Advances in Engineering Research : Possibilities and Challenges
ISSN Print: 3079-5192
ISSN Online: 3079-5206
https://doi.org/10.63313/AERpc.9058

Multivariate Data Fusion and Health Analysis of Coal Processing Facilities

Xiaolin Duan
Henan Polytechnic University, Henan, 454003, China

How to cite this paper: Duan, X. L. (2025). Multivariate Data Fusion and Health Analysis of Coal Processing Facilities. Advances in Engineering Research: Possibilities and Challenges, 2(3), 65-76. ISSN Print: 3079-5192; ISSN Online: 3079-5206.
https://doi.org/10.63313/AERpc.9058
Published: 2025-11-10

Abstract

This paper presents a comprehensive health diagnosis and predictive maintenance management system based on multi - source data fusion. Firstly, the research constructs a data acquisition framework covering multi - source sensors such as vibration, temperature, and pressure. Wavelet transform and empirical mode decomposition are employed for signal processing, and multi - dimensional features in the time domain, frequency domain, and time - frequency domain are extracted. Subsequently, the local preserving projection algorithm is introduced to achieve effective dimensionality reduction and visualization of high - dimensional features. At the diagnostic level, a fuzzy clustering - fuzzy integral fusion diagnostic algorithm based on fuzzy measure theory is innovatively proposed to address the challenge of composite fault identification. At the predictive level, by constructing a degradation feature evaluation system encompassing correlation, monotonicity, and robustness, the self - organizing map network is used to fuse multiple features into a comprehensive health index, and the long - short - term memory network model is utilized to accurately predict the evolution trend of the equipment's health status. Finally, based on the health status assessment and remaining useful life prediction, a full - life - cycle predictive maintenance strategy is formulated, and an intelligent maintenance decision - support system is designed.

Keywords

Coal processing plant; Multi-data fusion; LSTM health status prediction; Long Short-Term Memory Network; Optimization of maintenance decisions

1. Introduction

The intelligent operation and maintenance of industrial equipment is one of the core driving forces of "Made in China 2025" and Industry 4.0. As a key link in the clean and efficient utilization of coal, coal processing facilities (such as pulverized coal gasifiers) play an irreplaceable strategic role in the fields of energy, chemical industry, and metallurgy. These facilities are usually large-scale, complex in structure, and operate under harsh conditions of high temperature, high pressure, strong corrosion, and high wear for long periods. Their key transmission components (such as the chains, gears, and bearings of the slag remover) are prone to performance degradation and sudden failures. Once a failure occurs, it not only causes unplanned shutdowns, resulting in significant production losses, but also may trigger serious safety and environmental accidents.

Currently, in most industrial sites, the traditional approach of implementing planned maintenance based on fixed cycles and corrective maintenance after faults still prevails. The former often leads to "excessive maintenance", wasting maintenance resources and causing unnecessary downtime; the latter is a "passive response" that cannot prevent the serious consequences of faults. This traditional maintenance model is no longer able to meet the stringent requirements of modern industrial enterprises for high equipment availability, high safety, and low operating costs.

With the rapid development of Internet of Things, big data and artificial intelligence technologies, predictive maintenance has emerged. It precisely carries out maintenance activities before faults occur by monitoring the real-time operating status of equipment and making trend predictions. Thus, it achieves a paradigm shift from "treating diseases after they occur" to "preventing diseases before they happen". During this process, multi-data fusion technology has become the key to achieving precise state perception and intelligent diagnostic prediction. A single sensor's information dimension is limited and it is difficult to comprehensively depict the health status of complex equipment. By integrating data from various heterogeneous sensors such as vibration, temperature, pressure, and current, a more comprehensive and complementary equipment state profile can be formed, significantly improving the accuracy of fault diagnosis and the reliability of state prediction.

Although scholars both at home and abroad have achieved significant results in fields such as fault diagnosis and data fusion, for instance, Ben Ali et al. combined empirical mode decomposition with neural networks for bearing diagnosis, and Chen et al. developed a CNN-LSTM hybrid model for fault diagnosis in chemical processes, the existing research still has the following shortcomings: (1) The research objects are mostly concentrated on general rotating machinery (such as bearings, gearboxes), and there are relatively few systematic research cases specifically targeting core process equipment of coal processing plants (such as slag removal machines); (2) For the diagnosis of complex faults involving multiple concurrent fault modes and feature coupling, solutions are not yet mature; (3) The complete framework for integrated research on health diagnosis, state prediction, and maintenance decision optimization is relatively rare.

In view of this, this research takes the slag remover of a typical coal gasification unit - the pulverized coal gasifier as a specific case, aiming to construct a complete technical system and management framework that integrates "multivariate data fusion", "intelligent health diagnosis", "precise condition prediction" and "optimization maintenance decision-making". This is intended to address the pain points of the traditional maintenance model and provide a replicable theoretical model and practical path for the digital transformation and intelligent upgrade of industrial enterprises.

2. Problem Description and Model Assumptions

2.1. Core Problem Description

The health management of coal processing facilities involves a multi-dimensional and interdisciplinary complex system engineering issue. The core difficulties can be summarized as follows:

① Issues regarding the comprehensiveness and accuracy of status perception: The operational status information of equipment is scattered across various physical signals such as vibration, temperature, pressure, and current. These data possess characteristics such as multiple sources, heterogeneity, high dimensionality, and strong noise. How to effectively integrate and clean these data, and extract the features sensitive to equipment degradation from them, is the primary challenge.

② Problems related to the complexity and coupling of fault modes: In the actual operation of equipment such as slag removers, various faults such as chain wear, bearing failure, misalignment often do not occur independently but interact with each other and coexist simultaneously. Their vibration and temperature characteristics are superimposed and coupled in the signal space, resulting in a decline in the recognition rate of single fault diagnosis models and making the separation and identification of composite faults extremely difficult.

③ Uncertainties and long-term dependency in health status prediction: The performance degradation of equipment is a continuous, gradual, and trend-based process, subject to interference from factors such as working conditions and environmental changes, exhibiting nonlinear and non-stationary characteristics. Traditional time series prediction models struggle to capture its long-term evolution patterns, resulting in high uncertainty in prediction outcomes and affecting the lead time for maintenance decisions.

④ Multi-objective optimization and resource constraints in maintenance decisions: Maintenance decisions need to balance multiple conflicting goals such as equipment reliability, maintenance costs, production plans, spare parts inventory, and human resources. How to determine the optimal maintenance timing, maintenance strategy, and resource allocation plan under the constraints of limited resources is a typical NP-hard problem.

2.2. Model Framework and Core Hypotheses

To address the aforementioned issues, this study proposes the following overall research framework and core model assumptions:

Overall framework: Develop a closed-loop intelligent operation and maintenance framework from "data collection" to "decision execution". This framework consists of the following steps: multi-source data collection and preprocessing, multi-domain feature extraction and dimension reduction, intelligent health diagnosis, health status prediction,maintenance strategy formulation and optimization.

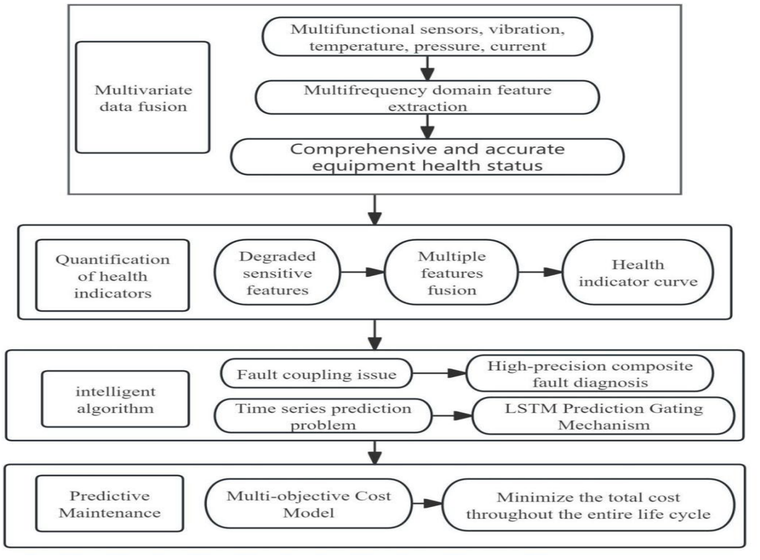

Figure 1. Technical implementation path and method

Based on the above assumptions, this study will successively carry out the core contents such as the construction of the data fusion management framework, the design of health diagnosis methods, the development and maintenance of state prediction models, and the optimization of maintenance strategies.

3. Content Analysis and Verification

3.1. Multi-data Fusion Health Diagnosis Methodology System

3.1.1. Data Fusion Management Framework

This study designed a hierarchical data fusion management framework. At the data collection layer, ICP acceleration sensors, PT100 temperature sensors, pressure transmitters, and Hall current sensors were deployed at key parts of the slag remover's motor, reducer, and guide wheel, forming a distributed monitoring network. The data processing layer is responsible for data synchronization, standardization, anomaly handling, and missing value filling to ensure data quality. The feature extraction layer is the core. We systematically extracted three major types of features from the vibration signals:

Time-domain characteristics: include 16 parameters such as mean, RMS value, peak value, kurtosis, skewness, pulse factor, margin factor, etc., which reflect the basic statistical characteristics and impact energy of the signal.

Frequency-domain characteristics: include characteristic frequencies such as center frequency, root mean square frequency, frequency variance, spectral kurtosis, and trans-frequency, mesh frequency, etc., whose amplitudes reveal the frequency structure of the signal.

Time-frequency domain characteristics: using wavelet packet decomposition and empirical mode decomposition, calculate the proportion of energy in each frequency band, wavelet packet entropy, IMF energy entropy, etc., to depict the energy distribution and complexity of the signal in the joint time-frequency domain.

3.1.2. Feature Dimension Reduction and Visualization

The initially extracted feature dimensions were too high, and there was a lot of redundant information. In this study, the local manifold projection algorithm was used for nonlinear dimensionality reduction. The core of the LPP algorithm lies in constructing an adjacency graph and optimizing the objective function to ensure that adjacent sample points in the high-dimensional space still remain adjacent in the low-dimensional space after reduction, thereby effectively preserving the local manifold structure of the data. As shown in the verification results, as presented in Table 2.

Table 2. Evaluation of LPP Dimension Reduction Effectiveness

Evaluation indicator	Original	After LPP	Comparision
Intra-class dispersion	15.67	8.34	46.8%
Inter-class dispersion	45.23	38.91	-14.0%
Separation index	2.89	4.67	61.6%
Compactness index	0.67	0.42	37.3%
Neighborhood retention rate	-	0.87	-
Visual quality	-	8.5/10	-

After dimensionality reduction by LPP to 2 dimensions, the intra-class dispersion decreased by 46.8%, while the separation index increased by 61.6%. As shown in Figure 3, different fault modes formed clearly defined clusters in the two-dimensional space, laying a solid foundation for subsequent diagnosis.

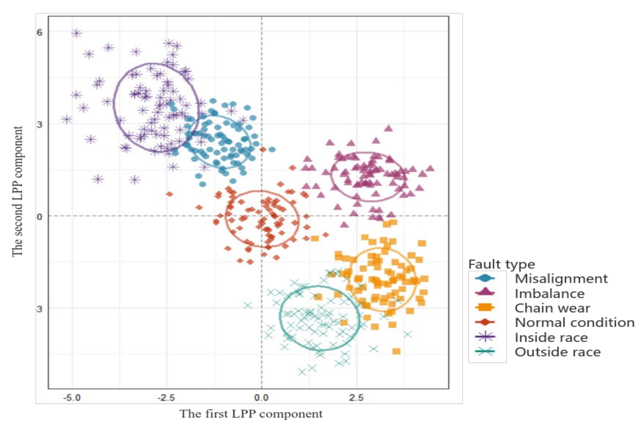

Figure 3. Distribution of Fault Modes Following LPP Dimension Reduction

3.1.3. FCM-FI Integrated Diagnostic Model

In view of the limitations of a single classifier in complex fault diagnosis, this paper proposes a fuzzy C-means clustering - fuzzy integral fusion diagnosis model.
Step 1: Base classifier training and preliminary diagnosis. Firstly, six classic classifiers such as support vector machine, random forest, and BP neural network are used to train on the reduced feature set, and the respective preliminary diagnosis results are obtained.
Step 2: Fuzzy measure calculation. This is the innovation point of the method. The fuzzy measure is different from the traditional weights. It measures the "importance" of individual classifiers and any combination of them for the final diagnosis decision. We adopt the Sugeno λ-measure, and by solving the specific fuzzy measure equation, the measure values of each classifier and its combinations are obtained. The calculation results are shown in Table 4.

Table 4. Fuzzy Measure Calculation Results

CC	FMV	WC	CD(%)	RR	OBC
{SVM}	0.893	0.285	28.5	0	-
{RF}	0.871	0.242	24.2	0	-
{NN}	0.856	0.218	21.8	0	-
{NB}	0.824	0.165	16.5	0	-
{KNN}	0.839	0.180	18.0	0	-
{DT}	0.817	0.152	15.2	0	-
{SVM,RF}	0.932	-	-	0.103	0.687
{SVM,NN}	0.924	-	-	0.127	0.642
{RF,NN}	0.918	-	-	0.145	0.598
{SVM,RF,NN}	0.951	-	-	0.089	0.723

CC,Classifier combination; FMV, Fuzzy measure value; WC, Weight coefficient; CD, Contribution degree; RR, Redundancy rate; OBC, Overall Bilateral Complementarity.
The individual fuzzy measure of SVM is the highest (0.893), while the combined measure of SVM and random forest reaches 0.932. Moreover, their complementary index is 0.687, indicating that the decision-making mechanisms of the two are significantly different. After integration, they can complement each other's strengths.
Step 3: Choquet fuzzy integral fusion. Based on the calculated fuzzy measures, the Choquet integral is used to fuse the output results of each classifier. The Choquet integral is a nonlinear integral that can fully consider the interaction (cooperation or redundancy) between classifiers, thereby making more reliable collective decisions.

3.1.4. Diagnostic Performance Verification

On the experimental platform of the slag remover that was set up, various working conditions such as normal operation, chain wear, faults in the inner and outer rings of the bearings, and misalignment were simulated. The verification results, as shown in Table 2, are very encouraging: the overall diagnostic accuracy of the FCM-FI fusion method reaches 95.7%, which is 6.4 percentage points higher than that of the most performing single SVM classifier (89.3%). Especially in the complex fault scenarios with severe feature crossover, the improvement in the recognition accuracy exceeds 15%. The confusion matrix heatmap, as shown in Figure5 clearly indicates that the misjudgment rates between different categories are extremely low, demonstrating the outstanding comprehensive diagnostic performance of this method.

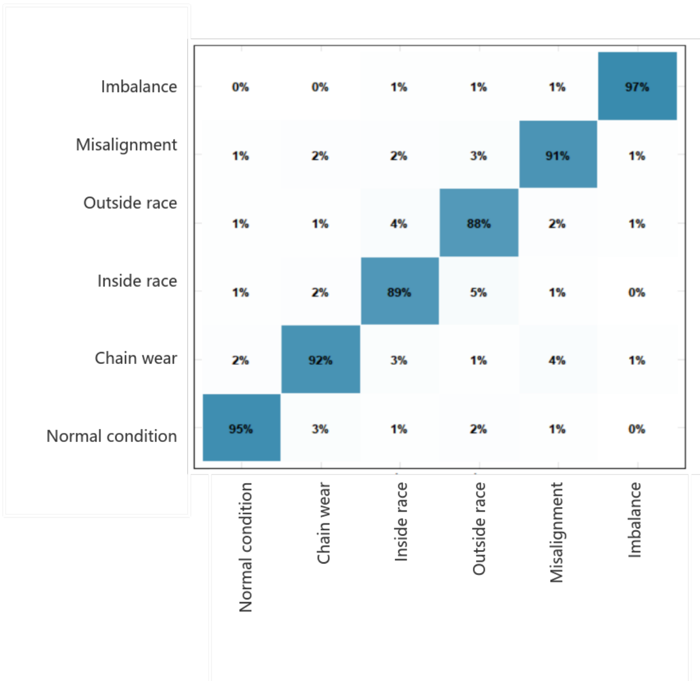

Figure 5. FCM-FI Method Confusion Matrix Heatmap

3.2. Construction and Validation of Health Status Prediction Model

3.2.1. Construction of Health Indicator Curve

The prerequisite for precise prediction is to establish a health indicator that can truly reflect the degradation process of the equipment. In this study, the degradation sensitivity of the multi-domain features extracted earlier was evaluated. The indicators included:

Correlation: The Pearson correlation coefficient between the feature value and the running time.

Monotonicity: The degree to which the feature value maintains a single increasing or decreasing trend over time.

Robustness: The ability of the feature value to resist noise interference and operational fluctuations.

Through weighted comprehensive scoring, the kurtosis, centroid frequency, effective value of vibration, high-frequency energy ratio, and wavelet packet entropy were selected as the five optimal degradation sensitive features.

Subsequently, an unsupervised fusion of these five features was performed using the Self-Organizing Map (SOM) network. The SOM network can map high-dimensional inputs onto a low-dimensional topological structure and measure the deviation between the current state and the normal state by calculating the minimum quantization error between the input vector and the weights of the best-matching unit. This MQE sequence constitutes the initial HI curve. To further enhance the smoothness and trend of the HI curve, wavelet packet decomposition was employed to extract its low-frequency trend components and perform normalization processing. Eventually, a smooth, monotonic HI curve with a value range of [0, 1] was obtained.

3.2.2. Health Status Prediction Based on LSTM

Considering the temporal dependence of the HI sequence, this study selects the Long Short-Term Memory (LSTM) network as the prediction model. LSTM can effectively control the flow of information through its carefully designed "forget gate", "input gate" and "output gate", avoiding the gradient disappearance/explosion problem in traditional recurrent neural networks when training long sequences. It is particularly suitable for handling long-term dependent problems such as equipment degradation.

The hyperparameters of LSTM (such as the number of hidden layer units, learning rate, batch size, sequence length, etc.) were optimized through grid search and Bayesian optimization. The optimal parameter combination was finally determined. The sliding window method was used to construct training samples, and the HI values of the previous 20 time steps were used to predict the HI value of the next step (single-step prediction) or multiple steps in the future (multi-step prediction).

3.2.3. Performance Verification of Prediction

The validation results on the test set indicate that the root mean square error of the single-step prediction of the LSTM model is 0.052, the average absolute percentage error is 4.67%, the coefficient of determination R^2 is as high as 0.973, and the accuracy of the prediction direction reaches 93.5%. These indicators all demonstrate that the model has extremely high prediction accuracy and trend capture ability.

To showcase the superiority of LSTM, we compared it with support vector regression models using different kernel functions (linear, RBF, polynomial) like Table6.

Table 6. The Performance Comparison of LSTM vs SVR Model

PI	LSTM	SVR-Linear	SVR-RBF	SVR-Poly	Compare
SSP-RMSE	0.052	0.089	0.067	0.078	29.0%
5SP0RMSE	0.089	0.156	0.134	0.145	33.6%
10SP0RMSE	0.128	0.234	0.198	0.223	35.4%
SSP-MAPE(%)	4.67	8.23	6.45	7.34	27.6%
5SP0MAPE(%)	8.34	14.67	12.89	13.78	35.3%
10SP0MAPE(%)	12.45	22.34	19.67	21.23	36.7%

TT(s)	156.7	45.6	78.9	67.3	-71.8%
PT(ms)	15.2	5.4	8.7	7.1	-74.6%
MU(MB)	89.4	23.5	34.7	28.9	-73.7%

PI,Performance Index;SSP-RMSE,Single-step prediction RMSE;5SP0RMSE,5-step prediction of RMSE;10SP0RMSE,10-step prediction of RMSE;SSP-MAPE，Single-step prediction MAPE；5SP0MAPE，5-step prediction of MAPE；10SP0MAPE，10-step prediction of MAPE；TT，Training Time；PT,predicted time；MU，Memory Usage.

The results showed that LSTM's RMSE in single-step, 5-step, and 10-step predictions was respectively 29.0%, 33.6%, and 35.4% lower than the optimal SVR model, and the advantage became more obvious as the prediction step length increased. This fully proves the powerful ability of LSTM in handling complex time series problems.

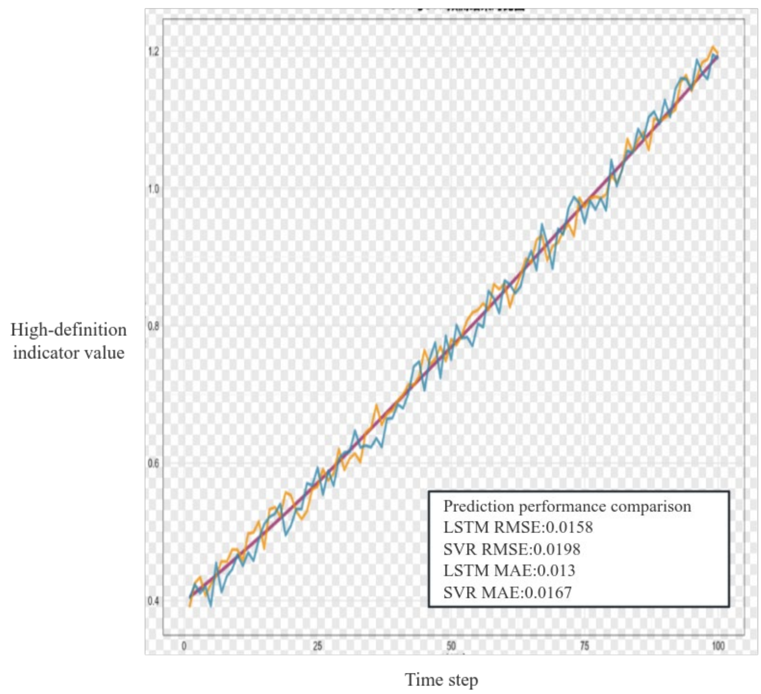

Figure 7. Comparison of LSTM and SVR prediction results

3.3. Design of Lifecycle Maintenance Strategy and Effect Evaluation

3.3.1. Construction of Maintenance Strategy System

Based on precise health status prediction, this study has constructed a complete lifecycle maintenance strategy system.

Health status classification: The equipment health status is divided into five levels: "Healthy, Good, Attention, Warning, and Dangerous", and corresponding HI thresholds (HI ≥ 0.8, 0.6, 0.4, 0.2) are set to provide clear guidance for maintenance actions

in different states.
Predictive Maintenance Decision Model: With remaining service life prediction as the core. The LSTM model is used to predict the future HI curve, and the intersection with the preset fault threshold (such as HI = 0.15) is found, which is the predicted fault time, and RUL is the duration from the current moment to that point. Combining RUL, fault risk, maintenance cost, and downtime loss, a multi-objective optimization function is established, and the genetic algorithm is used to solve the optimal maintenance timing.
Differentiated Maintenance and Inventory Optimization: Using the Failure Mode and Effects Analysis to calculate the risk priority number of each component, determine the maintenance priority. The ABC classification method is used to manage spare parts, and the (s, S) inventory strategy is implemented for high-value and critical A-class spare parts to balance inventory holding costs and shortage risks.

3.3.2. Verification of Maintenance Strategy Effect

To evaluate the actual effect of the proposed strategy, a 18-month on-site comparative experiment was designed. Three identical slag collectors were implemented with post-maintenance, planned maintenance, and predictive maintenance in the same workshopThe experimental results fully prove the significant advantages of predictive maintenance:
Equipment availability rate: The predictive maintenance group reached 96.4%, much higher than 82.3% of post-maintenance and 91.7% of planned maintenance.
Fault frequency: The annual fault frequency was 12 times for post-maintenance, 6 times for planned maintenance, and significantly reduced to 2 times.
Economic benefits: The annual maintenance cost was reduced by 39.0%, and downtime loss was reduced by 80.7%. After financial analysis, the net present value of the predictive maintenance system investment was as high as 15,473,000 yuan, the internal rate of return was 42.6%, the investment payback period was only 2.6 years, and the economic benefits were extremely significant.

4. Summary and Outlook

4.1. Research Summary

This study addressed the health management challenges faced by coal processing plants under complex operating conditions and successfully designed and verified an integrated intelligent operation and maintenance solution from state perception to decision execution. The main research results and conclusions are as follows:
An efficient multi-data fusion diagnostic method was proposed: By constructing a multi-domain feature system and LPP dimension reduction, the problem of high-dimensional data processing was solved; the innovative FCM-FI fusion diagnostic model precisely depicted the collaboration and complementarity relationship among multiple classifiers, increasing the accuracy of composite fault diagnosis to

95.7%, providing a new paradigm for resolving information conflicts and feature crossover issues.

A precise equipment health status prediction model was constructed: Through the SOM network integrating multiple degraded features to construct a health indicator curve, it has good monotonicity and robustness; the prediction model based on LSTM fully leveraged its advantages in handling long-term temporal dependencies, achieving a high accuracy of single-step prediction RMSE of 0.052, providing a reliable basis for predictive maintenance.

The outstanding effectiveness of the predictive maintenance strategy was verified: Through industrial experiments, it was confirmed that compared with the traditional mode, predictive maintenance can significantly improve equipment availability (up to 96.4%), significantly reduce the frequency of failures and maintenance costs (39.0%), and has excellent investment returns, providing a complete technical path and management framework for the transformation from "passive response" to "active prevention" in maintenance mode.

4.2. Future Outlook

Although this study achieved the expected results, there are still several directions worthy of in-depth exploration:

Technical deepening and integration: Future research can explore the application of graph neural networks to model the topological relationship of the equipment system, or introduce reinforcement learning to achieve adaptive optimization of maintenance strategies. Transfer learning and meta-learning can help quickly apply the model to new equipment or new operating conditions, enhancing generalization ability.

Integration of digital twin technology: Building a digital twin model that synchronizes with the physical equipment in real time, enabling fault simulation, maintenance plan verification and optimization in the virtual space, achieving true closed-loop intelligent operation.

Edge intelligent decision-making: With the development of 5G and edge computing technologies, lightweight diagnostic and prediction models can be deployed at the edge of the equipment, enabling real-time analysis with lower latency and local decision-making, improving system response speed and reliability.

Platformization and ecologization: Promoting predictive maintenance from a technical solution to a platform service, integrating equipment manufacturers, service providers and users, forming a collaborative innovation industrial ecosystem, further reducing costs through standardization and scale, and releasing data value.

In conclusion, this study not only provides a specific and feasible technical solution for intelligent operation and maintenance of coal processing plants, but also contributes valuable theoretical and practical references for the digital transformation of process industries. With the continuous maturation and integration of related technologies, intelligent operation and maintenance based on data will become a key pillar for enhancing the core competitiveness of the industrial sector.

References

[1] Xiaopeng Xie, Wei Feng, et al. Study on Condition Monitoring in Petrochemical Equipment Using Oil Analysis Technology[J]. Petrochemical Technology, 2018, 47(8): 23-29.

[2] Huihui Li, Linfeng Gou, et al. Intelligent Fault Diagnosis of Aeroengine Sensors Using Improved Pattern Gradient Spectrum Entropy[J]. International Journal of Aerospace Engineering, 2019, 2019: 1-15.

[3] V. Venkatasubramanian, R. Rengaswamy, et al. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods[J]. Computers & Chemical Engineering, 2003, 27(3): 293-311.

[4] Y. Lei. Deep Transfer Diagnosis Method for Machinery in Big Data Era[J]. Journal of Mechanical Engineering, 2019, 55(13): 1-8.

[5] J. Wang, X. Li, Z. Chen, et al. Attention-based multi-sensor fusion for autonomous systems[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(8): 3456-3468.

[6] Y. Li, J. Zhang, K. Wang, et al. Bayesian inference for multi-sensor data fusion in intelligent systems[J]. Expert Systems with Applications, 2019, 128: 245-258.

[7] Y. Chen, Z. Liu, X. Wang, et al. Support vector machine based multi-sensor classification for fault diagnosis[J]. Mechanical Systems and Signal Processing, 2019, 132: 244-259.

[8] Q. Wang, X. Liu, Y. Chen, et al. Adaptive filtering algorithms for robust multi-sensor data fusion[J]. IEEE Transactions on Signal Processing, 2020, 68: 3142-3155.

[9] César L. Melchor-Hernández, F. Rivas-Dávalos, S. Maximov, et al. A model for optimizing maintenance policy for power equipment[J]. International Journal of Electrical Power and Energy Systems, 2015, 68: 331-340.