Temperature Prediction for Electromechanical Equipment in Open-Pit Coal Mines Under Complex Working Conditions Using Wavelet Packet Decomposition and Graph Attention Network

Indexed by:
Web of Science Group

Feiyang Zhao[a], Yiping Yuan[a,*], Caifeng Chen[a], Lu Xiao[b]

[a] Xinjiang University, China
[b] Xinjiang Tianchi Energy Co., Ltd., China

Highlights

- The prediction model combines wavelet packet decomposition and graph attention network.
- Long-term and short-term features obtained by WPD of temperature series data.
- GAT can automatically capture features and assign the appropriate weights.
- The feasibility of the model is verified using belt conveyor motor temperature data.

Abstract

The electromechanical equipment in open-pit coal mines is influenced by perturbing factors such as load, ambient temperature, and frequent startups and shutdowns, which result in low accuracy and poor generalization performance of the prediction model for its operating state. This paper considers both long-term temperature fluctuations and episodic changes caused by these perturbing factors. Additionally, multidimensional time and spatial data are integrated to propose a temperature prediction model for mining electromechanical equipment based on wavelet packet decomposition and a graph attention network (WPD-GAT). Experimental validation is conducted using temperature data from electromechanical equipment in an open-pit coal mine in Xinjiang, with comparisons made to four other models. The results demonstrate that the proposed model outperforms the others to varying degrees, highlighting its feasibility and superiority for predicting the temperature of electromechanical equipment under complex working conditions.

Keywords

1. Introduction

Coal is a critical energy source and industrial raw material in China, playing an essential role in the national economy[1]. Open-pit mines are fundamental to China's coal industry and contribute significantly to national energy security[2]. By the end of 2022, China's open-pit coal mines produced approximately 1,057 million tons of coal, with around 30,000 mining and transportation units in operation daily[3].

In open-pit mining and transportation, the health of electromechanical equipment is directly linked to the safety and economic performance of coal mines. To evaluate equipment health, scholars commonly use various data types, such as vibration[4], current[5] and temperature[6]. Temperature data plays a critical role in equipment condition monitoring and health assessment due to its intuitive, real-time, and easily accessible nature. However, extended operation and external perturbations can lead to temperature fluctuations in electromechanical equipment, potentially

(*) Corresponding author.
E-mail addresses: F. Zhao (ORCID: 0009-0001-2209-4701) feiyangzhao0708@163.com, Y. Yuan (ORCID: 0000-0001-7736-840X) yipingyuan@xju.edu.cn , C. Chen (ORCID: 0000-0002-9926-8968) ccf@stu.xju.edu.cn , L. Xiao, xiaolu_hit@163.com ,

impacting performance and safety. Therefore, accurately predicting the temperature of electromechanical equipment in open-pit mines is essential.

Current temperature prediction methods for electromechanical equipment primarily rely on traditional statistical models or machine learning techniques. Zhang Yufei et al.[7] proposed an enhanced temperature prediction method for electrical equipment using Long Short-Term Memory (LSTM) neural networks, which utilizes a recurrent recursive layer based on LSTM loops to capture long-term features and deliver accurate temperature predictions. Teng Wei et al.[8] developed a combined model of extreme gradient boosting trees and a weighted fusion of LSTM networks for wind turbine temperature prediction. Kirchgässner et al.[9] introduced deep recurrent and convolutional neural networks with residual connections to forecast temperature changes in permanent magnet synchronous motors. Ma Xiaoyu et al.[10] designed a variable weight combination forecasting model that integrates feedback neural networks, grey models, and differential autoregressive moving average models to address low prediction accuracy in individual models. However, these approaches focus only on the long-term characteristics of temperature variations and do not account for temperature changes due to long-term operational trends and working condition disturbances.

Traditional methods, such as time series analysis, struggle to ensure prediction reliability and cannot handle predictions in complex environments[11][12]. Some scholars have attempted to improve the processing of time series data by extracting features from either temporal or spatial dimensions. Zhang Lei et al.[13] proposed preprocessing industrial data using mutual information theory to select the most relevant features as inputs, followed by a gated recurrent unit network for temperature prediction. This approach seeks to address the challenges of accurately detecting temperature anomalies in real-time. Zhai Naiju et al.[14] developed a temperature prediction model using a temporal convolutional network and generative adversarial loss for domain adaptation. Their model accurately forecasts the temperature of all heating zones within a furnace and, by incorporating a distillation network in a multi-task learning framework, mitigates the high delay in deep transfer networks. Qu et al.[15] introduced a stochastic vectorial function chain neural network with a sliding time window technique, improving prediction accuracy. Zou et al.[16] proposed constructing an ARIMA model using historical temperature data to forecast future temperatures of electrical equipment, allowing for effective and precise prediction of temperature trends. While these methods improve the accuracy and stability of temperature prediction to some extent, they still extract features from either the temporal or spatial dimension, neglecting the integration of both dimensions. As a result, the processed time series features are underutilized.

To address the aforementioned challenges, this paper presents a temperature prediction model that integrates wavelet packet decomposition and a graph attention network, aiming to more accurately predict temperature trends in electromechanical equipment in open-pit coal mines. Wavelet packet decomposition, with its finer division of the signal, offers higher resolution for the signal's high-frequency components, allowing for effective handling of non-stationary signals. It extracts both long-term features and short-term fluctuations caused by operational conditions from the temperature data[17]. Furthermore, these long-term and short-term features are fused through a graph structure, and the graph attention network's attention mechanism is employed to automatically capture complex relationships within the data, thereby improving prediction accuracy[18]. By combining these two methods, the limitations of single-model temperature prediction in motor systems are overcome, providing a more accurate basis for preventive maintenance of motors in open-pit coal mines. To validate the model's feasibility, a case study on temperature prediction for belt conveyor motors in large-scale open-pit mines is presented. The key innovations of this paper are outlined as follows:

(1) The proposed prediction model, which integrates wavelet packet decomposition and a graph attention network, addresses the limitations of single temperature prediction models.

(2) The temperature sequence is decomposed using wavelet packet decomposition to extract both long-term trends and short-term disturbances in equipment temperature, accounting for the operational trends of electromechanical equipment and temperature fluctuations caused by working condition disturbances.

(3) By fusing features from the temporal and spatial dimensions, the graph attention network automatically identifies the features with the most significant impact on temperature changes and assigns corresponding weights, thereby improving the model's prediction accuracy.

(4) A case study of a belt conveyor motor in a large-scale open-pit coal mine demonstrates the feasibility and effectiveness of the proposed prediction model.

2. Relevant Theory

2.1. Wavelet packet decomposition

Wavelet packet decomposition provides a more comprehensive approach than traditional wavelet decomposition. It divides the signal without redundancy, distinguishing between low-frequency approximation coefficients and high-frequency details within each sub-band, enabling more detailed signal analysis[19]. Additionally, wavelet packet decomposition introduces the concept of optimal basis functions into wavelet analysis. By selecting the optimal basis function based on the characteristics of the analyzed signal, the precision of the signal analysis is further enhanced.
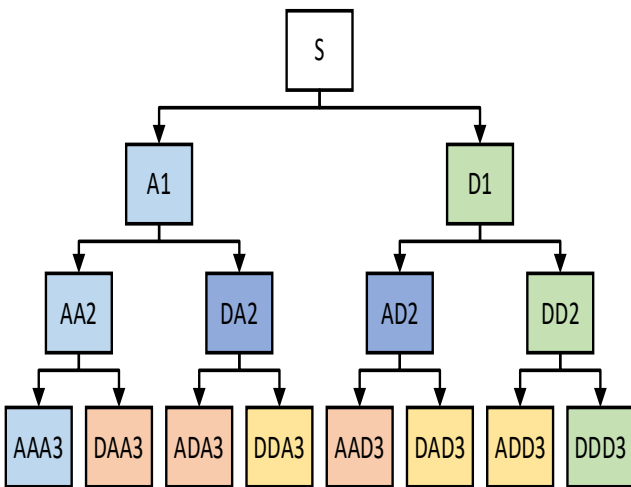

Fig.1. Schematic diagram of wavelet packet decomposition hierarchy.

Figure 1 illustrates a schematic diagram of a three-layer wavelet packet decomposition hierarchy. S denotes the analyzed signal, A represents the low-frequency component after decomposition, and D represents the high-frequency component. The numbers appended to A and D indicate the decomposition layers, starting from 1.

The low-frequency subsequence is characterized by a larger amplitude and shorter wavelength, effectively representing the long-term trends in the equipment temperature time series. In contrast, the high-frequency subsequence, with its gradually decreasing amplitude and longer wavelength, is better suited to capturing the subtle changes caused by working condition perturbations in the equipment temperature time series[20].

(1) Wavelet packet decomposition of the original signal[21]:

$$\begin{cases} \omega(t) = \sqrt{2}\sum_n h(n)\omega(2t-n) \\ \varphi(t) = \sqrt{2}\sum_n g(n)\varphi(2t-n) \end{cases} \tag{1}$$

Here, $\omega(t)$ represents the orthogonal scale function, $\varphi(t)$ represents the orthogonal wavelet function, and $h(n)$ and $g(n)$ denote the filter coefficients in the multiscale function.

(2) The recursive equation for the wavelet packet decomposition is:

$$\begin{cases} d_{i+1}^{2k} = \sum_n h(n-2t)d_i^k(n) \\ d_{i+1}^{2k+1} = \sum_n g(n-2t)d_i^k(n) \end{cases} \tag{2}$$

2.2. Graph Structure Data Construction

After wavelet packet decomposition, the equipment temperature data is separated into a low-frequency sequence that captures the overall temperature trend and a high-frequency sequence that reflects detailed temperature fluctuations. These sub-sequences can then form a network. $G = \{X, E, A\}$ represents the undirected graph formed by the subsequences, where denotes the feature matrix of the nodes, denotes the set of edges, and represents the graph's adjacency matrix[22]. Thus, all decomposed time subsequences can be structured as graph data.

The wavelet packet decomposition of the subsequence data is denoted as $X = [X_1, X_2, \dots, X_n]$, where n represents the number of sub-band data after wavelet packet

decomposition. First, each subsequence is treated as a node, forming a graph with multiple nodes. Next, cosine similarity is used to calculate the Euclidean distance between each pair of subsequences. Finally, the cosine similarity is compared to a predefined threshold, and if it exceeds this threshold, an edge is established between the corresponding nodes[23]. Thus, the presence of an edge between node x_i and node x_j can be determined by the following equation:

$$Ne(x_i) = \begin{cases} 1, if\ radius(x_i, x_j) > \varepsilon \\ 0, otherwise \end{cases} \qquad (3)$$

Here, $radius(x_i, x_j)$ represents the cosine similarity between nodes x_i and x_j, and ε denotes the threshold value. Cosine similarity measures the similarity between items on a scale from -1 to 1, where -1 indicates complete dissimilarity, 1 indicates perfect similarity, and higher values represent greater similarity.
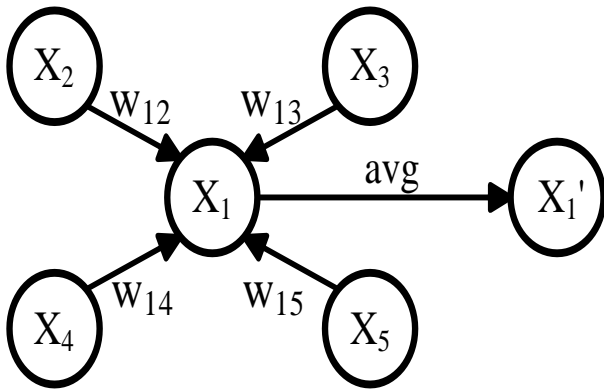

Fig.2. Diagram illustrating the working principle of the graph attention mechanism.

2.3. Graph Attention Networks

For the graph attention layer[24], the input to the GAT is $X = \{x_1, \ldots, x_n\}, X \in \mathbb{R}^{N \times F}$, where N represents the number of nodes and F denotes the number of features associated with each node. During feature extraction, the network linearly transforms the data based on the output size. For node x_i, attention is computed, and the attention coefficient is calculated as follows:

$$e_{ij} = Attn\left(W_{x_i}, W_{x_j}\right) = \sigma\left(a^T \cdot \left[W^c X_i \| W^c X_j\right]\right) \qquad (4)$$

Here, W represents the weight matrix, *Attn(·)* denotes the shared weights of the nodes parameterized by $a \in \mathbb{R}^D$ and $W^c = \mathbb{R}^{\frac{D^c}{2} \times F}$ within a neural network, D^c represents the

number of neurons in *Attn(·)*, $\|$ denotes the concatenation operation, and $\sigma(\cdot)$ represents the activation function. Formula (4) calculates the importance level between node x_i and node x_j. In a single layer of the graph attention network, this operation is performed for all neighboring nodes of a given node. Finally, the relationships between nodes are aggregated, and the formula is calculated as follows[25]:

$$a_{ij} = Soft\ max(e_{ij}) = \frac{exp(e_{ij})}{\sum_{k \in N_i} exp(e_{ik})} \qquad (5)$$

$$x_i' = \sigma\left(\sum_{j \in N_i} a_{ij} W_{x_{ij}}\right) \qquad (6)$$

Here, x_i' denotes the output of node feature x_i after passing through the graph attention layer and σ represents a non-linear activation function. The computed node feature is an aggregated representation of the correlation features of the node and its neighboring nodes. The operational mechanism of the graph attention layer is illustrated in Figure 2.

2.4. Temperature prediction model based on WPD-GAT

Wavelet packet decomposition offers multi-scale decomposition capabilities, allowing the original equipment temperature data to be broken down into various frequency sub-sequences. This process effectively extracts long-term features and short-term subtle fluctuations hidden in the temperature data, reduces the non-stationarity of the data, and mitigates its complexity. The graph attention network autonomously learns the importance of each neighboring node and flexibly captures relationships between nodes, enhancing the model's ability to mine original and associated features from the subsequences while fully utilizing the information contained in the temperature data. Therefore, this paper proposes a temperature prediction model based on wavelet packet decomposition and a graph attention network (WPD-GAT), with the model's principle illustrated in Figure 3. The detailed steps are as follows:

1) Decompose the original equipment temperature data into eight sub-sequences, including four low-frequency and four high-frequency sub-sequences, using wavelet packet decomposition;

2) Use the radius composition method to convert the

eight sub-sequences into graph-structured data;

3) Input the graph-structured data into the graph

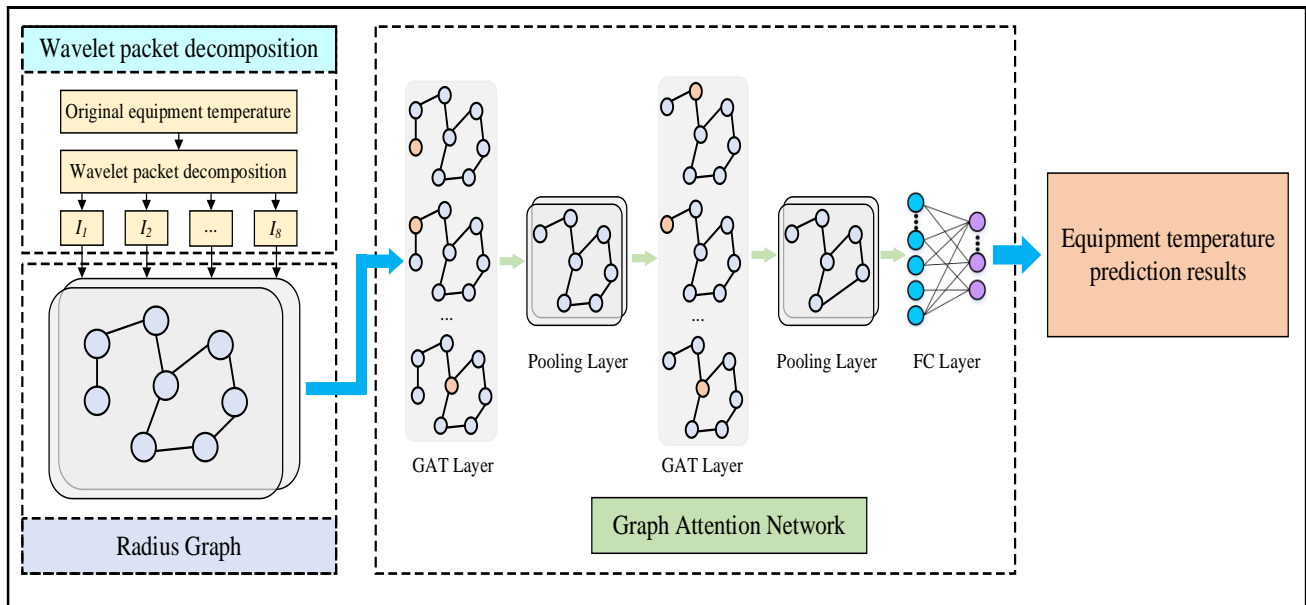attention network for prediction to obtain the equipment temperature prediction results.

Fig.3. WPD-GAT prediction model.

3. Experimental study

In this paper, the actual equipment production data from a 10-million-ton open-pit coal mine in Xinjiang is used as a case study to verify the effectiveness of the proposed model. Furthermore, the proposed model is compared against existing mainstream models, including the Convolutional Neural Network (CNN)[26], Long Short-Term Memory (LSTM)[27] and Graph Convolutional Neural Network (GCN)[28].

3.1. Data description

This paper examines the drive motors of the belt conveyor at a large open-pit coal mine. The belt conveyor is powered by three YXKK-400-4W model motors, as illustrated in Figure 4. In this paper, temperature monitoring data from the No. 1 motor of the belt conveyor, collected between February and December 2021, is selected for analysis. The data was sampled at a frequency of 1 sample per second, resulting in a total of 27,513,222 samples
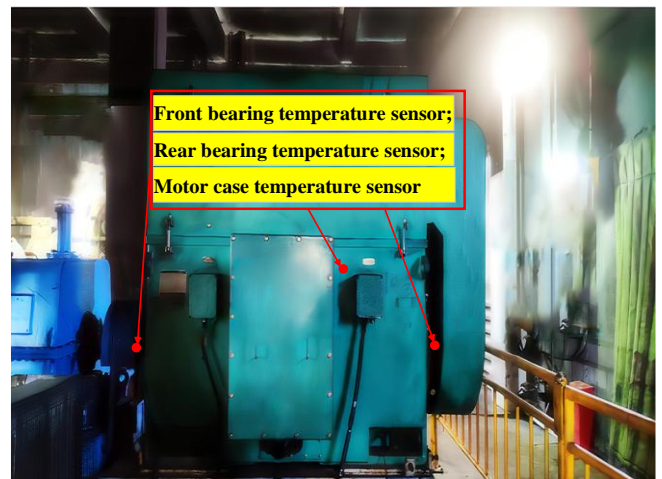

Fig.4 Belt conveyor drive motor.

The temperature changes of the equipment are illustrated in Figure 5. As shown in Figure 5, significant drops in motor temperature can be observed during extended downtime, such as the substantial decrease around 20:00 on April 16, 2021, due to a shutdown. Cyclical temperature changes are also evident, caused by frequent starts and stops, as seen between August 17 and August 31, 2021. Additionally, sensor failures led to anomalies, such as the instantaneous drop to -75°C near November 21, 2021. Therefore, data preprocessing is necessary before using this dataset to validate the predictive model.
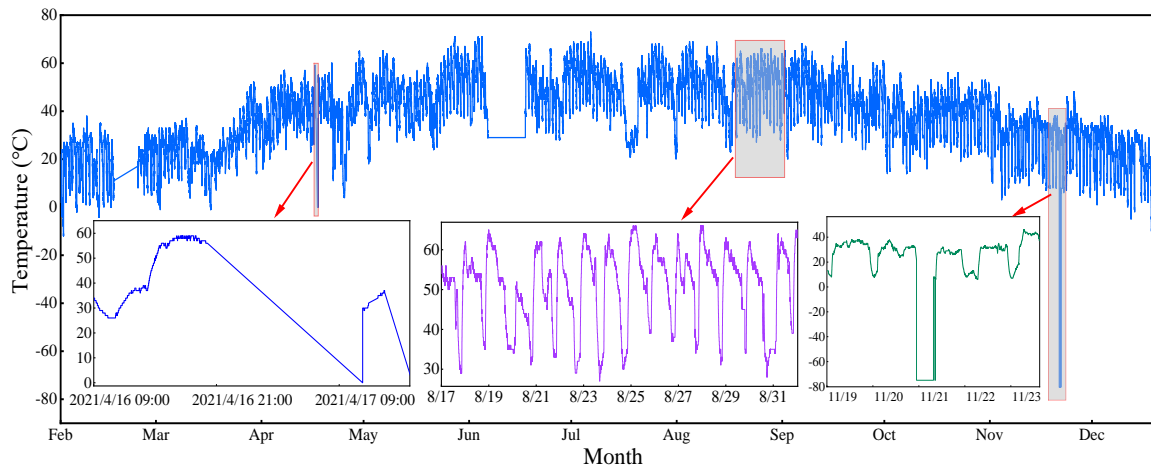
Fig.5. Raw Motor Temperature Data for February through December 2021.

3.2. Data processing

Data collected during actual production may contain missing values, duplicates, or anomalies, necessitating data cleaning before analysis. This includes removing duplicate temperature data and excluding downtime periods. After data cleaning, 15,133,240 data points remained. The temperature data was collected at a frequency of 1 sample per second, which exceeds the accuracy required for practical applications. At this frequency, temperature changes are typically gradual, especially in industrial settings, where instantaneous fluctuations are minimal, as shown in Figure 6. For most environmental and industrial applications, temperature data sampled per second shows little variation over a 10-minute interval. This allows for data dilution using 10-minute intervals, reducing the amount of data to be processed and improving processing efficiency. Additionally, averaging over this interval helps remove

short-term noise while preserving the underlying temperature trend. This approach maintains key characteristics of the data, such as short-term fluctuations and cyclical patterns, while preventing overfitting due to excessive detail. For long-term trend prediction models, frequent sampling may introduce unnecessary complexity and hinder the model's ability to identify long-term trends. Therefore, using 10-minute averages simplifies the input features while retaining essential information, making it easier for the model to capture long-term temperature trends and improving prediction accuracy.

The processed data, shown in Figure 7, consists of 24,508 temperature samples. The temperature trend aligns with that of the original dataset, with a minimum of -12°C recorded on February 7, 2021, at 23:20, and a maximum of 73°C on September 8, 2021, at 08:42. Additionally, the significant variability observed in the original data is retained in the processed data.

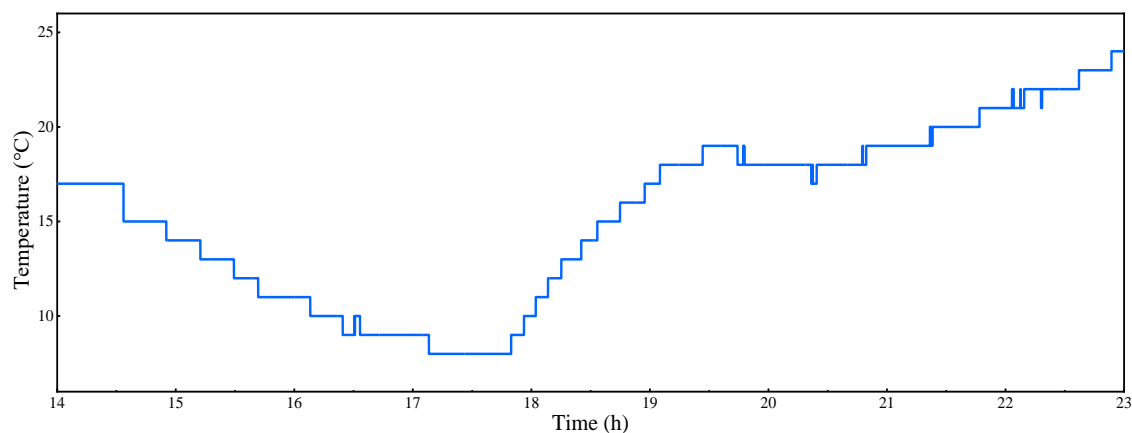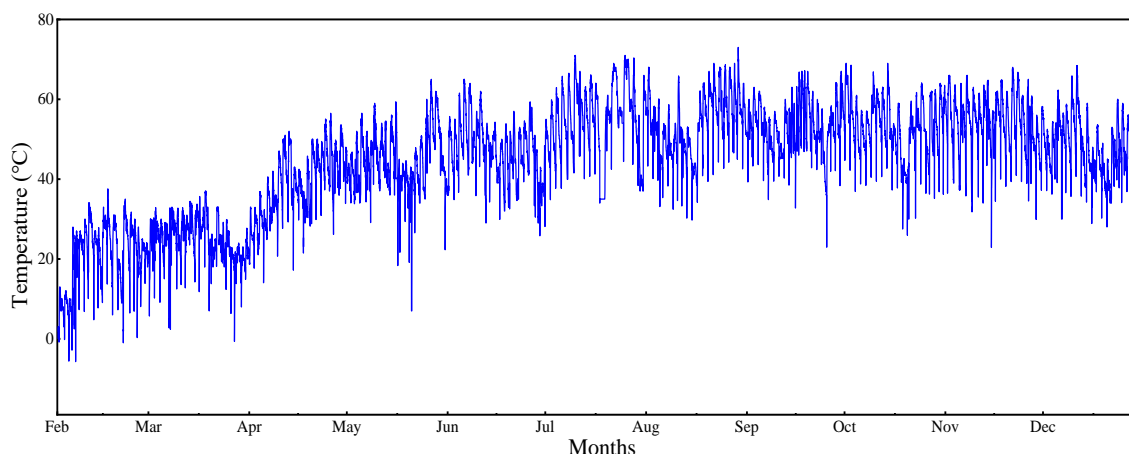Fig.6. Motor temperature change on February 23, 2021 after cleaning.

Fig.7. Preprocessed motor temperature data from February to December 2021.

3.3. Evaluation indicators

In this paper, three evaluation metrics are used to assess the prediction model: the coefficient of determination (R^2), the mean absolute error (MAE), and the root mean square error ($RMSE$)[28]. These metrics are calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{N}(Y_i - \bar{Y})^2} \qquad (7)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Y_i - \hat{Y}_i| \qquad (8)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2} \qquad (9)$$

Where N denotes the number of samples, $\hat{Y}_i$ represents the predicted temperature value of the i-th sample, Y_i represents the true temperature value of the i-th sample, and $\bar{Y}$ represents the average of the true temperature values across all samples.

The $RMSE$ represents the square root of the mean squared deviation between the predicted and actual temperature values, divided by the total number of samples. It is used to assess the error between predicted and actual temperatures, where smaller $RMSE$ values indicate higher prediction accuracy. The MAE calculates the average absolute difference between the predicted and actual temperature values. It computes the absolute difference between each predicted value and its corresponding true value, then averages these differences. Lower MAE values indicate fewer prediction errors in the model. The R^2 measures the correlation between the predicted and actual temperature values, with higher R^2 values signifying

a stronger correlation and better prediction accuracy.

Table 1. Output shapes of the WPD-GAT model.

Name of the model layer	Output shape
Input layer	[1024, 8]
GAT_1	[1024, 64]
BatchNorm_1	[1024, 64]
ReLU_1	[1024, 64]
Pooling_1	[825, 64]
GAT_2	[825, 128]
BatchNorm_2	[825, 128]
ReLU_2	[825, 128]
Pooling_2	[763, 128]
Linear_1	[128, 256]
Dropout	[128, 256]
Output	[128, 1]

3.4. Model training

Through several model parameter comparison experiments, the proposed WPD-GAT prediction model consists of two graph attention layers that weight and sum node features while learning associations between nodes via the attention mechanism. These graph attention layers are immediately followed by a batch normalization layer, which accelerates training and improves model stability. Additionally, the model includes an edge pooling layer to reduce the number of edges, thereby lowering the model's complexity. Finally, the model contains two fully connected layers, with a Dropout layer following the first fully connected layer to prevent overfitting. These fully connected layers are used to make predictions about the nodes. The shape of the output for each layer of the WPD-GAT prediction model is

presented in Table 1.

3.5. Analysis of Factors Influencing Wavelet Packet Decomposition

In signal processing, wavelet packet decomposition is a powerful multi-scale analysis tool, with its performance largely influenced by the choice of wavelet basis functions and the selection of decomposition levels. These two factors significantly affect the model's predictive accuracy and practical applicability.

3.5.1. Choosing Optimal Wavelet Basis Functions

The selection of wavelet basis functions is a crucial step in constructing a wavelet packet decomposition model. Ideally, the wavelet basis should possess mathematical properties such as orthogonality, compact support, and good regularity to ensure efficient and accurate signal analysis[29]. However, relying solely on mathematical attributes often fails to meet the diverse requirements of complex engineering applications. To address this, we systematically evaluate a range of commonly used wavelet basis functions, including the Haar wavelet and the Daubechies series (specifically, db2, db3, db4, and db5). Using ten-fold cross-validation and metrics such as $RMSE$, MAE, and R^2, we aim to identify the most suitable wavelet basis for the current analysis.
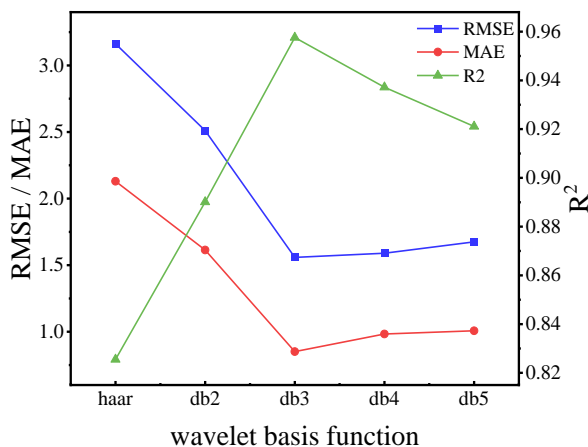

Fig.8 Comparison of experimental results using different wavelet basis functions

As shown in Figure 8, the db3 wavelet basis achieves the lowest MAE and $RMSE$ values while obtaining the highest R^2 value, indicating its optimal performance for this signal analysis task. Therefore, the db3 wavelet basis is selected as the final wavelet basis function for wavelet packet decomposition.

3.5.2. Determining the Optimal Decomposition Levels

The selection of decomposition levels is another critical factor influencing the effectiveness of wavelet packet decomposition. Excessive decomposition levels can dilute essential long-term trend information within the signal while significantly increasing computational complexity. Conversely, too few decomposition levels may fail to adequately distinguish between long-term variations and short-term disturbances, thereby reducing the accuracy of the analysis.

Using the selected db3 wavelet basis, this paper further investigates the impact of different decomposition levels (2, 3, 4, and 5) on the model's predictive performance. A ten-fold cross-validation analysis, combined with the three evaluation metrics, yields the experimental results presented in Figure 9. When the decomposition level is set to 3, the model demonstrates the best predictive performance, indicated by the lowest $RMSE$ and MAE values and the highest R^2 value. This suggests that for the dataset used in this study, a three-level decomposition strategy effectively retains useful signal information while maintaining computational efficiency. Consequently, the optimal number of decomposition levels for wavelet packet decomposition is determined to be 3.
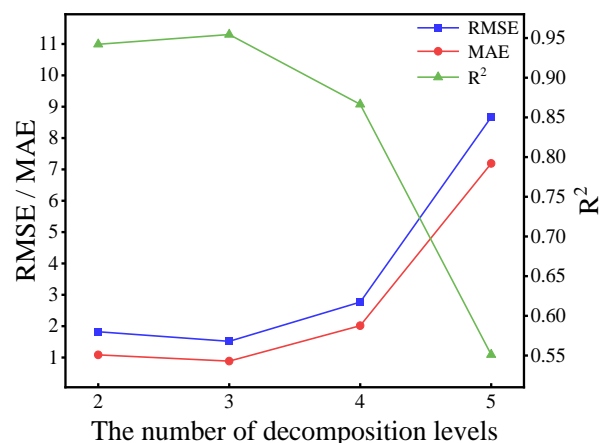

Fig.9. Comparison of experimental results for different decomposition levels.
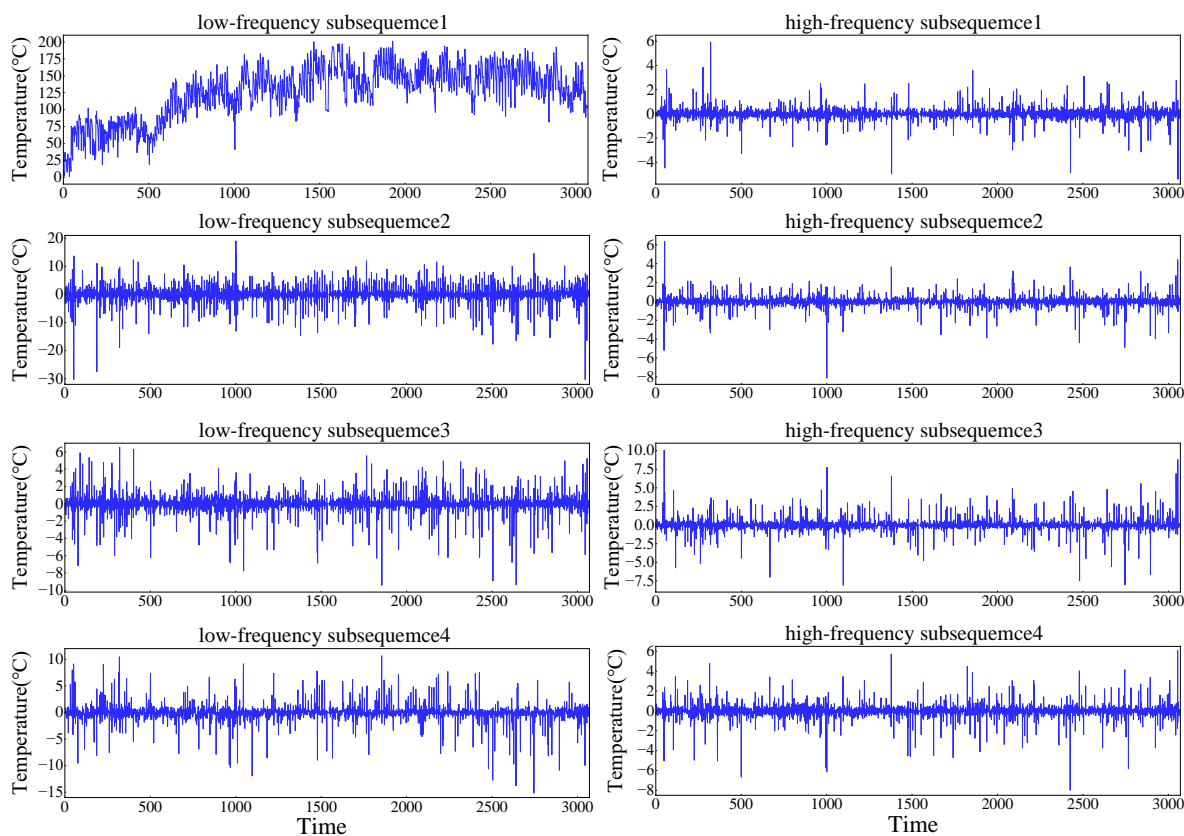
Fig.10. High and low frequency subsequences after wavelet packet decomposition.

Therefore, the preprocessed temperature sequence is decomposed using a 3-layer wavelet packet decomposition with db3 as the wavelet base. The resulting low-frequency and high-frequency subsequences are displayed in Figure 10. The trend of low-frequency subsequence 1 mirrors that of the original motor temperature data, while the high-frequency subsequence captures the subtle temperature fluctuations caused by complex working conditions. After each level of wavelet packet decomposition, the size of the resulting components is half of the original signal length plus two. Therefore, after three levels of decomposition, the length of the subsequence temperature data is reduced to 3,067.

3.6. Construction of graph-structured data

Considering the data correlation during the prediction period, this paper employs a sliding window approach to divide the time series temperature data into training samples. The time series data is split into fixed-length training samples, which serve as input for predicting the next temperature value. The sliding window shifts by removing the first value and appending the predicted value to the end, enabling continuous rolling predictions. In this paper, the window size is set to 30, meaning the temperature data from the past 5 hours is used to predict the device's average temperature for the next 10 minutes. A 3-layer wavelet packet decomposition is applied to the 30 temperature samples, producing 4 low-frequency subsequences and 4 high-frequency subsequences, each with a signal length of 8. The Radius composition method is then used to construct graph-structured data from the 8 subsequences, as illustrated in Figure 11.

0: AAA3
1: DAA3
2: ADA3
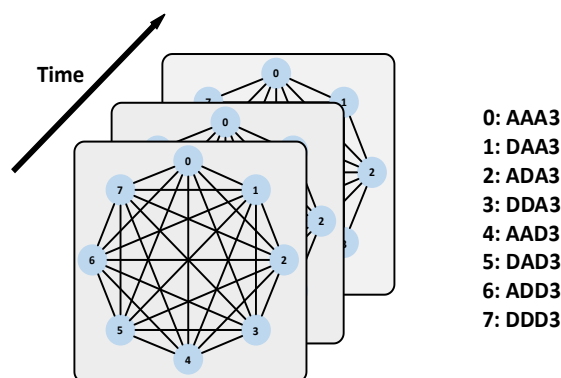3: DDA3
4: AAD3
5: DAD3
6: ADD3
7: DDD3

Fig.11. Graph structure data constructed from high and low frequency signal subsequences.

By continuously sliding the window and adding temperature values with different time stamps, a sufficient number of training samples is generated.
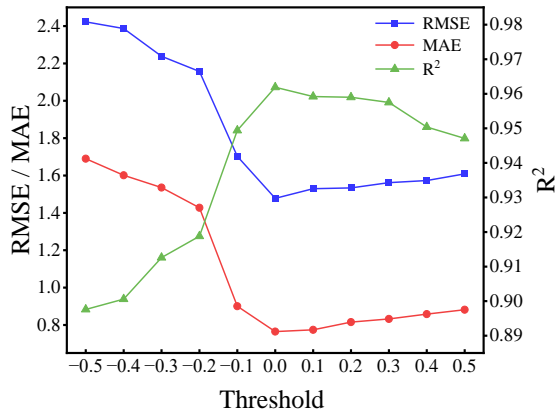

Fig.12. Comparison of experimental results for different thresholds.

In constructing graph-structured data using the Radius method, the existence of an edge between two nodes is determined by comparing their cosine similarity to a preset threshold. To systematically investigate the impact of different threshold settings on the model's predictive performance, a series of experiments were conducted, testing various thresholds, including -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, and 0.5. Each threshold was tested through 25 experimental repetitions. The results, shown in Figure 12, reveal the variations in predictive performance across different thresholds. The findings indicate that when the threshold is set to 0, the model achieves optimal predictive performance. Both the *RMSE* and *MAE* are minimized, indicating the smallest deviations between the model's predictions and actual values. Additionally, the R^2 value reaches its maximum, further confirming the model's strong fit and robust predictive capability. Therefore, the optimal cosine similarity threshold is determined to be 0, serving as the basis for constructing the graph structure.

4. Results and discussion

To evaluate and analyze the predictive performance of the proposed motor temperature prediction model, it is compared against existing prediction methods, including LSTM, CNN, WPD-CNN, and WPD-GCN. Each model was tested in 25 repeated experiments, with the average

results used to minimize the influence of randomness on the outcomes. Experimental validation is conducted using preprocessed motor temperature data collected from February to December 2021. The temperature data from February to November 2021 is used as the training set, while the data from December 2021 is used as the test set.
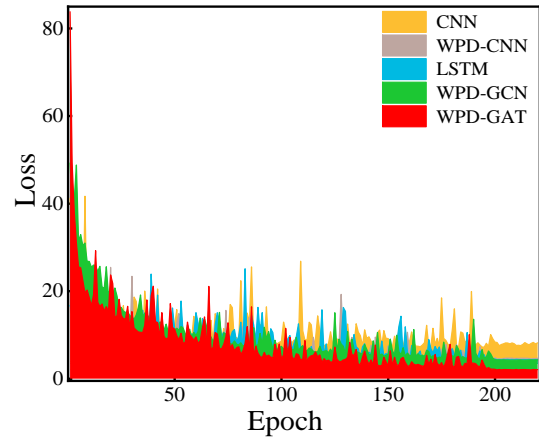
Fig.13 Comparison of the convergence curves of the loss functions of the models.

After several experimental comparisons, the optimizer was set to Adam, with the maximum number of iterations set to 220, a learning rate of 0.0005, and a batch size of 128. The mean-square error function was used as the objective function for model training. A comparison of the loss curves during the training process for each model is illustrated in Figure 13, showing that the WPD-GAT model exhibits a rapid convergence rate during the early training stages. As training progresses, the model's convergence rate gradually decreases, but the trend becomes increasingly smooth with no significant fluctuations, demonstrating good stability. Additionally, the model achieves a minimized loss value.

Fig.14. Comparison of evaluation indicators for models.

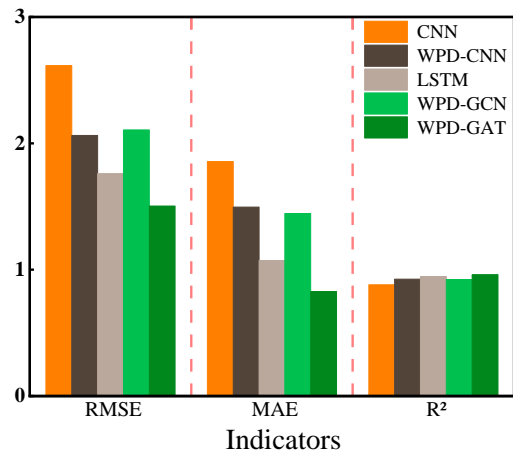The detailed results of the 25 repeated experiments for each model are presented in Table 2. To clearly demonstrate the average predictive performance of the different models, the average test results were visualized as a bar chart, shown in Figure 14. Based on the obtained prediction results, the analysis is as follows:

(1) As shown in Table 2, the proposed WPD-GAT model outperforms all five models, with an average $RMSE$ of 1.504, an average MAE of 0.827, and an average coefficient of determination (R^2) of 0.961. The prediction result curves for the test set are shown in Figure 15. Due to the large volume of data, a portion of Figure 15 is enlarged in Figure 16, demonstrating that the WPD-GAT model provides the best fit with the actual values, achieves the highest prediction accuracy, and closely follows the temperature trends of the real equipment.

(2) The use of WPD significantly enhances the model's predictive performance. As shown in Table 2, the evaluation metrics for the WPD-CNN model are superior to those of the CNN model, with an average $RMSE$ reduction of 21.11%, an average MAE reduction of 19.45%, and an average R^2 improvement of 5.11%.

(3) In processing graph data, GAT exhibits superior feature representation, global feature capture, and graph structure adaptation compared to GCN. When Comparing the evaluation metrics of the WPD-GAT model with those of the WPD-GCN model, the WPD-GAT model achieves an average $RMSE$ reduction of 28.58%, an average MAE reduction of 42.77%, and an average R^2 improvement of 4.22%. Furthermore, as shown in the prediction result graphs, the WPD-GAT model's predictions are more closely aligned with the actual values than those of the WPD-GCN model.
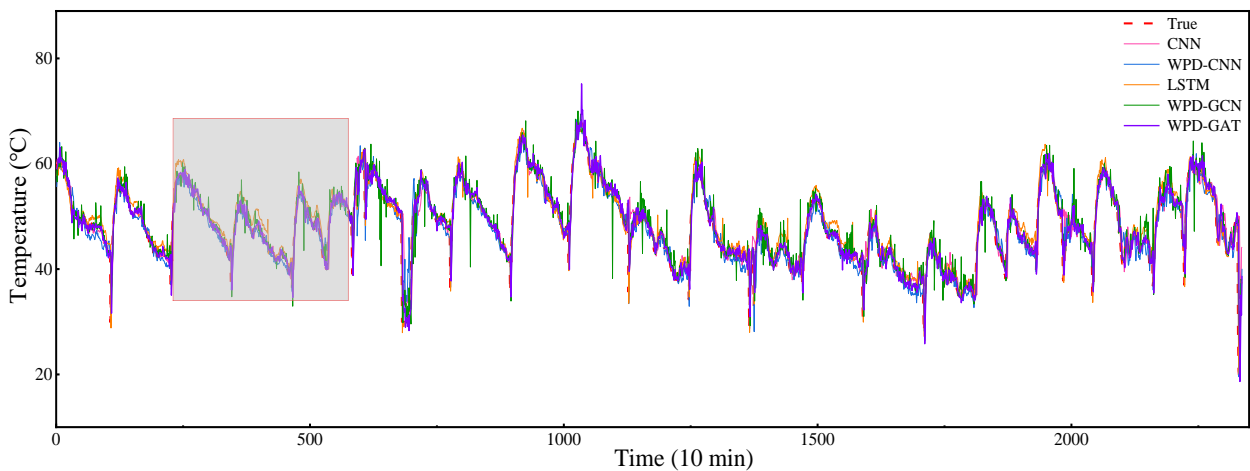

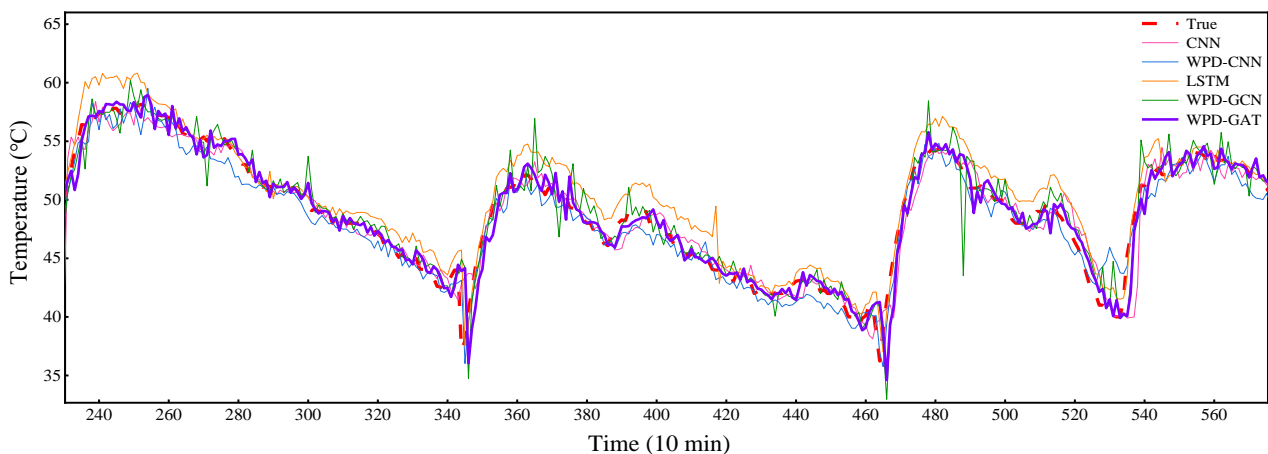Fig. 15. Comparison of prediction results of each model.

Fig. 16. Localized enlargement of the prediction results of each model.

Table 2. Evaluation indicators for each prediction model.

Number	CNN			WPD-CNN			LSTM			WPD-GCN			WPD-GAT		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
1	2.683	1.487	0.875	2.042	1.399	0.927	1.841	1.271	0.941	2.124	1.177	0.921	1.505	0.824	0.961
2	2.728	1.929	0.870	1.915	1.344	0.936	1.615	0.783	0.955	2.115	1.447	0.922	1.452	0.772	0.963
3	2.745	1.954	0.869	1.756	1.149	0.946	1.842	1.051	0.941	2.328	1.620	0.905	1.512	0.825	0.960
4	2.683	1.886	0.876	1.820	1.196	0.942	1.741	1.102	0.947	1.920	1.379	0.936	1.530	0.902	0.959
5	2.900	2.205	0.853	2.088	1.586	0.924	1.722	1.037	0.948	2.375	1.509	0.902	1.457	0.802	0.963
6	2.525	1.709	0.889	2.036	1.563	0.928	1.790	0.944	0.944	2.262	1.507	0.911	1.484	0.800	0.962
7	2.898	2.256	0.854	2.221	1.629	0.914	1.924	1.040	0.936	1.753	1.222	0.946	1.479	0.811	0.962
8	2.423	1.690	0.898	2.021	1.543	0.929	1.700	0.836	0.950	2.243	1.467	0.912	1.534	0.928	0.959
9	2.451	1.715	0.895	1.859	1.292	0.940	1.861	1.339	0.940	2.285	1.467	0.909	1.496	0.808	0.961
10	2.494	1.714	0.892	2.228	1.782	0.913	1.821	0.984	0.942	2.181	1.386	0.917	1.499	0.838	0.961
11	2.607	1.850	0.881	2.212	1.564	0.915	1.675	1.007	0.951	2.360	1.548	0.903	1.520	0.873	0.960
12	2.490	1.805	0.892	2.080	1.619	0.925	1.692	1.005	0.950	1.601	0.953	0.955	1.522	0.863	0.960
13	2.475	1.732	0.893	2.054	1.521	0.926	1.682	0.985	0.951	1.936	1.340	0.935	1.508	0.821	0.960
14	2.750	2.106	0.868	2.105	1.479	0.923	1.780	1.093	0.945	2.110	1.333	0.922	1.489	0.811	0.961
15	2.452	1.721	0.895	2.471	2.102	0.894	1.906	1.184	0.937	2.283	1.627	0.909	1.509	0.828	0.960
16	2.599	1.827	0.882	2.031	1.432	0.928	1.734	1.052	0.948	1.985	1.504	0.931	1.521	0.859	0.960
17	2.738	1.999	0.869	2.206	1.648	0.915	1.850	1.282	0.940	2.169	1.692	0.918	1.515	0.841	0.960
18	2.602	1.812	0.882	2.234	1.705	0.913	1.831	1.168	0.942	2.125	1.495	0.921	1.552	0.859	0.958
19	2.541	1.865	0.887	1.962	1.320	0.933	1.761	1.091	0.946	2.002	1.401	0.930	1.499	0.784	0.961
20	2.438	1.689	0.896	2.193	1.604	0.916	1.783	1.184	0.945	2.186	1.666	0.917	1.534	0.820	0.959
21	2.551	1.739	0.887	2.073	1.495	0.925	1.724	1.055	0.948	1.997	1.326	0.930	1.478	0.765	0.962
22	2.640	1.968	0.878	2.019	1.387	0.929	1.690	1.063	0.948	1.975	1.498	0.932	1.498	0.780	0.961
23	2.684	1.872	0.874	1.943	1.277	0.934	1.638	1.031	0.953	2.049	1.436	0.927	1.515	0.841	0.960
24	2.682	1.991	0.875	1.947	1.385	0.934	1.624	0.973	0.954	2.226	1.576	0.914	1.550	0.861	0.958
25	2.602	1.870	0.882	2.048	1.361	0.927	1.790	1.257	0.944	2.065	1.549	0.926	1.432	0.762	0.964
average	2.615	1.856	0.880	2.063	1.495	0.925	1.761	1.073	0.946	2.106	1.445	0.922	1.504	0.827	0.961

5. Conclusions

This paper proposes a temperature prediction model for key equipment in open-pit coal mines, based on wavelet packet decomposition and graph attention networks (GAT). First, the equipment temperature sequence is decomposed using wavelet packet decomposition. Cosine similarity is then applied to quantify the correlation between each high and low frequency subsequence, which is subsequently converted into graph-structured data using the Radius composition method. Finally, the graph-structured data samples are used as inputs for prediction in the GAT model.

To verify the effectiveness of the proposed model, this paper constructs four comparative models using historical data from electromechanical equipment in a large open-pit coal mine in Xinjiang and evaluates them using three performance metrics. The experimental results show that, compared to the CNN, WPD-CNN, LSTM, and WPD-GCN models, the WPD-GAT model improves the R^2 index by an average of 9.20%, 3.89%, 1.58%, and 4.23%, respectively. The WPD-GAT model accounts for both long-term temperature trends and short-term perturbations, performing feature fusion across both temporal and spatial dimensions, which enhances the model's prediction accuracy and robustness.

While the WPD-GAT model does not significantly outperform other comparative models in terms of training time, it does demonstrate a superior level of prediction accuracy. The high accuracy of the WPD-GAT model is critical in guiding real-world production operations. It provides insight into potential equipment failures and enables more proactive, targeted maintenance measures, thereby avoiding production interruptions and cost increases due to sudden equipment failures. This maintenance strategy, based on highly accurate predictions, not only improves productivity and equipment utilization, but also significantly enhances the overall operational stability and competitiveness of the enterprise. Therefore, the WPD-GAT model shows great potential and value in the field of equipment health management due to its excellent predictive performance.

References

1. Wang G, Xu Y, Ren H. Intelligent and ecological coal mining as well as clean utilization technology in China: Review and prospects. International Journal of Mining Science and Technology, 2019, 29: 161-169. https://doi.org/10.1016/j.ijmst.2018.06.005.

2. Cai Q, Chen Y. Review of 70 years' achievements and high-quality development architecture system of surface coal mining in China. Journal of China Coal Society, 2024, 49(01): 235-260. https://doi.org/10.13225/j.cnki.jccs.2023.1479.

3. Bai R, Fu E, Ma L, Zhao H, Chai S. Collaborative mining technological system of safety-green-high efficiency-low carbon for open pit coal mine[J]. Journal of China Coal Society, 2024, 49(01): 298-308. https://doi.org/10.13225/j.cnki.jccs.2023.1433.

4. Fan P, Yuan Y, Ma Z, Gao J, Zhang Y. Incipient fault prediction based on warning control limit self-learning for the rolling bearing. Computer Integrated Manufacturing Systems, 2024, 30(01): 227-238. https://doi.org/10.13196/j.cims.2021.0736.

5. Cao X, Duan Y, Zhao J, Yang X, Zhao F, Fan H. Summary of research on health status assessment of fully mechanized mining equipment. Industry and Mine Automation, 2023, 49(09):23-35,97. https://doi.org/10.13272/j.issn.1671-251x.18143.

6. Cao X, Lei Y, Gong Y, Zhang S, Luo X. Study on health assessment method of shearer based on combination weighting method. Coal Science and Technology, 2020, 48(6):135-141. https://doi.org/10.13199/j.cnki.cst.2020.06.017.

7. Zhang Y, Meng F, Wang Y, Wu Y, Li H. Research on temperature prediction method of power equipment based on improved LSTM. Journal of Electronic Measurement and Instrumentation, 2021, 35(12):167-173. https://doi.org/10.13382/j.jemi.B2104077.

8. Teng W, Huang Y, Wu S, Liu Y. Wind Turbine Generator Winding Temperature Prediction Based on XGBoost and LSTM. Electric Power, 2021, 54(06): 95-103. https://doi.org/10.11930/j.issn.1004-9649.202101090.

9. Kirchgässner W, Wallscheid O, Böcker J. Estimating Electric Motor Temperatures With Deep Residual Machine Learning. IEEE Transactions on Power Electronics, 2021, 36: 7480-7488. https://doi.org/10.1109/TPEL.2020.3045596.

10. Ma X, Zhang H, Liu Z, Zhang C. Temperature Prediction Model of Electrical Equipment Based on Variable Weight Combination. Control Engineering of China, 2023, 30(04):694-702. https://doi.org/10.14107/j.cnki.kzgc.20210840.

11. Moura M, Zio E, Lins I, Droguett E. Failure and reliability prediction by support vector machines regression of time series data. Reliability Engineering & System Safety, 2011, 96: 1527-1534. https://doi.org/10.1016/j.ress.2011.06.006.

12. Tealab A. Time series forecasting using artificial neural networks methodologies: A systematic review. Future Computing and Informatics Journal, 2018, 3: 334-340. https://doi.org/10.1016/j.fcij.2018.10.003.

13. Zhang L, Li J, Du J, Li L, Zou Y, Zhang E, Li S. Dynamic prediction of cylinder wall temperature for drum dryer based on DGRU network. Journal of Light Industry, 2022,37(06):85-91, 100. https://doi.org/10.12187/2022.06.11.

14. Zhai N, Zhou X, Li S, Shi H. Prediction method of furnace temperature based on transfer learning and knowledge distillation. Computer Integrated Manufacturing Systems, 2022, 28(06): 1860-1869. https://doi.org/10.13196/j.cims.2022.06.024.

15. Qu H, Fu S, Pang L, Ding L, Zhang H. Rapid temperature prediction method for electronic equipment cabin. Applied Thermal Engineering, 2018, 138: 83-93. https://doi.org/10.1016/j.applthermaleng.2018.04.001.

16. Zou Y, Wang T, Xiao J, Feng X. Temperature prediction of electrical equipment based on autoregressive integrated moving average model. 2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC): IEEE, 2017. https://doi.org/10.1109/YAC.2017.7967404.

17. Chi X, Xu Z, Jia X, Zhang W. WPD-ISSA-CA-CNN Model Based Carbon Emission Prediction in Power Plants. Control Engineering of China,1-8. https://doi.org/10.14107/j.cnki.kzgc.20220983.

18. Xi L, Wang R, Fan H, Zhang F. Sample-Correlation-Aware Unsupervised Deep Anomaly Detection Model. Chinese Journal of Computers,2021,44(11):2317-2331. https://doi.org/10.11897/SP.J.1016.2021.02317.

19. Liu Y, Zhao Y, Zhang L, Wu M. Photovoltaic Power Forecasting Based on Wavelet Packet Decomposition of Long Short-term Memory Network. Computer & Digital Engineering, 2022, 50(09): 2114-2118. https://doi.org/10.3969/j.issn.1672-9722.2022.09.044.

20. Wang J, Wang J. A New Hybrid Forecasting Model Based on SW-LSTM and Wavelet Packet Decomposition: A Case Study of Oil Futures

Prices. Computational Intelligence and Neuroscience, 2021, 2021: 1-22. https://doi.org/10.1155/2021/7653091.

21. Qu J, Cheng X, Liang P, Zheng L, Ma X. Fault Diagnosis of Bearings Using Wavelet Packet Energy Spectrum and SSA-DBN. Processes 2023,11,1875. https://doi.org/10.3390/pr11071875.

22. Wang Y, Jing C, Xu S, Guo T. Attention based spatiotemporal graph attention networks for traffic flow forecasting. Information Sciences, 2022, 607: 869-883. https://doi.org/10.1016/j.ins.2022.05.127.

23. Li T, Zhou Z, Li S, Sun C, Yan R, Chen X. The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. Mechanical Systems and Signal Processing, 2022, 168: 108653. https://doi.org/10.1016/j.ymssp.2021.108653.

24. Su G, Wang H, Zhang Y, Zhang W, Lin X. Simple and deep graph attention networks. Knowledge-based Systems, 2024, 293: 111649. https://doi.org/10.1016/j.knosys.2024.111649.

25. Wang X, Yao Y, Gao C. Wasserstein Distance- EEMD Enhanced Multi-Head Graph Attention Network for Rolling Bearing Fault Diagnosis Under Different Working Conditions. Eksploatacja i Niezawodnosc – Maintenance and Reliability, 2024: 26(2). http://doi.org/10.17531/ein/184037.

26. Karabacak Y E. Deep learning-based CNC milling tool wear stage estimation with multi-signal analysis. Eksploatacja i Niezawodnosc – Maintenance and Reliability, 2023: 25(3). http://doi.org/10.17531/ein/168082.

27. Zhan X, Liu Z, Yan H, Wu Z, Guo C, Jia X. A novel method of health indicator construction and remaining useful life prediction based on deep learning. Eksploatacja i Niezawodnosc – Maintenance and Reliability, 2023: 25(4). http://doi.org/10.17531/ein/171374.

28. Chen C, Yuan Y, Sun W, Zhao F. Multivariate multi-step time series prediction of induction motor situation based on fused temporal-spatial features. International Journal of Hydrogen Energy, 2024, 50: 1386-1394. https://doi.org/10.1016/j.ijhydene.2023.11.047.

29. Zhong M, Zhang C, Li H, Wang B. Improved de-noising method based on wavelet packet energy subsection threshold. Computer Engineering and Applications, 2015, 51(5): 204-207. https://doi.org/10.3778/j.issn.1002-8331.1306-0272.