Correlation-Enhanced Multi-Scale Residual Network for Bearing Fault Diagnosis in Noisy and Cross-Working Conditions

Panfeng Bao[1,3], Yue Zhu[1], Yufeng Shen[1], Jiashun Ou[1], Xuening Hu[2*]

[1] *Changsha Aeronautical Vocational and Technical College, Changsha 410000, China*
cshy0855@163.com
zhuyue19931277@163.com
15116220284@163.com
oujiashun1990@163.com

[2] *CRRC Changzhou Auto Parts Co., Ltd., Changzhou 213011, China*
[*]*Corresponding author: huxuening508@163.com*

[3] *Nanjing University of Science and Technology, National Key Laboratory of Transient Physics, Nanjing 210000, China*

ABSTRACT

Bearing fault diagnosis under noisy and cross-working conditions remains a challenging task due to complex signal variations and interference. To address this challenge, this paper proposes a Correlation-Enhanced Multi-Scale Residual Network (CE-MSRN), which effectively captures multi-scale fault features while enhancing correlation across different bearing faults. Our model integrates a residual learning framework with a multi-scale feature fusion mechanism, improving robustness against noise and generalization across diverse working conditions. Experimental evaluations on benchmark datasets demonstrate that CE-MSRN achieves superior diagnostic accuracy compared to mainstream methods, exhibiting strong adaptability to unseen fault patterns. These results confirm the potential of our approach for real-time and reliable bearing fault diagnosis in aero-engines and transmission systems.

1. INTRODUCTION

The bearing in the transmission systems of aero-engines and aerospace equipment, as core components of aircraft systems, directly affect flight safety (H. Wang, Liu, Peng, & Zuo, 2023). Under extreme operating conditions such as high speed, high temperature, and heavy loading (Zhao & Chen, 2022), bearings in the transmission systems are subjected to cyclic loads and severe vibrations, leading to early failures (H. Wang & Li, 2023). Such malfunctions not only cause abnormal engine vibration and a drastic drop in transmission efficiency but may also trigger secondary disasters such as blade fractures, resulting in catastrophic consequences (Hou, Wang, Lv, Xiong, & Peng, 2022). Therefore, real-time monitoring of bearing vibration signals combined with intelligent fault diagnosis models to precisely extract and classify weak fault features has become a critical technical requirement for ensuring safe operation of aerospace equipment (D. Peng, Wang, Desmet, & Gryllias, 2023).

Time-frequency domain analysis methods, such as wavelet transform (WT), Hilbert-Huang transform (HHT), and short-time Fourier transform (STFT), along with their enhanced algorithms(Z. Wang, Lu, Wang, Liu, & Fan, 2013; Rai & Mohanty, 2007; Kabla & Mokrani, 2016), have demonstrated the feasibility of installing accelerometers on equipment to identify bearing fault patterns through vibration signals. However, these algorithms rely on manual interpretation of frequency-domain variations for analysis. Machine learning-based approaches including support vector machine (SVM), extreme learning machine (ELM), and k-nearest neighbor (kNN) classifiers, along with their improved variants(Fernández-Francos, Martínez-Rego, Fontenla-Romero, & Alonso-Betanzos, 2013; J. Wang et al., 2021; Luo, Li, Zhang, Li, & An, 2016), enable automated extraction of critical diagnostic features. Yet under the severe noise levels and variable working conditions of aeroengine environments, these methods struggle to achieve practically viable performance outcomes.

Addressing the challenge of bearing fault diagnosis under intense noise, convolutional neural networks (CNNs) have been introduced to automatically learn latent fault patterns in time-domain vibration signals. Zhang et al. (W. Zhang, Peng, Li, Chen, & Zhang, 2017) proposed a deep CNN with a

https://doi.org/10.36001/IJPHM.2025.v16i2.4302

wide first-layer kernel (WDCNN) for bearing fault diagnosis. They found that enlarging convolutional kernels in the initial layer enhances the network's receptive field, thus achieving higher accuracy in noisy environments. Chen et al. (Chen, Zhang, & Gao, 2021) incorporated multi-sized convolutional kernels into CNNs to exploit multiscale convolutions for capturing fault patterns across different scales, thereby improving diagnostic precision under noise interference. Peng et al. (H. Peng, Wang, Gao, Wang, & Du, 2025) employed CNNs with residual connections for bearing fault diagnosis and utilized deeper network architectures to enhance feature learning capabilities in the presence of noise. The introduction of residual connections effectively mitigated degradation issues potentially caused by overly deep networks.

In addition, frequent variations in rotational speed and load of aero-engines, coupled with sensors installed at different locations, pose significant challenges for bearing fault diagnosis under cross-working-condition scenarios. Lu et al. (Lu, Yu, Han, & Wang, 2020) employed transfer learning methods to pre-train models on large datasets. Subsequently, by fixing the encoder parameters and fine-tuning the decoder (i.e., the fault classifier) using small datasets, they effectively achieved high accuracy with limited samples. Berghout et al. (Berghout & Benbouzid, 2023) proposed a collaborative selection-based incremental transfer learning approach, enhancing cross-working-condition fault diagnosis of aerospace bearings by increasing source-domain sample quantities during training. Qian et al. (Qian, Qin, Luo, Wang, & Wu, 2023) introduced a novel domain discrepancy measure in their transfer learning framework to improve feature distribution between source and target domains, thereby strengthening the model's capability for cross-working-condition fault diagnosis.

To tackle the aforementioned challenges in bearing fault diagnosis within the aviation domain, this paper proposes a Correlation-Enhanced Multi-Scale Residual Network (CE-MSRN). This network employs convolutional kernels of varying sizes to extract multi-scale features, improving diagnostic accuracy under noisy conditions. Residual connections are introduced to deepen the network, enhancing modeling capabilities while preventing degradation. Finally, a correlation-enhanced classifier is designed to boost transfer learning effectiveness. The primary contributions of this work are summarized as follows:

- **An encoder that integrates multi-scale convolution with residual connections.** This combination helps mitigate the impact of noise on fault diagnosis accuracy and enables the exploitation of deeper network architectures for extracting fault-relevant features, thereby facilitating transfer learning across various working conditions.
- **A correlation-enhanced decoder for fault classification.** This decoder leverages latent associations among different failure categories, thereby reducing computational costs while improving classification accuracy and enhancing performance in transfer learning.
- **Extensive experimental validation.** Our approach has been evaluated on two distinctly different bearing fault diagnosis datasets. The experimental results demonstrate that the proposed method maintains high accuracy under intense noise conditions and exhibits robust cross-working-condition transfer learning performance across varying sensors, loads, rotational speeds, and other operational parameters.

The remainder of this paper is organized as follows. Section 2 introduces related works. Section 3 details the implementation method of our proposed CE-MSRN. In Section 4, experiments are conducted using two distinct public datasets to fully demonstrate the effectiveness of our approach. Finally, Section 5 summarizes the contributions and outlines future research directions.

2. RELATED WORKS

2.1. Multi-Scale Convolution and Noise Effects

For aviation bearing fault diagnosis, the time-domain multiscale characteristics of vibration signals present a core challenge. Due to the nonlinear relationship between fault feature frequencies and rotational speed (Lai et al., 2024), the duration of vibrational impacts caused by identical faults under different speeds may differ by several-fold (H. Wang et al., 2023), rendering traditional single-resolution methods incapable of simultaneously capturing transient impulses and periodic oscillations (W. Zhang, Chen, Xiao, & Yin, 2023). A multiscale convolutional architecture addresses this challenge by employing parallel convolutional kernels with varying receptive fields to establish a multiresolution analysis framework (Dong, Jiang, Yao, Mu, & Yang, 2024). Shao et al. (Shao & Kim, 2024) proposed an adaptive multiscale attention-based convolutional network that achieves fault diagnosis using minimal labeled data. Ni et al. (Ni et al., 2024) utilized multiscale convolutions to compute frequency-specific similarities between test signals and fault-free reference signals, enabling the evaluation of bearing health conditions.

In real-world scenarios, bearing signals are often interfered by multisource noise (H. Wang, Liu, Peng, & Cheng, 2022). Based on the central limit theorem, existing studies (Li, Wang, Yao, Li, & Gao, 2024; X. Zhang, Sheng, Ouyang, & Zheng, 2023) simulate their statistical characteristics using zero-mean additive Gaussian noise. Multiscale convolution kernels essentially form bandpass filters: small-scale kernels suppress low-frequency noise through short-time windowing, while large-scale kernels remove high-frequency noise via long-term averaging to improve diagnostic accuracy in noisy environments. Miao et al. (Miao, Yu, & Zhao, 2022) developed an adaptive multiresolution mechanism utilizing convolutional ker-

nels with diverse receptive fields and sparse representation for noise filtering. Peng et al. (H. Peng, Du, Gao, Wang, & Wang, 2024) combined multiscale convolutions with a dynamic noise injection module to enhance fault diagnosis accuracy under high-intensity noise conditions.

2.2. Cross-Working Condition Transfer Learning and Residual Connection

In aviation equipment fault diagnosis, cross-working-condition transfer learning is a core method to address data scarcity and working condition differences. Facing the lack of actual fault samples and significant discrepancies between laboratory and real-world working conditions (Tang et al., 2024), transfer learning identifies domain-shared features to enable rapid model adaptation with limited target-domain data (Chang et al., 2024). Zheng et al. (Zheng et al., 2024) proposed a diversity-regularized transfer learning strategy to extract common features across differing working conditions. Zhang et al. (J. Zhang, Zhang, An, Luo, & Yin, 2024) developed a representation-learning-based migration approach for fault classification under few-shot scenarios.

Existing transfer learning frameworks commonly adopt an encoder parameter freezing strategy, training the decoder solely using target-domain data (Huo, Jiang, Shen, Zhu, & Zhang, 2023). The encoder is designed to extract domain-invariant fault features (Lei et al., 2023), while residual modules model input-output residuals to separate common signal characteristics from working-condition-related disturbances (Yan, Yan, Xu, & Yuen, 2023). This architecture preserves feature consistency and enhances robustness against varying working condition signals. Wan et al. (Wan, Li, Chen, Gong, & Li, 2022) employed an improved residual network to extract transferable features from vibration signals. Yu et al. (Yu et al., 2023) combined time-frequency analysis with residual connections and leveraged a self-attention mechanism to achieve transfer learning under changing working conditions.

3. PROPOSED METHOED

The proposed CE-MSRN architecture, as shown in Fig. 1, first employs a wide multi-scale convolutional layer to extract primary features. This is followed by two multi-scale residual layers for further extraction of fault-related features. Finally, a correlation-enhanced decoder is utilized for fault classification. In this section, a detailed introduction will be provided.

3.1. Wide First Layer

As shown in Figure 2, let the input signal be x. When this signal is fed into the Wide First Layer, an average pooling layer is initially applied to perform moving-average denoising, which can be represented as

$$y_i = \frac{1}{k}\sum_{m=-p}^{p} x_{i+m}, \quad \forall i \in [1,n], \tag{1}$$

where k denotes the convolution kernel size. The number of zero padding elements p is set to $(k-1)/2$ to ensure that the output size matches that of the input.

Assuming the observed signal is $s = x+n$, where x represents the true signal and n denotes zero-mean independent noise. The output of the moving-average is:

$$y_i = \frac{1}{k}\sum_{m=-p}^{p} s_{i+m} = \underbrace{\frac{1}{k}\sum_{m=-p}^{p} x_{i+m}}_{\text{signal smoothing term}} + \underbrace{\frac{1}{k}\sum_{m=-p}^{p} n_{i+m}}_{\text{noise attenuation term}}. \tag{2}$$

Now, the variance of the noise is

$$\mathrm{Var}\left(\frac{1}{k}\sum_{m=-p}^{p} n_{i+m}\right) = \frac{1}{k}\sigma_n^2. \tag{3}$$

The noise variance is scaled down to $1/k$ after averaging, resulting in a signal-to-noise ratio (SNR) improvement of $10\log_{10}(k)$ dB and enhancing the discriminability of the signal. The filter primarily attenuates high-frequency noise, while fault-related frequencies are mainly contained in low-frequency regions, thus minimizing adverse effects on fault diagnosis. Visualization results of the vibration signals in Figure 3 further validate our design.

After initial noise reduction, this paper employed convolutions for primary feature identification. Existing studies (W. Zhang et al., 2017) indicate that using larger convolution kernels in the first layer aids in expanding the receptive field, thereby enhancing diagnostic accuracy and robustness. However, selecting an appropriate kernel size is equally critical. To ensure our model's applicability across diverse working conditions, three wide convolutions with varying sizes are implemented. The input signal is denoted as $x \in \mathbb{R}^{1\times L}$, where L represents the signal length. Feature extraction is performed using three convolutional kernels of different sizes (k_1, k_2, k_3), with their respective outputs denoted as y_1, y_2, y_3. The convolution operation is formulated as

$$y_{i,j} = \sum_{m=0}^{k-1} w_j[m]\cdot x[i\cdot s + m - p] + b_j, \tag{4}$$

where i is the index of the output feature map, and j is the channel index of the convolution kernel (corresponding to the number of output channels, set to 16). k is the size of the convolution kernel (set to 32, 64, and 128, respectively). s is the stride (set to 16), and p is the padding. $w_j[m]$ and b_j are the weight and bias of the j-th convolution kernel, respectively. The output size of the convolutional layer can be expressed

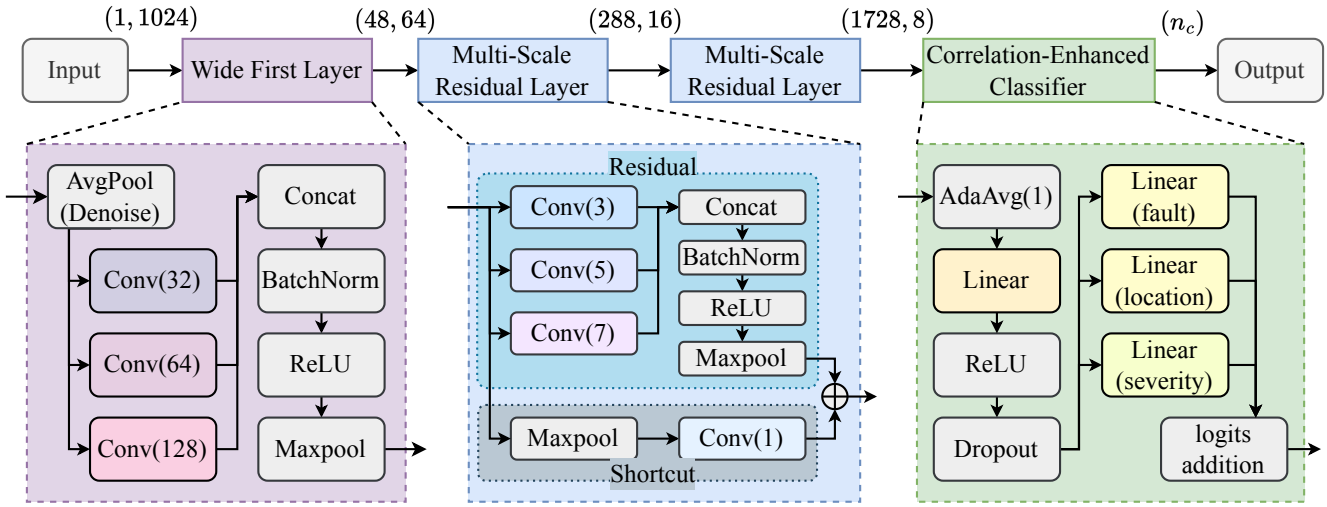


Figure 1. Our proposed Correlation-Enhanced Multi-Scale Residual Network. The wide first layer employs a larger receptive field to extract primary features. Multi-scale residual layers capture features at various scales while using residual connections to prevent degradation. A correlation-enhanced classifier is applied to enhance the transfer learning capability.

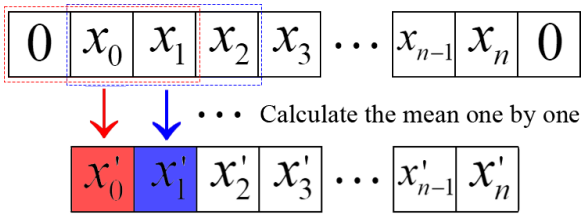


Figure 2. Noise reduction via moving average using average pooling. Initially, the signal is padded with zeros at both ends, and then the mean of each window is calculated.

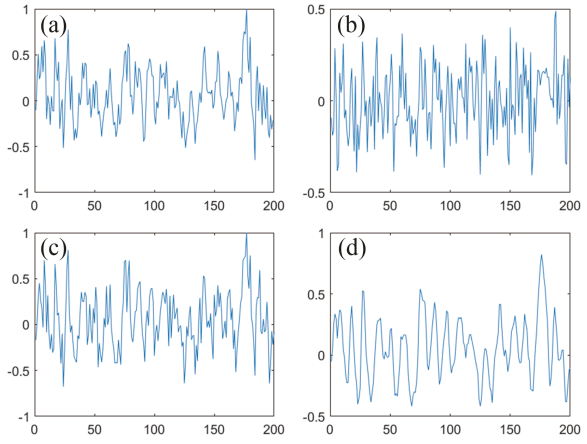


Figure 3. Visualization of the denoising layer's effects. (a) A vibration signal segment from the dataset. (b) Randomly generated Gaussian noise. (c) Observation values obtained by adding noise to the signal, where noise interference poses challenges for fault diagnosis. (d) Denoising results after applying moving average smoothing, which effectively suppresses high-frequency noise while preserving low-frequency fault-related vibration information.

as

$$L_{\text{out}} = \left\lfloor \frac{L_{\text{in}} + 2p - k}{s} \right\rfloor + 1. \tag{5}$$

By setting an appropriate padding p (8, 24, and 56, respectively), convolutions of different sizes can produce outputs of the same length, thus enabling concatenation along the channel dimension.

Next, Batch Normalization (BN) is used to process features at different scales. First, the batch data are standardized, so that the mean is 0 and the variance is 1:

$$\mu_B = \frac{1}{B}\sum_{i=1}^{B} x_i, \tag{6}$$

$$\sigma_B^2 = \frac{1}{B}\sum_{i=1}^{B}(x_i - \mu_B)^2, \tag{7}$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}}, \tag{8}$$

where B is the batch size, and ε is a small constant to prevent the denominator from becoming 0. To preserve the expressive power of the model, the normalized data is then scaled and shifted by learnable parameters γ and β:

$$y_i = \gamma\hat{x}_i + \beta. \tag{9}$$

BN reduces the internal covariate shift by normalizing the input of each layer, i.e., stabilizing the distribution of intermediate layers within the network. The normalized data maintains a stable distribution, which makes the gradient descent more efficient, thus accelerating the training convergence. Meanwhile, it mitigates the impact of extreme values and enhances the stability of gradient propagation. At the end of the wide

first layer, ReLU and max pooling are employed to enhance the non-linearity of the model.

3.2. Multi-Scale Residual Layer

Similar to the wide first layer, the multi-scale residual layer employs three convolutional kernels of different sizes (i.e., 3, 5, and 7) to extract multi-scale features. Their paddings are set to 1, 2, and 3, respectively, to ensure that the output length remains identical for concatenation.

By stacking multi-scale layers, the model's receptive field scale can be expanded exponentially. In our CE-MSRN, a total of 3 layers use multi-scale convolution. Each multi-scale convolution contains 3 different scales, resulting in a total number of receptive field scales being $3^3 = 27$. This significantly enhances the generalization ability and robustness of the model, enabling it to extract fault features under various working conditions and noise environments.

Additionally, residual connections (He, Zhang, Ren, & Sun, 2016) are introduced, which add the input to the output, to address the vanishing gradients and training difficulties in deep neural networks. The Multi-Scale Residual Layer can be expressed as:

$$\mathbf{y} = F(\mathbf{x}) + R(\mathbf{x}), \tag{10}$$

where $\mathbf{x}$ represents the input features of the Multi-Scale Residual Layer, F denotes the extraction of residuals using multi-scale convolution, R signifies the reshaping of the input to match the dimensions of the risidual with learnable parameters, and $\mathbf{y}$ is the final output after incorporating the residual.

Following the multi-scale convolution, the feature dimension increases while the sequence length decreases. To align of the Residual and Shortcut connections, Max Pooling is applied to adjust the sequence length and then a convolution with a kernel size of 1 to match the feature dimensions. This reshaping process minimizes the number of parameters used, thereby improving computational efficiency while retaining as much information from the original input features as possible.

The identity mapping through shortcut connections ensures that gradients do not completely vanish during backpropagation (since the gradient of the shortcut is theoretically always 1). This allows gradients to flow directly back to the shallow layers, avoiding the vanishing gradient problem and stabilizing training. Moreover, residual connections simplify the network's learning task from modeling a complete mapping to only capturing residual components, which reduces complexity and accelerates convergence. Even in scenarios where residuals are minimal (i.e., close to zero), the network can still directly pass the input to the next layer, avoiding performance degradation caused by excessively deep networks. Taking advantage of the above benefits, stacking multi-scale residual layers can increase the receptive field range of the model without encountering issues typically associated with very deep networks.

3.3. Correlation-Enhanced Classifier

In fault diagnosis tasks, existing research (Huo et al., 2023) based on the encoder-decoder structure first employs the encoder to map the vibration signal into state features, and then uses the decoder to classify these features, to determine the state category of the bearing. Taking the CWRU dataset (*Case Western Reserve University Bearing Fault Data, Available: https://engineering.case.edu/bearingdatacenter/download-data-file*, 2019) as an example, the bearing states are categorized into normal condition and 3 different location faults, and each fault is further divided into 3 severity levels, resulting in a total of $1 + 3 \times 3 = 10$ state categories. However, existing methods often use MLP and other methods to classify features into 10 distinct classes without considering the logical relationship among these categories.

This paper introduces a Correlation-Enhanced Classifier that transforms the second linear mapping layer of a two-layer MLP into three linear mappings: one for fault presence, another fir fault location, and the last for severity. This approach converts a single 10-class classification problem into three sub-problems: determining fault existence, classifying fault locations, and accessing severity levels. Assuming the hidden layer of the MLP is $\mathbf{h}$, this process can be expressed as:

$$\mathbf{f} = \mathbf{W}_\mathrm{f}\mathbf{h} + \mathbf{b}_\mathrm{f}, \quad \mathbf{f} \in \mathbb{R}^2, \tag{11}$$

$$\mathbf{l} = \mathbf{W}_\mathrm{l}\mathbf{h} + \mathbf{b}_\mathrm{l}, \quad \mathbf{l} \in \mathbb{R}^3, \tag{12}$$

$$\mathbf{s} = \mathbf{W}_\mathrm{s}\mathbf{h} + \mathbf{b}_\mathrm{s}, \quad \mathbf{s} \in \mathbb{R}^3, \tag{13}$$

Here, $\mathbf{f}_0$ denotes the fault-free logit, $\mathbf{f}_1$ denotes the faulty logit, $\mathbf{l}_i$ denotes the logit at the i-th location, and $\mathbf{s}_j$ denotes the logit of the j-th severity.

Assume that the existence, location, and severity of a fault are probabilistically conditionally independent, that is:

$$P(y = (i, j) \mid \mathrm{f} = 1) = P(\mathrm{l} = i \mid \mathrm{f} = 1) \cdot P(\mathrm{s} = j \mid \mathrm{f} = 1). \tag{14}$$

Convert it into logarithmic form:

$$\log P(y = (i, j) \mid \mathrm{f} = 1) = \log P(\mathrm{l} = i) + \log P(\mathrm{s} = j). \tag{15}$$

The classifier thus constructs a joint probability distribution over all failure categories by additively combining the logits for failure, location, and severity:

$$\mathrm{logits}_{\mathrm{fault}}^{(i,j)} = \mathbf{f}_1 + \mathbf{l}_i + \mathbf{s}_j, \quad \forall i, j \in \{0, 1, 2\}, \tag{16}$$

where, $\mathbf{f}_1$ represents the global contribution of fault existence, $\mathbf{l}_i$ represents the contribution from location i, and $\mathbf{s}_j$ represents the contribution from severity j. The logits for all fault categories are flattened into a 9-dimensional vector, which is then concatenated with the fault-free logit $\mathbf{f}_0$ to obtain the final 10-

dimensional logits:

$$\text{final\_logits} = [\mathbf{f}_0, \mathbf{f}_1 + \mathbf{l}_0 + \mathbf{s}_0, \mathbf{f}_1 + \mathbf{l}_0 + \mathbf{s}_1, \ldots, \mathbf{f}_1 + \mathbf{l}_2 + \mathbf{s}_2] \tag{17}$$

Finally, the probability of each bearing health state can be obtained using the softmax function:

$$P(y = c \mid \mathbf{x}) = \frac{\exp(\text{final\_logits}_c)}{\sum_{k=0}^{9} \exp(\text{final\_logits}_k)}. \tag{18}$$

As cross-entropy is chosen as the loss function, which inherently computes softmax during training (since it utilizes log probabilities), this paper retains the logits in their raw form. This correlation-enhanced classifier explicitly models the fault hierarchy, forcing the model to learn independent distributions of location and severity before combining them into a joint probability. Consequently, this improvement not only enhances fault diagnosis accuracy but also boosts the model's transfer learning capability.

4. EXPERIMENTAL STUDY

4.1. Experimental Setup

The experimental environment utilized a workstation equipped with hardware including an Intel Core i5-12600K CPU and Nvidia RTX4090 GPU, alongside software environments of Debian 12 and Python 3.12. This study implemented the proposed CE-MSRN along with other comparative models based on the PyTorch (Paszke et al., 2019) framework. In training all models, Adam (Kingma & Ba, 2017) served as the optimizer. Hyperparameters during training were configured as follows: batch size set to 128, number of epochs at 50, an initial learning rate of 1×10^{-4}, and this rate was halved every 10 epochs.

We used publicly accessible datasets from Case Western Reserve University (CWRU) and Paderborn University (PU) for experimentation, establishing a precision baseline for bearing failure diagnostics. The CWRU dataset comprised vibration signals sampled with a 12-kHz accelerometer (normal signals were originally sampled at 48 kHz and downsampled to 12 kHz). The faults encompass 3 different locations and 3 different severities, resulting in 10 distinct health state labels, detailed in Table 1. The 3 severity levels of faults were simulated using Electric Discharge Machining (EDM) to create artificial damage of 0.007, 0.014, and 0.021 inches.

The PU dataset simulates damage on the inner and outer rings of bearings using EDM, Electric Engraver, and Drilling. Both the Electric Engraver and Drilling methods involve two different levels. All health state labels are shown in Table 2. Compared to the CWRU dataset, the PU dataset has a higher sampling rate of 64 kHz. This means that for the same sample length (set to 1024 in the experiment), samples from the PU dataset have a shorter duration. This difference makes fault diagnosis with the PU dataset more challenging.

Table 1. Health state labels of the CWRU dataset

Label	Fault	Location	Severity
0	No (f=0)	-	-
1	Yes (f=1)	Inner Ring (l=0)	Minor (s=0)
2	Yes (f=1)	Inner Ring (l=0)	Medium (s=1)
3	Yes (f=1)	Inner Ring (l=0)	Severe (s=2)
4	Yes (f=1)	Ball (l=1)	Minor (s=0)
5	Yes (f=1)	Ball (l=1)	Medium (s=1)
6	Yes (f=1)	Ball (l=1)	Severe (s=2)
7	Yes (f=1)	Outer Ring (l=2)	Minor (s=0)
8	Yes (f=1)	Outer Ring (l=2)	Medium (s=1)
9	Yes (f=1)	Outer Ring (l=2)	Severe (s=2)

Table 2. Health state labels of the PU dataset

Label	Fault	Location	Severity
0	No (f=0)	-	-
1	Yes (f=1)	Outer Ring (l=0)	*EDM* (s=0)
2	Yes (f=1)	Outer Ring (l=0)	*EE1* (s=1)
3	Yes (f=1)	Outer Ring (l=0)	*EE2* (s=2)
4	Yes (f=1)	Outer Ring (l=0)	*D1* (s=3)
5	Yes (f=1)	Outer Ring (l=0)	*D2* (s=4)
6	Yes (f=1)	Inner Ring (l=1)	*EDM* (s=0)
7	Yes (f=1)	Inner Ring (l=1)	*EE1* (s=1)
8	Yes (f=1)	Inner Ring (l=1)	*EE2* (s=2)

Note: *EDM* represents Electric Discharge Machining, *EE1* represents Electric Engraver with level 1, *D1* represents Drilling with level 1, etc.

Both public datasets encompass a variety of working conditions, with specific descriptions detailed in Table 3. In the CWRU dataset, three distinct working conditions are employed to simulate variations in vibration signals resulting from different sensor installation positions. Conversely, the PU dataset utilizes four different working conditions to simulate vibration signal variations caused by changes in speed, torque, and load. This study conducted non-overlapping random sampling within the datasets to ensure sample diversity and prevent data leakage. All samples were normalized to the range of 0-1 before being input into the model to enhance training stability.

Under each working condition, these samples were divided into training, validation, and test sets in a 7:2:1 ratio. The training set was used to train the network weights. During testing, the validation set weights with the lowest loss were loaded. Due to the random processes involved in dataset partitioning and model weight initialization, each experiment was repeated 10 times, and the mean and variance of the results were calculated.

For transfer learning across different working conditions, this paper uses a common method: freezing the encoder parameters and only training the decoder parameters. In the transfer learning experiments, all models are trained using only 10%

Table 3. Working condition descriptions

Condition	Dataset	Description
0	CWRU	Sensor on drive end
1	CWRU	Sensor on fan end
2	CWRU	Sensor on base
3	PU	Baseline condition
4	PU	Lower speed
5	PU	Lower torque
6	PU	Lower load

of the training samples from the target working condition for 5 epochs. This is done to compare the rapid deployment capabilities of these models under cross-working condition scenarios.

Additionally, since the vibration signals from both public datasets were collected on experimental platforms, Gaussian noise was added to the signal samples to better simulate real-world scenarios. In bearing fault diagnosis under noisy conditions, the noise intensity is typically represented by the signal-to-noise ratio (SNR) in dB:

$$\mathrm{SNR}_{\mathrm{dB}} = 10\log_{10}\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right), \tag{19}$$

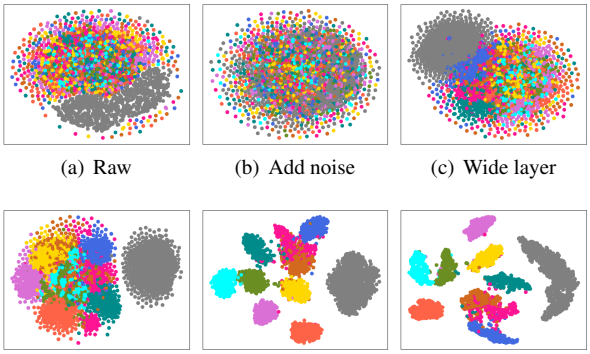
where, P_{signal} represents the signal power, and P_{noise} represents the noise power. A lower SNR indicates a higher noise intensity. When SNR < 0, the noise power exceeds the signal power.

To evaluate the performance of the proposed CE-MSRN, this paper selected several common models for comparison. Table 4 lists the number of parameters, computational complexity, and processing speed of these networks. The processing speed is measured by the number of signal samples processed per second, which evaluates whether the model can achieve online fault diagnosis.

Table 4. Comparison of computational performance

Model	Parameter	Flops	Samples/s
CNN	0.14M	0.78M	13.22K
WDCNN	0.04M	0.36M	7.81K
CE-MSRN	3.25M	12.60M	3.27K
ResNet18	3.98M	175.82M	1.69K
ResNet34	7.35M	359.32M	0.97K

Among all the compared models, our proposed CE-MSRN possesses a moderate number of parameters and achieves reasonable computational speed. Additionally, it surpasses the required sampling rate, enabling real-time fault diagnosis, which is particularly advantageous for applications in the aviation industry.



(a) Raw (b) Add noise (c) Wide layer

(d) Residual layer 1 (e) Residual layer 2 (f) Classifier

Figure 4. The T-SNE visualization of CE-MSRN

4.2. Performance in Noisy Environments

All comparative models are trained using vibration signal samples with added noise. Our proposed CE-MSRN demonstrates the fastest convergence speed and the highest diagnostic accuracy. Figure 5 illustrates the variations in training accuracy and validation accuracy between CE-MSRN and the comparative model (WDCNN) on two public datasets.

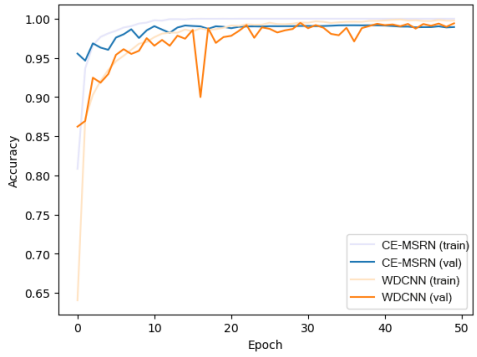
The CE-MSRN presented in this study employs a multi-scale fusion technique, enhancing the model's ability to extract features with greater precision and efficiency in learning. Additionally, the inclusion of residual connections contributes to enhanced training stability, effectively mitigating convergence issues that may arise due to gradient problems, as observed in the case of WDCNN depicted in Figure 5(b).

To provide a visual representation of the model's operation, the T-distributed Stochastic Neighbor Embedding (T-SNE) algorithm is employed for visualization. Each color in the scatterplot corresponds to a specific bearing health state, as shown in Figure 4. Notably, healthy bearings exhibit distinct characteristics in the raw signal, while the identification of states becomes more challenging when noise with the same power as the signal (SNR=0) is added.

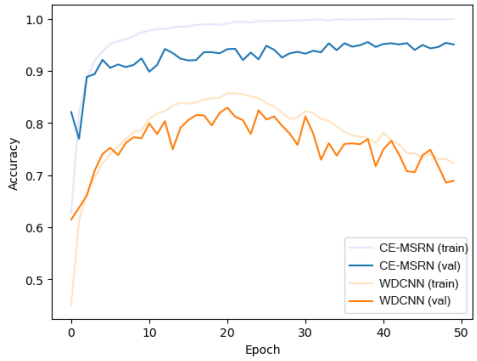
Following the wide first layers, the signals corresponding to healthy bearings are identified. Subsequently, two residual layers are employed to isolate other bearing faults. Ultimately, the fault diagnosis for bearing operation is accomplished through the classifier.

We trained the model using noisy signals under SNR conditions ranging from 10 dB to -10 dB. Table 1 shows the training accuracy for each condition. Here, SNR = None represents the baseline case without added noise.

As shown in Figure 6, the accuracy of all models decreases with the decline of the SNR. It is worth noting that the proposed CE-MSRN achieves both the highest accuracy and the smallest standard deviation across all noise conditions.



(a) CWRU dataset



(b) PU dataset

Figure 5. Training accuracy variation at SNR of 8dB

In practical applications, the test environment's noise often deviates from that encountered during training, leading to considerable performance fluctuations. Consequently, it becomes imperative to investigate the performance of the trained model under various noise conditions during testing.

The training-testing matrices of various models are presented in Figure 7. Through analysis, the following conclusions can be drawn. When the SNR of the training dataset is high, the primary objective of the model is to extract meaningful features from the vibration signal. Conversely, when the SNR is low, the primary task shifts to noise suppression. Consequently, the values near the main diagonal of the matrix

Table 5. The training accuracy of models (%)

SNR	CE-MSRN	WDCNN	CNN	ResNet18	ResNet34
None	96.55±4	92.02±8	94.32±6	79.72±18	79.52±9
10	95.05±7	89.61±10	91.75±8	77.30±13	72.25±14
8	94.28±7	85.90±13	90.44±8	76.32±11	78.47±11
6	93.08±8	86.32±12	89.35±10	76.52±12	60.94±9
4	91.38±10	84.01±13	85.34±10	68.01±8	65.24±9
2	88.33±10	81.89±14	83.30±11	64.76±14	64.00±14
0	86.07±13	79.24±17	79.97±13	61.34±10	61.05±10
-2	79.82±13	73.17±15	75.98±14	58.50±15	53.64±7
-4	72.41±14	66.29±18	68.85±15	47.20±9	49.43±14
-6	63.11±13	57.99±13	59.22±15	42.96±9	42.13±6
-8	48.92±11	50.12±15	47.94±11	39.77±10	35.73±13
-10	42.00±9	36.27±7	36.50±6	33.35±7	30.88±5

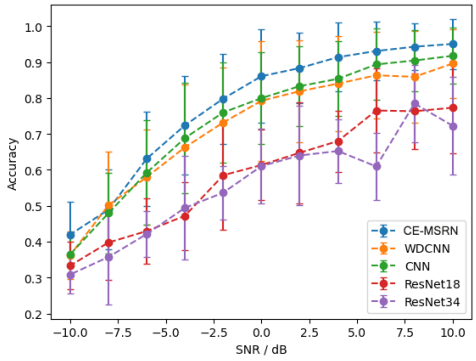


Figure 6. The training accuracy of models

tend to be higher, indicating that optimal test performance is achieved when the SNR in both training and testing is comparable or equal.

Furthermore, it is observed that the upper triangle of the matrix contains lower values compared to the lower triangle. This observation suggests that a model trained in a lower SNR environment exhibits adaptability when applied to a higher SNR scenario, but not vice versa.

To summarize, selecting a training dataset with a similar or lower SNR compared to the expected SNR range in the testing environment enhances the model's robustness in the presence of varying levels of noise.

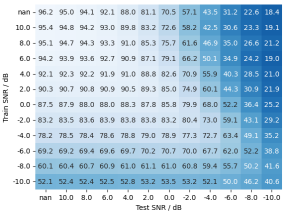
Table 6. Test accuracy comparison of models

Model	Accuracy
CE-MSRN	0.6677±0.22
WDCNN	0.5857±0.22
CNN	0.5865±0.23
ResNet18	0.4328±0.19
ResNet34	0.4239±0.18

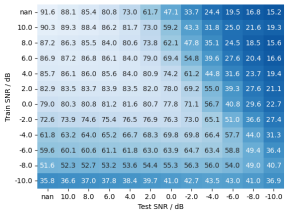
The mean and variance of all values in the above matrix are statistically obtained to establish a comprehensive comparison of the test performance between different models. The results are visually presented in Figure 8, and a detailed summary is provided in Table 6. It is worth noting that the CE-MSRN proposed in this study exhibits a significant advantage over alternative models, consistently maintaining commendable diagnostic performance even under extreme noise conditions.

4.3. Transfer Learning Performance

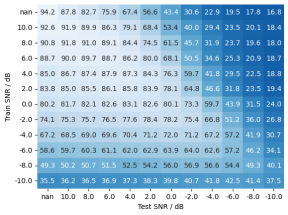
In practical applications within the aviation field, the working condition of bearings often changes, leading to variations in bearing vibration signals. Therefore, it is crucial to study the transfer learning performance of the model under different working conditions. Cross-transfer learning experiments are conducted under seven different working conditions across



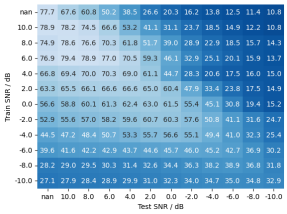
(a) CE-MSRN



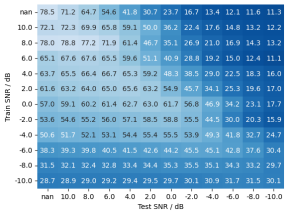
(b) WDCNN



(c) CNN



(d) ResNet18



(e) ResNet34

Figure 7. Training-testing matrices with different SNR

two datasets. The diagnostic accuracy matrix after transfer learning is shown in 9.

In Figure 9, 0-2 represent three working conditions in the CWRU dataset, while 3-6 correspond to four working conditions in the PU dataset. After analyzing the results, several notable findings can be drawn from the comparison. Through experimentation, the highest values appear on the main diagonal of the matrix, indicating that the similarity of signal characteristics is most evident when the source and target domains are the identical. However, transferring from the CWRU dataset to the PU dataset yields suboptimal performance, whereas the reverse transfer from the PU dataset to the CWRU dataset shows more favorable outcoms. This difference may be attributed to the a broader range of fault types in the PU dataset, which encompasse those found in the CWRU dataset. Consequently, this allows for a more comprehensive mapping from the time domain to the feature domain when transferring from PU to CWRU, but not vice versa.

All elements in the matrix above have been quantized and visualized in Figure 10, with more detailed numerical information provided in Table 7. Compared to the shallow CNN, ResNet with two different settings demonstrates superior performance, attributed to their deeper network architecture. The proposed CE-MSRN combines the advantages of multi-scale fusion and residual connections, yielding the most favorable

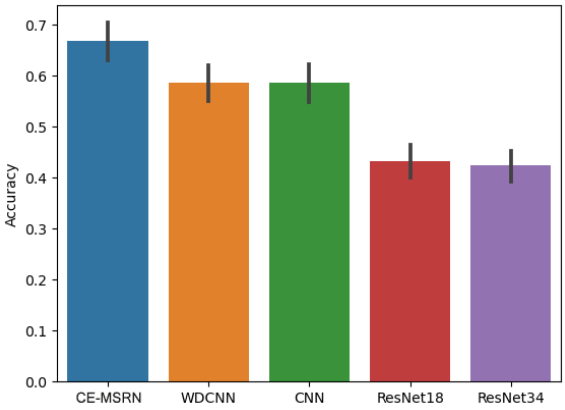


Figure 8. Test accuracy comparison of models

Table 7. Transfer accuracy comparison of models

Model	Accuracy
CE-MSRN	0.8000±0.13
WDCNN	0.6603±0.16
CNN	0.5224±0.24
ResNet18	0.7839±0.15
ResNet34	0.7672±0.16

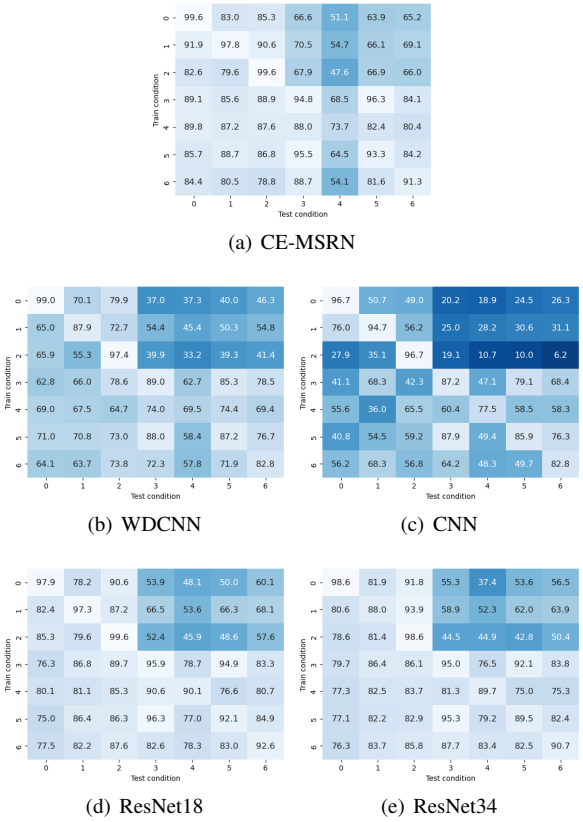
transfer learning results even with a reduced number of layers. Furthermore, the correlation-enhanced classifier explicitly models fault hierarchies, forcing the model to learn independent distributions of location and severity. This improvement facilitates the encoder in learning features independent of the working condition, thereby enhancing the usability of transfer learning.

5. CONCLUSION AND FETURE WORK

In aero-engine and transmission systems, intense noise significantly affects the characteristics of the signals, which brings difficulties to the fault diagnosis of key components such as bearings. Furthermore, practical issues such as the sharing of model weights under different working conditions further complicate the current applications. The existing diagnostic algorithms struggle with these complex integrated tasks.

To address the aforementioned issues, this paper introduces a novel approach called Correlation Enhancement-based Multi-Scale Residual Network (CE-MSRN). The network initially employs temporal moving average for denoising and utilizes multi-scale wide convolution kernels to broaden the model's receptive field, preliminarily extract signal features. Then, it continuously uses two multi-scale residual layers to further extract signal features and stabilize the training process. Finally, a correlation enhancement-based classification layer is used for health state classification.

When the working environment changes, fine-tuning through transfer learning enables achieving favorable recognition re-



(a) CE-MSRN

(b) WDCNN

(c) CNN

(d) ResNet18

(e) ResNet34

Figure 9. Transfer learning matrices of different models

sults with a limited number of samples. The experimental results demonstrate the following advantages of the proposed CE-MSRN:

- **Fewer parameters** compared to existing ResNets, results in reduced reliance on computational resources, and faster processing speed.
- **Stable and rapid convergence** during the model training process.
- **Strong feature extraction capabilities**, leading to effective training results in the presence of various types of noise. This approach also demonstrates advantages in transfer learning, particularly with a limited number of samples.
- **Robustness**, as the trained model maintains its effectiveness under varying noise conditions.

In future work, our proposed network can be further improved, such as by introducing an attention mechanism to assign different weights to convolutional results of different scales. By using methods such as diffusion models, we can preprocess noise signals to achieve more accurate fault diagnosis under noisy conditions. A feature-constrained loss function can further optimize the output of the encoder, thereby enhancing transfer learning performance. Integrating these methods can

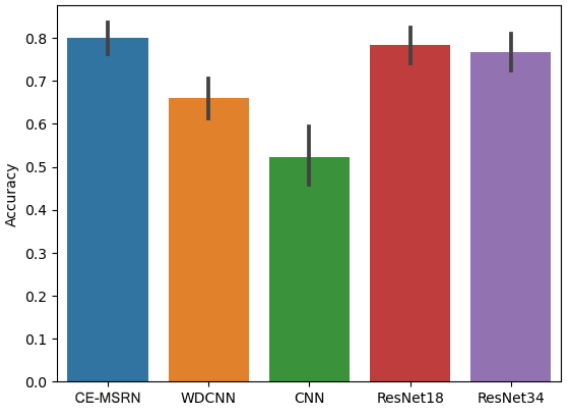


Figure 10. Transfer accuracy comparison of models

more effectively solve the problem of bearing fault diagnosis in the aviation field.

ACKNOWLEDGEMENT

This research was funded by Hunan Provincial Department of Education Excellent Youth Project (Grant No. 23B1104). The APC was funded by Hunan Provincial Department of Education.

REFERENCES

Berghout, T., & Benbouzid, M. (2023, January). Diagnosis and Prognosis of Faults in High-Speed Aeronautical Bearings with a Collaborative Selection Incremental Deep Transfer Learning Approach. *Applied Sciences*, *13*(19), 10916. doi: 10.3390/app131910916

Case western reserve university bearing fault data, available: https://engineering.case.edu/bearingdatacenter/download-data-file. (2019, December). doi: https://engineering.case.edu/bearingdatacenter/download-data-file

Chang, R., Ma, Y., Nie, W., Nie, J., Zhu, Y., & Liu, A.-A. (2024). Causal Disentanglement-Based Hidden Markov Model for Cross-Domain Bearing Fault Diagnosis. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15. doi: 10.1109/TNNLS.2024.3513329

Chen, X., Zhang, B., & Gao, D. (2021, April). Bearing fault diagnosis base on multi-scale CNN and LSTM model. *Journal of Intelligent Manufacturing*, *32*(4), 971–987. doi: 10.1007/s10845-020-01600-2

Dong, Y., Jiang, H., Yao, R., Mu, M., & Yang, Q. (2024, March). Rolling bearing intelligent fault diagnosis towards variable speed and imbalanced samples using multiscale dynamic supervised contrast learning. *Reliability Engineering & System Safety*, *243*, 109805. doi: 10.1016/j.ress.2023.109805

Fernández-Francos, D., Martínez-Rego, D., Fontenla-Romero, O., & Alonso-Betanzos, A. (2013, January). Automatic bearing fault diagnosis based on one-class ν-svm. *Computers & Industrial Engineering*, *64*(1), 357–365. doi: 10.1016/j.cie.2012.10.013

He, K., Zhang, X., Ren, S., & Sun, J. (2016, June). Deep residual learning for image recognition. In *2016 ieee conference on computer vision and pattern recognition (cvpr)* (pp. 770–778). Las Vegas, NV, USA: IEEE. doi: 10.1109/CVPR.2016.90

Hou, Z.-G., Wang, H.-W., Lv, S.-L., Xiong, M.-L., & Peng, K. (2022, December). Siamese multiscale residual feature fusion network for aero-engine bearing fault diagnosis under small-sample condition. *Measurement Science and Technology*, *34*(3), 035109. doi: 10.1088/1361-6501/aca044

Huo, C., Jiang, Q., Shen, Y., Zhu, Q., & Zhang, Q. (2023, May). Enhanced transfer learning method for rolling bearing fault diagnosis based on linear superposition network. *Engineering Applications of Artificial Intelligence*, *121*, 105970. doi: 10.1016/j.engappai.2023.105970

Kabla, A., & Mokrani, K. (2016). Bearing fault diagnosis using hilbert-huang transform (hht) and support vector machine (svm). *Mechanics & Industry*, *17*(3), 308. doi: 10.1051/meca/2015067

Kingma, D. P., & Ba, J. (2017, January). *Adam: A Method for Stochastic Optimization* (No. arXiv:1412.6980). arXiv. doi: 10.48550/arXiv.1412.6980

Lai, Z., Yang, C., Lan, S., Wang, L., Shen, W., & Zhu, L. (2024, June). BearingFM: Towards a foundation model for bearing fault diagnosis by domain knowledge and contrastive learning. *International Journal of Production Economics*, 109319. doi: 10.1016/j.ijpe.2024.109319

Lei, Z., Zhang, P., Chen, Y., Feng, K., Wen, G., Liu, Z., … Yang, C. (2023, October). Prior knowledge-embedded meta transfer learning for few-shot fault diagnosis under variable operating conditions. *Mechanical Systems and Signal Processing*, *200*, 110491. doi: 10.1016/j.ymssp.2023.110491

Li, X., Wang, Y., Yao, J., Li, M., & Gao, Z. (2024, May). Multi-sensor fusion fault diagnosis method of wind turbine bearing based on adaptive convergent viewable neural networks. *Reliability Engineering & System Safety*, *245*, 109980. doi: 10.1016/j.ress.2024.109980

Lu, T., Yu, F., Han, B., & Wang, J. (2020). A generic intelligent bearing fault diagnosis system using convolutional neural networks with transfer learning. *IEEE Access*, *8*, 164807–164814. doi: 10.1109/ACCESS.2020.3022840

Luo, M., Li, C., Zhang, X., Li, R., & An, X. (2016, November). Compound feature selection and parameter optimization of elm for fault diagnosis of rolling element bearings. *ISA Transactions*, *65*, 556–566. doi: 10.1016/j.isatra.2016.08.022

Miao, M., Yu, J., & Zhao, Z. (2022, March). A sparse domain adaption network for remaining useful life prediction of rolling bearings under different working conditions. *Reliability Engineering & System Safety*, *219*, 108259. doi: 10.1016/j.ress.2021.108259

Ni, Q., Ji, J. C., Feng, K., Zhang, Y., Lin, D., & Zheng, J. (2024, February). Data-driven bearing health management using a novel multi-scale fused feature and gated recurrent unit. *Reliability Engineering & System Safety*, *242*, 109753. doi: 10.1016/j.ress.2023.109753

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., … Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems* (Vol. 32). Curran Associates, Inc.

Peng, D., Wang, H., Desmet, W., & Gryllias, K. (2023, April). RMA-CNN: A Residual Mixed-Domain Attention CNN for Bearings Fault Diagnosis and its Time-Frequency Domain Interpretability. *Journal of Dynamics, Monitoring and Diagnostics*. doi: 10.37965/jdmd.2023.156

Peng, H., Du, J., Gao, J., Wang, Y., & Wang, W. (2024, May). Adversarial training of multi-scale channel attention network for enhanced robustness in bearing fault diagnosis. *Measurement Science and Technology*, *35*(5), 056204. doi: 10.1088/1361-6501/ad2828

Peng, H., Wang, W., Gao, J., Wang, Y., & Du, J. (2025). A Lightweight Triple-Stream Network With Multisensor Fusion for Enhanced Few-Shot Learning Fault Diagnosis. *IEEE Transactions on Reliability*, 1–14. doi: 10.1109/TR.2025.3540500

Qian, Q., Qin, Y., Luo, J., Wang, Y., & Wu, F. (2023, March). Deep discriminative transfer learning network for cross-machine fault diagnosis. *Mechanical Systems and Signal Processing*, *186*, 109884. doi: 10.1016/j.ymssp.2022.109884

Rai, V. K., & Mohanty, A. R. (2007, August). Bearing fault diagnosis using fft of intrinsic mode functions in hilbert–huang transform. *Mechanical Systems and Signal Processing*, *21*(6), 2607–2615. doi: 10.1016/j.ymssp.2006.12.004

Shao, X., & Kim, C.-S. (2024, February). Adaptive multi-scale attention convolution neural network for cross-domain fault diagnosis. *Expert Systems with Applications*, *236*, 121216. doi: 10.1016/j.eswa.2023.121216

Tang, H., Tang, Y., Su, Y., Feng, W., Wang, B., Chen, P., & Zuo, D. (2024, January). Feature extraction of multi-sensors for early bearing fault diagnosis using deep learning based on minimum unscented kalman filter. *Engineering Applications of Artificial Intelligence*, *127*, 107138. doi: 10.1016/j.engappai.2023.107138

Wan, L., Li, Y., Chen, K., Gong, K., & Li, C. (2022,

March). A novel deep convolution multi-adversarial domain adaptation model for rolling bearing fault diagnosis. *Measurement*, *191*, 110752. doi: 10.1016/j.measurement.2022.110752

Wang, H., & Li, Y.-F. (2023, September). Robust Mechanical Fault Diagnosis With Noisy Label Based on Multistage True Label Distribution Learning. *IEEE Transactions on Reliability*, *72*(3), 975–988. doi: 10.1109/TR.2022.3190942

Wang, H., Liu, Z., Peng, D., & Cheng, Z. (2022, September). Attention-guided joint learning CNN with noise robustness for bearing fault diagnosis and vibration signal denoising. *ISA Transactions*, *128*, 470–484. doi: 10.1016/j.isatra.2021.11.028

Wang, H., Liu, Z., Peng, D., & Zuo, M. J. (2023, July). Interpretable convolutional neural network with multilayer wavelet for Noise-Robust Machinery fault diagnosis. *Mechanical Systems and Signal Processing*, *195*, 110314. doi: 10.1016/j.ymssp.2023.110314

Wang, J., Zhang, Y., Zhang, F., Li, W., Lv, S., Jiang, M., & Jia, L. (2021, August). Accuracy-improved bearing fault diagnosis method based on avmd theory and awpso-elm model. *Measurement*, *181*, 109666. doi: 10.1016/j.measurement.2021.109666

Wang, Z., Lu, C., Wang, Z., Liu, H., & Fan, H. (2013, August). Fault diagnosis and health assessment for bearings using the mahalanobis–taguchi system based on emd-svd. *Transactions of the Institute of Measurement and Control*, *35*(6), 798–807. doi: 10.1177/0142331212472929

Yan, X., Yan, W.-J., Xu, Y., & Yuen, K.-V. (2023, November). Machinery multi-sensor fault diagnosis based on adaptive multivariate feature mode decomposition and multi-attention fusion residual convolutional neural network. *Mechanical Systems and Signal Processing*, *202*, 110664. doi: 10.1016/j.ymssp.2023.110664

Yu, X., Wang, Y., Liang, Z., Shao, H., Yu, K., & Yu, W. (2023). An Adaptive Domain Adaptation Method for Rolling Bearings' Fault Diagnosis Fusing Deep Convolution and Self-Attention Networks. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1–14. doi: 10.1109/TIM.2023.3246494

Zhang, J., Zhang, K., An, Y., Luo, H., & Yin, S. (2024, May). An Integrated Multitasking Intelligent Bearing Fault Diagnosis Scheme Based on Representation Learning Under Imbalanced Sample Condition. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(5), 6231–6242. doi: 10.1109/TNNLS.2022.3232147

Zhang, W., Chen, D., Xiao, Y., & Yin, H. (2023, October). Semi-Supervised Contrast Learning Based on Multiscale Attention and Multitarget Contrast Learning for Bearing Fault Diagnosis. *IEEE Transactions on Industrial Informatics*, *19*(10), 10056–10068. doi: 10.1109/TII.2023.3233960

Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2017, February). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*, *17*(2), 425. doi: 10.3390/s17020425

Zhang, X., Sheng, C., Ouyang, W., & Zheng, L. (2023, June). Fault diagnosis of marine electric thruster bearing based on fusing multi-sensor deep learning models. *Measurement*, *214*, 112727. doi: 10.1016/j.measurement.2023.112727

Zhao, Y.-P., & Chen, Y.-B. (2022, February). Extreme learning machine based transfer learning for aero engine fault diagnosis. *Aerospace Science and Technology*, *121*, 107311. doi: 10.1016/j.ast.2021.107311

Zheng, X., Nie, J., He, Z., Li, P., Dong, Z., & Gao, M. (2024, March). A fine-grained feature decoupling based multi-source domain adaptation network for rotating machinery fault diagnosis. *Reliability Engineering & System Safety*, *243*, 109892. doi: 10.1016/j.ress.2023.109892