Bearing Vibration Signal Fault Diagnosis Based on LSTM-Cascade CatBoost

Miaomiao Yang, Weizhi Liu, Wenxuan Zhang, Mei Wang, Xia Fang*

School of Mechanical Engineering, Sichuan University, China
yangmiaomiao@stu.scu.edu.cn, liuweizhi@stu.scu.edu.cn, zhangwenxuan@stu.scu.edu.cn, wangmei@scu.edu.cn, fangxia@scu.edu.cn

Abstract

Bearing is one of the most concerned parts in the field of fault diagnosis. At present, there are numerous excellent algorithms applied to bearing fault detection. This paper proposes a new fault bearings diagnosis model named LSTM-Cascade CatBoost, which can directly classify bearing vibration signals in the case of multiple granularity and high dimensions without signal processing. The model is based on gcForest, whose complexity can be adjusted automatically to the size of data set and it uses LSTM to extract features of time series signals instead of multi-grained scanning for improving the model's feature extraction ability. CatBoost is used as the base classifier of cascade forest to improve the classification accuracy of the model. Experimental results show the fact that this model is highly accurate in CWRU and XJTU-SY datasets. Besides, it not only proves that the feature extraction ability of LSTM is significantly better than that of multi-grained scanning, but CatBoost as a base classifier can further improve the accuracy of cascade forest.

Keywords: Fault diagnosis, Bearing, Long Short-Term Memory, gcForest, CatBoost

1 Introduction

Bearing is one of the significant parts of rotating machinery, which is prone to failure in the case of high speed or heavy load due to its structure. And bearing failure will definitely affect the proper functioning of rotating machinery, then it will cause economic losses even endangering worker's personal safety. Therefore, it is very important to monitor the state of bearings. At present, acoustic emission signals [1-2] and vibration signals [3-5] of bearings are used regularly in fault diagnosis field. Acoustic emission detection technology is superior to vibration detection at low speed, but the cost of vibration detection system is lower. Therefore, the methods based on vibration signals are more popular [6]. To improve the accuracy of the detection device, in addition to designing a stable hardware measurement system, it is also important to design an excellent fault diagnosis algorithm to distinguish signal types.

For bearing vibration signals, the general processing method is to use fast Fourier transform, wavelet transform, Hilbert transform or other signal processing methods to extract fault features, and then complete classification [7-9]. With the rapid development of deep learning, deep neural network is gradually applied to bearing fault diagnosis. Some researchers used wavelet transform to input time-frequency diagram of bearing time-domain signal transformation into convolutional neural network for classification [10-13]. Gan eta. used deep confidence network to build a hierarchical diagnosis model then they classified the wavelet packet energy characteristics of bearing vibration signals [14]. Hao eta. used the combined model containing 1D-CNN and LSTM for vibration signals' fault diagnosis of bearing [15].

GcForest is an integrated learning method based on decision tree proposed by Zhou, and its test accuracy in data sets such as MINIST and ORL Dataset is close to CNN [16]. However, the complexity of the model can be adjusted automatically according to the size of the data set and can adapt to different sizes of the data set. These days, there are also some cases of applying gcForest in the field of bearing fault diagnosis. Qin eta. modified the cascade layer of gcForest for classifying the original vibration data of bearings [17]. Xu eta. used wavelet transform to change the bearing vibration signal into a time-frequency diagram, and combined CNN with gcForest to classify the time-frequency diagram [18]. However, gcForest also has some shortcomings The multi-grained scanning structure of this model is weaker than CNN and RNN in feature extraction of image or sequence data, which will affect the accuracy of cascade forest classification to a certain extent. Therefore, Classifiers in cascading forests still have enhanced space.

In the view of the above deficiencies, we use the LSTM [19] layer to replace the multi-grained scanning structure for feature extraction, and change the base classifier in the cascade forest in to CatBoost [20] classifier. The advantage of this model is that the original signal is processed directly without signal processing steps, which simplifies the process of fault diagnosis. LSTM layer features the sequence data and then directly transmits it to cascade CatBoost for classification. Experimental results show that LSTM has better feature extraction capability than multi-granularity scan, and cascade CatBoost can further improve the accuracy of classification. The contribution of this study is to use LSTM layer instead of multi-grained scanning structure to complete classification. The contribution of this study is to propose the LSTM-Cascade Catboost model based on gcForest, which has better ability to classify bearing timing signals.

The main structural flow of this paper is as follows: The second part introduces LSTM, gcForest and the methods proposed in this paper; The third part is the experimental

*Corresponding Author: Xia Fang; E-mail: fangxia@scu.edu.cn
DOI: 10.53106/160792642022092305024

results; The fourth part is the conclusion and prospect of our research.

2 Basic Theory and Proposed Method

2.1 LSTM

RNN (Recurrent Neural Network) has been widely used when dealing with events related to timing information. However, when the length of the input signal is different and the information header is difficult to determine, RNN will have gradient disappearance and will not be able to catch long-term dependencies. As an improvement of RNN, LSTM (Long Short-Term Memory) overcomes this problem. LSTM is widely used in fault diagnosis and prognosis [21-22].

The structure of RNN is shown in Figure 1. Unroll the loop of RNN, it can be thought of as multiple identical basic units connected to each other. x_t is the input at time t. A is the basic unit. h_t is the output at time t. The information at time $t-1$ can be transmitted through the basic unit to time t. Thus, RNN can capture the association between the data. LSTM has the same chain structure as RNN, the difference lies in the calculation rules of each basic unit.

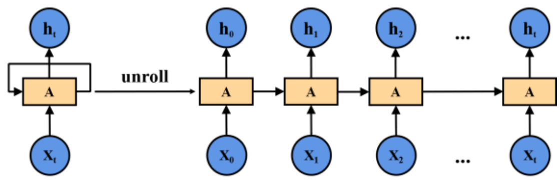


Figure 1. Unrolled of RNN

The basic units of the LSTM are shown in Figure 2. The line at the top of the diagram represents the cell state. At the bottom of the diagram is a forget gate, an input gate, and an output gate.

The forget gate f_t can selectively retain information about the h_{t-1}. σ is the sigmoid activation function and its output value is between 0 and 1. This value determines whether the information from the previous moment is retained.

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad \textbf{(1)}$$

The input gate i_t controls the input and determines how much information enters the cell state. Tanh will produce a candidate value $\tilde{C}_t$.

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \quad \textbf{(2)}$$

$$\widetilde{C}_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \quad \textbf{(3)}$$

The cell state C_t consists of the information of the forget gate and the input gate.

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \quad \textbf{(4)}$$

The output gate O_t determines the information in the cell state to be output at the current time. O_t and C_t constitute the output h_t of the basic unit.

$$O_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad \textbf{(5)}$$

where w_f, w_i, w_C, w_o are the weights of each gate. b_f, b_i, b_C, b_o are the biases of each gate respectively.

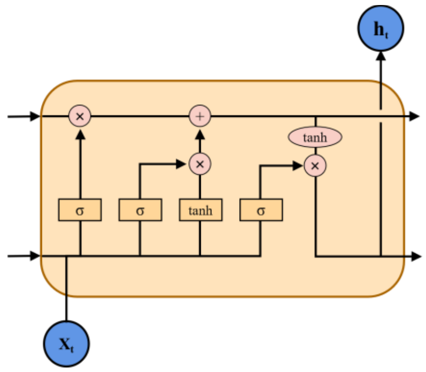


Figure 2. The structure of LSTM

2.2 GCForest

GCForest can be divided into two structures: one is Multi-Grained Scanning, which is responsible for extracting the features of the sample. The other is Cascade Forest Structure, which is responsible for classifying the high-dimensional features obtained from Multi-Grained Scanning.

2.2.1 Multi-Grained Scanning

The raw feature is segmented by the sliding window to get the feature vectors, in the Multi-Grained Scanning, then the feature vectors are input into the random forest [23] and the completely-random forest [24] to generate the class vectors, and finally all the class vectors are concatenated as the output of the Multi-Grained Scanning. As shown in Figure 3, a J-dimensional sample is fed into a Multi-Grained Scanning. Using a K-dimensional sliding window with a sliding step of λ to obtain L feature vectors, where $L = (J - K) / \lambda + 1$. Set the number of categories as x, the random forest and the completely-random tree forest generate Lx-dimensional class vectors respectively, and all class vectors are concatenated to get a $2 \cdot L \cdot x$-dimensional vector.

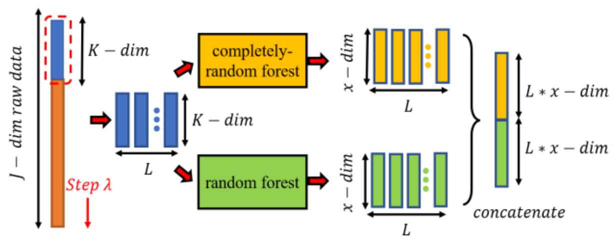


Figure 3. The structure of Multi-Grained Scanning

In the random forest, each decision tree will estimate the category distribution, and the class vector of input features can be obtained by averaging the category distribution generated by all decision trees in the forest. The random forest randomly selects $\sqrt{K}$ features from K-dimensional input vectors, calculates gini coefficient [25], evaluates the selected features, and selects the optimal nodes for splitting. The completely-random tree forest randomly selects features from input vectors for splitting until there is only one category of nodes.

2.2.2 Cascade Forest

The Cascade Forest consists of multiple random forests and completely-random tree forest. The feature vectors are processed layer-by-layer in cascade layer to get the final category prediction. As shown in Figure 4, if a cascade layer has two random forests and two completely-random tree forest, then the feature vector can generate four x-dimensional class vectors after the first cascade layer. Concatenate the four x-dimensional class vectors with the raw feature vectors and input them to the next cascade layer. If the cascade forest expands to the N th layer and stops, then the class vectors are averaged and the maximum value is taken as the final category prediction result. During the training, K-fold cross validation is used to weaken the overfit in the cascade forest. The number of layers in the cascade forest can be automatically expanded. If the accuracy rate does not improve within a certain number of layers, the expansion will stop and the training will be terminated.

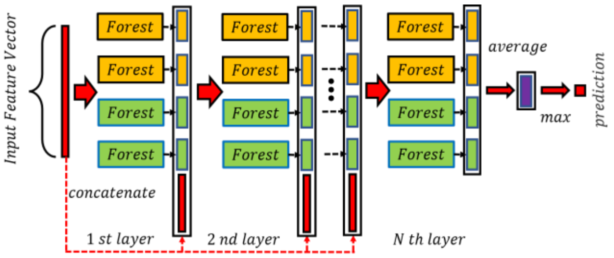


Figure 4. The structure of Cascade Forest

2.3 CatBoost

CatBoost is a machine learning algorithm based on GBDT [26] algorithm, and compared with other GBDT algorithms such as XGBoost [27] and LightGBM [28], it trains faster and more accurately than XGBoost, and CatBoost is not as fast as LightGBM, but it's more accurate [29].

2.4 Proposed Method

To directly classify bearing vibration signals in the case of multiple granularity and high dimensions, we combine the advantages of LSTM and CatBoost algorithms. The main structure of the model proposed in this paper is based on gcForest and improved by LSTM and CatBoost. The model structure is shown in Figure 5. Firstly, the bearing vibration signals are input into the model, and a LSTM layer is used to extract the features of those. Secondly, the extracted feature vectors are input into the cascade CatBoost. Each CatBoost layer will generate class vectors and then splice class vectors with feature vectors as the input of the next layer until the cascade layer is no longer extended. Next, the class vectors of the last layer are averaged and the maximum value is taken as the final category prediction result.

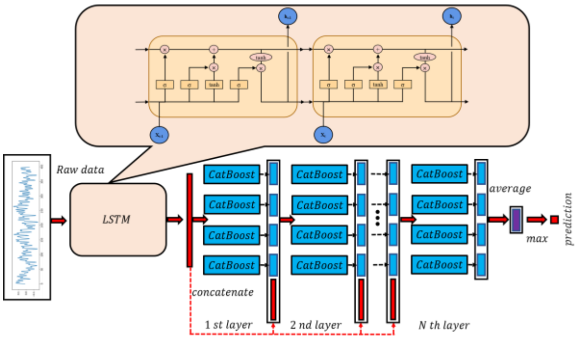


Figure 5. The structure of proposed method

3 Case Study

CWRU bearing dataset [30] and XJTU-SY bearing dataset [31] were used to verify the feasibility of applying the proposed method of bearing fault detection. The computer performance parameters are as follows: Intel Core I5-9400F CPU, NVIDIA GTX 1080Ti GPU, 32GB of RAM. The software framework used are TensorFlow 1.12, Keras 2.2.4 and Scikit-learn 0.24.

3.1 Case Study 1

The CWRU data set used in this experiment is published by the Bearing Data Center of Case Western Reserve University (CWRU) which is a well-known data set in the field of bearing fault diagnosis. The bearing data acquisition platform, as shown in Figure 6, consists of a 1.5kW motor, a torque sensor and a power tester. Two bearings support the operation of the motor spindle. The fan end bearing model is SKF6203, and the drive end bearing model is SKF6205. Electrical discharge machining (EDM) is used to destroy the inner ring, ball and outer ring of bearings with different fault diameters. The accelerating sensor is fixed on both the fan end and the driver end. Vibration signals of the bearing are collected in the sampling rate of 12000, and those of the driver end are also collected in the sampling rate of 48000 in the data set. We use the bearing signals collected at the driving end at 12000 sampling rate to make a data set. The data at 1730rpm, 1772rpm and 1797rpm is used for training, then the data at 1750rpm is used for testing. The ratio of training set to test set is 3:1. The length of each sample is 400, and the sliding sampling is carried out on the original data. The sliding step is 200 and 500 samples are extracted from each original data. The details of the dataset are shown in Table 1. There are four statuses of normal, ball fault (Ball), inner ring fault (IR) and outer ring fault (OR), and the fault diameter is 0.007mm, 0.014mm and 0.021mm, contains 20000 samples in 10 categories.

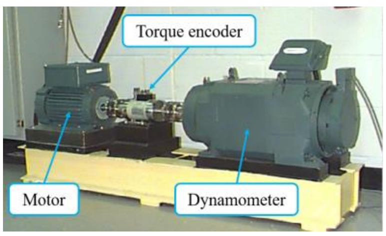


Figure 6. CWRU bearing test platform

Table 1. The details of the datasets from case study 1

Fault type	Label	Speed/rpm	Number (Train/Test)
Normal	0		1500 / 500
0.007-Ball	1	**Train:** 1730 1772 1797	1500 / 500
0.007-IR	2		1500 / 500
0.007-OR	3		1500 / 500
0.014-Ball	4		1500 / 500
0.014-IR	5		1500 / 500
0.014-OR	6	**Test:** 1750	1500 / 500
0.021-Ball	7		1500 / 500
0.021-IR	8		1500 / 500
0.021-OR	9		1500 / 500

The proposed model training process is shown as follows: first, the samples are standardized [32], then they are input into LSTM for feature extraction, finally the cascade CatBoost is used to complete the prediction of sample categories. The parameters of LSTM are shown in Table 2. Single-layer LSTM is used, the batch size is 128, the epoch is 200. The 256-dimensional eigenvector output from the LSTM layer is used as the input of the cascade CatBoost. The cascade layer adopts four CatBoost classifiers with the same parameters as shown in Table 3. Each classifier uses 10-fold cross-validation training.

Table 2. The parameters of LSTM

Layer	Value	Other parameters
Input layer	400×1	Learning rate: 0.001
LSTM	256	Loss: categorical_ crossentropy
Dropout	0.3	
Output layer	10	Optimizer: rmsprop

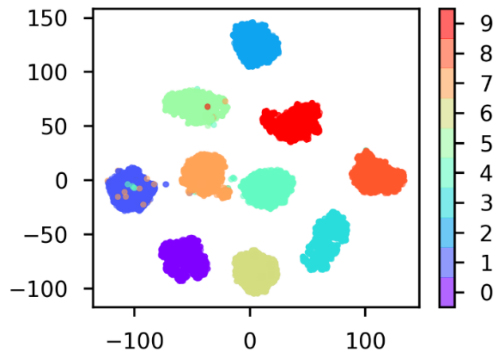
Table 3. The parameters of CatBoost

Parameters	Value
Iterations	80
Learning rate	0.3
Loss	MultiClass
depth	3

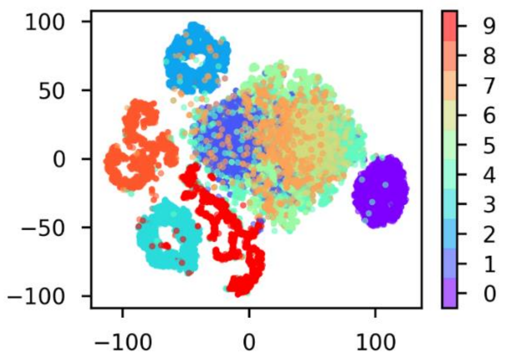
During the experiment LSTM-Cascade Forest was added as the control. The multi-grained scanning part of gcForest was replaced while the other structures remained unchanged. The test results of the experiment are shown in Table 4. The essential reason why the accuracy of gcForest is inferior to that of LSTM is that the feature extraction ability of multi-grained scanning is weaker than that of LSTM layer. It can be proved by the fact that LSTM-Cascade Forest has higher accuracy than gcForest. In order to intuitively illustrate that LSTM is superior to multi-grained scanning in feature extraction, T-SNE [33] is used to visualize the feature vectors generated by multi-grained scanning of LSTM layer and 25-dimensional sliding window respectively. The boundary of each cluster can be clearly seen in the T-SNE distribution diagram of LSTM shown in Figure 7. However, in the T-SNE distribution map of multi-grained scanning, the cluster boundary is fuzzy and some categories cannot be distinguished. The accuracy of LSTM-Cascade CatBoost is the highest in this experiment, indicating that replacing the random forest classifier in the Cascade layer with a more advanced CatBoost classifier can further improve the classification performance of the model. The test set confusion matrix of LSTM-Cascade CatBoost is shown in Figure 8.

Table 4. The test accuracy of case study 1

Method	Accuracy (%)
LSTM-Cascade CatBoost	99.72
LSTM-Cascade Forest	99.54
LSTM	99.23
gcForest	98.12
CatBoost	96.73



(a) t-SNE of LSTM layer



(b) t-SNE of Multi-Grained Scanning

Figure 7. The visualization of t-SNE distributions

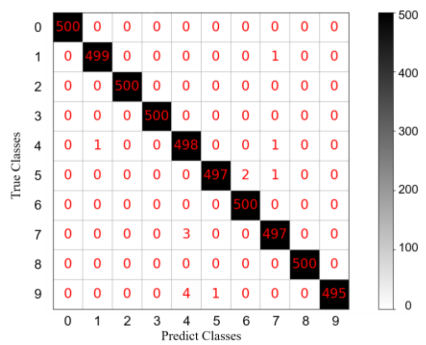


Figure 8. The confusion matrix of study 1

3.2 Case Study 2

Since the bearing failure of the CWRU dataset was artificially caused, the XJTU-SY dataset was used to test the robustness of the model under actual working conditions. The test platform is shown in Figure 9. The bearing model used in this experiment is LDK UER204, and two acceleration sensors are used to collect the horizontal and vertical signals of the bearing respectively. The test was carried out under three working conditions with 5 bearings of each type, and a total of 15 bearings' whole-life cycle vibration signals were obtained. The sampling rate was 25.6k, the single sampling time was 1.28s, and each sampling interval was 1min. The signals of normal bearing, outer ring failure, inner ring failure and cage failure are extracted from the data of three working conditions to make data sets. The length of each sample is 400 and the sliding step is 100. The total number of samples is 2400, the detailed information is shown in Table 5.

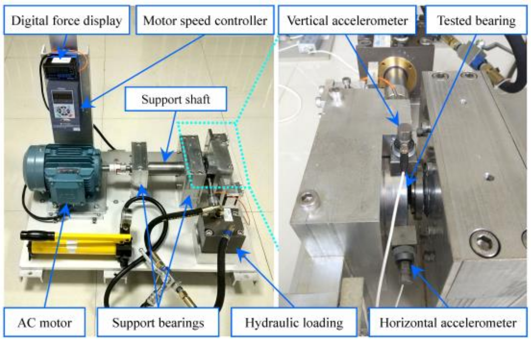


Figure 9. XJTU-SY bearing test platform

Table 5. The details of the datasets from case study 2

Fault type	Label	Number (Train/Test)
Normal	0	450 / 150
Cage	1	450 / 150
InnerRace	2	450 / 150
OuterRace	3	450 / 150

The method in Case Study 1 was used for training, the batch size and the epoch were adjusted to 64 and 180 respectively. Other parameters remain unchanged. The classification accuracy of the proposed model is shown in Table 6. From the experimental results, the accuracy is generally decreased, which is mainly caused by the change of data set size, but the decrease is not very large, and the accuracy of the proposed method is still the highest among the methods in the table.

Table 6. The test accuracy of case study 2

Method	Accuracy (%)
LSTM-Cascade CatBoost	99.33
LSTM	99.00
gcForest	97.86
CatBoost	96.16

4 Conclusion and Prospect

In this paper, we propose a bearing vibration signal fault diagnosis method based on LSTM-Cascade CatBoost. The CWRU bearing dataset and XJTU-SY bearing dataset were used to verify the method. By comparing the T-SNE distribution of LSTM and multi-grained scanning, it has been proved that LSTM has a better feature extraction ability, and the proposed method also achieves the highest accuracy in comparison with other methods, therefore, it proves the effectiveness of our proposed method. Two data sets of different sizes also illustrate the robustness of this method.

Although this method has achieved satisfactory results, this is not the end of the study. LSTM in deed has high accuracy, but its training speed is slow. GRU, as an improved method of LSTM, is not only faster, but also has similar accuracy. Next, we will study the algorithm of feature extraction to further improve the performance of this method.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 92060114, and in part by the Sichuan Province Key Research and Development Project of under Grant (2021YFG0198, 2021YFG0079).

References

[1] Z. Liu, X. Wang, L. Zhang, Fault Diagnosis of Industrial Wind Turbine Blade Bearing Using Acoustic Emission Analysis, *IEEE Transactions on Instrumentation and Measurement*, Vol. 69, No. 9, pp. 6630-6639, September, 2020.

[2] R. Chen, L. Tang, X. Hu, H. Wu, Fault Diagnosis Method of Low-Speed Rolling Bearing Based on Acoustic Emission Signal and Subspace Embedded Feature Distribution Alignment, *IEEE Transactions on Industrial Informatics*, Vol. 17, No. 8, pp. 5402-5410, August, 2021.

[3] H. Miao, D. He, A New Hybrid Deep Signal Processing Approach for Bearing Fault Diagnosis Using Vibration Signals, *Neurocomputing*, Vol. 396, pp. 542-555, July, 2020.

[4] J. Jiao, J. Yue, D. Pei, Z. Hu, Application of Feature Fusion Using Coaxial Vibration Signal for Diagnosis of Rolling Element Bearings, *Shock and Vibration*, Vol. 2020, pp. 1-14, October, 2020.

[5] A. Khairnar, A. Patange, S. Pardeshi, R. Jegadeeshwaran, Supervision of Carbide Tool Condition by Training of Vibration-based Statistical Model using Boosted Trees Ensemble, *International Journal of Performability Engineering*, Vol. 17, No. 2, pp. 229-240, February, 2021.

[6] M. Cerrada, R. V. Sánchez, C. Li, F. Pacheco, D. Cabrera, J. V. D. Oliveira, R. E. Vasquez, A Review on Data-driven Fault Severity Assessment in Rolling Bearings, *Mechanical Systems and Signal Processing*, Vol. 99, pp. 169-196, January, 2018.

[7] H. Lin, Y. Ye, Reviews of Bearing Vibration Measurement Using Fast Fourier Transform and Enhanced Fast Fourier Transform Algorithms. *Advances in Mechanical Engineering*, Vol. 11, No. 1, pp. 1-12, January, 2019

[8] W. Fan, Q. Zhou, J. Li, Z. Zhu, A Wavelet-based Statistical Approach for Monitoring and Diagnosis of Compound Faults with Application to Rolling Bearings, *IEEE Transactions on Automation Science and Engineering*, Vol. 15, No. 4, pp. 1563-1572, October, 2018.

[9] F. Ding, F. W. Qin, Application of Wavelet Denoising and Hilbert Transform in Fault Diagnosis of Motor Bearing, *Electric Machines and Control*, Vol. 21, No. 6, pp. 89-95, June, 2017.

[10] Y. Wang, M. Yang, Y. Li, Z. Xu, J. Wang, X. Fang, A Multi-Input and Multi-Task Convolutional Neural

Network for Fault Diagnosis Based on Bearing Vibration Signal, *IEEE Sensors Journal*, Vol. 21, No. 9, pp. 10946-10956, May, 2021.

[11] J. Guo, X. Liu, S. Li, Z. Wang, Bearing Intelligent Fault Diagnosis Based on Wavelet Transform and Convolutional Neural Network, *Shock and Vibration*, Vol. 2020, pp. 1-14, November, 2020.

[12] L. Yuan, D. Lian, X. Kang, Y. Chen, K. Zhai, Rolling Bearing Fault Diagnosis Based on Convolutional Neural Network and Support Vector Machine, *IEEE Access*, Vol. 8, pp. 137395-137406, July, 2020.

[13] M. Shuo, J. Kang, K. Chi, X. Die, Intelligent Fault Diagnosis of Gearbox based on Multiple Synchrosqueezing S-Transform and Convolutional Neural Networks, *International Journal of Performability Engineering*, Vol. 16, No. 4, pp. 528-536, April, 2020.

[14] M. Gan, C. Wang, C. Zhu, Construction of Hierarchical Diagnosis Network Based on Deep Learning and Its Application in the Fault Pattern Recognition of Rolling Element Bearings, *Mechanical Systems and Signal Processing*, Vol. 72-73, pp. 92-104, May, 2016.

[15] S. Hao, F. Ge, Y. Li, J. Jiang, Multisensor Bearing Fault Diagnosis Based on One-dimensional Convolutional Long Short-term Memory Networks, *Measurement*, Vol. 159, Article No. 107802, July, 2020.

[16] Z. Zhou, J. Feng, Deep Forest, *National Science Review*, Vol. 6, No. 1, pp. 74-86, January, 2019.

[17] X. Qin, D. Xu, X. Dong, X. Cui, S. Zhang, The Fault Diagnosis of Rolling Bearing Based on Improved Deep Forest, *Shock and Vibration*, Vol. 2021, pp. 1-13, June, 2021.

[18] Y. Xu, Z. Li, S. Wang, W. Li, T. Sarkodie-Gyan, S. Feng. A Hybrid Deep-learning Model for Fault Diagnosis of Rolling Bearings, *Measurement*, Vol. 169, Article No. 108502, February, 2021.

[19] S. Hochreiter, J. Schmidhuber, Long Short-term Memory, *Neural computation*, Vol. 9, No. 8, pp. 1735-1780, November, 1997.

[20] A. V. Dorogush, V. Ershov, A. Gulin, *CatBoost: Gradient Boosting with Categorical Features Support*, October, 2018. https://arxiv.org/abs/1810.11363

[21] H. Zhang, G. Niu, B. Zhang, Q. Miao, Cost-Effective Lebesgue Sampling Long Short-Term Memory Networks for Lithium-ion Batteries Diagnosis and Prognosis, *IEEE Transactions on Industrial Electronics*, Vol. 69, No. 2, pp. 1958-1967, February, 2022.

[22] S. Kumar, A. Damaraju, A. Kumar, S. Kumari, C. Chen, LSTM Network for Transportation Mode Detection, *Journal of Internet Technology*, Vol. 22, No. 4, pp. 891-902, July, 2021.

[23] L. Breiman, Random Forests, *Machine learning*, Vol. 45, No. 1, pp. 5-32, October, 2001.

[24] P. Geurts, D. Ernst, L. Wehenkel, Extremely Randomized Trees, *Machine learning*, Vol. 63, No. 1, pp. 3-42, April, 2006.

[25] W. Loh, Classification and Regression Trees, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 1, No. 1, pp. 14-23, January/ February, 2011.

[26] J. H. Friedman, Stochastic Gradient Boosting, *Computational statistics & data analysis*, Vol. 38, No. 4, pp. 367-378, February, 2002.

[27] T. Chen, C. Guestrin, Xgboost: A Scalable Tree Boosting System, *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, San Francisco, CA, USA, 2016, pp. 785-794.

[28] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T. Liu, Lightgbm: A Highly Efficient Gradient Boosting Decision Tree, *Advances in neural information processing systems*, Long Beach, CA, USA, 2017, pp. 3146-3154.

[29] C. Bentéjac, A. Csörgő, G. M. Muñoz, A Comparative Analysis of Gradient Boosting Algorithms, *Artificial Intelligence Review*, Vol. 54, No. 3, pp. 1937-1967, March, 2021.

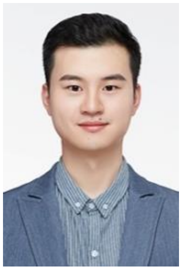
[30] W. A. Smith, R. B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study, *Mechanical systems and signal processing*, Vol. 64-65, pp. 100-131, December, 2015.

[31] B. Wang, Y. Lei, N. Li, N. Li, A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings, *IEEE Transactions on Reliability*, Vol. 69, No. 1, pp. 401-412, March, 2020.

[32] S. Ioffe, C. Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, *International conference on machine learning*, Lille, France, 2015, pp. 448-456.

[33] L. Maaten, G. Hinton, Visualizing Data using t-SNE, *Journal of machine learning research*, Vol. 9, pp. 2579-2605, November, 2008.

Biographies



Miaomiao Yang received the bachelor's degree from Wuhan Polytechnic University in 2019. He is currently pursuing the master's degree in mechanical engineering with Sichuan University. His main research interest includes machine learning and mechanical fault diagnosis.

Weizhi Liu is a sophomore in mechanical engineering from Sichuan University, China. His research interest includes the image processing with artificial intelligence.

Wenxuan Zhang is currently pursuing the Bachelor's Degree and Postgraduate Candidates Exempt from Admission Exam in the Institution of Mechanical Engineering of Sichuan University.

Mei Wang received her Ph.D. degree in mechanical engineering from Sichuan University. She is interested in interaction design, user experience research, and computer aided industrial design theory and technology.

Xia Fang received the Ph.D. degree in mechanical engineering from Sichuan University. He is now a Research Assistant with the School of Mechanical Engineering, Sichuan University. His research interests include pattern recognition, machine learning, and machine vision.