Research Article

Multidomain Feature Fusion Network for Fault Diagnosis of Rolling Machinery

Dewei Yang [1,2] **Kefa Zhou,**[1,2] **Feng Qi,**[1,3] **and Kai Dong**[1,4]

[1]*Nanjing Hydraulic Research Institute, Nanjing 210029, China*
[2]*Country Dam Safety Management Center of the Ministry of Water Resources, Nanjing 210029, China*
[3]*Najing R&D Hydro-Information Technology Co., Ltd, Nanjing 210029, China*
[4]*Sichuan University, Chengdu 610065, China*

Correspondence should be addressed to Dewei Yang; dwyang@nhri.cn

Mechanical vibration constitutes a valuable cue for performing fault diagnosis as it is directly related to the transient regime of rolling machinery. This study establishes a multidomain feature fusion network (MFFN) to extract and fuse multidomain features through a novel multistream architecture. Three primary features are simultaneously extracted from the time, frequency, and time-frequency domains. Then, highly representative features are extracted via three convolutional branches in one- or two-dimensional spaces. A novel squeeze-connection-excitation (SCE) module is proposed to adaptively fuse features in the three domains. The advantage offered by the proposed method is that it can leverage cues from the raw vibration signal, resulting in accurate fault diagnosis. Experimental results comprehensively demonstrate and analyze the high accuracy and generalization achieved by this MFFN-based fault diagnosis method.

1. Introduction

Rolling machinery is a foundational element in industrial infrastructures. Machinery faults are the main factors that significantly affect equipment and production safety. Intelligent fault diagnosis of rolling machinery has been a topic of interest in studies concerning vibration-based health monitoring of mechanical systems [1]. Previously, fault diagnosis was realized through a combination of traditional signal processing methods, such as Fourier and wavelet transforms (WTs), and shallow learning techniques, such as support vector machine (SVM) [2] and Bayes classifiers [3]. In general, these methods are physically analyzable; however, they provide an inadequate representation of faults, which may result in a low diagnosis accuracy. This problem has motivated the development of deep learning-based methods, such as deep belief networks (DBNs) [4], stacked autoencoders (SAEs) [5], convolutional neural networks (CNNs) [6], and long short-term memory (LSTM) [7]. The high representability offered by deep learning methods significantly improves fault diagnosis accuracy.

Recently, multistream architectures are being used for fault diagnosis. In contrast to single-stream architectures, multistream architectures can represent faults in terms of multiple aspects; thus, they can achieve further enhancements in the representability of intrinsic characteristics of machinery faults. This property may further improve the performance of fault diagnosis methods. However, current multistream architectures primarily focus on the multiscale characteristic of raw vibration signals [8] and ignore the various physical properties observed in multiple domains. A novel multistream architecture that can extract and fuse multidomain features is desirable to facilitate accurate fault diagnosis.

This study proposes a novel multidomain feature fusion network (MFFN) for fault diagnosis. To this end, three one-dimensional (1D) and two-dimensional (2D) convolutional streams are designed and combined to construct the multistream architecture. Two 1D streams manage the data in the time and frequency domains, while a 2D stream extracts the time-frequency feature. At the backend joint, three representative features are fused by the squeeze-connection-

excitation (SCE) module. Finally, the fused features are used for fault classification. The contributions of this study to fault diagnosis include the following:

(i) A novel multistream architecture that can process multidomain features in an organized and comprehensive manner

(ii) A novel SCE module that can adaptively fuse multidomain features

(iii) A novel feature type that offers high representability of fault patterns and improves the decision-making capabilities of fault diagnosis

The remainder of this paper is organized as follows. Section 2 reviews related works. The overall framework of MFFN is presented in Section 3. Section 4 describes the proposed MFFN-based fault recognition method. Experimental comparisons and analysis are discussed in Section 5. Section 6 provides the summary and conclusions.

2. Related Works

Various fault diagnosis methods have been proposed to classify faults in mechanical systems. These methods generally collect vibration signals as the source data because vibrations directly relate to the transient state of running elements. Various existing shallow learning models, such as the hidden Markov model [9], k-nearest neighbors [10], and SVMs [11], have been applied in fault classification. Recently, deep learning-based methods have demonstrated excellent performance in fault diagnosis. The advantage afforded by deep learning methods is the high representability of faults. For example, the DBN model [4], which is a typical probabilistic generative model, has been introduced to solve the problems of nonlinear dynamics and discrete failure patterns. However, experimental results have revealed that DBN architectures are susceptible to overfitting. The SAE method, which is a popular deep learning-based fault diagnosis method, can incrementally learn new samples without a retraining process [5].

Another key issue in fault diagnosis is that feature extraction. Previously, temporal and frequency analyses were the two main approaches toward fault feature extraction [12]. However, they cannot represent the temporal variation of a vibration signal accurately [13]. This problem has been solved via methods including short-time Fourier transforms [13], Wigner Ville distributions [14], and WTs [15]. Among these, WT is the most practical because its relaxed structure can decompose signals with varying temporal resolutions. Moreover, WT can produce 2D feature maps such that successful image classification methods can be transformed into fault diagnosis methods.

Deep learning-based fault diagnosis has attracted considerable attention recently [16]. The advantage of deep learning lies in its excellent ability to abstract signals by performing layer-wise nonlinear calculations, thereby enabling the deeper layer to generate more representative features. This encourages the utilization of various deep learning methods in fault recognition. The DBN is one of the most widely used deep learning methods because it can adapt to a wide range of problems, including those of nonlinear dynamics and discrete failure patterns [17]. To leverage valuable cues for fault diagnosis, an adaptive spatiotemporal feature learning architecture with multiple measurements was proposed [18]. Subsequently, the generalization of deep learning architecture was considered. For example, a domain generalization-based hybrid diagnosis network was established, which could be deployed in unseen working conditions instead of in real-world working conditions [19]. Moreover, a domain adversarial transfer network has been evaluated for application in fault diagnosis, wherein a transfer learning mechanism can be implemented to enhance the generalization of deep learning-based fault diagnosis [20]. Recently, a novel convolutional neural network is established to diagnose faults from small samples. Based on the domain adaption, this method won success when the vibration data are not available in abundant [21]. Different from this previous strategy, our study in this paper aims to solve another problem in fault diagnosis—feature representation and fusion.

Also, other types of signal have been introduced in faults diagnosis. For example, the thermographic information has been utilized in fault diagnosis of ventilation in BLDC motors [22, 23]. In contrast to the vibration signal, the thermographic signals provide additional informative clues which help to increase the accuracy of the fault diagnosis. Moreover, the thermographic signal is relatively simple in contrast to the vibration signal, such that it can better identify the fault types. However, the main drawback of the thermographic signal-based strategy lies in that it is commonly hysteretic to reflect the machinery statement. In practice, it is observed that the temperature significantly increases after a while of the fault occurrence. Alternatively, the acoustic signal has been investigated in the field of fault diagnosis [24]. The main advantage of the acoustic signal-based strategy lies in the noncontact measurement that we can efficiently deploy the acoustic sensors to diagnose faults. However, the acoustic signal is likely affected by the environmental noises. As the result, noise removal is the main issue of the acoustic signal processing.

This study aims to leverage valuable cues from multiple domains for fault diagnosis. To this end, a novel MFFN is proposed. This network comprises three streams that can comprehensively extract highly representative features in multiple domains, such as the temporal, frequency, and time-frequency domains. The MFFN can obtain more valuable cues for fault diagnosis than those of current single-stream and multistream architectures. Moreover, the novelty of the proposed architecture lies in its ability to adaptively fuse 1D and 2D features using the SCE block.

3. Proposed Fault Diagnosis Scheme

3.1. MFFN. To achieve high fault representability, this study proposes a novel multistream architecture for extracting and classifying three types of features in the temporal, time-frequency, and frequency domains. The block diagram of MFFN is shown in Figure 1. The sliding window block is
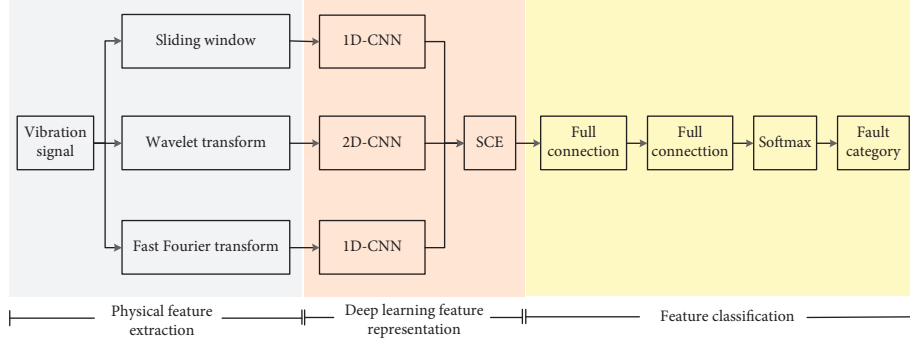
FIGURE 1: Block diagram of multidomain feature fusion network.

applied to segment the vibration signal into sequence vectors at the first stream. Time-frequency features are extracted using WT at the second stream. The third stream extracts frequency features via fast Fourier transform (FFT). These primary features are subsequently enhanced in terms of their representability through layer-wise convolutional calculations. Finally, these highly representative features are adaptively fused via the SCE module. The backend classifier is established using two fully connected layers and a Softmax calculation block. The data used in this study were obtained from public datasets. The fault diagnosis platform consists of a motor, torque transducer/encoder, dynamometer, and control electronics [21]. The reason for selecting these datasets lies in that they provide a baseline to fairly evaluate and compare different methods.

3.2. Primary Feature Extraction.

The primary feature extraction process is shown in Figure 2. The sliding window is used for extracting temporal features. L denotes the window length, and M is the sliding step. The frequency spectrum is extracted via the following FFT:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi/Nkn}, \quad k = 0, 1, \ldots, N - 1, \quad (1)$$

where N denotes the length of the signal segmentation.

A limitation of FFT is that it analyses the frequency spectrum pattern of the vibration signal exclusively from a global perspective; therefore, it is not suitable for an amplitude-modulated or nonstationary signal. This drawback can be addressed by wavelet package transform (WPT), which is a time-frequency analysis method that can analyze vibration signals with flexible temporal resolutions at both high and low frequencies [15]. Therefore, the WT is operated with a wavelet packet tree that decomposes a signal into several levels of wavelet packets. A three-layer wavelet packet tree is used in our method. As a result, eight sub-bands are obtained, and the energy value of each sub-band signal can be calculated through the following equation:

$$E_n^j = \int \left| C_n^j(t) \right|^2 \mathrm{d}t = \sum_{k=1}^{N_i} \left| x_j^k \right|^2, \quad (2)$$

where $C_n^j(t)\,(n = 3, \quad j = 0, 1, \ldots, 7)$ is the reconstructed sub-band signal, N_i is the length of the reconstructed signal,

and $x_j^k\,(k = 1, 2, \ldots, r)$ is the amplitude of the jth reconstructed signal. The energy spectrum feature of a sub-band signal can be presented through a normalized value, as shown in the following:

$$\left[\frac{E_4^0}{\sqrt{E_r}}, \frac{E_4^1}{\sqrt{E_r}}, \ldots, \frac{E_4^7}{\sqrt{E_r}} \right], \quad (3)$$

where E_r is the square root of the summed square values of the sub-band signal energy and is expressed as follows:

$$E_r = \sqrt{\sum_{j=1}^{M} \left(E_n^j \right)^2}. \quad (4)$$

3.3. Extraction and Fusion of Highly Representative Features.

In the first and third streams, two 1D-CNNs are connected to the primary feature extractor to manage features in the temporal and frequency domains, while a 2D-CNN is connected to the WT module to process the feature in the time-frequency domain. These highly representative features are then fused at the backend joint by the SCE module. In general, there are four successive phases in the SCE module, namely, squeeze, connection, excitation, and reweight, as shown in Figure 3. The feature matrices of the three domains are the input to the SCE module. The temporal, frequency, and time-frequency feature matrices are presented as follows:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1C_1} \\ \vdots & \ddots & \vdots \\ x_{L_1 1} & \cdots & x_{L_1 C_1} \end{bmatrix}_{L_1 \times C_1},$$

$$Y = \begin{bmatrix} y_{11} & \cdots & y_{1C_2} \\ \vdots & \ddots & \vdots \\ y_{L_2 1} & \cdots & y_{L_2 C_2} \end{bmatrix}_{L_2 \times C_2}, \quad (5)$$

$$Z = \begin{bmatrix} z_{11} & \cdots & z_{1C_3} \\ \vdots & \ddots & \vdots \\ z_{L_3 1} & \cdots & z_{L_3 C_3} \end{bmatrix}_{L_3 \times C_3},$$

where L_1, L_2, and L_3 are the feature dimensions and C_1, C_2, and C_3 are the number of feature channels.
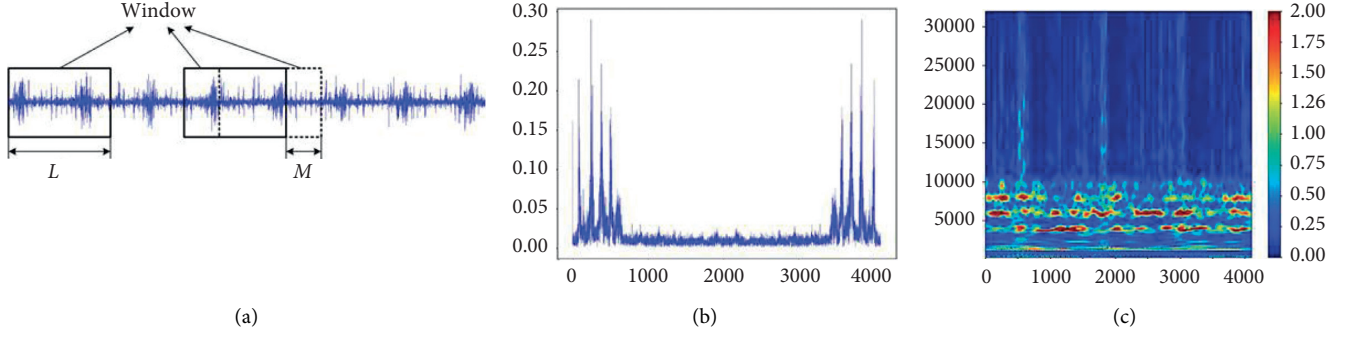
Figure 2: Primary feature extraction: (a) temporal feature, (b) frequency feature, and (c) time-frequency feature.
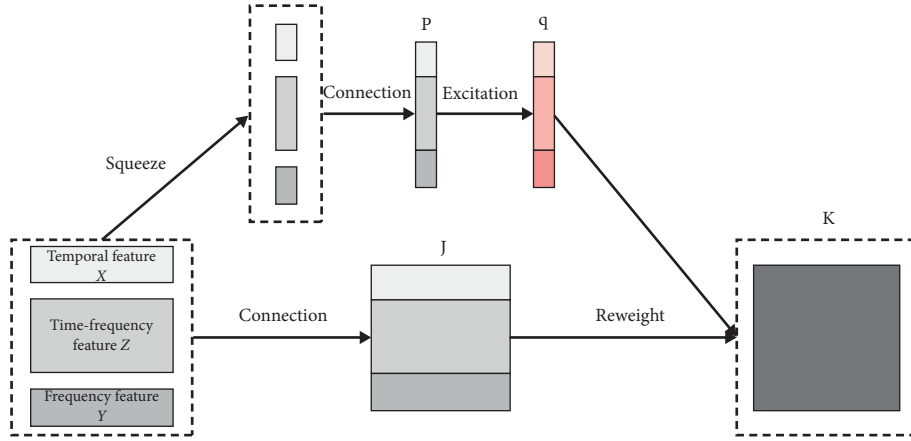

Figure 3: SCE module.

3.4. Squeeze.

A pooling operation is applied to squeeze the feature matrix. As a result, three 1D feature vectors are generated to present the feature matrix across three domains, shown as follows:

$$l_j = F_{sq}(X) = \frac{1}{L_1}\sum_{i=1}^{L_1} x_{ij}, \quad j = (1,2,\dots,C_1), \mathbf{l} = (l_1, l_2, \dots, l_{C_1}),$$

$$m_j = F_{sq}(Y) = \frac{1}{L_2}\sum_{i=1}^{L_2} y_{ij}, \quad j = (1,2,\dots,C_2), \mathbf{m} = (m_1, m_2, \dots, m_{C_2}),$$

$$n_j = F_{sq}(Z) = \frac{1}{L_3}\sum_{i=1}^{L_3} z_{ij}, \quad j = (1,2,\dots,C_3) \mathbf{n} = (n_1, n_2, \dots, n_{C_3}).$$

(6)

3.5. Connection.

The feature vectors are fused by the following concatenation operation:

$$\begin{aligned}\mathbf{p} &= F_c(\mathbf{l}, \mathbf{m}, \mathbf{n}) \\ &= (l_1, l_2, \dots, l_{C_1}, \quad m_1, m_2, \dots, m_{C_2}, n_1, n_2, \dots, n_{C_3}) \\ &= (p_1, p_2, \dots, p_C),\end{aligned}$$

(7)

$$\mathbf{J} = F_c(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \begin{bmatrix} j_{11} & \cdots & j_{1C} \\ \vdots & \ddots & \vdots \\ j_{L1} & \cdots & j_{LC} \end{bmatrix}_{L \times C}.$$

3.6. Excitation.

Multilayer mapping is performed to achieve excitation, as follows:

$$\mathbf{q} = F_{\text{ex}}(\mathbf{p}, \mathbf{W}) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{p})).$$

(8)

In the above equation, σ is the sigmoid function, δ is the ReLU activation function, and $\mathbf{W}, \mathbf{W}_1$, and $\mathbf{W}_2$ are the full-connection weights.

3.7. Reweight.

The learned weight is added to feature channels to generate the weighted feature for final classification:

$$\mathbf{K} = F_r(\mathbf{J}, \mathbf{q}) = \mathbf{J} \times \mathbf{q}.$$

(9)

The advantage of the *reweight* calculation is similar to that of the global average pooling operation in the *squeeze* process that can generate channel-wise statistics. Subsequently, this global information is embedded by the *excitation* process to generate the channel descriptor $\mathbf{q}$, which comprehensively captures channel-wise dependencies. As a result, the most important feature can be emphasized by multiplying the feature channels with the channel descriptor. In this regard, SCE blocks intrinsically introduced dynamics conditioned on the input, thereby helping boost feature discriminability of specific fault patterns [5].

4. MFFN-Based Fault Diagnosis

The MFFN-based fault diagnosis method comprises four modules. The first module performs primary data processing, wherein the physical significance of the vibration signal is presented with respect to the temporal, frequency, and time-frequency features. In the second module, the high-representation features are extracted through layer-wise mapping. Adaptive feature fusion is realized in the third module, wherein the most credible factor is enhanced, while feature redundancy is reduced considerably. Finally, the fourth module is designed for fault classification, wherein a shallow architecture is established with two fully connected layers and a Softmax calculation block. The details of our proposed MFFN are presented in Table 1.

5. Experimental Evaluation and Discussion

5.1. Setup. To evaluate the performance of the proposed MFFN, experiments were conducted on defective bearing datasets provided by the Case Western Reserve University Bearing Data Center (CWRU dataset) [25], Jiangnan University (Jiangnan dataset) [26], and Paderborn University (Paderborn dataset) [27]. The bearing system platform in the Case Western Reserve University Bearing Data Center includes a 2 HP motor, torque transducer, dynamometer, and load motor. The vibration signal was collected via an accelerometer at a sampling frequency of 12 kHz. In addition to the normal state, nine categories of fault state data were included in this database: single-point faults with sizes of 0.007, 0.14, and 0.021 were individually identified on the inner race (IR), outer race (OR), and rolling elements (REs), respectively. For each state, 120,000 samples were collected in 10 s. The data from Jiangnan University include four categories of running states: the normal state and fault states separately seeded on the bearing at IR, OR, and RE. All data were collected at a 50 kHz sampling frequency at rolling speeds of 600, 800, and 1000 rpm. For the normal state, 1800 samples were randomly collected, while 600 samples were collected for each fault state. The data from the Paderborn University were provided via measurements concerning six healthy and 26 damaged bearings at IR and OR. All data were collected at a 64 kHz sampling frequency at rolling speeds of 900 and 1500 rpm. For each state, 256,0000 samples are collected in 4 s. The training and testing samples for experimental evaluations are shown in Tables 2 to 4.

5.2. Model Pretraining and Fine-Tuning. An adequate number of epochs in the training period is important for model training. Excessive epochs may result in overfitting, while the learning outcome may be poor in the case of insufficient epochs. Figure 4 illustrates the training times of the three datasets and reveals that that the MFFN can converge rapidly on all three datasets. Thirty iterations are sufficient for model learning with the CWRU and Jiangnan datasets, while 25 iterations are required for model learning with the Paderborn dataset.

5.3. MFFN-Based Fault Diagnosis. Confusion matrices were utilized for evaluating the performance of the proposed MFFN. Figure 5(a) presents the confusion matrix for the CWRU dataset. This result demonstrates that our fault diagnosis method is highly accurate. Only two samples were erroneously classified; the rest were identified correctly. The classification results regarding the Jiangnan dataset are satisfactory (Figure 5(b)); only one sample was misclassified. A similar outcome was observed in the results on the Paderborn dataset (Figure 5(c)) with one error.

5.4. Comparison against Existing Deep Learning Methods. We evaluated the proposed MFFN by comparing it to state-of-the-art methods. The temporal-, frequency-, and time-frequency feature-based methods are comprehensively catalogued for experimental comparison. For example, the 1D-CNN was used to classify 1D temporal features [28] and denoted as "TF + 1D-CNN." Furthermore, temporal feature + WDCNN (TF + WDCNN) [29], frequency feature + 1D-CNN (FF + 1D-CNN) [30], frequency feature + SDAE (FF + SDAE), time-frequency feature + 2D-CNN (TFF + 2D-CNN), and time-frequency feature + VGG16 (TFF + VGG16) [31] were included in the experimental comparison. Each dataset was divided into 10 subsets for experimental evaluation with respect to working conditions. Figure 6 reveals that no salient performance variation was observed among the 10 evaluations for MFFN; the maximum differences among evaluations were 0.08% on the CWRU dataset, 0.12% on the Jiangnan dataset, and 0.13% on the Paderborn dataset. Thus, experimental analysis demonstrates the stability of our MFFN compared with that of the other methods, which exhibit lesser model stability owing to significant performance variations across tested subsets.

The average accuracies of the compared fault diagnosis methods are listed in Table 5. Two observations can be made based on this table. First, the feature in the time-frequency domain outperforms the temporal- and frequency-domain features. This is because the time-frequency domain feature can identify details of the frequency spectrum of the vibration signal, which facilitates an improved fault diagnosis. Second, fusing features in multiple domains is preferable for fault diagnosis. This is because the feature fusion results can represent machinery faults from multiple aspects and allow more valuable cues to be leveraged for fault diagnosis. As a result, intraclass fault differences are enlarged, while interclass clustering is enhanced, which theoretically explains the better performance of the proposed MFFN.

5.5. Visualization. Aiming to comprehensively understand the benefits of our proposed MFFN, the t-SNE technique was applied to reduce the dimensionality of the learned features

TABLE 1: Parameters of MFFN.

Module	Name	Size/step/number	Parameter size	Output size
1D-CNN	Input_1	—	0	(None,4096,1)
	ResBlock_1	3/1/16	912	(None,4096,16)
	ResBlock_2	3/1/16	1568	(None,4096,16)
	ResBlock_3	3/1/4	441	(None,4096,4)
	Max_Pooling	2/1/-	0	(None,2048,4)
2D-CNN	Input_2	—	0	(None,128,128,3)
	Conv2D_1	30/5/256	691200	(None,20,20,256)
	Conv2D_2	6/2/256	2359552	(None,8,8,256)
	Inception_1	(1,3,5,7)/1/32	688768	(None,8,8,128)
	Reshape_1	—	0	(None,2048,4)
SCE	Global_Average_Pooling_1	—	0	(None,4)
	Global_Average_Pooling_2	—	0	(None,4)
	Global_Average_Pooling_3	—	0	(None,4)
	Concatenate_1	—	0	(None,2048,12)
	Concatenate_2	—	0	(None,12)
	Dense_1	6/-/-	78	(None,6)
	Dense_2	12/-/-	84	(None,12)
	Multiply	—	0	(None,2048,12)
Classifier	Flatten_1	—	0	(None,24576)
	Dense_3	100/-/-	2457700	(None,100)
	Dense_4	10,4,3/-/-	1010,404,303	(None,10),(None,4),(None,3)

TABLE 2: Samples in the CWRU dataset.

Category	Normal	Outer race			Inner race			Rolling element		
Size	—	0.007	0.014	0.021	0.007	0.014	0.021	0.007	0.014	0.021
Label	0	1	2	3	4	5	6	7	8	9
Training	1200	1200	1200	1200	1200	1200	1200	1200	1200	1200
Validation	400	400	400	400	400	400	400	400	400	400
Testing	400	400	400	400	400	400	400	400	400	400

TABLE 3: Samples in the Jiangnan dataset.

Category	Normal	Outer race	Inner race	Rolling element
Label	0	1	2	3
Training	1200	1200	1200	1200
Validation	400	400	400	400
Testing	400	400	400	400

TABLE 4: Samples in the Paderborn dataset.

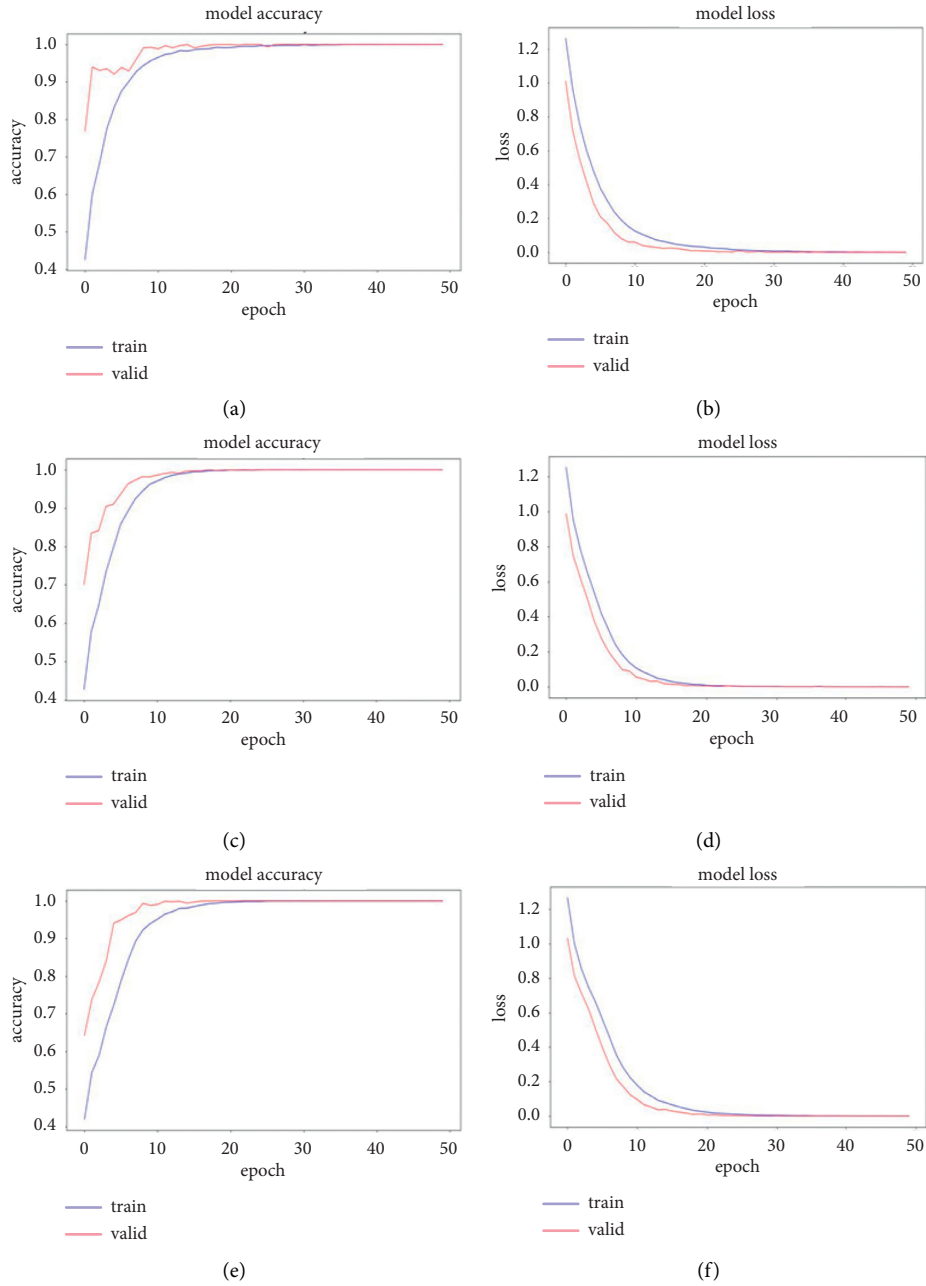| Category	Normal	Outer race	Inner race
Label	0	1	2
Training	1200	1200	1200
Validation	400	400	400
Testing	400	400	400

Figure 4: Training-time investigation for CWRU, Jiangnan, and Paderborn datasets. CWRU: (a) accuracy and (b) loss; Jiangnan: (c) accuracy and (d) loss; Paderborn: (e) accuracy and (f) loss.

to two for facilitating map generation. The resulting 2D feature maps are shown in Figure 6, wherein different colors represent various fault or normal categories. As shown in Figure 7, after MFFN feature learning, a fault-category clustering effect is observed in contrast to the raw

distribution, along with linear margins between fault categories. This result is desirable and enables simpler classification. This further demonstrates that using the MFFN architecture can significantly improve the accuracy of fault diagnosis.
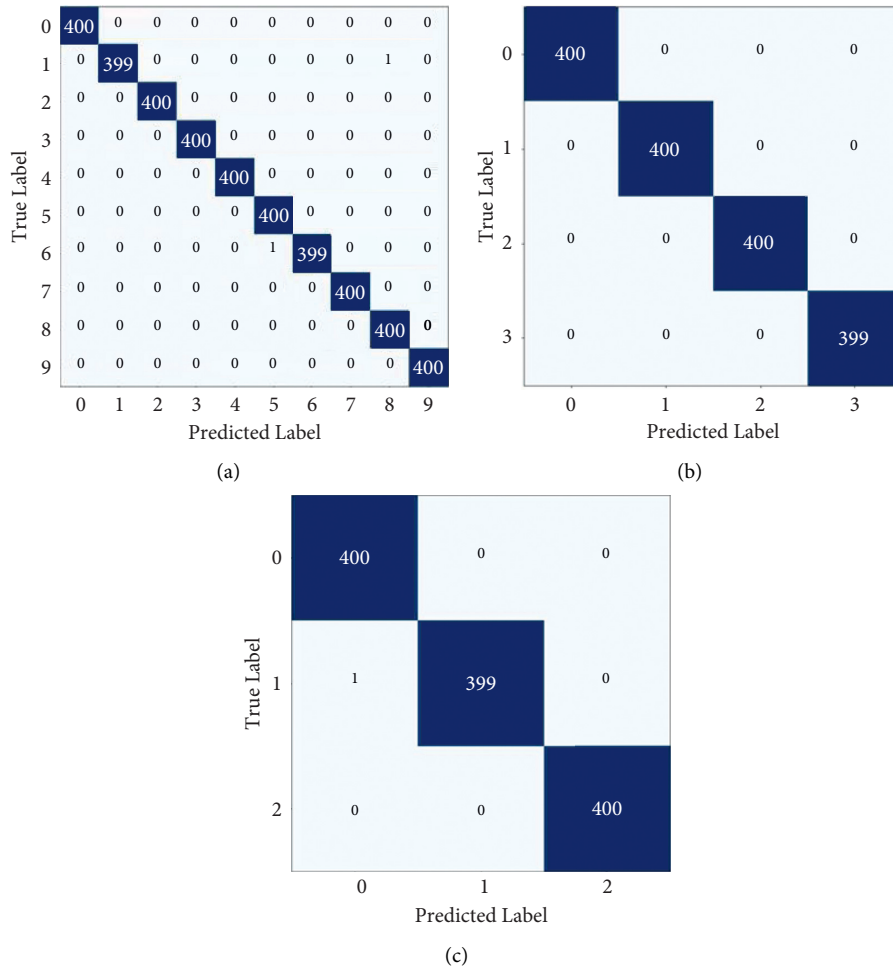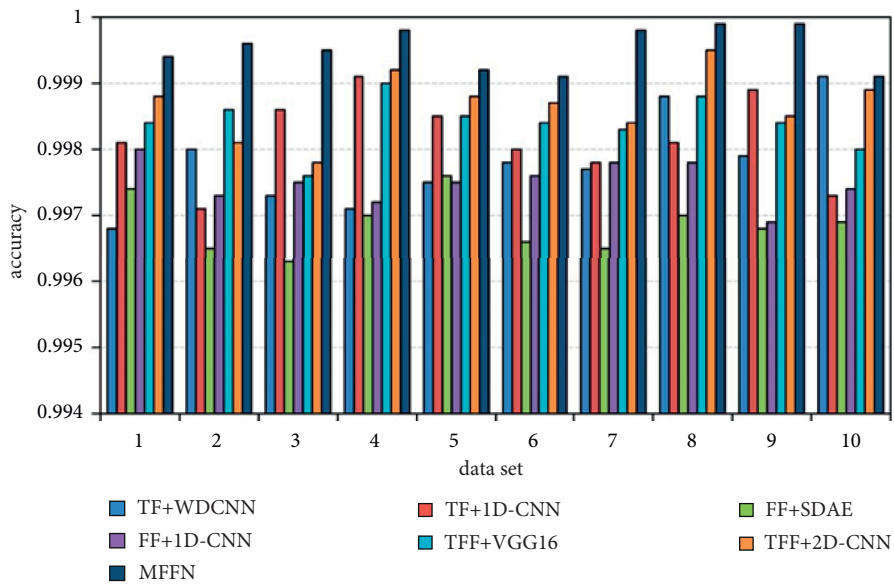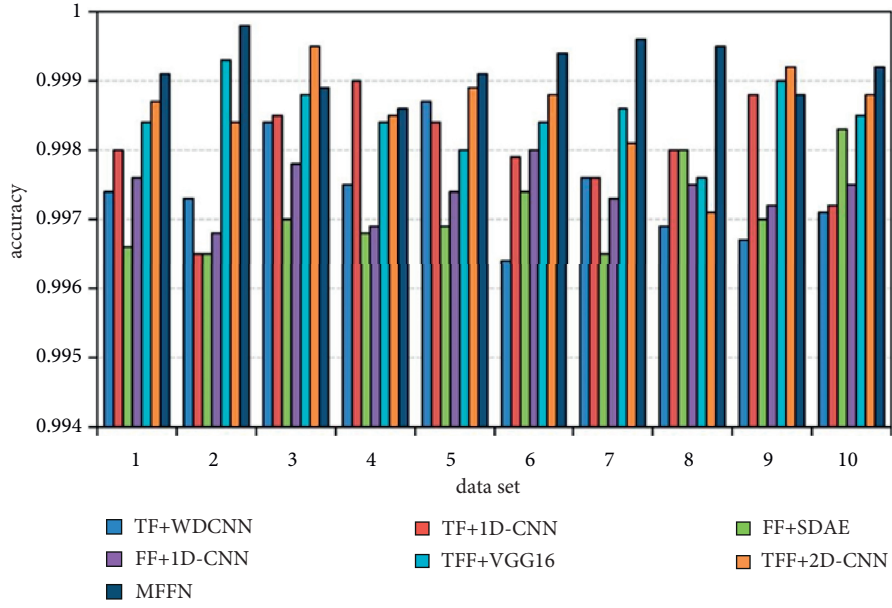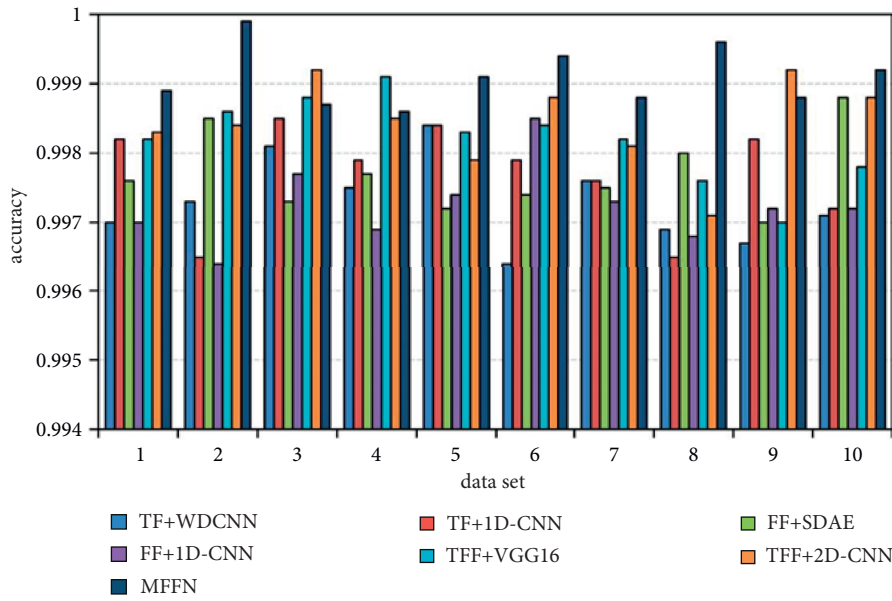
(a)

(b)

(c)

Figure 5: Confusion matrices for (a) CWRU, (b) Jiangnan, and (c) Paderborn datasets.

(a)

Figure 6: Continued.

(b)

(c)

Figure 6: Experimental comparisons on (a) CWRU, (b) Jiangnan, and (c) Paderborn datasets.

Table 5: The average performance in different domains.

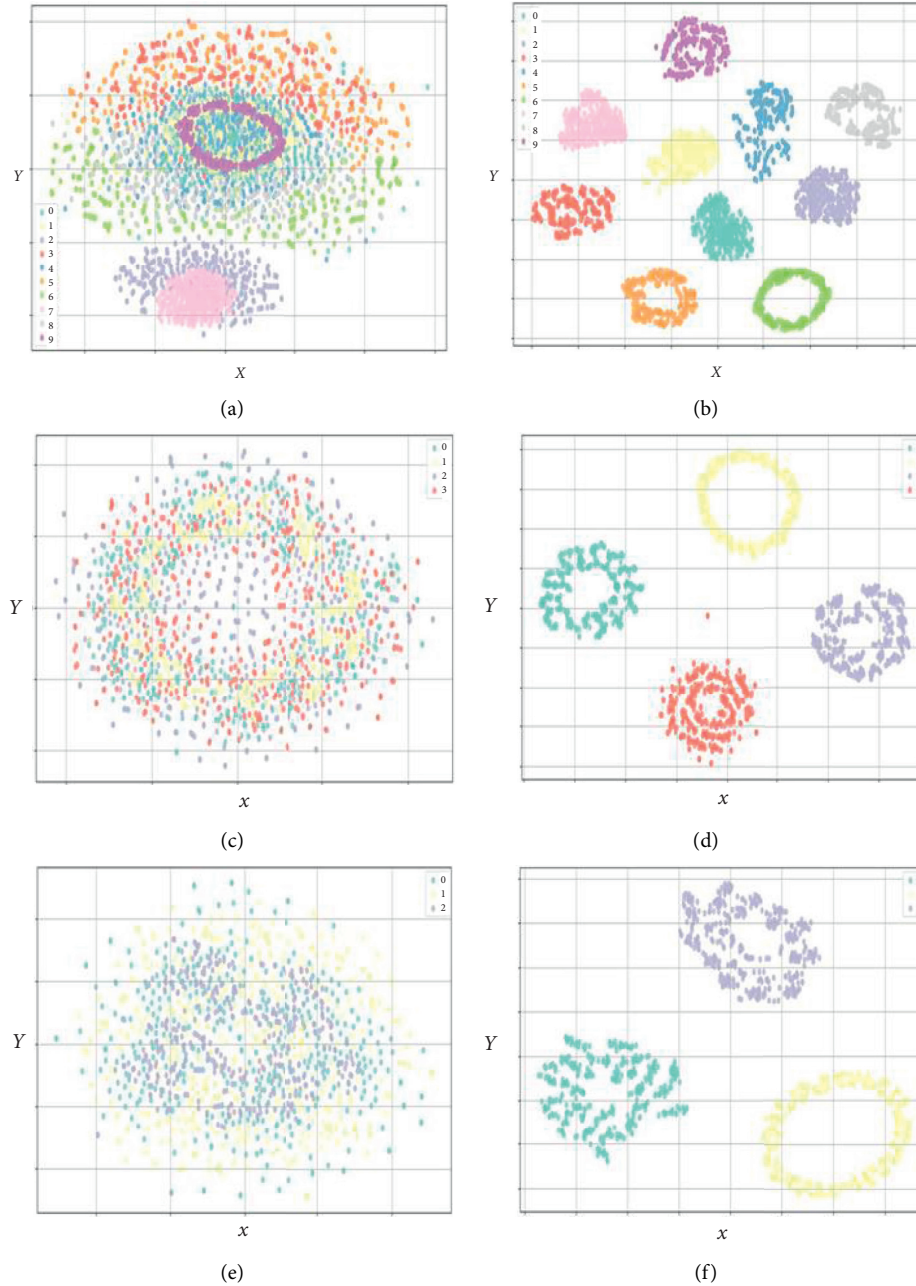| Input	Model	Average accuracy		
		CWRU (%)	Jiangnan (%)	Paderborn (%)
TF	WDCNN	99.78	99.74	99.73
	1D-CNN	99.81	99.80	99.77
FF	SDAE	99.69	99.71	99.78
	1D-CNN	99.75	99.74	99.72
TFF	VGG16	99.84	99.85	99.82
	2D-CNN	99.87	99.86	99.84
TF + FF + TFF	MFFN	99.95	99.92	99.91

(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 7: Feature visualization with (a) raw data on the CWRU dataset, (b) learned features on the CWRU dataset, (c) raw data on the Jiangnan dataset, (d) learned features on the Jiangnan dataset, (e) raw data on the Paderborn dataset, and (f) learned features on the Paderborn dataset.

6. Conclusions

A novel MFFN-based fault diagnosis method is proposed. The proposed MFFN can fuse features in different domains, such as the temporal, frequency, and time-frequency domains. Sufficient cues are comprehensively leveraged through the deep learning process of MFFN. The main contribution of MFFN is that it can improve the representability of faults, leading to a significant improvement in the accuracy of fault diagnosis. Theoretically, features in multiple domains depict faults from multiple perspectives,

which are complementary in physical significance. Moreover, the importance of features in multiple domains varies with respect to the tasks on hand. Intrinsically, our proposed MFFN adopts a feature fusion strategy using adaptive weights. Features extracted in multiple domains are weighted and fused, leading to a comprehensive utilization of their advantages. Consequently, MFFN achieves higher accuracy compared with existing architectures.

Using our proposed MFFN, exceptional accuracy can be achieved, enabling its utilization in many practical

applications. In our future work, we will train the MFFN model to handle more signal types, such as thermal imaging and acoustic data, which contain much more valuable features for diagnosing faults. Moreover, we will evaluate MFFN in real-world applications, especially in online fault diagnosis.

Data Availability

The experiment data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics-A comprehensive review," *IEEE Access*, vol. 8, pp. 29857–29881, 2020.

[2] Q. Shi and H. Zhang, "Fault diagnosis of an autonomous vehicle with an improved SVM algorithm subject to unbalanced datasets," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 7, pp. 6248–6256, 2020.

[3] Y. Hu, F. Cui, and X. Tu, "Bayesian estimation of instantaneous speed for rotating machinery fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 9, pp. 8842–8852, 2021.

[4] X. Zhao, M. Jia, and Z. Liu, "Semisupervised graph convolution deep belief network for fault diagnosis of electormechanical system with limited labeled data," *IEEE Transactions on Industrial Informatics*, vol. 17, pp. 5450–5460, 2020.

[5] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, 2016.

[6] S. Tang, S. Yuan, and Y. Zhu, "Convolutional neural network in intelligent fault diagnosis toward rotatory machinery," *IEEE Access*, vol. 8, pp. 86510–86519, 2020.

[7] J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on Long Short-term memory networks," *Renewable Energy*, vol. 133, pp. 422–432, 2019.

[8] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 4, pp. 3196–3207, 2019.

[9] H. Ocak and K. A. Loparo, "HMM-based fault detection and diagnosis scheme for rolling element bearings," *Journal of Vibration and Acoustics*, vol. 127, no. 4, pp. 299–306, 2005.

[10] Q. Wang, Y. B. Liu, X. He, S. Y. Liu, and J. Liu, "Fault diagnosis of bearing based on KPCA and KNN method," *Advanced Materials Research*, vol. 1491, pp. 986–987, 2014.

[11] X. Zhang, Y. Liang, and J. Zhou, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, 2015.

[12] A. Krishnakumari, M. Saravanan, M. Ramakrishnan, S. M. Ponnuri, and R. Srinadh, "Vibration condition monitoring of spur gear using feature extraction of EMD and hilbert-huang transform," *Intelligent Manufacturing and Energy Sustainability*, vol. 169, pp. 133–145, 2020.

[13] L. Zhang, Y. Li, L. Dong et al., "Gearbox fault diagnosis using multiscale sparse frequency-frequency distributions," *IEEE Access*, vol. 9, pp. 113089–113099, 2021.

[14] J. A. Rosero, L. Romeral, J. A. Ortega, and E. Rosero, "Short-circuit detection by means of empirical mode decomposition and wigner-ville distribution for PMSM running under dynamic condition," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 11, pp. 4534–4547, 2009.

[15] J. Li, H. Wang, X. Wang, and Y. Zhang, "Rolling bearing fault diagnosis based on improved adaptive parameterless empirical wavelet transform and sparse denoising," *Measurement*, vol. 152, Article ID 107392, 2020.

[16] H. Li, G. Hu, J. Li, and M. Zhou, "Intelligent fault diagnosis for large-scale rotating machines using binarized deep neural networks and random forests," *IEEE Transactions on Automation Science and Engineering*, vol. 19, pp. 1–11, 2021.

[17] B. Cai, L. Huang, and M. Xie, "Bayesian networks in fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2227–2240, 2017.

[18] T. Han, C. Liu, L. Wu, S. Sarkar, and D. Jiang, "An adaptive spatiotemporal feature learning approach for fault diagnosis in complex systems," *Mechanical Systems and Signal Processing*, vol. 117, pp. 170–187, 2019.

[19] T. Han, Y.-F. Li, and M. Qian, "A hybrid generalization network for intelligent fault diagnosis of rotating machinery under unseen working conditions," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[20] Z. Chen, G. He, J. Li, Y. Liao, K. Gryllias, and W. Li, "Domain adversarial transfer network for cross-domain fault diagnosis of rotary machinery," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 11, pp. 8702–8712, 2020.

[21] A. Kumar, G. Vashishtha, C. P. Gandhi, Y. Zhou, A. Glowacz, and J. Xiang, "Novel convolutional neural network (NCNN) for the diagnosis of bearing defects in rotary machinery," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.

[22] A. Glowacz, "Thermographic fault diagnosis of ventilation in BLDC motors," *Sensors*, vol. 21, no. 21, Article ID 7245, 2021.

[23] A. Glowacz, "Ventilation diagnosis of angle grinder using thermal imaging," *Sensors*, vol. 21, no. 8, Article ID 2853, 2021.

[24] A. Glowacz, R. Tadeusiewicz, S. Legutko et al., "Fault diagnosis of angle grinders and electric impact drills using acoustic signals," *Applied Acoustics*, vol. 179, Article ID 108070, 2021.

[25] Case western Reserve university, "Case western Reserve university bearing data centre website[DB/OL]," 2017, https://csegroups.case.edu/bearingdatacenter/pages/download-data-file.

[26] K. Li, "Jiangnan university bearing dataset [DB/OL]," 2018, https://mad-net.org:8765/explore.html.

[27] Paderborn University, "Paderborn University Website," 1972, https://mb.uni-paderborn.de/en/kat/main-research/datacenter/bearing-datacenter/data-sets-and-download.

[28] W. Zhang, F. Zhang, W. Chen, Y. Jiang, and D. Song, "Fault state recognition of rolling bearing based fully convolutional

network," *Computing in Science & Engineering*, vol. 21, pp. 132–147, 2018.

[29] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, pp. 425–446, 2017.

[30] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Processing*, vol. 130, pp. 377–388, 2017.

[31] M. T. Pham, J.-M. Kim, and C. H. Kim, "Accurate bearing fault diagnosis under variable shaft speed using convolutional neural networks and vibration spectrogram," *Applied Sciences*, vol. 10, no. 18, pp. 6385–6397, 2020.