Proceedings of the 3rd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2025.29603

Rolling Bearing Fault Detection and Classification of Train Bogie Systems Based on Deep Learning Algorithms

Yuankang Jiang[1], Youmin Guo[1*]

[1]School of Mechatronic Engineering, Lanzhou Jiaotong University, Lanzhou, China
**Corresponding Author. Email: 2079037437@qq.com*

Abstract. The rolling bearings of the train bogie system, as the core basic components of rail transit trains, are widely used in the bogie system and are key components to ensure the safety of train operation and the reliability of train operation. This paper proposes a classification algorithm for rolling bearings in train bogie systems based on KELM-LSTM-Transformer. During the research process, data preprocessing and data analysis are carried out first, followed by correlation analysis, and then this algorithm is compared with multiple machine learning algorithms. From the comparison results, it can be seen that the KELM-LSTM-Transformer significantly outperforms other single models and integrated models in all key evaluation dimensions Its Accuracy is 0.855, Recall is 0.855, Precision is 0.851, and F1 value is 0.847, all of which are the highest among all models. Not only does it far exceed models with relatively weak performance such as the BP neural network with an F1 value of only 0.641 and the GBDT with an F1 value of 0.732, It also outperforms better-performing models such as CatBoost with an Accuracy of 0.811 and an F1 value of 0.792, as well as XGBoost with an Accuracy of 0.792 and an F1 value of 0.795. In terms of the AUC metric for measuring the model's discrimination ability, the algorithm's value of 0.983 is also ahead of the integrated models such as CatBoost with an AUC of 0.974 and XGBoost with an AUC of 0.97, demonstrating a stronger ability to distinguish positive and negative samples. The proposal of this algorithm provides an effective technical solution for improving the accuracy of condition monitoring and fault classification of rolling bearings in the bogies of rail transit trains, which is of great practical significance for ensuring the safe and stable operation of high-speed trains and promoting the intelligent upgrade of the rail transit field.

Keywords: Train bogie system, Rolling bearing, LSTM, Transformer

1. Introduction

Rolling bearings of train bogie systems, as the core basic components of rail transit trains, are widely used in the bogie systems of rail transit trains. They are the core components for the safety and reliability of train operation, and their operating conditions directly affect the smoothness of train operation, braking performance and operational safety [1]. In train operation, the rolling bearings of the bogie system play a crucial role in transmitting loads, supporting the car body and reducing wheel pair friction. They operate for a long time under conditions such as high speed, variable loads and complex vibrations, and are prone to faults such as wear, cracks, spalling and

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.29603

adhesion due to factors such as material fatigue, lubrication failure and assembly errors [2]. According to statistics, approximately 40% of mechanical failures in rail transit trains are related to the failure of rolling bearings in the bogie system. If such failures are not detected in time, they may lead to sudden stops, delays of the train, and even cause major safety accidents such as derailment, endangering the lives of passengers and causing huge economic losses.

In the field of rail transit, the traditional train maintenance mode is mainly planned maintenance, that is, the rolling bearings of the bogie system are shut down for inspection or replacement according to a fixed operating mileage or time cycle. At present, the fault diagnosis of rolling bearings in train bogie systems mainly relies on traditional methods, including manual inspection, vibration signal analysis, temperature monitoring, etc [3]. Manual inspection requires the interruption of train operations, which is inefficient and has limited ability to identify weak faults in the early stage. With the development of sensor technology in rail transit, the amount of monitoring data that can be collected during train operation has grown exponentially, and traditional methods have become difficult to meet the demands of real-time processing and in-depth analysis of massive data [4].

"Made in China 2025" explicitly proposes to promote the intelligence of the manufacturing process, requiring the improvement of the intelligence level of key procedures and the realization of real-time monitoring and intelligent diagnosis of equipment status. In the field of rail transit, machine learning technology, with its powerful data mining and autonomous learning capabilities, can automatically extract fault features from massive multi-source monitoring data, achieving early identification, type classification, and severity assessment of rolling bearing faults in train bogie systems, providing core technical support for condition-based maintenance [5]. Compared with traditional methods, the fault diagnosis method based on machine learning has higher diagnostic accuracy, stronger anti-interference ability and wider adaptability to working conditions, and can effectively solve the problem of fault diagnosis of rolling bearings in bogie systems in complex rail transit environments [6].

In conclusion, conducting research on the fault diagnosis method of rolling bearings in train bogie systems based on machine learning can not only break through the bottlenecks of traditional diagnostic techniques, enhance the safe operation level of rail transit trains, and reduce maintenance costs, but also promote the development of intelligent monitoring technology in rail transit and facilitate the transformation and upgrading of the rail transit industry. It has significant theoretical research value and broad engineering application prospects.

2. Related research

As a core component of rail transit trains, the fault detection of rolling bearings in train bogie systems has always been a key link in train operation and maintenance safety. Early fault detection mainly relies on manual disassembly and inspection. Technicians judge the condition of rolling bearings by visual inspection and touch. This method requires the interruption of train operation, is inefficient and difficult to detect early weak faults [7]. With the development of rail transit detection technology, a series of traditional non-destructive testing methods have gradually been applied, such as bogie vibration signal analysis method, axle temperature monitoring method, and oil analysis method, etc. These methods have achieved the initial identification of faults to a certain extent, but due to the influence of the dynamic operating conditions of the train, they have problems such as slow detection speed, low sensitivity to early faults, and susceptibility to environmental interference [8]. In recent years, with the improvement of sensor technology and data processing capabilities in rail transit, the fault detection technology of rolling bearings in train bogie systems based on data-

DOI: 10.54254/2755-2721/2025.29603

driven has become a research hotspot. Compared with traditional methods, the data-driven approach achieves fault diagnosis through the analysis of train operation monitoring data, featuring high levels of automation and fast response speed. It has been initially applied in fields such as high-speed railways and subways. However, the existing technology still has deficiencies in terms of robustness under complex working conditions (such as curved driving and uneven tracks) and accurate classification of multiple fault types, which urgently need further optimization [9].

2.1. A fault detection method for rolling bearings in train bogie systems based on traditional signal processing

The traditional signal processing method is the basis for the fault diagnosis of rolling bearings in the train bogie system. Its core idea is to extract features from the original signal and then combine threshold judgment or simple classifiers to achieve fault identification. Typical methods include:

Fourier Transform (FT) : By converting time-domain signals to the frequency domain, it identifies fault characteristic frequencies, but it is difficult to handle non-stationary vibration signals during train start-up/braking. Wavelet Transform (WT) : It captures the local mutation characteristics of signals through multi-scale decomposition and is suitable for the analysis of non-stationary signals. It is widely used in the detection of early cracks and wear faults in bogie rolling bearings [10]. Empirical Mode Decomposition (EMD) : Adaptively decomposes the signal into an inherent mode function (IMF), effectively extracting fault features from nonlinear and non-stationary signals. However, it is prone to mode aliasing problems due to the interference of train wheel-rail vibration [11]. Scholars at home and abroad have conducted extensive research in this field: Smith et al. utilized wavelet packet decomposition to extract the energy characteristics of the vibration signals of bogie rolling bearings and combined the threshold method to achieve the detection of early wear faults [12]; Domestic scholars Zhang Wei et al. identified the crack characteristics of the outer ring of the rolling bearing of the metro bogie through EMD decomposition and Hilbert spectrum analysis, with an accuracy rate of 82% [13].

2.2. A fault detection method for rolling bearings in train bogie systems based on machine learning

With the development of machine learning technology, its application in the fault diagnosis of rolling bearings in train bogie systems has gradually matured. The core process is "signal preprocessing→feature extraction→classifier training". Through the classifier's autonomous learning of the mapping relationship between features and faults, the intelligent level of diagnosis is enhanced. Typical methods include:

Support Vector Machine (SVM) : It transforms a linearly inseparable problem in a high-dimensional feature space into a linearly separable problem in a low-dimensional space through kernel function mapping, and performs exceptionally well in the case of small samples. For instance, Li et al. classified the small-wave characteristics of the vibration signals of bogie rolling bearings using SVM, achieving the identification of crack and spalling faults with an accuracy rate of 90% [14]; Artificial Neural Network (ANN) : It learns features through nonlinear mapping of multiple layers of neurons and is suitable for complex pattern recognition. Wang et al. constructed a BP neural network to train the time-domain features (such as mean and variance) of the vibration signals of bogie rolling bearings, achieving a multi-fault classification accuracy rate of 86% [15]. Ensemble learning methods, such as AdaBoost and random Forest, enhance generalization ability by fusing multiple weak classifiers. Domestic scholars Liu Qiang et al. used the random forest

DOI: 10.54254/2755-2721/2025.29603

algorithm to conduct fusion diagnosis on multi-sensor data (vibration, shaft temperature, speed) of high-speed rail bogies, which improved the accuracy rate by 12% compared with a single classifier [16].

3. Data sources

This dataset contains fault diagnosis data of rolling bearings in the train bogie system, with a total of 527 records. The data encompasses multi-dimensional information. The basic information dimensions include sampling time, bearing model and operating time. The operating environment includes train speed, track type, environmental temperature, humidity and line condition score. Vibration characteristics distinguish between the time domain and the frequency domain. The time domain features include vibration peak value, vibration root mean square, and kurtosis coefficient. The frequency domain features include characteristic frequency amplitude and dominant frequency. It also includes physical parameters such as shaft temperature, noise level, load percentage, and maintenance interval; At the same time, mark whether it is a fault and the type of fault and other fault-related labels. The data were statistically analyzed as shown in Table 1.

Table 1. The statistical results of the dataset

Indicator	Mean	Median	Variance
Running time (hours)	1495.98	1502.62	235109.7
Train speed (km/h)	83.25	82.93	1486.76
Ambient temperature (°C)	25.5	25.44	99.43
Humidity (%)	60.18	60.02	349.04
Shaft temperature (°C)	49.62	46.73	288.6
Vibration peak (g)	0.46	0.33	0.15
Vibration RMS (g)	0.29	0.21	0.06
Kurtosis coefficient	3.85	3.15	3.27
Characteristic frequency amplitude (mm/s)	1.1	0.88	0.58
Dominant frequency (Hz)	149.48	149.38	2281.88
Noise level (dB)	74.41	74.61	99.98
Load percentage (%)	68.16	68.88	400.29
Track condition score (1-10)	5.63	6	8.61
Maintenance interval (days)	89.22	88.95	967.79

Correlation analysis was conducted on each variable, and a correlation heat map was drawn, as shown in Figure 1. From the correlation data of each variable with whether there is a fault or not, it can be seen that variables such as vibration peak value, vibration root mean square, kurtosis coefficient and shaft temperature have a strong positive correlation with it, with correlation coefficients of 0.80, 0.77, 0.87 and 0.73 respectively. This indicates that the higher the values of these variables are, the greater the possibility of a failure occurring. Variables such as vibration peak value, root mean square of vibration, kurtosis coefficient and shaft temperature show a strong negative correlation with fault types, with correlation coefficients of -0.50, -0.51, -0.56 and -0.48 respectively. This might imply that when these variables are at different levels, the corresponding types of faults are different.

DOI: 10.54254/2755-2721/2025.29603

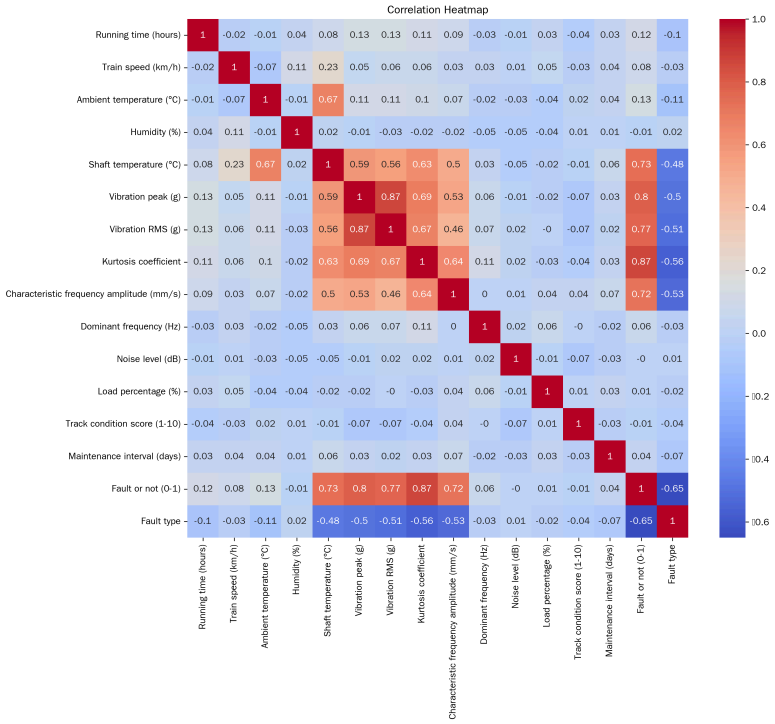


Figure 1. Correlation analysis

4. Method

4.1. KELM

KELM (Kernel Extreme Learning Machine) is an improved single-hidden-layer feedforward neural network based on ELM (Extreme Learning Machine). Its core lies in solving the problem of unstable generalization ability that may be caused by the random initialization of input weights in ELM through kernel functions. It retains the efficiency of ELM in randomly generating input weights and analytively solving output weights. At the same time, it introduces kernel techniques: there is no need to manually set the number of hidden layer nodes and activation functions. Instead, a kernel matrix is used to replace the hidden layer output matrix, mapping samples to a high-dimensional feature space to handle nonlinear problems. The output weights are directly calculated through the Momoore -Penrose generalized inverse, avoiding the iterative training of traditional neural networks and featuring both fast training speed and strong generalization performance.

4.2. LSTM

LSTM is a time series model designed to solve the problem of vanishing/exploding gradients in long sequences of traditional RNNS. Its core lies in the gating mechanism and cell states. The cell state is similar to a conveyor belt, capable of transmitting information for a long time and only subject to gate fine-tuning. Three gates control the flow of information through the sigmoid function: the forget gate determines how many historical cell states to discard, the input gate filters new information and updates the cell states, and the output gate generates the current hidden state based on the cell states. This structure enables LSTM to capture long-term dependencies. The network structure of LSTM is shown in Figure 2.

DOI: 10.54254/2755-2721/2025.29603

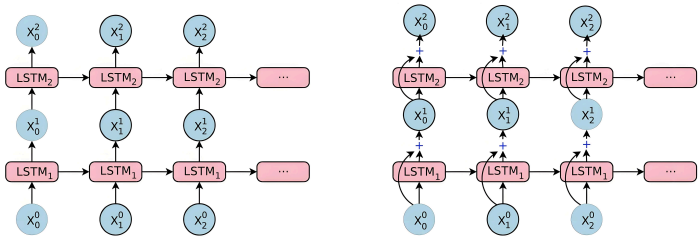


Figure 2. Correlation analysis

4.3. Transformer

Transformer is a model that is entirely based on the self-attention mechanism, discarding the temporal dependence of RNN to achieve efficient parallel computing. The core is self-attention. By calculating the attention weight of each input token and all tokens, the degree of association between tokens is quantified, thereby capturing global dependencies. Multi-head attention captures feature associations from different dimensions through multiple parallel self-attention heads, enhancing expressive power. The model is divided into an encoder and a decoder (a mask self-attention is added on the basis of the encoder to prevent the utilization of future information and is responsible for generating the output). The network structure of the Transformer is shown in Figure 3.

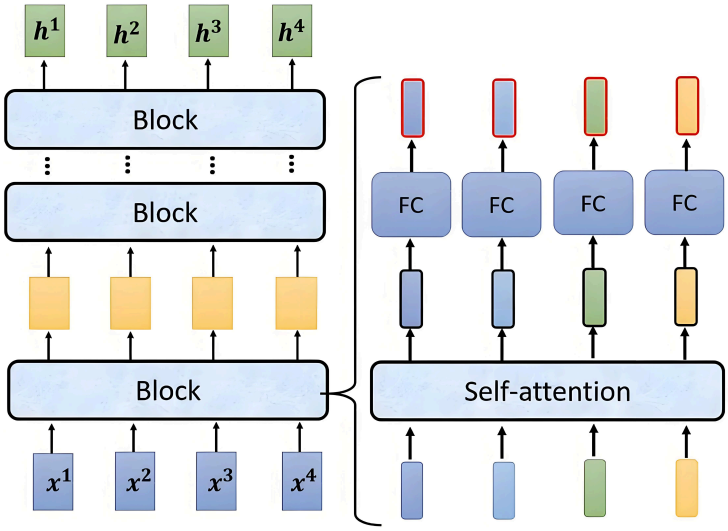


Figure 3. Correlation analysis

4.4. KWLM-LSTM-Transformer

The KELM-LSTM-Transformer classification algorithm is a hybrid model integrating Kernel Extreme Learning Machine (KELM), Long Short-Term Memory Network (LSTM), and Transformer. Its core is to enhance classification performance by collaboratively mining the temporal patterns and global correlations of sequential data through module collaboration. The process is as follows: Firstly, input the data into the LSTM. Utilize the gating mechanism of the LSTM to capture local long and short-term temporal dependencies, avoid the gradient vanishing problem of traditional RNN, and initially extract temporal features. Then, the feature sequence

DOI: 10.54254/2755-2721/2025.29603

output by LSTM is passed into the Transformer encoder. The global correlation weights between features are calculated through the self-attention mechanism to accurately capture the global dependencies across positions and make up for the local perception limitations of LSTM. Finally, the global feature vectors output by the Transformer are input into the KELM. The KELM uses the kernel function to map the features to a high-dimensional space, enabling rapid classification without iterative training and balancing generalization ability and efficiency. The algorithm flowchart of KELM-LSTM-Transformer is shown in Figure 4.

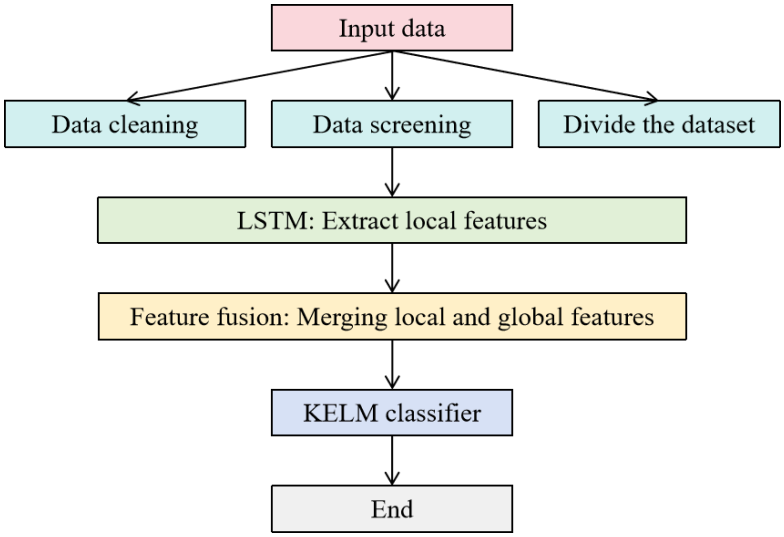


Figure 4. Correlation analysis

5. Result

In terms of data division, the dataset is divided into a training set and a test set in a ratio of 7:3. The input channel number of the Transformer-LSTM model is the feature dimension, the maximum position encoding is 256×2, the number of self-attention mechanism heads is 4, and the number of key channels for each head is 4×32. It includes a sequence input layer, a position embedding layer, an addition layer, two self-attention layers, an LSTM layer, a relu activation layer, a dropout layer, an index layer, a fully connected layer, a softmax layer and a classification layer. In terms of training parameters, the optimizer adam is selected, with a maximum of 800 training rounds and a batch size of 256. Each training session will disrupt the dataset. The initial learning rate is 0.01, the learning rate decline factor is 0.1, the decline period is 650, the L2 regularization coefficient is 0.001, and the gradient clipping threshold is 10. The regularization coefficient of KELM is 20, the kernel parameter is 6, and the kernel function type is rbf.

For the comparative models, this paper selects Decision tree, AdaBoost, GBDT, CatBoost, BP neural network, XGBoost and SVM. Regarding the evaluation parameters, This article employs Accuracy, Recall, Precision, F1 and AUC. Output the comparison results of each model as shown in Table 2. The bar chart of the comparison results of each indicator is shown in Figure 5.

DOI: 10.54254/2755-2721/2025.29603

Table 2. The results of the comparative experiment

Model	Accuracy	Recall	Precision	F1	AUC
Decision tree	0.799	0.799	0.795	0.789	0.883
AdaBoost	0.78	0.78	0.805	0.792	0.951
GBDT	0.748	0.748	0.716	0.732	0.949
CatBoost	0.811	0.811	0.777	0.792	0.974
BP neural network	0.748	0.748	0.56	0.641	0.843
XGBoost	0.792	0.792	0.799	0.795	0.97
SVM	0.742	0.742	0.822	0.765	0.956
KELM-LSTM-Transformer	0.855	0.855	0.851	0.847	0.983

From the perspective of model performance metrics, the KELM-LSTM-Transformer significantly outperforms other single models and integrated models in all key evaluation dimensions: Its Accuracy (0.855), Recall (0.855), Precision (0.851), and F1 value (0.847) are all the highest among all models. Not only do they far exceed models with weaker performance such as BP neural network (F1 only 0.641) and GBDT (F1 0.732), It is also superior to the better-performing CatBoost (Accuracy 0.811, F1 0.792) and XGBoost (Accuracy 0.792, F1 0.795).

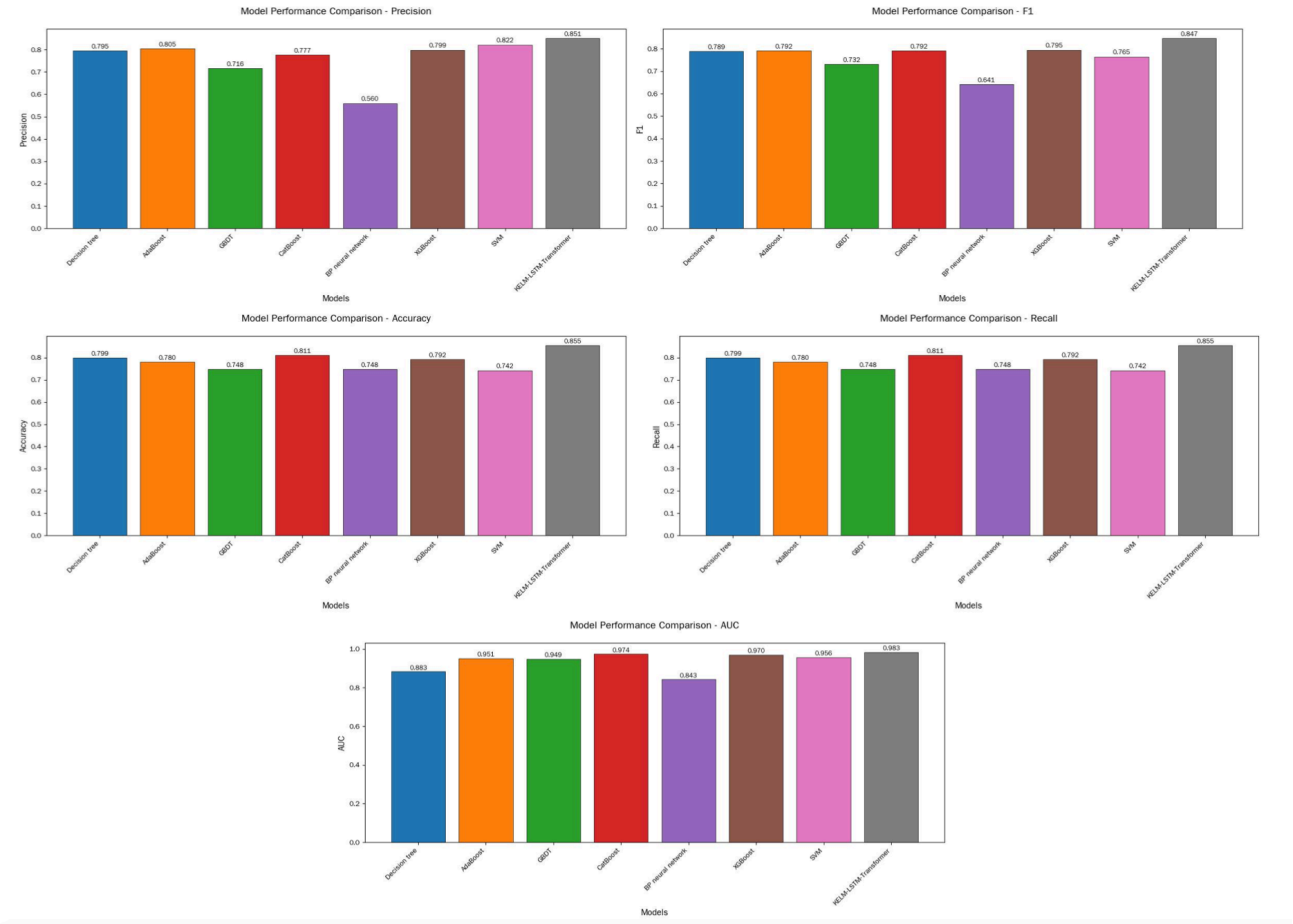


Figure 5. The comparison bar chart of the indicators of each model

DOI: 10.54254/2755-2721/2025.29603

In terms of the AUC metric for measuring the model's discrimination ability, its value of 0.983 also leads that of integrated models such as CatBoost (0.974) and XGBoost (0.97), demonstrating a stronger ability to distinguish positive and negative samples.

6. Conclusion

The rolling bearings of the train bogie system are the core basic components of the bogie system in rail transit trains and are widely used in it. They are also the key to ensuring the safety and reliability of train operation. With the continuous advancement of intelligent rail transit, high-speed trains are gradually developing towards high speed, high reliability and long endurance, which puts forward more stringent requirements for the safety of rolling bearings in the bogie system. This paper proposes a classification algorithm for rolling bearings in train bogie systems based on KELM-LSTM-Transformer. In the research, data preprocessing and data analysis are carried out first, followed by correlation analysis, and multiple machine learning algorithms are also used for comparative experiments. From the results, it can be seen that the KELM-LSTM-Transformer significantly outperforms other single models and integrated models in all key evaluation dimensions. Its Accuracy is 0.855, Recall is 0.855, Precision is 0.851, and F1 value is 0.847 All are the highest among all models, not only significantly surpassing models with weaker performance such as the BP neural network (with an F1 value of only 0.641) and GBDT (with an F1 value of 0.732), It also outperforms the better-performing CatBoost (Accuracy of 0.811, F1 value of 0.792) and XGBoost (Accuracy of 0.792, F1 value of 0.795). In terms of the AUC metric for measuring the model's discrimination ability, the algorithm's value of 0.983 is also ahead of CatBoost's 0.974 and XGBoost's 0.97, demonstrating a stronger ability to distinguish positive and negative samples. The proposal of this algorithm provides more reliable technical support for the safety monitoring of rolling bearings in the bogie system of rail transit trains, which is conducive to further enhancing the safety guarantee level of high-speed trains and promoting the in-depth application of intelligent rail transit technology in the field of core component monitoring.

References

[1] Zhang, Shen, et al. "Deep learning algorithms for bearing fault diagnostics—A comprehensive review." IEEE access 8 (2020): 29857-29881.
[2] He, Miao, and David He. "Deep learning based approach for bearing fault diagnosis." IEEE Transactions on Industry Applications 53.3 (2017): 3057-3065.
[3] Hoang, Duy-Tang, and Hee-Jun Kang. "A survey on deep learning based bearing fault diagnosis." Neurocomputing 335 (2019): 327-335.
[4] Li, Bo, "Neural network-based motor rolling bearing fault diagnosis." IEEE transactions on industrial electronics 47.5 (2002): 1060-1069.
[5] Chen, Xiaohan, et al. "Deep transfer learning for bearing fault diagnosis: A systematic review since 2016." IEEE Transactions on Instrumentation and Measurement 72 (2023): 1-21.
[6] Lou, Xinsheng, and Kenneth A. Loparo. "Bearing fault diagnosis based on wavelet transform and fuzzy inference." Mechanical systems and signal processing 18.5 (2004): 1077-1095.
[7] Chen, Zhiqiang, et al. "Deep neural networks-based rolling bearing fault diagnosis." Microelectronics Reliability 75 (2017): 327-333.
[8] Wei Wenbo, et al. "Research on Fault Diagnosis of Train Bearings Based on Spectral Recursive Image Coding Technology." Rolling Stock (1002-7602) 63.4 (2025).
[9] Li, Chuan, A systematic review of fuzzy formalisms for bearing fault diagnosis. IEEE Transactions on Fuzzy Systems 27.7 (2018): 1362-1382
[10] Li, Xiang, Wei Zhang, and Qian Ding. "Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism." Signal processing 161 (2019): 136-154.

DOI: 10.54254/2755-2721/2025.29603

[11] YaRai, V. K., and A. R. Mohanty. "Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform." Mechanical systems and signal processing 21.6 (2007): 2607-2615.
[12] Jin, Zhenzhen, "Bearing fault diagnosis based on VMD and improved CNN." Journal of Failure Analysis and Prevention 23.1 (2023): 165-175.
[13] Cui Hongju, et al. "Fault Diagnosis of Axle Box Bearings for Rail Vehicles Based on Inverted Spectrum Pre-whitening." Computer Measurement & Control 26.9 (2018).
[14] Pan, Honghu, et al. "An improved bearing fault diagnosis method using one-dimensional CNN and LSTM." Strojniski Vestnik-Journal of Mechanical Engineering 64.7-8 (2018): 443-453.
[15] Mohanty, Satish, Karunesh Kumar Gupta, and Kota Solomon Raju. "Comparative study between VMD and EMD in bearing fault diagnosis." 2014 9th international conference on industrial and information systems (ICIIS). IEEE, 2014.
[16] Hoang, Duy Tang, and Hee Jun Kang. "A motor current signal-based bearing fault diagnosis using deep learning and information "IEEE Transactions on Instrumentation and Measurement 69.6 (2019): 3325-3333.