



Research article

A novel multiscale hybrid neural network for intelligent fine-grained fault diagnosis

Chuanjiang Li¹, Shaobo Li^{1,2,*}, Lei Yang¹, Hongjing Wei³, Ansi Zhang^{1,2} and Yizong Zhang¹

¹ School of Mechanical Engineering, Guizhou University, Guiyang, Guizhou 550025, China

² State Key Laboratory of Public Big Data, Guizhou University, Guiyang, Guizhou 550025, China

³ School of Mechanical Engineering, Guizhou Institute of Technology, Guiyang, Guizhou 550003, China

* **Correspondence:** Email: lishaobo@gzu.edu.cn.

Abstract: Various intelligent methods for condition monitoring and fault diagnosis of mechanical equipment have been developed over the past few years. However, most of the existing deep learning (DL)-based fault diagnosis models perform well only when applied to deal with limited types of general failures, and these models fail to accurately distinguish fine-grained faults under multiple working conditions. To address these challenges, we propose a novel multiscale hybrid model (MSHM), which takes the raw vibration signal as input and progressively learns representative features containing both spatial and temporal information to effectively classify fine-grained faults in an end-to-end way. To simulate fine-grained failure scenarios in practice, more than 100 classes of faults under different working conditions are constructed based on two benchmark datasets, and the experimental results demonstrate that our proposed MSHM has advantages over state-of-the-art methods in terms of accuracy in identifying fine-grained faults, generality in handling fault classes of different granularity, and learning ability with limited data.

Keywords: Fine-grained fault diagnosis; multiscale hybrid network; various working conditions; ResCNN; LSTM

1. Introduction

Bearings play an important role in supporting rotating shafts, reducing friction and ensuring rotational accuracy, and are key components of high-end equipment such as aero engines [1], wind turbines [2] and high-speed trains [3]. At the same time, due to the high temperature, high pressure and high-speed working conditions, bearings are also the most prone to failures, and common failure types include but are not limited to ball failures, inner ring failures and outer ring failures, accounting for approximately 30% of the total rotating machinery failures [4]. Therefore, timely and accurate bearing fault diagnosis is basic and necessary to ensure the safe operation of equipment.

Initial research in bearing fault diagnosis mainly focused on signal processing methods, applying techniques such as autoregressive (AR) modelling, fast Fourier transform (FFT), and wavelet packet transform (WPT) to manually extract fault features from the time, frequency, and time-frequency domains [5,6]. On the other hand, these methods are always laborious, time-consuming and require much experience when dealing with large amounts of monitoring data. The successful applications of deep learning (DL) in the field of speech recognition [7], image classification [8] and natural language processing [9] have shown the advantages of DL in automatic feature extraction, pattern recognition and less reliance on expert knowledge, which has inspired scholars to explore intelligent fault diagnosis methods based on DL algorithms.

Diverse advanced DL models have been widely applied in bearing fault diagnosis with their respective features and advantages. Deep Belief Neural Network (DBN) is one of the earliest kinds of generative networks that follows an unsupervised training procedure and has been used to reconstruct fault signals [10]. Convolutional Neural Network (CNN) has the advantage of learning spatial features from monitoring data with a shared-weight architecture of convolutional kernels [11], and Recurrent Neural Network (RNN) is good at capturing temporal and dynamic relationships from sequences and is often used to predict the health conditions of devices [12]. For example, Shao et al. designed a deep wavelet auto-encoder (DWAE) with extreme learning machine (ELM) for unsupervised feature learning from vibration signals [13]. Zhang et al. [14] designed a one-dimensional (1-D) CNN model with wide first-layer kernels (WDCNN) for bearing fault diagnosis with ten classes of faults, achieving an average accuracy of 95.9% in domain adaptation experiments. Hoang et al. [15] applied a vibration image-based CNN model to classify four health statuses of bearings in the Case Western Reserve University (CWRU) dataset: normal condition, inner race failure, ball fault failure, and outer race failure. Chen et al. [16] proposed a multisensory data fusion technique that used sparse autoencoder and deep belief network (SAE-DBN) for feature extraction and classification, respectively. Jiang et al. [17] developed an improved deep recurrent neural network (DRNN) model for identifying both fault category and fault severity of rolling bearings. More details about these methods can be found in Table 1.

Table 1. DL-based bearing fault diagnosis models in existing research.

| Authors | Feature extractor | Faults | Results | Dataset | Ref |
|--------------|---------------------|--------------|-----------------------------------|-------------|------|
| Shao et al. | Wavelet autoencoder | 12 classes | 95.20% | CWRU | [13] |
| Zhang et al. | 1D CNN | 10 classes | Average accuracy = 95.90% | CWRU | [14] |
| Hoang et al. | 2D CNN | 4 classes | Highest accuracy = 100% | CWRU | [15] |
| Chen et al. | Stacked AE | 7 classes | Average accuracy = 97.82% | Own dataset | [16] |
| Jiang et al. | Stacked RNN layers | 12 classes | Average accuracy = 96.53% | CWRU | [17] |
| Fang et al. | 1D CNN | 10/8 classes | Average accuracy = 94.12% | CWRU | [18] |
| Wang et al. | 1D CNN | 11 classes | Accuracy = 99.40% for SNR = 10 dB | Own dataset | [19] |

Although the above methods have achieved a high degree of accuracy in fault diagnosis, most of them are only good at classifying a small number of bearing faults (4–12 categories). However, the reality of faults in the industry is far more complex: modern equipment is highly integrated by multiple components, and each production line may contain many devices and work under different conditions. These factors significantly increase the probability and variety of faults. As the number of fault types increases, the inter-class distance of some faults from the same system or component becomes smaller and their fault characteristics more and more similar [20], which poses challenges to the classification ability of the traditional DL models described above. Therefore, there is an urgent need to propose advanced models to perform large-scale fine-grained fault diagnosis to provide specific and explicit decision-making information for equipment maintenance and repair [21].

Fortunately, several researchers have conducted exploratory studies on this topic, Sun et al. [22] combined multisynchrosqueezing transform (MSST) with sparse feature coding based on dictionary learning (SFC-DL) to achieve fine-grained fault diagnosis considering fault type and fault severity. Wang et al. [23] proposed a three-stage fault diagnosis method that first extracted knowledge from coarse-grained tasks, then transferred the knowledge and finally fine-tuned it on fine-grained tasks, and obtained good performance in large-scale fault diagnosis with 66 classes. From these works, we found that the extraction of discriminative features is the key to solving large-scale fine-grained fault diagnosis.

This paper further puts forward a multiscale hybrid model (MSHM), which consists of multiscale 1-D residual convolutional neural networks (ResCNN) and long short-term memory (LSTM) for fine-grained fault diagnosis, with five main implementation steps: noise filtering, spatial feature encoding, multiscale feature fusion, temporal feature learning, and fine-grained fault identification. The following highlights the main innovations in this paper:

(1) This paper aims to address the challenges of fine-grained fault diagnosis under different working conditions and proposes an intelligent solution of MSHM.

(2) Bearing faults with different fault locations and different fault severities under different operating conditions are considered to be fine-grained faults whose vibration signals are collected for diagnostic purposes.

(3) The discriminative features of fine-grained faults are extracted by integrating the end-to-end spatial and temporal feature encoding capabilities of multiscale 1D ResCNN and LSTM.

(4) Extensive experiments are conducted based on two bearing datasets, and the results of comparison with other popular diagnostic models prove the superiority of the proposed method.

The rest of the paper is organized as follows: Section 2 presents the theoretical background and

the details of the proposed method. Dataset introduction and experimental setup are described in Section 3. In Section 4, the experimental results from two case studies are analyzed. Section 5 concludes this paper and provides some future research directions.

2. Methodology

This part begins by introducing the basic theories of 1D CNN, residual learning and LSTM in processing vibrational signals, and then presents the structure and learning process of our method.

2.1. 1D CNN

CNN is a classical artificial neural network that uses convolutional operations to filter information and produces feature maps from the input data. 1D CNNs are a modified version of 2D CNNs that have some advantages in dealing with sequence signals: automatically learning underlying information of different signals; and processing high-dimensional data with low computational complexity based on a shared-weight architecture. As displayed in Figure 1, the learning process of a conventional CNN with the input vibration signal usually includes the following operations: 1) convolution (Conv), which acts as a filter on the input; 2) batch normalization (BN), which rescales the input of each layer to speed up the training of the model; 3) nonlinear activation (ReLU, sigmoid, etc.) improves the expression ability by introducing nonlinear transformation; and 4) pooling (P), down sampling to reduce the number of parameters. In the design process of CNN, the size of the kernel is an important parameter that determines the receptive field for feature extraction, and the calculation of the receptive field R^l of the l -th CNN layer can be described as:

$$R^l = \prod_{i=1}^{l-1} s^i + \sum_{i=1}^{l-1} [(k^i - s^i) \times \prod_{j=1}^{i-1} s^j], \quad \{l \in \mathbf{N} | l \geq 2, R^1 = 1, s^0 = 1\} \quad (2.1)$$

where k^i and s^i represent respectively the kernel size and the stride of the i -th layer.

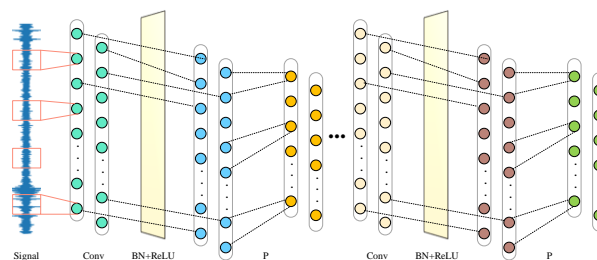


Figure 1. Learning process of basic convolutional neural network based on a 1D vibration signal.

2.2. Residual learning

Although traditional 1D CNNs are good at extracting fault features from time-series signals, their performance degrades significantly when using shallow networks to process nonlinear and high-

dimensional data obtained from variable speed and non-stationary conditions. Therefore, it is often necessary to build deeper neural networks to extract representative features, but the increase in the number of network layers introduces problems such as gradient explosion. To address this problem, He et al. [24] proposed to reduce the error during training by skipping one or several layers through the design of shortcut connections.

For a deep network structure with an input x and a learned feature $H(x)$ the network can learn the residuals $F(x) = H(x) - x$ instead of learning mappings of each stacked layer. The output y of a residual learning block can be described with the following equation and the corresponding schematic diagram.

$$y = F(x, W_i) + x, \quad (2.2)$$

where $F(\cdot)$ and W_i represent the residual mapping function and the model parameters, respectively. The residual learning block shown in Figure 2 consists of two 1D convolution layers and two ReLU activation layers.

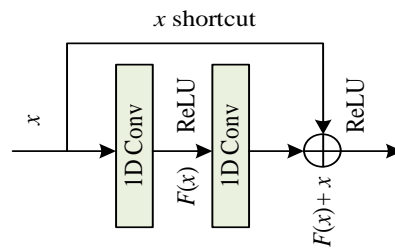


Figure 2. Schematic diagram of a residual learning block.

2.3. LSTM

RNN is another popular neural network that is adept at processing variable length of time-series signals with the memory operation. Moreover, the LSTM is a variant of the RNN that specially designed to avoid the vanishing gradient problem by controlling the memory state. Each typical LSTM layer has four major components: input gate i^t , forget gate f^t , state gate s^t , and output gate p^t . Their corresponding mathematical formulas are listed as follows.

$$i^t = g(W^i h^{(t-1)} + U^i x^{(t)} + b^i), \quad (2.3)$$

$$f^t = g(W^f h^{(t-1)} + U^f x^{(t)} + b^f), \quad (2.4)$$

$$p^t = g(W^p h^{(t-1)} + U^p x^{(t)} + b^p), \quad (2.5)$$

$$s^{t*} = \tanh(W^s h^{(t-1)} + U^s x^{(t)} + b^s), \quad (2.6)$$

$$s^t = f^t \cdot e^{s^{t-1}} + s^{t*} \cdot e^{i^t}, \quad (2.7)$$

$$h^{(t)} = p^t \cdot e^{\tanh(s^t)}, \quad (2.8)$$

Where g denotes the gating function, e represents the element-wise multiply operation, and $h^{(t)}$ is the hidden state. In addition, Figure 3 describes the feature extraction process of LSTM for time series signal x .

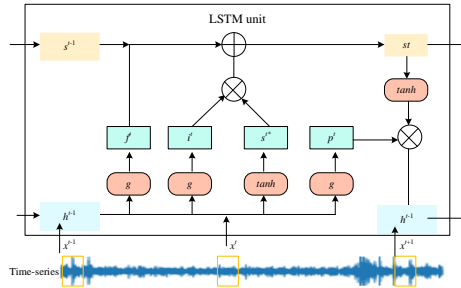


Figure 3. Feature extraction of time-series signal with LSTM network.

2.4. End-to-end MSHM architecture

As shown in Figure 4, the proposed MSHM model is designed with a multiscale and hybrid structure and consists of five main components: a noise-filter module, a multiscale spatial feature encoder, a feature fusion module, a temporal feature encoder and a fault classifier. As described in Eq (2.9), the noise-filter module is a combination of a convolution layer (Conv), a batch normalization layer (BN), a ReLU activation function, and a max-pooling layer (MP). The point is the application of the wide kernel principle [14], which expands the receptive field of the first convolutional layer to improve the sample quality by filtering the noise in the vibration signal x .

$$I = \text{MP}(\text{ReLU}(\text{BN}(\text{Conv}(x)))) \quad (2.9)$$

The information I obtained from the noise-filter module is then fed to the following multiscale spatial feature encoder, where three ResCNN blocks with different kernel scales are established in parallel to extract the spatial features of different failure modes. Specifically, each ResCNN block consists of two subblocks, as depicted in the right part of Figure 4, where residual learning is introduced between the second BN and the ReLU layer to extend the feature learning ability. The kernel sizes of the convolutional layers of the three ResCNN blocks are designed to be 1×3 , 1×5 , and 1×7 . The specific architecture was selected based on two reasons: first, the odd kernel size can avoid alignment errors; second, the fault features extracted from different scales cover fault information from low to high frequencies. After that, a global average pooling (GAP) in each block is applied to avoid overfitting by reducing the spatial feature, and finally three feature maps containing different fault patterns are obtained as follows:

$$F_i = \text{ResCNN block}_i(I), i = 1, 2, 3 \quad (2.10)$$

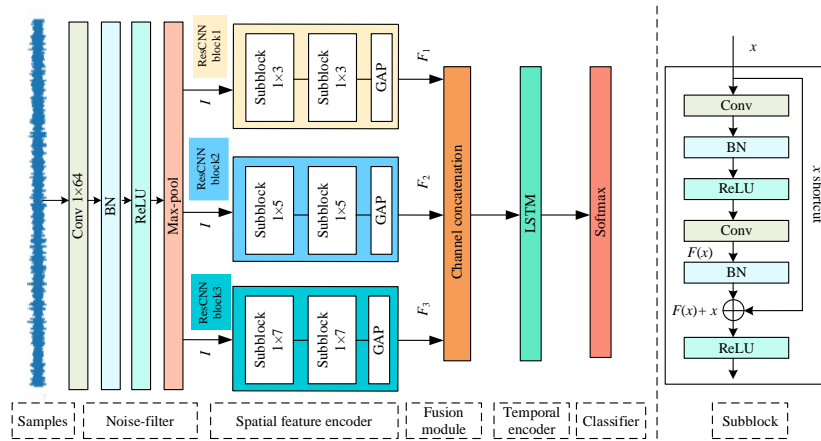


Figure 4. Architecture of the proposed MSHM model.

In order to aggregate the information extracted from the previous layers, the feature fusion module uses the concatenation operation F_{con} to merge the feature map F_1 , F_2 and F_3 in the channel dimension, which can be described as:

$$F_{con} = \text{concat}(F_1, F_2, F_3) \quad (2.11)$$

Although multiscale spatial features have been obtained, they focus only on local features and ignore the sequential relationships hidden in the time-series signals. Therefore, an LSTM-based temporal feature encoder is added, which takes the fused features F_{con} as input and adds important fault information to the cell states or removes redundancy by applying the gating mechanism, and then outputs the state of the last hidden layer, as described in Eq (2.8).

At last, one dense layer with a softmax function plays the role of the classifier to convert the predictions into the probability distributions of fine-grained faults. The softmax is defined as:

$$q(z_j) = \text{softmax}(z_j) = \frac{e^{z_j}}{\sum_{k=1}^n e^{z_k}} \quad (2.12)$$

where z_i represents the output of the j -th neuron and n is the number of the fine-grained faults. One point to note is that we use only one LSTM layer and one dense layer for feature extraction and final classification, which is much more lightweight than existing methods utilizing multiple network layers with a large number of parameters [14,25].

The MSHM-based fine-grained fault diagnosis framework is shown in Figure 5 and its implementation steps include: (1) Acquiring vibration signals of large-scale faults based on the data acquisition system. (2) Constructing fine-grained fault samples for training, validation and testing based on raw signals. (3) The MSHM model is trained to adaptively extract fault features and identify faults with the cross-entropy loss function and the Adam optimization algorithm. (4) In the testing phase, the trained model is used to predict the classes of fine-grained faults and the results are analyzed.

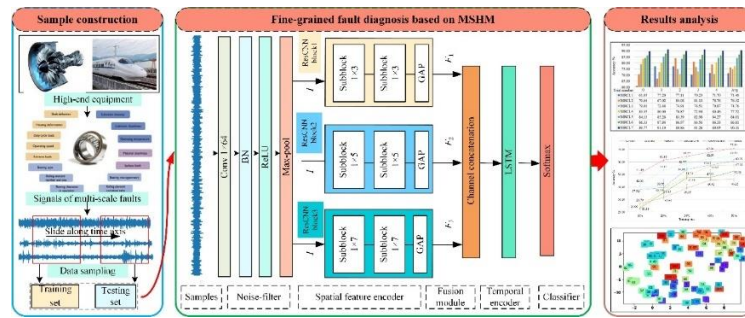


Figure 5. Flowchart of MSHM-based fine-grained bearing fault diagnosis.

3. Datasets and experimental settings

This section describes the two bearing datasets as well as the experimental settings for the validation experiments of large-scale fine-grained bearing fault diagnosis under different working conditions.

3.1. Dataset introduction

This paper studies the problem of fine-grained fault diagnosis with multiple classes. Although there is no standard fine-grained fault dataset as in the field of computer vision, we innovate the sample organization based on two benchmark datasets to simulate the large-scale fine-grained bearing faults in practice. Inspired by the sample organization method of [23], we consider various practical factors such as load, speed, health status, bearing type, damage size, and treat each failure under different working conditions as a class of fine-grained fault.

CWRU dataset. The Case Western Reserve University Bearing Fault Database (CWRU) is commonly used in evaluating bearing fault diagnosis methods [24]. The test bench used in the experiments is shown in Figure 6a and includes: one fan-end bearing, one drive-end bearing, a 2 hp motor, a torque encoder and a dynamometer. The bearings with different fault types (normal, ball fault, inner and outer raceway faults) with different severities (0.007 to 0.028 inches) were seeded using electro-discharge machining, and four loads (0 to 3 hp) and four speeds (1730 to 1797 RPM) were chosen to conduct the experiments. As listed in Table 2, 109 classes of fine-grained faults under different conditions are constructed based on the CWRU dataset in this paper.

PU dataset. Paderborn University (PU) bearing dataset is another popular benchmark with a higher level of complexity as it contains not only artificially damaged faults but also naturally damaged faults. Artificially damaged faults were caused by electric discharge machining, drilling and manual electric engraving, and the artificial fault signals were obtained with the test rig shown in the left part of Figure 6b. As for the real bearing faults, they were generated by accelerated lifetime test, as depicted in the left part of Figure 6b. Failure data of 32 different bearings were obtained under each working condition: 6 healthy bearings, 12 artificial damaged bearings and 14 accelerated lifetime tested bearings. As listed in Table 3, we organized 128 categories of health statuses as fine-grained faults based on PU dataset.

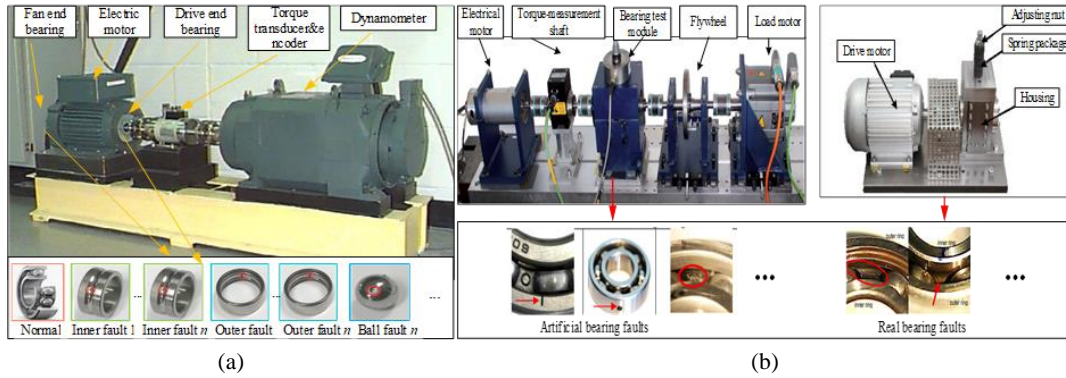


Figure 6. Test bench used in: (a) CWRU dataset, (b) PU dataset.

Table 2. Fault and label information of CWRU dataset.

| Bearing type | Load/hp | Speed/rpm | Health status | Damage size/ inches | Classes | Labels |
|------------------------------|---------|-----------|---------------|------------------------|-------------------------|--------|
| 6205-2RS JEM SKF | | | | | $4 \times 15 + 4 = 64$ | 0–63 |
| 12k frequency (Drive end) | 0/ | 1797/ | Normal, | 0.007, | | |
| | 1/ | 1772/ | ball, | 0.014, | | |
| 6203-2RS JEM SKF | 2/ | 1750/ | inner race, | 0.021, | | |
| 12k frequency (Fan end) | 3 | 1730 | outer race | 0.028 | $12 + 11 \times 3 = 45$ | 64–108 |

Table 3. Fault and label information of PU dataset.

| Load torque/Nm | Radial Force/N | Speed/rpm | Health status | Damage level | Classes | Labels |
|----------------|-------------------|-----------|----------------------|-----------------|--------------------------|--------|
| 0.7/ | 1000/ | 1500/ | Normal, | 1, | $(6 + 12 + 14) \times 4$ | |
| 0.7/ | 1000/ | 900/ | Inner ring, | 2, | $= 128$ | 0–127 |
| 0.1/ | 1000/ | 1500/ | outer ring, | 3, | | |
| 0.7 | 400 | 1500 | Inner and outer ring | | | |

3.2. Experimental settings

To avoid the information loss and maintain sequential characteristics, our proposed method directly utilizes raw vibration signal as input and employs the time-window-based sequence sampling strategy as shown in Figure 7: first, each vibration signal is divided into two disjoint parts according to the time order of the signal generation, and then a time-window is slid along the time-axis of the two split parts to generate samples for training and test sets in turn. The length and the shift step of the time-window are set to be 1024 and 100, and the validation set is generated by taking 20% of samples from the training set. Specifically, 160, 40, and 20 samples of each type of fault were randomly selected from the corresponding sets for training, validation and testing, respectively.

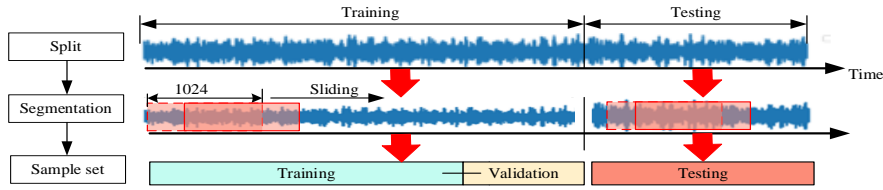


Figure 7. Schematic diagram of the sequence sampling strategy.

In the following experiments, the effectiveness of our proposed method is verified by comparing it with two conventional DL models—1D CNN and Bidirectional LSTM (BiLSTM) [26], and three state-of-the-art methods, including WDCNN [14], a multiscale kernel based residual convolutional neural network (MK-ResCNN) [27] and a multi-scale CNN and LSTM model (MCNN-LSTM) [25]. All these models were trained on an NVIDIA GeForce RTX 2080 Ti GPU using the PyTorch 1.7 framework, and the learning rate and batch size are set to be 0.005 and 1024, respectively, through exploratory experiments.

In addition, accuracy and F1 score are introduced as the evaluation indicators for the model performance. Accuracy is defined by the ratio of the correctly predicted samples to the total number of samples, which describes the degree of closeness of predictions to the true fault classes. The F1 score measures the model performance on each class. The mathematical formulas for these two indicators are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (13)$$

$$F_1 = \frac{2TP}{2TP + FP + FN}, \quad (14)$$

where TP, TN, FP, FN are the amount of true positive, true negative, false positive and false negative, respectively.

4. Validation experiments and analysis

Two case studies based respectively on the CWRU and the PU datasets, are conducted to validate the model performance in large-scale fine-grained bearing fault diagnosis, and the experimental results are analyzed.

4.1. Case study 1: Validation based on the CWRU dataset

4.1.1. MSHM model construction

Our proposed MSHM is a multiscale and hybrid model whose construction involves many parameters, and in order to search for the optimal model structure, we used the control variable method to adjust some important parameters to observe the changes of model performance. According to

Table 4, seven MSHM models with different parameters were built and compared based on the CWRU dataset with 109 fine-grained faults, and each model was performed five times and the average accuracy was calculated.

Table 4. Architecture of MSHM model.

| Model | Architecture and parameters |
|-------|--|
| MSHM1 | C(32/15)–(R ₁ × ₃ (32,64,128), R ₁ × ₅ (32,64,128), R ₁ × ₇ (32,64,128))–L(128,2)–FC(128) |
| MSHM2 | C(64/15)–(R ₁ × ₃ (32,64,128), R ₁ × ₅ (32,64,128), R ₁ × ₇ (32,64,128))–L(128,2)–FC(128) |
| MSHM3 | C(128/15)–(R ₁ × ₃ (32,64,128), R ₁ × ₅ (32,64,128), R ₁ × ₇ (32,64,128))–L(128,2)–FC(128) |
| MSHM4 | C(64/15)–(R ₁ × ₃ (32,64,128), R ₁ × ₅ (32,64,128), R ₁ × ₇ (32,64,128))–L(256,2)–FC(256) |
| MSHM5 | C(64/15)–(R ₁ × ₃ (64,128), R ₁ × ₅ (64,128), R ₁ × ₇ (64,128))–L(256,2)–FC(256) |
| MSHM6 | C(64/15)–(R ₁ × ₃ (64,128), R ₁ × ₅ (64,128), R ₁ × ₇ (64,128))–L(256,1)–FC(256) |
| MSHM7 | C(64/64)–(R ₁ × ₃ (64,128), R ₁ × ₅ (64,128), R ₁ × ₇ (64,128))–L(256,1)–FC(256) |

*Note: Explanation (Take MSHM1 as an example): C(32/15) indicates that the first convolutional layer has 32 convolution kernels of size 15; R₁ ×₃(32,64,128) represents the number of convolution kernels of the 3 convolution layers contained in the ResCNN block with a convolution kernel of 1 × 3, as do R₁ ×₅(32,64,128) and R₁ ×₇(32,64,128); L(128,2) represents the size and number of hidden layers in the LSTM layer are 128 and 2, respectively; FC(128) means the input size of the final fully connected layer is 128. All convolutional layers have a stride of 2.

As can be seen from Figure 8, different parameter choices have a huge impact on model performance. MSHM1 performs poorly because there are only 32 convolution kernels in the noise-filter module, which cannot effectively filter the noise in the signal. As the number of convolution kernels increases, the model performance improves the most when the number is 64. Comparing MSHM5 with the first 4 models, we found that multiscale spatial feature encoder improves the model performance by more than 7% when configured with 64 and 128 convolution kernels for the first and the second subblock in each ResCNN block. MSHM6 demonstrates that the single LSTM layer is suitable for temporal feature encoding and also reduces the model complexity. MSHM7 is the best model for fine-grained fault classification with an average accuracy of 90.41%, and therefore we finally use it as the architecture of our proposed model in this paper.

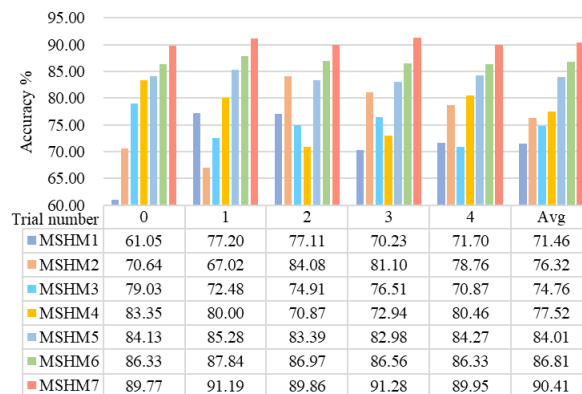


Figure 8. Comparison results of the MSHM models with different parameters.

4.1.2. Comparison of model performance with state-of-the-art methods

The fine-grained bearing fault diagnosis task was built considering various factors present in the actual production, such as load, speed, bearing type and damage size, and finally a total of 109 classes of faults counted based on the CWRU dataset. In comparison with other state-of-the-art diagnosis models, each experiment was performed five times, and the minimum accuracy, maximum accuracy, average accuracy and F1 score were used for a comprehensive evaluation. The results are shown in Figure 9.

The experimental results of fine-grained fault diagnosis clearly demonstrate the effectiveness of our proposed MSHM model, which outperforms the other five DL diagnostic models with its end-to-end discriminative feature encoding capability. MSHM has achieved the highest accuracy of 91.28%, and even its minimum accuracy of 89.77% is higher than other models' maximum accuracy. The WDCNN, on the other hand, ranks second in model performance, with F1 score and average accuracy of 87.05% and 88.13%, respectively, which are 2.84% and 2.28% lower than MSHM. Meanwhile, MK-ResCNN shows the lowest performance in classifying 109 faults, where it loses 41.05%, 35%, 27.96% and 38.15% respectively when compared with MSHM. There is a considerable performance gap, and this because MK-ResCNN was designed for diagnosing a limited class of faults and it is unable to extract sufficient fault features from fine-grained faults under different conditions.

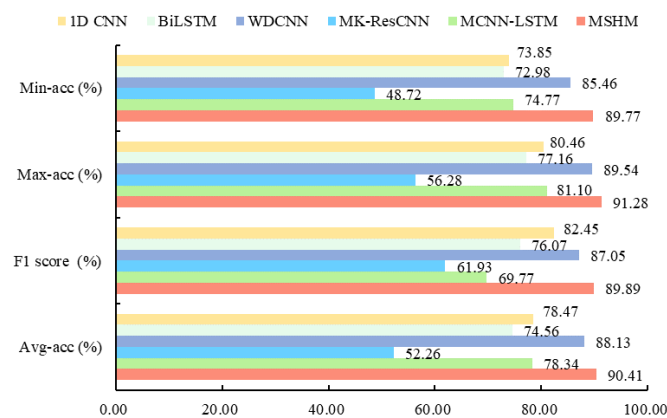


Figure 9. Comparison results of different fault diagnosis models.

4.1.3. Model performance on bearing fault diagnosis with different granularities

In practice, the scale of faults that may occur is related to the complexity of the equipment system and the working conditions. Therefore, a good diagnostic model should have excellent generalization ability to deal with faults of different levels of complexity. Here, we conducted a series of experiments based on the CWRU dataset to validate the performance of the model in classifying bearing faults with different granularities-coarse, medium and fine-grained, as determined by the number of various faults under different working conditions. Specifically, as listed in Table 2, samples of ball, inner race and outer race faults with damage sizes of 0.007, 0.014 and 0.021 inches and normal status under the 1730

rpm working condition, at a total of 10 statuses, are used to establish the coarse-grained fault classification task. Samples of drive-end bearing with all health statuses and all damage sizes under four working conditions are then used to construct the medium-grained fault diagnosis with 64 categories. Finally, samples from all 109 categories of faults from the bearings at the drive end and the fan end are applied to simulate the fine-grained fault classification. Table 5 summarizes the results of the model performance comparison on different granularity fault diagnosis, with the best model indicated in bold.

Table 5. Model performance when classifying bearing faults with different granularities.

| Bearing fault diagnosis with coarse-grained (10) categories | | | | |
|---|-------------|-------------|--------------|-----------------|
| Methods | Max-acc (%) | Min-acc (%) | Avg-acc (%) | F_1 score (%) |
| 1D CNN | 99.50 | 97.50 | 98.90 ± 0.89 | 99.50 ± 1.00 |
| BiLSTM | 100.00 | 99.00 | 99.70 ± 0.45 | 99.00 ± 2.01 |
| WDCNN | 100.00 | 97.00 | 99.30 ± 1.30 | 100.00 ± 0 |
| MK-ResCNN | 100.00 | 96.00 | 96.60 ± 3.97 | 99.50 ± 1.00 |
| MCNN-LSTM | 92.50 | 75.00 | 81.60 ± 6.54 | 91.32 ± 10.64 |
| MSHM | 100.00 | 99.00 | 99.80 ± 0.45 | 100.00 ± 0 |
| Bearing fault diagnosis with medium-grained (64) categories | | | | |
| 1D CNN | 83.36 | 72.11 | 78.84 ± 5.41 | 82.83 ± 22.08 |
| BiLSTM | 77.34 | 73.83 | 75.19 ± 1.37 | 76.67 ± 25.91 |
| WDCNN | 88.98 | 82.42 | 85.70 ± 2.72 | 82.69 ± 22.13 |
| MK-ResCNN | 70.08 | 56.41 | 63.36 ± 5.99 | 63.13 ± 28.30 |
| MCNN-LSTM | 80.55 | 69.84 | 76.98 ± 4.42 | 80.95 ± 21.25 |
| MSHM | 93.52 | 92.03 | 92.95 ± 0.56 | 91.45 ± 17.34 |
| Bearing fault diagnosis with fine-grained (109) categories | | | | |
| 1D CNN | 80.46 | 73.85 | 78.47 ± 2.72 | 82.45 ± 21.24 |
| BiLSTM | 77.16 | 72.98 | 74.56 ± 1.65 | 76.07 ± 23.90 |
| WDCNN | 89.54 | 85.46 | 88.13 ± 1.71 | 87.05 ± 17.09 |
| MK-ResCNN | 56.28 | 48.72 | 52.26 ± 2.91 | 61.93 ± 29.29 |
| MCNN-LSTM | 81.10 | 74.77 | 78.34 ± 2.78 | 69.77 ± 28.30 |
| MSHM | 91.28 | 89.77 | 90.41 ± 0.76 | 89.89 ± 19.38 |

It can be seen that the difficulty of fault classification increases and the model performance decreases from coarse to medium to fine granularity, but our proposed MSHM model achieves the best performance among the models in diagnosing faults with different granularities, with the classification accuracy remaining above 90%. Among the 10 classes of coarse-grained fault diagnosis, all models with the exception of MCNN-LSTM obtained an average accuracy of over 90% and an F_1 score of over 99%, as the fault types are limited, are generated under the same working condition and are easily distinguished from each other. MSHM achieved 100%, 99%, 99.80% and 100% in terms of minimum accuracy, maximum accuracy, average accuracy and F_1 score, 7.50%, 24%, 18.20% and 8.68%, respectively, higher than the performance of MCNN-LSTM model. In medium-grained fault diagnosis with the fault scale of 64 classes, the trend of the average accuracy and the F_1 score indicates that the increase in operating conditions leads to a diversity and similarity of faults, causing a certain degree of degradation in feature extraction of the model. However, the multiscale and hybrid feature encoder design of MSHM model improves the feature extraction capability by simultaneously learning spatial and temporal fault information, and its average accuracy and F_1 score reaches at around 92%, which

are 10% higher than WDCNN's. In the third experiment with fine-grained fault diagnosis, failure samples of another bearing at the fan-end were added and for a total of 109 categories of faults, MSHM maintains its superiority with average accuracy of 90.41% and an F1 score of 89.89%. In a word, the above experiments indicate that our proposed method has better classification and generalization performance in identifying small and large-scale faults with different granularities.

To further explore the feature learning process of the MSHM model, we take the medium-grained fault diagnosis with 64 categories as an example to visualize the feature distribution of all test samples of the spatial feature encoder, the temporal feature encoder and the last fully-connected layer via t-SNE method, as shown in Figure 10. It is clear from the visualization that at the beginning the signals clustered together and afterwards they are fed into the MSHM, each component contributes to the fault feature extraction, producing increasingly clear boundaries between different types of faults, allowing the classifier to eventually achieve accurate fine-grained fault classification.

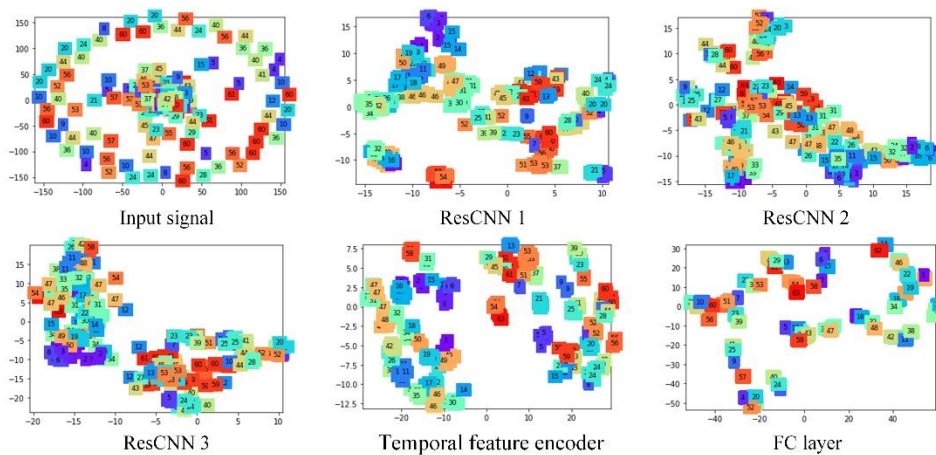


Figure 10. Visualization of feature distribution extracted from MSHM via t-SNE method.

4.2. Case study 2: Validation based on the PU dataset

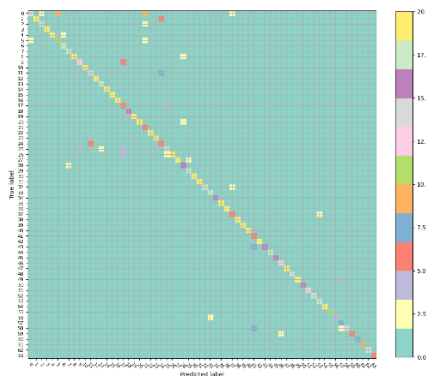
4.2.1. Comparison of model performance with state-of-the-art methods

In this case study, to further verify the model performance, we implemented validation experiments on the PU dataset, which has a higher fault complexity than the CWRU, as reflected in three aspects: first, the PU dataset includes not only artificially generated faults, but also real faults obtained by accelerated life tests; the PU dataset has both single and multiple damage forms; the categories of fine-grained faults of PU dataset are more than that of CWRU. The following experiments were conducted for five times with the same setup as in case study 1, and the model performance is evaluated with the minimum accuracy, maximum accuracy, average accuracy and F1 score, as shown in Table 6.

Table 6. Model performance on fine-grained fault diagnosis based on PU dataset.

| Methods | Max-acc (%) | Min-acc (%) | Avg-acc (%) | F_1 score (%) |
|-----------|-------------|-------------|------------------|-------------------|
| 1D CNN | 61.02 | 56.45 | 59.03 ± 1.71 | 58.16 ± 26.05 |
| BiLSTM | 74.61 | 72.81 | 73.89 ± 0.66 | 74.47 ± 17.80 |
| WDCNN | 70.16 | 67.07 | 68.84 ± 1.31 | 69.70 ± 19.14 |
| MK-ResCNN | 73.52 | 70.55 | 71.98 ± 1.19 | 69.38 ± 22.47 |
| MCNN-LSTM | 69.77 | 63.01 | 67.59 ± 2.82 | 68.91 ± 21.84 |
| MSHM | 80.12 | 78.01 | 79.25 ± 0.81 | 76.95 ± 21.43 |

It is obvious that the overall performance of all models in fine-grained fault diagnosis based on the PU dataset is much lower than that of the CWRU. Specifically, our proposed MSHM model achieves the highest accuracy of 80.12% on the PU dataset, which is 11.16% lower than that obtained on CWRU, while for other comparative methods, especially for 1D CNN and WDCNN, their performance drops by nearly 20%. Such obvious performance difference reflects the difficulty of the PU dataset, and even so, MSHM performs best with an average accuracy of 79.25% and an F1 score of 76.95%, thanks to its strength in feature extraction. BiLSTM performs second best, with a lower standard deviation of both average accuracy and F1 score than MSHM, indicating that temporal features are important and useful for fine-grained fault diagnosis. One thing to note is that MK-ResCNN improves its performance on the PU dataset, obtaining better results than 1D CNN, WDCNN and MCNN-LSTM, which may be due to the fact that MK-ResCNN is designed for fault diagnosis under complex working conditions and the PU dataset maximizes its performance. In addition, the confusion matrix in Figure 11 shows the detailed accuracy of the proposed method for the first 64 classes of fine-grained faults. MSHM can correctly classify most of the faults but performs poorly in classifying the fault classes of 17, 21, 24, 37, 41, 59, and 63, which are mostly bearing faults with actual damage and are more difficult to distinguish than other faults.

**Figure 11.** Confusion matrix of the proposed method for the top 64 categories of fine-grained faults.

4.2.2. Model learning ability with limited training data

The previous experiments are based on the assumption that there are sufficient training samples, but it is challenging to collect massive data for fine-grained faults, and the labelling also requires much more expert knowledge. Thus, in this section, we verify the model learning ability under limited data

condition by reducing the training samples from 50% to 10% of the original amount.

The experimental results in Figure 12 show the influence of the amount of training data on model performance in fine-grained fault diagnosis. When trained with 10% size of the training set and only 16 samples per class of fault are provided for model learning, the compared 1D CNN, WDCNN and MCNN-LSTM model perform poorly with an average accuracy below 30%, MK-ResCNN and BiLSTM perform better with accuracies of 37.92% and 42.08%, which are respectively 13.08% and 8.92% lower than our proposed MSHM method. With the growth of the training size, the performance of the model is improved due to the increase of fault information available during the training process. Specifically, the model performance of MSHM and BiLSTM improves the most when the training sample is increased to 20%, while the rest comparison models obtain the largest performance gain with the 30% training size. Further, the proposed MSHM model is able to achieve impressive performance with an average accuracy of 70.94% when trained with 50% training samples and it is 5% to 20% higher than that of the other models. These results fully demonstrate the excellent learning ability of MSHM with limited samples, we attribute this to the adaptive multi-scale feature extraction and fusion of MSHM, which can compensate for the problem of incomplete information due to insufficient samples.

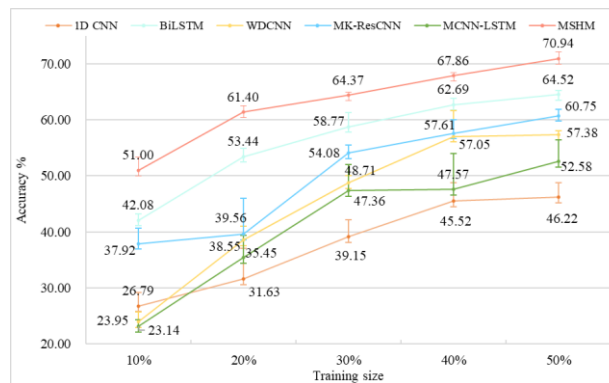


Figure 12. Model performance with limited training data.

5. Conclusion

Most of the existing DL-based models are designed only for the diagnosis of a limited number of faults, and this paper aims to fill the research gap in fine-grained bearing fault diagnosis under various working conditions. Considering the difficulty of diagnosing fine-grained faults due to similar feature patterns, this paper proposes a novel deep multiscale hybrid model consisting of multiscale 1D ResCNNs and LSTM as an intelligent solution, where the noise in the input signal is first removed by a noise-filter module with a wide kernel, and then the extraction and fusion of the spatial and temporal features are realized by the encoders and the fusion layer of the proposed MSHM model, respectively. The performance of the MSHM is comprehensively evaluated on two benchmarks with more than 100 classes of fine-grained faults, and the results of the comparison experiments verify the importance of multiscale and hybrid fault features for fine-grained fault diagnosis and demonstrate the superiority of the MSHM over other mainstream DL models. Additionally, experiments with different fault

granularity prove the general ability and the application potential of the MSHM in practice. In addition, there are two issues need to be further studied in future research, one is to improve the accuracy in diagnosing fine-grained faults with real damages, and the other is to explore the anti-noise capability of the proposed model.

Acknowledgments

This work was supported in part by the Guizhou Province Higher Education Project [No. QJH KY [2020]005, QJH KY [2020]009], and in part by China Scholarship Council [No. 202106670003]. Thanks for the computing support of the State Key Laboratory of Public Big Data, Guizhou University.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. C. J. Li, S. B. Li, A. S. Zhang, L. Yang, E. Zio, M. Pecht, et al., A Siamese hybrid neural network framework for few-shot fault diagnosis of fixed-wing unmanned aerial vehicles, *J Comput Des Eng*, **9** (2022), 1511–1524. <https://doi.org/10.1093/jcde/qwac070>
2. J. L. Chen, J. Pan, Z. P. Li, Y. Y. Zi, X. F. Chen, Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals, *Renewable Energy*, **89** (2016), 80–92. <https://doi.org/10.1016/j.renene.2015.12.010>
3. D. D. Peng, Z. L. Liu, H. Wang, Y. Qin, L. M. Jia, A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains, *IEEE Access*, **7** (2018), 10278–10293. <https://doi.org/10.1109/ACCESS.2018.2888842>
4. R. X. Chen, L. L. Tang, X. L. Hu, H. N. Wu, Fault diagnosis method of low-speed rolling bearing based on acoustic emission signal and subspace embedded feature distribution alignment, *IEEE Trans Ind Inf*, **17** (2020), 5402–5410. <https://doi.org/10.1109/TII.2020.3028103>
5. A. Rai, S. H. Upadhyay, A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings, *Tribol Int*, **96** (2016), 289–306. <https://doi.org/10.1016/j.triboint.2015.12.037>
6. Z. J. Wang, W. H. Du, J. Y. Wang, J. Zhou, X. F. Han, Z. Y. Zhang, et al., Research and application of improved adaptive MOMEDA fault diagnosis method, *Measurement*, **140** (2019), 63–75. <https://doi.org/10.1016/j.measurement.2019.03.033>
7. A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, K. Shaalan, Speech recognition using deep neural networks: A systematic review, *IEEE Access*, **7** (2019), 19143–19165. <https://doi.org/10.1109/ACCESS.2019.2896880>
8. Y. Li, H. K. Zhang, X. Z. Xue, Y. N. Jiang, Q. Shen, Deep learning for remote sensing image classification: A survey, *WIREs Data Min Knowl Discovery*, **8** (2018), e1264. <https://doi.org/10.1002/widm.1264>

9. PM. Lavanya, E. Sasikala, Deep learning techniques on text classification using Natural language processing (NLP) in social healthcare network: A comprehensive survey, *2021 3rd International Conference on Signal Processing and Communication (ICSPC) (IEEE)*, (2021), 603–609. <https://doi.org/10.1109/ICSPC51351.2021.9451752>
10. C. Zhong, J. S. Wang, W. Z. Sun, Fault diagnosis method of rotating bearing based on improved ensemble empirical mode decomposition and deep belief network, *Meas Sci Technol*, **33** (2022), 085109. <https://doi.org/10.1088/1361-6501/ac6cc9>
11. C. J. Li, S. B. Li, A. S. Zhang, Q. He, Z. Liao, J. J. Hu, Meta-learning for few-shot bearing fault diagnosis under complex working conditions, *Neurocomputing*, **439** (2021), 197–211. <https://doi.org/10.1016/j.neucom.2021.01.099>
12. J. L. Li, R. X. Chen, X. Z. Huang, A sequence-to-sequence remaining useful life prediction method combining unsupervised LSTM encoding-decoding and temporal convolutional network, *Meas Sci Technol*, **33** (2022), 085013. <https://doi.org/10.1088/1361-6501/ac632d>
13. H. D. Shao, H. K. Jiang, X. Q. Li, S. P. Wu, Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine, *Knowledge-Based Systems*, **140** (2018), 1–14. <https://doi.org/10.1016/j.knosys.2017.10.024>
14. W. Zhang, G. L. Peng, C. H. Li, Y. H. Chen, Z. J. Zhang, A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, *Sensors*, **17** (2017), 425. <https://doi.org/10.3390/s17020425>
15. D. T. Hoang, H. J. Kang, Rolling element bearing fault diagnosis using convolutional neural network and vibration image, *Cognit Syst Res*, **53** (2019), 42–50. <https://doi.org/10.1016/j.cogsys.2018.03.002>
16. Z. Y. Chen, W. H. Li, Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network, *IEEE Trans Instrum Meas*, **66** (2017), 1693–1702. <https://doi.org/10.1109/TIM.2017.2669947>
17. H. K. Jiang, X. Li, H. D. Shao, K. Zhao, Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network, *Meas Sci Technol*, **29** (2018), 065107. <https://doi.org/10.1088/1361-6501/aab945>
18. H. R. Fang, J. Deng, B. Zhao, Y. Shi, J. Y. Zhou, S. Y. Shao, LEFE-Net: A lightweight efficient feature extraction network with strong robustness for bearing fault diagnosis, *IEEE Trans Instrum Meas*, **70** (2021), 1–11. <https://doi.org/10.1109/TIM.2021.3067187>
19. Z. L. Liu, H. Wang, J. J. Liu, Y. Qin, D. D. Peng, Multitask learning based on lightweight 1DCNN for fault diagnosis of wheelset bearings, *IEEE Trans Instrum Meas*, **70** (2020), 1–11. <https://doi.org/10.1109/TIM.2020.3017900>
20. X. Chu, Y. Lin, Y. S. Wang, X. T. Wang, H. L. Yu, X. Gao, et al., Distance metric learning with joint representation diversification, *International Conference on Machine Learning (PMLR)*, (2020), 1962–1973.
21. C. P. Pang, Management optimization of equipment maintenance and spare parts for automobile intelligent manufacturing enterprises, *Int J Front Eng Technol*, (2022), 4. <https://dx.doi.org/10.25236/IJFET.2022.040507>

22. G. D. Sun, Y. Gao, K. Lin, Y. Hu, Fine-grained fault diagnosis method of rolling bearing combining multisynchrosqueezing transform and sparse feature coding based on dictionary learning, *Shock Vib*, **2019** (2019), 1–13. <https://doi.org/10.1155/2019/1531079>
23. Y. Wang, R. N. Liu, D. Lin, D. Y. Chen, P Li, Q. H. Hu, et al., Coarse-to-Fine: progressive knowledge transfer-based multitask convolutional neural network for intelligent large-scale fault diagnosis, *IEEE Trans Neural Networks Learn Syst*, (2021), 1–14. <https://doi.org/10.1109/TNNLS.2021.3100928>
24. K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, Deep residual learning for image recognition, *Comput Vision Pattern Recognit*, (2021), 770–781. <https://doi.org/10.48550/arXiv.1512.03385>
25. X. H. Chen, B. K. Zhang, D. Gao, Bearing fault diagnosis base on multi-scale CNN and LSTM model, *J Intell Manuf*, **32** (2021), 971–987. <https://doi.org/10.1007/s10845-020-01600-2>
26. Z. B. Zhao, T. F. Li, J. Y. Wu, C. Sun, S. B. Wang, R. Q. Yan, et al., Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study, *ISA Trans*, **107** (2020), 224–255. <https://doi.org/10.1016/j.isatra.2020.08.010>
27. R. N. Liu, F. Wang, B. Y. Yang, S. J Qin, Multiscale kernel based residual convolutional neural network for motor fault diagnosis under nonstationary conditions, *IEEE Trans Ind Inf*, **16** (2019), 3797–3806. <https://doi.org/10.1109/TII.2019.2941868>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)