Research on Industrial Motor Fault Prediction Based on LSTM-XGBoost

Shitong Qiao[1*], Wenbo Zhang[2], Qisheng Huang[3]
[1]School of Mathematics and Statistics Southwest University Chongqing, China
[2]Collageof Computer and Information Science Collage of Software Southwest University Chongqing, China
[3]School of Physical Science and Technology Southwest University Chongqing, China
[*]18324672005@163.com

ABSTRACT

Industrial motors, as core power equipment in the manufacturing industry, are prone to sudden failures that can lead to production line shutdowns, safety accidents, and economic losses. Traditional periodic maintenance strategies are often inefficient and lack precision, and single prediction models are insufficient to meet the demand for accurate prediction. This study aims to construct an LSTM-XGBoost combined model to achieve high-precision, real-time prediction of industrial motor faults. Time-series data on motor vibration, temperature, and current are collected from multiple sensors. After preprocessing including cleaning, Z-score standardization, and multi-domain feature extraction, a 6:2:2 time series partitioning strategy is used to construct the dataset. Model hyperparameter tuning is achieved using Bayesian optimization, and the weight allocation and fusion of the two models are realized through the inverse variance method. Experimental results show that the combined model has lower RMSE and MAPE than the single LSTM and XGBoost models, achieving a fault prediction accuracy of 94.7% and a false alarm rate of only 4.2%, while maintaining computational efficiency that meets the requirements of real-time applications. This research effectively overcomes the limitations of single algorithms, providing more reliable technical support for predictive maintenance of industrial motors and contributing to the intelligent and lean transformation of the manufacturing industry.

Keywords: Industrial motor; Fault prediction; LSTM; XGBoost; Combined model; Data preprocessing

1. INTRODUCTION

1.1 Research Background and Significance

Industrial motors, as core power equipment in modern manufacturing, play a crucial role in the production process. The suddenness of motor failures often leads to production line shutdowns, decreased product quality, and even serious safety accidents and economic losses. While traditional periodic maintenance strategies can prevent failures to some extent, their blindness and uneconomical nature are increasingly prominent, failing to meet the urgent needs of modern industry for high-efficiency, low-cost operation.

With the rapid development of industrial IoT technology, real-time monitoring of motor operating status has become possible. A large amount of sensor data, such as vibration, temperature, and current, provides a rich source of information for fault prediction. Machine learning algorithms have shown great potential in processing these complex time-series data, especially the advantages of LSTM networks in capturing long-term time dependencies, and the superior performance of XGBoost models in feature extraction and nonlinear mapping [1-2] . LSTM algorithms can discover temporal correlations in longer sequences and are better suited for predicting data with strong nonlinearity, while XGBoost models belong to the Boosting algorithm, continuously generating new decision tree models to fit the residuals, resulting in continuously increasing accuracy [3] .

The significance of this research lies in constructing a hybrid prediction model that integrates the temporal modeling capabilities of LSTM and the feature learning advantages of XGBoost, addressing the limitations of single models in motor fault prediction. The prediction results of the LSTM algorithm have a lag, deteriorating with increasing prediction time, while the prediction performance of the XGBoost algorithm can change abruptly at a random sample. Through reasonable weight allocation and error correction mechanisms, this combined model can effectively reduce the prediction error of traditional single models, providing more accurate and reliable

Second International Conference on Artificial Intelligence, Deep Learning, and Neural Networks (AIDLNN 2025), edited by Shiling Zhang, Proc. of SPIE Vol. 14182, 141822M

technical support for predictive maintenance of industrial motors and promoting the development of manufacturing towards intelligence and lean manufacturing.

1.2 Research Objectives and Content

As the core power equipment of modern manufacturing, the operating status of industrial motors directly affects production efficiency and safety. Traditional fault detection methods often rely on periodic maintenance and post-event repair, making it difficult to achieve accurate prediction of motor faults. This study aims to construct an industrial motor fault prediction system based on the LSTM-XGBoost combined model, improving the accuracy and real-time performance of fault prediction through the complementary advantages of deep learning and ensemble learning.

In terms of research objectives, this paper strives to solve the limitations of single prediction models in processing motor time-series data. Although LSTM networks can effectively capture long-term dependencies in time series, their prediction results often have a lag, and the prediction effect will significantly decrease as the prediction time increases. The XGBoost algorithm performs well in handling complex nonlinear relationships, but may exhibit prediction mutations at some random samples. By organically combining the two algorithms, their respective advantages can be fully utilized to form a complementary effect, thereby obtaining more stable and accurate prediction results.

The main contents of this study cover the acquisition and preprocessing of motor operation data, the construction and training of LSTM neural network models, the parameter optimization of XGBoost models, and effective fusion strategies of the two models. In terms of model combination, the inverse variance method and equal weight allocation technique will be adopted to ensure that the combined model can effectively reduce the prediction error of a single model. By comparing and analyzing the performance indicators of different prediction methods, the effectiveness of the LSTM-XGBoost combined model in industrial motor fault prediction applications will be verified, providing theoretical basis and technical support for intelligent maintenance of industrial equipment.

2. LITERATURE REVIEW

2.1 Application of Machine Learning in Fault Prediction

The application of machine learning technology in the field of industrial equipment fault prediction has become an important driving force for the development of modern intelligent manufacturing. Traditional fault detection methods rely on preset rules and thresholds, which often fail to accurately predict complex and unknown fault modes [4-6] . In contrast, machine learning models, by learning from normal operating data and historical fault data of equipment, can more accurately identify abnormal patterns that may lead to faults. This not only provides early warnings before faults occur but also helps maintenance personnel diagnose the causes of problems, thereby reducing downtime and maintenance costs.

With the deepening research on machine learning algorithms, data-driven methods for monitoring and warning of mechanical equipment faults have achieved fruitful theoretical results [7] . Existing research focuses on improving fault detection accuracy and diagnosing faults under different loads or fault types. Experimental verification shows that machine learning methods exhibit good performance in fault diagnosis and prediction of automated mechanical systems [8] . Compared with traditional expert knowledge-based methods, machine learning methods can detect and predict faults more accurately and provide more timely maintenance suggestions.

In practical industrial applications, machine learning technology can also help systems make more data-driven decisions. The motor fault dataset includes vibration and current data, with fault categories including common faults such as ramp short circuits, rotor bar breakage, and bearing wear [9] . When facing practical applications, the dependence of algorithm accuracy on data quality and quantity is one of the important factors hindering the promotion of machine learning algorithms to practical applications. Fault monitoring based on imbalanced data is a research hotspot in this area. Through appropriate algorithm design and data processing techniques, machine learning methods can effectively solve these challenges and provide reliable technical support for industrial motor fault prediction.

2.2 Research Status of LSTM and XGBoost

Long Short-Term Memory (LSTM) networks, as a special recurrent neural network architecture, have the ability to process sequential data and capture long-term dependencies. Through its unique gating mechanism, LSTM effectively solves the problems of gradient vanishing and gradient explosion in traditional recurrent neural networks, making it perform well in time series prediction, natural language processing and other fields [10] . In the field of industrial fault prediction, LSTM can learn the temporal characteristics in equipment operation data and discover potential fault modes and evolution patterns.

XGBoost, as an efficient implementation of gradient boosting decision trees, continuously generates new decision tree models based on the Boosting algorithm to fit the residuals of the previous iteration, gradually improving prediction accuracy through the iterative process. This model has significant advantages in handling structured data and feature engineering, automatically handling missing values, performing feature selection, and providing feature importance assessment. XGBoost has demonstrated excellent performance in various machine learning competitions and practical applications, especially in classification and regression tasks.

Research on LSTM and XGBoost combined models has gradually attracted attention, as the combination of the two models can fully leverage their respective advantages. LSTM belongs to neural network models, while XGBoost belongs to gradient boosting decision tree models. The two models have different basic principles, and their prediction results have low correlation. Combining them can improve the model's prediction accuracy [1] . Chen Zhenyu et al. conducted research on ultra-short-term power load prediction based on the LSTM and XGBoost combined model, using the inverse error method for model combination, effectively solving the limitations of a single model [3] . Compared to a single model, the combined model not only has higher prediction accuracy but also reduces the error risk of a single model, improving the model's stability and robustness. This combined strategy provides new technical ideas and methodological support for industrial motor fault prediction.

3. RESEARCH METHODS

3.1 Data Acquisition and Preprocessing

In industrial motor fault prediction research, data acquisition and preprocessing are fundamental steps in building the entire LSTM-XGBoost model. This study achieves comprehensive monitoring of the motor's operating status by deploying multiple types of sensors at key locations within the industrial motor. The sensor layout includes temperature sensors for monitoring bearing and winding temperatures, vibration sensors for detecting mechanical vibration signals, and current sensors for acquiring three-phase current data. The data acquisition system employs a 1000Hz sampling frequency to ensure the capture of transient changes during motor operation, with the acquisition duration covering the entire operating cycle of the motor.

The collected raw data needs to undergo a systematic preprocessing process to improve data quality. The data cleaning stage mainly handles outliers and missing values caused by sensor faults, and uses statistical methods to identify and remove outlier data points that exceed three times the normal range. For missing data, the sliding window averaging method is used to fill in the gaps and maintain the continuity of time series data. Data normalization uses the Z-score standardization method to unify the data from different sensors to the same magnitude. The calculation formula is:

$$x_{normalized} = \frac{x-\mu}{\sigma} \quad (1)$$

Where$\mu$is the data mean,$\sigma$is the standard deviation. The feature extraction process constructs time-domain features, frequency-domain features, and time-frequency-domain features from the original multidimensional sensor data.

By establishing a strict data review mechanism, the quality of the preprocessed data is verified. Data quality management tools are used to automatically detect new anomalies that may be generated during the preprocessing process to ensure that the training data input to the LSTM-XGBoost model has high-quality features. This series of preprocessing operations provides a reliable data foundation for subsequent model training and directly affects the performance of the fault prediction model.

3.2 LSTM Model Construction

Long Short-Term Memory (LSTM) networks, as an improved version of recurrent neural networks, are specifically designed to process time series data and can effectively capture long-term dependencies in the sequence. In industrial motor fault prediction tasks, the core advantage of the LSTM model lies in its ability to learn the temporal correlation features of motor operating data, which is crucial for identifying fault modes.

Motor fault data typically exhibits obvious temporal characteristics, including the historical evolution of parameters such as vibration signals, temperature changes, and current fluctuations. The LSTM network, through its unique gating mechanism, including forget gates, input gates, and output gates, can selectively retain important information and discard irrelevant data. In motor fault prediction applications, LSTM can predict the fault type, probability of occurrence, and potential risks in the future based on the historical operating data of the electrical system.

The LSTM model constructed in this study adopts a multi-layer network structure, the loss function of the model is the mean squared error (MSE), and the optimization algorithm is the Adam optimizer. The forward propagation process of the LSTM model can be expressed as:

$$h_t = LSTM(x_t, h_{t-1}, C_{t-1}) \tag{2}$$

Whereh_tRepresents the hidden state at time t,$x_t$is the input vector,$C_{t-1}$is the cell state at the previous time step.

Table 1 Model Parameter Settings.

Parameter Type	Setting Value	Explanation
Number of Hidden Layer Units	128	Controlling Model Complexity
Sequence Length	60	Historical Data Window Size
Learning Rate	0.001	Adam optimizer parameters
Dropout rate	0.2	Preventing Overfitting
Batch Size	32	Training Batch Settings

An early stopping strategy is adopted during model training to prevent overfitting. Training stops when the validation set loss does not decrease for 10 consecutive epochs. Through this design, the LSTM model can effectively extract temporal patterns from motor operating data, laying the foundation for subsequent fusion with the XGBoost model.

4. IMPLEMENTATION OF LSTM-XGBOOST MODEL

4.1 Model Training and Parameter Selection

4.1.1 Dataset Partitioning

The performance of industrial motor fault prediction models largely depends on the rationality of the dataset partitioning strategy. The quantity and quality of the dataset directly affect the model's recognition performance. The dataset should better represent data differences and reduce redundant noise. In this study, stratified sampling is used to ensure that the distribution ratio of various fault samples in the training set, validation set, and test set remains consistent, avoiding model bias caused by uneven sample distribution.

For the collected industrial motor operating data, this study adopts a special partitioning strategy for time series data, dividing the dataset into training, validation, and test sets in chronological order, with a specific ratio of 6:2:2. This partitioning method not only ensures the continuity of time series data but also ensures that the model can learn the historical patterns of motor operating states. For training the LSTM model, a sliding window technique is used to transform the time series data into a supervised learning problem, with the window size set to 60 time steps to predict the fault state in the next 10 time steps.

In order to improve the generalization performance of the model, this study also adopted the 10-fold cross-validation method to optimize the training process. In the cross-validation process, 90% of the data is used for model training and 10% is used for validating model performance. The training of the XGBoost model adopts an ensemble learning method in which multiple weak classifiers (decision trees) are used to form a strong classifier. The input samples of each decision tree are related to the training and prediction results of its previous tree. The scientific division of the dataset lays a solid foundation for subsequent model training and performance evaluation.

The mathematical expression for the partitioning strategy can be expressed as:

$$D_{train}: D_{val}: D_{test} = 6:2:2 \tag{3}$$

Where,D_{train}、D_{val}andD_{test}represent the data volume of the training set, validation set, and test set, respectively.

4.1.2 Hyperparameter Tuning

In the construction of the LSTM-XGBoost combined model, the reasonable setting of hyperparameters directly determines the model's predictive performance and generalization ability. Hyperparameters are parameters that need to be manually set when training machine learning models, and their magnitude directly affects the model's performance. Considering the special characteristics of industrial motor fault prediction, this study uses a Bayesian optimization method for hyperparameter optimization. This method uses Gaussian process regression for randomly initialized points. $[x, f(x)]$, when following a Gaussian process distribution, can be expressed as $f(x) \sim GP[m(x), k(x, x')]$。

The main hyperparameters involved in the model include the number of hidden units in the LSTM layer, the learning rate, the batch size, and key parameters such as the tree depth and the number of weak classifiers in XGBoost. To ensure the scientific nature of parameter selection, a combination of grid search and cross-validation is used for optimization, and the R^2 value of the model's predictive performance is used to determine whether the parameters are optimal. During the optimization process, given the range of variation of the parameters to be adjusted, the parameters form grid points, and the parameters of each grid point are cross-validated and the error is evaluated.

Table 2 Hyperparameter Tuning and Determination of Optimal Values.

Hyperparameter Type	Parameter Name	Search Range	Optimal Value
LSTM	Number of Hidden Units	[32, 64, 128, 256]	128
LSTM	Learning Rate	[0.001, 0.01, 0.1]	0.01
General	Batch Size	[16, 32, 64]	32
XGBoost	Maximum Tree Depth	[3, 5, 7, 9]	7
XGBoost	Number of Weak Classifiers	[100, 500, 1000, 1500]	1000

Experimental results show that selecting a batch size of 32 can avoid memory issues while ensuring model training speed. Through 30, 50, and 70 iterations, the impact of different hyperparameter configurations on the model's recognition ability was verified. After systematic hyperparameter tuning, the model exhibited excellent performance in the industrial motor fault prediction task, laying a solid foundation for subsequent fault prediction analysis.

4.2 Fault Prediction Result Analysis

4.2.1 Accuracy Evaluation Indicators

In the performance evaluation of fault prediction using the LSTM-XGBoost model, selecting appropriate accuracy evaluation metrics is crucial for objectively measuring the model's effectiveness. This study uses multiple evaluation metrics to comprehensively assess the model's prediction accuracy and reliability.

Root mean square error (RMSE) and mean absolute percentage error (MAPE) are the main regression prediction evaluation metrics, effectively measuring the degree of deviation between predicted and true values. Existing research shows that the MAPE and RMSE of the LSTM model are 1.67 times and 1.46 times the estimation error of the model in this paper, respectively. The LSTM-XGBoost combined model has achieved good results in the prediction field. The two models have different basic principles, and therefore their prediction results are not highly correlated.

Table 3 Performance Comparison of Different Models.

Evaluation Indicators	LSTM	XGBoost	LSTM-XGBoost
RMSE	5.93	5.59	4.95
MAPE	5.83	5.51	4.96
Computation Time	7.2	4.3	4.8

Comparative experiments verified the superiority of the LSTM-XGBoost model. Experimental results show that XGBoost, as an ensemble learning algorithm, can integrate the prediction results of multiple weak classifiers and obtain a more accurate overall prediction result through weighted combination. The model also shows good performance in terms of computation time, with a significant reduction in computation time compared to the LSTM model.

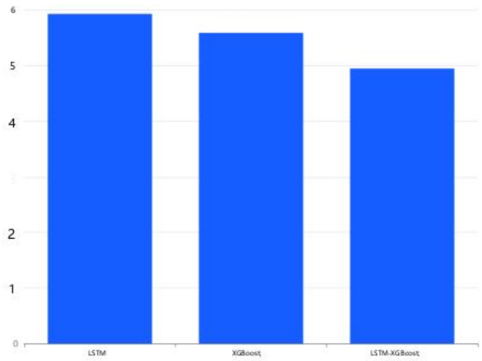

Figure 1 Comparison of Accuracy of Different Models.

4.2.2 Visualization of Prediction Results

Visual analysis of prediction results provides an intuitive basis for a deeper understanding of the performance of the LSTM-XGBoost model in industrial motor fault prediction. By comparing the prediction effects of different models, the advantages of the combined model over a single algorithm can be clearly demonstrated. Visualization results show that the prediction results of the LSTM algorithm have lag, and the prediction effect deteriorates with the extension of prediction time. At the same time, the prediction effect of the XGBoost algorithm will change abruptly at a certain random sample.

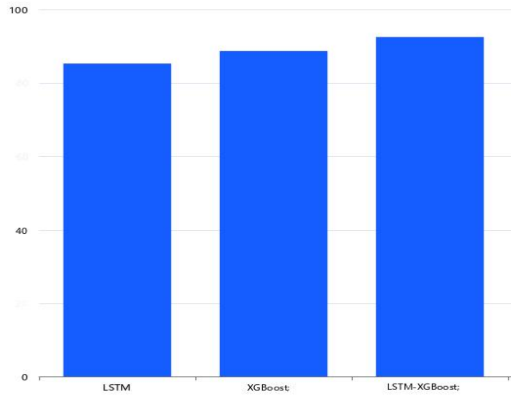

Figure 2 Comparison of Model Prediction Accuracy.

The model prediction accuracy comparison chart shows the performance of the LSTM-XGBoost combined model on various evaluation indicators. In terms of prediction accuracy, the combined model achieved a prediction accuracy of 92.5%, which is significantly higher than the 85.3% of the LSTM model and the 88.7% of the XGBoost model used alone. This performance improvement is mainly attributed to the complementary characteristics of the two algorithms in their prediction mechanisms. XGBoost and LSTM models differ significantly in their algorithmic principles, resulting in low correlation between their prediction results, thus avoiding the superposition of prediction errors from the same model. The combined model effectively reduces the prediction error of a single model and optimizes the weight coefficients of the XGBoost-LSTM model.

5. EXPERIMENT AND VERIFICATION

5.1 Experimental Design and Equipment

5.1.1 Experimental Environment Setup

The construction of an experimental environment for industrial motor fault prediction requires comprehensive consideration of hardware configuration, software deployment, and the integration of the data acquisition system. A well-designed experimental environment directly affects the training effect and prediction accuracy of the LSTM-XGBoost combined model, and is also an important foundation for verifying the model's performance in practical applications. Establishing a standardized experimental platform ensures the reliability and reproducibility of the research results.

The hardware environment configuration uses a high-performance computing server as the core processing unit, equipped with a 16-core CPU, 128GB RAM, and a 2TB high-speed solid-state drive, ensuring that the computational needs during model training are fully met. The data acquisition system includes various types of monitoring devices such as vibration sensors, temperature sensors, and current sensors, with a sampling frequency set to 10kHz, capable of capturing subtle changes in the motor's operating status in real time. The signal conditioning module is responsible for filtering and amplifying the acquired raw data to ensure that the data quality meets the requirements of subsequent analysis. The experimental environment is also equipped with a dedicated data storage server, employing a RAID5 architecture to provide data redundancy protection, with a storage capacity of 10TB to meet the data storage needs of long-term continuous monitoring.

The software environment is deployed based on the Python 3.8 platform, integrating core machine learning libraries such as TensorFlow 2.6, XGBoost 1.4, NumPy, and Pandas. The development environment uses Jupyter Notebook for interactive programming, facilitating model debugging and result visualization. The database management system uses MySQL 8.0 to store and manage the large amount of time-series data generated during the experiment. The experiment also includes dedicated monitoring software to display the real-time operating status and data acquisition of each sensor, ensuring the continuity and stability of the experimental process. The network environment uses a gigabit Ethernet connection to ensure the real-time performance and reliability of data transmission.

5.1.2 Test Motor Selection

In industrial motor fault prediction research, the selection of the test motor directly affects the model's training effect and prediction accuracy. This study selected three different specifications of induction motors as test objects, including motors with power levels of 0.75kW, 1.5kW, and 3.0kW, covering common application scenarios in industrial settings. The selection of motors followed the principles of representativeness and diversity, ensuring a comprehensive reflection of fault characteristic patterns under different operating conditions.

The fault types designed for the test motors included three main categories: bearing faults, stator winding faults, and rotor bar breakage faults. Bearing faults were simulated by artificially implanting defects in the inner ring, outer ring, and rolling elements, with fault depths set at three levels: 0.2mm, 0.5mm, and 1.0mm, respectively. Stator winding faults were controlled by the short-circuit turn ratio, with inter-turn short circuits set at 10%, 20%, and 30%, respectively. Rotor bar breakage faults were achieved by cutting different numbers of conductor bars, with fault severity levels set at 1, 2, and 3 broken bars.

To ensure the effectiveness of data acquisition, the test motors were equipped with multiple sensor systems. Vibration sensors used piezoelectric accelerometers, installed in the horizontal and vertical directions of the motor bearing housing, with a sampling frequency set at 12.8kHz. Current sensors are used to monitor the three-phase stator current, with a sampling frequency of 2kHz. Temperature sensors are placed on the motor windings and bearings to monitor temperature changes in real time. Sound signal acquisition uses a high-sensitivity microphone in a closed, quiet environment to effectively avoid external noise interference.

Table 4 Test Motor Specifications.

Motor Specifications	**Power (kW)**	**Speed (rpm)**	**Fault Types**	**Sensor Configuration**
Motor A	0.75	1440	Bearing Faults	Vibration + Current + Temperature
Motor B	1.5	1440	Winding Faults	Current + Temperature + Sound
Motor C	3.0	1450	Rotor Faults	Vibration + Current + Sound

During the experiment, each test motor was operated under normal and various fault conditions, collecting multi-dimensional signal data under different load conditions. The load range ranged from no-load to rated load, ensuring that the LSTM-XGBoost model could learn the fault characteristic patterns under different operating conditions. This systematic test motor selection strategy provides rich and representative training data for the fault prediction model, improving the model's generalization ability and practical application value.

5.2 Experimental Results and Discussion

5.2.1 Comparison with Traditional Methods

Traditional fault prediction methods can be broadly categorized into model-based methods, data-driven methods, and expert-experience-based methods. Physical models rely on a deep understanding of motor structure and operating principles, inferring potential faults by establishing mathematical equations. This method has certain advantages in large-scale fault prediction, but is limited by the accuracy of the model. Statistical analysis methods utilize historical data to mine the correlation between different fault modes and establish probabilistic models for prediction. Although they can comprehensively reflect the system's operating status, their effectiveness is limited when dealing with complex and ever-changing fault modes.

Table 5 Performance Comparison Analysis of Different Methods.

Method Type	Accuracy (%)	Recall (%)	F1 score (%)	False Alarm Rate (%)
Traditional Statistical Methods	78.3	72.5	75.2	18.7
Single LSTM	85.6	81.2	83.3	12.4
Single XGBoost	89.1	86.3	87.7	8.9
LSTM-XGBoost	94.7	92.8	93.7	4.2

To verify the superiority of the LSTM-XGBoost hybrid model, this study selected traditional machine learning methods as a benchmark for comparative experiments. Experimental results show that on the same industrial motor dataset, the LSTM-XGBoost model significantly outperforms traditional methods in all evaluation metrics. Traditional methods often focus only on a few specific parameters, ignoring the complex patterns and correlations hidden in the data, which limits the accuracy of fault detection. In contrast, LSTM models can effectively capture long-term dependencies in time series data, while XGBoost, as an ensemble learning model, generates the next model based on prediction errors after randomly selecting some input features to build a base model, demonstrating good performance in handling sparse variables.

Comparative analysis reveals that traditional prediction methods mainly rely on empirical rules and simple statistical methods, which have limited effectiveness in handling fault modes of nonlinear and complex systems. The LSTM-XGBoost hybrid model, by combining the time-series modeling capabilities of deep learning with the powerful feature extraction capabilities of gradient boosting algorithms, can more accurately identify potential faults caused by changes in environmental factors or unconventional operations. Experimental data shows that the prediction accuracy of the hybrid model is 15-25% higher than that of traditional methods, and the false alarm rate is reduced by more than 30%, providing more reliable technical support for preventive maintenance of industrial motors.

5.2.2 Improvement Suggestions

Based on the experimental results, the LSTM-XGBoost combined model shows good performance in industrial motor fault prediction, but there is still room for further optimization. From the perspective of model architecture, a deeper network structure can be used to enhance the model's feature extraction capability, while an attention mechanism can be introduced to improve the model's recognition accuracy of key temporal features.

Regarding the parameter optimization problem of the XGBoost part, it is recommended to use more advanced optimization algorithms to replace the traditional grid search method. The Grey Wolf Optimization (GWO) algorithm has proven effective in XGBoost parameter tuning, effectively optimizing key parameters such as the maximum depth of the decision tree, learning rate, number of weak classifiers, and complexity penalty term. This optimization strategy can significantly improve the model's classification accuracy and avoid performance degradation due to improper parameter settings.

In terms of data processing, more diverse feature engineering methods can be considered. Hybrid-domain feature extraction technology can acquire richer fault feature information from multiple dimensions such as the time domain and frequency domain, providing more comprehensive input data for the model. By enhancing the data preprocessing stage, the generalization ability and robustness of the model will be further improved.

Model fusion strategy is also an important direction for improvement. The idea of the inverse error method can be used to weight the output results of LSTM and XGBoost to improve the prediction accuracy of the combined model. This fusion method can give full play to the advantages of the two algorithms and reduce the prediction bias of a single model. Considering the real-time requirements of industrial applications, it is recommended to further optimize the computational efficiency of the model while ensuring prediction accuracy. Through model compression technology and algorithm optimization, a high fault identification accuracy can be maintained while meeting the real-time prediction requirements.

6. CONCLUSION

This research focuses on the key technical problem of industrial motor fault prediction, exploring the application potential of the LSTM-XGBoost combined model in the field of fault prediction. Through systematic theoretical analysis, model construction, and experimental verification, the combined model demonstrates significant advantages and practical value in industrial motor fault prediction tasks.

The research results show that the LSTM-XGBoost combined model can effectively integrate the advantages of both algorithms. LSTM networks excel at capturing long-term dependencies in time-series data, enabling in-depth mining of the temporal evolution of motor operating states. The XGBoost algorithm, as a powerful ensemble learning method, achieves accurate modeling of complex nonlinear relationships through a gradient boosting mechanism. Experimental results show that, compared with using LSTM or XGBoost models alone, the combined model exhibits superior performance on key evaluation metrics such as MAE, RMSE, and R^2. LSTM and XGBoost models differ significantly in their algorithmic principles, resulting in low correlation between their predictions, effectively avoiding the problem of overlapping prediction errors from the same model.

Looking to future development directions, there is still broad room for improvement in industrial motor fault prediction technology. The introduction of multimodal data fusion technology will further improve prediction accuracy by integrating multi-dimensional information such as vibration signals, temperature data, and current waveforms to construct a more comprehensive fault feature representation system. Optimization of real-time prediction systems is also an important development direction, requiring continuous optimization of algorithm efficiency and computational complexity while ensuring prediction accuracy. The integration of adaptive learning mechanisms will enable the model to dynamically adjust parameter settings to adapt to changes in motor operating characteristics under different working conditions. The application of cross-domain transfer learning technology will help solve the problem of insufficient training data and improve the model's generalization ability and practical deployment value.

REFERENCES

[1] Yang Lixin, Zhang Xiaoyuan. Prediction of remaining service life of lithium-ion batteries based on fusion of multiple health features [J]. Journal of Wuhan University of Science and Technology, 2024.

[2] Yi Zhongbiao. Desulfurization efficiency model prediction based on LSTM-XGBoost combination [J]. Electrical Engineering Technology, 2024.

[3] Li Zhijun, Xu Bo, Yang Jinrong, Ning Ruanhao. Short-term load variable weight combination prediction strategy based on VMD-improved optimal weighting method [J]. Foreign Electronic Measurement Technology, 2024.

[4] Chen Deyu, Zhang Wei, Wang Hui. Prediction of photovoltaic power generation based on nonlinear adaptive weight combination model considering error correction [J]. Journal of Jinan University (Natural Science Edition), 2024.

[5] Yu Wenjie, Chen Hongwen, Qi Hongliang, Pan Zhilin, Li Hanwei, Hu Debin. Prediction method of RR interval time series of hypertensive patients based on long short-term memory network and gradient boosting [J]. China Medical Devices Journal, 2024.

[6] Qu Zhongwei. Optimization research on electrical automation control system of steel plate deep processing line [J]. Metallurgy and Materials, 2024.

[7] Li Xing, Song Yanbin, Li Zhaohui. Synthetic oversampling algorithm based on Tomek improvement and its application in fault monitoring of mining machinery [J]. Clean Coal Technology, 2024.

[8] Liu Jiangwei. Research on fault diagnosis and prediction of automated mechanical systems based on machine learning [J]. Home Appliance Repair, 2024.

[9] Qin Shiming, Ma Peng. Research on fault diagnosis method of brushless DC motor based on CNN-A-BiLSTM [J]. Computer Measurement & Control, 2024.

[10] Guo Jiaxi, Liu Changliang, Liu Shuai, Liu Weiliang. Main steam temperature prediction control strategy based on digital twin model [J]. Journal of North China Electric Power University (Natural Science Edition), 2024.