

Research Article

A Novel Method for Diagnosis of Bearing Fault Using Hierarchical Multitasks Convolutional Neural Networks

Yong-Zhi Liu ¹, Yi-Sheng Zou ¹, Yu-Liang Jiang,¹ Hui Yu ² and Guo-Fu Ding¹

¹School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China

²School of Creative Technologies, University of Portsmouth, Portsmouth PO1 2DJ, UK

Correspondence should be addressed to Yi-Sheng Zou; zysapple@swjtu.edu.cn

Received 3 June 2020; Revised 14 October 2020; Accepted 24 October 2020; Published 4 November 2020

Academic Editor: M. Z. Naser

Copyright © 2020 Yong-Zhi Liu 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.

Intelligent mechanical fault diagnosis has developed very fast in recent years due to the advancement and application of deep learning technologies. Thus, there are many deep learning network models that have been explored in fault classification and diagnosis. However, there are still limitations in research on the relationship between fault location, fault type, and fault severity. In this paper, a novel method for diagnosis of bearing fault using hierarchical multitask convolution neural networks (HMCNNs) is proposed, taking into account the mentioned relationships. The HMCNN model includes a main task and multiple subtasks. In the HMCNN model, a weighted probability is used to reduce the classification error propagation among multitasks to improve the fault diagnosis accuracy. The validity of the proposed method is verified on bearing datasets. Experimental results show that the proposed method is very effective and superior to the existing methods.

1. Introduction

Rolling bearings, as the key parts of mechanical equipment, are widely used in rail transit equipment, construction machinery, precision machine tools, instrumentation, and other fields. According to statistics, about 40% of rotating machinery faults are caused by bearing faults. Once bearing faults occur, they will seriously affect the normal operation of equipment, and they may even cause accidents and economic losses. Therefore, it is necessary to diagnose and monitor bearing faults before anything goes wrong [1, 2]. At present, bearing fault diagnosis is usually based on data-driven methods. By collecting motor current signals or bearing vibration signals, fault diagnosis methods are applied to complete fault identification [3, 4].

Data-driven fault diagnosis generally includes two steps: fault feature extraction and fault classification. The common methods of feature extraction include Fast Fourier Transformation (FFT) [5], Wavelet Transform (WT) [6], Empirical Mode Decomposition (EMD) [7], Local Mean Decomposition (LMD) [8], and Variational Mode

Decomposition (VMD) [9]. The common fault classification algorithms include support vector machine (SVM) [10], BP neural networks [11], Bayesian classifier [12], K-Nearest Neighbor (KNN) [13], Random Forest (RF) [14], and Classification and Regression Tree (CART) [15]. Seryasat et al. [16] presented a diagnosis method based on wavelet transform and FFT to extract energy and root mean square of different frequency bands, which could accurately and effectively identify bearing faults. Yan and Jia [17] proposed a multidomain feature classification algorithm based on optimized SVM, which included three stages: multidomain feature extraction, feature selection, and fault recognition. The algorithm has high diagnostic accuracy for rolling bearings under different working conditions. Zhang et al. [18] proposed a new method of rolling bearing fault diagnosis based on Variational Mode Decomposition and compared the performance of VMD and EMD in extracting bearing defect features from rolling bearing simulation signals. The VMD method can accurately extract the main mode of bearing fault signal and is superior to EMD in bearing defect feature extraction. Liu et al. [19] presented a fault diagnosis method for wind turbine bearings based on

integral extension local mean decomposition (IELMD), which could effectively process nonstationary signals. Kankar et al. [20] extracted the statistical characteristics of wavelet coefficients and completed the classification of bearing faults combined with an artificial neural network. Jiang et al. [21] proposed a fault diagnosis method for rolling bearings based on high-order cumulant and BP neural network in view of the fact that the vibration signals of rolling bearings were susceptible to the influence of Gaussian noise.

The basic steps of these traditional fault diagnosis methods can be summarized as follows: acquiring fault signals, analyzing the characteristics of fault signals, extracting appropriate features, and selecting appropriate classifiers according to the specific diagnosis problems. This process requires high professional knowledge and experience of signal processing for fault diagnosis personnel. With the development of modern industry, fault monitoring equipment obtains a large amount of data, and the data types are diverse, which brings great challenges to traditional fault diagnosis methods.

Deep learning has been widely used in recent years. The characteristic of the deep learning method is that it can automatically complete the task of feature extraction and classification [22]. Deep learning has also been introduced into the field of mechanical fault diagnosis to overcome the shortcomings of traditional methods recently. Zhang et al. [23–25] proposed a fault diagnosis method for rolling bearings based on deep convolution neural networks (CNN), which avoided manual feature extraction and realized automatic feature learning. Shao et al. [26] proposed an enhanced depth feature fusion method for fault diagnosis of rotating machinery. A new depth autoencoder method was constructed by combining Denoising Autoencoder (DAE) with Contractive Autoencoder (CAE) to improve the learning ability of features. Jia et al. [27] proposed a deep normalized convolution neural networks (DNCNN), which could effectively deal with unbalanced classification problems. Liu et al. [28] proposed an unsupervised fault diagnosis method for rolling bearings based on the generative adversarial networks. This method has higher generalization accuracy under noisy and varied workload situations. These deep learning methods have been successfully applied to bearing fault diagnosis from different perspectives and application scenarios. Compared with traditional methods, they have higher diagnostic accuracy. However, the problem of low generalization ability of deep neural network model remains unsolved.

The fault diagnosis of bearing includes fault location, fault type, and fault degree. In the existing fault diagnosis methods, all kinds of samples are generally used as training samples of the training model to achieve fault diagnosis, and the relationship between them and the impact on the final fault diagnosis results are less considered. For the hierarchical classification of deep learning, Yan et al. [29] first proposed the hierarchical deep convolution neural network model to classify images. It first classifies the easily separated classes roughly and then classifies them at a fine level [28]. Based on this idea, Guo et al. [30] and Qu et al. [31] proposed

a hierarchical intelligent fault diagnosis algorithm based on an adaptive deep convolution neural network model (ADCNN), which classified bearing fault location first and then classified fault severity. The design of this hierarchical classification model requires training multiple CNN recognition models, taking pretraining, and fine-tuning. It can thus lead to a cumbersome training process, more training samples, interlayer error propagation, and difficulty in model level expansion.

Therefore, a hierarchical multitask bearing fault diagnosis method based on the deep convolution neural networks is proposed in this paper. By adding multilearning tasks to the convolution neural network, the multitask learning of bearing fault diagnosis is realized, and the generalization performance of the proposed model is improved. The main contributions of this paper are as follows:

- (1) Based on the CNN fault classification model, the HMCNN model is formed by adding several related classification tasks representing different dimension information for parallel auxiliary diagnosis. In the proposed model, final classification results are obtained by fusing the classification results of the main task and subtasks of different dimensions according to the weight obtained by training, which can reduce the interlayer error propagation of tasks.
- (2) The proposed model can extract more valuable features from fewer training samples for fault classification and improve the classification accuracy of the model. In the proposed model, multiple tasks can share network parameters and information, and only one network structure needs to be trained that reduced computational consumption and training complexity. The parallel structure of multiple tasks also has good scalability.

The remainder of the paper is structured as follows. The structure of the hierarchical multitask convolutional neural network (HMCNN) and some main techniques used in HMCNN is introduced in Section 2. In Section 3, experiments are carried out to prove that HMCNN has better performance than traditional intelligent methods and some typical deep learning models. After that, the structure of HMCNN model is extended to demonstrate its ability to extend. Lastly, the diagnostic results of HMCNN model are compared with CNN model and analyzed visually to explore its mechanism. The conclusion of this paper is presented in Section 4.

2. Proposed Method

In this paper, a novel method called hierarchical multitask convolutional neural network (HMCNN) is proposed for the intelligent fault diagnosis of bearings. The proposed model includes four parts: CNN model, hierarchical classifiers, multitask learning, and hierarchical multitask convolutional neural network. The HMCNN model only needs to train a network model to realize multitask classification, in which the sharing layer can reduce the number of network parameters, thereby reducing the computational load. The

multitask learning, hierarchical classification, and joint classification layer design in the HMCNN model can improve network generalization ability. More details are described in the following sections.

2.1. Introduction to CNN Model. Convolutional neural network (CNN) is a kind of feed-forward multistage neural network. It mainly contains three kinds of layers: convolutional layer, pooling layer, and fully connected layer. The convolution layer is designed to extract different features of input data. The pooling layer following the convolutional layer is to reduce the parameters of the network through extracting the local mean or maximum value of input data. A fully connected layer is usually built in the last part of the hidden layer of the convolutional neural network. Its main function is to connect all features and send the output value to the classifier. Convolutional neural network (CNN) is one of the common deep learning models. It is used to extract features and classify vibration signals in bearing fault diagnosis.

2.2. Hierarchical Classification. The main idea of hierarchical fault diagnosis based on convolutional neural network is proposed in this paper, as shown in Figure 1. The main structure of hierarchical classification includes the sharing layer, coarse classification layer, fine classification layer, and joint classification layer, in which the sharing layer can reduce the number of network parameters and thus reduce the computation. The coarse classification layer is mainly used for coarse classification of bearing fault location, such as bearing recognition as health, inner ring fault, and outer ring fault. The fine classification is achieved through the fine classification layer. The joint classification layer which receives fine classification results as well as coarse classification results produces a weighted probability as the final classification results; it can be described as follows:

$$p(x_j) = \sum_{j=1}^N \frac{P_c(x_j)}{\sum_{i=1}^N P_c(x_i)} p_F(x_j), \quad (1)$$

where $p_c(x_j)$ is the probability of coarse classification made by the coarse classification layer. $p_F(x_j)$ is the fine classification made by the fine classification layer. N is the number of hierarchical tasks.

The coarse classification layer, fine classification layer, and joint classification layer are all based on softmax classification function for the final classification tasks. In this way, the output of the network is transformed into a probability distribution; the softmax function is described as follows:

$$p_{c,F}(x_j) = \text{softmax}(x_j) = \frac{e^{x_j}}{\sum_{k=1}^n e^{x_k}}, \quad (2)$$

where z_j is the logits of the j th output. n is the number of categories.

2.3. Multitask Learning. In this paper, bearing fault diagnosis has multiple learning tasks, such as fault location, fault type, and fault severity. From the perspective of machine learning, multitask learning can be regarded as inductive transfer learning, which can improve the learning performance of the model by using multiple related tasks, including improving generalization accuracy, learning speed, and comprehensibility of the learning model. In this paper, the related learning tasks are fault or fault location, which helps the final classification of severity tasks of fault categories. In the training process of the model, the joint training method is adopted, which combines the loss functions of multiple tasks and carries out the optimization training together. The loss function is described as

$$\begin{aligned} \text{loss} &= \sum_{i=1}^N k_i \text{loss}_i, \\ \text{loss}_i &= -\frac{1}{m} \sum_{k=1}^m \sum_j p_k^j \log q_k^j, \end{aligned} \quad (3)$$

where k_i is the coefficient, loss_i is the loss function for each task, N is the number of hierarchical tasks, m is the size of training minibatch, p_k^j is the true predicted output value, and q_k^j is the one-hot type vector with target distribution.

2.4. Hierarchical Multitask Convolutional Neural Network (HMCNN) Model. The HMCNN model is shown in Figure 1, which consists of the following four parts.

The first part is the sharing layer, which consists of two modules. The two modules are composed of 2 convolution layers and 1 pooling layer. The 3×1 size convolution kernel with stride of 2 is used in all convolutional layers. For pooling layers, the 8×1 sized max-pooling with the stride of 8 is done.

The second part is the coarse classification layer, which is connected to the shared layer. It consists of 2 full connection layers and 1 softmax layer, which are used to complete the classification of bearing fault location, fault type, and fault severity.

The third part is the fine classification layer, which is connected to the shared layer and consists of 3 convolution layers, 1 pooling layer, 2 full connection layers, and 1 softmax layer. It is used to complete the fine classification of bearing fault.

The fourth part is the joint classification layer, which receives fine classification results as well as coarse classification results and produces a weighted probability as the final classification results. The HMCNN model training parameters are shown in Table 1.

3. Experimental Verification

3.1. Data Description. Experimental data were collected from the bearing test rig of Paderborn University in Germany [32]. The experimental data were obtained by the test rig for condition monitoring of rolling bearings. The test rig consisted of several modules: an electric motor (1), a torque-

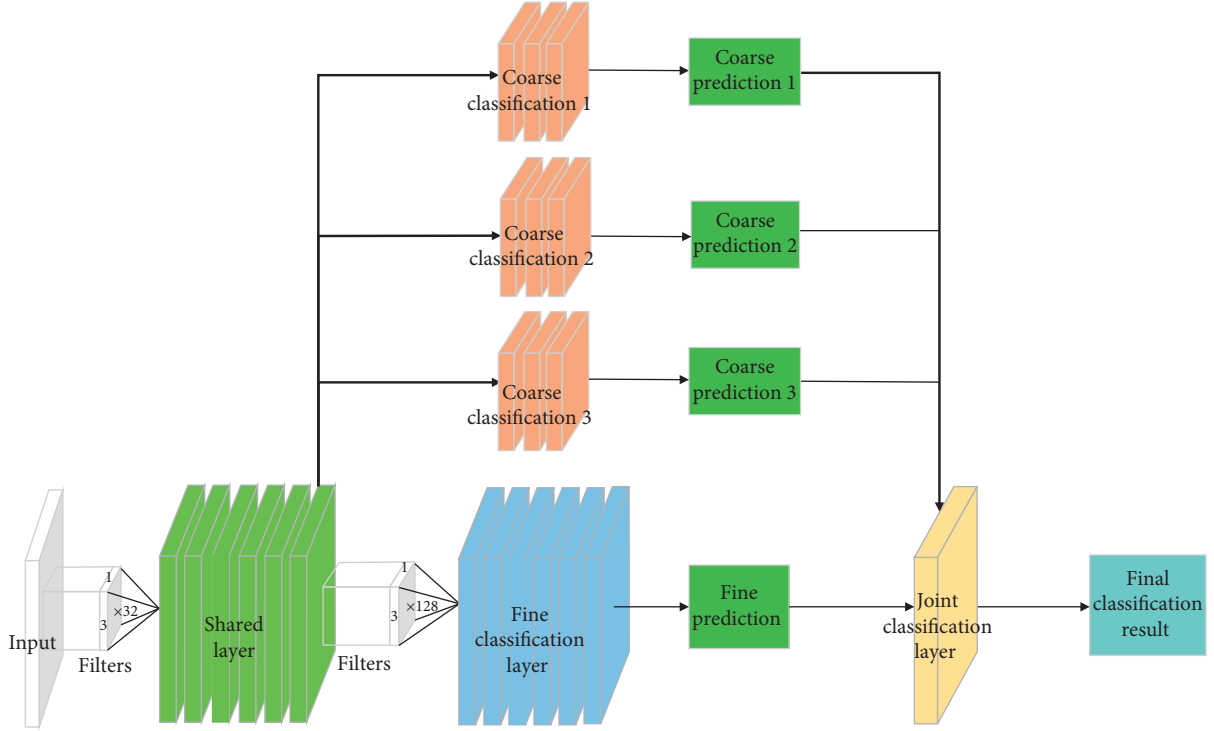


FIGURE 1: Hierarchical multitasks convolutional neural network (HMCNN) model.

TABLE 1: The HMCNN training parameters.

Learning rate	Training steps	Batch size	Optimizer
0.001	20000	256	Adam

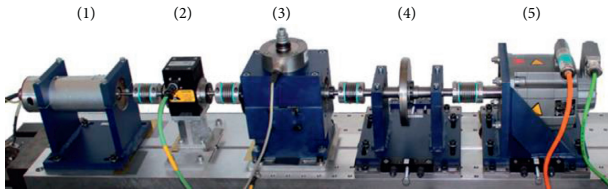


FIGURE 2: Test rig for condition monitoring of rolling bearings.

measurement shaft (2), a rolling bearing test module (3), a flywheel (4), and a load motor (5), as shown in Figure 2.

The test bearing was ball bearings of type 6203. Bearings are run at a rotational speed of 900 rpm with a load torque of 0.7 Nm and a radial force on the bearing of 1000 N. The frequency of the data acquisition system is 64 kHz. The bearing temperature was kept roughly at 45–50°C. Three kinds of bearing states are used in this experiment: inner ring damage, outer ring damage, and healthy. The detailed situation of data is shown in Table 2. In Table 2, the bearing fault location, damage method, and fault severity are listed. For fault location, H is the bearing with no fault, IR is the bearing with an inner race fault, and OR is the bearing with an outer race fault. The damage methods of bearing are shown in Figure 3. The bearing damage used in this paper was caused by three different methods: electric discharge

machining (trench of 0.25 mm length in rolling direction and depth of 1-2 mm), drilling (diameter: 0.9 mm, 2 mm, and 3 mm), and manual electric engraving (damage length from 1 to 4 mm).

As described in the document of the dataset, each bearing acquired 20 original vibration time-series signals, each of which recorded about 256,000 data points. In this experiment, the 2048 data points were used to construct a sample. For each health condition of bearings, 2000 samples were used in the training set and 500 samples were used in the test set. The vibration signals of each health state are shown in Figure 4. The experimental data are normalized by maximum and minimum normalization, and the normalization formula is as follows:

$$x^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad (4)$$

where x_i is the value of the i -th point of sample data. x_{\max} is the maximum value of sample data. x_{\min} is the minimum value of sample data.

In this paper, the proposed model is based on tensorflow deep learning framework. The experiment was completed on a computer with CPU i7 8700, 16 GB memory, and NVIDIA GTX 1070 GPU.

3.2. Diagnosis Results of HMCNN. The experiments are divided into three parts. The first part is a comparison among the proposed model, traditional method, and intelligent algorithm based on deep learning to demonstrate the superiority of the proposed model in terms of generalization performance. The second part is to extend and compare the

TABLE 2: Situation of damaged bearings.

Bearing code	K001	KA03	KA07	KA08	KI01	KI03	KI07
Fault location	H	OR	OR	OR	IR	IR	IR
Damage method	No fault	Electric engrave	Drilling	Drilling	EDM	Electric engraver	Electric engraver
Fault severity	0	2	1	2	1	1	2

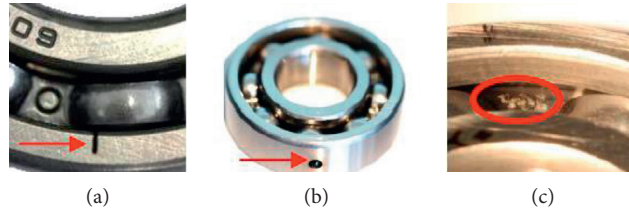


FIGURE 3: The methods of bearing damage. (a) EDM, (b) drilling, and (c) electric engraver.

network structure of the proposed model to verify the effectiveness of multitask learning. The third part is a comparison between the proposed model (HMCNN) and ADCNN model, studies the propagation of errors between model levels, and analyzes the reasons for the high recognition accuracy of the proposed model.

3.2.1. The Comparative Analysis of HMCNN Model with Other Models. In the first part, the proposed model (HMCNN) is compared with the commonly used fault diagnosis models such as support vector machine (SVM), backpropagation neural networks (BPNN), convolutional neural networks (CNN), and long short-term memory (LSTM). The inputs of SVM and BPNN are multidimensional features extracted after ensemble empirical mode decomposition. The inputs of CNN, LSTM, and HMCNN methods are normalized original signals. Experimental comparison results are shown in Figure 4. From Figure 5, it can be seen that the classification accuracy of traditional models such as SVM and DNN is less than 80%, while the classification accuracy of deep learning model such as CNN and LSTM is about 90%. The bearing fault diagnosis accuracy of HMCNN model reaches 99.7%. Compared with traditional bearing fault diagnosis models, it does not need to extract features, has higher diagnosis accuracy, and has better generalization ability than the current deep learning models.

The HMCNN model and CNN model are also compared and analyzed. The CNN and HMCNN models use the same optimization method and training parameters in this paper. The accuracy of per 50 steps in HMCNN and CNN model is shown in Figure 6. Figure 6 shows that the bearing fault diagnosis accuracy of HMCNN model is 99.7% and that of CNN model is 92.1%. Compared with CNN model, HMCNN not only has higher bearing fault diagnosis accuracy but also has fewer training steps to achieve the highest diagnosis accuracy. The confusion matrix of experimental results is compared as shown in Figure 7. From Figure 7, compared with CNN model, HMCNN model mainly reduces the confusion degree between outer ring bearing faults, so as to improve the diagnosis accuracy of bearing faults.

In the training processes, the learning speed of HMCNN model and CNN model is compared, as shown in Figure 8. The convergence rate (learning speed) of HMCNN model is about twice that of CNN model.

In order to further verify the generalization performance of HMCNN model, the diagnosis accuracy of HMCNN with other models under different training sets is also compared here. The comparison results are shown in Table 3 and Figure 9. As the number of training samples decreases, the recognition accuracy of all methods decreases to varying degrees. From Table 4, it can be seen that the HMCNN model has better recognition accuracy in fewer training sets, and the recognition accuracy can reach 96.7% in the case of only 500 training samples.

In addition, the performance differences between HMCNN model and SVM, BPNN, CNN, and LSTM model under noise are compared. Comparison results are shown in Table 3 and Figure 10. It can be seen that the diagnosis accuracy of HMCNN is 99.1% in noise environment (SNR = 10 dB), while the diagnosis accuracy of other models is not more than 90%. At the same time, the diagnosis accuracy of HMCNN model in noise environment (SNR = -2 dB) is more than 90%. So HMCNN has good antinoise performance.

3.2.2. The Comparative Analysis of HMCNN with Different Tasks Numbers. In the second part, we study and analyze the relationship between the hierarchical tasks' number of HMCNN model and its diagnosis accuracy. According to the dataset, the HMCNN model is used to learn one, two, and three classification tasks (bearing fault location, fault type, and fault severity), which are named HMCNN1, HMCNN2, and HMCNN3, respectively. The comparison results are shown in Figures 11 and 12. The results show that both HMCNN3 and HMCNN2 model can achieve a high diagnosis accuracy. The accuracy of HMCNN2 model is 0.8% higher than that of HMCNN1 model. Compared with CNN, HMCNN contributes more to the accuracy improvement by adding the task of bearing fault location. In HMCNN3 model, the diagnosis accuracy of fault type reaches 99.7%, which shows that the task of bearing fault position, fault type, and fault severity diagnosis is effective and feasible, and the final diagnosis accuracy is improved to a certain extent.

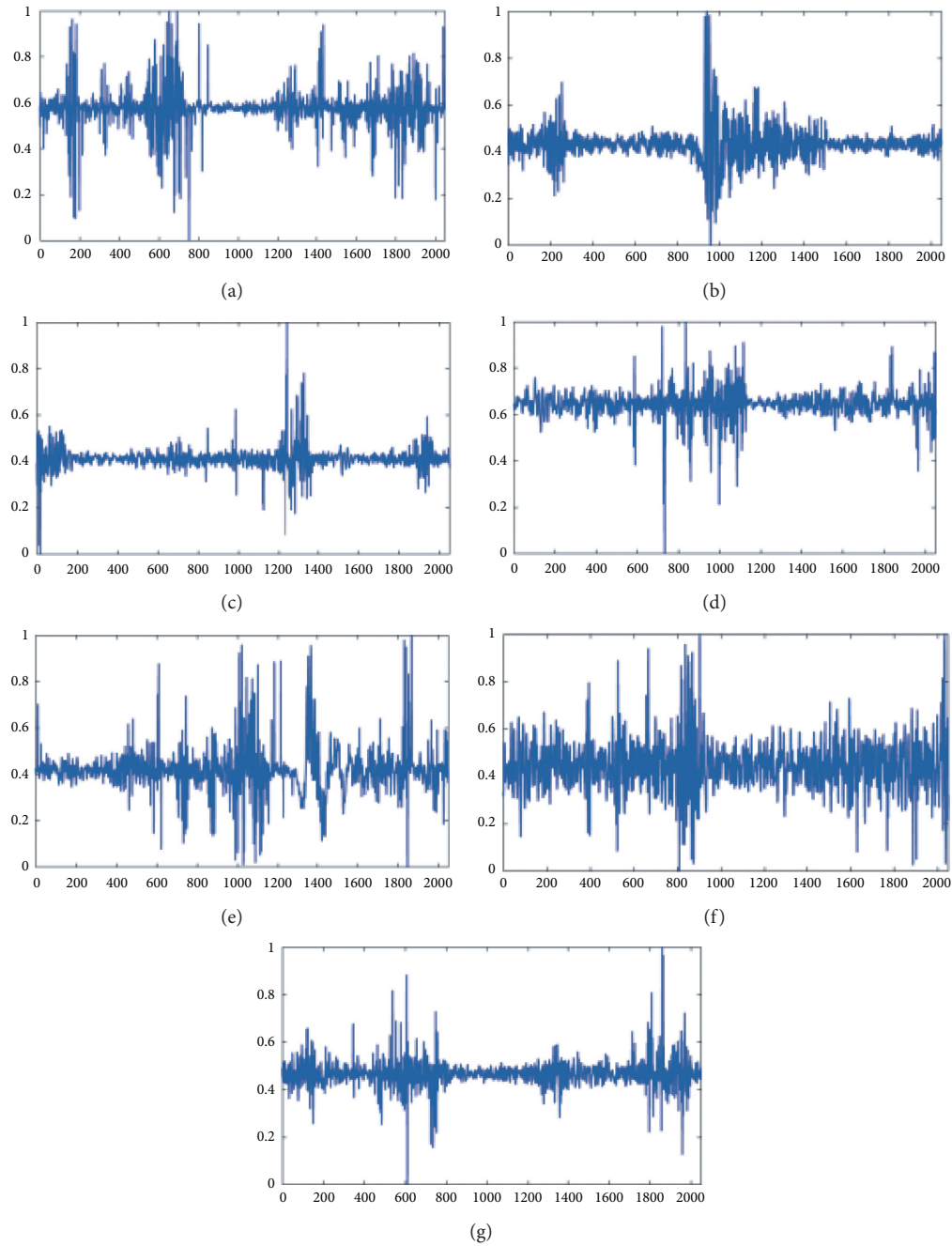


FIGURE 4: Vibration signals for each health condition of bearings. (a) K001, (b) KA03, (c) KA07, (d) KA08, (e) KI01, (f) KI03, and (g) KI07.

The comparison between HMCNN1, HMCNN2, and HMCNN3 models proves that the proposed model can be extended to diagnosis multiple tasks. In practical application, the location, type, and severity of bearing fault can be output in HMCNN model, which provides more detailed guidance for fault maintenance.

3.2.3. The Comparative Analysis of HMCNN Model with ADCNN Model. In the third part, the ADCNN model proposed by Guo et al. [30] is compared with HMCNN model. The idea of ADCNN model for bearing fault diagnosis is to identify

the location of bearing fault first and then identify the fault severity of each location of bearing on this basis. Accuracy results of ADCNN and HMCNN models for diagnosis bearing fault locations and final fault severity are shown in Tables 5 and 6, respectively. We can see that the ADCNN model's hierarchical diagnosis of the bearing will make the error of bearing fault location diagnosis spread to the result of bearing fault severity diagnosis, and the more the number of tasks, the more serious the error propagation. The HMCNN model has a shared layer and a weighted joint classification layer, which can solve the problem of error propagation and make the model more scalable.

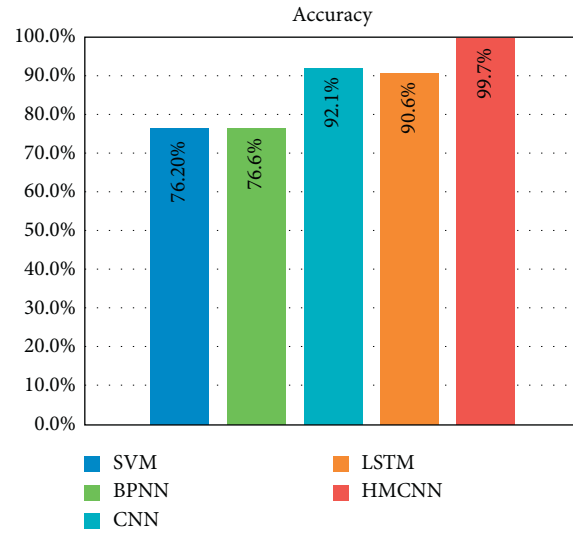


FIGURE 5: Accuracy of SVM, BPNN, CNN, LSTM, and HMCNN.

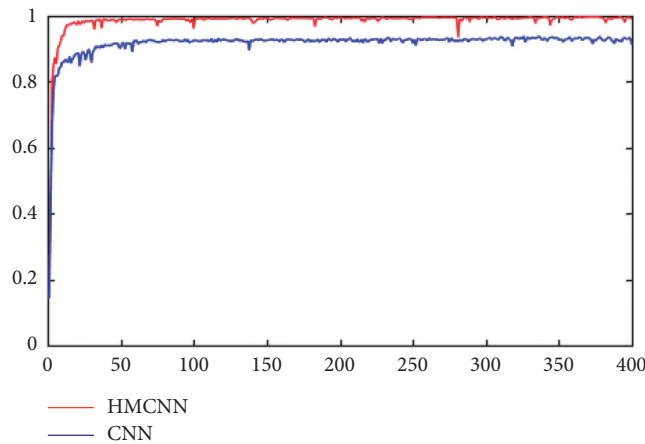


FIGURE 6: Accuracy per 50 steps in HMCNN and CNN model.

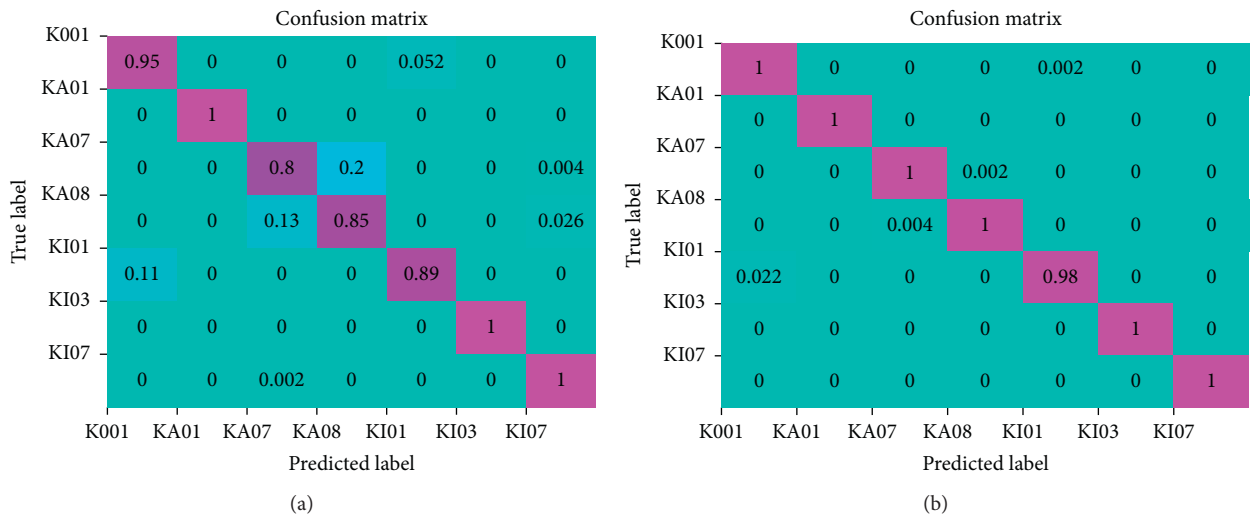


FIGURE 7: Confusion matrix of (a) CNN and (b) HMCNN.

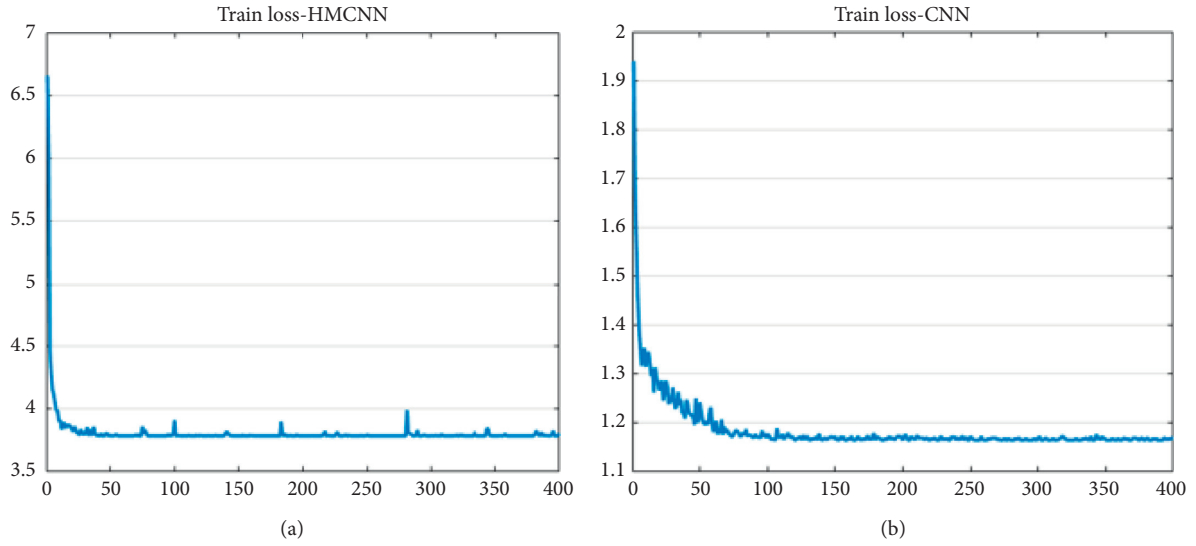


FIGURE 8: Train loss of (a) HMCNN and (b) CNN.

TABLE 3: Comparison of diagnosis accuracy with other models in different noise environments.

Methods	SNR							
	-4	-2	0	2	4	6	8	10
SVM	0.663	0.724	0.735	0.761	0.812	0.822	0.823	0.826
BPNN	0.601	0.623	0.656	0.683	0.702	0.832	0.845	0.854
CNN	0.651	0.687	0.787	0.811	0.832	0.856	0.864	0.886
LSTM	0.652	0.663	0.752	0.787	0.822	0.834	0.850	0.852
HMCNN	0.845	0.916	0.932	0.950	0.972	0.983	0.985	0.991

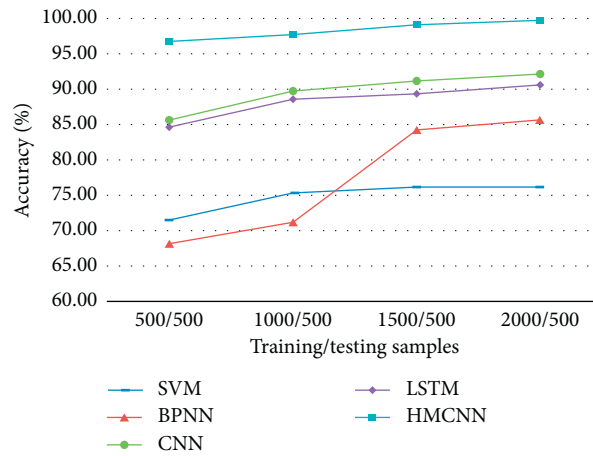


FIGURE 9: Performance comparison of accuracy metric between other models and the proposed HMCNN model on bearing fault dataset.

3.3. Visualization Analysis. The principle of HMCNN model is further analyzed by t-SNE visualization technology. In this paper, the test set data are used as input, and the output data of the pooling layer in the HMCNN model are extracted as output. These output data are reduced to two-dimensional feature vectors by t-SNE, and then these outputs are plotted as scatter plots, representing their classes with different colors, as shown in Figure 13. The visualization results show

that, with the increase of network layers of HMCNN model, the separation degree of features extracted from original signals becomes more and more obvious. At last, the output features of softmax classifier have seven distinct distributions (the final classification of bearing faults).

By comparing the output of two identical pooling layers of HMCNN and CNN models as shown in Figure 14, it can be seen that HMCNN has a better classification effect than CNN

TABLE 4: Performance comparison of accuracy metric between other models and the proposed HMCNN model on bearing fault dataset.

Case	Number training samples	Number testing samples	Accuracy				
			SVM	BPNN	CNN	LSTM	HMCNN
1	500	500	0.716	0.682	0.856	0.846	0.967
2	1000	500	0.753	0.712	0.897	0.886	0.978
3	1500	500	0.761	0.842	0.912	0.893	0.991
4	2000	500	0.762	0.856	0.921	0.906	0.997

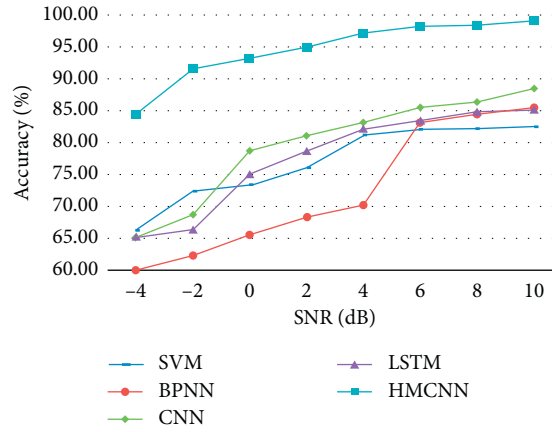


FIGURE 10: Comparison of diagnosis accuracy with other models in different noise environments.

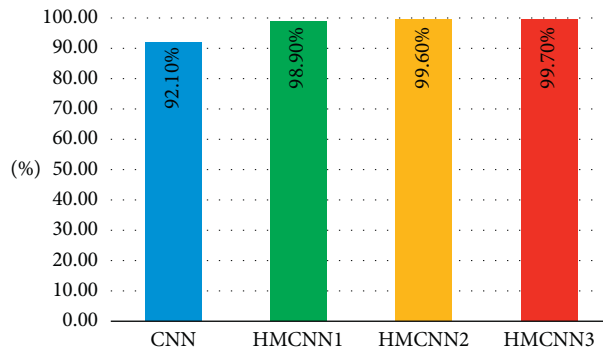


FIGURE 11: Accuracy of CNN, HMCNN1, HMCNN2, and HMCNN3.

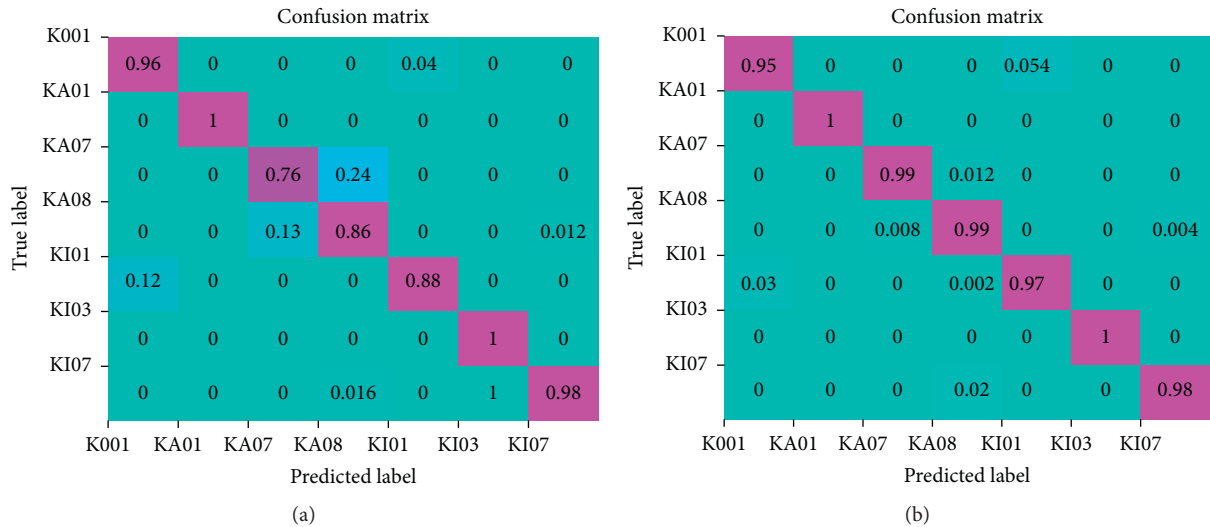


FIGURE 12: Continued.

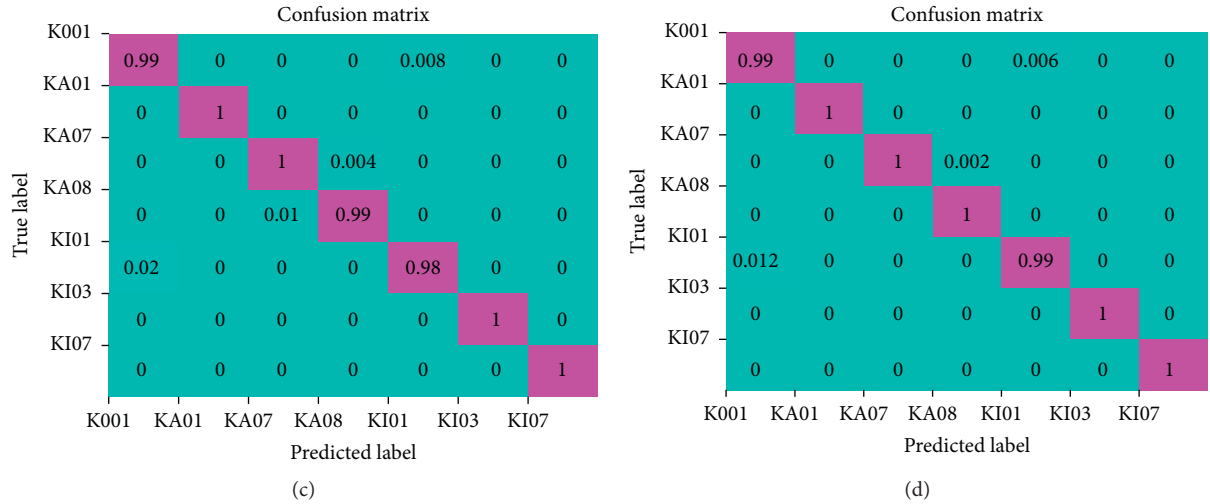


FIGURE 12: Confusion matrix of (a) CNN, (b) HMCNN1, (c) HMCNN2, and (d) HMCNN3.

TABLE 5: The accuracy of fault location for ADCNN and HMCNN model.

Case	ADCNN (%)	HMCNN (%)
Inner race fault	81	98
Outer race fault	92	100
Health	86	100
Overall accuracy	86.3	99.4

TABLE 6: The accuracy of fault severity for ADCNN and HMCNN model.

Case	ADCNN (%)	HMCNN (%)
Fault severity accuracy	81.7	99.7

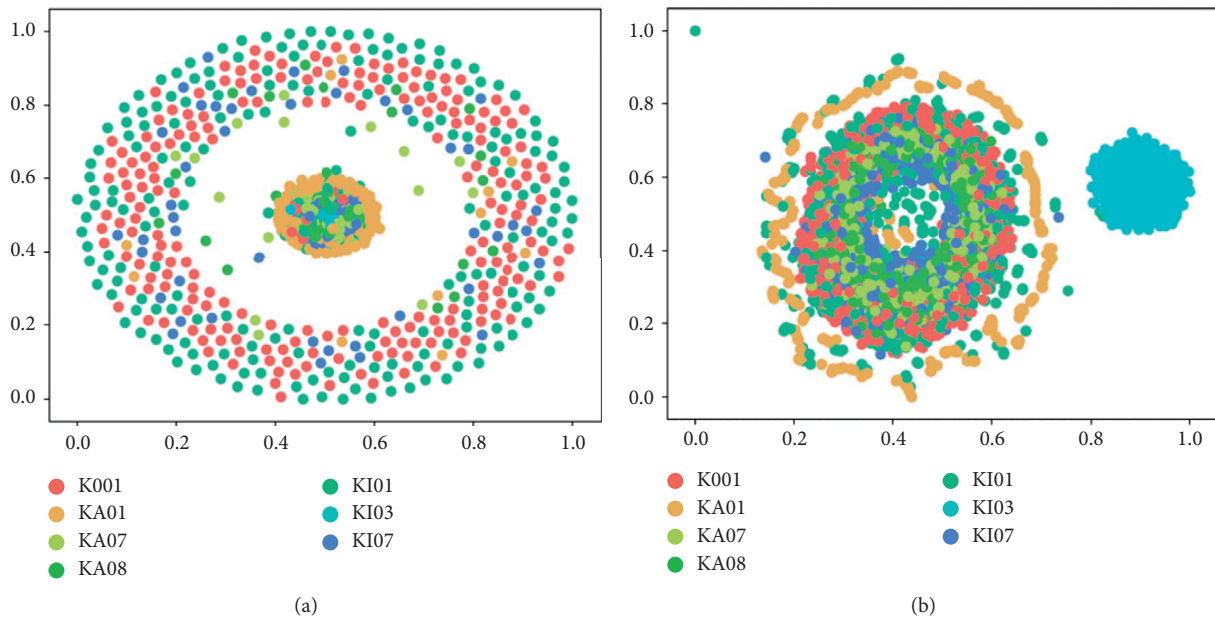


FIGURE 13: Continued.

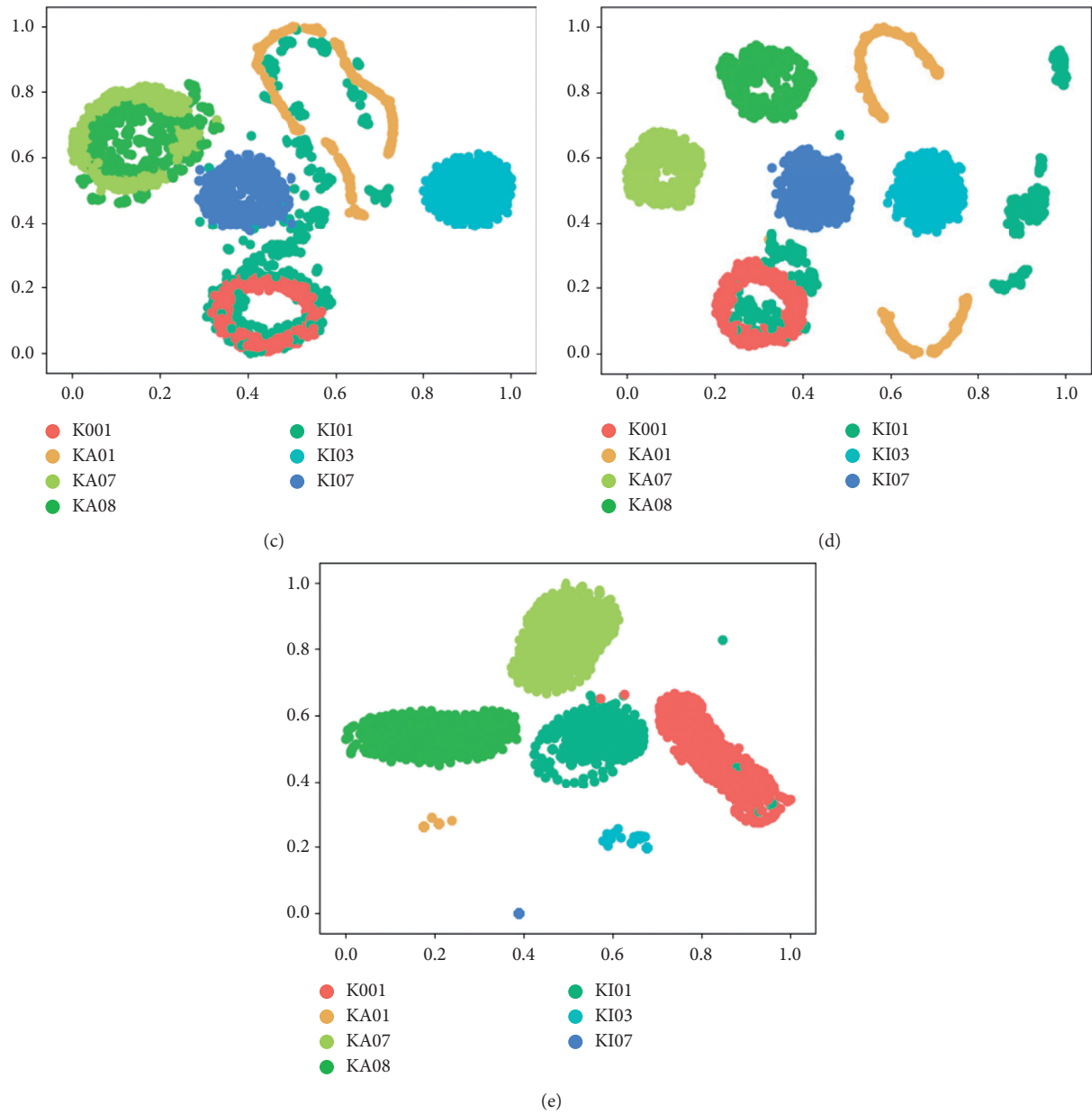


FIGURE 13: The visualization of learned features for HMCNN: (a) the input layer, (b) the first pooling layer, (c) the second pooling layer, (d) the third pooling layer, and (e) the softmax layer.

model. Compared with the second pool layer of HMCNN and CNN model, the classification of KI07 (Inner Race Fault 2) bearings, KA07 (Outer Race Fault 1) bearings, and KA08 (Outer Race Fault 2) bearings by CNN model is not obvious, but HMCNN model has been able to separate KI07 bearings obviously. This shows that, in the second pooling layer of HMCNN model, the fault location of bearing is well classified. The visual output of HMCNN and CNN third pool layer shows that CNN model has no obvious effect on the diagnosis of KA07 and KA08 bearings, while HMCNN model has an obvious effect on the diagnosis of KA07 and KA08 bearings. This proves that, in the third pooling layer of HMCNN model, the bearing fault severity is clearly classified. From the network of HMCNN, the second pooling layer is essentially the last layer of shared layer.

In the shared layer, the proposed model learns the features of bearing fault location, fault type, and fault severity.

HMCNN model can share more fault information than CNN model through multitasks learning. In particular, the task of fault location focuses the network attention on the possible neglected fault location information, which enhances the classification ability of bearing fault location of HMCNN model and the KI07 bearing can be separated obviously by HMCNN model in the second pool layer. After shared layer, the HMCNN model only needs to recognize KA07 and KA08 bearings, but CNN model also needs to recognize KI07, KA07, and KA08 bearings, which may lead to the final recognition accuracy of the CNN model which is lower than that of the HMCNN model.

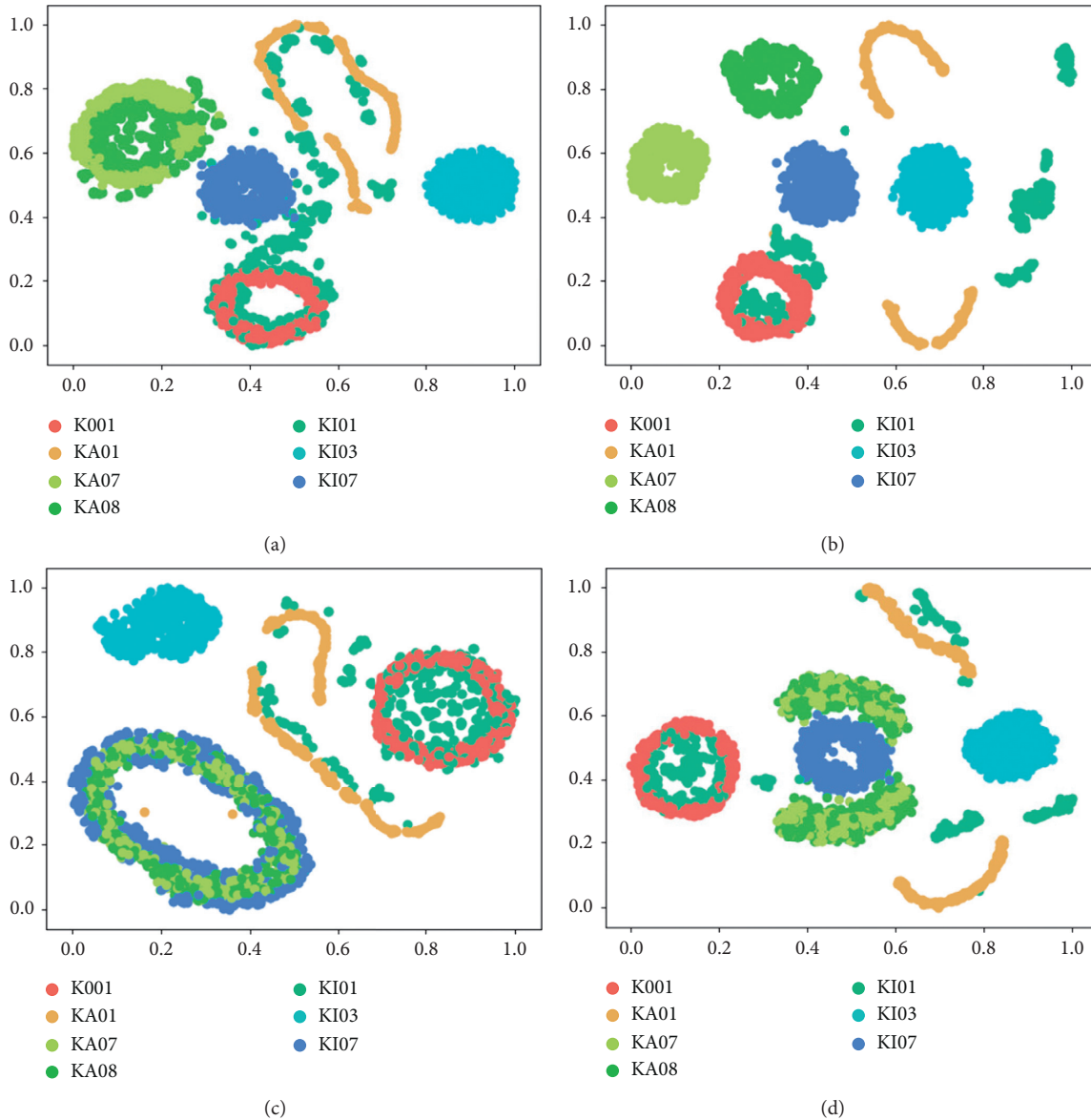


FIGURE 14: The visualization of learned features for CNN and HMCNN: (a) the second pooling layer of HMCNN, (b) the third pooling layer of CNN, (c) the second pooling layer of HMCNN, and (d) the third pooling layer of CNN.

From the analysis of the learning process of HMCNN model, multitask learning in HMCNN model may improve the generalization accuracy of the model by using the information hidden in training signals of multiple tasks as an inductive bias. Multitask learning plays the same role as regularization and reduces the risk of model overfitting. At the same time, it reduces the ability to fit random noise and makes the model have better generalization performance.

4. Conclusions

The hierarchical multitask learning CNN model (HMCNN) is proposed, which reflects hierarchical classification. Only one model needs to be trained to achieve a multitask classification. Compared with the experimental results of other models, the HMCNN model can improve the accuracy

of the final fault diagnosis, and the diagnosis accuracy reaches 99.7%. Compared with CNN model, the HMCNN model has a faster learning speed. We compare the diagnosis accuracy of HMCNN with other models in different training samples and noise environments. HMCNN model has better diagnosis accuracy than other models. It is proposed that the HMCNN model can be extended to diagnosis multiple tasks. The fault location, fault type, and fault severity of bearing fault diagnosis are given, which can provide more detailed guidance for fault maintenance. Compared with the ADCNN model, the HMCNN model solves the problem of error propagation and makes the model scalable.

Through the visual analysis of the HMCNN and CNN model learning process, the reason why HMCNN has higher generalization accuracy is further explored. HMCNN model shares more fault information than the CNN model. In

particular, the task of fault location focuses the network attention on the possible neglected fault position information, which enhances the classification ability to bear fault location of the HMCNN model.

Data Availability

The data that support the findings of this study are available at https://mb.uni-paderborn.de/en/kat/main-research/datacenter/bearing-datacenter/data-sets-and-download/?tsourcectag=s_pcqq_aiomsg. At the same time, the data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by National Key R&D Program of China (2017YFB1201201).

References

- [1] M. Cerrada, R. V. Sánchez, C. Li et al., "A review on data-driven fault severity assessment in rolling bearings," *Mechanical Systems and Signal Processing*, vol. 99, pp. 169–196, 2018.
- [2] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1–2, pp. 108–126, 2013.
- [3] B. Muruganatham, M. A. Sanjith, S. A. V. Satya Murty, and S. S. Murty, "Roller element bearing fault diagnosis using singular spectrum analysis," *Mechanical Systems and Signal Processing*, vol. 35, no. 1–2, pp. 150–166, 2013.
- [4] S. S. Refaat, H. Abu-Rub, M. S. Saad, E. M. Aboul-Zahab, and A. Iqbal, "ANN-based for detection, diagnosis the bearing fault for three phase induction motors using current signal," in *Proceedings of the 2013 IEEE International Conference on Industrial Technology (ICIT)*, IEEE, Cape Town, South Africa, February 2013.
- [5] V. K. Rai and A. R. Mohanty, "Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert-Huang transform," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2607–2615, 2007.
- [6] X. Lou and K. A. Loparo, "Bearing fault diagnosis based on wavelet transform and fuzzy Inference," *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 1077–1095, 2004.
- [7] Y. Yang, D. J. Yu, and J. S. Cheng, "A roller bearing fault diagnosis method based on EMD energy entropy and ANN," *Journal of Sound and Vibration*, vol. 294, no. 1–2, pp. 269–277, 2006.
- [8] Y. Li, M. Xu, R. Wang, and W. Huang, "A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy," *Journal of Sound and Vibration*, vol. 360, pp. 277–299, 2016.
- [9] C. L. Liu, Y. J. Wu, and C. G. Zhen, "Rolling bearing fault diagnosis based on variational mode decomposition and fuzzy C means clustering," *Proceedings of the CSEE*, vol. 35, no. 13, pp. 3358–3365, 2015.
- [10] Y. Yang, D. J. Yu, and J. S. Cheng, "A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM," *Measurement*, vol. 40, no. 9–10, pp. 943–950, 2007.
- [11] Z. R. Hou, "Rolling bearing fault diagnosis based on wavelet packet and improved BP neural network for wind turbines," *Applied Mechanics and Materials*, vol. 347–350, pp. 117–120, 2013.
- [12] V. Muralidharan and V. Sugumaran, "A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis," *Applied Soft Computing*, vol. 12, no. 8, pp. 2023–2029, 2012.
- [13] F. F. Chen, M. Li, B. J. Chen et al., "Fault diagnosis of roller bearing based on hybrid feature set and weighted KNN," *Journal of Mechanical Transmission*, vol. 40, no. 8, pp. 138–143, 2016.
- [14] Z. Wang, Q. Zhang, J. Xiong, M. Xiao, G. Sun, and J. He, "Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests," *IEEE Sensors Journal*, vol. 17, no. 17, pp. 5581–5588, 2017.
- [15] M. Seera and C. P. Lim, "Online motor fault detection and diagnosis using a hybrid FMM-CART model," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 4, pp. 806–812, 2014.
- [16] O. R. Seryasat, M. A. Shoorehdeli, F. Honarvar, and A. Rahmani, "Multi-fault diagnosis of ball bearing using FFT, wavelet energy entropy mean and root mean square (RMS)," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pp. 4295–4299, Istanbul, Turkey, October 2010.
- [17] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, 2018.
- [18] M. Zhang, Z. Jiang, and K. Feng, "Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump," *Mechanical Systems and Signal Processing*, vol. 93, no. 1, pp. 460–493, 2017.
- [19] W. Y. Liu, Q. W. Gao, G. Ye, R. Ma, X. N. Lu, and J. G. Han, "A novel wind turbine bearing fault diagnosis method based on integral extension LMD," *Measurement*, vol. 74, pp. 70–77, 2015.
- [20] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Rolling element bearing fault diagnosis using wavelet transform," *Neurocomputing*, vol. 74, no. 10, pp. 1638–1645, 2011.
- [21] L. Y. Jiang, Q. Q. Li, J. G. Cui, and J. H. Xi, "Rolling bearing fault diagnosis based on higher-order cumulants and BP neural network," in *Proceedings of the 27th Chinese Control and Decision Conference*, pp. 2664–2667, Qingdao, China, May 2015.
- [22] H. Zhao, J. Zheng, W. Deng, and Y. Song, "Semi-supervised broad learning system based on manifold regularization and broad network," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 3, pp. 983–994, 2020.
- [23] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, 2017.
- [24] F. Wang, H. Jiang, H. Shao et al., "An adaptive deep convolutional neural network for rolling bearing fault diagnosis," *Measurement Science and Technology*, vol. 28, no. 9, Article ID 095005, 2017.
- [25] L. Wen, X. Li and Li. Gao, Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method,"

- IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2018.
- [26] H. Shao, H. Jiang, F. Wang, and H. Zhao, “An enhancement deep feature fusion method for rotating machinery fault diagnosis,” *Knowledge-Based Systems*, vol. 119, pp. 200–220, 2017.
 - [27] F. Jia, Y. Lei, N. Lu, and S. Xing, “Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization,” *Mechanical Systems and Signal Processing*, vol. 110, pp. 349–367, 2018.
 - [28] H. Liu, J. Zhou, Y. Xu, Y. Zheng, X. Peng, and W. Jiang, “Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks,” *Neurocomputing*, vol. 315, pp. 412–424, 2018.
 - [29] Z. Yan, H. Zhang, R. Piramuthu et al., “HD-CNN: hierarchical deep convolutional neural networks for large scale visual recognition,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2740–2748, Santiago, Chile, December 2015.
 - [30] X. Guo, L. Chen, and C. Shen, “Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis,” *Measurement*, vol. 93, pp. 490–502, 2016.
 - [31] J. L. Qu, L. Yu, T. Yuan et al., “A novel hierarchical intelligent fault diagnosis algorithm based on convolutional neural network,” *Control and Decision*, vol. 12, pp. 1–11, 2018.
 - [32] C. Lessmeier, J. K. Kimotho, D. Zimmer et al., “Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification,” in *Proceedings of the European Conference of the Prognostics and Health Management Society*, pp. 1–17, Bilbao, Spain, July 2016.