s22155681>

Academic Editors: Yongbo Li, Bing Li, Jinchun Ji and Hamed Kalhori

Received: 4 July 2022

Accepted: 27 July 2022

Published: 29 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As a kind of special mechanical parts, rolling bearing has a decisive impact on the operation and reliability of mechanical equipment [1]. Once damaged, it will cause major losses to industrial production and personal property. Detecting abnormalities in the early stages of bearing fault, and performing accurate and reliable detection and diagnosis, will help to take timely measures for maintenance and avoid major accidents. Incipient fault detection is a key link in the Prognostics and Health Management (PHM) for rolling bearings [2].

For signal analysis-based incipient fault detection methods, noise elimination and noise utilization are conducted at first for vibration signal. Then, time domain, frequency domain or time-frequency domain analysis are performed to extract and compare fault characteristics [3–5]. Vibration spectra were conducted in [6] with consideration of a set of different recurrence indicators to describe the response of the bearing to the optimal clearance. Acoustic emission and lubricating oil characteristics are also very helpful for condition monitoring of bearings. Liu et al. [7] based on acoustic emission signal proposed a modified time-dependent excitation (TDE) model to detect defects of angular contact ball bearings. Chen et al. [8] made contributions on low-speed rolling bearing fault detection with subspace embedded feature distribution alignment and Structural Risk Minimization framework based on acoustic emission signal. Maroua et al. [9] analyzed the performance of different kinds of rolling bearings under five fully formulated axle gear oils with different viscosity and different formulations.

Those methods can extract incipient fault features from the original signal, which work as the input feature vector of classifier or as the indicator of rolling bearing incipient faults. However, the de-noise method has the disadvantage of weakening the fault information. In addition, these time-frequency domain methods cannot adaptively extract features, which lead to the weak ability of bearings early detection.

In recent years, machine learning-based methods are widely applied in many industries. In reference [10], an impact time-frequency dictionary was built to extract signal features with short-time matching method first, and then support vector machine (SVM) worked as classifier for incipient fault states. The supervised local Fisher discriminant and K-nearest neighbor method were introduced for weak fault diagnosing in [11], in which they are working for feature reduction and incipient fault state classification respectively. Ocak et al. [12] proposed a Hidden Markov Model (HMM) for rolling bearing fault detection and diagnosis, which can identify and detect early failures by tracking the probability changes of the pre-trained HMM under normal conditions. These methods usually use part of the normal state data in the initial stage to establish a single classification model, or employ existing normal state samples to construct abnormal discrimination criteria. However, the bearing has certain noise in the normal state, this kind of methods cannot automatically adapt to the irregular data fluctuations caused by the various noise, which may cause false alarms.

In the past decade, deep learning has already become an efficient way to detect and diagnose fault in many fields [13–17]. According to the authors' literature research, deep learning is still in its infancy on incipient fault detection. Lu et al. [18] used deep neural networks (DNN) and long short-term memory (LSTM) to construct an online data distribution estimator, and used the prediction bias value generated by the estimator to identify incipient fault location. A two-way Gated Recurrent Unit (GRU) network with local features is proposed for different types of faults to realize effective identification of incipient faults [19]. A new framework for rotor-bearing system fault diagnosis under varying working conditions is proposed in [20], it introduced stochastic pooling and Leaky rectified linear unit to overcome the training problems in classical CNN. Chen and et al. studied mixed faults diagnosis from multiple components by combining two 1-D convolutional neural networks (CNNs) [21]. Mao et al. [22] proposed an on-line detection method based on self-adaptive deep feature matching for incipient faults of rolling bearings. However, due to weak incipient fault, the ability of feature representation is poor. At the same time, due to the constraints of online application scenarios, the amount of available target-bearing data are limited. These methods have insufficient normal state data information, thus there are some certain obstacles in accurate decision boundary construction of online detection models.

Machine learning is used to dig out regular information from training data and learn pattern recognition knowledge. The parameter number of current deep neural networks is always very huge. It requires sufficient training data for training in order to obtain ideal results. In the case of a limited amount data, data enhancement technology can be used to increase the diversity of training data. Meanwhile, transformation operations can improve the feature representation ability of training data, therefore the feature information of training data is more sufficient and the problem of model over fitting can be avoided. Existing methods of data enhancement include: geometric transformation [23], color space enhancement [24], kernel filter [25], as well as generative adversarial networks [26] based on the idea of antagonistic thoughts and neural style transfer [27] and other methods. A geometric transformation is applied in [28] to solve the data imbalance problem for bearing fault detection.

To solve these problems, an incipient fault detection method based on multi-scale Deep-SVDD model with data enhancement is proposed in this research. First, the training data are transformed into multiple subspaces through data enhancement technology. Second, the prototype clustering method is introduced to improve the robustness of features under the framework of regularized dual averaging (RDA) algorithm, and then a robust multi-scale

Deep-SVDD hyper sphere model is constructed. Finally, the product of the probability that the transformed sample is located in its respective subspace is calculated as the anomaly score to achieve early online fault detection. The effectiveness of the proposed method is verified by experiments on the IEEE PHM Challenge 2012 bearing dataset and XJTU-SY dataset. The contribution of this paper can be summarized as follows. (1) A robust multi-scale Deep-SVDD hyper sphere model is proposed for online anomaly detection. The data information based on data enhancement technology is enriched. By extracting robust low-rank deep features, this method can enhance the capacity of multi-scale features representation and has good robustness. (2) An anomaly alarm indicator is built for online scenarios. This indicator is based on the robust low-rank features extraction, and then can measure abnormality. Therefore, this indicator is very effectively suitable for online applications. The details of this work are as follows.

2. Deep Support Vector Data Description

Deep support vector data description (Deep-SVDD) [29] is a representative method of using deep learning for anomaly detection in recent years. The nonlinear high-dimensional mapping is replaced by a neural network in this method, which improved the ability in dealing with high-dimensional and very large data sets. Deep-SVDD can take advantage of deep learning to deal with high-dimensional representation and processing of massive data.

Deep-SVDD constructs a neural network mapping. The method minimizes the volume of the hyper sphere containing the data features in the network when solving, and obtains the high-dimensional space common feature representation of normal data. The objective function is:

$$\min_W \frac{1}{n} \sum_{i=1}^n \|\phi(x_i; W) - c\|_2^2 + \frac{\lambda}{2} \sum_{l=1}^L \|w^l\|_F^2 \quad (1)$$

The objective function consists of two items. The first item is the quadratic loss of the distance between the penalty sample feature and the center of the hyper sphere, and the second is the regular item that constrains the network weight to prevent over fitting, where ϕ is neural network mapping function, $x_i, i = 1, 2, \dots, n$ is the sample data, W is the set of weight parameters of the network, $W = \{w^1, \dots, w^L\}$. c is the center of hyper sphere, L is the number of layers of the neural network, $l = \{1, 2, \dots, L\}$. λ is the hyper parameters that control the weight decay. W^l is the weight of l th hidden layer.

Optimizing the first item lets the network learn parameters W such that data points are closely mapped to the center c , and optimizing the second item is to minimize the volume of the hypersphere.

Center c is fixed in the neighborhood of the initial network outputs, which makes stochastic gradient descent (SGD) convergence faster and more robust [30].

The abnormal score of the Deep-SVDD algorithm evaluation sample can be calculated by the following:

$$s(x) = \|\phi(x; W^*) - c\|^2 \quad (2)$$

where W^* is the network parameter of a trained model.

The larger $s(x)$, the farther the sample is from the center of the hyper sphere, the higher the degree of abnormality of the sample.

3. The Robust Multi-Scale Deep-SVDD Model of Incipient Fault Detection

In this section, the proposed incipient Fault detection method for bearing is divided into offline stage and online stage. In the offline stage, employing data enhancement technology to transform a small number of normal samples into multiple feature spaces, based on this, the prototype clustering loss and multi-hyper sphere Deep-SVDD center loss are introduced to train the robust multi-scale Deep-SVDD model, and obtain each transformed model in the feature space, the distance-based cross entropy is used to determine the distance score threshold of the normal period data. In the online stage, the test samples to be detected are subjected to the same transformation enhancement, and then they are put

into the trained deep model to extract the deep features. The extracted deep features are used with each prototype center to calculate the distance-based anomaly score, and finally combined with the threshold value. When the score is less than the threshold, the sample is regarded as normal, otherwise, it is judged to be abnormal. Each step of the proposed robust multi-scale deep-SVDD model is elaborated in the following. The detailed flow chart is shown in Figure 1.

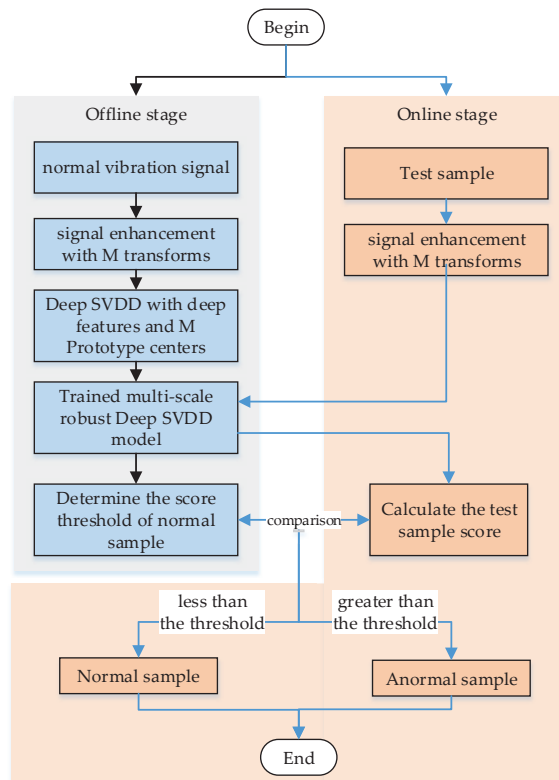


Figure 1. Flowchart of the Robust Multi-scale Deep-SVDD Model.

3.1. Signal Enhancement

Vibration signal is a special one-dimensional datum and there is no specific neighborhood or order. Thus, traditional geometric transformations such as translation and rotation cannot be performed. In order to enable the transformation-based method to process vibration signal data, we propose a data transformation method for vibration signals. Specifically, we propose two transformations of vibration signals from the perspective of graphics.

3.1.1. Horizontal Scaling

First, we crop the length of $p\%$ ($0 \leq p \leq 50$) from either end of the original signal. To ensure the same dimension of the feature space after transformation, we use the resampling method to sample the cropped signal to the length of the original signal, which is equivalent to stretching the original signal in the horizontal direction from a graphical point of view. In addition, for a vibration signal sample, to reduce information loss, we cut the two ends of the signal to obtain two sets of data of equal length, which are used as the two channels of the transformed data, for signal samples are displayed in the same feature space at different

scales. As shown in Figure 2, the original signal length is 1280, the cropping parameter p is set to 30, and the two channels of the transformed sample are obtained by cropping from the left and right ends respectively.

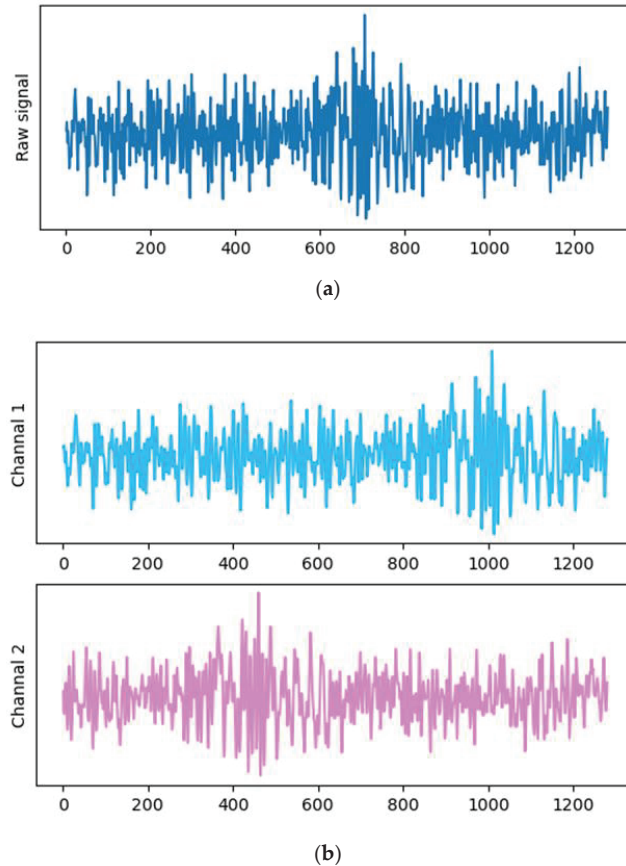


Figure 2. Schematic diagram of the horizontal scaling of the vibration signal with (a) the original signal and (b) two transformed channel signals.

3.1.2. Vertical Scaling

We set scaling parameters $0 < \alpha < 1$, transforming the original signal as follows:

$$f(x_i) = \begin{cases} (1 + \alpha)x_i, & x_i < 0 \\ (1 - \alpha)x_i, & x_i \geq 0 \end{cases} \quad (3)$$

Vertical scaling does not change the length of the signal and the signal can extend or shorten in the vertical direction from the graph. This transformation is also the display of different scales of the signal. Similarly, to ensure the consistency of the feature space after transformation, the samples after vertical transformation are set to be two channels.

By setting different parameters and different combinations of horizontal and vertical transformation, we can obtain a variety of transformation methods to process the original signal. In this paper, the original vibration signal sample space X is transformed to obtain M subspaces X_1, \dots, X_M . The transformed sample is represented as $T(x, i), \dots, T(x, M)$.

3.2. Prototype Clustering

Prototype clustering is a clustering algorithm that uses the prototype to represent the center of the cluster. The prototype clustering algorithm usually needs to initialize the prototype cluster center and then employs the idea of iterative solution to find the cluster prototype.

Learning Vector Quantization (LVQ) [31] is a typical prototype clustering method. The LVQ algorithm uses prototype vectors to represent clusters. The sample is assumed to be labeled, and then the label information is working as an aid in the iterative optimization process to find the optimal prototype vector, which represents the cluster structure. The high-dimensional clustering space is divided into n clusters, and each prototype vector represents a cluster. The solution steps of the LVQ algorithm are as follows (Algorithm 1):

Algorithm 1 Learning Vector Quantization

| | |
|----------------|---|
| Input: | Sample set $M = (x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$; Suppose that, q is the number of prototype vectors, t_1, t_2, \dots, t_q is the initial category of each prototype vector, and $\eta \in (0, 1)$ is the learning rate. |
| Output: | prototype vector v_1, v_2, \dots, v_q |
| 1: | Initialize the prototype vector v_1, v_2, \dots, v_q . |
| 2: | Loop |
| 3: | Select samples (x_j, y_j) from sample set M randomly. |
| 4: | Calculate the distance between x_j and $v_i (1 \leq i \leq q)$: $d_{ji} = \ x_j - v_i\ _2$ |
| 5: | Find the prototype vector v_i closest to x_j , $i^* = \operatorname{argmin}_{i \in \{1, 2, \dots, q\}} d_{ji}$ |
| 6: | If $y_j = t_{i^*}$ |
| 7: | $v' = v_{i^*} + \eta(x_j - v_{i^*})$ |
| 8: | Else |
| 9: | $v' = v_{i^*} - \eta(x_j - v_{i^*})$ |
| 10: | End if |
| 11: | Update v_i to v' |
| 12: | Until the stop criterion is reached |

The algorithm finally learns a set of prototype vectors. Moreover, each of them represents the center of a certain area, which is equivalent to the center point of Voronoi division in space geometry. This center point is the center point of transformed sample in the neural network feature space.

3.3. Distance-Based Cross Entropy Loss

The original sample space X undergoes M transformations to obtain the transformed sample $T(x, i), \dots, T(x, M)$. For each transformed sample $T(x, i)$, calculate the following conditional probability:

$$p(T(x, i) \in X_i) = \frac{e^{-\|E_\theta(T(x, i)) - c_i\|_2^2}}{\sum_{i=1}^M e^{-\|E_\theta(T(x, i)) - c_i\|_2^2}} \quad (4)$$

where E_θ is the network for feature extraction, c_i is the center of X_i .

The distance-based cross entropy (dce) is expressed as:

$$\text{loss}_{dce} = -\log p(T(x, i) \in X_i) \quad (5)$$

Minimizing distance-based cross-entropy loss can map data samples to the class feature space near the prototype center, and improve the separability between classes. Compared with the softmax loss in traditional neural networks, it is more robust.

3.4. Robust Multi-Scale Deep-SVDD

The main idea of the robust multi-scale Deep-SVDD method proposed in this section is to perform data enhancement on the normal samples in the single-class anomaly detection and generate multiple transformations to construct SVDD hyper spheres, and use each transformation in multiple Deep-SVDD hyper spheres. The comprehensive score is used to measure the degree of sample abnormality.

First, to improve the robustness of feature extraction, we select robust deep auto-encoding as the main framework, in which the robust deep auto-encoding encoder is conducted for feature extraction, so as to map the original samples to the low-rank feature space. Second, the learning vector method is used to find out the prototype centers c_1, \dots, c_M of the transformed M samples subspace in the robust deep auto-encoder low-rank space. On this basis, the Deep-SVDD center loss is added, so that all normal samples are as close as possible to the center of each prototype, and the intra-class aggregation degree of each subspace is constrained. The final optimization function is as follows:

$$\min_{\theta, S} \|L_D - D_\theta(E_\theta(L_D))\|_2^2 + loss_{dce} + \mu \|E_\theta(L_D) - c_i\|_2^2 + \lambda \|S\|_{2,1} \quad (6)$$

$$s.t. \quad X - L_D - S = 0$$

where $\mu > 0$, $\lambda > 0$ are regularization coefficient. Increasing the value of μ will make the normal sample features move closer to the center of each prototype, and vice versa, it will weaken the effect of the features gathering to the center.

3.5. Calculation of Anomaly Score

After the above steps, the training model can extract features from the input sample data after specific transformations and obtain the corresponding set of prototype centers c_1, \dots, c_M , and then we can measure the degree of abnormality of the test sample. In the test stage, the test sample x undergoes M transformations to obtain the transformed sample $T(x, i), \dots, T(x, M)$. Put the transformed samples into a robust deep self-encoding encoder to extract features, according to Formula (4), the distance between all transformed samples and the centers of all prototypes is calculated to obtain the probability that they are located in their respective subspaces. Then the probability of the test normal sample x is the product of the probabilities that all the transformed samples are located in their respective subspaces, and the final anomaly score is expressed as:

$$\text{Score}(x) = -\sum_i^M \log P(T(x, i) \in X_i) \quad (7)$$

where the score represents the abnormal score of test sample x . The higher the score, the more abnormal.

Finally, for the bearing incipient fault detection, we need to determine a threshold for the abnormal score of a normal sample to determine whether the calculated abnormal score of the test sample meets the abnormal standard, that is, whether the bearing operating state is abnormal. In this paper, the maximum value of the training data anomaly score is directly used as the threshold standard.

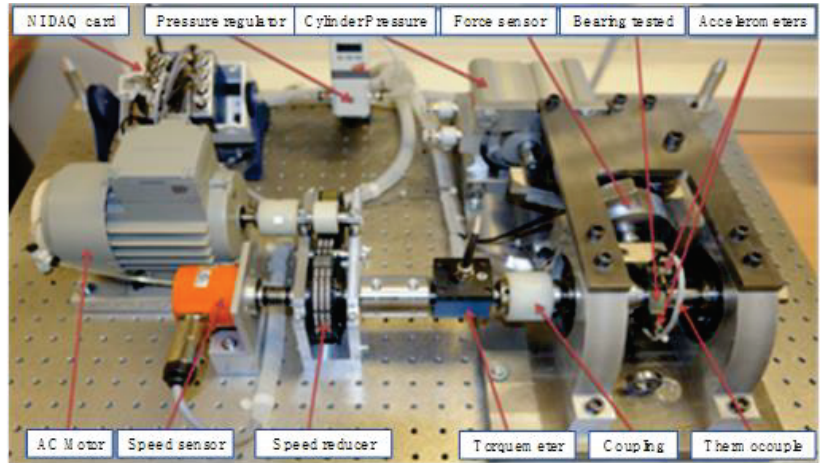
4. Experiment

Experiments on the IEEE PHM Challenge and XJTU-SY datasets are performed to verify the effectiveness of the proposed method. The programming environment is Python 3.6.0, Guido van Rossum, Beijing, China. The computer used in the experiment is configured with i5-8400 processor and 16 G memory.

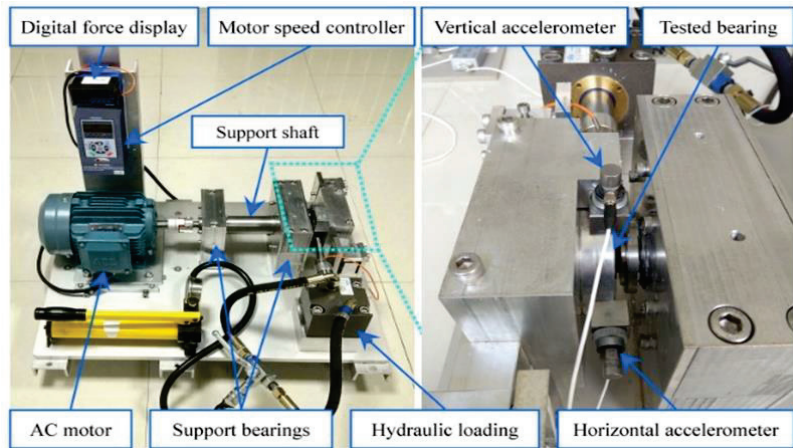
4.1. Dataset Introduction

IEEE PHM Challenge 2012 dataset is collected from the PRONOSTIA platform (shown in Figure 3a) [32], which specially designed and implemented by the AS2M department

of the French FEMTO-ST Institute. It provides the entire life cycle data of rolling bearings through accelerated life degradation experiments. Bearings are working under three different working conditions in these experiments, (1) the engine speed is 1800 rpm and the load is 4000 N, (2) the engine speed is 1650 rpm and the load is 4200 N, (3) the engine speed is 1500 rpm and the load is 5000 N.



(a)



(b)

Figure 3. Test platforms of (a) PRONOSTIA [32] and (b) XJTU-SY [33].

The XJTU-SY dataset is provided by the Institute of Design Science and Basic Component at Xi'an Jiaotong University (XJTU) [33], China and the Changxing Sumyoung Technology Co., Ltd. (SY), Zhejiang province, China. The platform is shown in Figure 3b. Three kinds of experimental working conditions were designed in this experiment, and five bearings were tested in each working condition. (1) The engine speed is 2100 rpm and the load is 12 kN. (2) The engine speed is 2250 rpm and the load is 11 kN. (3) The speed is 2400 rpm and the load is 10 kN.

4.2. Model Parameter Settings

The same data transformation (Horizontal and Vertical scaling) is conducted in the experiments. The parameter of Horizontal scale p is set to be 16, the value set is $\{0, 2, 4, 6, 8, 11, 14, 17, 20, 23, 27, 31, 35, 39, 43, 47\}$. The parameter of vertical scale α is set to be 3, and the value set is $\{0, 0.3, 0.7\}$. There are 48 combinations of these two transformations. The neural network structure used in the experiment is a deep residual network [34]. In the multi-scale robust Deep-SVDD, $\mu = 0.0001$, $\lambda = 0.5$, $\gamma = 0.002$, the training iteration number is 100, and the size of each training batch is 8.

4.3. Incipient Fault Detection Results

Bearing 1_2 and bearing 1_3 in the IEEE PHM Challenge 2012 dataset, as well as the bearing 1_1 and the bearing 2_2 in the XJTU-SY dataset, are the target bearings, as shown in Table 1. The first 100 samples are selected for data transformation to obtain signals of 48 different scales in this experiment, and then the obtained data are put into the multi-scale robust Deep-SVDD model for training to complete the model training. In the test stage, the test samples are first subjected to data transformation, and then input into the model to calculate the abnormal score of each sample. The results of abnormal score and the RMS value are shown in Figures 4 and 5.

Table 1. Experiment dataset.

| Dataset | Sample | Number of Sample | Training Sample | Testing Sample | The Real Sample Point of Incipient Fault | Number of Early Fault Samples |
|---------------------------------|------------------------|------------------|-----------------------|-----------------------|--|-------------------------------|
| IEEE PHM Challenge 2012 dataset | Condition 1 Bearing1_2 | 871 | The first 100 samples | The rest 771 samples | - | 479 |
| | Condition 1 Bearing1_3 | 2375 | The first 100 samples | The rest 2275 samples | 1348th | 1027 |
| XJTU-SY dataset | Condition 1 Bearing1_1 | 1476 | The first 100 samples | The rest 1376 samples | 634th | 839 |
| | Condition 2 Bearing2_2 | 1932 | The first 100 samples | The rest 1832 samples | - | 942 |

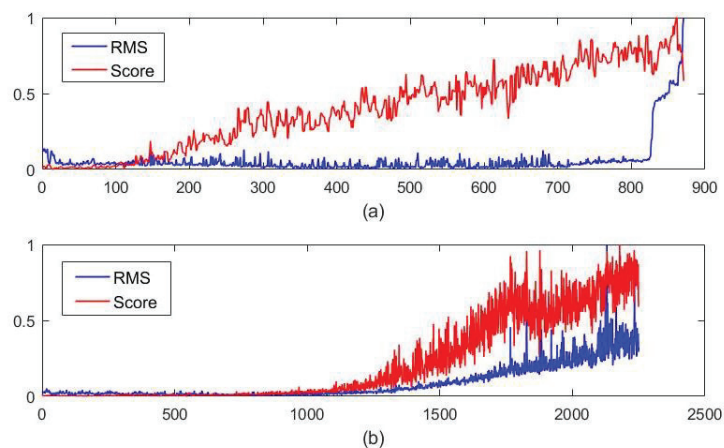


Figure 4. Result comparison of abnormal score and the RMS value with (a) PHM1_2 and (b) PHM1_3.

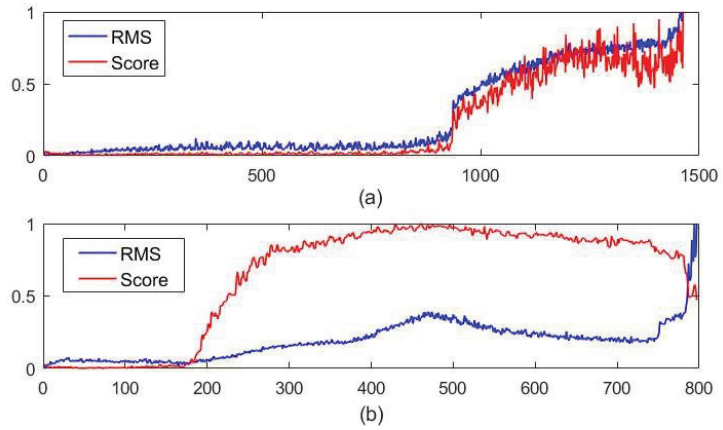


Figure 5. The comparison results of abnormal score and the RMS value with (a) XJTU1_1 and (b) XJTU 2_2.

4.4. Comparative Results

To verify the superiority of the proposed algorithm, comparison between five other widely used methods for incipient fault diagnosis and detection and the proposed method is made. Among them, bandwidth empirical mode decomposition and adaptive multiscale morphological analysis (BEMD-AMMA) [35] is a typical method based on weak signal analysis, local outlier factor (LOF) and isolation forest (iFOREST) are two classic anomaly detection algorithms, meanwhile, Self-Adaptive Deep Feature Matching method (SDFM) [22] and Sparse Dictionary Representation (SDR) [36] methods are also used for comparison.

The Spectrum of bearing fault at different sample points are shown in Figure 6, where Figure 6a is for PHM1_3, and Figure 6a is for XJTU1_1. As we can see from both Figure 6a,b, the fault frequency gradually shows up with time.

We define a deviation rate of incipient fault detection (*DA*) to evaluate the methods' performance mentioned above.

$$DA = \frac{|p_d - p_r|}{p_e - p_r} \times 100\% \tag{8}$$

where p_d is the detected sample point of incipient fault through method, p_r is the real sample point of incipient fault, and p_e is the end sample point of bearing in whole life.

The anomaly detection result is shown in Table 2.

Table 2. Comparison of anomaly detection results.

| Comparison Methods | PHM1_3 | | XJTU1_1 | |
|------------------------|---------------------------|---|---------------------------|---|
| | The Detected Sample Point | Deviation Rate of Incipient Fault Detection | The Detected Sample Point | Deviation Rate of Incipient Fault Detection |
| 1. BEMD-AMMA | 1600 | 55.79% | 1320 | 57.33% |
| 2. LOF | 1236 | 20.35% | 944 | 12.51% |
| 3. iFOREST | 1341 | 30.57% | 1041 | 24.08% |
| 4. SDFM | 1156 | 12.56% | 1137 | 35.52% |
| 5. SRD | 1160 | 12.95% | 1013 | 20.74% |
| 6. The proposed method | 997 | 2.9% | 826 | 1.55% |

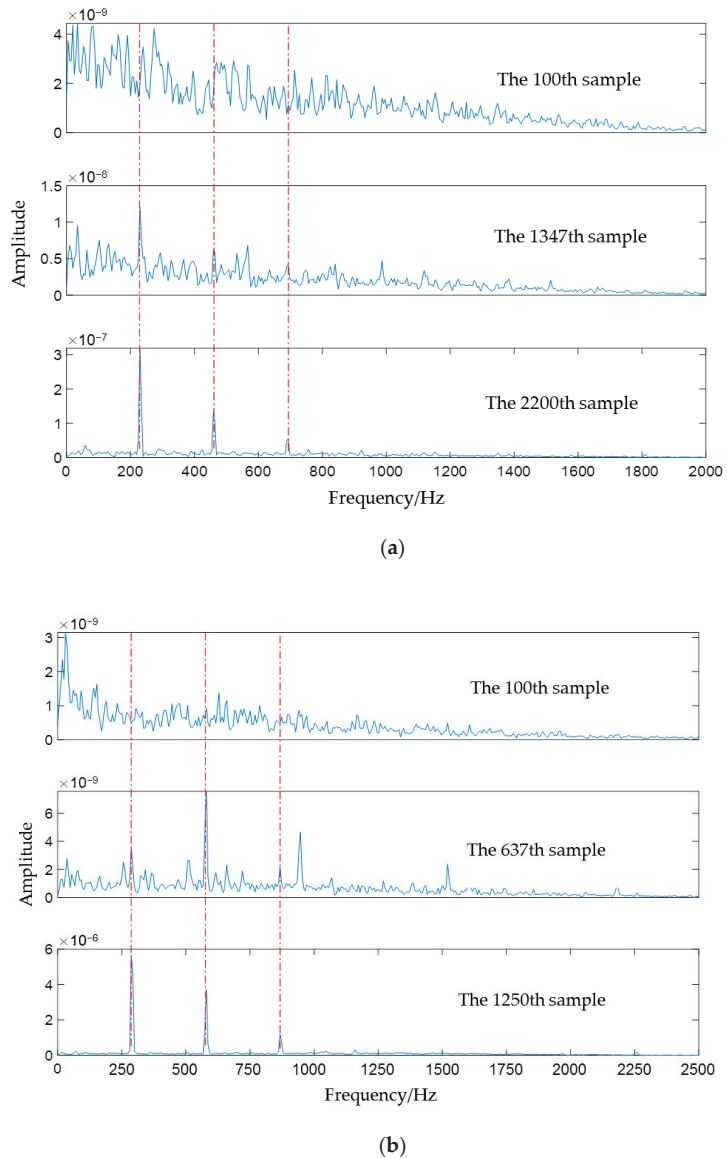


Figure 6. Spectrum of bearing fault at different sample points (a) PHM1_3 (b) XJTU1_1.

As shown in Table 2, the detection result of proposed method is the best one in the comparison. It indicates that employing multi-scale signal samples can enhance feature representation and make incipient fault more sensitive. The robust low-rank deep features extracted by multi-scale robust Deep-SVDD hyper sphere model have strong anti-noise ability for signal fluctuations. Thus, the stability and accuracy of detection results are relatively high.

5. Conclusions

This paper proposes a multi-scale robust incipient fault detection method of rolling bearing with data enhancement. The data enhancement technology is incorporated into

the framework of the robust deep auto-encoding network, to improve the anti-noise ability. It makes the extracted features more robust. Moreover, the constructed robust multi-scale Deep-SVDD model is with good stability by adopting the multi-scale vibration signal features. From the experimental results, the proposed method is more sensitive to incipient faults and has lower false alarm number. The proposed method significantly improves the performance of incipient fault detection of rolling bearings.

Author Contributions: Conceptualization, Methodology, Software, Investigation, Formal Analysis, Writing—Original Draft, Writing—review and editing, L.K.; Visualization, Software, Writing—Original Draft, J.C.; Validation, Investigation and Supervision, Y.Q.; Conceptualization, Data Curation, Supervision, Writing—Review and Editing, Project Administration, W.M. 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: Not applicable.

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

References

- Mao, W.; Feng, W.; Liang, X. A novel deep output kernel learning method for bearing fault structural diagnosis. *Mech. Syst. Signal Process.* **2019**, *117*, 293–318. [\[CrossRef\]](#)
- Tsui, K.; Chen, N.; Zhou, Q.; Hai, Y.; Wang, W. Prognostics and Health Management: A review on data driven approaches. *Mathematical. Probl. Eng.* **2015**, *2015*, 793161. [\[CrossRef\]](#)
- Lei, Y.; Qiao, Z.; Xu, X.; Lin, J.; Niu, S. An underdamped stochastic resonance method with stable-state matching for incipient fault diagnosis of rolling element bearings. *Mech. Syst. Signal Process.* **2017**, *94*, 148–164. [\[CrossRef\]](#)
- Huang, D.; Ke, L.; Mi, B.; Zhao, L.; Sun, G. A new incipient fault diagnosis method combining improved RLS and LMD algorithm for rolling bearings with strong background noise. *IEEE Access* **2018**, *6*, 26001–26010.
- Ma, J.; Wu, J.; Wang, X. Incipient fault feature extraction of rolling bearings based on the MVMD and Teager energy operator. *ISA Trans.* **2018**, *80*, 297–311. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ambrokiewicz, B.; Syta, A.; Gassner, A.; Anthimos, G.; Grzegorz, L.; Nicolas, M. The influence of the radial internal clearance on the dynamic response of self-aligning ball bearings. *Mech. Syst. Signal Process.* **2022**, *171*, 108954. [\[CrossRef\]](#)
- Liu, J.; Xu, Y.; Pan, G. A combined acoustic and dynamic model of a defective ball bearing. *J. Sound Vib.* **2021**, *501*, 116029. [\[CrossRef\]](#)
- Chen, R.; Linlin, L.; Hu, X.; Wu, H. Fault Diagnosis Method of Low-speed Rolling Bearing Based on Acoustic Emission Signal and Subspace Embedded Feature Distribution Alignment. *IEEE Trans. Ind. Inform.* **2021**, *17*, 5402–5410. [\[CrossRef\]](#)
- Maroua, H.; Ramiro, M.; Carlos, F.; Jorge, S.; Mohamed, S.; Mohamed, H. Friction torque in rolling bearings lubricated with axle gear oils—ScienceDirect. *Tribol. Int.* **2018**, *119*, 419–435.
- Liu, R.; Yang, B.; Zhang, X.; Wang, S.; Chen, X. Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis. *Mech. Syst. Signal Process.* **2016**, *75*, 345–370. [\[CrossRef\]](#)
- Li, F.; Wang, J.; Chyu, M.K.; Tang, B. Weak fault diagnosis of rotating machinery based on feature reduction with Supervised Orthogonal Local Fisher Discriminant Analysis. *Neurocomputing* **2015**, *168*, 505–519. [\[CrossRef\]](#)
- Ocak, H.; Loparo, K.A. A new bearing fault detection and diagnosis scheme based on hidden Markov modeling of vibration signals. In Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, Salt Lake City, UT, USA, 7–11 May 2001; Volume 5, pp. 3141–3144.
- Jia, F.; Lei, Y.; Guo, L.; Lin, J.; Xing, S. A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines. *Neurocomputing* **2018**, *272*, 619–628. [\[CrossRef\]](#)
- Shao, H.; Jiang, H.; Zhang, X.; Niu, M. Rolling bearing fault diagnosis using an optimization deep belief network. *Meas. Sci. Technol.* **2015**, *26*, 115002. [\[CrossRef\]](#)
- Mao, W.; He, J.; Zuo, M.J. Predicting Remaining Useful Life of Rolling Bearings Based on Deep Feature Representation and Transfer Learning. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 1594–1608. [\[CrossRef\]](#)
- Mao, W.; Liu, Y.; Ding, L.; Safian, A.; Liang, X. A New Structured Domain Adversarial Neural Network for Transfer Fault Diagnosis of Rolling Bearings. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–13.
- Zhang, E.; Li, M.; Yiu, S.; Du, J.; Zhu, J.; Jin, G. Fair hierarchical secret sharing scheme based on smart contract. *Inf. Sci.* **2021**, *546*, 166–176. [\[CrossRef\]](#)
- Lu, W.; Li, Y.; Cheng, Y.; Meng, D.; Liang, B.; Zhou, P. Early Fault Detection Approach with Deep Architectures. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 1679–1689. [\[CrossRef\]](#)

19. Zhao, R.; Wang, D.; Yan, R.; Mao, K.; Shen, F.; Wang, J.J. Machine Health Monitoring Using Local Feature-Based Gated Recurrent Unit Networks. *IEEE Trans. Ind. Electron.* **2018**, *65*, 1539–1548. [[CrossRef](#)]
20. Shao, H.; Xia, M.; Han, G.; Zhang, Y.; Wan, J. Intelligent Fault Diagnosis of Rotor-Bearing System Under Varying Working Conditions With Modified Transfer Convolutional Neural Network and Thermal Images. *IEEE Trans. Ind. Inform.* **2021**, *17*, 3488–3496. [[CrossRef](#)]
21. Chen, S.; Meng, Y.; Tang, H.; Tian, Y.; He, N.; Shao, C. Robust Deep Learning-Based Diagnosis of Mixed Faults in Rotating Machinery. *IEEE/ASME Trans. Mechatron.* **2020**, *25*, 2167–2176. [[CrossRef](#)]
22. Mao, W.; Chen, J.; Liang, X.; Zhang, X. A New Online Detection Approach for Rolling Bearing Incipient Fault via Self-Adaptive Deep Feature Matching. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 443–456. [[CrossRef](#)]
23. Paschali, M.; Simson, W.; Roy, A.G.; Naeem, M.F.; Gobl, R.; Wachinger, C.; Navab, N. Data augmentation with manifold exploring geometric transformations for increased performance and robustness. *arXiv* **2019**, arXiv:1901.04420v1.
24. Mikóajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the 2018 International Interdisciplinary PhD Workshop, Swinoujscie, Poland, 9–12 May 2018.
25. Kang, G.; Dong, X.; Zheng, L.; Yang, Y. PatchShuffle Regularization. *Computer Vision and Pattern Recognition. arXiv* **2017**, arXiv:1707.07103.
26. Goodfellow, I.J.; Shlens, J.; Szegedy, C. Explaining and Harnessing Adversarial Examples. *arXiv* **2014**, arXiv:1412.6572.
27. Gatys, L.; Ecker, A.; Bethge, M. A Neural Algorithm of Artistic Style. *J. Vis.* **2016**, *16*, 326. [[CrossRef](#)]
28. Han, S.; Oh, J.; Jeong, J. Bearing Fault Detection with Data Augmentation Based on 2-D CNN and 1-D CNN. In Proceedings of the BDIOT 2020: 2020 the 4th International Conference on Big Data and Internet of Things, Singapore, 22–24 August 2020.
29. Ruff, L.; Vandermeulen, R.A.; Grnitz, N.; Deecke, L.; Siddiqui, S.A.; Binder, A.; Müller, E.; Kloft, M. Deep One-Class Classification. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 4390–4399.
30. Ruff, L.; Kauffmann, J.R.; Vandermeulen, R.A.; Montavon, G.; Samek, W.; Kloft, M.; Dietterich, T.G.; Müller, K.-R. A Unifying Review of Deep and Shallow Anomaly Detection. *Proc. IEEE* **2021**, *109*, 756–795.
31. Li, R.P.; Masao, M. Proportional learning vector quantization. *J. Jpn. Soc. Fuzzy Theory Syst.* **1998**, *10*, 1129–1134. [[CrossRef](#)]
32. Nectoux, P.; Gouriveau, R.; Medjaher, K. PRONOSTIA: An experimental platform for bearings accelerated life test. In Proceedings of the IEEE International Conference on Prognostics and Health Management, Denver, MA, USA, 1–8 June 2012.
33. Wang, B.; Lei, Y.; Li, N.; Li, N. A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings. *IEEE Trans. Reliab.* **2020**, *69*, 401–412. [[CrossRef](#)]
34. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
35. Li, Y.; Xu, M.; Liang, X.; Huang, W. Application of Bandwidth EMD and Adaptive Multiscale Morphology Analysis for Incipient Fault Diagnosis of Rolling Bearings. *IEEE Trans. Ind. Electron.* **2017**, *64*, 6506–6517. [[CrossRef](#)]
36. Guo, X.; Liu, S.; Li, Y. Fault Detection of Multi-mode Processes Employing Sparse Residual Distance. *Acta Autom. Sin.* **2019**, *45*, 617–625.

Article

Bearing Fault Diagnosis Using Piecewise Aggregate Approximation and Complete Ensemble Empirical Mode Decomposition with Adaptive Noise

Lei Hu ^{1,2,*}, Ligui Wang ¹, Yanlu Chen ¹, Niaoqing Hu ³ and Yu Jiang ³¹ College of Railway Transportation, Hunan University of Technology, Zhuzhou 412007, China² Hunan Provincial Key Laboratory of Health Maintenance for Mechanical Equipment, Hunan University of Science and Technology, Xiangtan 411201, China³ Laboratory of Science and Technology on Integrated Logistics Support, National University of Defense Technology, Changsha 410073, China

* Correspondence: hulei0819@foxmail.com

Abstract: Complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) effectively separates the fault vibration signals of rolling bearings and improves the diagnosis of rolling bearing faults. However, CEEMDAN has high memory requirements and low computational efficiency. In each iteration of CEEMDAN, fault vibration signals are added with noises, both the vibration signals added with noises and the added noises are decomposed with classical empirical mode decomposition (EMD). This paper proposes a rolling bearing fault diagnosis method that combines piecewise aggregate approximation (PAA) with CEEMDAN. PAA enables CEEMDAN to decompose long signals and to achieve enhanced diagnosis. In particular, the method first yields the vibration envelope using bandpass filtering and demodulation, then compresses the envelope using PAA, and finally decomposes the compressed signal with CEEMDAN. Test data verification results show that the proposed method is more effective and more efficient than CEEMDAN.

Keywords: rolling bearings; fault diagnosis; piecewise aggregate approximation; CEEMDAN

Citation: Hu, L.; Wang, L.; Chen, Y.; Hu, N.; Jiang, Y. Bearing Fault Diagnosis Using Piecewise Aggregate Approximation and Complete Ensemble Empirical Mode Decomposition with Adaptive Noise. *Sensors* **2022**, *22*, 6599. <https://doi.org/10.3390/s22176599>

Academic Editors: Yongbo Li, Bing Li, Jinchen Ji and Hamed Kalhori

Received: 30 June 2022

Accepted: 24 August 2022

Published: 1 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rolling bearings are one of the most widely used components in rotating machinery. Failure of rolling bearings are one of the most frequent reasons for machine breakdown. Thus, fault diagnosis of rolling bearings is crucial to ensure the operational efficiency and reliability of engineering systems [1,2]. When a fault bearing rotates, a localised defect on the outer or inner race is struck by the rollers, or a localised defect on a roller strikes the inner and outer races. High-frequency resonances are excited and presented as impact transients. The periodicity of the successive impact transients is expressed as characteristic fault frequencies [2]. The vibration of fault bearing is recognised as the modulation between the components of low fault frequency f_F and high natural frequency f_N , as shown in Figure 1. It is the most classic bearing fault diagnosis method to obtain the envelope spectrum or squared envelope spectrum using bandpass filtering and demodulation [3]. Finding the optimal frequency band for filtering is critical for the envelope analysis [4]. Some successful tools, such as fast Kurtogram [5], the improved Kurtogram based on wavelet packet transform [6], protruogram [7], and Autogram [8], have been developed for finding the optimal band.

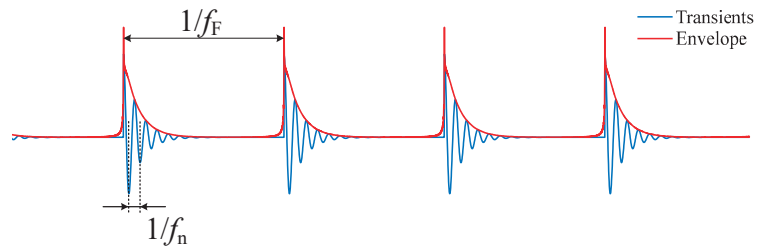


Figure 1. Transient response of bearing defects.

Empirical mode decomposition (EMD) is another widely used method for bearing fault diagnosis. EMD decomposes a signal into a set of intrinsic mode functions (IMFs) and a residue signal [9]. The IMFs are narrow-band components and indicate the natural oscillatory mode imbedded in the original signal [10]. As EMD is effective for nonlinear, non-stationary signals with both Gaussian and non-Gaussian noise, it has been applied with success in different fields, including bearing fault diagnosis [11], planetary gearbox fault diagnosis [12], railway structural wavelength identification [13], automatic sleep scoring [14], etc. However, EMD suffers from endpoint effects and mode mixing. As for mode mixing, a single IMF consists of signals of widely disparate scales, or a signal of a similar scale resides in different IMF components [15]. The mode mixing is the major drawback of the EMD. Ensemble EMD (EEMD) is developed to suppress mode mixing by adding assisted noises to improve the extrema distribution of the signal [16]. However, the IMFs generated by EEMD contain residual noise, and different numbers of IMFs can be generated as different assisted noises are added to the signal to be decomposed. In order to solve the problem that the IMFs are contaminated by residue noise, complementary EEMD (CEEMD) is presented via adding noises in pairs with opposite signs to the targeted signal [17,18]. However, the completeness property is not proven, and different noisy copies of the signal can produce a different number of modes. How to choose proper parameters is also a problem for CEEMD. A further improved algorithm named CEEMD with adaptive noise (CEEMDAN) is proposed to solve the problem of incomplete decomposition by adding particular noise to the signal, which in turn reduces the residual noise in the IMFs [19]. CEEMDAN has been applied in the fields of biomedical engineering [20], energy economics [21], and fault diagnosis [22,23]. In each iterative layer of CEEMDAN, N signals added with noises, as well as the N assisted noises (N is the number of overall averages), are decomposed. Thus, CEEMDAN takes up a lot of memory, and is of low computational efficiency especially for long signal analysing.

A longer signal brings more robust information. For a signal of $x(t)$, its Fourier transform is:

$$F(\omega) = \mathcal{F}[f(t)] = \int_{t_1}^{t_2} f(t)e^{-j\omega t} dt. \quad (1)$$

It can be seen from Equation (1) that a frequency component reflects the average energy of the periodic component over the entire test period of $t_2 - t_1$. Theoretically, the longer the signal is, the more times that a component is averaged, and the clearer the spectrum will be. Figure 2a shows the simulation signal of a bearing with background noise. The signal length is $L = 100$ s and the signal noise ratio (SNR) is -18.091 dB. The fault characteristic frequency is 15 Hz. Signals with a length of 2 s, 10 s, 30 s, and 100 s are selected for envelope analysis. The corresponding envelope spectra yielded are shown in Figure 2b–e, respectively. It can be seen that the harmonics of fault frequency, which cannot be seen in the spectrum of 2 s, can be seen in the spectra of 10 s, 30 s, and 100 s. Although the harmonics of 10 s, 30 s, and 100 s have nearly equal amplitudes, the harmonics become increasingly clearer from Figure 2c–e, as the longer the signal length is, the better the background noise is reduced.

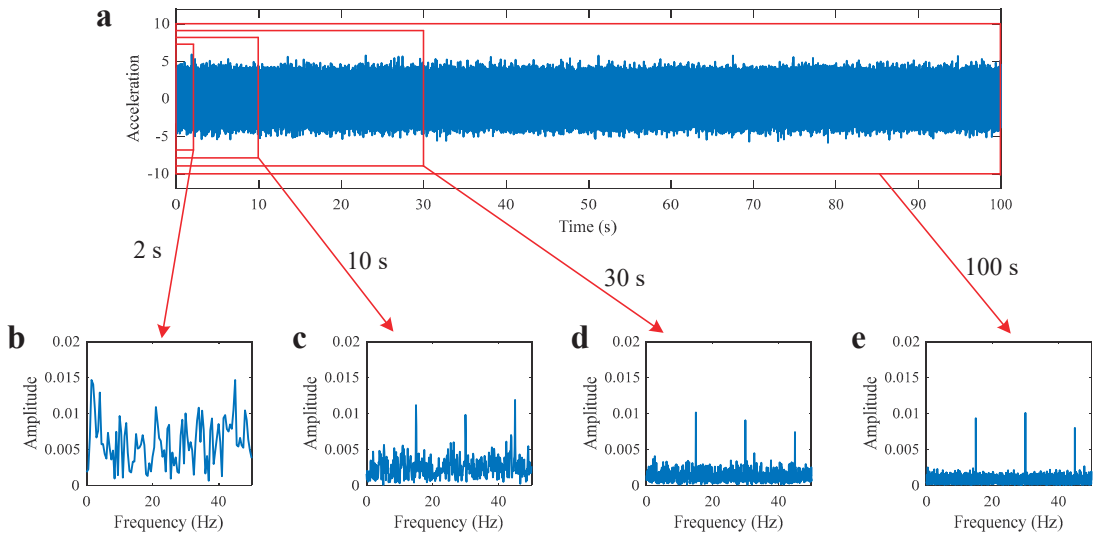


Figure 2. Fault bearing simulation signal and its envelope spectra: (a) Time domain signal with noise; (b) envelope spectrum of 2 s signal; (c) envelope spectrum of 10 s signal; (d) envelope spectrum of 30 s signal; (e) envelope spectrum of 100 s signal.

However, longer signals of high sampling frequency also increase the requirement of computing hardware, which can be a challenge especially for the application cases of edge computing. Particularly, as the natural frequency f_n is as high as thousands (or even tens of thousands) of Hz, the sampling frequency of bearing vibration, f_s , is set to be tens of thousands of Hz according to the Nyquist sampling theorem. Thus, it is natural to compress the signal before processing it using algorithms of high complexity. The technique of compressed sensing achieves data acquisition and compression at the same time. The measurements that compressed sensing obtains are nonadaptive linear projections of the original signals. And the original signals can be reconstructed with the measurements using recovery algorithms [24]. Compressed sensing is originally used for image processing in the fields of medical imaging [25–27], radar imaging [28,29], astronomy [30,31], face recognition [32,33], etc. Compressed sensing is also introduced for machinery fault diagnosis to obtain sparse representation of original signals and to extract fault features from the compressed signals [34–36]. The major drawback of compressed sensing for fault diagnosis is that the compression is not supervised with prior knowledge. Some classical diagnosis methods, such as envelope analysis and EMD, are not applicable any more for the compressed signals. In addition, loss of fault information is inevitable when reconstructing the original signals from the compressed signals.

Piecewise aggregate approximation (PAA) is a far easier method that can be used for signal compression [37,38]. An improved PAA is proposed to take fluctuating trends into account as well [39]. PAA first divides the time series into N segments equally and uses the average of each segment as an approximate representation of that segment. In this process, the original time series with L samples is compressed into a signal of N samples, which can be regarded as a process of dimensionality reduction. The equivalent sampling frequency of the compressed signal is $f_{ES} = f_s \times N/L$, where f_s is the sampling frequency of the original signal. Thus, there is information loss for components whose frequencies are larger than $f_{ES}/2.56$.

In order to obtain reliable diagnostic results using long signals, a method combining PAA and CEEMDAN is proposed. In order to overcome the problem that CEEMDAN has large memory requirements and low computational efficiency, PAA is introduced to

compress the signals before decomposing them. Moreover, in order to avoid information loss caused by signal compression, the traditional envelope analysis method is applied and PAA is performed on the envelopes instead of the original signals. Validations are carried out with signals collected from real rolling bearings.

2. Methodology

2.1. Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)

CEEMDAN is an improved algorithm of EMD and EEMD, which overcomes the shortcomings of EEMD, as mentioned in Section 1. The flow chart of CEEMDAN is shown in Figure 3.

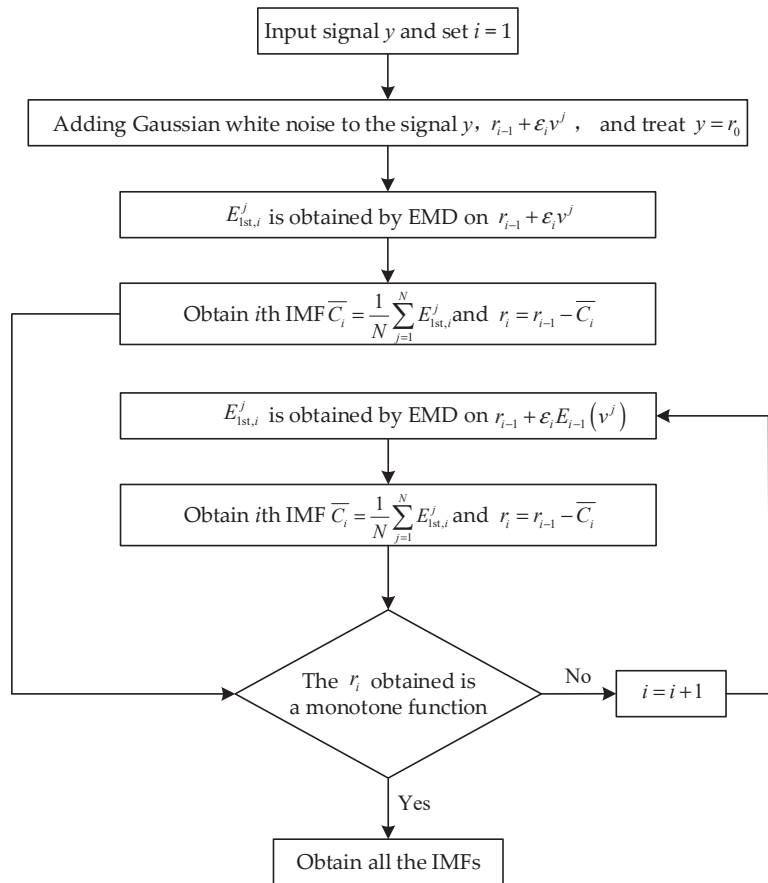


Figure 3. Flow chart of the CEEMDAN algorithm.

Assuming y is the signal to be decomposed, CEEMDAN is performed to decompose the signal y , and the IMF obtained by layer i decomposition is expressed as \bar{C}_i , $i = 1, 2, \dots, I$, where I is the number of layers of decomposition, and the decomposition steps are as follows:

- (1) First layer decomposition, i.e., $i = 1$.

① Adding white noise v^j to the signal of y yields a new signal of $y + \varepsilon_{i=1}v^j$, where $j = 1, 2, \dots, N$, N is the number of adding white noise, and $\varepsilon_{i=1}$ is the amplitude of white noise.

② Decomposing the new signal of $y + \varepsilon_{i=1}v^j$ with EMD yields a series of IMFs, and the first IMF is presented as $E_{1st,i=1}^j$.

③ Ensemble averaging of N IMFs $E_{1st,i=1}^j$ yields the i th ($i = 1$) IMF of CEEMDAN:

$$\overline{C_{i=1}} = \frac{1}{N} \sum_{j=1}^N E_{1st,i=1}^j \tag{2}$$

④ Removing the first IMF of $\overline{C_{i=1}}$ from y yields the residual of $r_{i=1}$:

$$r_{i=1} = y - \overline{C_{i=1}} \tag{3}$$

(2) Second layer decomposition, $i = 2$.

① Decomposing v^j with EMD yields a series of IMFs, the first of which is presented as $E_1(v^j)$. Adding $E_1(v^j)$ as noise to the residual r_{i-1} yields a new signal of $r_{i-1} + \varepsilon_i E_1(v^j)$.

② Decomposing the new signal $r_{i-1} + \varepsilon_i E_1(v^j)$ with EMD yields a series of IMFs, the first of which is presented as $E_{1st,i}^j$.

③ Ensemble averaging of N IMFs $E_{1st,i}^j$ yields the i th IMF of CEEMDAN:

$$\overline{C_i} = \frac{1}{N} \sum_{j=1}^N E_{1st,i}^j \tag{4}$$

④ Removing the i th IMF of $\overline{C_i}$ from r_{i-1} yields the residual of r_i :

$$r_i = r_{i-1} - \overline{C_i} \tag{5}$$

(3) The above steps are repeated until the residual signal obtained is a monotone function and cannot be further decomposed, at which point the algorithm ends. At last, the signal to be decomposed is presented as:

$$y = \sum_{i=1}^I \overline{C_i} + r_I \tag{6}$$

in which I is the number of IMFs and r_I is last residual signal.

2.2. Piecewise Aggregate Approximation

It can be seen that CEEMDAN has large memory requirements and low computational efficiency, as in each iteration of CEEMDAN, tens of fault vibration signals added with assisted noises, as well as the assisted noises, are decomposed with classical EMD. To solve the problem of low computational efficiency, PAA is introduced to compress the signals before performing CEEMDAN.

PAA compresses a large amount of time series data while keeping as many original features of the data as possible. Assuming that the test signal is $x = \{x_i\}$, the sampling frequency is f_s , and the signal length is L . PAA defines a constant window w , then divides the sample sequence x into N equal segments, $N = \lfloor L/w \rfloor$, and finally calculates the mean of each segment:

$$p_n = \frac{1}{w} \sum_{i=w(n-1)+1}^{nw} x_i, n = 1, 2, \dots, N \tag{7}$$

The new sequence $p = (p_1, p_2, \dots, p_N)$ is the obtained compressed signal. It can be seen that the equivalent sampling frequency of the compressed signal is $f_{ES} = f_S/w$. The larger w is, the smaller the samples that the compressed signal has obtained.

2.3. Diagnosis Flowchart

Figure 4 shows the flow chart of the proposed method, which consists of five main steps: optimal band selection for filtering, bandpass filtering and demodulation, PAA, CEEMDAN, and spectra analysis. The steps are depicted as follows:

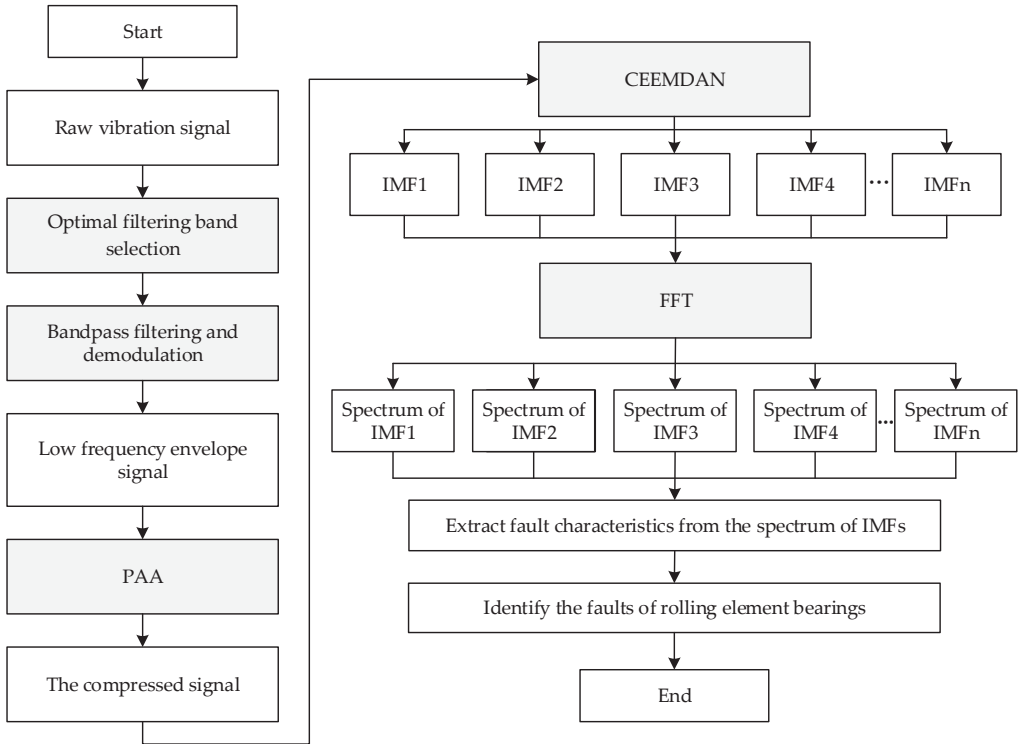


Figure 4. Flow chart of the proposed method.

(1) Optimal filtering band selection.

In order to enhance the modulation signal of low fault frequency and high natural frequency, finding an optimal resonance band for bandpass filtering is critical. The fast Kurtogram, which finds the optimal band according to the kurtosis of the filtered time signal in different filter banks, has been proven to be a practical tool in bearing fault diagnosis. Thus, the fast Kurtogram is introduced for optimal filtering band selection.

(2) Bandpass filtering and demodulation.

Bandpass filtering enhances the modulation signal of low fault frequency and high natural frequency, while demodulation obtains the envelope signal y of low fault frequency, $y = |x + iH(x)|$, where x is the filtered signal and $H(x)$ is the Hilbert transform of x . The envelope consists of components of low frequencies, including the harmonics of fault frequencies. As the fault frequencies are far smaller than the natural frequency, the envelope can be compressed to obtain a signal whose equivalent sampling frequency is far smaller than the original sampling frequency.

(3) Signal compression.

PAA is introduced to compress the envelope yielded in the second step. PAA first divides the envelope into N segments of equal length w , $N = \lfloor L/w \rfloor$, where L is the length of the envelope. Then, PAA uses the mean p_i of each segment as an approximate representation of the segment. The obtained compressed signal is $\mathbf{p} = \{p_i\}$.

The window size, or the segment length, w , is the only unknown parameter of PAA. In addition, w can be set according to the requirement for the equivalent sampling frequency f_{ES} of the compressed signal. As for the envelope of the bearing fault signal, the interesting components are the harmonics of bearing fault characteristic frequencies, which include the ball pass frequency of outer race f_{BPFO} , the ball pass frequency of inner race f_{BPFI} , the ball spin frequency f_{BS} , and the cage frequency f_C . The maximum of the bearing fault characteristic frequencies, $f_{max} = (f_{BPFO}, f_{BPFI}, f_{BS}, f_C)$, is generally f_{BPFI} . According to the Nyquist sampling theorem, the equivalent sampling frequency of the compressed signal should satisfy the condition of $f_{ES} \geq 2.56Zf_{max}$, in which Z is the max order of fault frequency harmonics. Therefore, the window size meets the inequality of:

$$w \leq f_s / (2.56Zf_{max}). \quad (8)$$

(4) CEEMDAN.

Following the steps of CEEMDAN described in Section 2.1, the compressed signal is decomposed, and a series of IMFs is obtained.

(5) Spectrum analysis.

Spectrum analysis is performed on the IMFs obtained to find the interesting IMFs whose frequency bands cover the fault characteristic frequency. Fault diagnosis of rolling element bearing is finally achieved according to the spectra of the interesting IMFs.

2.4. Remarks

PAA is simple, but the envelope waveform of impact transients is well retained in the compressed signal. The reason is that signal compression is supervised with prior knowledge. Particularly, PAA compresses the envelope instead of the original signal. The series of impact transients produced successively by a localised defect are recognised as the modulation between the low-frequency fault components and high-frequency resonances. Thus, bearing vibration is collected with high sampling frequency, and compressing the original vibration signal causes the information loss of the high-frequency resonance; while the diagnostic information in the demodulated envelope is the low-frequency fault components, and the information will be kept in the compressed signal as long as the equivalent sampling frequency is larger than 2.56 multiples of interesting frequencies.

3. Experiment Validation

Bearing fault simulation tests are carried out on the test bench, as shown in Figure 5. The test bench consists of a driving motor, a bearing-supported rotating shaft, an inertia wheel for providing radial load, a belt drive mechanism, a gearbox, a crank connecting rod mechanism, and a reciprocating mechanism. The bearing seeded with defect is mounted in the bearing housing closer to the motor. The seeded defect is a localised crack with both a width and depth of 0.2 mm. The bearing is a deep groove ball contact bearing, the model is MB-ER-10K. The fault characteristic frequencies are $f_{BPFO} = 3.052f_r$, $f_{BPFI} = 4.948f_r$, $f_{BS} = 1.992f_r$, and $f_C = 0.382f_r$, where f_r is the shaft frequency. Vibration signals were collected using accelerometers of the PCB Model 608A11, whose bandwidth is of 0.5 Hz~9 kHz. The sampling frequency was set as $f_s = 25.6$ kHz.

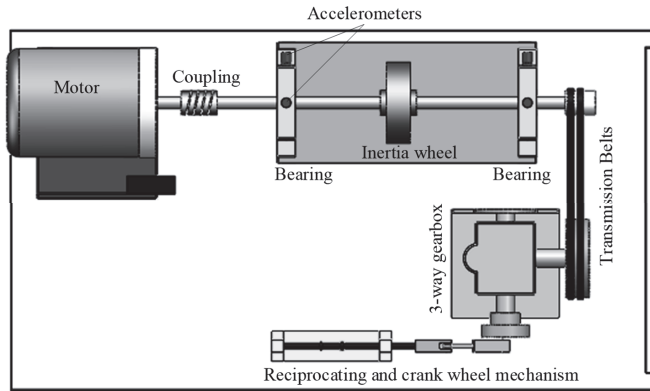


Figure 5. Machinery fault simulation bench.

The maximum of the bearing fault characteristic frequencies is $f_{max} = f_{BPFI}$. Assuming that $Z = 5$ orders of fault frequency harmonics are supposed to be retained in the compressed signals, it yields the condition of the window length, $w \leq 404.20/f_r$, according to Equation (8).

3.1. Validation for Outer Race Defect Case

The vibration signal of a bearing with an outer race defect is shown in Figure 6a. The signal length is $L = 19$ s, the shaft speed is $f_r = 14.1184$ Hz, and the corresponding fault frequency is $f_{BPFO} = 43.0894$ Hz. The proposed method combining CEEMDAN and PAA was used to analyse the signal. Firstly, analysing the signal with fast Kurtogram yields the diagram, as shown in Figure 7. It can be seen that the center frequency of the optimal band is 10,667 Hz, the bandwidth is 4267 Hz, and the corresponding optimal filtering band is 8533.5~12,800.5 Hz. The filtered signal for the optimal filtering band is shown in Figure 6b. The envelope of the filtered signal is shown in Figure 6c.

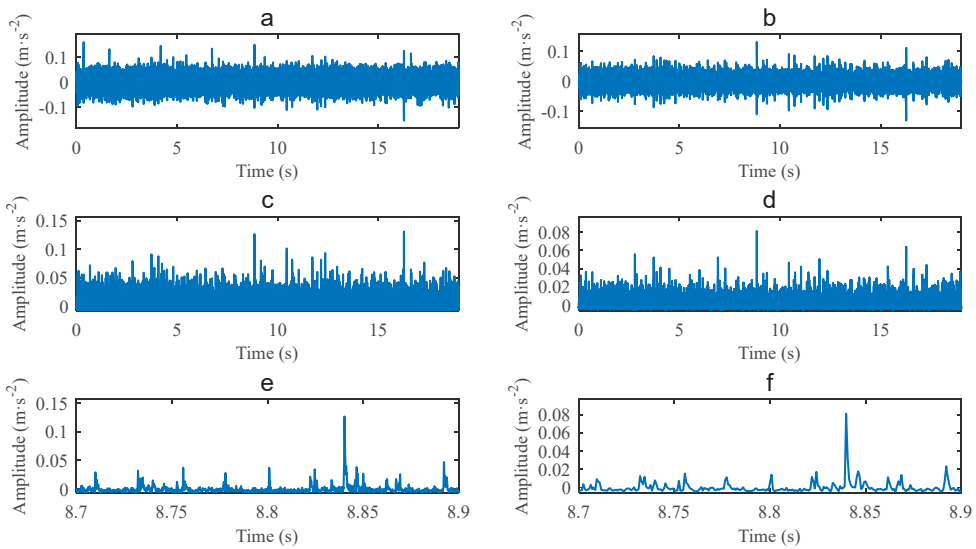


Figure 6. Case 1 for outer race defect: (a) Original signal; (b) filtered signal; (c) envelope; (d) compressed signal; (e) partial enlarged envelope; (f) partial enlarged compressed signal.

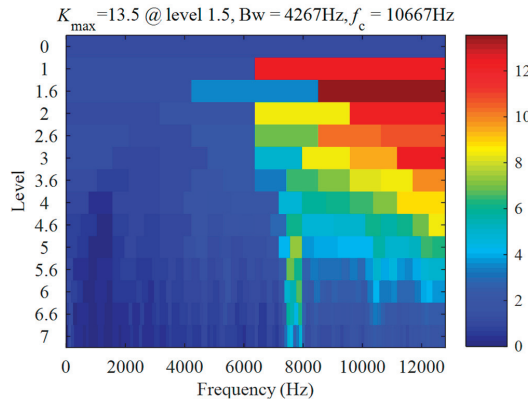


Figure 7. Case 1 for outer race defect: Kurtogram results of the vibration signal.

Performing PAA to compress the envelope yields the result shown in Figure 6d. The window length is set to be $w = 20$ as $w \leq 404.20/f_r$ and $f_r = 14.1184$ Hz. The equivalent sampling frequency of the compressed signal is $f_{ES} = 1.28$ kHz. Partial enlarging the envelope of Figure 6c yields Figure 6e, and partial enlarging the compressed signal of Figure 6d yields Figure 6f. Comparing Figures 6d and 6c, and Figures 6f and 6e, it can be seen that although the compressed signal has smaller amplitudes than the envelope does, they share the same waveform of impulses.

Decomposing the compressed signal with CEEMDAN yields 16 IMFs. The spectra of the IMF 2~IMF 7 are shown in Figure 8, from which the component of f_{BPFO} and its high order harmonics can be seen clearly. Particularly, the spectrum band of IMF 6 is concentrated around f_{BPFO} , IMF 5 is around f_{BPFO} and $2f_{BPFO}$, IMF 4 is around $2f_{BPFO}$ and $3f_{BPFO}$, and IMF 3 is around $3f_{BPFO}$ and $4f_{BPFO}$. These peaks of the fault frequency harmonics illustrate the tested bearing with outer race defects.

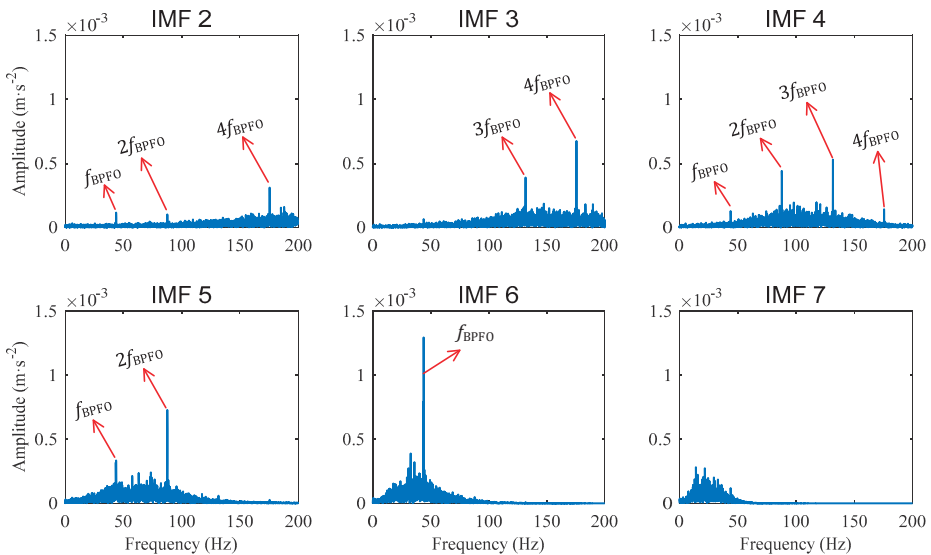


Figure 8. Case 1 for outer race defect: Amplitude spectra of IMFs obtained from compressed signal.

The time length of the signal is $L = 19$ s, and the original signal is of $f_s \times L = 486,400$ samples. CEEMDAN was used to decompose the original signal directly, and the algorithm was still running after 24 h of operation (the computer processor is I5 2.5 g dual-core, and the operating memory is 8G). The compressed signal is of $f_{ES} \times L = 24,320$ samples, which equals the original signal of 0.95 s. Performing CEEMDAN to decompose the compressed signal 10 times, the mean operation time is 359.2 s.

For a segment of the original signal, which is of 0.95 s, it consists of the same 24,320 samples as the compressed signal does. Performing CEEMDAN to decompose the signal segment also yields 14 IMFs. The spectra of IMF 7~IMF 12 are shown in Figure 9. It can be seen that the spectrum band of IMF 10 is concentrated around f_{BPFO} , IMF 9 is around $2f_{BPFO}$, and IMF 8 is around $3f_{BPFO}$ and $4f_{BPFO}$. However, none of these harmonics can be seen from these spectra. The reason is that the signal segment to be decomposed is too short, and the times that these harmonics are averaged during FFT are not enough to reduce background noises.

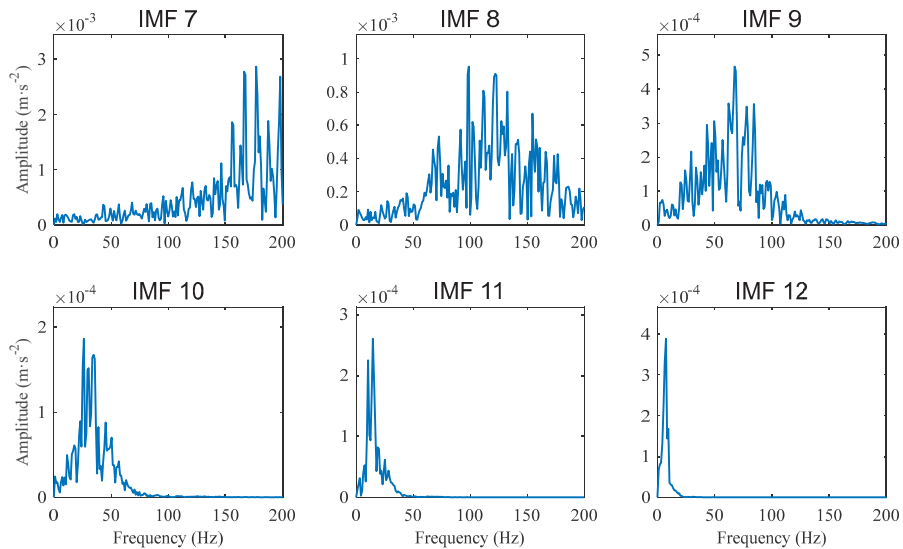


Figure 9. Case 1 for outer race defect: Amplitude spectra of IMFs obtained from an original signal segment that has the same samples as the compressed signal.

Comparing Figures 8 and 9, it can be seen that PAA enables CEEMDAN to decompose long signals and to yield enhanced diagnostic results.

3.2. Validation for Inner Race Defect Case

The vibration signal of a bearing with an inner race defect is shown in Figure 10a. The signal length is $L = 19$ s, the shaft speed is $f_r = 19.7$ Hz, and the characteristic frequency of the inner race fault is $f_{BPFI} = 97.48$ Hz. Analysing the signal with fast Kurtogram yields the result shown in Figure 11. The diagram is different from the one in Figure 7. The same band is selected, with a center frequency of 10,667 Hz, and a bandwidth of 4267 Hz.

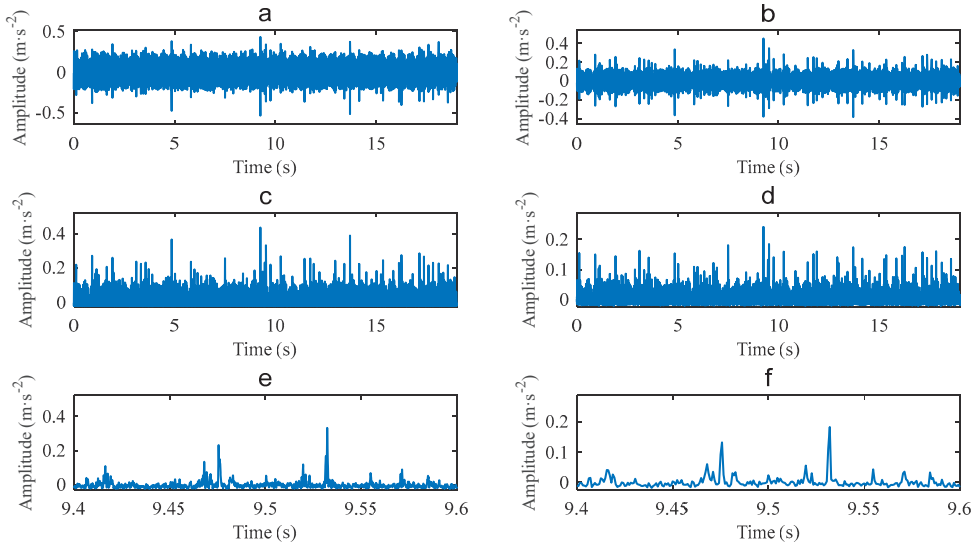


Figure 10. Case 2 for inner race defect: (a) Original signal; (b) filtered signal; (c) envelope; (d) compressed signal; (e) partial enlarged envelope; (f) partial enlarged compressed signal.

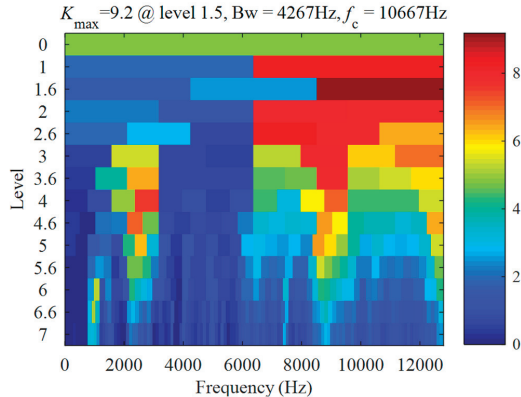


Figure 11. Case 2 for inner race defect: Kurtogram results of the vibration signal.

Figure 10b shows the filtered signal for the filtering band, Figure 10c shows the envelope of the filtered signal, and Figure 10d shows the compressed signal of the envelope obtained with PAA. The window length of PAA is also set to be $w = 20$, which satisfies $w \leq 404.20/f_r$. The equivalent sampling frequency of the compressed signal is also $f_{ES} = 1.28$ kHz. Figure 10e,f shows the partial enlarged envelope and the partial enlarged compressed signal, respectively. It can be seen that the compressed signal keeps the waveform of low frequency components in the envelope.

Decomposing the compressed signal yields 16 IMFs. The spectra of IMF 3~IMF 6 are shown in Figure 12. It can be seen that the spectrum band of IMF 4 is concentrated around the inner race fault frequency f_{BPFI} . The harmonic of f_{BPFI} and its sidebands of $f_{BPFI} \pm f_r$, and $f_{BPFI} + 2f_r$ are clearly presented in the spectrum of IMF 4. The reason for the modulation frequency of f_r is that the inner race defect passes the bearing load zone once every rotation of the shaft, and the transient amplitudes change periodically.

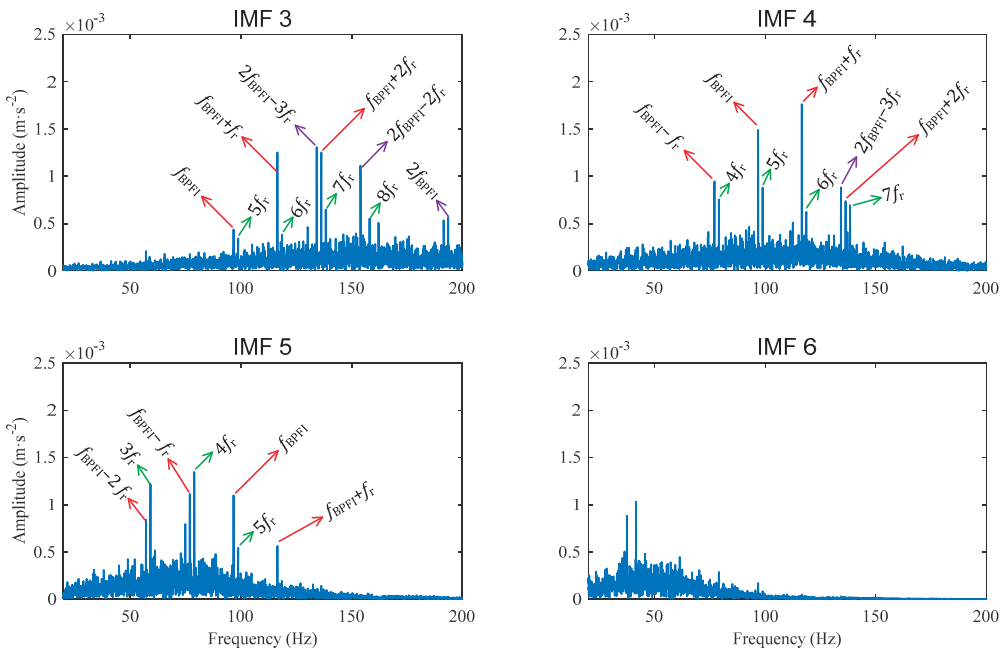


Figure 12. Case 2 for inner race defect: Amplitude spectra of IMFs obtained from the compressed signal.

The spectrum of IMF 5 is concentrated around the frequency of $f_{BPFI} - f_r$, and clearly shows the harmonics of f_{BPFI} , $f_{BPFI} \pm f_r$, and $f_{BPFI} - 2f_r$. The spectrum of IMF 3 is concentrated around the band of $[f_{BPFI}, 2f_{BPFI}]$. The harmonic of the fault frequency f_{BPFI} and its sidebands of $f_{BPFI} + f_r$ and $f_{BPFI} + 2f_r$, as well as the second order fault frequency of $2f_{BPFI}$ and its sidebands of $2f_{BPFI} - 3f_r$ and $2f_{BPFI} - 2f_r$, can be clearly seen from the spectrum. The sideband of $2f_{BPFI} - 3f_r$ can also be seen in the spectrum of IMF 4.

It is worth noting that the characteristic frequency of $f_{BPFI} = 4.948f_r$ is very close to the 5th order harmonic of shaft frequency $5f_r$. Thus, the fault frequency harmonics and their sidebands are very close to the high order shaft frequencies. In any case, the components of f_{BPFI} , $2f_{BPFI}$, and their sidebands illustrate that the tested bearing has inner race defects.

4. Conclusions

In this paper, a rolling bearing fault diagnosis method that combines PAA and CEEM-DAN is proposed. The method firstly extracts the envelope signal from an original signal using bandpass filtering and demodulation, then compresses the envelope with PAA, decomposes the compressed signal with CEEMDAN, and finally investigates the spectra of IMFs. Validation results with real bearings show that the proposed method is effective and efficient.

The interesting components in the original signal for fault diagnosis are the modulation between the fault frequencies and the resonance natural frequencies. The natural frequencies are as high as thousands of Hz, or even tens of thousands of Hz, while the interesting components in the envelope are the fault frequency harmonics demodulated from the original signal. As the fault frequencies are far lower than the natural frequencies, compressing the envelope instead of the original signal avoids information loss.

The spectra of IMFs reflect the average energy over the entire test period. The longer the signal is, the more times the spectra are averaged during FFT, and the better the background noise is reduced. However, in each iteration of CEEMDAN, an IMF is yielded by decomposing tens of signals added with assisted noises, as well as the assisted noises

themselves. Thus, CEEMDAN has large memory requirements and low computational efficiency for long signals. Compressing the envelope with PAA enables the use of CEEMDAN for long signals to achieve enhanced diagnosis.

Author Contributions: Conceptualization and supervision, L.H.; data curation, L.W. and N.H.; investigation, software and validation, L.W. and Y.C.; methodology and writing, L.H. and L.W.; review and editing, L.H. and Y.J. 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 (Grant No. 51575518).

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

References

- Randall, R.B. *Vibration-Based Condition Monitoring: Industrial, Aerospace and Automotive Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
- Hu, L.; Zhang, L.; Gu, F.; Hu, N.; Ball, A. Extraction of the largest amplitude impact transients for diagnosing rolling element defects in bearings. *Mech. Syst. Signal Process.* **2019**, *116*, 796–815. [[CrossRef](#)]
- Abboud, D.; Antoni, J.; Sieg-Zieba, S.; Eltabach, M. Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment. *Mech. Syst. Signal Process.* **2017**, *84*, 200–226. [[CrossRef](#)]
- Hu, L.; Xu, Y.; Gu, F.; He, J.; Hu, N.; Ball, A. Autocorrelation ensemble average of larger amplitude impact transients for the fault diagnosis of rolling element bearings. *Energies* **2019**, *12*, 4740. [[CrossRef](#)]
- Antoni, J. Fast computation of the Kurtogram for the detection of transient faults. *Mech. Syst. Signal Process.* **2007**, *21*, 108–124. [[CrossRef](#)]
- Lei, Y.; Lin, J.; He, Z.; Zi, Y. Application of an improved Kurtogram method for fault diagnosis of rolling element bearings. *Mech. Syst. Signal Process.* **2011**, *25*, 1738–1749. [[CrossRef](#)]
- Barszcz, T.; Jablonski, A. A novel method for the optimal band selection for vibration signal demodulation and comparison with the Kurtogram. *Mech. Syst. Signal Process.* **2011**, *25*, 431–451. [[CrossRef](#)]
- Moshrefzadeh, A.; Fasana, A. The Autogram: An effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis. *Mech. Syst. Signal Process.* **2017**, *105*, 294–318. [[CrossRef](#)]
- Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [[CrossRef](#)]
- Lei, Y.; Lin, J.; He, Z.; Zuo, M.J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2013**, *35*, 108–126. [[CrossRef](#)]
- Yang, Y.; Dejie, Y.U.; Cheng, J. A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *J. Sound Vib.* **2006**, *294*, 269–277.
- Pan, H.; Zheng, J.; Yang, Y.; Cheng, J. Nonlinear sparse mode decomposition and its application in planetary gearbox fault diagnosis. *Mech. Mach. Theory* **2020**, *155*, 104082. [[CrossRef](#)]
- Song, Y.; Liu, Z.; Rønquist, A.; Nåvik, P.; Liu, Z. Contact wire irregularity stochastics and effect on high-speed railway pantograph–catenary interactions. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 8196–8206. [[CrossRef](#)]
- Hassan, A.R.; Bhuiyan, M.I.H. Automatic sleep scoring using statistical features in the EMD domain and ensemble methods. *Biocybern. Biomed. Eng.* **2016**, *36*, 248–255. [[CrossRef](#)]
- Wu, Z.; Huang, N. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [[CrossRef](#)]
- Randall, R.B.; Antoni, J. Rolling element bearing diagnostics—A tutorial. *Mech. Syst. Signal Process.* **2011**, *25*, 485–520. [[CrossRef](#)]
- Imaouchen, Y.; Kedadouche, M.; Alkama, R.; Thomas, M. A frequency-weighted energy operator and complementary ensemble empirical mode decomposition for bearing fault detection. *Mech. Syst. Signal Process.* **2017**, *82*, 103–116. [[CrossRef](#)]
- Zhao, L.; Yu, W.; Yan, R. Gearbox fault diagnosis using complementary ensemble empirical mode decomposition and permutation entropy. *Shock. Vib.* **2016**, *2016*, 3891429. [[CrossRef](#)]
- Torres, M.E.; Colominas, M.A.; Schlotthauer, G.; Flandrin, P. A complete ensemble empirical mode decomposition with adaptive noise. In Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech, 22–27 May 2011; pp. 4144–4147.
- Hassan, A.R.; Bhuiyan, M. Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating. *Biomed. Signal Process. Control.* **2016**, *24*, 1–10. [[CrossRef](#)]
- Afanasyev, D.O.; Fedorova, E.A. The long-term trends on the electricity markets: Comparison of empirical mode and wavelet decompositions. *Energy Econ.* **2016**, *56*, 432–442. [[CrossRef](#)]
- Wang, L.; Shao, Y. Fault feature extraction of rotating machinery using a reweighted complete ensemble empirical mode decomposition with adaptive noise and demodulation analysis. *Mech. Syst. Signal Process.* **2020**, *138*, 106545. [[CrossRef](#)]

23. Lei, Y.; Liu, Z.; Ouazri, J.; Lin, J. A fault diagnosis method of rolling element bearings based on CEEMDAN. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2017**, *231*, 1804–1815. [[CrossRef](#)]
24. Candes, E.J.; Wakin, M.B. An introduction to compressive sampling. *IEEE Signal Process. Mag.* **2008**, *25*, 21–30. [[CrossRef](#)]
25. Candes, E.J.; Romberg, J.; Tao, T. Robust uncertainty principles: Exact signal recognition from highly incomplete frequency information. *IEEE Trans. Info. Theory* **2006**, *52*, 489–509. [[CrossRef](#)]
26. Provost, J.; Lesage, F. The application of compressed sensing for photo-acoustic tomography. *IEEE Trans. Med. Imaging* **2009**, *28*, 585–594. [[CrossRef](#)]
27. Hu, S.; Lustig, M.; Chen, A.P.; Crane, J.; Kerr, A.; Kelley, D.A.; Hurd, R.; Kurhanewicz, J.; Nelson, S.J.; Pauly, J.M.; et al. Compressed sensing for resolution enhancement of hyperpolarized ¹³C flyback 3D-MRSI. *J. Magn. Reson.* **2008**, *192*, 258–264. [[CrossRef](#)]
28. Baraniuk, R.; Steeghs, P. Compressive radar imaging. In Proceedings of the IEEE Radar Conference, Boston, MA, USA, 17–20 April 2007; pp. 128–133.
29. Herman, M.; Strohmer, T. High-resolution radar via compressed sensing. *IEEE Trans. Signal Process.* **2009**, *57*, 2275–2284. [[CrossRef](#)]
30. Bobin, J.; Starck, J.L.; Ottensamer, R. Compressed sensing in astronomy. *IEEE J. Sel. Top. Signal Process.* **2008**, *2*, 718–726. [[CrossRef](#)]
31. Wenger, S.; Magnor, M.; Pihlström, Y.; Bhatnagar, S.; Rau, U. SparseRI: A compressed sensing framework for aperture synthesis imaging in radio astronomy. *Astron. Soc. Pac.* **2010**, *122*, 1367–1374. [[CrossRef](#)]
32. Vo, N.; Vo, D.; Challa, S.; Moran, B. Compressed sensing for face recognition. In Proceedings of the IEEE Symposium on Computational Intelligence for Image Processing, Nashville, TN, USA, 30 March 2009–2 April 2009; pp. 104–109.
33. Andrés, A.M.; Padovani, S.; Tepper, M.; Jacobo-Berlles, J. Face recognition on partially occluded images using compressed sensing. *Pattern Recognit. Lett.* **2014**, *36*, 235–242. [[CrossRef](#)]
34. Sun, J.; Yu, Y.; Wen, J. Compressed-sensing reconstruction based on block sparse Bayesian learning in bearing condition monitoring. *Sensors* **2017**, *17*, 1454. [[CrossRef](#)]
35. Zhang, X.; Hu, N.; Hu, L.; Chen, L.; Cheng, Z. A bearing fault diagnosis method based on the low-dimensional compressed vibration signal. *Adv. Mech. Eng.* **2015**, *7*, 1687814015593442. [[CrossRef](#)]
36. Tang, G.; Hou, W.; Wang, H.; Luo, G.; Ma, J. Compressive sensing of roller bearing faults via harmonic detection from under-sampled vibration signals. *Sensors* **2015**, *15*, 25648–25662. [[CrossRef](#)]
37. Yi, B.K.; Faloutsos, C. Fast time sequence indexing for arbitrary Lp norms. In Proceedings of the 26th VLDB Conference, Cairo, Egypt, 10–14 September 2000; pp. 385–394.
38. Keogh, E.; Chakrabarti, K.; Pazzani, M.; Mehrotra, S. Dimensionality reduction for fast similarity search in large time series databases. *Knowl. Inf. Syst.* **2001**, *3*, 263–286. [[CrossRef](#)]
39. Hung NQ, V.; Anh, D.T. An improvement of PAA for dimensionality reduction in large time series databases. In Proceedings of the Pacific Rim International Conference on Artificial Intelligence: Trends in Artificial Intelligence, Hanoi, Vietnam, 15–19 December 2008; pp. 698–707.

Article

Simultaneous Sensor and Actuator Fault Reconstruction by Using a Sliding Mode Observer, Fuzzy Stability Analysis, and a Nonlinear Optimization Tool

Samira Asadi, Mehrdad Moallem* and G. Gary Wang

School of Mechatronic Systems Engineering, Simon Fraser University, Surrey, BC V3T 0A3, Canada

* Correspondence: mehrdad_moallem@sfu.ca

Abstract: This paper proposes a Takagi–Sugeno (TS) fuzzy sliding mode observer (SMO) for simultaneous actuator and sensor fault reconstruction in a class of nonlinear systems subjected to unknown disturbances. First, the nonlinear system is represented by a TS fuzzy model with immeasurable premise variables. By filtering the output of the TS fuzzy model, an augmented system whose actuator fault is a combination of the original actuator and sensor faults is constructed. An H_∞ performance criteria is considered to minimize the effect of the disturbance on the state estimations. Then, by using two further transformation matrices, a non-quadratic Lyapunov function (NQLF), and `fmincon` in MATLAB as a nonlinear optimization tool, the gains of the SMO are designed through the stability analysis of the observer. The main advantages of the proposed approach in comparison to the existing methods are using nonlinear optimization tools instead of linear matrix inequalities (LMIs), utilizing NQLF instead of simple quadratic Lyapunov functions (QLF), choosing SMO as the observer, which is robust to the uncertainties, and assuming that the premise variables are immeasurable. Finally, a practical continuous stirred tank reactor (CSTR) is considered as a nonlinear dynamic, and the numerical simulation results illustrate the superiority of the proposed approach compared to the existing methods.

Keywords: actuator and sensor faults; TS fuzzy system; sliding mode observer (SMO); H_∞ performance; non-quadratic Lyapunov function (NQLF); `fmincon`; fault reconstruction

Citation: Asadi, S.; Moallem, M.; Wang, G.G. Simultaneous Sensor and Actuator Fault Reconstruction by Using a Sliding Mode Observer, Fuzzy Stability Analysis, and a Nonlinear Optimization Tool. *Sensors* **2022**, *22*, 6866. <https://doi.org/10.3390/s22186866>

Academic Editors: Yongbo Li, Bing Li, Jinchun Ji and Hamed Kalhori

Received: 12 July 2022

Accepted: 7 September 2022

Published: 10 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over the past few decades, the reliability and safety of industrial systems has attracted considerable attention. As a consequence, fault-tolerant control (FTC) has received considerable attention in different fields [1,2]. There are different classifications for FTCs. In general, FTCs are classified into passive and active classifications. Active fault-tolerant controllers compensate for the effects of the occurred faults by using early information obtained from fault detection and isolation (FDI) schemes, which leads to a more flexible dynamic [3]. Consequently, FDI is becoming an attractive topic in different research fields. Observer-based methods are one of the most popular model-based FDIs. The main idea of observer-based FDIs is to construct a residual based on the measured output of the systems or to reconstruct the fault directly. Sliding mode observer (SMO) works based on the second approach, which detects the faults while determining the dynamic behavior [4,5]. SMOs are more insensitive to the unknown uncertainties occurring in the system compared to other observers like unknown input observers (UIOs) [6].

First, SMO observers were developed for linear dynamic systems; however, most actual physical systems are often nonlinear. Currently, lots of SMO-based fault reconstruction methods have been developed for uncertain nonlinear systems. In ref. [7], by considering a filter of the measured output vector, the original system with sensor and actuator faults is transformed into an augmented system with just the actuator fault and unknown inputs.

Nevertheless, the classes of nonlinear systems considered in most of the papers are limited and cannot represent a general model for real systems [8,9].

Takagi–Sugeno (TS) fuzzy models can represent the behavior of nonlinear systems while keeping the simplicity of the linear models. A TS fuzzy representation is a convex nonlinear aggregation of several linear systems. Because the parameters of a TS fuzzy representation satisfy the convex sum, it is interesting to investigate the properties of the TS system based on its local linear vertices. With the advent of TS fuzzy systems, TS-based FDI techniques emerged to tackle a broader range of nonlinear systems [10]. By changing a nonlinear system to a TS system, some local linear systems are created, representing the behavior of the nonlinear system in a specific operating area. These local linear systems can be aggregated by using an interpolation mechanism. Thus, TS fuzzy models can represent the actual nonlinear behavior while maintaining the simplicity of linear models. Thus, an efficient FDI can be obtained by combining the SMO, which is robust to the uncertainties, and the TS fuzzy model, which causes simplicity in the design process. Recently, several researchers have utilized TS-based SMOs for fault detection and isolation in continuous-time and discrete-time systems [11,12]. However, in the methods developed in these articles, it is assumed that the premise variables are measurable, which reduces the applicability of these approaches. To deal with this problem, an FDI approach for stability analysis of the TS fuzzy systems with immeasurable premise variables was proposed in [13,14].

In [15], simultaneous actuator and sensor faults in a nonlinear system represented by a TS fuzzy model are reconstructed by using an SMO and considering H_∞ performance criteria to reduce the effect of disturbance, whereas [16] does the same procedure for the fault reconstructions and both of the exogenous disturbance and the system faults are reconstructed. However, in refs. [15,16] quadratic Lyapunov functions (QLFs) are used to design the observers. By using the QLF for TS fuzzy systems with a large number of fuzzy rules can cause undesired performance or unfeasible solutions. Consequently, refs. [17,18] offered to use a non-quadratic Lyapunov function (NQLF) to design the TS-based SMO for the FDI purposes. In all these papers, a linear optimization approach based on linear matrix inequalities (LMIs) is utilized, making the stability analysis more complex and using some approximations and lemmas to prove the stability conditions.

In this paper, a TS fuzzy-based SMO with immeasurable premise variables is designed to reconstruct simultaneous actuator and sensor faults in a nonlinear system exposed to an unknown disturbance. Then, the states and faults are estimated. The stability of the proposed observer is guaranteed by using the NQLF and `fmincon` as a nonlinear optimization tool in MATLAB. In addition, H_∞ performance criteria are considered to minimize the effect of disturbances and uncertainties on the estimation error and the fault estimations. By using the NQLF, a generalized eigenvalue problem is proposed, which maximizes the admissible Lipschitz constant and minimizes the disturbance effects on the estimation error through a nonlinear optimization problem.

The main advantages of the proposed approach over the existing methods can be summarized as follows:

- Using nonlinear optimization tools instead of LMIs, which results in better accuracy.
- Utilizing NQLF, which leads to less conservative optimization conditions than simple quadratic Lyapunov functions.
- Assuming that the premise variables are immeasurable, which makes the proposed method applicable to a broader class of TS fuzzy systems.

This paper is organized as follows. Section 2 presents a TS fuzzy model with simultaneous actuator and sensor faults and disturbance and how to construct a fictitious system with just an actuator fault. In Section 3, the main results of this paper, including the sliding mode observer design and the sufficient conditions of stability of the estimation errors, are proposed and guarantee the H_∞ performance simultaneously. Section 4 discusses the procedure of the actuator and sensor fault reconstructions. In Section 5, simulation results

are given, and comparisons are discussed. Finally, in Section 6, the concluding remarks are given.

2. Preliminaries

Assume that a continuous-time nonlinear system affected by actuator and sensor faults and disturbance is given as

$$\begin{cases} \dot{x}(t) = f(x(t), u(t), f_a(t), d(x(t), u(t), t)) \\ y(t) = Cx(t) + Nf_s(t) \end{cases}, \tag{1}$$

where $x(t) \in R^n, u(t) \in R^m, y(t) \in R^p, f_a(t) \in R^q, f_s(t) \in R^h$ and $d(x(t), u(t), t) \in R^l$ are the state, input, output, unknown actuator, and sensor faults, and the system uncertainty vectors, respectively. f and g are nonlinear smooth functions. By using sector nonlinearity transformation, the nonlinear model (1) can be replaced by the following TS fuzzy model

$$\begin{cases} \dot{x}(t) = \sum_{i=1}^r \mu_i(\xi(t)) \left\{ \begin{matrix} A_i x(t) + B_i u(t) + M_i f_a(t) \\ + D_i d(x(t), u(t), t) \end{matrix} \right\} \\ y(t) = Cx(t) + Nf_s(t) \end{cases}, \tag{2}$$

where C and N are known full rank matrices with appropriate dimensions. $A_i, B_i, M_i,$ and D_i are real known matrices, r represents the number of fuzzy rules and $\mu_i(\xi(t))$ are the fuzzy membership functions depending on the unmeasurable variable vector $\xi(t)$ and satisfy the following so-called convex sum property

$$\begin{cases} 0 \leq \mu_i(\xi(t)) \leq 1 \\ \sum_{i=1}^r \mu_i(\xi(t)) = 1 \end{cases}. \tag{3}$$

In the rest of the paper, (t) is dropped from the equations, d, μ_i and $\hat{\mu}_i$ denote $d(x, u, t), \mu_i(\xi(t)),$ and $\mu_i(\hat{\xi}(t))$ and the mark $(*)$ denotes the transposed element in a symmetric matrix.

To build a system with just an actuator fault and then use the actuator fault reconstruction concepts, the output is passed through an orthogonal matrix $T_r \in R^{p \times p}$ and an augmented TS system of order $n + h$ can be obtained as

$$\begin{cases} \dot{X} = \sum_{i=1}^r \mu_i \{ A_i X + B_i u + D_i d + M_i f_a + N f_s \} \\ Y = C X \end{cases}, \tag{4}$$

where $X = [x^T \ z^T]^T \in R^{n+h}, Y = [y_1^T \ z^T]^T \in R^p,$ and

$$A_i = \begin{bmatrix} A_i & 0 \\ A_f C_2 & -A_f \end{bmatrix}, B_i = \begin{bmatrix} B_i \\ 0 \end{bmatrix}, D_i = \begin{bmatrix} D_i \\ 0 \end{bmatrix}, M_i = \begin{bmatrix} M_i \\ 0 \end{bmatrix}, N = \begin{bmatrix} 0 \\ A_f N_2 \end{bmatrix}, C = \begin{bmatrix} C_1 & 0 \\ 0 & I_h \end{bmatrix}. \tag{5}$$

$-A_f \in R^{h \times h}$ is an arbitrary stable matrix, $z \in R^h$ and $N_2 \in R^{h \times h}$. T_r can be obtained by QR reduction of the matrix N .

By defining

$$\phi := \sum_{i=1}^r (\mu_i - \hat{\mu}_i) \{ A_i X + B_i u + D_i d + M_i f_a + N f_s \}, \tag{6}$$

where \hat{x} is the estimation of the x , the TS system (4) can be derived as

$$\begin{cases} \dot{X} = \sum_{i=1}^r \hat{\mu}_i \{ A_i X + B_i u + D_i d + M_i f_a + N f_s + \phi \} \\ Y = C X \end{cases}. \tag{7}$$

Moreover, the nonlinear term ϕ is assumed to satisfy the Lipschitz condition as

$$\|\phi\| \leq \gamma \|x - \hat{x}\|, \quad \forall x, \hat{x} \in \mathbb{R}^n. \tag{8}$$

To design a sliding mode observer, some assumptions and lemmas are needed as follows.

Assumption 1.

$$\text{rank}(C[\mathcal{M}_i \ \mathcal{N}]) = q + h \tag{9}$$

Assumption 2.

$$n > p \geq q + h \tag{10}$$

Assumption 3.

$$\text{rank} \begin{bmatrix} sI_{n+h} - \mathcal{A}_i & \mathcal{M}_i & \mathcal{N} \\ C & 0 & 0 \end{bmatrix} = n + 2h + q \tag{11}$$

for all s satisfying $\text{Re}(s) \geq 0$ holds.

Lemma 1.

(a) If Assumptions 1 and 2 are satisfied, then there exist changes of coordinates \bar{T}_i such that

$$\mathcal{A}_i = \begin{bmatrix} \mathcal{A}_{11.i} & \mathcal{A}_{12.i} \\ \begin{bmatrix} \mathcal{A}_{211.i} \\ \mathcal{A}_{212.i} \end{bmatrix} & \mathcal{A}_{22.i} \end{bmatrix}, \mathcal{M}_i = \begin{bmatrix} 0 \\ \mathcal{M}_{2.i} \end{bmatrix}, \mathcal{N} = \begin{bmatrix} 0 \\ \mathcal{N}_2 \end{bmatrix}, \mathcal{D}_i = \begin{bmatrix} \mathcal{D}_{1.i} \\ \mathcal{D}_{2.i} \end{bmatrix}, C = [0 \ T_0], \tag{12}$$

where $\mathcal{A}_{11.i} \in \mathbb{R}^{(n+h-p) \times (n+h-p)}$, $\mathcal{A}_{211.i} \in \mathbb{R}^{(p-q-h) \times (n+h-p)}$, $\mathcal{D}_{2.i} \in \mathbb{R}^{p \times l}$, and $T_0 \in \mathbb{R}^{p \times p}$ is an orthogonal matrix. Matrices $\mathcal{M}_{2.i} \in \mathbb{R}^{p \times q}$, $\mathcal{N}_2 \in \mathbb{R}^{p \times h}$ can have the following structure:

$$\mathcal{M}_{2.i} = \begin{bmatrix} 0 \\ \mathcal{M}_{0.i} \end{bmatrix}, \mathcal{N}_2 = \begin{bmatrix} 0 \\ \mathcal{N}_0 \end{bmatrix}. \tag{13}$$

With $\mathcal{M}_{0.i} \in \mathbb{R}^{(q+h) \times q}$, $\mathcal{N}_0 \in \mathbb{R}^{(q+h) \times h}$ are nonsingular.

(b) The pairs $(\mathcal{A}_{11.i}, \mathcal{A}_{21.i})$ are detectable if and only if the invariant zeros of $\{\mathcal{A}_i, [\mathcal{M}_i \ \mathcal{N}], C\}$ lie in \mathbb{C}^- and it happens if and only if Assumption 3 is satisfied.

Assumption 4. The unknown vectors f_a and f_s and the derivatives of the μ_i for $i \in \{1, \dots, r\}$ are assumed to be norm bounded by some known constants. Therefore,

$$\|f_a\| \leq \rho_a; \quad \|f_s\| \leq \rho_s; \quad \|\dot{\mu}_i\| \leq \rho_{mi}. \tag{14}$$

Lemma 2. Ref. [19] parameterized linear matrix inequality (PLMI) $\sum_{i=1}^r \sum_{j=1}^r \mu_i \mu_j Q_{ij} < 0$ is fulfilled if the following conditions hold:

$$\begin{cases} R_{ii} < 0 & \text{for } i = 1, \dots, r \\ \frac{2}{r-1} R_{ii} + R_{ij} + R_{ji} < 0 & \text{for } i \neq j = 1, \dots, r \end{cases}. \tag{15}$$

3. TS Fuzzy-Based Sliding Mode Observer Design

The proposed TS sliding mode observer for the nonlinear system (2) in the new coordinate (10) is as follows:

$$\begin{cases} \dot{\hat{X}} = \sum_{i=1}^r \hat{\mu}_i \{ \mathcal{A}_i \hat{X} + \mathcal{B}_i u + \mathcal{G}_{l,i} e_Y + \mathcal{G}_{n,i} v_{a,i} + \mathcal{G}_{n,i} v_s \} \\ \hat{Y} = C \hat{X} \end{cases} \tag{16}$$

where $\mathcal{G}_{n,i}$ and $\mathcal{G}_{l,i}$ are design matrices of the observer that will be derived through Theorem 1. $e_Y := Y - \hat{Y}$ represents the output error estimation, $v_{a,i}$ and v_s are the equivalent output error injections that are used to compensate the errors due to the actuator fault and sensor fault, respectively, and have the following structure:

$$v_{a,i} = \begin{cases} \eta_{a,i} \frac{\|e_Y\|}{e_Y} & e_Y \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

$$v_s = \begin{cases} \eta_s \frac{\|e_Y\|}{e_Y} & e_Y \neq 0 \\ 0 & \text{otherwise} \end{cases}'$$

where $\eta_{a,i}$ and η_s are two positive scalars such that

$$\eta_{a,i} \geq \rho_a \|T_0 \mathcal{M}_{2,i}\| \max_j \left(\frac{\|P_{2,j}\|}{\lambda_{\min}(P_{2,j})} \right) + w_{a,i}$$

$$\eta_s \geq \rho_s \|T_0 \mathcal{N}_2\| \max_j \left(\frac{P_{2,j}}{\lambda_{\min}(P_{2,j})} \right) + w_s \quad \forall i, j \in \{1, \dots, r\}. \tag{18}$$

$w_{a,i}$ and w_s are two arbitrary positive constants.

The observer (16) guarantees that the state estimation error converges to a pre-designed sliding surface in finite time and then, asymptotically to zero. Define state estimation error as $e := X - \hat{X}$. By subtracting the observer dynamics from the system dynamic (7) in the new coordinate (12), the state estimation error dynamic can be given as

$$\dot{e} = \sum_{i=1}^r \hat{\mu}_i \left\{ \begin{matrix} (\mathcal{A}_i - \mathcal{G}_{l,i} \mathcal{C})e + \mathcal{M}_i f_a - \mathcal{G}_{n,i} v_{a,i} \\ + \mathcal{N}_i f_s - \mathcal{G}_{n,i} v_s + \mathcal{D}_i d + \phi \end{matrix} \right\}. \tag{19}$$

By partitioning ϕ as $\phi = [\phi_1^T \ \phi_2^T]^T$ and applying a further change of coordinates

$$T_{L,i} = \begin{bmatrix} I_{n+h-p} & L_i \\ 0 & T_0 \end{bmatrix}, \quad L_i = [\bar{L}_i \quad 0] \in \mathbb{R}^{(n+h-p) \times p} \tag{20}$$

where $\bar{L}_i \in \mathbb{R}^{(n+h-p) \times (p-q-h)}$ is a stabilizing gain matrix, it is straightforward to see that

$$\left\{ \begin{matrix} \tilde{\mathcal{A}}_i = \begin{bmatrix} \mathcal{A}_{11,i} + L_i \mathcal{A}_{21,i} & \tilde{\mathcal{A}}_{12,i} \\ T_0 \mathcal{A}_{21,i} & \tilde{\mathcal{A}}_{22,i} \end{bmatrix} \\ \tilde{\mathcal{M}}_i = \begin{bmatrix} 0 \\ T_0 \mathcal{M}_{2,i} \end{bmatrix} \\ \tilde{\mathcal{N}} = \begin{bmatrix} 0 \\ T_0 \mathcal{N}_2 \end{bmatrix} \\ \tilde{\mathcal{D}}_i = \begin{bmatrix} \mathcal{D}_{1,i} + L_i \mathcal{D}_{2,i} \\ T_0 \mathcal{D}_{2,i} \end{bmatrix} \\ \tilde{\mathcal{C}} = [0 \quad I_p] \\ \tilde{\mathcal{G}}_{n,i} = \begin{bmatrix} 0 \\ I_p \end{bmatrix} \\ \tilde{\mathcal{G}}_{l,i} = \begin{bmatrix} \tilde{\mathcal{A}}_{12,i} \\ \tilde{\mathcal{A}}_{22,i} - \mathcal{A}_{s,i} \end{bmatrix} \\ \tilde{\phi} = \begin{bmatrix} T_{L,i} \phi_1 \\ T_{L,i} \phi_2 \end{bmatrix} \end{matrix} \right., \tag{21}$$

where $\mathcal{A}_{s,i}$ are arbitrary stable design matrices. Through the new coordinate, the error dynamic (19) can be re-written as

$$\dot{\tilde{e}} = \begin{bmatrix} \dot{\tilde{e}}_1 \\ \dot{\tilde{e}}_Y \end{bmatrix} = \sum_{i=1}^r \hat{\mu}_i \left\{ \begin{bmatrix} \tilde{\mathcal{A}}_{t,i} \tilde{e} + T_{L,i} \phi + \tilde{\mathcal{M}}_i f_a - \\ \tilde{\mathcal{G}}_{n,i} v_{a,i} + \tilde{\mathcal{N}} f_s - \tilde{\mathcal{G}}_{n,i} v_s + \tilde{\mathcal{D}}_i d \end{bmatrix} \right\} \tag{22}$$

where

$$\tilde{\mathcal{A}}_{t,i} = \begin{bmatrix} \mathcal{A}_{11,i} + L_i \mathcal{A}_{21,i} & 0 \\ T_0 \mathcal{A}_{21,i} & \mathcal{A}_{s,i} \end{bmatrix} . \tag{23}$$

The goal is to design the matrices L_i such that the asymptotic stability of (22) is assured while the following specified H_∞ performance is guaranteed:

$$\| \tilde{e}^2 \| \leq \vartheta^2 \| d^2 \| . \tag{24}$$

The following theorem provides sufficient conditions to ensure asymptotic stability of the state estimation error (22) with maximized admissible Lipschitz constant γ in (8) and minimized H_∞ performance gain ϑ in (24).

Theorem 1. *If there exist feasible solutions for the following optimization problem with a fixed scalar $0 \leq \lambda \leq 1$*

$$\begin{aligned} & \min[\lambda(\sigma + \varepsilon) + (1 - \lambda)\vartheta] \\ & \text{Subject} \\ & \text{eig}(R_{ii}) < 0 \qquad \text{for } i = 1, \dots, r \\ & \text{eig}\left(\frac{2}{r-1}R_{ii} + R_{ij} + R_{ji}\right) < 0 \qquad \text{for } i \neq j = 1, \dots, r \\ & -\text{eig}(P_{1,i}) < 0 \qquad \text{for } i = 1, \dots, r \\ & -\text{eig}(P_{2,i}) < 0 \qquad \text{for } i = 1, \dots, r \\ & -\varepsilon < 0 \\ & -\sigma < 0 \\ & -\vartheta < 0 \end{aligned} \tag{25}$$

where

$$\begin{aligned} R_{ij} &= \begin{bmatrix} \Phi_{1,ij} & (P_{2,j} T_0 \mathcal{A}_{21,i})^T & \Phi_{3,ij} \\ P_{2,j} T_0 \mathcal{A}_{21,i} & \Phi_{2,ij} & P_{2,j} T_0 \mathcal{D}_{2,i} \\ \Phi_{3,ij}^T & (P_{2,j} T_0 \mathcal{D}_{2,i})^T & -\beta I_i \end{bmatrix} \\ \Phi_{1,ij} &= (\mathcal{A}_{11,i} + L_i \mathcal{A}_{21,i})^T P_{1,j} + P_{1,j} (\mathcal{A}_{11,i} + L_i \mathcal{A}_{21,i}) + \varepsilon^{-1} P_{1,j} P_{1,j} r \\ & \quad + (\sigma^{-1} + 1) I_{n+h-p} + \sum_{k=1}^r q_{mk} P_{1,k} \\ \Phi_{2,ij} &= \mathcal{A}_{s,i}^T P_{2,j} + P_{2,j} \mathcal{A}_{s,i} + \varepsilon^{-1} P_{2,j} P_{2,j} + (\sigma^{-1} + 1) I_p + \sum_{k=1}^r q_{mk} P_{2,k} \\ \Phi_{3,ij} &= P_{1,j} \mathcal{D}_{1,i} + P_{1,j} L_i \mathcal{D}_{2,i} \end{aligned} \tag{26}$$

and eig represents eigenvalues of a matrix, then, the estimation error (22) is asymptotically stable with the maximized admissible Lipschitz constant $\gamma^* = \max(\gamma) = \frac{1}{\|T_L\| \|T_L^{-1}\| \sqrt{\varepsilon \sigma}}$ and the derived L_i matrices can be used for the purpose of simultaneous fault reconstruction.

Proof. The proof of this theorem is done by using a positive NQLF as follows

$$V = \tilde{e}^T \left(\sum_{j=1}^r \hat{\mu}_j P_j \right) \tilde{e}, \tag{27}$$

where $P_j = \text{diag}(P_{1j}, P_{2j})$ with $P_{1j} \in R^{(n+h-p) \times (n+h-p)}$ and $P_{2j} \in R^{p \times p}$ are symmetric positive definite matrices. The time derivative of the candidate Lyapunov function along the trajectory (22) is given by

$$\dot{V} = \sum_{i=1}^r \sum_{j=1}^r \hat{\mu}_i \hat{\mu}_j \{ \tilde{e}^T (\mathcal{A}_{t,i}^T P_j + P_j \mathcal{A}_{t,i} + \sum_{k=1}^r \dot{\hat{\mu}}_k P_k) \tilde{e} + 2\tilde{e}^T P_j (T_{L,i} \phi + \tilde{\mathcal{M}}_i f_a - \tilde{\mathcal{G}}_{n,i} v_{a,i} + \tilde{\mathcal{N}} f_s - \tilde{\mathcal{G}}_{n,i} v_s + \tilde{\mathcal{D}}_i d) \}. \tag{28}$$

From (14), (17), (18) and (21), one has:

$$\begin{aligned} & \tilde{e}^T P_j (\tilde{\mathcal{M}}_i f_a - \tilde{\mathcal{G}}_{n,i} v_{a,i}) \\ &= e_Y^T P_{2,j} T_0 \mathcal{M}_{2,i} f_a - \eta_{a,i} \frac{e_Y^T P_{2,j} e_Y}{\|e_Y\|} \\ &\leq \|e_Y^T P_{2,j} T_0 \mathcal{M}_{2,i} f_a\| - \eta_{a,i} \frac{e_Y^T P_{2,j} e_Y}{\|e_Y\|} \\ &\leq \|e_Y^T P_{2,j} T_0 \mathcal{M}_{2,i} f_a\| - \eta_{a,i} \lambda_{\min}(P_{2,j}) \|e_Y\| \\ &\leq \|e_Y\| (\rho_a \|P_{2,j}\| \|T_0 \mathcal{M}_{2,i}\| - \eta_{a,i} \lambda_{\min}(P_{2,j})) \\ &\leq -w_{a,i} \lambda_{\min}(P_{2,j}) \|e_Y\| \leq 0 \\ & \tilde{e}^T P_j (\tilde{\mathcal{N}} f_s - \tilde{\mathcal{G}}_{n,i} v_s) \\ &= e_Y^T P_{2,j} T_0 \mathcal{N}_{2,i} f_s - \eta_s \frac{e_Y^T P_{2,j} e_Y}{\|e_Y\|} \\ &\leq \|e_Y^T P_{2,j} T_0 \mathcal{N}_{2,i} f_s\| - \eta_s \frac{e_Y^T P_{2,j} e_Y}{\|e_Y\|} \\ &\leq \|e_Y^T P_{2,j} T_0 \mathcal{N}_{2,i} f_s\| - \eta_s \lambda_{\min}(P_{2,j}) \|e_Y\| \\ &\leq \|e_Y\| (\rho_s \|P_{2,j}\| \|T_0 \mathcal{N}_{2,i}\| - \eta_s \lambda_{\min}(P_{2,j})) \\ &\leq -w_{s,i} \lambda_{\min}(P_{2,j}) \|e_Y\| \leq 0. \end{aligned} \tag{29}$$

From (14), one has

$$\sum_{k=1}^r \dot{\hat{\mu}}_k P_k \leq \sum_{k=1}^r \rho_{mi} P_k. \tag{30}$$

By considering the fact that $2\mathbb{P}^T \mathbb{Q} \leq \frac{1}{\epsilon} \mathbb{P}^T \mathbb{P} + \epsilon \mathbb{Q}^T \mathbb{Q}$ with $\epsilon > 0$ and using (8), one obtains

$$2\tilde{e}^T P_j T_L \phi \leq \frac{1}{\epsilon} \tilde{e}^T P_j P_j \tilde{e} + \epsilon \phi^T T_L^T T_L \phi \leq \frac{1}{\epsilon} \tilde{e}^T P_j P_j \tilde{e} + \epsilon \alpha^2 \|\tilde{e}\|^2, \tag{31}$$

where $\alpha := \|T_L\| \|T_L^{-1}\| \gamma$. By Substituting (29)–(31) into (28), one has

$$\dot{V} \leq \sum_{i=1}^r \sum_{j=1}^r \hat{\mu}_i \hat{\mu}_j \left\{ \tilde{e}^T \left(\mathcal{A}_{t,i}^T P_j + P_j \mathcal{A}_{t,i} + \frac{1}{\epsilon} P_j P_j + \epsilon \alpha^2 I_{n+h} + \sum_{k=1}^r \rho_{mk} P_k \right) \tilde{e} + 2\tilde{e}^T P_j \tilde{\mathcal{D}}_i d \right\}. \tag{32}$$

By defining parameter $\sigma := (\epsilon \alpha^2)^{-1}$ and the cost function as $J := \dot{V}(\tilde{e}) + \tilde{e}^T \tilde{e} - \theta^2 d^T d$, one has

$$J \leq \sum_{i=1}^r \sum_{j=1}^r \hat{\mu}_i \hat{\mu}_j \left\{ \tilde{e}^T (\mathcal{A}_{t,i}^T P_j + P_j \mathcal{A}_{t,i} + \epsilon^{-1} P_j P_j + \sigma^{-1} I_{n+h} + I_{n+h} + \sum_{k=1}^r \rho_{mk} P_k) \tilde{e} + 2\tilde{e}^T P_j \tilde{\mathcal{D}}_i d - \beta d^T d \right\}, \tag{33}$$

where $\beta := \theta^2$. By placing (23) in (33) and considering the diagonal structure of P_j , the inequality (33) is continued as

$$J \leq \sum_{i=1}^r \sum_{j=1}^r \hat{\mu}_i \hat{\mu}_j \begin{bmatrix} e_1 \\ e_Y \\ \zeta \end{bmatrix}^T \Lambda \begin{bmatrix} e_1 \\ e_Y \\ \zeta \end{bmatrix} < 0, \tag{34}$$

where

$$\Lambda = \begin{bmatrix} \Phi_{1,ij} & (P_{2,j}T_0\mathcal{A}_{21,i})^T & \Phi_{3,ij} \\ P_{2,j}T_0\mathcal{A}_{21,i} & \Phi_{2,ij} & P_{2,j}T_0\mathcal{D}_{2,i} \\ \Phi_{3,ij}^T & (P_{2,j}T_0\mathcal{D}_{2,i})^T & -\beta I_l \end{bmatrix}$$

$$\Phi_{1,ij} = (\mathcal{A}_{11,i} + L_i\mathcal{A}_{21,i})^T P_{1,j} + P_{1,j}(\mathcal{A}_{11,i} + L_i\mathcal{A}_{21,i}) + \varepsilon^{-1}P_{1,j}P_{1,j} + (\sigma^{-1} + 1)I_{n+h-p} + \sum_{k=1}^r q_{mk}P_{1,k}$$

$$\Phi_{2,ij} = \mathcal{A}_{s,i}^T P_{2,j} + P_{2,j}\mathcal{A}_{s,i} + \varepsilon^{-1}P_{2,j}P_{2,j} + (\sigma^{-1} + 1)I_p + \sum_{k=1}^r q_{mk}P_{2,k}$$

$$\Phi_{3,ij} = P_{1,j}\mathcal{D}_{1,i} + P_{1,j}L_i\mathcal{D}_{2,i}.$$
(35)

Based on the Congruence [20], the inequality (35) is satisfied by

$$\sum_{i=1}^r \sum_{j=1}^r \hat{\mu}_i \hat{\mu}_j \begin{bmatrix} \Phi_{1,ij} & (P_{2,j}T_0\mathcal{A}_{21,i})^T & \Phi_{3,ij} \\ P_{2,j}T_0\mathcal{A}_{21,i} & \Phi_{2,ij} & P_{2,j}T_0\mathcal{D}_{2,i} \\ \Phi_{3,ij}^T & (P_{2,j}T_0\mathcal{D}_{2,i})^T & -\beta I_l \end{bmatrix} < 0.$$
(36)

By utilizing Lemma 2, the summations and the fuzzy membership functions will be omitted from inequalities (36). Finally, the results are going to be used for fmincon function which is a nonlinear optimization tool in MATLAB software and finds the minimum of a problem specified by

$$\begin{aligned} & \min_x f(x) \\ & \text{subject to } \begin{cases} c(x) \leq 0 \\ ceq(x) \leq 0 \\ A \cdot x \leq b \\ Aeq \cdot x = beq \\ lb \leq x \leq ub \end{cases} \end{aligned}$$
(37)

The matrix inequalities (36) should be changed to some one-dimensional inequalities, and the optimization problem can be defined as (25) and (26). In addition, from the α and σ found by the optimization problem, the maximum admissible Lipschitz constant and the minimum can be calculated as

$$\gamma^* = \frac{1}{\|T_L\| \|T_L^{-1}\| \sqrt{\sigma\varepsilon}}.$$
(38)

□

4. Simultaneous Fault Reconstruction

In Section 3, an H_∞ sliding mode observer is designed in which two discontinuous terms (19) are considered to reconstruct simultaneous faults in the presence of an unknown disturbance based on the measured signals u and y . Along the sliding surface $e_Y = \dot{e}_Y = 0$. Consequently, (22) on the sliding surface changes to

$$\sum_{i=1}^r \hat{\mu}_i \left\{ T_0\mathcal{A}_{21,i}e_1 + T_0\phi_2 + T_0\mathcal{M}_{2i}f_a - v_{eqs,i} + T_0\mathcal{D}_{2i}d \right\} = 0,$$
(39)

where $v_{eqa,i}$ and v_{eqs} are approximations of the equivalent output error injection terms (17) required to maintain the sliding motion and can be defined as

$$v_{eqa,i} = \eta_{a,i} \frac{e_Y}{\|e_Y\| + \delta_a}; \quad v_{eqs} = \eta_s \frac{e_Y}{\|e_Y\| + \delta_s},$$
(40)

where δ_f and δ_d are small positive constants. Consequently, (40) leads to

$$0 = \sum_{i=1}^r \hat{\mu}_i \left\{ -T_0^{-1}v_{eqa,i} + \mathcal{N}_{2i}f_s - T_0^{-1}v_{eqs} + \mathcal{D}_{2i}d \right\}.$$
(41)

On the other hand, using (8) and (24) can show that the term $\mathcal{A}_{21,i}e_1 + \phi_2 + \mathcal{D}_{2,i}d$ is bounded as

$$\begin{aligned} & \| \mathcal{A}_{21,i}e_1 + \phi_2 + \mathcal{D}_{2,i}d \| \\ & \leq (\| \mathcal{A}_{21,i} \| + \gamma \| T_L^{-1} \|) \| e_1 \| + \| \mathcal{D}_{2,i} \| \| d \| \\ & \leq (\| \mathcal{A}_{21,i} \| + \gamma \| T_L^{-1} \|) \| \tilde{e} \| + \| \mathcal{D}_{2,i} \| \| d \| \\ & \leq \epsilon \| d \|, \end{aligned} \tag{42}$$

where $\epsilon = \mu(\| \mathcal{A}_{21,i} \| + \gamma \| T_L^{-1} \|) + \| \mathcal{D}_{2,i} \|$. Therefore, for small values of $\epsilon \| d \|$, the actuator and sensor faults can be estimated as

$$\hat{f}_a = \left(\sum_{k=1}^r \hat{\mu}_i \{ \mathcal{M}_{2,i} \} \right)^\dagger T_0^{-1} \sum_{k=1}^r \hat{\mu}_i \left\{ \eta_{a,i} \frac{e_Y}{\| e_Y \| + \delta_a} \right\} \tag{43}$$

$$\hat{f}_s = \mathcal{N}_2^\dagger T_0^{-1} \eta_s \frac{e_Y}{\| e_Y \| + \delta_s}, \tag{44}$$

where † shows the pseudo-inverse of a matrix.

Remark 1. The numerical solution of Theorem 1 can be summarized as follows:

- Find the orthogonal transfer matrix $T_r \in R^{p \times p}$ by using the QR reduction of matrix N and obtain the augmented TS system (4).
- Find the changes of coordinates \bar{T}_i and obtain the system matrices in the format(12) and(13).
- Compute the scalars σ , ϵ , and θ and also the matrices L_i using the `fmincon` function in MATLAB software and solving the nonlinear optimization problem(25).
- Compute the maximized admissible Lipschitz constant as $\gamma^* = \max(\gamma) = \frac{1}{\bar{T}_L T_L^{-1} \sqrt{\epsilon \sigma}}$.
- Reconstruct the sensor and actuator faults using Equations(43) and(44).

5. Numerical Example

In this section, a three-state variable continuous stirred tank reactor (CSTR) system is utilized to show the effectiveness of the proposed sliding mode observer in both actuator and sensor faults reconstruction in the presence of an unknown disturbance. To show the performance improvement of the proposed approach, the obtained results are compared to the LMI approach presented in ref. [17].

Consider a well-mixed variable CSTR in which a multi-component chemical reaction $A \rightleftharpoons B \rightarrow C$ is being carried out. The nonlinear dynamics of the CSTR is given by the following model [21],

$$\dot{x} = \begin{bmatrix} -4 & 0.8796 & 0 \\ 3 & -3.6388 & 0 \\ 0 & 1.7592 & -1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u + \begin{bmatrix} 0.5x_2^2 \\ -1.5x_2^2 \\ x_2^2 \end{bmatrix}, \tag{45}$$

where $x = [x_1 \ x_2 \ x_3]^T$, and the states represent the concentrations of the species A , B , and C , respectively. To check the advantage of the proposed method, two faults and a disturbance are added to the dynamic (45) as

$$\begin{cases} \dot{x} = \begin{bmatrix} -4 & 0.8796 + 0.5x_2 & 0 \\ 3 & -3.6388 - 1.5x_2 & 0 \\ 0 & 1.7592 + x_2 & -1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u \\ \quad + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} f_a + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \zeta \\ y = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} f_s \end{cases} \tag{46}$$

It is supposed that the concentration of B is dimensionless, which means that $x_2 \in [-1\ 1]$. Consequently, by using TS rules, two membership functions can be defined as

$$h_1 = \frac{1 - x_2}{2}; \quad h_2 = \frac{1 + x_2}{2}. \tag{47}$$

Therefore, the local linear TS matrices can be determined as

$$\begin{aligned} A_1 &= \begin{bmatrix} -4 & 0.8796 - 0.5 & 0 \\ 3 & -3.6388 + 1.5 & 0 \\ 0 & 1.7592 - 1 & -1 \end{bmatrix}; B_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}; M_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}; D_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ A_2 &= \begin{bmatrix} -4 & 0.8796 + 0.5 & 0 \\ 3 & -3.6388 - 1.5 & 0 \\ 0 & 1.7592 + 1 & -1 \end{bmatrix}; B_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}; M_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}; D_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \end{aligned} \tag{48}$$

The TS fuzzy system matrices satisfy all the assumptions; therefore, the TS fuzzy sliding observer (16) can be designed.

For simulation, the parameters and input signal are chosen as $u = \sin(t)$, $A_f = 1$, $\mathcal{A}_s = -5I$, $\eta_{d,i} = \eta_a = 25$, $\eta_s = 25$, $\delta_a = 0.01$ and $\delta_s = 0.01$. and the initial conditions are chosen as $X_0 = [1\ 1.2\ 1\ 0]^T$ and $\hat{X}_0 = [1.5\ 2.8\ 0.5\ 0]^T$. Moreover, the disturbance is chosen as $d = 0.1 \sin(0.2t)x_3$ and the shape is shown in Figure 1.

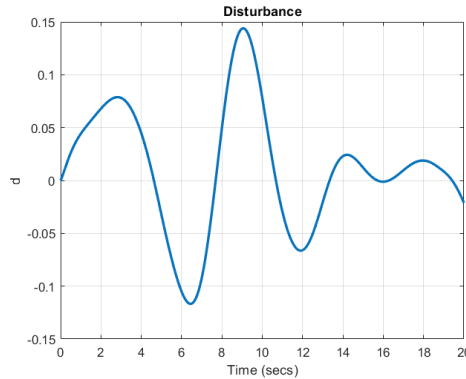


Figure 1. Disturbance $d(t)$.

The maximum Lipschitz constant and the minimum H_∞ performance gain obtained through fmincon function in MATLAB on Theorem 1 are $\gamma^* = 0.8358$ and $\vartheta^* = 0.2982$. The observer matrices are derived as

$$\begin{aligned} \mathcal{G}_{l,1} &= \begin{bmatrix} 0.4499 & 1 & 0 \\ 3.3912 & 3 & 0 \\ 4.8998 & 0 & 0 \\ 1.1852 & 0 & 4 \end{bmatrix}, \mathcal{G}_{l,2} = \begin{bmatrix} 2.4723 & 1 & 0 \\ -0.2487 & 3 & 0 \\ 8.9447 & 0 & 0 \\ 1.7921 & 0 & 4 \end{bmatrix}, \\ \mathcal{G}_{n,1} &= \begin{bmatrix} 0 & 1 & 0 \\ 1.1852 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathcal{G}_{n,2} = \begin{bmatrix} 0 & 1 & 0 \\ 1.7921 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

It should be noted that the initial point for fmincon is chosen based on the results of the related published papers. Figure 2 shows the state estimation error which converges to a neighborhood close to zero due to the unknown disturbance.

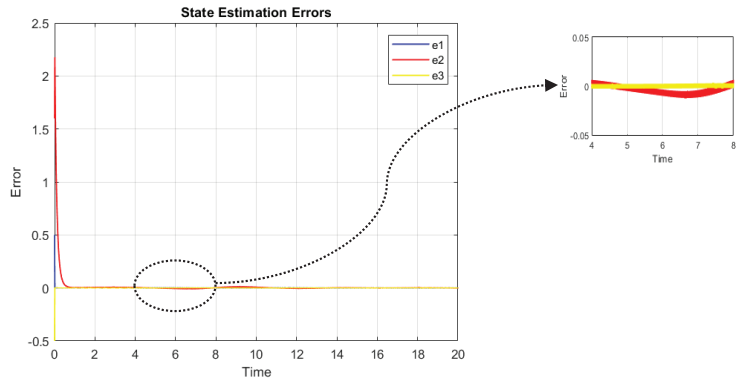


Figure 2. State estimation error in the presence of faults and disturbance.

Figures 3 and 4 show that the proposed TS-based SMO is able to reconstruct the simultaneous faults with a small error in the presence of an unknown disturbance.

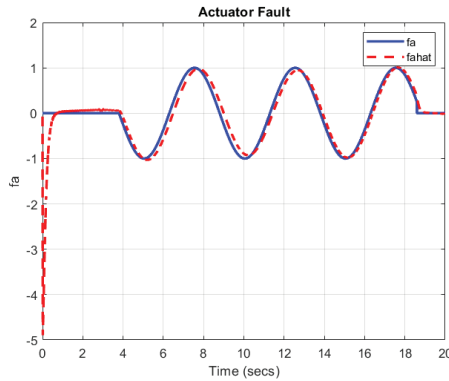


Figure 3. Actuator fault $f_a(t)$ (by blue solid line) and its estimation $\hat{f}_a(t)$ (by red dashed line).

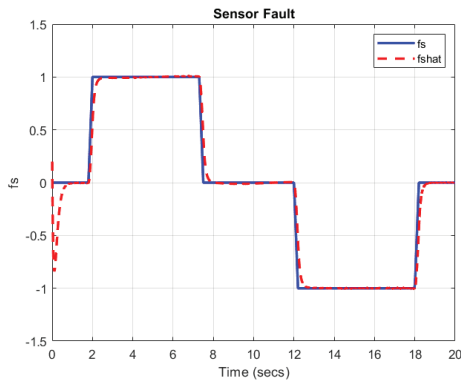


Figure 4. Sensor fault $f_s(t)$ (by blue solid line) and its estimation $\hat{f}_s(t)$ (by red dashed line).

The proposed approach is compared with another non-quadratic Lyapunov-based approach using linear optimization analysis based on LMIs [17]. Figure 5 describes the fault estimation errors using both approaches.

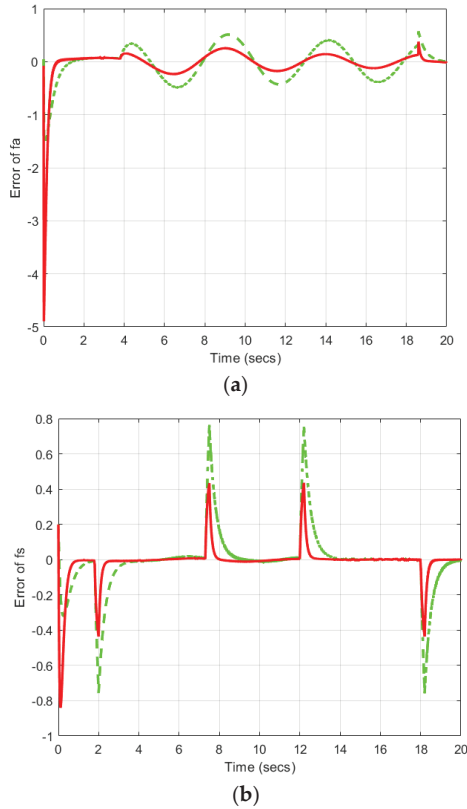


Figure 5. Fault estimation errors (a). Actuator fault, (b). Sensor fault (the proposed approach by red solid line and ref. [17] by green dashed line).

As can be seen, the proposed nonlinear approach is less conservative and can estimate both actuator and sensor faults with smaller errors. In addition, the proposed approach has less computational burden. In Table 1, a quantitative comparison between the proposed approach and the LMI approach presented in ref. [17] is considered. In this table, the Euclidean and infinity norms of the fault error estimations are compared and the improvements are calculated as

$$Improvement (\%) = \left(\frac{F_l - F_n}{F_l} \right) * 100, \tag{49}$$

where F_n and F_l represent the $\| Error\ of\ f \|$ using the LMI approach [17] and the nonlinear proposed approach, respectively.

Table 1. The norm specifications of the fault reconstruction errors for two different approaches.

| | $\ ef_a\ _2$ | $\ ef_a\ _\infty$ | $\ ef_s\ _2$ | $\ ef_s\ _\infty$ |
|--------------------------|--------------|-------------------|--------------|-------------------|
| Proposed Approach | 23.9858 | 0.3820 | 15.3073 | 0.4376 |
| [17] | 53.3519 | 0.5704 | 33.6076 | 0.7679 |
| Improvement (%) | +55.04 | +33.03 | +54.45 | +43.01 |

As can be seen in Table 1, the proposed approach improves the fault estimation accuracies by more than 30%.

6. Discussion

In this paper, a nonlinear optimization approach for simultaneous actuator and sensor fault reconstruction in nonlinear systems subjected to unknown disturbances was proposed. First, an augmented system with just an actuator fault was created. Then, by using the fuzzy Lyapunov stability analysis and two changes of coordinates, the parameters of a sliding mode observer were designed through a nonlinear optimization problem while maximizing the Lipschitz constant and minimizing the H_∞ performance index. The optimization problem was solved by using `fmincon` in MATLAB as a nonlinear optimization tool. By utilizing the optimum points, both actuator and sensor faults were reconstructed properly. Finally, the simulation results showed a considerable increase in the fault reconstruction accuracy with constraints with smaller dimensions.

Author Contributions: Conceptualization, S.A.; methodology, S.A., M.M., and G.G.W.; software, S.A.; validation, S.A., M.M.; formal analysis, S.A.; investigation, S.A.; resources, S.A.; writing—original draft preparation, S.A. and G.G.W.; writing—review and editing, S.A. and M.M.; visualization, S.A. and M.M.; supervision, M.M. and G.G.W.; project administration, M.M. and G.G.W. 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: Not applicable.

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

References

- Zhang, Z.; Hu, Y.; Luo, G.; Gong, C.; Liu, X.; Chen, S. An Embedded Fault-Tolerant Control Method for Single Open-Switch Faults in Standard PMSM Drives. *IEEE Trans. Power Electron.* **2022**, *37*, 8476–8487. [[CrossRef](#)]
- Tawfiq, K.B.; Ibrahim, M.N.; Sergeant, P. An Enhanced Fault-Tolerant Control of a Five-Phase Synchronous Reluctance Motor Fed from a Three-to-Five-Phase Matrix Converter. *IEEE J. Emerg. Sel. Top. Power Electron.* **2022**, *10*, 4182–4194. [[CrossRef](#)]
- Alwi, H.; Edwards, C.; Tan, C.P. *Fault Detection and Fault-Tolerant Control Using Sliding Modes*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2011.
- Falcon, R.; Rios, H.; Dzul, A. A Robust Fault Diagnosis for Quad-Rotors: A Sliding-Mode Observer Approach. *IEEE/ASME Trans. Mechatron.* **2022**, 1–10. [[CrossRef](#)]
- Zhang, K.; Jiang, B.; Yan, X.-G.; Mao, Z. Incipient Fault Detection for Traction Motors of High-Speed Railways Using an Interval Sliding Mode Observer. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 2703–2714. [[CrossRef](#)]
- Zhang, K.; Jiang, B.; Shi, P.; Cocquempot, V. *Observer-Based Fault Estimation Techniques*; Springer International Publishing: Cham, Switzerland, 2018; Volume 127. [[CrossRef](#)]
- Yang, J.; Zhu, F.; Wang, X.; Bu, X. Robust sliding-mode observer-based sensor fault estimation, actuator fault detection and isolation for uncertain nonlinear systems. *Int. J. Control Autom. Syst.* **2015**, *13*, 1037–1046. [[CrossRef](#)]
- Nguyen, V.-C.; Vo, A.-T.; Kang, H.-J. A Non-Singular Fast Terminal Sliding Mode Control Based on Third-Order Sliding Mode Observer for a Class of Second-Order Uncertain Nonlinear Systems and its Application to Robot Manipulators. *IEEE Access* **2020**, *8*, 78109–78120. [[CrossRef](#)]
- Ziyabari, H.S.; Shoorehdeli, M.A. Fuzzy robust fault estimation scheme for a class of nonlinear systems based on an unknown input sliding mode observer. *J. Vib. Control* **2018**, *24*, 1861–1873. [[CrossRef](#)]
- Gu, Z.; Yue, D.; Park, J.H.; Xie, X. Memory-Event-Triggered Fault Detection of Networked IT2 T-S Fuzzy Systems. *IEEE Trans. Cybern.* **2022**, 1–10. [[CrossRef](#)] [[PubMed](#)]
- Li, S.; Wang, H.; Aitouche, A.; Christov, N. Sliding mode observer design for fault and disturbance estimation using Takagi–Sugeno model. *Eur. J. Control* **2018**, *44*, 114–122. [[CrossRef](#)]
- Asadi, S.; Khayatian, A.; Dehghani, M.; Vafamand, N. Simultaneous fault reconstruction of TS fuzzy systems using robust sliding mode observer and non-quadratic stability analysis. In Proceedings of the 2017 Iranian Conference on Electrical Engineering (ICEE), Tehran, Iran, 2–4 May 2017; pp. 823–828. [[CrossRef](#)]
- Salem, A.; Kardous, Z.; Braiek, N.B.; Ragot, J. On the state observer based stabilization of Takagi–Sugeno systems with immeasurable premise variables. *Int. J. Control Autom. Syst.* **2012**, *10*, 1153–1163. [[CrossRef](#)]
- Gómez-Peñate, S.; Valencia-Palomo, G.; López-Estrada, F.; Astorga-Zaragoza, C.; Osornio-Rios, A.R.; Ruiz, I.D.L.S. Sensor Fault Diagnosis Based on a Sliding Mode and Unknown Input Observer for Takagi–Sugeno Systems with Uncertain Premise Variables. *Asian J. Control* **2019**, *21*, 339–353. [[CrossRef](#)]

15. Ben Brahim, A.; Dhahri, S.; Ben Hmida, F.; Sellami, A. An H_∞ sliding mode observer for Takagi–Sugeno nonlinear systems with simultaneous actuator and sensor faults. *Int. J. Appl. Math. Comput. Sci.* **2015**, *25*, 547–559. [[CrossRef](#)]
16. Ben Brahim, A.; Dhahri, S.; BEN Hmida, F.; Sellami, A. Simultaneous Actuator and Sensor Faults Reconstruction Based on Robust Sliding Mode Observer for a Class of Nonlinear Systems. *Asian J. Control* **2017**, *19*, 362–371. [[CrossRef](#)]
17. Asadi, S.; Khayatian, A.; Dehghani, M.; Vafamand, N.; Khooban, M.H. Robust sliding mode observer design for simultaneous fault reconstruction in perturbed Takagi-Sugeno fuzzy systems using non-quadratic stability analysis. *J. Vib. Control* **2020**, *26*, 1092–1105. [[CrossRef](#)]
18. Asadi, S.; Vafamand, N.; Moallem, M.; Dragicevic, T. Fault Reconstruction of Islanded Nonlinear DC Microgrids: An LPV-Based Sliding Mode Observer Approach. *IEEE J. Emerg. Sel. Top. Power Electron.* **2021**, *9*, 4606–4614. [[CrossRef](#)]
19. Vafamand, N.; Asemani, M.H.; Khayatian, A. A robust L1 controller design for continuous-time TS systems with persistent bounded disturbance and actuator saturation. *Eng. Appl. Artif. Intell.* **2016**, *56*, 212–221. [[CrossRef](#)]
20. Scherer, C.; Weiland, S. Linear matrix inequalities in control. In *Lecture Notes*; Dutch Institute for Systems and Control: Delft, The Netherlands, 2000; Volume 3.
21. Vafamand, N.; Khayatian, A. Model predictive-based reset gain-scheduling dynamic control law for polytopic LPV systems. *ISA Trans.* **2018**, *81*, 132–140. [[CrossRef](#)] [[PubMed](#)]

Article

Vibro-Acoustic Distributed Sensing for Large-Scale Data-Driven Leak Detection on Urban Distribution Mains

Lili Bykerk * and Jaime Valls Miro

Robotics Institute (UTS:RI), University of Technology Sydney, Ultimo, NSW 2007, Australia

* Correspondence: lili.bykerk@uts.edu.au

Abstract: Non-surfacing leaks constitute the dominant source of water losses for utilities worldwide. This paper presents advanced data-driven analysis methods for leak monitoring using commercial field-deployable semi-permanent vibro-acoustic sensors, evaluated on live data collected from extensive multi-sensor deployments across a sprawling metropolitan city. This necessarily includes a wide variety of pipeline sizes, materials and surrounding soils, as well as leak sources and rates brought about by external factors. The novel proposition for structural pipe health monitoring shows that excellent leak/no-leak classification results (>94% accuracy) can be observed using Convolutional Neural Networks (CNNs) trained with Short-Time Fourier Transforms (STFTs) of the raw audio files. Most notably, it is shown how this can be achieved irrespective of the sensor used, with four models from different manufactures being part of the investigation, and over time across extended densely populated areas.

Keywords: water distribution network; vibro-acoustic sensors; leak detection; structural health monitoring; feature extraction; signal processing; machine learning; binary classification; data-driven; neural network

Citation: Bykerk, L.; Valls Miro, J. Vibro-Acoustic Distributed Sensing for Large-Scale Data-Driven Leak Detection on Urban Distribution Mains. *Sensors* **2022**, *22*, 6897. <https://doi.org/10.3390/s22186897>

Academic Editors: Yongbo Li, Bing Li, Jinchun Ji and Hamed Kalhori

Received: 9 June 2022

Accepted: 7 September 2022

Published: 13 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Potable water mains are critical components of water infrastructure. Many water utilities worldwide are managing underground pipes that have been in use for centuries. Given their age and environmental surroundings, pipes are susceptible to failures often caused by tree roots, corrosion, and/or ground movement. In addition to pipe failures, leaks can also emerge from appurtenances in the pipe network such as hydrants, valves, pipe joints, main tapping points, or service lines. Depending on the environment, water from some leaks may never surface, and will remain hidden, resulting in large water losses. When a leak becomes visible, reactive repairs are undertaken; causing disruption to customers and costly maintenance, which can be challenging for utilities to manage.

Distributed IoT sensors such as digital meters are being increasingly used by utilities to remotely monitor the performance of their network in (near) real-time. This allows the monitoring of water usage habits, and establishing the potential for leaks in the main tap and service line connection to a home. In the distribution network, IoT flow meters have been explored to identify leakage. A small experimental laboratory study contrasting various machine learning algorithms (random forest, decision trees, neural networks, and Support Vector Machine) revealed the former as the best at detecting leaks with a 75% accuracy [1]. These sensors require access to the water column to operate, a non-trivial exercise in distribution networks, thus severely limiting their leak identification and localisation capabilities. They have not been widely adopted by the industry, whose preference is for non-intrusive and portable sensing methods, such as contact acoustics-based signalling. As water discharges from a leak in the pipe network, vibrations are induced and propagated along the pipe wall. To detect hidden leaks, utilities commonly schedule Active Leak Detection (ALD) teams to periodically sweep areas of pipelines using acoustic leak detection equipment such as listening sticks and real-time correlators [2]. The

success of these ALD sweeps can be hindered by the prevalence of environmental and water usage noises during the day, when the sweeps are conducted, and the experience of the user [3]. Depending on the length of the utility's pipeline network, the time that elapses between ALD sweeps may result in hidden leaks remaining undetected for long periods of time, or missed entirely. For the continuous monitoring of the network, alternative methods of leak detection are also employed, such as Minimum Night Flow (MNF) and pressure transient analysis using existing network hardware (flow meters and pressure gauges). These methods, however, are only capable of detecting possible leakage in a given area, and will not provide any means of locating or pinpointing a leak location.

Vibro-acoustic sensing has been widely adopted by water utilities [4,5], mainly due to the relative low cost, ease of implementation, flexibility, and passive nature of the system, whereby no permanent changes to the water pipeline network are required for the technology to function. These semi-permanent devices can be used to effectively and remotely monitor the water mains for leakages—generally between 2 and 4 a.m.—when there is low network activity (the time period when MNF is calculated) and low levels of environmental noise. However, there are several challenges and uncertainties in analysing the acoustic sensor data for leak detection: (1) a leak noise can be attenuated due to fittings, joints, junctions, and service connections which are often undocumented; (2) the presence of environmental noises, and water usage in the network; (3) the signal recorded by the acoustic sensor is directly related to the pipe material and diameter, proximity to the leak noise and the quality of the sensor's mounting point on the asset [6,7].

Semi-permanent vibro-acoustic noise loggers have in-built algorithms which raise leak alarms based on the intensity and consistency of the recorded noise [8]. Using this method, a large number of false positive leak alarms are raised by the system, and quieter leaks are missed (false negatives). By understanding the limitations of these in-built leak detection algorithms, and the uncertainties affecting the data recorded by an acoustic logger, there is a motivation and need for a more advanced analysis of the acoustic data to achieve accurate and reliable leak detection. Signal processing and data-driven machine learning methods are common techniques to increase the reliability of leak detection using vibro-acoustic noise loggers. Most leak detection approaches in the literature extract features from an audio recording, which is either directly used to interpret signals for leakage [7,9–11], or used to train machine learning classifiers. Models trained with simple features such as the absolute noise level recorded by loggers [12], or cross-correlation and coherence signals from neighbouring correlating noise loggers [13] have also demonstrated high accuracies in leak localisation and classification, respectively. Other methods rely on having collected baseline signals or signals before and after a leak has been repaired [14–17], to establish leak detection thresholds. Due to the persistent nature of a leak signal in an audio recording, Recurrence Plots (RPs) offer an alternative input for a binary classifier, with RPs of leak noises showing strong deterministic properties [18].

Data-driven machine learning studies have leveraged frequency-domain features of acoustic signals for training such as the Power Spectrum Density (PSD) [14,19] or Intrinsic Mode Functions (IMFs) [20]. Whilst these features may prove effective for classification in controlled laboratory tests, they are easily influenced by a temporary ambient noise which can mask a persistent leak noise in the PSD, leading to decreased classification performance [21]. This limitation is critical for sensor deployments on functioning pipeline networks, where both persistent and transient non-leak noises are prevalent, leak noises are not controlled, and the pipe network can be complex. Many of these studies are conducted in controlled laboratory environments [22–25], with few examples of data sets obtained from real pipeline networks. Data collected from in-field deployments of vibro-acoustic sensors have predominantly contained unbalanced data sets, with small amounts of leak samples [18,26,27] or data collected with minimal interference noises, where Gaussian White Noise (GWN) with different Signal-to-noise Ratios (SNRs) are added to augment the data sets [21]. Unbalanced data sets remain a limitation in evaluating the success of any leak detection classifier, particularly for real-world sensor deployments where pipe materials,

diameters, soil properties, service lines, and offtakes, amongst other geospatial features, can vary significantly and heavily influence the signals recorded by the vibro-acoustic sensors.

Time–frequency features generated using discrete Short-time Fourier Transforms (STFTs), such as spectrograms, reveal the temporal nature of a signal that is not captured by analysing frequency-domain features alone. STFTs can provide rich features for machine learning; however, STFTs as standalone input features are rarely used for acoustic signal analysis, due to a limitation in the time–frequency resolution [28]. In an effort to balance the relationship between the time and frequency resolutions, a Time–Frequency Convolutional Neural Network (TFCNN), with three different spectrogram resolutions as inputs is proposed to study the efficacy of classification under varying SNR conditions in real pipeline networks [21]. The TFCNN model is compared against a range of other common classifiers, including a CNN trained with Fast Fourier Transform (FFT) data (Frequency Convolutional Neural Network (FCNN)). It is reported that the spectrogram contains sufficient defining characteristics of a leak signal (as opposed to time, or frequency-based features alone), and is therefore more favourable and reliable as an input to a leak detection system. Mel-frequency spectrograms, which closely align with the human perception of sound, are also commonly used as features in machine learning applications, including leak classification problems [29,30].

This paper evaluates state-of-the-art data-driven methods for leak classification using data collected from semi-permanent vibro-acoustic logger deployments in small reticulation mains across metropolitan Sydney over the course of up to 24 months. Data from a range of commercially available types of vibro-acoustic sensors deployed in different metropolitan areas of a utility-managed water network are used to evaluate the efficacy of existing data-driven methods (FCNN and TFCNN models [21]) for reliable leak detection in urban distribution mains.

The paper is organised as follows. Section 2 details the vibro-acoustic sensors and data loggers, data collection, signal processing, data curation, feature extraction and binary classification methods. Section 3 presents the results and discussion. Finally, the conclusions and future work are presented in Section 4.

2. Materials and Methods

2.1. Vibro-Acoustic Sensors and Data Loggers

Vibro-acoustic logging hardware consists of a vibro-acoustic sensor, data logger, and other peripherals such as GSM transmitters and antennas to send the collected data to the cloud. Vibro-acoustic sensors function on the premise that when water leaks through a pipe it creates vibrations due to the pressure differential between the inside and the outside of a pipe. The waves can travel thorough both pipe material and water, allowing the sensors to measure the vibration inflicted on the material, or directly in the water column. Standard manufacturer specifications indicate that vibro-acoustic sensors are effective in recording leakage noises on reticulation mains typically smaller than 375 mm in diameter, and can correlate over distances of up to 150 m between adjacent loggers.

In December 2019, a range of vibro-acoustic sensors deployments commenced across six Central Business District (CBD) areas in metropolitan Sydney (summarised in Table 1). In these CBD areas, five different types of commercially available semi-permanent vibro-acoustic loggers (see Figure 1) were deployed. These could not be collocated in the same spots to compare performance given the chamber’s physical limitations, and the extent of exposed asset to mount them on (see some examples Figure 2), and were thus distributed to cover separate areas and zones (when within the same area). It should also be noted that, given the attachment coupling of the sensor to the appurtenance, they can not physically measure the exact same point regardless, so arranging them over an extended geographical coverage of the city is more representative of a realistic deployment in a practical sense for comparison, and more effective to search for as many leaks as possible over a given time period for a more robust validation of the proposed scheme.



Figure 1. Range of vibro-acoustic loggers installed across metropolitan Sydney: (a) HWM PermaNET SU, (b) HWM PermaNET+, (c) SebaKMT Sebalog N-3, (d) Von Roll ORTOMAT-MTC, (e) Primayer Enigma3m.

Table 1. Vibro-Acoustic Logger Deployment Details.

| Manufacturer | Model | Deployment Area # | Frequency Range (Hz) | Sampling Rate (Hz) | Audio Recording Duration (s) | # Leaks Detected |
|--------------|-------------|-------------------|----------------------|--------------------|------------------------------|------------------|
| HWM | PermaNET+ | 1,2,3 | 0–2048 | 4096 | 10 | 19 |
| HWM | PermaNET SU | 1 | 0–2048 | 4096 | 10 | 18 |
| Von Roll | ORTOMAT-MTC | 2,4 | 0–2340 | 4681 | 10 | 16 |
| SebaKMT | Sebalog N-3 | 1,5 | 0–3277 | 6554 | 2.5 | 23 |
| Primayer | Enigma3m | 4,6 | 0–2500 | 5000 | 10 | 11 |

Each of the five different vibro-acoustic sensors and data loggers are functionally equivalent, whereby vibrations in the pipeline network are detected by the sensors and recorded with the data logging hardware. The key differences between the loggers are the quality of the hardware used, the level of processing of the data, both on the logger itself and the cloud-based portals, and the user programmable settings (e.g., audio recording duration and time).

The sensors have mostly been installed on appurtenances (valves and hydrants) attached to Cast Iron Cement Lined (CICL) or Steel Cement Lined (SCL) pipelines, ranging in diameter from 100 mm to 450 mm and up to more than 100 years old. Depending on the available space in a hydrant or valve chamber and the condition of the assets, the sensors are often mounted with differing orientations and mounting points, as shown in Figure 2.



Figure 2. HWM PermaNET SU loggers deployed on hydrant control valves in different locations.

2.2. Data Collection

Noises in the pipe network are measured every day at a time of low water usage and theoretically low environmental noise (between 2–4 a.m.). With the exception of the Sebalog N-3 vibro-acoustic sensors, all of the deployed sensors were programmed to record a 10-s duration audio file daily. The Sebalog N-3 units have limited configuration settings, thus, despite recording a 2.5 s duration audio clip every day, the audio file is only sent to the

cloud if the logger itself determines that a leak is present (through a noise level threshold algorithm). In addition to audio recordings, other noise-level data are also available for analysis from most of the loggers; however, these were not used in this study. All loggers are equipped with integrated modems and transmit data to the cloud, with the raw acoustic data (audio files) available through the sensor manufacturers FTP servers, or accessible through API calls.

The collected data consist of ‘leak’ and ‘no-leak’ audio recordings originating from a range of leak sources across the six deployment areas. Approximately 70% of the detected leaks were hidden, many of which were in built-up areas and estimated to have been present for up to 10 years. The detected leaks were found to have emerged from a range of sources, including hydrants ~30%, valves ~20%, main taps ~22%, private ~11%, service lines ~12%, mains (leaks/breaks) ~2.5%, and meter taps ~2.5%. Some examples of hidden leaks detected by the vibro-acoustic sensors are shown in Figure 3.

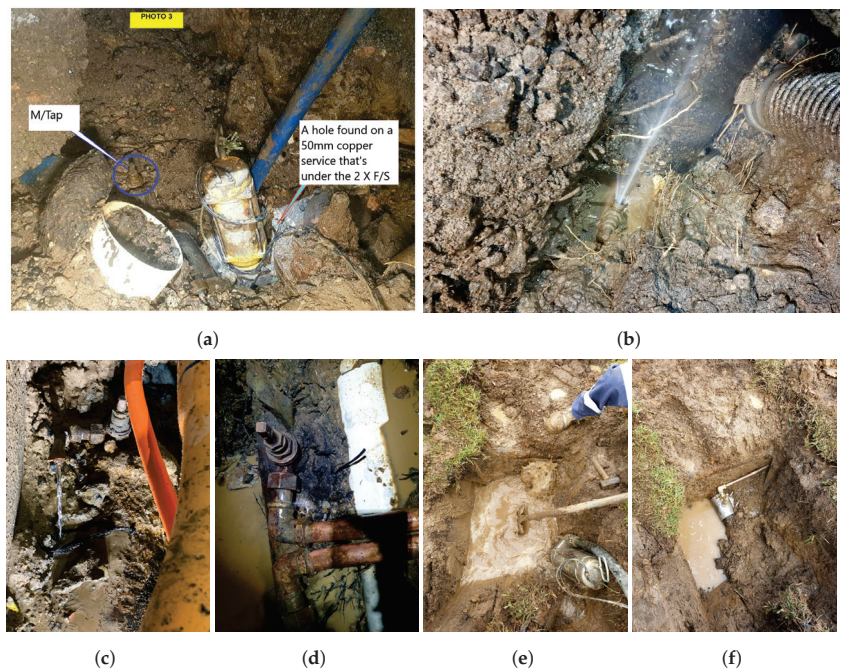


Figure 3. Examples of detected hidden leaks from a range of vibro-acoustic sensors and deployment areas (pictures supplied by utility field crews, taken during repairs): (a) Copper service leak, (b) Main tap leak, (c) Main tap leak (clamped service line), (d) Leak on main tap coupling piece, (e) leaking main tap (excavation site), and (f) leaking main tap repaired with full circumference pipe clamp.

The four logger data sets (HWM, Von Roll, SebaKMT, Primayer) mostly include loggers that recorded leak noises from the first day they were deployed. These existing leaks were monitored for several days to confirm the likelihood of the presence of a leak, prior to raising these locations for in-field investigation by the water utility. The leaks were confirmed on-site by skilled network technicians through use of listening sticks and pinpointed using real-time correlators. Significant delays were experienced with some repair jobs, due to the complex locations of some leaks. Consequently, many of the recorded leak signals contain the same underlying persistent leak noise, occasionally overlaid with transient environmental noises. As existing leaks were gradually repaired and baseline noise levels could be achieved, the emergence and evolution of new leaks were able to be identified and the data sets grew further in size over the course of the deployments. Since

only a small subset of all of the deployed loggers detected leaks, only these loggers were included in the data sets (both before and after leak repairs), to ensure a relative balance of the data sets. To improve on the robustness of the classification in the presence of other environmental noises, those loggers which only recorded ‘no-leak’ signals for the duration of their deployment could also be used.

2.3. Data Analysis-Signal Processing

Across the six deployment areas, a wide variety of leak noises were recorded. Some sensors were located very close to the leak source, and others at a distance, with variations in pipe diameters and materials, and several offtakes between. Using STFT signal processing techniques, acoustic signals can be best visualised by generating spectrograms, which reveal temporal changes to the frequency and power of a signal. If the audio recording contains persistent noise, without the presence of any intermittent external noises, PSD line plots can also provide a simple means of signal comparison. As leaks are continuous noise sources, their higher-power frequencies are persistent in the spectrum, for the duration of an audio recording. On the other hand, non-leak noises—such as those from environmental sources, or water usage—are mostly transient in nature, with intermittent frequency components. Some environmental noises, however, can be persistent, such as mechanical or electrical equipment which commonly emit high-power, low-frequency noises usually with narrow frequency bands. Due to these characteristic features, persistent and intermittent ‘no-leak’ signals are easily distinguishable from ‘leak’ signals in a spectrogram (see Figure 4 for an example). Due to the close coupling of the sensors to the water main, leaks generally have a distinguishing pattern in the audio spectrum, even in the presence of other intermittent noises.

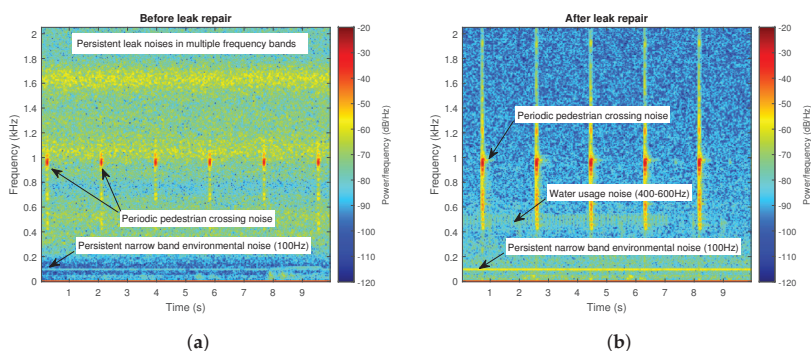


Figure 4. Spectrograms for a sensor detecting a variety of noises: (a) before and (b) after a leak was repaired ~21 m away from the sensor. The sensor was situated <1 m from a pedestrian crossing.

By clustering the loggers in the pipeline networks to ensure neighbouring loggers are able to correlate, often more than one logger was able to record noise from a single leak source; one such example is shown in Figure 5, where six vibro-acoustic sensors were able to detect the leak noise caused by a broken back on the pipe (main break). The shift in the dominant leak frequency can be observed with increased distance between the leak and the sensor. Other contributing factors to the frequency shift could also include pipe material change and junctions and offtakes between the leak and the sensor. In general, the further away the sensor is from the leak location, the more the higher-frequency components of the spectrum are attenuated, and the lower frequency noises are more prevalent. With increased distance between the logger and the noise source, the intensity (power) of the noise also decays.

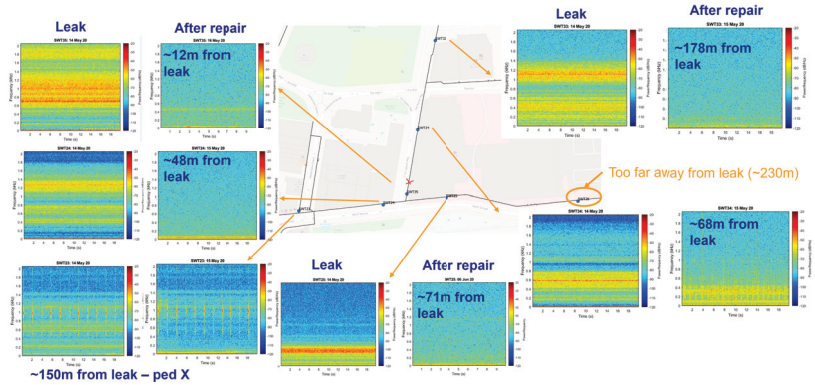


Figure 5. An illustration of a single leak source originating from a broken back pipe, and the noise spectrograms picked up by 6 HWM PermaNET+ loggers in the vicinity. The 7th sensor on the right is located too far away to pick it up.

A leak located close to the hydrant where the logger is installed will typically have elevated noise across the spectrum, often with higher power in high frequency band/s. Figure 6 shows PSD line plots from HWM vibro-acoustic sensors detecting leaks at the hydrant they were installed on. All leaks were on screw-down-type hydrants, and suspected to be of varying leak rates. The vibro-acoustic sensors were installed in different orientations and with different contact points on the hydrants, similar to those mounting configurations shown in Figure 2. There is a significant difference in the PSDs of each hydrant leak. The difference in signals could be attributed to many factors including the quality of the attachment point of the sensor on the asset or the magnitude of the leak. Comparing these signals to a ‘quiet’, baseline signal with no leak present, it is noted that all four leak signals show elevated power across almost all recorded frequencies, and clear peaks in the spectrum at certain frequencies. This indicates that despite leaking hydrant signals being inconsistent across multiple loggers/hydrants, there is still a significant deviation from a baseline ‘no-leak’ noise that is sufficient to detect a leak.

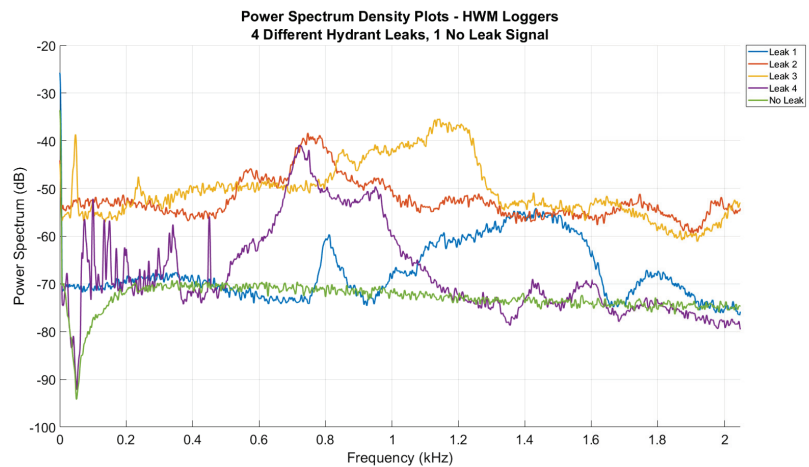


Figure 6. PSD plot: four different hydrant leak signals (loggers on leaking hydrants).

2.4. Data Analysis-Data Curation

In order to curate the collected data to train machine learning classifiers, the raw acoustic data were analysed in the time, frequency, and time–frequency domains using the signal processing and visualisation techniques (PSD, STFT, FFT) described in Section 2.3. Analysis of the vibro-acoustic data, in conjunction with feedback from the utility field crews, allowed for a database to be compiled with key information pertaining to the leaks. The collated and curated data consist of the audio file name, date of audio recording and binary class label ('leak' or 'no-leak'). Other collected information not used for the binary classification includes the leak source, and the distance, pipe material/s, and diameter/s between the leak and logger.

Most of the detected leaks were present prior to the loggers being installed; however, there were some instances where new leaks emerged during the logger deployment time. For the leaks that were already present, the collected acoustic signals were generally stable and unchanged in their frequency. In some instances, noticeable frequency/power shifts in the spectrum were observed (see Figure 7)—possibly from a leak worsening, or the sensor being slightly shifted/dislodged on the asset due to environmental factors or human intervention. These cases were carefully analysed to ensure that the data was representative of a true 'leak' or 'no-leak' signal, and the logger had not been dislodged from the asset.

The curated data from individual loggers were compiled into complete data sets for each logger *manufacturer* (for a total of four discrete data sets). Due to the slightly differing frequency ranges and audio recording duration (as listed in Table 1), individual classifiers were trained for each sensor manufacturer and were evaluated individually. With nearly 300 loggers deployed across the six deployment areas over the course of two years, the complete data sets from each logger manufacturer are vast. To ensure a relative balance of data for each data set, only data from loggers which recorded both 'leak' and 'no-leak' signals throughout their deployment are included in the data sets.

2.5. Feature Extraction and Binary Classification

To evaluate the performance of a binary classifier for each of the data sets, an extensive literature review on the topic of data-driven leak detection methods with acoustic data was first conducted. A critical criteria in determining the suitability of a classifier was the reported performance with data collected from real pipeline networks. With limited studies and evaluations utilising data from deployments of loggers outside of controlled laboratory environments, it was found that CNN-based classifiers leveraging features obtained from FFTs and STFTs (spectrograms) had the best reported performance, compared with other common binary classification models.

Both the FCNN and TFCNN models from [21] are trained and evaluated in this paper, using the four discrete data sets collected from the six deployment areas. The data sets were first prepared by augmenting [31] (splitting) each audio file into several 1 s audio chunks. For the SebaKMT loggers, only the first two seconds of the 2.5 s duration audio recordings were used. All other loggers (with 10 s duration) audio recordings were split into 10 individual audio chunks. Due to the vast array of samples, including various 'no-leak' noise sources, it was not deemed necessary to further augment the data sets by adding GWN with different SNRs into the raw signals. To extract the frequency bands of interest where leaks are most common, the 1 s duration audio samples are also bandpass filtered (100–2000 Hz). With the data sets collected and curated, finally, a random 80% of each complete data set (for each logger type) was used for training and 20% for testing. The models (whose structures are shown in Figure 8) were implemented in Python 3.9 using Keras [32] and TensorFlow [33] version 2.6.0.

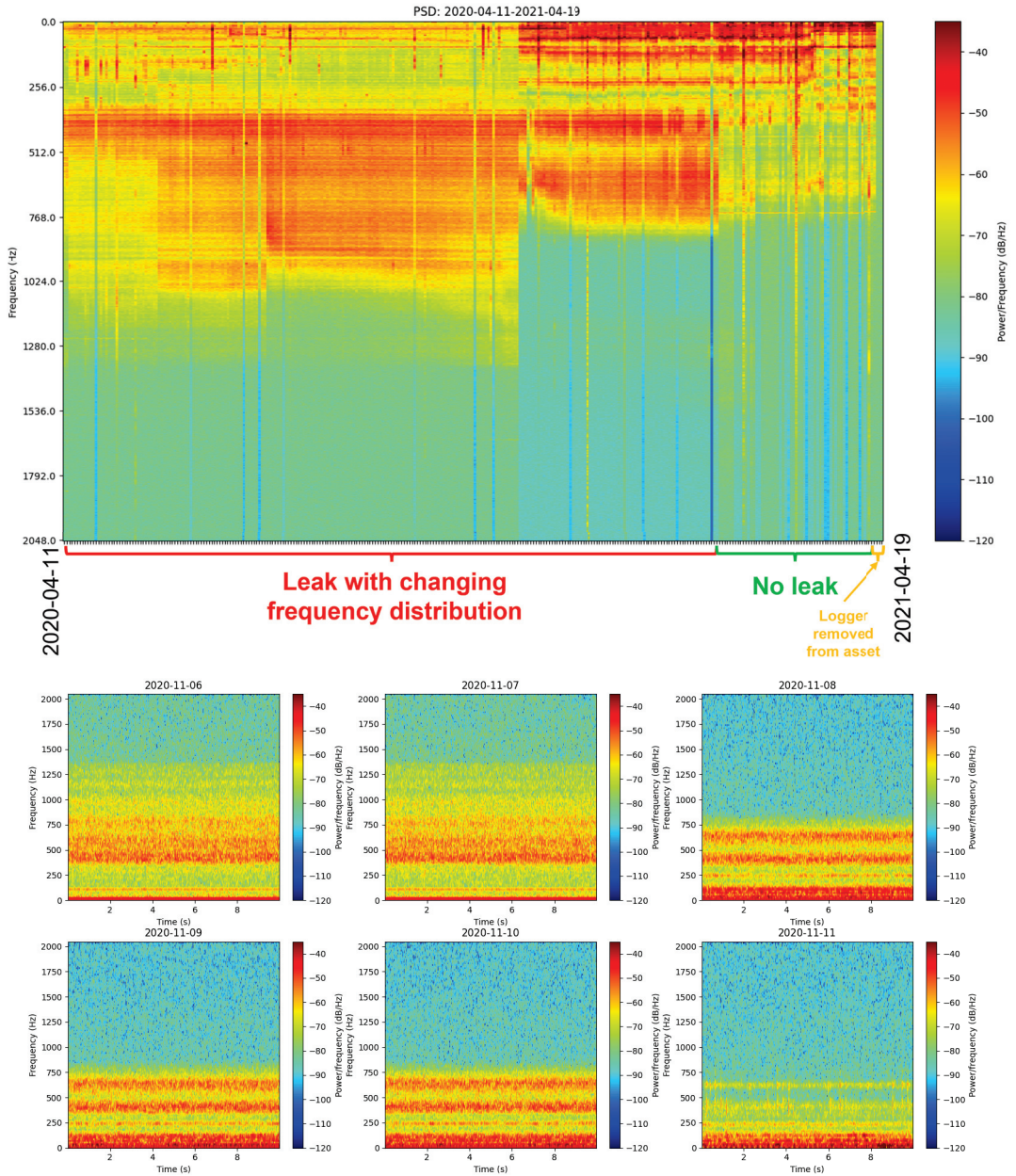


Figure 7. Leaking main tap-changing frequency distribution is visible in the PSD (top) and spectrograms (bottom) from consecutive days. The logger was situated approximately 59 m away from the leak location, with noise being propagated along a straight section of 150 mm diameter CI/CL pipe.

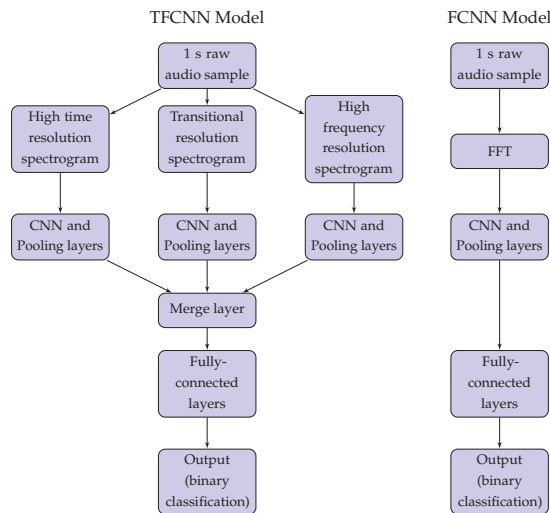


Figure 8. TFCNN and FCNN model structures.

The input to the FCNN model is purely frequency-domain based—a FFT of the 1 s audio signal. The inputs to the TFCNN model are three spectrograms generated from the same 1 s audio signal. Each spectrogram is generated with a different time–frequency resolution (high time, transitional, high frequency) and is intended to improve the leak detection performance, since ‘no-leak’ and ‘leak’ noises have different time–frequency components. A high-time-resolution spectrogram reflects the change of the signal in the time-domain, where a leak signal is most stable. In these spectrograms, the presence of any transient noises are most obvious. The high-frequency-resolution spectrogram reflects the spectral structure and energy distribution of the signal in the frequency domain. Whilst transient noises can still be observed in these spectrograms, the leak frequency or frequencies are best represented. Finally, the transitional time–frequency resolution is intended to balance the relationship between the time and frequency resolutions. Due to different sampling rates of the four sets of loggers, the dimensions of the three spectrograms which are the inputs for the TFCNN model differ slightly, as listed in Table 2.

Table 2. TFCNN model spectrogram matrix sizes for different resolutions.

| Logger Manufacturer | Audio Sampling Rate (Hz) | High Time | Spectrogram Resolution | |
|---------------------|--------------------------|-----------|------------------------|----------------|
| | | | Transitional | High Frequency |
| HWM | 4096 | [94, 60] | [186, 28] | [372, 12] |
| Von Roll | 4681 | [99, 70] | [197, 33] | [394, 15] |
| SebaKMT | 6554 | [72, 99] | [142, 48] | [283, 22] |
| Primayer | 5000 | [94, 75] | [184, 36] | [369, 16] |

3. Results and Discussion

Tables 3 and 4 summarise the results of the FCNN and TFCNN classification models for the four logger data sets. The metrics used to evaluate the model performance were accuracy, sensitivity, and specificity. The following abbreviations are used to simplify the presentation of the equations: True Positive (TP); True Negative (TN); False Positive (FP); False Negative (FN). Accuracy is the measure of the classifier’s overall correct classification performance: $TP + TN / (TP + TN + FP + FN)$. Sensitivity is the classifier’s ability to label a ‘leak’ signal as ‘leak’ (recall of the positive class): $TP / (TP + FN)$. Specificity is the classifier’s ability to label a ‘no-leak’ signal as ‘no-leak’ (recall of the negative class): $TN / (TN + FP)$.

Despite the excellent performance of the FCNN model, as was reported in [21], it was found that the TFCNN model consistently outperformed the FCNN model across each of the performance metrics studied (with the exception of the specificity of the HWM loggers). This indicates that the spectrogram-based inputs are more effective than purely frequency-domain-based inputs in representing the characteristics of both ‘leak’ and ‘no-leak’ signals for binary classification.

Table 3. FCNN Results.

| Logger Manufacturer | Total # Files | # Leak Files | # No Leak Files | Accuracy (%) | Specificity (%) | Sensitivity (%) |
|---------------------|---------------|--------------|-----------------|--------------|-----------------|-----------------|
| HWM | 67,730 | 36,310 | 31,420 | 98.18 | 99.19 | 97.32 |
| Von Roll | 37,160 | 14,850 | 22,310 | 96.60 | 98.57 | 93.60 |
| Seba KMT | 3072 | 1026 | 2046 | 87.97 | 95.09 | 71.66 |
| Primayer | 32,020 | 2950 | 29,070 | 96.83 | 98.94 | 75.77 |

Table 4. TFCNN Results.

| Logger Manufacturer | Total # Files | # Leak Files | # No Leak Files | Accuracy (%) | Specificity (%) | Sensitivity (%) |
|---------------------|---------------|--------------|-----------------|--------------|-----------------|-----------------|
| HWM | 67,730 | 36,310 | 31,420 | 98.46 | 98.92 | 98.07 |
| Von Roll | 37,160 | 14,850 | 22,310 | 98.51 | 99.75 | 96.61 |
| Seba KMT | 3072 | 1026 | 2046 | 94.63 | 95.33 | 93.05 |
| Primayer | 32,020 | 2950 | 29,070 | 97.95 | 99.59 | 81.62 |

Figures 9 and 10 show the confusion matrices for each of the four different TFCNN and FCNN trained models, respectively. For a practical leak detection system that water utilities can rely on, high accuracy but also high specificity (true negative) and sensitivity (true positive) rates are key performance metrics. A reliable leak detection system will minimise the false positive leak alarms, to ensure that any follow-up field investigations are for real leak events, maximising the efficiency for utilities.

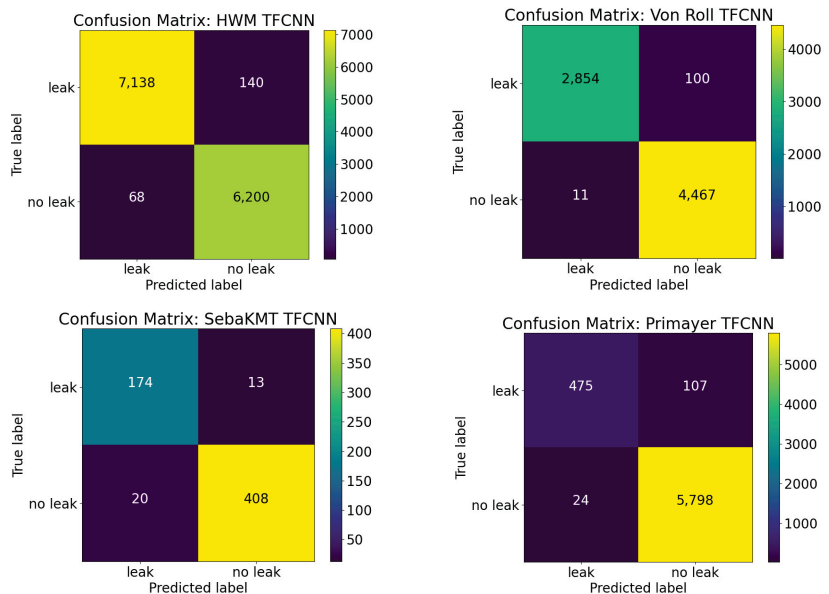


Figure 9. Confusion matrices for TFCNN models. HWM (top left), Von Roll (top right), Seba KMT (bottom left) and Primayer (bottom right).

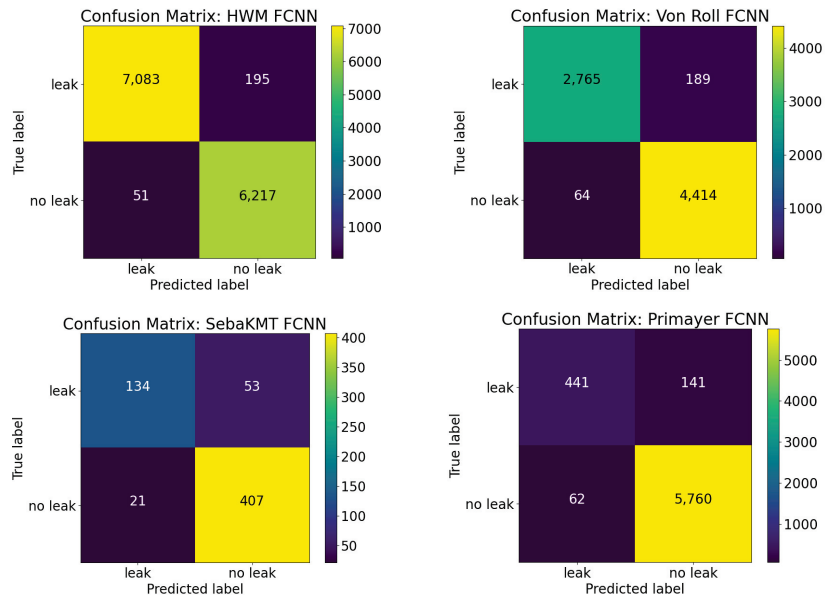


Figure 10. Confusion matrices for FCNN models. HWM (top left), Von Roll (top right), Seba KMT (bottom left) and Primayer (bottom right).

Despite the limited data available from SebaKMT Sebalog N-3 loggers and a data imbalance with ‘leak’ and ‘no-leak’ signals across 3/4 of the data sets, the results indicate that the type of sensor used (different vibro-acoustic sensor with different sampling rate, sensitivity, etc.) does not affect the performance of the classifier. Furthermore, the results demonstrate that a leak detection system using either the FCNN or TFCNN model can be effectively trained with data from a single location both before and after a leak repair.

The excellent classification results show that—irrespective of the type of vibro-acoustic sensor used—the classifiers have been able to learn sufficiently with data from a range of deployment areas, where leak sources, pipe sizes and materials as well as soil conditions have varied widely. The results indicate that this is particularly relevant for identifying leaks in built-up CBD areas, where a variety of ‘no-leak’ persistent and transient environmental noises are prevalent, even in the early hours of the morning. Considering all of the factors that affect the recorded vibro-acoustic signals, the results presented show great promise for water utilities looking to integrate the use of semi-permanent vibro-acoustic sensors into their business-as-usual practice for structural pipe health monitoring. Through the use of vibro-acoustic sensors and early detection of hidden leaks, proactive maintenance can be scheduled and conducted, with minimal impact to the customer.

The classification performance may be improved by including a large number of ‘no-leak’ signals from elsewhere in the pipeline network during a deployment i.e., by including those other loggers that did not record both ‘leak’ and ‘no-leak’ signals in the data set. This will help further train the classifier to better discriminate between ‘leak’ and ‘no-leak’ noises, further increasing the reliability and robustness of the classification.

4. Conclusions

This paper studied and analysed the performance of a range of different semi-permanent vibro-acoustic sensors deployed in six CBD areas across wider Sydney for extended periods of time. Following careful collation, analysis and curation of the collected acoustic data, two state-of-the-art CNN-based classification models (FCNN and TFCNN) were trained and tested for each of the four logger types.

The results presented point towards the potency of FFT and STFT signal processing for CNN-based classification of vibro-acoustic measurements. Moreover, they represent the first known documented comparison of a variety of different semi-permanent sensing hardware, with a special underscore on the study having been undertaken on live deployments. The results demonstrate that these state-of-the-art methods are not only applicable to one particular make and model of semi-permanent acoustic sensor, as was previously documented in the single relevant case study found in the literature. Classification accuracies in the range of [94.63–98.51%] were achieved with the best performer, the TFCNN model, for all the sensors studied.

Future work to enhance the results of this study would involve obtaining further validated data collected from a wider variety of deployment locations and CBD areas. As indicated in Section 3, the robustness and reliability of these classifiers may also be improved by adding further existing ‘no-leak’ audio recordings into the data sets. Finally, despite their sensing hardware similarities, a comparison of the classification performance of semi-permanent and Lift and Shift (L&S) vibro-acoustic sensors (intended for short-term deployments, rather than continuous monitoring) would also provide further insights into the potential success and value of implementing smart leak detection methods for utilities.

Author Contributions: Conceptualisation, L.B. and J.V.M.; methodology, L.B. and J.V.M.; software, L.B.; validation, L.B. and J.V.M.; formal analysis, L.B.; investigation, L.B.; resources, J.V.M.; data curation, L.B.; writing—original draft preparation, L.B.; writing—review and editing, J.V.M.; visualisation, L.B.; supervision, J.V.M.; project administration, L.B. and J.V.M.; funding acquisition, J.V.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study cannot be made publicly available due to confidentiality; readers should contact the corresponding author for details.

Acknowledgments: The authors would like to thank the authors of [21] for their feedback and discussions about their TFCNN work during the development of the work presented in this manuscript.

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

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|-------|---|
| ALD | Active Leak Detection |
| API | Application Programming Interface |
| CBD | Central Business District |
| CICL | Cast Iron Cement Lined |
| CNN | Convolutional Neural Network |
| FCNN | Frequency Convolutional Neural Network |
| FFT | Fast Fourier Transform |
| FTP | File Transfer Protocol |
| GSM | Global System for Mobile communication |
| GWN | Gaussian White Noise |
| IMF | Intrinsic Mode Function |
| L&S | Lift and Shift (vibro-acoustic sensors) |
| MNF | Minimum Night Flow |
| PSD | Power Spectrum Density |
| RP | Recurrence Plot |
| SCL | Steel Cement Lined |
| SNR | Signal-to-noise Ratio |
| STFT | Short-time Fourier Transform |
| TFCNN | Time–Frequency Convolutional Neural Network |

References

- Alves Coelho, J.; Glória, A.; Sebastião, P. Precise Water Leak Detection Using Machine Learning and Real-Time Sensor Data. *IoT* **2020**, *1*, 474–493. [\[CrossRef\]](#)
- Hamilton, S.; Charalambous, B. *Leak Detection: Technology and Implementation*, 1st ed.; IWA Publishing: London, UK, 2013; pp. 13–14.
- Hunaidi, O.; Wang, A.; Bracken, M.; Gambino, T.; Fricke, C. Acoustic methods for locating leaks in municipal water pipe networks. In Proceedings of the International Conference on Water Demand Management, Dead Sea, Jordan, 30 May–3 June 2004.
- Age of Data-driven Leak Management Dawns. *Water Online*, 17 July 2020. Available online: www.wateronline.com/doc/age-of-data-driven-leak-management-dawns-0001 (accessed on 30 August 2022).
- Noise Loggers—The Technology Which Is Transforming Leak Detection for the United Kingdom’s Pipe Network. Available online: piperepair.co.uk/2020/06/30/noise-loggers-the-technology-which-is-transforming-leak-detection-for-the-united-kingdoms-pipe-network/ (accessed on 30 August 2022).
- Fuchs, H.V.; Riehle, R. Ten Years of Experience with Leak Detection by Acoustic Signal Analysis. *Appl. Acoust.* **1991**, *33*, 1–19. [\[CrossRef\]](#)
- Gong, J.; Lambert, M.F.; Stephens, M.L.; Cazzolato, B.S.; Zhang, C. Detection of Emerging through-Wall Cracks for Pipe Break Early Warning in Water Distribution Systems Using Permanent Acoustic Monitoring and Acoustic Wave Analysis. *Water Resour. Manag.* **2020**, *34*, 2419–2432. [\[CrossRef\]](#)
- Pilcher, R.; Hamilton, S.; Chapman, H.; Field, D.; Ristovski, B.; Stapely, S. *Leak Location and Repair: Guidance Notes; Version 1*; International Water Association: London, UK, 2007; p. 16.
- Stephens, M.; Gong, J.; Zhang, C.; Marchi, A.; Dix, L.; Lambert, M.F. Leak-Before-Break Main Failure Prevention for Water Distribution Pipes Using Acoustic Smart Water Technologies: Case Study in Adelaide. *J. Water Resour. Plann. Manag.* **2020**, *146*, 05020020. [\[CrossRef\]](#)
- Marmarokopos, K.; Doukakis, D.; Frantziskonis, G.; Avlonitis, M. Leak Detection in Plastic Water Supply Pipes with a High Signal-to-Noise Ratio Accelerometer. *Meas. Control* **2018**, *51*, 27–37. [\[CrossRef\]](#)
- Martini, A.; Troncosi, M.; Rivola, A. Automatic Leak Detection in Buried Plastic Pipes of Water Supply Networks by Means of Vibration Measurements. *Shock Vib.* **2015**, *2015*, 165304. [\[CrossRef\]](#)
- El-Abbasy, M.S.; Mosleh, F.; Senouci, A.; Zayed, T.; Al-Derham, H. Locating Leaks in Water Mains Using Noise Loggers. *J. Infrastruct. Syst.* **2016**, *22*, 04016012. [\[CrossRef\]](#)
- Burtea, V.; Murray, P. How Well Can Machine Learning Support Pipeline Leak Monitoring? In Proceedings of the Pipelines 2021, Virtual Conference, 3–6 August 2021.
- El-Zahab, S.; Asaad, A.; Abdelkader, E.M.; Zayed, T. Collective Thinking Approach for Improving Leak Detection Systems. *Smart Water* **2017**, *2*, 3. [\[CrossRef\]](#)
- Cody, R.A.; Tolson, B.A.; Orchard, J. Detecting Leaks in Water Distribution Pipes Using a Deep Autoencoder and Hydroacoustic Spectrograms. *J. Comput. Civ. Eng.* **2020**, *34*, 04020001. [\[CrossRef\]](#)
- Martini, A.; Troncosi, M.; Rivola, A. Leak Detection in Water-Filled Small-Diameter Polyethylene Pipes by Means of Acoustic Emission Measurements. *Appl. Sci.* **2017**, *7*, 2. [\[CrossRef\]](#)
- Tariq, S.; Bakhtawar, B.; Zayed, T. Data-driven application of MEMS-based accelerometers for leak detection in water distribution networks. *Sci. Total Environ.* **2022**, *809*, 151110. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nam, Y.W.; Arai, Y.; Kunizane, T.; Koizumi, A. Water leak detection based on convolutional neural network using actual leak sounds and the hold-out method. *Water Supply* **2021**, *21*, 3477–3485. [\[CrossRef\]](#)
- Teruhi, S.; Yamaguchi, Y.; Akahani, J. Water Leakage Detection System for Underground Pipes by Using Wireless Sensors and Machine Learning. *J. Disaster Res.* **2017**, *12*, 557–568. [\[CrossRef\]](#)
- Butterfield, J.D.; Meyers, G.; Meruane, V.; Collins, R.P.; Beck, S.B.M. Experimental investigation into techniques to predict leak shapes in water distribution systems using vibration measurements. *J. Hydroinf.* **2018**, *20*, 815–828. [\[CrossRef\]](#)
- Guo, G.; Yu, X.; Liu, S.; Ma, Z. Leakage Detection in Water Distribution Systems Based on Time–Frequency Convolutional Neural Network. *J. Water Resour. Plann. Manag.* **2021**, *147*, 04020101. [\[CrossRef\]](#)
- El-Zahab, S.; Abdelkader, E.M.; Zayed, T. An accelerometer-based leak detection system. *Mech. Syst. Sig. Process.* **2018**, *108*, 276–291. [\[CrossRef\]](#)
- Ismail, M.I.M.; Dziyauddin, R.A.; Salleh, N.A.A.; Muhammad-Sukki, F.; Bani, N.A.; Izhar, M.A.M.; Latiff, L.A. A Review of Vibration Detection Methods Using Accelerometer Sensors for Water Pipeline Leakage. *IEEE Access* **2019**, *7*, 51965–51981. [\[CrossRef\]](#)
- Martini, A.; Rivola, A.; Troncosi, M. Autocorrelation analysis of vibroacoustic signals measured in a test field for water leak detection. *Appl. Sci.* **2018**, *8*, 2450. [\[CrossRef\]](#)
- Chi, Z.; Li, Y.; Wang, W.; Xu, C.; Yuan, R. Detection of water pipeline leakage based on random forest. *J. Phys. Conf. Ser.* **2021**, *1978*, 012044. [\[CrossRef\]](#)
- Ravichandran, T.; Gavahi, K.; Ponnambalam, K.; Burtea, V.; Jamshid Mousavi, S. Ensemble-based machine learning approach for improved leak detection in water mains. *J. Hydroinf.* **2021**, *23*, 307–323. [\[CrossRef\]](#)
- Tijani, I.A.; Abdelmageed, S.; Fares, A.; Fan, K.H.; Hu, Z.Y.; Zayed, T. Improving the leak detection efficiency in water distribution networks using noise loggers. *Sci. Total Environ.* **2022**, *821*, 153530. [\[CrossRef\]](#) [\[PubMed\]](#)

28. Ahadi, M.; Bakhtiar, M.S. Leak detection in water-filled plastic pipes through the application of tuned wavelet transforms to Acoustic Emission signals. *Appl. Acoust.* **2010**, *71*, 634–639. [[CrossRef](#)]
29. Chuang, W.; Tsai, Y.; Wang, L. Leak Detection in Water Distribution Pipes Based on CNN with Mel Frequency Cepstral Coefficients. In Proceedings of the 3rd International Conference on Innovation in Artificial Intelligence, Suzhou, China, 15–18 March 2019.
30. Müller, R.; Illium, S.; Ritz, F.; Schröder, T.; Platschek, C.; Ochs, J.; Linnhoff-Popien, C. Acoustic Leak Detection in Water Networks. In Proceedings of the 13th International Conference on Agents and Artificial Intelligence (Volume 2), Vienna, Austria, 4–6 February 2021.
31. Kang, J.; Park, Y.; Lee, J.; Wang, S.; Eom, D. Novel Leakage Detection by Ensemble CNN-SVM and Graph-Based Localization in Water Distribution Systems. *IEEE Trans. Ind. Electron.* **2018**, *65*, 4279–4289. [[CrossRef](#)]
32. Chollet, F. Keras. Available online: github.com/fchollet/keras (accessed on 30 August 2022).
33. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. Tensorflow: A System for Large-scale Machine Learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.

Article

Composite Multiscale Transition Permutation Entropy-Based Fault Diagnosis of Bearings

Jing Guo ^{1,2}, Biao Ma ¹, Tiangang Zou ², Lin Gui ² and Yongbo Li ^{3,*}¹ School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China² China North Vehicle Research Institute, Beijing 100072, China³ MIIT Key Laboratory of Dynamics and Control of Complex System, School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China

* Correspondence: yongbo@nwpu.edu.cn

Abstract: When considering the transition probability matrix of ordinal patterns, transition permutation entropy (TPE) can effectively extract fault features by quantifying the irregularity and complexity of signals. However, TPE can only characterize the complexity of the vibration signals at a single scale. Therefore, a multiscale transition permutation entropy (MTPE) technique has been proposed. However, the original multiscale method still has some inherent defects in the coarse-grained process, such as considerably shortening the length of time series at large scale, which leads to a low entropy evaluation accuracy. In order to solve these problems, a composite multiscale transition permutation entropy (CMTPE) method was proposed in order to improve the incomplete coarse-grained analysis of MTPE by avoiding the loss of some key information in the original fault signals, and to improve the performance of feature extraction, robustness to noise, and accuracy of entropy estimation. A fault diagnosis strategy based on CMTPE and an extreme learning machine (ELM) was proposed. Both simulation and experimental signals verified the advantages of the proposed CMTPE method. The results show that, compared with other comparison strategies, this strategy has better robustness, and can carry out feature recognition and bearing fault diagnosis more accurately and with improved stability.

Citation: Guo, J.; Ma, B.; Zou, T.; Gui, L.; Li, Y. Composite Multiscale Transition Permutation Entropy-Based Fault Diagnosis of Bearings. *Sensors* **2022**, *22*, 7809. <https://doi.org/10.3390/s22207809>

Academic Editor: Giovanni Betta

Received: 17 August 2022

Accepted: 21 September 2022

Published: 14 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: composite multiscale transition permutation entropy; bearing; fault diagnosis; feature extraction

1. Introduction

Rotating machinery is essential mechanical equipment which has been widely used in large-scale industries, such as aerospace, vehicle engineering, electrical engineering, machinery manufacturing, and so on. Bearings are an important part of electric and power transmissions, and a bearing fault is one of the main causes of rotating machinery faults [1,2]. Bearing fault diagnosis is vital for the healthy maintenance and reliable operation of rotating machinery. Bearing health detection can reduce the occurrence of rotating machinery failures, thus ensuring system safety and reducing maintenance costs [3].

In bearing health monitoring, vibration signal analysis is a commonly used fault feature extraction method. This is because the vibration signal contains a wealth of useful fault information [4]. In recent decades, bearing vibration signal processing and pattern recognition have become a research hotspot in the field of fault diagnosis. Kankar et al. [5] used an artificial neural network (ANN) and a support vector machine (SVM) to diagnose bearing faults, and verified that machine learning can be used for the automatic diagnosis of a bearing fault. With the development of deep learning algorithms, neural networks such as convolutional neural networks (CNN) have been effectively used in bearing fault diagnosis [6,7]. The authors of [8] proposed a deep learning model to preprocess the original signal for noise removal to overcome the shortcoming of the traditional intelligent method being greatly affected by noise. However, in practice, the vibration signal of the bearing has obvious nonlinear and non-stationary characteristics. Therefore, the analysis of

nonlinear dynamic behavior and the extraction of useful and reliable fault features have become the key steps in fault diagnosis.

Entropy methods can quantify the dynamic trend and randomness of a nonlinear time series. In recent years, the use of entropy-based methods has become an important tool for analyzing signal complexity and feature extraction, and has been effectively used in fault diagnosis [9]. At present, approximate entropy (AE), sample entropy (SE), permutation entropy (PE), fuzzy entropy (FE), and diversity entropy (DE) methods are widely used in fault diagnosis of rotating machinery. AE was proposed by Pincus [10], and can be used to measure the regularity of a time series. Richman et al. [11] proposed SE, which uses the association dimension to induce SE to show relative consistency, and its complexity analysis performance is better than that of AE. However, SE has the disadvantage of relying heavily on data length [12]. The FE method proposed by Chen et al. [13] is an improvement on the SE method. Other researchers [14] have used FE to measure the complexity of vibration signals, and they verified the excellent dynamic tracking performance of FE and its ability to obtain a more accurate complexity estimation. In consideration of noise resistance and computational efficiency, Wang et al. [15] proposed DE, which uses cosine similarity to measure the divergence of orbits. Bandt et al. [16] proposed PE to calculate the state probability of track arrangement order, showing it has high computational efficiency and good feature extraction effects in signal processing.

In order to overcome the problem of insufficient information analysis when using SE to evaluate the dynamic characteristics and randomness of complex data, Costa et al. [17] proposed using multiscale sample entropy (MSE) to evaluate the complexity of time series over a range of scales. MSE has been successfully applied to analyze vibration signals generated by various dynamic behaviors [18–20]. Based on the same coarsening process as MSE, FE, PE, and DE can be extended to multiscale fuzzy entropy (MFE) [21–23], multiscale permutation entropy (MPE) [24–26], and multiscale diversity entropy (MDE) [15]. Through coarse-grained processing, the original time series can be divided into several short time series. The coarse-grained time series can represent the dynamic distribution characteristics of the original signal at a certain scale. Therefore, multiscale processing enhances the performance of entropy methods in evaluating signal complexity. On a multiscale basis, the combination of the symbol dynamic filtering process and the entropy method can not only remove noise, but also significantly improve the computational efficiency and feature extraction ability [27–29].

Recently, Zhang et al. [30] proposed a novel complexity estimation method, transition permutation entropy (TPE). TPE is different from the other methods in that it extracts the features of a time series from the transition probability matrix of ordinal patterns. Because the eigenvalue is very important when analyzing the dynamic behavior, TPE uses the positive eigenvalue of the transition probability matrix to calculate the entropy. This improves the feature identification performance of a time series. However, TPE only analyzes a time series using a single scale, which reduces the accuracy and comprehensiveness of the information analysis. Therefore, in this work we extended TPE to multiscale analysis. In the traditional multiscale calculation method, the coarse-grained time series is obtained by calculating the arithmetic mean of adjacent data points on the original time series without overlapping. The length of the coarse-grained time series obtained in this way is too short at large scale, and the accuracy and stability will be affected. Therefore, in this work, composite multiscale transition permutation entropy (CMTPE) was proposed as a way to solve these obstacles. When the composite multiscale method is used to coarse-grain the original time series, a coarse-grained time series with different starting points can be obtained at each scale, and the number is equal to the scale factor. Each coarse-grained time series can characterize the dynamic characteristics and randomness of the original signal, which can effectively enhance the accuracy and stability of TPE. CMTPE not only had excellent feature extraction performance, but also had better robustness to noise. The superiority of the proposed CMTPE method was verified by the simulation and experimental signals of bearing faults. The main contributions of this study are given as follows:

- (1) In order to enhance the accuracy of feature extraction and the comprehensiveness of information analysis, CMTPE was proposed as a strategy to quantify the complexity of time series.
- (2) A fault diagnosis strategy based on CMTPE and ELM was proposed for bearing fault type identification to identify the fault types of bearings.
- (3) The advantages of CMTPE in feature extraction were verified by simulation and experimental signals. Comparing TPE, MPE, and MTPE, the results showed that CMTPE has the highest fault diagnosis performance.

The rest of this paper is organized as follows: The concept of CMTPE is introduced in Section 2. In Section 3, results from simulation signals used to validate the superiority of CMTPE are reported. In Section 4, the effectiveness of CMTPE was verified using experimental signals. Finally, the conclusion of this article is provided in Section 5.

2. Methodology

In this section, the theories of TPE and MTPE are introduced in detail. In addition, the concept of the CMTPE algorithm is proposed.

2.1. Transition Permutation Entropy (TPE)

A time series of length N can be written as $X = \{x_1, x_2, \dots, x_N\}$. The TPE algorithm is introduced as follows:

Step 1. According to the phase space embedding theory, reconstruct X into a series of vectors with embedding dimension m . The reconstructed phase space is as follows:

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_{N-m+1} \\ x_2 & x_3 & \cdots & x_{N-m+2} \\ \vdots & \vdots & \ddots & \vdots \\ x_m & x_{m+1} & \cdots & x_N \end{bmatrix} \tag{1}$$

The reconstructed vectors can be expressed as $X_i = \{x_i, x_{i+1}, \dots, x_{i+m-1}\}, 1 \leq i \leq N - m + 1$.

Step 2. Compare the size relationship of the elements in the vector, so as to identify the ordinal pattern of each vector. When the embedding dimension is m , there are $m!$ possible ordinal patterns for any vector. For example, if the embedding dimension $m = 3$, there are 6 ordinal patterns. The size relationship of all vectors can be expressed by the size relationship of 0, 1, 2. For vector $(x_{k-1}, x_k, x_{k+1}) = (18, 3, 15)$, the element size relationship is $x_k < x_{k+1} < x_{k-1}$, and its corresponding ordinal pattern is $\pi = 2, 0, 1$.

Step 3. Calculate the transition probability between the ordinal patterns corresponding to all vectors to obtain the following transition probability matrix P :

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \tag{2}$$

where $n = m!$, and p_{ij} represents the probability of an transition from pattern π_i to pattern π_j .

Step 4. Calculate the TPE using the positive eigenvalue of matrix P . If the eigenvalue is complex number, its real part is taken. If P has n positive eigenvalues λ_i , TPE is calculated as follows:

$$\text{TPE}(X, m) = - \sum_{i=1}^n \frac{\lambda_i}{m!} \log \frac{\lambda_i}{m!} \tag{3}$$

2.2. Multiscale Transition Permutation Entropy (MTPE)

The entropy calculated from a single scale can only provide poor fault information. Multiscale analysis can extract more useful information from time series of different scales. The MTPE algorithm consists of two steps: (1) Conducting a coarse-graining process to obtain the series of the original time series at different scales; and (2) Calculating the TPE of each coarse-grained time series. First, divide the time series $X = \{x_1, x_2, \dots, x_N\}$ into multiscale time series $Y = \{Y_1, Y_2, \dots, Y_\tau\}$. The scale factor τ is a positive integer. The time series at any scale is $Y_\tau = [y_{1,\tau}, y_{2,\tau}, \dots, y_{j,\tau}]$, $j = N/\tau$, and the calculation is as follows:

$$y_{s,\tau} = \frac{1}{\tau} \sum_{i=\tau(s-1)+1}^{\tau s} x_i \tag{4}$$

Then, the time series of all scales obtained from the above process can be substituted into the TPE algorithm to calculate the MTPE as follows:

$$\text{MTPE}(X, m, \tau) = \text{TPE}(Y_\tau, m) \tag{5}$$

2.3. Composite Multiscale Transition Permutation Entropy (CMTPE)

In order to further improve the accuracy and stability of MTPE, CMTPE was proposed. When the scale factor is τ , τ different time series can be obtained. MTPE only considers the first coarse-grained time series at each scale, while CMTPE considers all τ coarse-grained time series. As shown in Figure 1, when the scale factor $\tau = 3$, MTPE only calculates one coarse-grained time series $y_1^{(3)}$, while CMTPE calculates three coarse-grained time series $y_1^{(3)}$, $y_2^{(3)}$, and $y_3^{(3)}$. Divide the time series $X = \{x_1, x_2, \dots, x_N\}$ into multiscale time series $Y = \{Y_1, Y_2, \dots, Y_\tau\}$. The time series at any scale is $Y_\tau = [y_{1,\tau}^1, y_{2,\tau}^2, \dots, y_{j,\tau}^k, \dots, y_{j,\tau}^\tau]$, where $y_{j,\tau}^k = [y_{1,\tau}^k, y_{2,\tau}^k, \dots, y_{j,\tau}^k]$. The calculation is as follows:

$$y_{j,\tau}^k = \frac{1}{\tau} \sum_{i=\tau(j-1)+k}^{\tau j+k-1} x_i \tag{6}$$

CMTPE is the mean of the TPE values for all coarse-grained time series, that is,

$$\text{CMTPE}(X, m, \tau) = \frac{1}{\tau} \sum_{k=1}^{\tau} \text{TPE}(y_{j,\tau}^k, m) \tag{7}$$

CMTPE considers all τ different coarse-grained time series, at each scale factor τ . Therefore, CMTPE can extract more fault information from the original time series. The entropy calculated by this method is more accurate and stable than that calculated by MTPE.

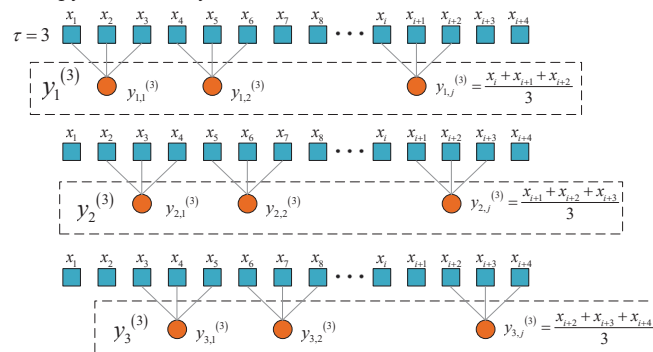


Figure 1. Schematic diagram of multiscale coarsening process when scale factor $\tau = 3$.

2.4. CMTPE Based Fault Diagnosis Strategy

In this work, a fault diagnosis strategy based on CMTPE was proposed. In this strategy, an ELM classifier was used to identify different fault types. The overall fault diagnosis framework is shown in Figure 2 [31]. The main steps were as follows:

Step 1. The vibration signals of bearings under different health conditions are measured by sensors.

Step 2. CMTPE is used for feature extraction of vibration signals. Each health condition will provide the corresponding entropy characteristics, representing the complexity of different vibration signals.

Step 3. A part of the fault features is randomly selected as a training set to train the ELM classifier.

Step 4. The remaining features are used as the test set to test the trained ELM, and the fault recognition rate is obtained. Steps 3 and 4 are run 20 times to obtain the average test accuracy.

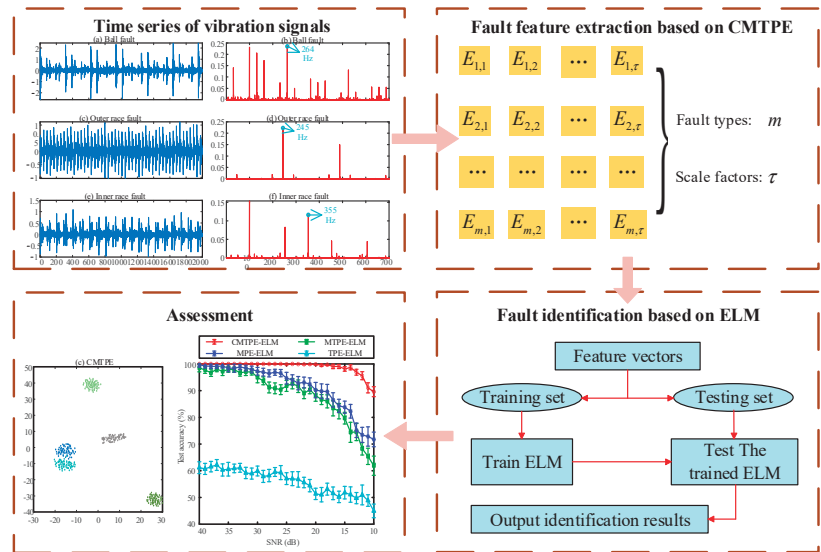


Figure 2. The overall fault diagnosis framework.

3. Simulation Evaluation

3.1. Simulated Bearing Signal

In this section, in order to verify the effectiveness and advantages of the proposed CMTPE, we detail the three types of simulated bearing faults which were designed: outer race fault, inner race fault, and ball fault models. The schematic diagram of the three simulated faults is shown in Figure 3.

In the load area, as shown in Figure 3, the sensor was installed at the maximum load density. Figure 3a shows the fault model of an outer race fault. Since the location of a localized defect will not change with time, the impulse force can be regarded as an ideal force. Figure 3b shows the fault model of an inner race fault, which has the same basic assumptions as the outer race fault model. At the peak of the load area, the ball will contact with a localized defect, resulting in the first impulse. After that, the localized defect will rotate with the inner race, so the contact position between the ball and the inner race will change with time. This type of contact will generate an impulse force only when it occurs in the load area. Figure 3c shows the ball fault model, which also has the same basic assumptions as the outer race fault model. In contrast to the inner race fault, a localized

defect will rotate with the ball, and the defect will continuously contact the inner and outer races to continuously generate impulse force [32].

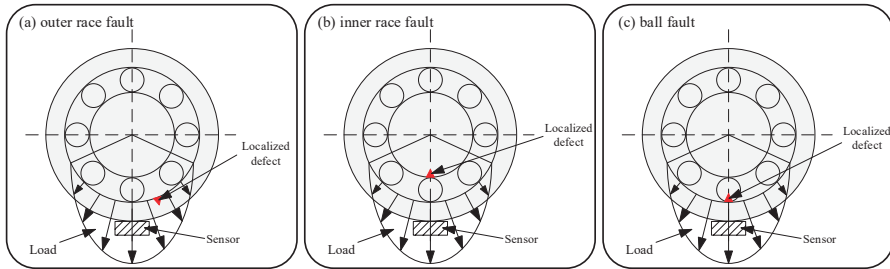


Figure 3. The schematic of simulated bearing faults.

The simulated bearing type was an N205 cylindrical roller bearing. The rotating speed was 3000 rpm. The sampling frequency was 10,240 Hz. The detailed bearing dimensions are shown in Table 1.

Table 1. Bearing parameters.

| Parameter | Value |
|------------------------------|-----------|
| Pitch circle diameter | 35.5 mm |
| Roller diameter | 6.5 mm |
| Rotating speed | 3000 rpm |
| Number of rollers | 12 |
| Sample frequency | 10,240 Hz |
| Natural frequency of bearing | 4000 Hz |
| Contact angle | 0° |

The fault frequency can be calculated according to the parameters in Table 1. Main parameters of the bearing: roller diameter $d = 6.5\text{mm}$; pitch circle diameter $D = 35.5\text{mm}$; number of rollers $Z = 12$; contact angle $\alpha = 0$; rotating speed $v = 3000\text{rpm}$. The fault frequency was calculated as follows:

(1) Outer race fault characteristic frequency f_0

$$f_0 = \frac{1}{2}Z \left(1 - \frac{d}{D} \cos \alpha \right) \frac{v}{60} = 245.0704(\text{Hz}) \quad (8)$$

(2) Inner race fault characteristic frequency f_i

$$f_i = \frac{1}{2}Z \left(1 + \frac{d}{D} \cos \alpha \right) \frac{v}{60} = 354.9296(\text{Hz}) \quad (9)$$

(3) Ball fault characteristic frequency f_e

$$f_e = \frac{D}{d} \left(1 - \left(\frac{d}{D} \right)^2 \cos^2 \alpha \right) \frac{v}{60} = 263.9220(\text{Hz}) \quad (10)$$

Figure 4 shows the time domain and envelope spectrum of each of the three simulated fault types. Among the data, Figure 4a,c,e depicts the time domain diagrams of the three faults, and Figure 4b,d,f shows the corresponding envelope spectrum diagrams. The fault frequency is marked with a blue arrow in the envelope spectrum.

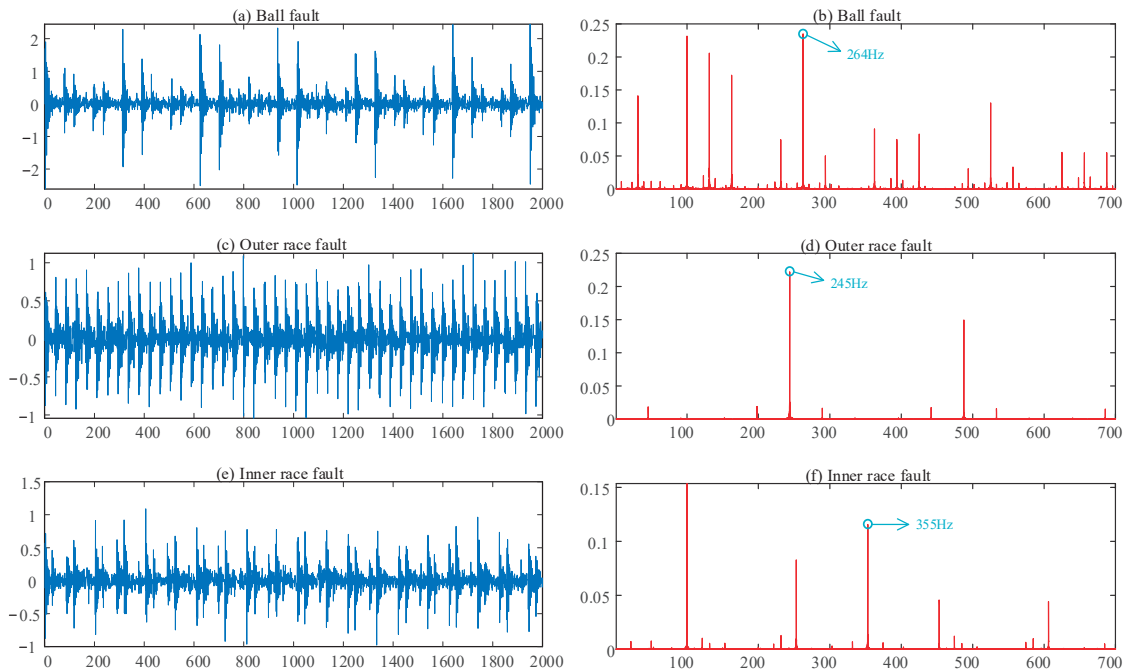


Figure 4. Time domains and spectra of simulated bearing faults: (a) time domain of ball fault; (b) envelope spectrum of ball fault; (c) time domain of outer race fault; (d) envelope spectrum of outer race fault; (e) time domain of inner race fault; (f) envelope spectrum of inner race fault.

3.2. Analysis of Simulation Results

In the practical working environment, the operation of the equipment is influenced by noise. Therefore, the simulated bearing fault signal was added to Gaussian white noise to simulate the actual working conditions. SNR ranged from 10 dB to 40 dB, in 1 dB steps.

In this simulation, MTPE, TPE, MPE, and the proposed CMTPE were used to extract fault features from simulation signals. For the selection of the main parameters when using the above method, there were the following considerations: if the embedding dimension m is small, the dynamic process of reconstruction will contain non detailed dynamic information, while if the value of m is too large, the number of vectors will decrease, which will lead to the loss of information. In addition, a large value of the scale factor τ will lead to information redundancy, and a small value of τ will lead to the loss of fault information. Therefore, the recommended value for parameter τ is 10–20 [29]. The values for the parameters of the entropy methods used in this study were set as $m = 3$ and $\tau = 20$.

The fault diagnosis strategies of each of the four methods combined with ELM were used to identify three simulated bearing faults. For each fault type, the original signal was sliced into 100 samples without overlap, and the data length of each sample was 2048. Therefore, the data set had a total of 300 samples. Among them, 50 samples of each fault type were randomly selected as the ELM training set, and the rest of the samples were used to test the trained ELM. The ELM was run 20 times and the average test accuracy was taken as the final test accuracy. Higher test accuracy means better performance of fault diagnosis strategy, and smaller test variance means better stability. The robustness of the method against noise can be obtained by comparing the test accuracies for each different SNR value.

The test results are shown in Figure 5. It is obvious that no matter the SNR value, the test accuracy of CMTPE was always higher than that of the other methods, and its error bar was also smaller than that of other methods. This shows that CMTPE had the best bearing fault diagnosis performance and the highest test stability. When the signal-to-noise ratio was high, all strategies except TPE had high accuracy. For example, when the SNR range was 30 to 40, the test accuracy of CMTPE, MTPE, and MPE was higher than 95%. However, as the SNR gradually decreased, the test accuracy of MTPE and MPE decreased at a faster speed and a larger range, while CMTPE still had 100% test accuracy, even when the SNR value decreased to 20 dB. Moreover, when the SNR was reduced to 10 dB, the test accuracy of MTPE and MPE was only 61.90% and 71.73%, respectively, while the test accuracy of CMTPE was still 89.60%. This shows that the proposed CMTPE method has better robustness to noise.

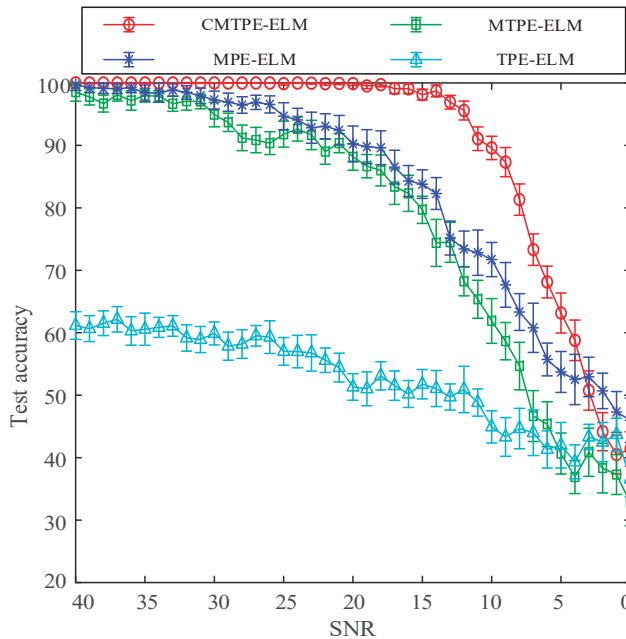


Figure 5. Test results with different SNR value.

When the SNR was less than 5, the classification accuracy of all methods was less than 60%, and the accuracy near 0 was less than 50%. When the SNR was negative, the accuracy of the classifier was no longer referential, because at this time, the simulation signal had been submerged by noise, and none of the four methods could correctly identify faults. However, when the noise was relatively weak, CMTPE still had the highest classification accuracy and the smallest error bar, compared with other methods, and had better stability. In the case of negative SNR, we used filtering and other noise reduction methods to preprocess the signal to achieve better anti-noise effects.

4. Experimental Evaluation

In this section, we report the testing of the effectiveness of the CMTPE using bearing fault data. CMTPE was compared with TPE, MTPE, and MPE to verify the superiority of the CMTPE-based fault diagnosis strategy.

4.1. Bearing Test Rig and Experimental Data Illustration

The experimental data were collected on an HD-FD-H-03X rotor rolling bearing fault test rig. The appearance and structure of the platform are shown in Figure 6. In the experiment, the speed of the motor was 1000 rpm and no load was applied. In order to verify the effectiveness of the proposed method, five different health conditions were designed, as shown in Figure 7. One of them was designated as normal, and the other four fault types were inner race crack 4mm (IRC), outer race crack 4 mm (ORC), inner race pitting 3 mm (IRP), and outer race pitting 3 mm (ORP).

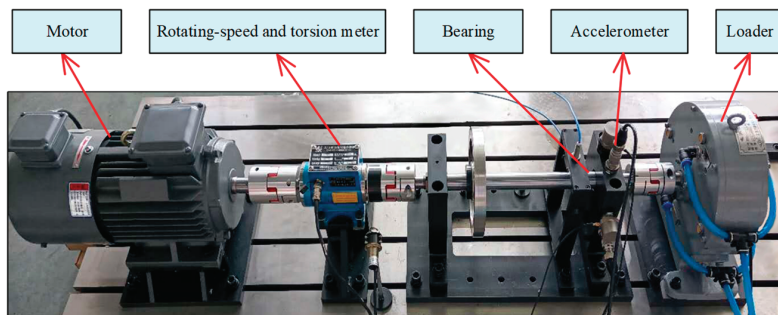


Figure 6. The bearing fault test rig used in the experiment.

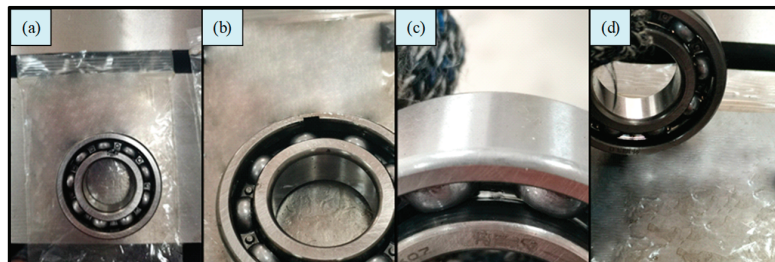


Figure 7. Four types of bearing fault designed in the experiment: (a) IRC; (b) ORC; (c) IRP; (d) ORP.

The vibration signals of the different health conditions were collected through the acceleration sensor, in which the sampling frequency was 10,240 Hz. Figure 8 shows the time domain and envelope spectrum of each of the four fault states. Figure 8a,c,e,g shows the time domain diagrams of three faults; Figure 8b,d,f,h displays the corresponding envelope spectrum diagrams. The vibration signal of each state was divided into 75 samples for feature extraction, and the length of each sample was 2048. Then, 25 samples of each state were randomly selected to train the ELM classifier, and the remaining samples were used for testing [33]. Therefore, the total number of training and test sets was 125 and 250, respectively.

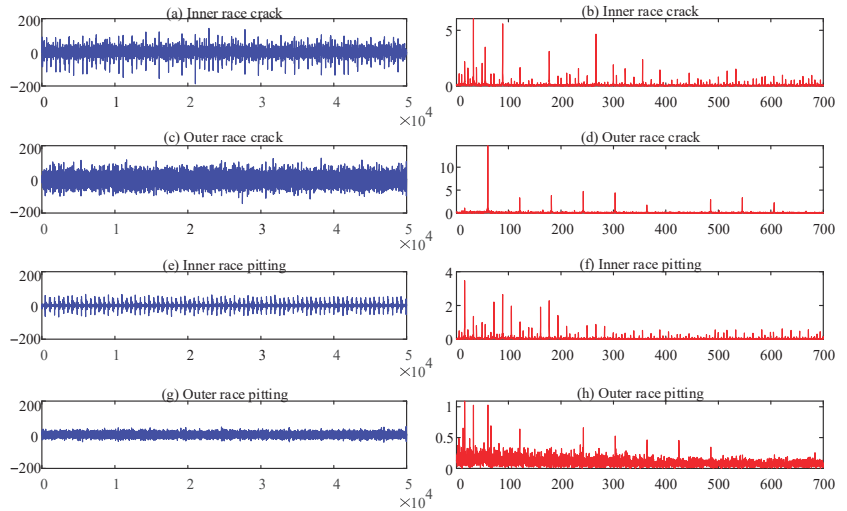


Figure 8. Time domains and spectra of the four fault states: (a) time domain of IRC; (b) envelope spectrum of IRC; (c) time domain of ORC; (d) envelope spectrum of ORC; (e) time domain of IRP; (f) envelope spectrum of IRP; (g) time domain of ORP; (h) envelope spectrum of ORP.

4.2. Comparison Analysis

According to the proposed fault diagnosis strategy, CMTPE was used to extract the fault features of bearing vibration signals. In addition, in order to prove that CMTPE has better feature extraction and fault diagnosis ability, MPE, TPE, and MTPE were used for comparison. The main parameters of these methods were as follows: the scale factor of the multi-scale methods was $\tau = 20$ and the embedding dimension of all methods was $m = 3$. The features extracted by the above four methods were used to train and test the ELM. In order to reduce the error caused by randomness, each method was run 20 times, and then the average test accuracy was taken as the final classification result. The test accuracy of the four strategies over 20 runs is shown in Figure 9. Table 2 shows the classification accuracy and variance. The criteria were: higher accuracy represents better feature extraction ability, and lower variance represents better stability.

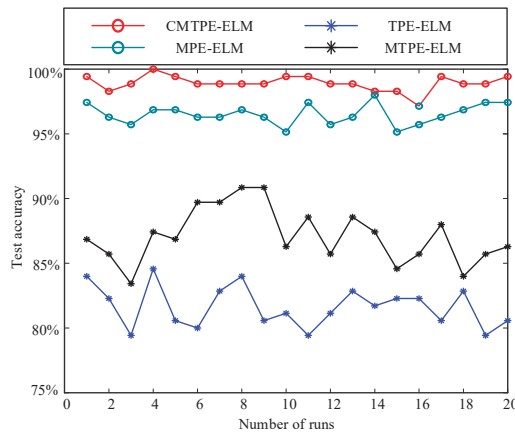


Figure 9. Test accuracy of the four strategies.

Table 2. Average test accuracy and variance of the four methods.

| Methods | CMTPE | TPE | MTPE | MPE |
|---------------------------|-------|-------|-------|-------|
| Average test accuracy (%) | 98.60 | 82.16 | 86.37 | 96.51 |
| Variance (%) | 0.65 | 1.39 | 1.92 | 0.94 |

From Table 2 and Figure 9, it can be observed that the fault identification accuracy of TPE was lower than that of the other multiscale methods, and that the identification accuracy of CMTPE reached 98.60%, which was the highest among the multiscale methods. This shows that the multiscale analysis could extract more abundant fault information when processing vibration signals. Moreover, the coarse-grained process also affects the quality of the fault features. The higher identification accuracy proves that the coarse-grained process of the proposed CMTPE method can better grasp the key information related to bearing faults, and it had the best feature extraction effect. Furthermore, the variance of CMTPE was only 0.65%, which was lower than the other methods. This verifies that the proposed fault diagnosis strategy based on CMTPE not only had excellent fault feature extraction performance, but also had the best stability.

As shown in Figure 10, the confusion matrix of the four methods can intuitively visualize the classification performance of each method combined with ELM [34,35]. As shown in Figure 10b,c, TPE and MTPE exhibited many misclassifications; in particular, it was difficult to distinguish between IRC and IRP, while, as shown in Figure 10a,d, CMTPE and MPE have better classification performance. CMTPE had only four misclassifications, and the classification accuracy reached 98.4%, which was the highest among all methods.

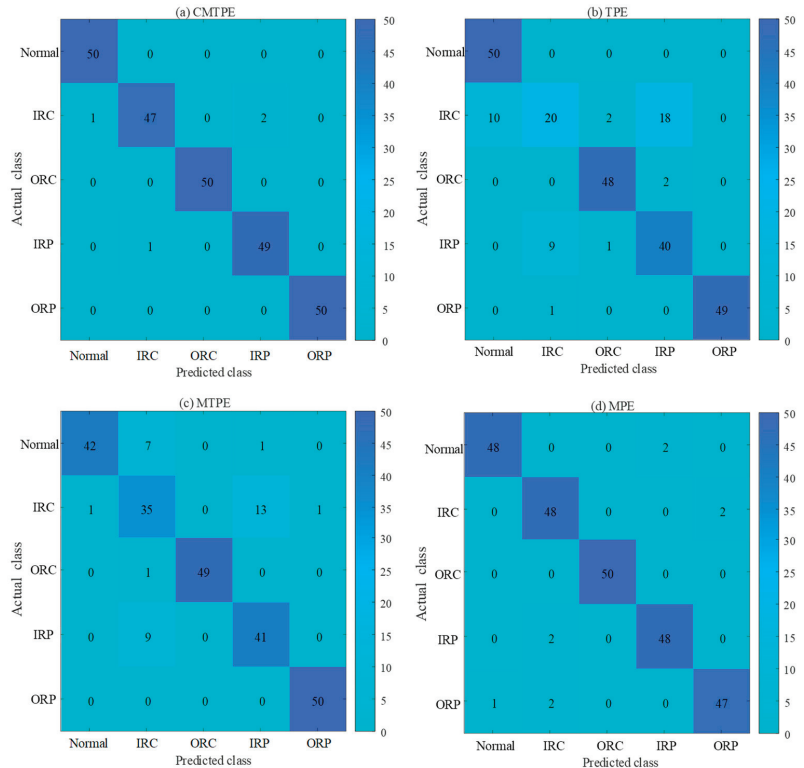


Figure 10. Confusion matrix of four methods.

In this work, 25 training samples and 50 test samples were selected. In order to eliminate the contingency brought by the specific number of training samples, a performance test of CMTPE with different numbers of training samples was carried out. The classification accuracy was tested with 15, 25, 35, 45, 55 and 65 training samples, and each case was run 20 times to reduce randomness. The results are shown in Figure 11. Obviously, with an increase in the number of training samples, the accuracy of various methods will increase, but CMTPE always had the highest test accuracy compared to the other methods.

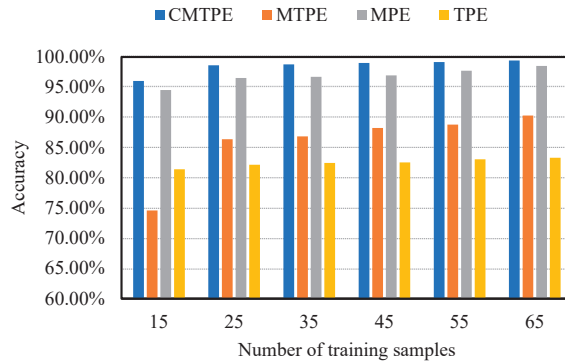


Figure 11. Effect of the number of training samples on the performance of CMTPE, MTPE, MPE and TPE.

In order to further intuitively compare the feature extraction capabilities of the three multiscale methods, we carried out visual processing on the extracted fault features. In this work, the scale factor of multiscale method was $\tau = 20$. Therefore, the t-SNE visualization method was used to reduce the dimension of fault features to two dimensions [36]. The results of this feature visualization are shown in Figure 12. The criterion of feature extraction effect is: the closer the distance between clusters of the same type of features, the farther the distance between clusters of different types of features, which proves that the feature extraction effect of this method is better.

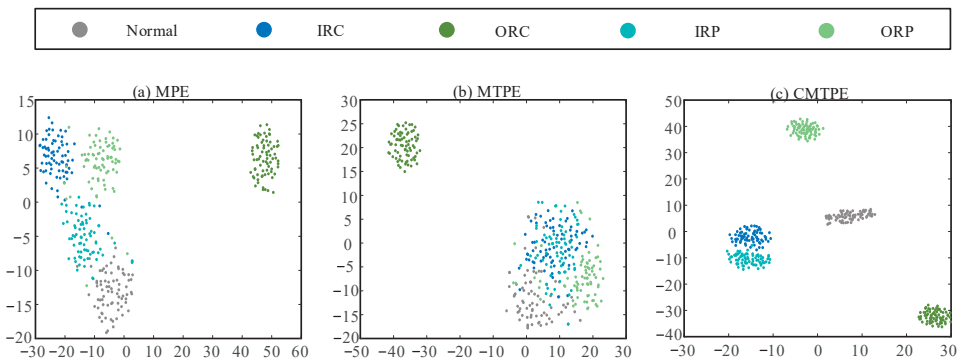


Figure 12. Visualization of features extracted by three multiscale entropy methods: (a) MPE method; (b) MTPE method; (c) CMTPE method.

As can be seen from Figure 12b, the features of the states, other than ORC, were obviously mixed, which indicates that MTPE has poor feature extraction performance. Figure 12a demonstrates that the features extracted by MPE could better distinguish between most fault characteristics. However, the distance between clusters of different

state features was small, some feature points were mixed, and the distance within clusters was large, so the effect of feature extraction were poor. In contrast, it can be seen from Figure 12c that CMTPE had the largest inter-cluster distance and the smallest intra-cluster distance, and that the feature extraction performance was the best.

In order to test the minor fault identification ability of the proposed method, the following two cases were designed with different extents of faults. Case 1 included a normal control and eight different degrees of inner and outer race crack faults: normal, inner race crack 0.2 mm, outer race crack 0.2 mm, inner race crack 1 mm, outer race crack 1 mm, inner race crack 2.7 mm, outer race crack 2.7 mm, inner race crack 4 mm, and outer race crack 4 mm. Case 2 included a normal control and six different degrees of pitting faults: normal, inner race pitting 1 mm, outer race pitting 1 mm, inner race pitting 2 mm, outer race pitting 2 mm, inner race pitting 3 mm, and outer race pitting 3 mm.

The vibration signals of each state were again divided into 75 samples, with 25 used for training the ELM, and the rest used for testing. The classification accuracy and variance under the two cases for the four methods are shown in Tables 3 and 4. It was found that CMTPE still had an excellent classification effect for the more minor faults, especially for pitting faults of different degrees; the test accuracy reached 99.67%. CMTPE also still had the highest stability. However, the other three methods had decreased test accuracy for minor faults, in particular, the classification effect of TPE was very poor. This also shows that the composite multiscale method can avoid the information loss caused by the single scale method, and also overcome the problem of low accuracy of entropy estimation caused by the traditional multiscale method. Therefore, CMTPE can effectively identify minor faults.

Table 3. Average test accuracy and variance of the four methods for different degrees of crack faults.

| Methods | CMTPE | TPE | MTPE | MPE |
|---------------------------|-------|-------|-------|-------|
| Average test accuracy (%) | 96.46 | 44.71 | 70.46 | 94.31 |
| Variance (%) | 0.74 | 1.95 | 1.85 | 1.01 |

Table 4. Average test accuracy and variance of the four methods for different degrees of pitting faults.

| Methods | CMTPE | TPE | MTPE | MPE |
|---------------------------|-------|-------|-------|-------|
| Average test accuracy (%) | 99.67 | 60.49 | 86.39 | 94.97 |
| Variance (%) | 0.34 | 1.51 | 1.90 | 1.06 |

5. Conclusions

In this study, a method using CMTPE for quantifying the complexity of time series was proposed. CMTPE takes into consideration the transition probability matrix of an ordinal pattern and performs composite multiscale processing on the original time series. This avoids the loss of information caused by single-scale analysis, and overcomes the problem where the traditional multiscale method will greatly shorten the time series in large scale, resulting in low accuracy of entropy evaluation. Composite multiscale analysis improved the performance of CMTPE feature extraction, the accuracy of entropy estimation, and the robustness against noise. Compared with MTPE, TPE, and MPE, the superiority of CMTPE was verified by both simulation and experimental data. The results show that CMTPE has better robustness, can effectively identify bearing faults, and has the highest diagnostic accuracy and stability.

Moreover, in the case of negative SNR, it was necessary to use filtering and other noise reduction methods to preprocess the signal to achieve better anti-noise effect. Thus, in future work, we will test the effectiveness of combining CMTPE with other noise reduction methods.

Author Contributions: Data curation, B.M.; Formal analysis, L.G.; Funding acquisition, Y.L.; Investigation, T.Z.; Methodology, J.G. and B.M.; Resources, T.Z.; Supervision, Y.L.; Validation, J.G.; Visualization, L.G.; Writing—original draft, J.G.; Writing—review & editing, B.M., T.Z., L.G. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research was supported by the National Natural Science Foundation of China under Grants 12172290 and 52250410345.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

References

- Gao, D.; Zhu, Y.; Wang, X.; Yan, K.; Hong, J. A Fault Diagnosis Method of Rolling Bearing Based on Complex Morlet CWT and CNN. In Proceedings of the 2018 Prognostics and System Health Management Conference (PHM-Chongqing), Chongqing, China, 26–28 October 2018; pp. 1101–1105. [\[CrossRef\]](#)
- Mohanty, S.; Gupta, K.K.; Raju, K.S. Comparative study between VMD and EMD in bearing fault diagnosis. In Proceedings of the 2014 9th International Conference on Industrial and Information Systems (ICIIS), Gwalior, India, 15–17 December 2014; pp. 1–6. [\[CrossRef\]](#)
- Sohaib, M.; Kim, C.H.; Kim, J.M. A hybrid feature model and deep-learning-based bearing fault diagnosis. *Sensors* **2017**, *17*, 2876. [\[CrossRef\]](#)
- Chegini, S.N.; Bagheri, A.; Najafi, F. Application of a new EWT-based denoising technique in bearing fault diagnosis. *Measurement* **2019**, *144*, 275–297. [\[CrossRef\]](#)
- Kankar, P.K.; Sharma, S.C.; Harsha, S.P. Fault diagnosis of ball bearings using machine learning methods. *Expert Syst. Appl.* **2011**, *38*, 1876–1886. [\[CrossRef\]](#)
- Youcef Khodja, A.; Guersi, N.; Saadi, M.N.; Boutasseta, N. Rolling element bearing fault diagnosis for rotating machinery using vibration spectrum imaging and convolutional neural networks. *Int. J. Adv. Manuf. Technol.* **2020**, *106*, 1737–1751. [\[CrossRef\]](#)
- Ali, J.B.; Fnaiech, N.; Saidi, L.; Chebel-Morello, B.; Fnaiech, F. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl. Acoust.* **2015**, *89*, 16–27.
- Zhang, W.; Li, C.; Peng, G.; Chen, Y.; Zhang, Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Signal Process.* **2018**, *100*, 439–453. [\[CrossRef\]](#)
- Li, Y.; Wang, X.; Liu, Z.; Liang, X.; Si, S. The Entropy Algorithm and Its Variants in the Fault Diagnosis of Rotating Machinery: A Review. *IEEE Access* **2018**, *6*, 66723–66741. [\[CrossRef\]](#)
- Pinkus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [\[CrossRef\]](#) [\[PubMed\]](#)
- Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wu, S.D.; Wu, C.W.; Lin, S.G.; Lee, K.Y.; Peng, C.K. Analysis of complex time series using refined composite multiscale entropy. *Phys. Lett. A* **2014**, *378*, 1369–1374. [\[CrossRef\]](#)
- Chen, W.; Wang, Z.; Xie, H.; Yu, W. Characterization of Surface EMG Signal Based on Fuzzy Entropy. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2007**, *15*, 266–272. [\[CrossRef\]](#) [\[PubMed\]](#)
- Deng, W.; Zhang, S.; Zhao, H.; Yang, X. A Novel Fault Diagnosis Method Based on Integrating Empirical Wavelet Transform and Fuzzy Entropy for Motor Bearing. *IEEE Access* **2018**, *6*, 35042–35056. [\[CrossRef\]](#)
- Wang, X.; Si, S.; Li, Y. Multiscale Diversity Entropy: A Novel Dynamical Measure for Fault Diagnosis of Rotating Machinery. *IEEE Trans. Ind. Inform.* **2021**, *17*, 5419–5429. [\[CrossRef\]](#)
- Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [\[CrossRef\]](#) [\[PubMed\]](#)
- Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* **2002**, *89*, 068102. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhang, L.; Xiong, G.; Liu, H.; Zou, H.; Guo, W. Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference. *Expert Syst. Appl.* **2010**, *37*, 6077–6085. [\[CrossRef\]](#)
- Li, Y.; Feng, K.; Liang, X.; Zou, M.J. A fault diagnosis method for planetary gearboxes under non-stationary working conditions using improved Vold-Kalman filter and multi-scale sample entropy. *J. Sound Vib.* **2019**, *439*, 271–286. [\[CrossRef\]](#)
- Hsieh, N.K.; Lin, W.Y.; Young, H.T. High-speed spindle fault diagnosis with the empirical mode decomposition and multiscale entropy method. *Entropy* **2015**, *17*, 2170–2183. [\[CrossRef\]](#)
- Li, Y.; Xu, M.; Wang, R.; Huang, W. A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy. *J. Sound Vib.* **2016**, *360*, 277–299. [\[CrossRef\]](#)
- Li, Y.; Wei, Y.; Feng, K.; Wang, X.; Liu, Z. Fault diagnosis of rolling bearing under speed fluctuation condition based on Vold-Kalman filter and RCMFE. *IEEE Access* **2018**, *6*, 37349–37360. [\[CrossRef\]](#)
- Zheng, J.; Cheng, J.; Yang, Y.; Luo, S. A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination. *Mech. Mach. Theory* **2014**, *78*, 187–200. [\[CrossRef\]](#)

24. Li, Y.; Xu, M.; Wei, Y.; Huang, W. A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree. *Measurement* **2016**, *77*, 80–94. [[CrossRef](#)]
25. Wang, X.; Si, S.; Wei, Y.; Li, Y. The optimized multi-scale permutation entropy and its application in compound fault diagnosis of rotating machinery. *Entropy* **2019**, *21*, 170. [[CrossRef](#)]
26. Wu, S.-D.; Wu, P.-H.; Wu, C.-W.; Ding, J.-J.; Wang, C.-C. Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Entropy* **2012**, *14*, 1343–1356. [[CrossRef](#)]
27. Li, Y.; Yang, Y.; Li, G.; Xu, M. A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mRMR feature selection. *Mech. Syst. Signal Process.* **2017**, *91*, 295–312. [[CrossRef](#)]
28. Li, Y.; Wang, S.; Yang, Y.; Deng, Z. Multiscale symbolic fuzzy entropy: An entropy denoising method for weak feature extraction of rotating machinery. *Mech. Syst. Signal Process.* **2022**, *162*, 108052. [[CrossRef](#)]
29. Li, Y.; Wang, S.; Li, N.; Deng, Z. Multiscale symbolic diversity entropy: A novel measurement approach for time-series analysis and its application in fault diagnosis of planetary gearboxes. *IEEE Trans. Ind. Inform.* **2021**, *18*, 1121–1131. [[CrossRef](#)]
30. Zhang, B.; Shang, P. Transition Permutation Entropy and Transition Dissimilarity Measure: Efficient Tools for Fault Detection of Railway Vehicle Systems. *IEEE Trans. Ind. Inform.* **2021**, *18*, 1654–1662. [[CrossRef](#)]
31. Patange, A.D.; Jegadeeshwaran, R.; Bajaj, N.S.; Apoorva, K. Application of Machine Learning for Tool Condition Monitoring in Turning. *Sound Vib.* **2022**, *56*, 127–145. [[CrossRef](#)]
32. Li, Y. Investigation of Fault Feature Extraction and Early Fault Diagnosis for Rolling Bearings. Ph.D. Thesis, Harbin Institute of Technology, Harbin, China, 2017.
33. Patange, A.D.; Jegadeeshwaran, R. Application of bayesian family classifiers for cutting tool inserts health monitoring on CNC milling. *Int. J. Progn. Health Manag.* **2020**, *11*. [[CrossRef](#)]
34. Patange, A.D.; Jegadeeshwaran, R. A machine learning approach for vibration-based multipoint tool insert health prediction on vertical machining centre (VMC). *Measurement* **2021**, *173*, 108649. [[CrossRef](#)]
35. Bajaj, N.S.; Patange, A.D.; Jegadeeshwaran, R.; Kulkarni, K.A.; Ghatpande, R.S.; Kapadnis, A.M. A Bayesian optimized discriminant analysis model for condition monitoring of face milling cutter using vibration datasets. *J. Nondestruct. Eval. Diagn. Progn. Eng. Syst.* **2022**, *5*, 021002. [[CrossRef](#)]
36. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Sensors Editorial Office
E-mail: sensors@mdpi.com
www.mdpi.com/journal/sensors



MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34

www.mdpi.com



ISBN 978-3-0365-6463-0