Article citation info
Wang K, Duan J, Sun Q, Han L, Tang J, Li J. Intelligent diagnosis and fault prediction of rock breaking process in milling machines based on big data analysis. Diagnostyka. 2026;27(1):2026105. https://doi.org/10.29354/diag/217441.

DIAGNOSTYKA, 2026, Vol. 27, No. 1

e-ISSN 2449-5220
DOI: 10.29354/diag/217441

INTELLIGENT DIAGNOSIS AND FAULT PREDICTION OF ROCK BREAKING PROCESS IN MILLING MACHINES BASED ON BIG DATA ANALYSIS

Jingde WANG [1,*], Jian DUAN [2], Qingsong SUN [1], Lu HAN [1], Jian TANG [3], Jian LI [3]

[1] State Grid Hebei Construction Company, Shijiazhuang 050021, China
[2] State Grid Hebei Company, Shijiazhuang 050081, China
[3] CGE(Chongqing) Exploration Machinery Co., Ltd., Chongqing 400030, China
*Email of Corresponding Author: zenengydb6951@163.com

Abstract

This study develops an intelligent method for diagnosing and predicting disc cutter deflection wear faults during rock excavation processes. The proposed method integrates wavelet time-frequency graph encoding with an Inception-Bidirectional Gated Recurrent Unit (Bi-GRU) model through big data analysis techniques. The approach utilizes large-scale vibration signal data to effectively extract time-frequency characteristics of non-stationary signals through wavelet time-frequency graphs, enhancing the model's fault identification capability under complex working conditions. The one-dimensional vibration signal is mapped into a two-dimensional time-frequency graph and input into the Inception-BiGRU model. The spatial features are extracted by using the multi-scale convolution of the Inception module. Meanwhile, the dynamic evolution law is captured by combining the Bi-GRU bidirectional time series modeling to achieve the deep integration of spatial-temporal features. The Inception-BiGRU model demonstrates excellent robustness in multi-condition tests, maintaining a diagnostic accuracy consistently exceeding 98% and reaching a maximum of 99.35%. In multi-class fault identification, the accuracy for normal and severely deflected cutters reaches 99%. Meanwhile, minor misjudgments occur in cases of mild and moderate deflection, with overall identification precision exceeding 97%. A comparative evaluation against various encoding methods and traditional algorithms demonstrates that the Inception-BiGRU model outperforms baseline models in both accuracy and stability, achieving an average accuracy of 98.63% with a variance of only 0.03. The proposed model improves average accuracy by 1.84% and reduces variance by 50% compared to baseline models. The research results verify the effectiveness and engineering application potential of the proposed method in diagnosing and predicting disc cutter faults. This study provides technical support for advancing the intelligent operation and maintenance of milling machines while offering a feasible path for future multimodal sensor fusion and early fault warning.

Keywords: milling machine; disc cutter deflection wear; fault diagnosis; inception; gated recurrent unit; wavelet time-frequency graph

1. INTRODUCTION

The milling machine serves as a critical piece of equipment in modern tunnel and underground engineering, with its rock-breaking efficiency being closely related to the health condition of disc cutter components. As the core element directly contacting rock masses and performing cutting tasks, the wear state of disc cutters significantly affects construction progress, equipment stability, and maintenance costs [1-3]. Under long-term complex working conditions, disc cutters often experience various forms of wear, including edge wear, local spalling, and skew wear. In particular, skew wear tends to lead to concentrated stress on the cutter ring, deviation of cutting trajectories, and additional vibration and impact loads. This form of wear results from axial misalignment caused by uneven force distribution on the cutter handle or cutterhead. This makes it one of the most significant wear types affecting equipment performance. Skew wear is characterized by strong concealment, rapid development, and wide-ranging impacts. In its early stage, fault features are often difficult to identify through manual observation. As

Received 2025-08-08; Accepted 2026-01-27; Available online 2026-01-28

skew wear further intensifies, it significantly reduces rock-breaking efficiency; it also causes fatigue of structural components, increased unstable vibration of the cutterhead, and even serious equipment accidents such as abnormal detachment of disc cutters. Additionally, unplanned downtime caused by skew wear directly leads to construction delays and economic losses, seriously affecting project progress and construction safety. Therefore, intelligent diagnosis and prediction technologies for disc cutters' skew wear hold significant engineering importance for enhancing the reliability and intelligence level of large-scale tunneling equipment. Traditional fault identification methods based on manual experience or simple threshold judgments fail to accurately reflect the disc cutters' actual working conditions, exhibiting problems such as time lag and high misjudgment rates [4-6]. Consequently, conducting intelligent fault diagnosis and prediction research focusing on abnormal wear behavior of disc cutters of milling machines during rock-breaking processes has become a crucial approach for enhancing construction intelligence and ensuring operational safety.

With the widespread application of high-frequency vibration sensors and multi-channel data collection systems, vibration signals during rock-breaking processes of milling machines can be monitored in real time with high temporal resolution. Thus, it forms large-scale time-series data characterized by non-stationarity, multiple working conditions, and high levels of background noise. Since skew wear changes the contact state and load transfer path of the cutterhead-rock mass system, it correspondingly presents typical dynamic features in vibration signals; these features include energy band transfer, enhanced transient impact, and changes in low-frequency characteristics. However, traditional signal processing methods such as wavelet packet decomposition have limited descriptive capabilities when dealing with highly non-stationary signals; this makes it difficult to fully reflect the time-frequency evolution laws of skew wear [7, 8]. Deep learning methods such as the artificial neural network (ANN), convolutional neural network (CNN), and recurrent neural network (RNN) have shown remarkable advantages in time-series data modeling and image identification, providing technical support for disc cutter fault diagnosis [9, 10]. In disc cutter fault diagnosis, integrating time-frequency feature extraction capability with time-series modeling capability has become an important direction for improving diagnosis accuracy.

Existing disc cutter fault diagnosis methods have improved recognition efficiency to a certain extent. However, they still generally have three limitations. 1. Most methods rely on manually constructed time-domain (TD) or frequency-domain (FD) features, which fail to fully characterize the non-stationarity, transient impact, and multi-scale dynamic changes caused by skew wear. 2. Some deep learning models lack adaptability to signal distribution differences under diverse working conditions, leading to significant performance degradation under actual conditions such as variable loads and variable speeds. 3. When CNN or RNN is used alone, either they cannot effectively capture time-frequency multi-scale structural features, or they face problems such as gradient decay and high computational cost in long-sequence modeling. To address the above shortcomings, this study proposes a model based on Continuous Wavelet Transform (CWT) and Inception-Bidirectional Gated Recurrent Unit (Bi-GRU). It aims to achieve systematic improvements in feature expression capability, working condition adaptability, and model efficiency. (1) CWT is used to construct two-dimensional (2D) time-frequency graphs, overcoming the insufficient description of non-stationary signals by traditional features and realizing fine-grained expression of skew wear features. (2) The Inception structure with multi-scale convolution is introduced; this enables the model to extract complex weak features from different receptive fields and breaks through the limitation of insufficient recognition capability of traditional single-scale networks. (3) Bi-GRU with higher computational efficiency is adopted to realize time-series dependency modeling while reducing computational overhead, improving the feasibility of on-site engineering deployment of the model. Through the above improvements, this study enhances the recognizability and state discrimination ability of skew wear of disc cutters in milling machines. It also provides a promotable technical path for intelligent monitoring of large-scale construction equipment under complex working conditions.

2. RELATED WORK

With the continuous development of mechanical equipment fault diagnosis technologies, researchers have made significant progress in improving fault identification accuracy and system robustness. Chen et al. developed an enhanced transfer learning method for mechanical fault diagnosis by incorporating a dynamic softmax function with an angular margin penalty to adjust the model's feature representation capability. They incorporated trend block learning to capture trend characteristics in vibration signals, thereby improving the diagnostic network's robustness and feature richness for target data [11]. This method provided theoretical references for enhancing feature representation and robustness in this study. Sahu et al. systematically reviewed 190 studies over the past two decades. They categorized data-driven diagnostic methods into supervised, semi-supervised, and

unsupervised learning approaches while thoroughly analyzing their respective advantages and applicable scenarios [12]. Their work offered systematic guidance for establishing the supervised learning framework in this study. Kibrete et al. conducted a comprehensive analysis of multi-sensor data fusion technology in rotating machinery fault diagnosis. Their research presented fundamental concepts and conducted in-depth application analyses, offering valuable resources for researchers, practitioners, and policymakers to advance intelligent fault diagnosis technology [13]. Their multi-source information integration concept inspired methodological considerations for complex signal feature extraction.

In the field of deep learning-driven fault diagnosis and prediction research, Wang et al. integrated deep transfer learning with digital twin technology. Their work achieved significant improvements in both clarity and authenticity of magnetic resonance imaging [14]. This concept indirectly validated the adaptation capability of deep models to complex feature structures and provided methodological support for the design of deep networks and image encoding in this study. He et al. proposed a cross-domain framework for motor-level compound fault diagnosis that required no target domain information, employing a time-frequency self-contrastive learning strategy to enhance domain-invariant feature extraction. Their approach generated homogeneous and heterogeneous information from time and frequency domains as self-contrastive pairs. Subsequently, the data is integrated through multi-scale spatial convolutional structures and cross-time-frequency information interaction strategies [15]. Thus, their research offered valuable references for multi-scale feature fusion and time-frequency information interaction in this study. Mahesh et al. constructed an optimized 1D CNN utilizing varying kernel sizes. By automatically learning the relevant features from raw sensor data, complex patterns could be captured without manual feature engineering, enabling effective prediction and identification of bearing problems [16]. Their research provided technical insights for reducing manual intervention and improving automatic feature extraction capability in this study.

Existing studies still have limitations in time-frequency feature extraction of vibration signals, model adaptability under multiple working conditions, and the design of deep network structure. On one hand, most methods are insufficiently sensitive to fine-grained changes in non-stationary signals, making it difficult to extract complex fault features such as disc cutter deflection wear. On the other hand, some models exhibit poor adaptability to working condition changes and insufficient generalization ability. Concurrently, most studies rely on single encoding methods or ignore time-series dependency information. This study constructs a diagnostic model by fusing CWT time-frequency graphs with an Inception-BiGRU deep architecture. It also combines parallel convolution for multi-scale feature extraction and bidirectional gated recurrence for time-series dependency modeling. This approach significantly enhances the accuracy and robustness of disc cutter deflection wear fault identification, expanding the application boundaries of deep learning in intelligent diagnosis of complex mechanical equipment.

3. DIAGNOSTIC AND PREDICTIVE MODEL OF DISC CUTTER DEFLECTION WEAR FAULTS IN MILLING MACHINES

3.1 Fault diagnosis and prediction model based on Inception-BiGRU

(1) Inception network structure

To effectively extract time-frequency features from vibration signals, this study first employs CWT to convert 1D disc cutter vibration signals into 2D wavelet time-frequency graphs. This method preserves the TD and FD information of non-stationary signals within a unified representation space. Compared with traditional time-series analysis methods such as the Fourier transform (FT) or autoregressive models, CWT time-frequency graphs can simultaneously present the transient impact characteristics and long-term trend changes of signals. They are suitable for non-stationary vibration signals of disc cutter skew wear faults under complex working conditions. Through a visualized 2D time-frequency representation, the model can directly capture the dynamic energy distribution of vibration signals across different frequency bands, thus improving the distinguishability of fault categories and early recognition capability. Since the transformed data essentially takes the form of images with clear spatial structural features, using CNN as the primary feature extraction tool offers significant advantages. The CNN is a feedforward neural network structure widely applied in deep learning; it is characterized by local connectivity, weight sharing, and spatial dimensionality reduction, making it particularly suitable for feature extraction tasks on image-like data [17]. The convolutional layer performs sliding operations with convolutional kernels over input feature maps to extract spatial features from local regions. The activation layer introduces nonlinear representation capabilities to improve the network's ability to fit complex patterns. Common activation functions (e.g., Sigmoid and Tanh) suffer from gradient vanishing as their derivatives approach zero for large input absolute values [18]. This study employs ReLU as an activation function, which features a constant gradient

in the positive region, alleviating the gradient vanishing issue and improving training stability for deep networks [19]. The joint operation of the convolutional and activation layers can be expressed as:

$$x_j^{(l)} = f(\sum_{i \in M_j} x_i^{(l-1)} \cdot k_{ij}^{(l)} + b_j^{(l)}) \quad (1)$$

$x_j^{(l)}$ represents the jth output feature map of the lth layer; $x_i^{(l-1)}$ refers to the ith input feature map of the previous layer; $k_{ij}^{(l)}$ is the corresponding convolution kernel; $b_j^{(l)}$ stands for the bias term; $f(*)$ denotes the activation function.

The pooling layer is employed to downsample the convoluted feature map to decrease the dimensionality, reduce the computational effort, and suppress the risk of overfitting [19]. Maximum pooling is the most widely used pooling method, which is calculated as follows:

$$p_{i,j} = \max\{a_{i,n} | jW \le n \le jW + w\} \quad (2)$$

$p_{i,j}$ refers to the output of the jth pooling unit in the ith feature map; $a_{i,n}$ denotes the input element in the pooling window; w is the width of the pooling window; W represents the pooling step.

In CNN, increasing network depth and width can remarkably improve model expressiveness, but it also brings the risk of parameter explosion and overfitting. The Inception network structure introduces a parallel mechanism of multi-scale convolutional kernels without significantly increasing computational complexity, enhancing the model's adaptability to features of different sizes [20-22]. A typical Inception module integrates 1×1, 3×3, and 5×5 convolution operations as well as 3×3 max-pooling operations. Each branch processes the same input in parallel and fuses features in the channel dimension, achieving the unity of width expansion and feature diversity. Figure 1 shows the structural diagram of a typical Inception module. Among them, parallel paths jointly construct a multi-scale feature learning framework, providing high-quality feature inputs for subsequent sequence modeling.

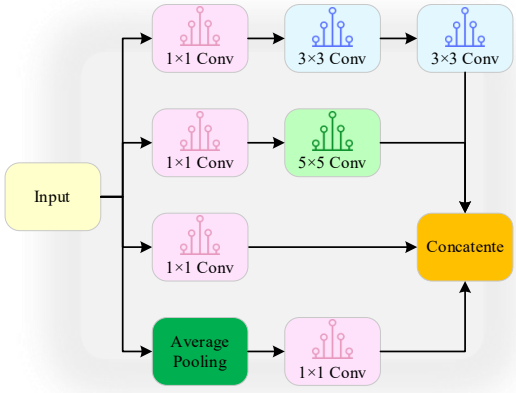

Fig. 1. Inception network structure

The Inception network in Figure 1 draws inspiration from the Network in Network (NIN) architecture, using 1×1 convolutions for dimensionality reduction in the depth direction. This approach reduces the number of parameters and introduces additional nonlinear transformation capabilities, enhancing the network's representational efficiency. The Inception module can simultaneously extract local details and global contextual features. Thus, it can demonstrate high robustness and identification ability when processing complex nonlinear time-frequency patterns such as disc cutter deflection wear faults.

(2) Bi-GRU

The time-frequency graphs generated by wavelet transforms contain rich image structural information in the spatial dimension while preserving the dynamic evolution laws of signals in the temporal dimension. For such 2D representational structures with temporal dependency, relying solely on CNN for spatial feature extraction is insufficient to capture the evolutionary patterns in the temporal dimension. The gated recurrent unit (GRU), an improved RNN structure, is characterized by fewer parameters, high training efficiency, and suitability for time-series modeling. It exhibits stronger stability in modeling long-term dependencies and effectively mitigates the gradient vanishing problem [23]. The bidirectional GRU (Bi-GRU) further expands information acquisition paths by encoding sequences through parallel forward and backward GRU networks. Thus, Bi-GRU extracts global temporal features from both past and future directions to enhance the model's representational capability and prediction performance [24].

In GRU, the hidden state update at each moment is controlled by two gate structures: the update and reset gates. The update gate regulates the proportion of information fusion between the current and previous states. In contrast, the reset gate determines the degree of reliance on the current input in relation to the previous state. The calculation for the update gate is as follows:

$$z_t = \sigma(W_{xz} x_t + b_{xz} + W_{hz} h_{t-1} + b_{hz}) \quad (3)$$

The calculation for the reset gate is:

$$r_t = \sigma(W_{xr} x_t + b_{xr} + W_{hr} h_{t-1} + b_{hr}) \quad (4)$$

The candidate hidden state is calculated as follows:

$$\tilde{h}_t = \tanh(W_{xn} x_t + b_{xn} + r_t \circ (W_{hn} h_{t-1} + b_{hn})) \quad (5)$$

The equation for updating the final hidden state is:

$$h_t = (1 - z_t) \circ \tilde{h}_t + z_t \circ h_{t-1} \quad (6)$$

x_t represents the input vector at moment t; h_{t-1} indicates the hidden state of the previous moment, and the initial state is set to an all-zero vector; $\sigma(*)$ stands for the Sigmoid activation function; $\tanh(*)$ refers to the hyperbolic tangent activation function; W and b are the weight matrices and bias terms corresponding to each gate. $\circ$ denotes the Hadamard product.

The Bi-GRU structure comprises two GRUs, which independently process the input data in the forward and reverse directions of the time series. Hidden state sequences in the two directions are obtained, which are denoted as $\vec{h}_t$ and $\overleftarrow{h}_t$. Finally, a comprehensive output representation is generated by concatenation:

$$H_t = [\vec{h}_t; \overleftarrow{h}_t] \tag{7}$$

The bidirectional structure can integrate the context information before and after the current moment. Moreover, it has a stronger ability to capture the complex temporal dynamic characteristics of the disc cutter vibration signal under abnormal wear and tear. Figure 2 illustrates the network structure of the Bi-GRU model.

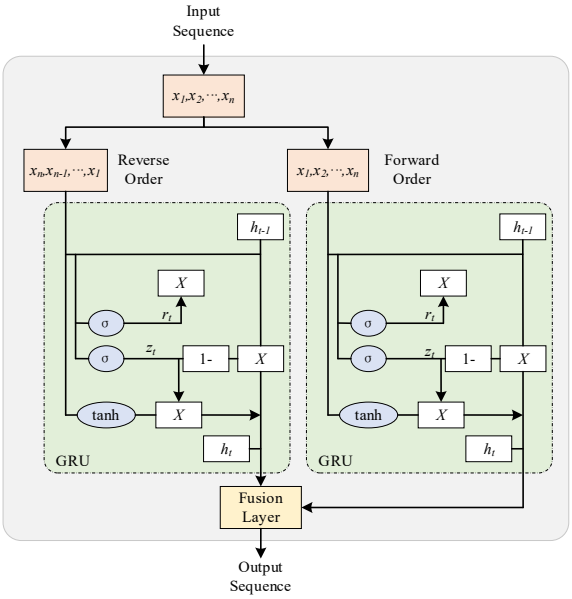

Fig. 2. The Bi-GRU structure

The output of the Bi-GRU model in Figure 2 provides multi-dimensional temporal feature representations for each time instant, offering sufficient temporal dependency support for subsequent fault identification. This effectively enhances the model's capability to capture the time-series characteristics of disc cutter vibration signals, improving the accuracy and robustness of fault diagnosis.

(3) The Inception-BiGRU fault diagnosis and prediction model

During the rock-breaking process of disc cutters, vibration signals exhibit complex non-stationary characteristics, whose time-frequency features are influenced by both multi-scale spatial variations and temporal dependencies. While the CNN is suitable for extracting local spatial structure information, it struggles to model long-term dependencies in time series. To comprehensively extract the discriminative information of disc cutter deflection wear faults from vibration time-frequency graphs, this study proposes a fault diagnosis and prediction model fusing the Inception module with Bi-GRU. The overall architecture is plotted in Figure 3. The model fully integrates the spatial feature extraction capability of Inception with the temporal modeling capability of Bi-GRU to achieve joint learning of spatial and temporal information.

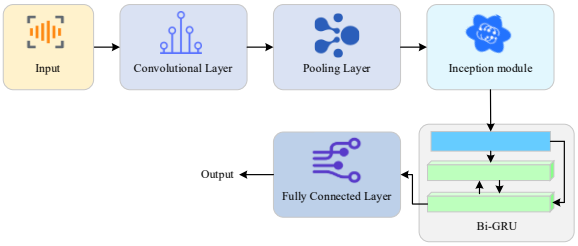

Fig. 3. Schematic diagram of the structure of the Inception-BiGRU model

The input to the Inception-BiGRU model in Figure 3 is a 2D time-frequency graph processed by CWT. The input first enters the convolutional layer to extract basic spatial features, including underlying textures and edges, combined with a max-pooling layer to perform down-sampling on local regions. This helps reduce feature dimensions and suppress local noise disturbances. The Inception module is then introduced; its internal parallel 1×1, 3×3, 5×5 convolution kernels and pooling branches allow the model to capture multi-scale features at different receptive fields. Thus, the model's ability is enhanced to represent complex textures and irregular structures. To reduce computational complexity, the Inception module introduces 1×1 convolution in the channel dimension for dimensionality reduction. This reduces the number of parameters while maintaining multi-scale feature extraction capability. The entire module is designed to be lightweight, making it suitable for deployment in resource-constrained milling machine control systems. This structure improves the collaborative perception of network width and depth, further excavating the spatial hierarchical features in time-frequency graphs. The high-dimensional feature maps output by the Inception module are further compressed and semantically enhanced through convolution and pooling operations, then unfolded into 1D vectors for input to the Bi-GRU module.

The Bi-GRU performs sequence modeling in both forward and backward temporal directions, obtaining sequence representations containing contextual dependency information through the concatenation of bidirectional hidden states. Compared with traditional Long Short-Term Memory (LSTM) networks, Bi-GRU has fewer parameters and higher computational efficiency, while retaining the ability to model long-sequence dependencies. This ensures diagnostic accuracy while reducing inference time. To further reduce the computational burden, the model performs

down-sampling on the high-dimensional feature maps output by the Inception module before inputting them into Bi-GRU. This compresses the sequence length and reduces feature dimensions while preserving key discriminative information. Besides, this model uses an end-to-end feature sharing strategy. The fault diagnosis and fault prediction modules share the same feature representation, avoiding redundant calculations and improving overall inference efficiency. After the spatial and temporal feature extraction is completed, two parallel fully connected modules are introduced at the end of the model to undertake the tasks of fault diagnosis and fault prediction, respectively. Based on Softmax, the fault diagnosis module outputs the current status category, and the output is:

$$P(y=i|x)=\frac{\exp(W_i^T x+b_i)}{\sum_j \exp(W_j^T x+b_j)} \quad (8)$$

x denotes the Bi-GRU output vector; W_i and b_i refer to the parameter of the fully connected layer; $P(y=i|x)$ indicates the probability that the input image belongs to category i. This module is responsible for fault diagnosis and prediction.

The overall model is trained in an end-to-end manner through joint optimization. The loss function considers both classification loss and regression error to ensure the simultaneous enhancement of diagnostic and predictive performance. The spatial multi-scale feature extraction capability of the Inception network, the temporal dependency modeling capability of Bi-GRU, and the strategies of feature map down-sampling and end-to-end feature sharing are deeply integrated. Thus, the Inception-BiGRU model can accurately identify and predict abnormalities in the early stage of disc cutter skew wear. It has good engineering practicality and generalization ability, while maintaining efficient operation in on-site environments with limited computational resources.

3.2 Experimental materials and settings

A standardized rock-breaking test platform is constructed under laboratory conditions for data collection to improve the fault trend prediction model's generalization ability and disc cutter deflection wear state identification. Each disc cutter undergoes five groups of rock-breaking tests with controlled variables, where the penetration depth increases by 0.5 mm sequentially. During each group of experiments, vibration acceleration signals of the disc cutter in three axial directions are synchronously collected. It obtains original TD data covering four working conditions: normal, mild, moderate, and severe deflection wear. To handle potential sample imbalance among different skew wear states, the following strategies are adopted in the data preprocessing and training phases. First, sliding windows and overlapping sampling are used to expand the number of samples for all categories. The sampling step is set to 4096 points, and the length of a single sample is 16384 points. Through this method, 3000 independent samples are constructed for each disc cutter skew wear state. Second, weighted processing is applied to the class imbalance problem during model training. Slightly higher weights are assigned to severe skew wear samples to ensure the loss function fully considers the contribution of a small number of samples during optimization, thereby reducing classification bias. Additionally, data augmentation techniques are employed to perform random, slight horizontal translation (±5% pixels), small-scale scaling (scaling factor 0.9-1.1), and random Gaussian noise injection (standard deviation of 1% of the maximum image value) on time-frequency graphs. This further enhances the diversity of minority class samples and improves the model's generalization ability for small sample sizes across different working conditions.

After collecting raw vibration signals, denoising is first performed. The wavelet threshold denoising method is adopted to apply soft threshold processing to high-frequency noise coefficients. This effectively suppresses sensor noise and environmental interference while retaining transient impact characteristics and low-frequency trend information in disc cutter vibration signals. The denoised signals are then subjected to sliding window overlapping sampling to generate training samples, ensuring sufficient and representative samples for each skew wear state. To eliminate dimensional differences between different channels and samples, all samples are normalized before being input into the model. After processing, CWT maps one-dimensional vibration signals into 2D time-frequency graphs, fully capturing the transient characteristics and frequency evolution laws of the signals. CWT's time-frequency transformation of the signal $x(t)$ can be written as:

$$w(a,b)=\frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} x(t)\varphi\left(\frac{t-b}{a}\right)dt \quad (9)$$

a refers to the scale factor, which controls the frequency resolution; b denotes the translation factor, which represents the offset on the time axis; $\varphi(t)$ is the selected wavelet basis function. The transform exhibits adaptive scaling and multi-resolution properties, making it well-suited for capturing transient features in vibration signals. In CWT, the Morlet wavelet basis function is selected. The range of its center frequency and scale factor is adjusted through experiments, and finally determined as $a\in[1, 64]$. This range can cover the main frequency feature bands of disc cutter vibration signals, ensuring that both low-frequency impacts and high-frequency transient information are fully expressed. By adjusting different scale combinations, it is found that the time-frequency graphs generated within this range achieve the best performance in fault state

distinguishability and model training convergence speed. In a specific implementation, the 2D images generated after wavelet transformation are unified to a size of 60×60 pixels for constructing model inputs. Samples of each deflection wear state are divided into training and test sets at a ratio of 4:1, with 2400 samples for training and 600 samples for model testing. During the training phase, the model performs classification tasks and introduces a prediction module for the disc cutter performance degradation trend. This module outputs the amplitude response or characteristic index change trends in several future time windows through a regression approach, enhancing the model's predictive capability.

The Inception-BiGRU model training is developed using the PyTorch deep learning framework. All experiments are completed in the local environment. In terms of network structure design, the model complexity is controlled to match the data scale. The Inception module performs dimensionality reduction through 1×1 convolution, effectively reducing the number of parameters. Meanwhile, Dropout layers are introduced after each convolutional layer in the Inception module and before the Bi-GRU layer, with a dropout rate of 0.5 to prevent model overfitting. Parameters such as the convolution kernel size, Inception module branches, pooling stride, and number of convolutional layer channels in the model network structure are optimized through grid search and cross-validation. The combination of convolution kernel sizes 1×1, 3×3, and 5×5 is selected to balance local texture and global structural feature extraction; the maximum pooling stride and number of convolution output channels achieve a balance between validation set accuracy and training time, ensuring high precision while reducing computational load. The Bi-GRU's number of hidden units and layers is determined through multiple rounds of experiments on the training and validation sets; a single-layer Bi-GRU structure with 64 units is finally adopted. Sensitivity analysis results comparing different hidden units [32, 64, 128] and layers [1, 2, 3] show that the single-layer Bi-GRU with 64 units achieves the optimal balance in accuracy, F1-score, and training efficiency; it avoids parameter redundancy and overfitting risks caused by excessive layers or overly large unit counts. Specific results are exhibited in Table 1.

Table 1. Sensitivity analysis results of hidden units and layers of the Bi-GRU

Number of hidden units	Layers	Average accuracy (%)	F1 score	Average training time/epoch (s)
32	1	0.965	0.956	18.5
32	2	0.968	0.958	29.2
32	3	0.970	0.960	41.0
64	1	0.986	0.985	22.1
64	2	0.987	0.986	34.5
64	3	0.987	0.986	48.3
128	1	0.987	0.986	37.8
128	2	0.988	0.987	55.4
128	3	0.988	0.987	74.1

Table 2. Experimental environment and parameter settings

Item/Parameter name		Configuration/Setting values
Experimental platform environment	Operating system	Windows 10
	Programming language	Python 3.8
	Deep learning framework	PyTorch 1.x
	Central processing unit	Intel Core i5-10400F
	Memory	16 GB
	Graphics processing unit	NVIDIA GeForce GTX1650 4 GB
Model training parameters	Sample length	16384 points
	Overlap step size	4096 points
	Image size	60×60
	Batch size	32
	Learning rate	5.7×10^{-4}
	Epoch	200
	Loss function	Cross-Entropy Loss
	Optimizer	Adam

In the model training phase, two key strategies actively prevent overfitting. (1) Early stopping: The loss on the validation set is continuously monitored. Training is terminated when the validation set loss stops decreasing for 15 consecutive epochs, and the model weights with the best performance on the validation set are rolled back. This effectively prevents the model from overlearning noise and irrelevant details in the training set. (2) L2 weight decay: An L2 regularization term is set in the Adam optimizer with a weight decay coefficient of 1×10^{-5}. It encourages the model to learn simpler and more generalizable patterns by penalizing large weight values of the model. Before training, all image data undergoes unified normalization processing to ensure consistent distribution across input channels and eliminate training biases caused by dimensional differences. The entire model simultaneously outputs fault state classification results during the training phase, adopting a joint loss optimization strategy to enhance the model's multi-task learning capability. Accuracy and F1 score are selected as the main evaluation indicators for the model's diagnostic and predictive performance. These indicators comprehensively measure the classification model's accuracy and robustness in identifying multi-category disc cutter deflection wear states.

4. RESULTS AND ANALYSIS

4.1 Analysis of model fault diagnosis and prediction results

Figure 4 illustrates the variations in accuracy and loss function throughout the training and testing phases of the Inception-BiGRU model.

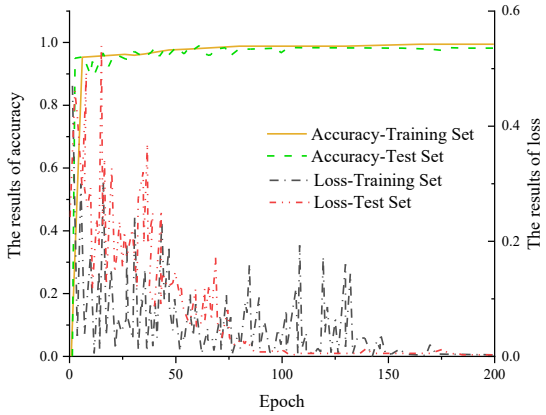

Fig. 4. The curve of the training process curve of the Inception-BiGRU model

Figure 4 shows that the accuracy of the Inception-BiGRU model steadily increases to approximately 99% during training, accompanied by a pronounced reduction in the loss function, confirming the model's strong fitting ability. The test set accuracy remains stable at around 95% with minimal fluctuations, indicating strong generalization ability. In the later training stage, the loss further decreases, and the accuracy stabilizes, without obvious overfitting. This confirms the effectiveness and robustness of the model structure in disc cutter vibration signal identification tasks. High generalization performance means the model can adapt to signal changes under different stratum conditions, impact loads, and vibration interference. It can provide a reliable decision-making basis for the online monitoring system of milling machines, thereby reducing unplanned downtime and improving construction efficiency.

To verify the model's diagnostic performance under variable working conditions, test experiments are conducted with different disc cutter speeds and penetration depths, with diagnostic results indicated in Figure 5.

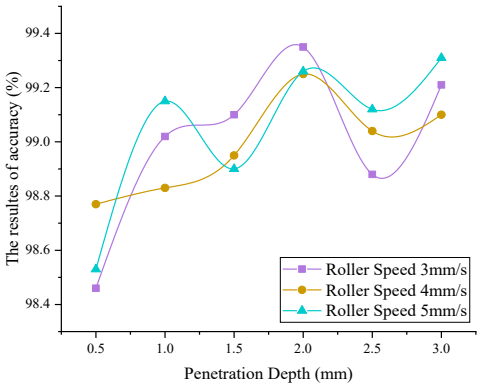

Fig. 5. Diagnosis results of different working conditions

In Figure 5, the Inception-BiGRU model maintains an identification accuracy of over 98% across various working conditions of diverse disc cutter speeds and penetration depths. The model's maximum value reaches 99.35%, demonstrating stable performance. With the increase in penetration depth, the accuracy slightly improves overall, indicating that deep rock-breaking signals contain more identifiable features. Different disc cutter speeds have little impact on the results. The model shows good robustness to speed changes, strong adaptability to complex working conditions, and practical application potential. This indicates that the model is not only applicable to laboratory conditions but also competent for highly complex on-site environments such as tunnel construction and underground pipe gallery excavation. Particularly, its ability to maintain high reliability under deep penetration and high-load working conditions enables it to identify the wear trend of disc cutters. This capability is crucial when stratum conditions change suddenly, or operation intensity increases, as it helps avoid reduced rock-breaking efficiency or equipment damage caused by aggravated wear.

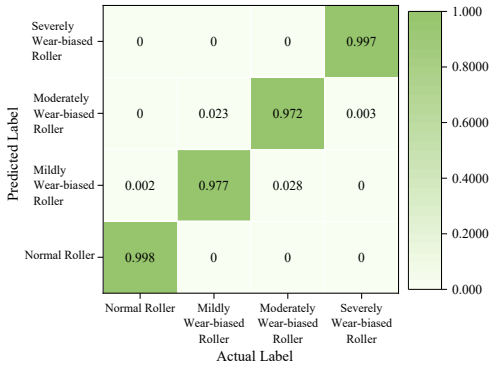

Fig. 6. Accuracy of the Inception-BiGRU model's fault prediction confusion matrix

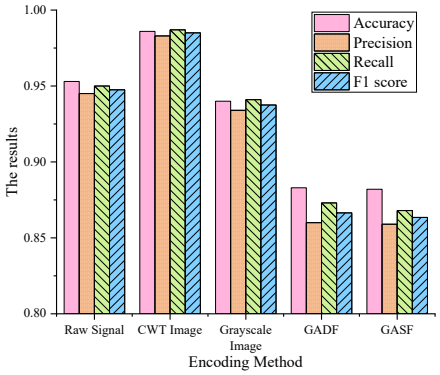

Fig. 7. The influence of encoding methods on identification performance

The model's identification capability in multi-classification tasks is further evaluated by drawing a fault prediction confusion matrix based on the test set, as demonstrated in Figure 6.

Figure 6 shows that the proposed Inception-BiGRU model achieves a prediction accuracy of 99% for both normal and severely deflected wear disc cutters. The proposed model has a minimal probability of misjudgment, demonstrating extremely high identification reliability. Partial confusion occurs in mild and moderate deflection wear disc cutters, with prediction accuracies of 97.7% and 97.2%, respectively, indicating some feature overlap in the feature space. However, the overall identification accuracy remains above 97%, confirming the model's good discrimination capability and engineering practical applicability. Severe skew wear is usually the main factor leading to disc cutter failure, abnormal equipment vibration, and construction interruption. The model's highly reliable identification of severe skew wear can effectively support proactive maintenance and tool replacement strategies, thereby reducing economic losses caused by unplanned downtime. Meanwhile, the model's ability to distinguish between mild and moderate skew wear means it can identify wear evolution trends. This capability transforms maintenance from a passive response to an active prediction, allowing it to serve the equipment health management system.

4.2 Comparative analysis of diverse coding methods

The wavelet time-frequency graph is compared with commonly used image encoding methods in the fault diagnosis. These methods include grayscale images, Gramian Angular Difference Field (GADF), Gramian Angular Summation Field (GASF), and original vibration signals. The accuracy and F1 score results are suggested in Figure 7.

Figure 7 reveals that by comparing different encoding methods, the CWT image performs best in fault diagnosis, with an accuracy of 0.986 and an F1 score of 0.985. Meanwhile, CWT represents improvements of 3.46% and 3.96% compared to the original signals, respectively, and substantially outperforms other image encoding methods. The grayscale image performs slightly worse with an F1 score of 0.937, while GADF and GASF exhibit similar performance, both significantly lower than CWT. This demonstrates that CWT can more effectively extract the time-frequency features of non-stationary vibration signals, facilitating accurate discrimination of disc cutter deflection wear states. Therefore, the CWT image encoding method shows remarkable advantages in feature representation capability and classification performance. This finding illustrates that CWT is an optimal choice at the signal analysis level; it also demonstrates the potential to improve diagnostic stability and accuracy in engineering applications. Since disc cutter vibration signals are inherently strongly non-stationary, wavelet time-frequency graphs can capture transient impacts and wear features. This enables the model to accurately identify different wear levels and improve the reliability of on-site diagnosis. Therefore, the CWT-encoded model is better suited for integration into real-time monitoring systems, contributing to enhanced operational and maintenance intelligence.

4.3 Comparative analysis of different algorithms

Common models such as Support Vector Machine (SVM), DNN, CNN, Residual Network (ResNet), CNN-BiGRU, and CNN-LSTM are selected for performance comparison with the proposed Inception-BiGRU model. All CNN hyperparameters are consistent with those of the Inception-BiGRU model. Each model undergoes ten identical training and testing cycles to reduce the impact of random errors, with the ten test results depicted in Figure 8.

Table 3. Comparison of downtime and economic losses before and after applying the intelligent diagnostic method

Condition	Average downtime per incident	Monthly failure frequency	Economic loss per downtime	Total monthly economic loss
Without using predictive methods	6 hours	8	20,000 yuan	160,000 yuan
Using predictive methods	4 hours	8	13,300 yuan	106,400 yuan

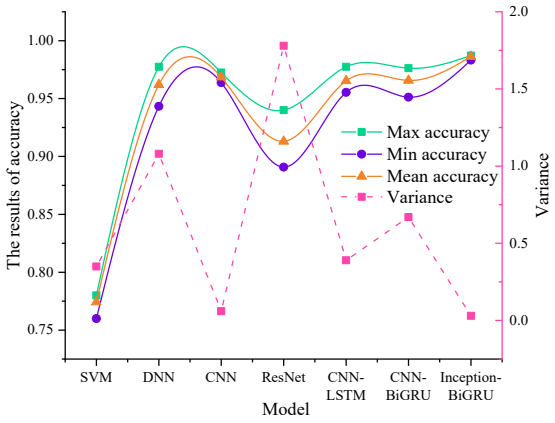

Fig. 8. The results of ten tests of different models

In Figure 8, the Inception-BiGRU model demonstrates the best comprehensive performance, with an average accuracy of 0.9863 and a variance of only 0.03, indicating stable performance. Compared with the best-performing CNN, the average accuracy is improved by 1.84%, and the variance is reduced by 50%. In contrast, SVM has the lowest accuracy of 0.7742 with large fluctuations (variance = 0.35). Although DNN, CNN-LSTM, and CNN-BiGRU all achieve accuracies above 0.96, their stability is slightly poorer, with variances of 1.08, 0.39, and 0.67, respectively. ResNet shows the most unstable performance, with a variance of 1.78. In summary, Inception-BiGRU outperforms other models in both accuracy and robustness, making it the most suitable for disc cutter fault diagnosis. Stable recognition performance helps reduce false alarms and missed detections, providing a more reliable basis for predictive maintenance on engineering sites while improving operational continuity and safety.

4.4 Practical application and economic benefit analysis

Skew wear faults of milling machine cutterheads often lead to equipment downtime in tunnel boring and underground engineering, resulting in substantial non-productive time and directly increasing construction costs and economic losses. Based on engineering practice experience, each fault shutdown delays the construction schedule while increasing labor, material, and maintenance costs. To intuitively evaluate the proposed method's economic value in engineering applications, a simulation case is designed and calculated based on typical milling machine operating parameters and fault frequencies. It is assumed that in a tunnel construction project, the milling machine operates 8 hours per day and 20 days per month. Without predictive intervention, the cutterhead experiences mild to moderate skew wear faults on average once every two weeks, with each shutdown for maintenance taking approximately 6 hours. The direct economic loss caused by a single shutdown (including labor, equipment idleness, and construction delay costs) is about 20,000 yuan.

After applying the Inception-BiGRU intelligent diagnosis method for real-time monitoring and prediction, it is assumed that 80% of potential faults can be identified in advance. This enables planned maintenance and reduces the unplanned downtime per fault from 6 hours to 4 hours. According to simulation calculations, the comparison of monthly downtime and economic losses is detailed in Table 3.

The simulation case shows that through early diagnosis and planned maintenance, the monthly economic loss can be reduced from 160,000 yuan to approximately 106,400 yuan; it saves 53,600 yuan (about 33.5% of the total loss without the method). This analysis indicates that the Inception-BiGRU intelligent diagnosis method has economic benefits in practical engineering. It reduces unplanned downtime while improving the equipment utilization rate and construction efficiency of milling machines. Meanwhile, this simulation result provides a quantitative reference for the promotion and application of the method in intelligent operation and maintenance systems. This further highlights its engineering value and practical application potential.

5. CONCLUSION

This study focuses on the common fault of skew wear in disc cutters in milling machines, which has significant engineering implications for tunnel and underground construction projects. It proposes an Inception-BiGRU-based intelligent diagnosis and prediction model. It systematically explores the application potential of multi-scale time-frequency features of vibration signals in identifying complex wear states. Affected by complex loads, rock mass inhomogeneity, and dynamic impact conditions, the

fault characteristics of disc cutter skew wear are characterized by strong concealment, complex evolution patterns, and significant signal non-stationarity. This study proposes encoding vibration signals into 2D time-frequency graphs via CWT to present the dynamic evolution process of skew wear. It aims to address the problem that traditional methods struggle to capture their time-frequency variation laws. By combining the multi-scale convolution structure of Inception with the temporal dependency modeling capability of Bi-GRU, more effective characterization of different wear degrees is achieved. Multi-working condition tests reveal that the model exhibits strong robustness to different disc cutter speeds and penetration depths, with diagnostic accuracy stably maintaining above 98%, illustrating excellent adaptability. In the fault multi-classification experiment, the model accurately identifies normal and severe deflection wear states. Concurrently, minor confusion exists between mild and moderate deflection wear states, reflecting the continuity and complexity of fault states in actual working conditions. Comparisons with multiple encoding methods suggest that wavelet time-frequency graphs can more effectively capture the time-frequency features of vibration signals, markedly improving the model's diagnostic performance. Multi-model comparisons validate the advantages of Inception-BiGRU in terms of accuracy and stability, demonstrating its strong potential for application in disc cutter fault diagnosis.

The Inception-BiGRU model, constructed based on vibration signals in this study, exhibits high accuracy and robustness in diagnosing and predicting disc cutter skew wear. However, relying on a single vibration data source still has certain limitations. Vibration characteristics affect the wear and fault development of disc cutters. They are also closely related to multiple factors, including cutting force, cutterhead temperature, equipment load, rock mass hardness, and construction conditions. A single vibration signal may fail to reflect the interaction among these complex factors. Especially in the early stages of minor or sudden faults, diagnostic features may not be significant enough, thus affecting the model's sensitivity and prediction accuracy. Consequently, future work can consider multi-modal fusion of vibration signals with data from other sensors (such as force sensors and temperature sensors) and equipment operating parameters. Through collaborative modeling of multi-source information, the model's ability to identify minor or complex faults and its robustness can be further improved. This extension helps comprehensively characterize the dynamic evolution process of disc cutter wear. Meanwhile, it provides a development direction for constructing a more accurate and intelligent fault diagnosis and prediction system for milling machines.

Source of funding: *This study was supported by the State Grid Corporation of China Science and Technology Project Funding: Research on Key Technologies for Mechanized Construction of Large Aperture Foundations in Mountainous Areas Based on Cableway Transportation Conditions (SGHEJSOOJGJS2400186).*

Authors contributions: *Research concept and design, J.W., J.D.; Collection and/or assembly of data, Q.S., L.H., J.T., J.L.; Data analysis and interpretation, J.D., Q.S., L.H.J.T., J.L.; Writing the article, J.W.; Critical revision of the article, J.W.; Final approval of the article, J.W., J.D.*

Declaration of competing interest: *The author declares no conflict of interest.*

REFERENCES

1. Zhou H, Huang Q, Zhou C, He P, Zhe N, Wang H. Rotating machinery fault diagnosis method based on temporal-spatial vibration feature fusion extraction. IEEE Sensors Journal. 2024;25(1):1184-1197. https://doi.org/10.1109/JSEN.2024.3496776.
2. Tambake N, Deshmukh B, Pardeshi S, Sachin Salunkhe S, Cep R, Nasr EA. Fault diagnosis of a CNC hobbing cutter through machine learning using three axis vibration data. Heliyon. 2025;11(2):e41637. https://doi.org/10.1016/j.heliyon.2025.e41637.
3. Jang JG, Noh CM, Kim SS, Shin SC, Lee SS, Lee JC. Vibration data feature extraction and deep learning-based preprocessing method for highly accurate motor fault diagnosis. Journal of Computational Design and Engineering. 2023;10(1):204-220. https://doi.org/10.1093/jcde/qwac128.
4. Divya D, Marath B, Santosh Kumar MB. Review of fault detection techniques for predictive maintenance. Journal of Quality in Maintenance Engineering. 2023; 29(2):420-441. https://doi.org/10.1108/JQME-10-2020-0107.
5. Zhou H, Yan P Huang Q, Wu D, Pei J, Zhang L. Weighted average selective ensemble strategy of deep convolutional models based on grey wolf optimizer and its application in rotating machinery fault diagnosis. Expert Systems with Applications. 2023;234:121076. https://doi.org/10.1016/j.eswa.2023.121076.
6. Mohammed OD. Dynamic modelling and fault diagnosis of a high contact ratio gear. In: Belhaq M. (eds) Advances in Nonlinear Dynamics and Control of Mechanical and Physical Systems. CSNDD INCREASE. 2023. Springer Proceedings in Physics. 2023;301. https://doi.org/10.1007/978-981-99-7958-5_16.
7. Cong F, Zhou Q, Chen L, Lin F, Lin X, Zhou Y. Hob wear state condition monitoring based on statistical distribution law. CIRP Journal of Manufacturing Science and Technology. 2023;44:16-26. https://doi.org/10.1016/j.cirpj.2023.04.007.
8. Emtaubel S. Bearing misalignment and eccentric wear:

A study on condition monitoring. International Journal of Engineering Research. 2024;3(1):77-94.
9. Pandiyan M, Babu TN. Systematic review on fault diagnosis on rolling-element bearing. Journal of Vibration Engineering & Technologies. 2024;12(7): 8249-8283. https://doi.org/10.1007/s42417-024-01358-4.
10. Salunkhe VG, Khot SM, Yelve NP, Jagadeesha T, Desavale RG. Rolling element bearing fault diagnosis by the implementation of Elman neural networks with long short-term memory strategy. Journal of Tribology. 2025;147(8):084301. https://doi.org/10.1115/1.4067382.
11. Chen Y, Zhang D, Yan R. A trend domain adaptation approach with dynamic decision for fault diagnosis of rotating machinery equipment. IEEE Transactions on Industrial Informatics. 2024;21(3):2084-2093. https://doi.org/10.1109/TII.2024.3488780.
12. Sahu AR, Palei SK, Mishra A. Data-driven fault diagnosis approaches for industrial equipment: A review. Expert Systems. 2024;41(2):e13360. https://doi.org/10.1111/exsy.13360.
13. Kibrete F, Woldemichael DE, Gebremedhen HS. Multi-Sensor data fusion in intelligent fault diagnosis of rotating machines: A comprehensive review. Measurement. 2024;232:114658. https://doi.org/10.1016/j.measurement.2024.114658.
14. Wang J, Qiao L, Lv H, Lv Z. Deep transfer learning-based multi-modal digital twins for enhancement and diagnostic analysis of brain MRI image. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2022;20(4):2407-2419. https://doi.org/10.1109/tcbb.2022.3168189.
15. He Y, Zhao C, Shen W. Cross-domain compound fault diagnosis of machine-level motors via time–frequency self-contrastive learning. IEEE Transactions on Industrial Informatics. 2024;20(7):9692-9701. https://doi.org/10.1109/TII.2024.3384603.
16. Mahesh TR, Chandrasekaran S, Ram VA, Kumar VV, Vivek V, Guluwadi S. Data-driven intelligent condition adaptation of feature extraction for bearing fault detection using deep responsible active learning. IEEE Access. 2024;12:45381-45397. https://doi.org/10.1109/ACCESS.2024.3380438.
17. Hameed S, Junejo F, Amin I, Qureshi AK, Tanoli IK. An intelligent deep learning technique for predicting hobbing tool wear based on gear hobbing using real-time monitoring data. Energies. 2023;16(17):6143. https://doi.org/10.3390/en16176143.
18. Zhang B, Li H, Kong W, Fu M, Ma J. Early-stage fault diagnosis of motor bearing based on kurtosis weighting and fusion of current–vibration signals. Sensors. 2024; 24(11):3373. https://doi.org/10.3390/s24113373.
19. Saeed A, Khan MA, Akram U, Obidallah WJ, Jawed S, Ahmad A. Deep learning based approaches for intelligent industrial machinery health management and fault diagnosis in resource-constrained environments. Scientific Reports. 2025;15(1):1114. https://doi.org/10.1038/s41598-024-79151-2.
20. Xu L, Teoh SS, Ibrahim H. A deep learning approach for electric motor fault diagnosis based on modified InceptionV3. Scientific Reports. 2024;14(1):12344. https://doi.org/10.1038/s41598-024-63086-9.
21. Wei L, Peng X, Cao Y. Enhanced fault diagnosis of rolling bearings using an improved inception-LSTM network. Nondestructive Testing and Evaluation. 2025;40(7):3274-3293. https://doi.org/10.1080/10589759.2024.2402549.
22. Shang Z, Zhang J, Li W, Qian S, Gao M. A domain adversarial transfer model with inception and attention network for rolling bearing fault diagnosis under variable operating conditions. Journal of Vibration Engineering & Technologies. 2024;12(1):1-17. https://doi.org/10.1007/s42417-022-00823-2.
23. Lee B, Kim Y, Lee H, Kang C. Bidirectional gated recurrent unit neural network for fault diagnosis and rapid maintenance in medium-voltage direct current systems. Sensors. 2025;25(3):693. https://doi.org/10.3390/s25030693.
24. Dong Z, Zhao D, Cui L. Rotating machinery fault classification based on one-dimensional residual network with attention mechanism and bidirectional gated recurrent unit. Measurement Science and Technology. 2024;35(8):086001. https://doi.org/10.1088/1361-6501/ad41fb.

Jingde WANG, born in 1990, holds a Master's degree. He is engaged in the construction management of power transmission and transformation projects, with expertise in mechanized construction and digital applications of power grid infrastructure projects.
e-mail: zenengydb6951@163.com

Jian DUAN, born in 1982, holds a Master's degree. He is engaged in the construction and management of power transmission and transformation projects, with expertise in mechanized construction and digital applications of power grid infrastructure projects.
e-mail: choucx5@163.com

Qingsong SUN, born in 1995, holds a Master's degree. He is engaged in the construction management of power transmission and transformation projects, with expertise in mechanized construction of power grid infrastructure projects.
e-mail: daimts56459@163.com

Lu HAN, born in 1996, holds a Master's degree. He is engaged in the construction management of power transmission and transformation projects, with expertise in mechanized construction of power grid infrastructure projects.
e-mail: gdtu3185@163.com

Jian TANG, born in 1986, holds a Bachelor's degree. He is engaged in the design and development of project drilling equipment and drilling tools, with expertise in electronic control systems, hydraulic systems and marketing management.
e-mail: niix35800@163.com

Jian LI, born in 1978, holds a Bachelor's degree. He is engaged in the design and development of project drilling equipment and drilling tools, with expertise in electronic control systems, hydraulic systems and marketing management.
e-mail: lijian2025@ari.ac.cn