

A rolling bearing fault diagnosis method under insufficient samples condition based on MSLSTM transfer learning

Ping Zhang¹, Debo Liu²

¹Institute of Intelligent Manufacturing, Xinxiang Vocational and Technical College, Xinxiang, 453006, China

²Faculty of Engineering, Huanghe Science and Technology College, Zhengzhou, 450063, China

²Corresponding author

E-mail: ¹1990716501@qq.com, ²xmjxlwzy@126.com

Received 17 October 2024; accepted 17 December 2024; published online 22 January 2025

DOI <https://doi.org/10.21595/jve.2024.24621>



Copyright © 2025 Ping Zhang, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract. It usually affects the accuracy and reliability of deep learning based intelligent diagnosis methods under the condition of insufficient samples. Existing methods for handling insufficient samples often have problems such as requiring rich expert experience or consuming a lot of time. To solve the above problems, a rolling bearing fault diagnosis method under insufficient samples condition based on multi-scale long-term and short-term memory network (MSLSTM) transfer learning is proposed, which mainly consists of an improved long-term and short-term memory network named as MSLSTM and transfer learning. By introducing multi-scale convolution operation into the traditional LSTM to improve its drawback that only extracts single type of fault feature information, which leads to poor diagnostic performance in noisy environments. Besides, the pooling layer and global average pooling layer in traditional LSTM are replaced with convolution operation to avoid the problem of information loss. Subsequently, the MSLSTM is combined with transfer learning, and a rolling bearing fault diagnosis method under insufficient samples condition based on MSLSTM transfer learning is proposed, which fine tunes the model parameters using a small amount of target domain data. Feasibility of the proposed method is verified through two kinds of experiments. The proposed method has stronger feature extraction ability and training efficiency compared with other models.

Keywords: multi-scale, convolution operation, long-term and short-term memory networks, global average pooling, insufficient data.

1. Introduction

Rolling bearing is an important part of rotating machinery, and monitoring its running state and timely fault diagnosis of it are very important to avoid heavy loss. Unfortunately, its components are prone to failure and difficult to be detected at early stage. Once fault occurs in rolling bearing, the rotating machinery may lose all or part of its functions and fail completely. Therefore, it is very important to monitor and diagnose the rolling bearing regularly so that the potential problems could be solved timely to ensure the normal operation of mechanical equipment [1].

Deep learning has achieved remarkable success in the field of fault diagnosis of rotating machinery, and its data-driven ability makes it a powerful tool [2-3]. However, it is difficult to obtain enough data sets due to the condition constraints in practical application and cost factors [Add 4-6], which has become a major challenge for the development of intelligent diagnosis methods. Currently, the mainstream methods for addressing the problem of sample shortages could be classified into the following three categories roughly: the first type is data enhancement, which increases the number of training samples by synthesizing data sets, and GANS [7] and data interpolation method [8] are the typical representations. A data enhancement framework is proposed to solve the problem of data imbalance, in which a model called GAP is designed, and

a new discriminator is constructed by spectrum normalization. Subsequently, the training process is stabilized by using double time-scale updating rules [9]. The self-attention mechanism is introduced to improve the quality of synthetic data [10]. A one-dimensional vibration signal enhancement method is proposed by combining self-attention mechanism with BPNN [11]. The data resampling method is used to restructure the dataset into a balanced dataset to improve sample utilization [12]. A variational autoencoder is added into the convolutional neural network framework to realize data enhancement [13]. A generative adversarial one-off diagnosis method is introduced, which utilizes health signals to train a bidirectional generative adversarial network. Subsequently, a random forest model facilitates fault diagnosis in industrial robots [14]. An adaptive variational self-encoding generative adversarial network is devised, which integrates an adaptive loss mechanism for fault data augmentation to promote interaction between model loss and function gradients within the network [15]. A state-of-the-art review is delivered and the recent advances of GANs and continuous-variant GANs, including architectures, pros and cons of GANs in terms of theoretical perspective and prognostic and health management (PHM) are presented [16].

The second type for solving sample shortages is transfer learning [17], whose core idea is to use the pre-trained model and adapt to the current problems through fine-tuning. In this way, the training and optimization of the diagnosis model can be realized more quickly. A feature discrimination enhancement method based on multi-scale entropy is proposed, and a transfer learning model based on balanced regularization is designed [18]. The sparsity of the generated countermeasure network is improved, and the trained discriminator is applied on the target domain by using parameter transfer learning [19]. A one-dimensional dual residual squeezing excitation transfer learning network is proposed for unsupervised intelligent diagnosis using unlabeled small samples under cross conditions [20]. The one-dimensional convolutional neural network, a portal circulation network and a neural network of attention mechanism are combined, and one novel transfer learning method based on the hybrid model is constructed [21]. A transfer learning model with convolutional encoders being used as feature extractors is constructed successfully [22]. A fine-grained transfer learning method is developed to solve the imbalanced domain data, which achieves the cross-domain fault diagnosis of bearing and gearbox based on CNN and correlation alignment [23]. A feature correlation matching deep transfer learning method is developed to resolve domain discrepancy and identify bearing defects [24]. Considering the data difference and distribution imbalance in source domain, a deep multi-source transfer learning model is developed for classification of four fault types of bearings [25].

The third type for solving sample shortages is active learning [26], which selects the most representative and critical samples by manual intervention to improve the effect of model training. A convolution-based gated cyclic network is proposed based on the framework of active learning, in which the convolutional neural network is combined with gated cyclic network, and tanh function is used as activation function [27]. A fault diagnosis method based on CNN-BiGRU twin network is introduced by calculating the L1 distance among the pairs of bearing data samples to measure similarity [28]. A rolling bearing fault diagnosis method based on regularization kernel maximum boundary projection weft reduction is proposed [29]. A modified active learning intelligent fault diagnosis method is proposed for rolling bearings with unbalanced samples, which could adeptly employ a limited number of labeled samples to intelligently label the unlabeled samples [30]. The uncertainty-based active learning is used for aiding experts through intelligent fault diagnosis, which search for potential samples of a new type of fault [31]. A data-driven framework is established for integrated design of active fault diagnosis and control while ensuring the tracking performance [32].

However, the above-mentioned methods need spending a lot of time and experience to select the appropriate network structure and super-parameters. Besides, the selected network model is usually only suitable for specific data sets and lacks of universality. Correspondingly, a rolling bearing fault diagnosis method under insufficient samples condition based on MSLSTM transfer learning is proposed, in which a multi-scale long-term and short-term memory network model

(MSLSTM) is constructed, and convolution operation is used instead of pooling layer to increase the ability of feature extraction. Then model transfer learning is used to transfer the model to the target domain, and the sub-model parameters are fine-tuned by using the target domain data.

The main contributions of the paper are as follows:

(1) The multi-scale convolution operation is introduced into traditional LSTM to overcome its drawback that only extracts single type of fault feature information, which usually leads to poor diagnostic performance in noisy environments.

(2) The pooling layer and global average pooling layer in traditional LSTM are replaced with convolution operation to avoid information loss problem.

(3) The proposed MSLSTM is combined with transfer learning, and a rolling bearing fault diagnosis method under insufficient samples condition based on MSLSTM transfer learning is proposed, whose effectiveness and advantage are verified through experiments and comparison.

The rest of the paper are organized as follows: Section 2 introduces the related basic theories used in the paper and the proposed method. Experiment verification and analysis results are presented in Section 3, and conclusion are obtained in Section 4 at last.

2. Theoretical basis

2.1. Multiscale convolution

A single-scale convolution kernel is used by traditional CNN to extract features, so it can only extract a single type of fault feature information, which leads to poor diagnosis effect in noisy environment [33]. Therefore, a multi-scale convolution module is constructed to extract features from different receptive fields and make full use of the excellent feature extraction ability of the convolution kernel in CNN. The multi-scale convolution module is mainly composed of three parts, naming MSC1, MSC2 and MSC3. The convolution kernels with different sizes are used in the three MSC modules. The convolution kernels are 1×3 , 1×5 and 1×7 , respectively.

2.2. LSTM

LSTM is a variant of recurrent neural network (RNN), which is used to process sequence data and has the ability of strong memory and long-term dependence modeling. Compared with the traditional RNN, LSTM introduces memory cells, input gates, forgetting gates and output gates to solve the problems of gradient disappearance and gradient explosion effectively. This makes LSTM perform well in natural language processing, speech recognition and time series prediction. Therefore, it has become one of the most important models in deep learning [34]. Its structure is shown in Fig. 1.

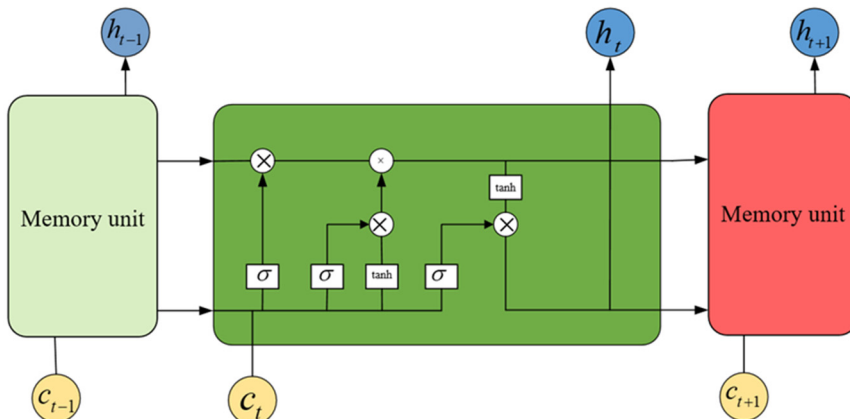


Fig. 1. LSTM structure diagram

Firstly, the calculation formula of input gate is defined as following:

$$i_t = \sigma(W_{xi} \cdot X_t + W_{hi} \cdot h_{t-1} + W_{ci} \cdot c_{t-1} + b_i), \quad (1)$$

where X_t is the input at the current moment, h_{t-1} is the hidden state at the previous moment, c_{t-1} is the neuron state at the previous moment, W_{xi} , W_{hi} and W_{ci} represent the weight matrixes, b_i is the bias vector and σ is the sigmoid function.

The definition of Forgetting Gate is as following:

$$f_t = \sigma(W_{xf} \cdot X_t + W_{hf} \cdot h_{t-1} + W_{cf} \cdot c_{t-1} + b_f), \quad (2)$$

where W_{xf} , W_{hf} and W_{cf} represent the weight matrixes.

The definition of neuron state:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc} \cdot X_t + W_{hc} \cdot h_t), \quad (3)$$

in which \odot represents the element-by-element multiplication, and \tanh represents the hyperbolic tangent function

The definition of Out-gate:

$$o_t = \sigma(W_{xo} \cdot X_t + W_{ho} \cdot h_{t-1} + W_{co} \cdot c_{t-1} + b_o), \quad (4)$$

where W_{xo} , W_{ho} and W_{co} denote weight matrixes.

Finally, the hidden layer is defined:

$$h_t = o_t \odot \tanh(c_t). \quad (5)$$

2.3. Convolution operation instead of pooling layer

Pooling layer is used widely in deep learning model. It can reduce the size of feature map and extract the most important features. The advantage of this method is that it can reduce the number of parameters, prevent over-fitting, and extract the translation invariant features of analyzed data.

Common pooling operations are maximum pooling and average pooling. Maximum pooling is to select the largest eigenvalue in each sliding window as the output, while average pooling is to calculate the average value of eigenvalues in each sliding window as the output. Fig. 2(a) and Fig. 2(b) are the schematic diagram of maximum pooling and average pooling.

Traditional pooling function reduces the size and dimension of feature map by aggregating or compressing the feature values in local areas. However, some detailed information will be lost in this process no matter the maximum pooling or the average pooling. In order to overcome this disadvantage, this paper replaces the pooling layer with convolution layer operation, which not only can achieve similar dimensionality reduction effect as the pooling layer, but also learn more abundant and complex feature information, capture the detailed information in the sequence better than the pooling layer. Fig. 2(c) is a schematic diagram of convolution operation instead of pool layer. The specific formula of convolution instead of pooling is as follows:

$$M = \frac{N - K}{S + 1}, \quad (6)$$

where M is the length of the output sequence, N is the length of the input sequence, K is the convolution kernel size, and S is the step size. It can be seen that the output length will be reduced after convolution operation.

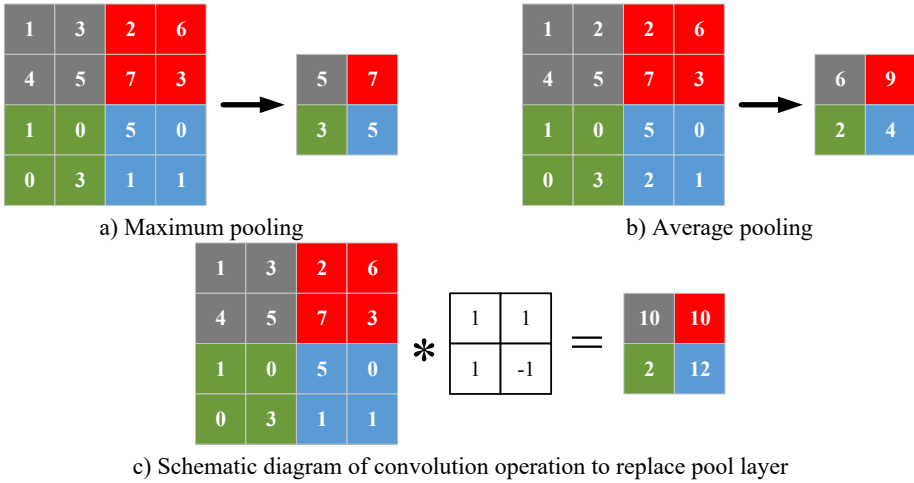


Fig. 2. Commonly used pooling and the proposed scheme

2.4. MSLSTM and transferring learning

Taking full account of spatial features while extracting multi-scale time series features, the Multi-scale convolution layer is combined with LSTM, and MSLSTM is proposed. In addition, the convolution operation is used instead of pooling layer, which not only reduces the dimension, but also avoids the loss of information. Because convolution operation instead of pooling layer will increase the calculation of the network, the traditional pooling layer is replaced with the global average pooling layer after multi-scale long-term and short-term memory network feature fusion. Compared with the fully connected layer, the global average pooling layer retains feature information. Finally, the fault of rotating machinery equipment is identified by softmax classifier. The structure of MSLSTM diagnosis model is shown in Fig. 3.

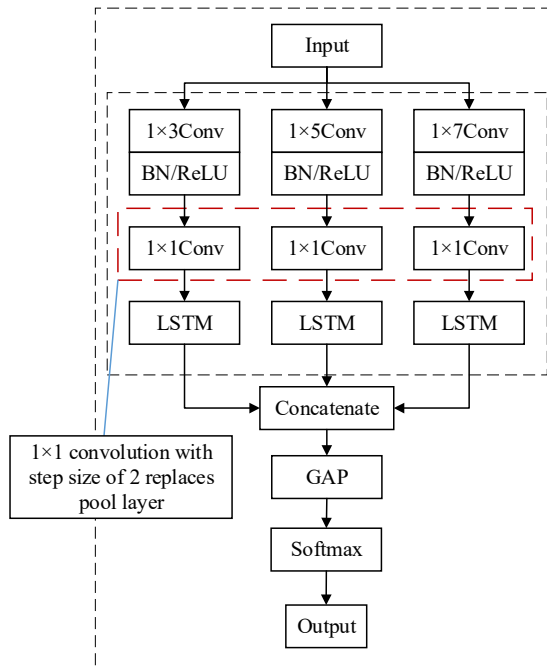


Fig. 3. Structure diagram of MSLSTM

The data in transfer learning is divided into source domain data $X_t = (x_i^t, y_i^t)$ and target domain data $X_r = (x_i^r)$. The source domain data is the data in the knowledge base, that is, the field with data labeling. The target domain is the domain to be learned, which is unlabeled. Transfer learning refers to the transfer of knowledge that has been learned through source domain data to the target domain, and the schematic diagram of transfer learning is shown in Fig. 4.

The specific implementation steps of the transfer learning model based on multi-scale long-term and short-term memory network are as follows:

- (1) Divide the collected vibration signals into source domain and target domain.
- (2) Establish a multi-scale long-term and short-term memory network model and initialize the model parameters.
- (3) Train the model through the source domain data, and store the pre-trained model parameters.
- (4) Use a small amount of target data to fine-tune the model parameters, and use the remaining data for diagnosis.

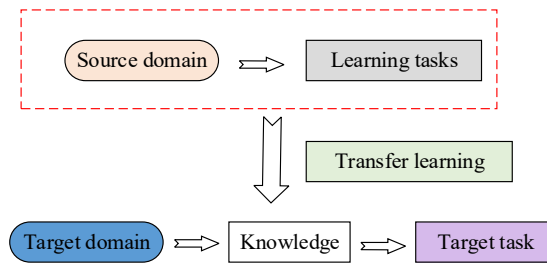


Fig. 4. Transfer learning structure diagram

3. Experiment verification

In order to verify the fault diagnosis performance of the proposed method under the condition of insufficient sample data, the rolling bearing (CWRU) data set of Western Reserve University and QPZZ-II rolling bearing data set are used. Firstly, the diagnostic performance under the varying condition is verified by CWTU data set, and then the domain adaptability is verified by CWRU data set and QPZZ-II data set.

3.1. Data description

3.1.1. CWRU data set

The main components of the rolling bearing test-bed are motor, sensor and dynamometer. SKF6205 bearing is selected as the experimental bearing, and the sampling frequency is 12000 Hz. The length of each sample is set as 1024 data points in order to ensure the reliability of the data set. The three load conditions are 1 HP, 2 HP and 3 HP respectively. Normal (NC) state and three single-point faults, namely inner ring fault (IF), outer ring fault (OFS) and rolling body fault (BFS) are introduced. There are three fault types with different fault size, and the fault diameters are 0.1778 mm, 0.3556 mm and 0.5334 mm respectively. There are 10 kinds of status labels (1 health label and 9 fault labels).

3.1.2. QPZZ-II bearing data set

The QPZZ-II rotating machinery vibration analysis and fault diagnosis test-bed is used to collect rolling bearing signals, which is mainly composed of signal acquisition box, acceleration sensor, driving motor, bearing and other components.

The deep groove ball bearing 6206-2z is selected as the experimental bearing, and the sampling

frequency is 12800 Hz. Normal (NC) state and two kinds of single point faults, namely inner ring fault and outer ring fault, are introduced. There are three fault types, the fault degree is 0.2 mm wide and 0.3 mm deep, 0.4 mm wide and 0.3 mm deep respectively. There are 10 kinds of status labels (1 health label and 9 fault labels), and each type of fault signal contains 1000 samples.

In order to show the superiority of MSLSTM, it is compared with a relative new deep Learning method (ResNet) [35], one traditional migration method (GFK) [36] and one deep migration method (Maximum Mean Difference, MMD) [37]. The backbone structure of MMD network is CNN, and MMD is used for domain adaptation. The ability of feature extraction is enhanced by superimposing multiple channel attention mechanism blocks.

3.2. Fault diagnosis without transfer learning

Firstly, the diagnostic performance of MSLSTM under different workload conditions is verified, and the data of three different loads are used for experiments under the condition of without migration learning, and each type of experiment is done for 10 times. The experimental results are plotted as a box diagram, and the results are shown in Fig. 5(a): it can be seen that the average accuracy rate under three different loads is above 99 %. It verifies that MSLSTM model has high diagnostic accuracy under different working conditions without migration learning. In order to better show the training process of MSLSTM model under different loads, the test results are visualized. Taking the 1HP data set of CWRU as an example, the corresponding accuracy curve and loss curve are shown in Fig. 6(a) and (b), and it can be seen that when the training iteration reaches 15 times, the accuracy curve tends to be stable at 99.8 % and the loss curve tends to be stable. The S-SNE visualization diagram and confusion matrix of the 1HP data set are presented in Fig. 6(c) and (d) respectively, and they further show the excellent performance of MLSTM.

To verify the superiority of MSLSTM over LSTM and improve verification efficiency, inner ring fault data from three different loads in the CWRU dataset are classified based on LSTM. The visualization results are shown in Fig. 7(a) and (b), respectively. Whether based on visual intuitive results or confusion matrix quantization result (53.8 %), there is a significant gap compared to MSLSTM diagnostic results. Similarly, for the QPZZ-II bearing dataset, data corresponding to the same fault size under three operating states of normal, inner ring fault, and outer ring fault are classified based on LSTM. The visualization results are shown in Fig. 7(c) and (d), respectively, based on which only 66.1 % diagnosis accuracy is obtained. The above analysis fully demonstrates that MSLSTM can effectively improve diagnostic accuracy by introducing multi-scale convolution operations into LSTM.

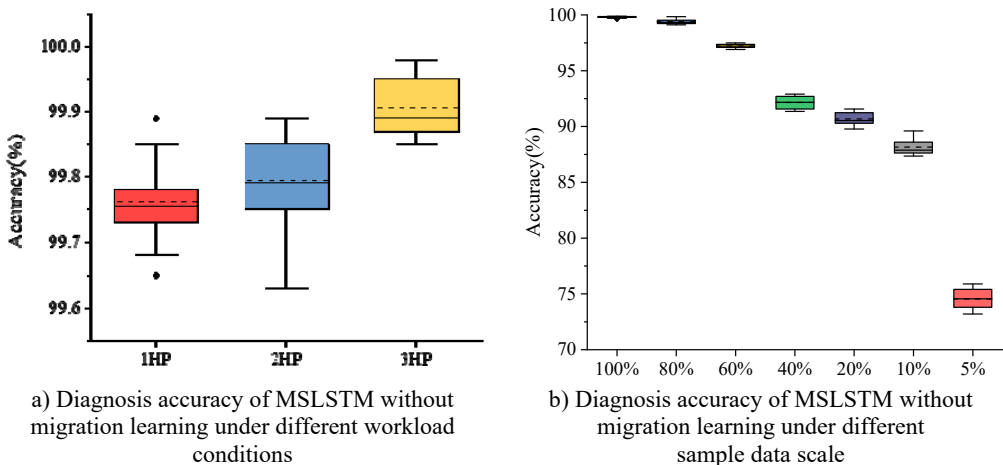


Fig. 5. Diagnosis accuracies of MSLSTM without migration learning under different conditions

Fault diagnosis experiments on CWRU data based on MSLTSM without migration learning by using different sample scales are further carried out. The 1HP data of CWRU is selected as the experimental data set, and 100 %, 80 %, 60 %, 40 %, 20 %, 10 % and 5 % of the total data set are taken respectively. Seven kinds of data set with different sample scales are compared and analyzed, and 10 experiments are carried out on the different selected sample scales. The experimental results are shown in Fig. 5(b), and it can be seen that the diagnostic accuracy is above 90 % when the proportion of all samples is more than 20 %. However, when the proportion is 5 %, the diagnostic accuracy is only 75.8 %. Those verifies that when the sample data is insufficient, the diagnosis performance of MSLSTM will be reduced greatly, so the transfer learning is needed under the condition of insufficient sample data.

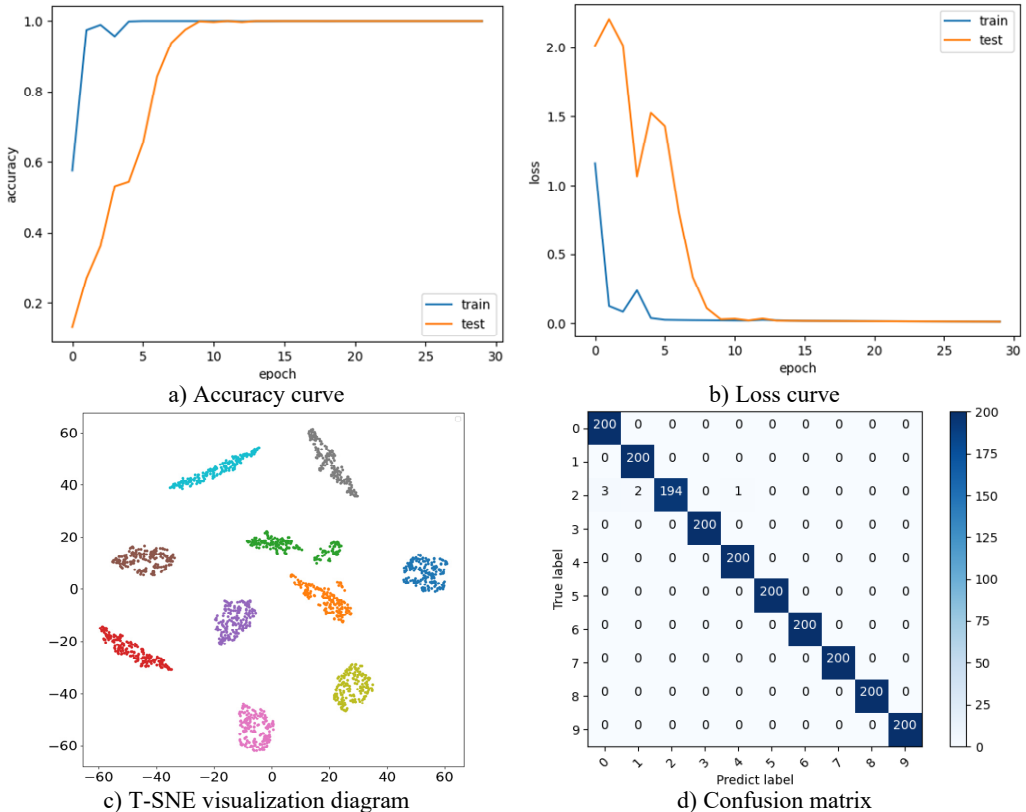


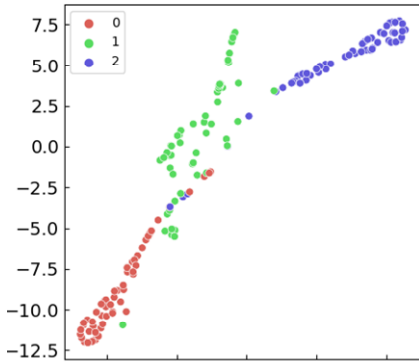
Fig. 6. Accuracy, loss curves and visualization result of MSLSTM without migration learning

3.3. Fault diagnosis with transfer learning

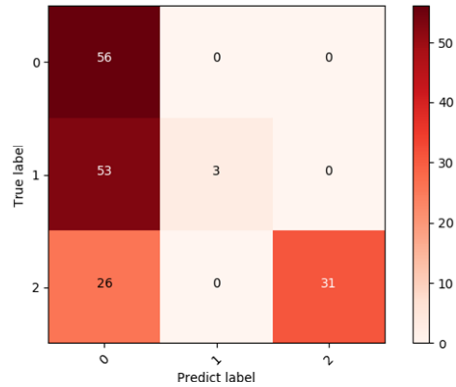
3.3.1. Fault diagnosis with transfer learning on the same experiment platform

The CWRU data set is used to simulate the variable load scenario, and three fault data sets under different loads are selected to verify the diagnosis effect of the proposed model under variable load. Data sets A, B and C represent the data under 1 horsepower load, 2 horsepower load and 3 horsepower load respectively. Taking A→B as an example, data set A is regarded as source domain data and data set B is regarded as target domain data. The model is trained with 100 % source domain data, and is fine-tuned with 5 % data. At the same time, it is compared with previous compared methods, that are ResNet, GFK and MMD. The test results are the average result of 20 calculations, which are shown in Fig. 8. It can be seen that the diagnostic accuracy of MSLSTM with an average accuracy of 93.7 % is higher than the other three methods, indicating that the

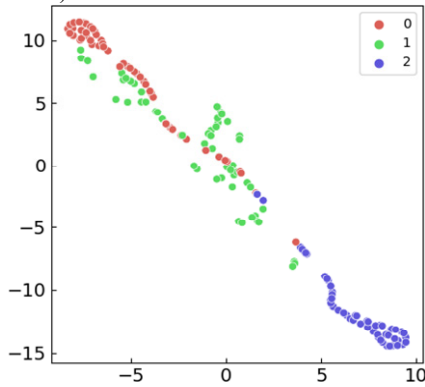
MSLSTM transfer learning diagnosis model has excellent generalization performance and robustness in the condition of variable load scenario.



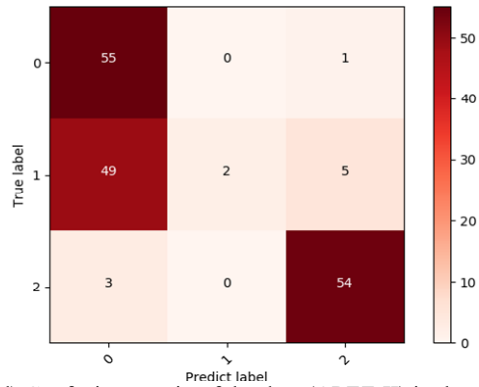
a) T-SNE visualization diagram of the data (CWRU) in three different loads based on LSTM



b) Confusion matrix of the data (CWRU) in three different loads based on LSTM



c) T-SNE visualization diagram of the data (QPZZ-II) in three different running states based on LSTM



d) Confusion matrix of the data (QPZZ-II) in three different running states based on LSTM

Fig. 7. Fault diagnosis results of the two data set based on LSTM



Fig. 8. The diagnosis accuracy of transfer learning under the condition of variable load scenario

3.3.2. Fault diagnosis with transfer learning on cross-platform migration

In the cross-working condition experiment, CWRU data set is used as source domain data and QPZZ-II data set is used as target domain data, and this experiment is called Experiment A. QPZZ-II data set is used as source domain data and CWRU data set is used as target domain data, and this experiment is called Experiment B. In the same experiment, 100 % data is used as the source domain and 5 % data is used as the target domain data. In order to verify the superiority of MSLSTM in cross-working condition diagnosis, it is compared with ResNet. Each method is tested for 10 times in order to eliminate the random initialization error in the network training process, and the results are the average of 20 tests. The experimental results are shown in Fig. 9, based on which it can be seen that the proposed MSLSTM has the highest diagnostic accuracy in both experiment A and experiment B, indicating that the MSLSTM model still has high diagnostic accuracy and domain adaptability despite of the differences between the two groups of data.

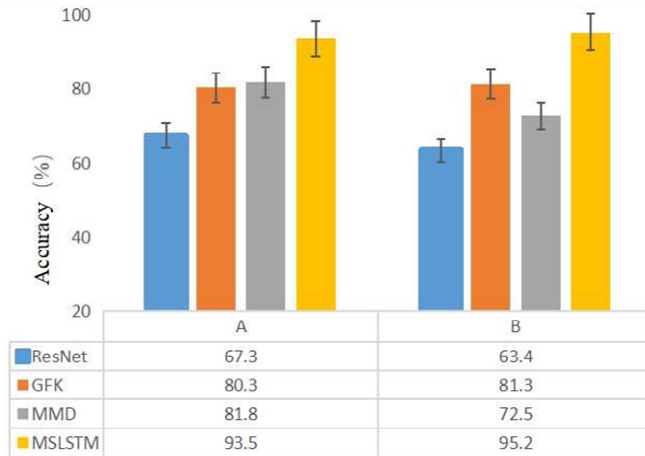


Fig. 9. Cross-platform diagnostic accuracy of participating comparison networks

3.4. Ablation experiment

In order to show the contribution of each branch network of MSLSTM, ablation experiments are conducted among these models: MSLSTM transfer learning (MSLSTM-TL) model, MSLSTM without transfer learning model, multi-scale convolution transfer learning model (MSCNN), non-convolution instead of pooling model (MSLSTM-Maxpooling) and non-global average pooling layer model (MSLSTM-FC). The structures of the comparison networks are shown in Table 1. All the training parameters of the comparison networks are consistent in order to ensure the rigor of the experiment. Taking experiment A and experiment B as examples, 100 % source domain data are also used for training and 5 % target domain data is used for testing. The experimental results are shown in Table 2. It can be seen that the average diagnostic accuracy of MSLSTM-TL is 94.3 %, which is the highest among all the comparison networks. The average diagnostic accuracy of MSLSTM without transfer learning model is 68.3 %, which is 26 % lower than that of transfer learning model, indicating that transfer learning plays an important role in the condition of small sample. The transfer learning model MSCNN is 4.3 % lower than that of the transfer learning model MSLSTM, because MSLSTM pays attention to multi-scale information and avoids the loss of spatial features. The transfer learning models MSLSTM-Maxpooling and MSLSTM-FC are 2.7 % lower than the transfer learning model MSLSTM. It shows that convolution plays an important role in replacing pool layer and GAP layer.

The contribution of each network branch is clearly demonstrated by t-SNE visualization technology. Fig. 10 is the t-SNE visualization diagram of the proposed MSLSTM-TL model, and

it can be seen that 10 health states are clearly separated. In the MSLSTM model without migration, only 5 health states are separated effectively. Besides, the compared three methods, namely MSCNN, MSLSTM-Maxpooling and MSLSTM-FC, most data features are classified and clustered obviously, and only a few sample features cannot be separated in the output layer, which shows that transfer learning plays a key role in small samples, and convolution replaces pool layer and GAP layer to optimize the network.

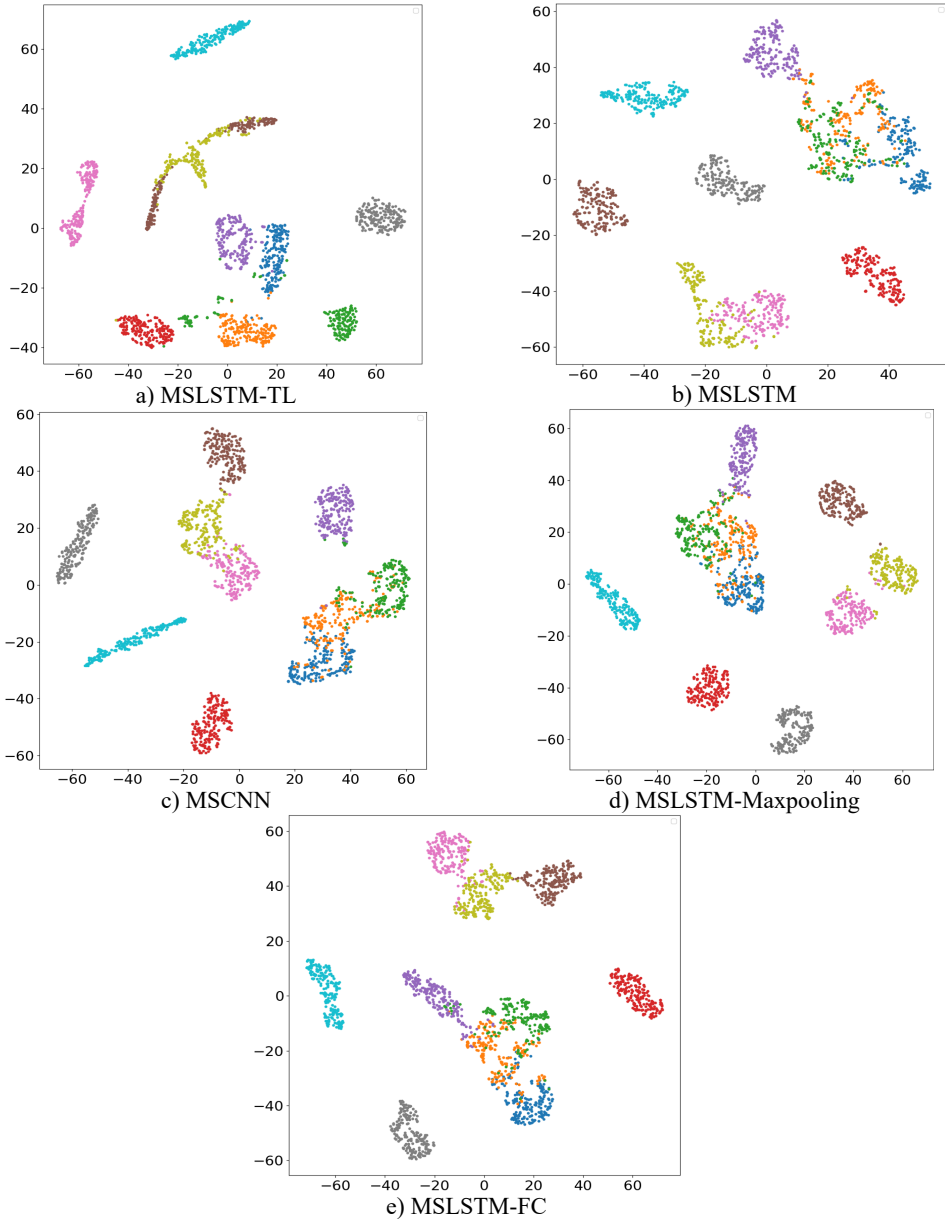


Fig. 10. t-SNE visualization of ablation experiment

Table 1. Branch structure details of ablation experiment

| Network structure | | | | |
|-------------------|-------------------|------|--|-----|
| Model | Transfer learning | LSTM | Convolution instead of pooling operation | GAP |
| MSLSTM-TL | Yes | Yes | Yes | Yes |
| MSLSTM | No | Yes | Yes | Yes |
| MSCNN | Yes | No | Yes | Yes |
| MSLSTM-Maxpooling | Yes | Yes | No | Yes |
| MSLSTM-FC | Yes | Yes | Yes | No |

Table 2. Experimental results of ablation of each branch network

| Model | A | B | Average accuracy (%) |
|-------------------|----------|----------|----------------------|
| MSLSTM-TL | 93.5±0.7 | 95.2±0.7 | 94.3 |
| MSLSTM | 65.7±1.2 | 70.5±1.8 | 68.1 |
| MSCNN | 91.5±0.8 | 88.5±0.6 | 90 |
| MSLSTM-Maxpooling | 92.7±0.5 | 91.4±0.6 | 92 |
| MSLSTM-FC | 93.3±0.8 | 91.8±0.7 | 92.5 |

4. Conclusions

Aiming at solving the problem of deep learning methods dependence on sufficient data sets, a small sample mechanical fault diagnosis model based on MSLSTM-TL is proposed. Firstly, the LSTM model is improved and the MSLSTM network model is constructed, which mainly includes the combination of multi-scale convolution and LSTM, and then the convolution operation is used to replace the pool layer, and the GAP is used to replace the full connection layer. The MSLSTM migration learning model is verified by CWRU bearing data set and QPZZ-II gearbox data set. Under the condition of no migration learning, the CWRU data set is used to verify the feature extraction ability and fault identification ability of MSLSTM. Subsequently, the CWRU data set is used to carry out the off-duty experiments and the cross-duty experiments of QPZZ-II data, and the corresponding results show that the proposed model not only has domain adaptive ability, but also has high accuracy in the case of insufficient sample data. The conclusions derived from this paper are as follow.

- 1) The proposed MSLSTM has superior diagnostic performance when the source domain samples are sufficient, which has been proven through data collected from CWRU that it can achieve a diagnostic accuracy of over 99 % even under variable load condition.
- 2) When the source domain samples are insufficient, the diagnostic accuracy of the proposed MSLSTM will decrease even under the condition of stable load. This indicates that when there are insufficient source domain samples, MSLSTM needs to be combined with transfer learning strategy to solve the problem of decreased diagnostic accuracy.
- 3) By combining the proposed MSLSTM with transfer learning for fault diagnosis in the case of insufficient source domain samples, two different experiments have shown that a diagnostic accuracy of over 93 % can still be achieved under variable load conditions

Although the proposed MSLSTM transfer learning methodology has been effectively applied on fault diagnosis of rolling bearing under varying load conditions, the time-varying speed and time-varying load running condition of rolling bearing also arises simultaneously in real industry applications. Thereby, future works will focus on further developing the proposed method for fault diagnosis of rolling bearing under even more complex running conditions.

Acknowledgements

The authors have not disclosed any funding.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions

Ping Zhang: the overall architecture and the writer of the paper. Debo Liu: the theory study and the algorithm implementation.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- [1] H. Chen, W. Meng, Y. Li, and Q. Xiong, "An anti-noise fault diagnosis approach for rolling bearings based on multiscale CNN-LSTM and a deep residual learning model," *Measurement Science and Technology*, Vol. 34, No. 4, p. 045013, Apr. 2023, <https://doi.org/10.1088/1361-6501/acb074>
- [2] M. Aghaee, A. Mishra, S. Krau, I. M. Tamer, and H. Budman, "Artificial intelligence applications for fault detection and diagnosis in pharmaceutical bioprocesses: a review," *Current Opinion in Chemical Engineering*, Vol. 44, p. 101025, Jun. 2024, <https://doi.org/10.1016/j.coche.2024.101025>
- [3] A. Gong, Z. Qiao, X. Li, J. Lyu, and X. Li, "A review on methods and applications of artificial intelligence on Fault Detection and Diagnosis in nuclear power plants," *Progress in Nuclear Energy*, Vol. 177, p. 105474, Dec. 2024, <https://doi.org/10.1016/j.pnucene.2024.105474>
- [4] J. Cui, P. Xie, X. Wang, J. Wang, Q. He, and G. Jiang, "M2FN: An end-to-end multi-task and multi-sensor fusion network for intelligent fault diagnosis," *Measurement*, Vol. 204, p. 112085, Nov. 2022, <https://doi.org/10.1016/j.measurement.2022.112085>
- [5] Y. Guan, Z. Meng, D. Sun, J. Liu, and F. Fan, "2MNet: Multi-sensor and multi-scale model toward accurate fault diagnosis of rolling bearing," *Reliability Engineering and System Safety*, Vol. 216, p. 108017, Dec. 2021, <https://doi.org/10.1016/j.res.2021.108017>
- [6] C. Wang, C. Xin, Z. Xu, M. Qin, and M. He, "Mix-VAEs: A novel multisensor information fusion model for intelligent fault diagnosis," *Neurocomputing*, Vol. 492, pp. 234–244, Jul. 2022, <https://doi.org/10.1016/j.neucom.2022.04.044>
- [7] Z. Li, D. Jiang, H. Wang, and D. Li, "Video image moving target recognition method based on generated countermeasure network," *Computational Intelligence and Neuroscience*, Vol. 2022, pp. 1–8, Aug. 2022, <https://doi.org/10.1155/2022/7972845>
- [8] X.-L. Song, Y.-L. He, X.-Y. Li, Q.-X. Zhu, and Y. Xu, "Novel virtual sample generation method based on data augmentation and weighted interpolation for soft sensing with small data," *Expert Systems with Applications*, Vol. 225, p. 120085, Sep. 2023, <https://doi.org/10.1016/j.eswa.2023.120085>
- [9] H. Zhang, R. Wang, R. Pan, and H. Pan, "Imbalanced fault diagnosis of rolling bearing using enhanced generative adversarial networks," *IEEE Access*, Vol. 8, pp. 185950–185963, Jan. 2020, <https://doi.org/10.1109/access.2020.3030058>
- [10] S. Liu, H. Jiang, Z. Wu, and X. Li, "Data synthesis using deep feature enhanced generative adversarial networks for rolling bearing imbalanced fault diagnosis," *Mechanical Systems and Signal Processing*, Elsevier BV, 2022.
- [11] Y. Long, W. Zhou, and Y. Luo, "A fault diagnosis method based on one-dimensional data enhancement and convolutional neural network," *Measurement*, Vol. 180, p. 109532, Aug. 2021, <https://doi.org/10.1016/j.measurement.2021.109532>
- [12] Z. Meng et al., "Fault diagnosis of rolling bearing based on secondary data enhancement and deep convolutional network," *Journal of Mechanical Engineering*, Vol. 57, No. 23, p. 106, Jan. 2021, <https://doi.org/10.3901/jme.2021.23.106>
- [13] D. Zhao et al., "Enhanced data-driven fault diagnosis for machines with small and unbalanced data based on variational auto-encoder," *Measurement Science and Technology*, Vol. 31, No. 3, p. 035004, Mar. 2020, <https://doi.org/10.1088/1361-6501/ab55f8>
- [14] H. Chen, J. Wei, H. Huang, L. Wen, Y. Yuan, and J. Wu, "Novel imbalanced fault diagnosis method based on generative adversarial networks with balancing serial CNN and Transformer (BCTGAN),"

- Expert Systems with Applications*, Vol. 258, p. 125171, Dec. 2024, <https://doi.org/10.1016/j.eswa.2024.125171>
- [15] X. Wang, H. Jiang, Z. Wu, and Q. Yang, "Adaptive variational autoencoding generative adversarial networks for rolling bearing fault diagnosis," *Advanced Engineering Informatics*, Vol. 56, p. 102027, Apr. 2023, <https://doi.org/10.1016/j.aei.2023.102027>
- [16] Q. Li, Y. Tang, and L. Chu, "Generative adversarial networks for prognostic and health management of industrial systems: A review," *Expert Systems with Applications*, Vol. 253, No. 1, p. 124341, Nov. 2024, <https://doi.org/10.1016/j.eswa.2024.124341>
- [17] D. Yang, W. Zhang, and Y. Jiang, "Mechanical fault diagnosis based on deep transfer learning: a review," *Measurement Science and Technology*, Vol. 34, No. 11, p. 112001, Nov. 2023, <https://doi.org/10.1088/1361-6501/ace7e6>
- [18] Q. Hu, X. Si, A. Qin, Y. Lv, and M. Liu, "balanced adaptation regularization based transfer learning for unsupervised cross-domain fault diagnosis," *IEEE Sensors Journal*, Vol. 22, No. 12, pp. 12139–12151, Jun. 2022, <https://doi.org/10.1109/jsen.2022.3174396>
- [19] C. Peng, L. Li, Q. Chen, Z. Tang, W. Gui, and J. He, "A fault diagnosis method for rolling bearings based on parameter transfer learning under imbalance data sets," *Energies*, Vol. 14, No. 4, p. 944, Feb. 2021, <https://doi.org/10.3390/en14040944>
- [20] J. Tong, C. Liu, J. Zheng, H. Pan, X. Wang, and J. Bao, "1D-DRSETL: a novel unsupervised transfer learning method for cross-condition fault diagnosis of rolling bearing," *Measurement Science and Technology*, Vol. 33, No. 8, p. 085110, Aug. 2022, <https://doi.org/10.1088/1361-6501/ac6f46>
- [21] C.-Y. Hsu and Y.-W. Lu, "Virtual metrology of material removal rate using a one-dimensional convolutional neural network-based bidirectional long short-term memory network with attention," *Computers and Industrial Engineering*, Vol. 186, p. 109701, Dec. 2023, <https://doi.org/10.1016/j.cie.2023.109701>
- [22] Q. Qian, Y. Qin, Y. Wang, and F. Liu, "A new deep transfer learning network based on convolutional auto-encoder for mechanical fault diagnosis," *Measurement*, Vol. 178, p. 109352, Jun. 2021, <https://doi.org/10.1016/j.measurement.2021.109352>
- [23] J. Dong, D. Su, Y. Gao, X. Wu, H. Jiang, and T. Chen, "Fine-grained transfer learning based on deep feature decomposition for rotating equipment fault diagnosis," *Measurement Science and Technology*, Vol. 34, No. 6, p. 065902, Jun. 2023, <https://doi.org/10.1088/1361-6501/acc04a>
- [24] B. Wang, B. Wang, and Y. Ning, "A novel transfer learning fault diagnosis method for rolling bearing based on feature correlation matching," *Measurement Science and Technology*, Vol. 33, No. 12, p. 125006, Dec. 2022, <https://doi.org/10.1088/1361-6501/ac8d20>
- [25] M. Huang, J. Yin, S. Yan, and P. Xue, "A fault diagnosis method of bearings based on deep transfer learning," *Simulation Modelling Practice and Theory*, Vol. 122, p. 102659, Jan. 2023, <https://doi.org/10.1016/j.simpat.2022.102659>
- [26] J. Shim and S. Kang, "Domain-adaptive active learning for cost-effective virtual metrology modeling," *Computers in Industry*, Vol. 135, p. 103572, Feb. 2022, <https://doi.org/10.1016/j.compind.2021.103572>
- [27] P. Yang and Y. C. Su, "Fault diagnosis of rolling bearing based on convolutional Gated recurrent network," *Journal of Aerospace Power*, Vol. 34, No. 11, pp. 2432–2439, Nov. 2019, <https://doi.org/10.13224/j.cnki.jasp.2019.11.015>
- [28] Z. H. Zhao et al., "Bearing fault diagnosis method based on a CNN-BiGRU Siamese network," *Journal of Vibration and Shock*, Vol. 42, No. 6, pp. 166–171, Jul. 2023, <https://doi.org/10.13465/j.cnki.jvs.2023.06.020>
- [29] X. L. Zhao et al., "Fault diagnosis of rolling bearings based on the dimension reduction using the regularized kernel maximum margin projection," *Journal of Vibration and Shock*, Vol. 36, No. 14, pp. 104–110, Nov. 2017, <https://doi.org/10.13465/j.cnki.jvs.2017.14.016>
- [30] J. Lu, W. Wu, X. Huang, Q. Yin, K. Yang, and S. Li, "A modified active learning intelligent fault diagnosis method for rolling bearings with unbalanced samples," *Advanced Engineering Informatics*, Vol. 60, p. 102397, Apr. 2024, <https://doi.org/10.1016/j.aei.2024.102397>
- [31] L. H. P. D. Silva, L. H. S. Mello, A. Rodrigues, F. M. Varejão, M. P. Ribeiro, and T. Oliveira-Santos, "Active learning for new-fault class sample recovery in electrical submersible pump fault diagnosis," *Expert Systems with Applications*, Vol. 212, p. 118508, Feb. 2023, <https://doi.org/10.1016/j.eswa.2022.118508>

- [32] Z. Yan, F. Xu, J. Tan, H. Liu, and B. Liang, "Reinforcement learning-based integrated active fault diagnosis and tracking control," *ISA Transactions*, Vol. 132, pp. 364–376, Jan. 2023, <https://doi.org/10.1016/j.isatra.2022.06.020>
- [33] Z. Shao, W. Li, H. Xiang, S. Yang, and Z. Weng, "Fault diagnosis method and application based on multi-scale neural network and data enhancement for strong noise," *Journal of Vibration Engineering and Technologies*, Vol. 12, No. 1, pp. 295–308, Jan. 2023, <https://doi.org/10.1007/s42417-022-00844-x>
- [34] Y. Han, N. Ding, Z. Geng, Z. Wang, and C. Chu, "An optimized long short-term memory network based fault diagnosis model for chemical processes," *Journal of Process Control*, Vol. 92, pp. 161–168, Aug. 2020, <https://doi.org/10.1016/j.jprocont.2020.06.005>
- [35] Z. Chen, Y. Wang, J. Wu, C. Deng, and W. Jiang, "Wide residual relation network-based intelligent fault diagnosis of rotating machines with small samples," *Sensors*, Vol. 22, No. 11, p. 4161, May 2022, <https://doi.org/10.3390/s22114161>
- [36] C. Wang, Q. Zhang, and L. Deng, "An unsupervised domain adaptation method for detecting blades icing for multiple wind turbines," *Engineering Applications of Artificial Intelligence*, Vol. 138, p. 109396, Dec. 2024, <https://doi.org/10.1016/j.engappai.2024.109396>
- [37] X.-Y. Zhang, L. He, X.-K. Wang, J.-Q. Wang, and P.-F. Cheng, "Transfer fault diagnosis based on local maximum mean difference and K-means," *Computers and Industrial Engineering*, Vol. 172, p. 108568, Oct. 2022, <https://doi.org/10.1016/j.cie.2022.108568>



Ping Zhang received Bachelor's degree in Huanghe Science and Technology College, Zhengzhou, China, in 2007. Associate Professor. Now she works at Xinxiang Vocational and Technical College. Her current research interests include composite materials, mechanical equipment control technology and fault diagnosis.



Debo Liu received Master degree in Zhengzhou University, Zhengzhou, China, in 2015. Associate Professor. Now she works at Huanghe Science and Technology College. He current research interests include Control Engineering Technology and fault diagnosis.