Renewable Energy and Power Quality Journal
https://repqj.com/index.php/repqj/index
RE&PQJ, Vol. 23, No. 3, 2025

Construction of Power Equipment Fault Prediction and Diagnosis Model Based on LightGBM and Particle Swarm Optimization

Jing Wang[1], Ya Yang[1,*], Qin Ding[2]

[1]School of Intelligent Manufacturing, Wuhu Institute of Technology, Wuhu, 241006, Anhui Province, China
[2]Department of Neurology, The Second Affiliated Hospital of Wannan Medical College, Wuhu, 241000, Anhui Province, China
[*]Corresponding author's email: yangyayy68@hotmail.com

Abstract. With the increasing intelligence of power systems and the complexity of equipment, traditional fault prediction based on experience or statistical methods is difficult to cope with multi-dimensional nonlinear data. Existing research has limited accuracy in high-dimensional features, complex working conditions and cross-device applications. This paper combined LightGBM with particle swarm optimization (PSO), used LightGBM (Light Gradient Boosting Machine) to model nonlinear features and handle sample imbalance, and used PSO to achieve automatic hyperparameter tuning to build a high-precision and highly adaptable power equipment fault prediction and diagnosis model. In the design of the intelligent fault identification system, this paper first processed the raw data such as voltage, current, temperature, vibration, etc., and extracted time series statistics, frequency domain features and change rate features to construct input variables. Then, the LightGBM model is used to establish a nonlinear mapping relationship between the equipment operation status and the fault mark, and the PSO algorithm is introduced to guide the particle iterative search for the optimal hyperparameter combination of LightGBM by minimizing the validation set loss function as the fitness function. Finally, this paper used the SMOTE (Synthetic Minority Over-sampling Technique) method to enhance fault class samples to alleviate the sample imbalance problem, and set category weights to improve recognition accuracy. Experiments show that the proposed method has a maximum fault identification accuracy of 0.84 and a recall rate of 0.95 (partial discharge). The measured detection delay on the edge computing platform is as low as 9.1 seconds (insulation degradation), and the fault classification F1 score is 0.94 (partial discharge), which are significantly better than traditional models. The area under the ROC curve is 0.87, achieving high-precision and low-latency power fault warning.

Key words. Power Equipment Fault Prediction, Machine Learning Models, LightGBM; Particle Swarm Optimization, Imbalanced Sample

1. Introduction

The complexity and diversity of the operating status of power system equipment are constantly increasing, making it difficult for traditional fault diagnosis methods that rely on manual experience or linear statistical feature modeling to meet the needs of actual engineering applications. Modern power equipment widely integrates sensor systems, which can collect multi-source and multi-dimensional operating parameter data in real time. However, in the context of the rapid growth of multi-source data, how to accurately identify potential faults has become an important challenge [1,2]. Equipment failures can cause power outages and economic losses, and are more likely to bring about systemic risks. Therefore, high-precision and scalable prediction models are critical to ensuring the safe operation of power grids [3,4].

The currently widely used machine learning methods have limited processing capabilities for high-dimensional features, are prone to accuracy fluctuations under multiple working conditions, and have weak generalization capabilities [5-7]. Although traditional methods such as SVM (Support Vector Machine), random forest, and neural network can capture some nonlinear relationships, the models are prone to overfitting or recognition bias in situations where data distribution is complex and power conditions are dynamically changing [8,9], making it difficult to meet the dual requirements of stability and sensitivity in practical applications. Researchers have attempted to improve the prediction effect by introducing more complex model structures or optimization algorithms, and have achieved performance improvements to a certain extent. However, most existing models rely on static parameter settings and empirical feature engineering, lacking dynamic adaptive capabilities, which limits their application value in multiple devices and scenarios [10]. At the same time, some studies have ignored the impact of sample imbalance in prediction tasks. In power systems, normal operating data is far more than fault data, which makes the model more likely

to be biased towards the main class samples during training, resulting in missed detection and reducing the effectiveness of actual warnings [11,12]. In addition, some methods introduce complex deep network structures, which improve the prediction performance, but also bring high computing resource consumption and deployment complexity, which is not conducive to rapid application in edge devices or real-time systems.

Existing research mainly focuses on the modeling and prediction mechanism design of power equipment status. Most studies extract statistical features, transformation features or spectral features from status data as input variables and use machine learning algorithms for training. In terms of model selection, traditional methods such as random forest, support vector machine and gradient boosting tree are widely used and show good results in some specific scenarios. In terms of method optimization, some studies have manually set model parameters or used grid search to adjust parameters, but this process has high computational cost and is prone to fall into local optimality, making it difficult to adapt to large-scale data processing and real-time prediction requirements.

In order to improve the adaptability of the prediction model under non-stationary and multi-operating conditions, an adaptive optimization strategy is introduced to adjust the model hyperparameters. For example, using swarm intelligence methods such as genetic algorithms, ant colony algorithms, and PSO algorithms to automatically search for the optimal configuration in the parameter space can reduce the degree of manual intervention and improve model performance. Among them, PSO has attracted attention due to its simple structure, fast convergence speed, and adaptability to high-dimensional optimization problems. By simulating the group collaboration behavior of individuals in the solution space, the search path is gradually updated iteratively, the optimal point is found in the parameter combination space, and the overall performance stability of the model is improved.

In recent years, gradient boosting tree algorithms have been increasingly used in power data modeling. Among them, LightGBM has become an important tool for power equipment fault prediction research due to its high computational efficiency, support for large-scale features, ability to handle missing values, and built-in category feature processing mechanism [13,14]. Compared with XGBoost (Extreme Gradient Boosting), LightGBM shows better speed and memory performance when processing large-scale training data, and reduces modeling costs while maintaining accuracy. Many studies have shown that LightGBM is highly sensitive to nonlinear and cross-features, and can more effectively capture the implicit associations between complex states and fault markers. After combining with intelligent optimization algorithms, the prediction effect and generalization ability of LightGBM are further enhanced, especially for scenarios with high-dimensional input and unbalanced label distribution [15,16].

Although the above methods have achieved certain breakthroughs in performance, most current studies still rely on static strategies in the parameter adjustment stage, ignoring the nonlinear effects of dynamic changes and parameter interactions during model training, which makes the model prone to local optimality [17,18]. Some studies introduced PSO to adjust parameters, but in the actual modeling process, there is a lack of systematic integration of search space, particle behavior and objective function structure, which affects the algorithm convergence speed and parameter stability [19,20]. In addition, power equipment data has significant time series and non-stationarity. Some studies have not properly processed the sample partitioning strategy and time series feature structure, resulting in information leakage risks between the training set and the validation set [21,22], which ultimately affects the authenticity of the evaluation results and the actual effect after the model is deployed.

This paper aims at the fault prediction needs of power equipment under complex working conditions and constructs a prediction model that integrates LightGBM and PSO. In the overall process, the multi-dimensional raw data is first subjected to sliding window denoising, outlier processing and normalization operations to construct a set of input variables containing time domain statistics, frequency domain transformation and rate of change characteristics. PSO uses the loss function on the cross-validation set as the fitness evaluation indicator, and automatically adjusts the key hyperparameters in LightGBM such as the learning rate, maximum depth, and number of leaves through global particle search and local optimal update mechanism until the set convergence conditions are met or the maximum number of iterations is reached, thereby improving the robustness and generalization ability of the model. In terms of sample processing, this paper introduces the SMOTE method to expand minority class samples and combines the category weight mechanism to solve the recognition offset problem caused by sample imbalance. In the model verification stage, a timestamp-based sample stratification strategy is adopted to ensure that there is no time overlap between the training and verification sets, thereby improving the credibility and generalizability of the results. The proposed method performs well in fault identification, with the highest accuracy reaching 0.84 and the recall rate of partial discharge fault reaching 0.95. The detection delay of insulation degradation fault is as low as 9.1 seconds, and the F1 score of partial discharge fault classification reaches 0.94, which are significantly better than traditional technologies. The area under the ROC curve reaches 0.87, indicating that the method can achieve high-precision and low-latency power fault warning.

In view of the strong time series and multi-condition change characteristics of power equipment operation data, this paper uses LightGBM to perform structured modeling of multi-dimensional features, and uses its gradient-based splitting strategy to effectively identify nonlinear patterns across time periods. Compared with traditional sequence models, LightGBM has higher

advantages in training speed and computing resource usage, and is particularly suitable for edge deployment environments that are sensitive to response time. Through the above methods, this paper aims to improve the prediction accuracy and diagnostic response capabilities of power equipment under multiple working conditions and multiple feature scenarios, and to build a set of intelligent fault prediction frameworks with practical value and scalability.

The main contributions of this paper include: (1) Proposing a joint optimization framework based on LightGBM and PSO to solve the local optimal problem of traditional models under complex working conditions through dynamic parameter search; (2) Integrating SMOTE and category weight calibration strategy to systematically alleviate the recognition bias caused by the imbalance of fault samples; (3) Designing a time series-aware data partitioning strategy to avoid the risk of information leakage and improve the generalization ability of the model in actual deployment. The subsequent chapters of this paper are arranged as follows: Section 2 describes the framework design of the intelligent fault identification system in detail; Section 3 shows the experimental results and performance comparison; Section 4 summarizes the research conclusions and looks forward to future directions.

2. Design of Intelligent Fault Identification System

Regarding the overall method design of the intelligent fault identification system for power equipment, the framework diagram is shown in Figure 1:

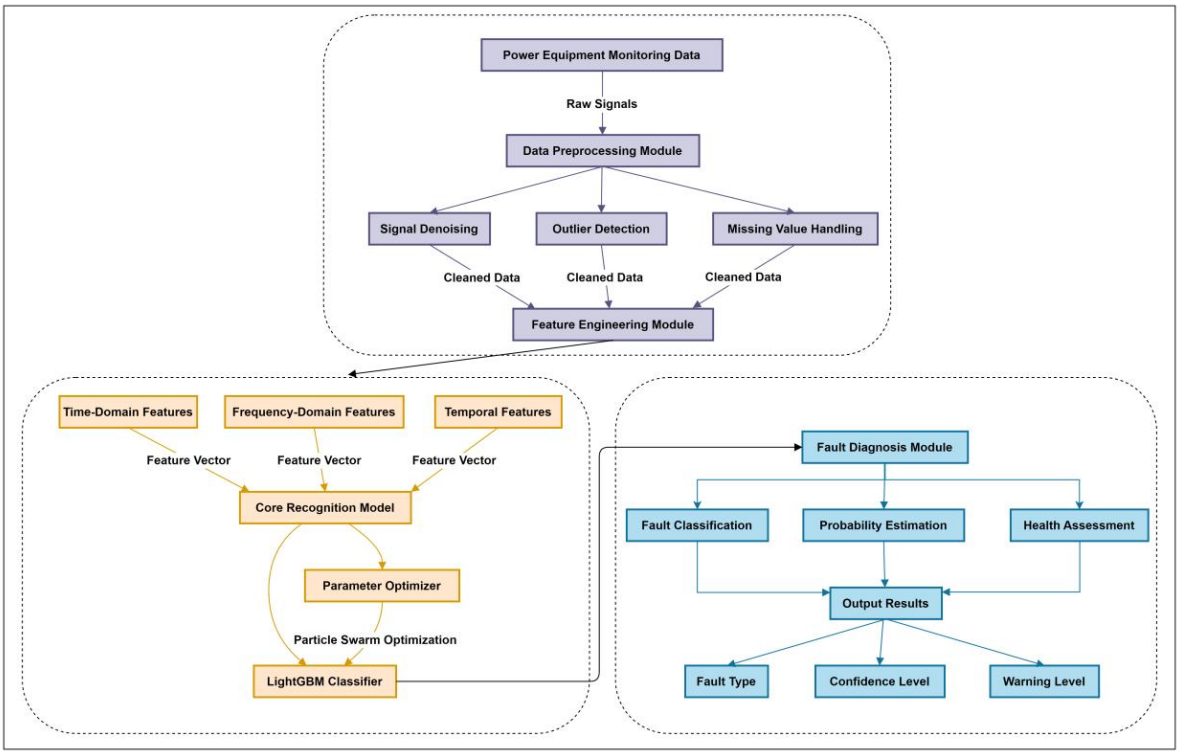

Figure 1. Overall framework of the intelligent identification system for power equipment faults.

Figure 1 shows the overall framework of the intelligent identification system for power equipment faults, which includes three core modules: the data acquisition and preprocessing module is responsible for raw signal denoising, anomaly detection and missing value processing, and generates high-quality feature data. The intelligent identification model module uses the LightGBM classifier combined with PSO to implement feature analysis and complete fault feature extraction. The diagnosis and output module accurately classifies the fault type and generates confidence assessment and warning level. The system adopts modular design, and through feature engineering optimization and parameter adaptive adjustment, it realizes the whole process processing from raw monitoring data to fault diagnosis results, ensuring that the output results have practical engineering value.

A. Abnormal Data Cleaning and Feature Construction

1) Multi-Dimensional Raw Monitoring Data Preprocessing Process

The power equipment status data is collected in real time by a variety of sensors deployed on site, mainly including multiple types of continuous variables such as current, voltage, temperature, vibration acceleration, etc. Considering the interference factors such as missing, anomaly, outlier and noise fluctuation in the original data, data quality control and standardization should be carried out before modeling [23,24]. This study firstly verifies the time integrity of each type of signal and removes the sections with unstable sampling frequency or interrupted recording. On the premise of keeping the time series information intact, the sliding window averaging method

(Window Size = 5) is used to smooth the noise of each variable to suppress the influence of high-frequency interference items on the fault feature judgment; at the same time, the triple median absolute deviation method (Median Absolute Deviation, MAD) is introduced to identify local mutation points, remove or replace outlier samples, and ensure the consistency of sample distribution.

In the standardization phase, the Z-score normalization method is used to process all continuous variables, that is, each feature is subtracted from its sample mean and divided by the standard deviation to convert it into a standard distribution with a mean of 0 and a standard deviation of 1, which is used to improve the robustness of the LightGBM model to feature dimension differences. For possible missing values, local mean interpolation is used to repair them to avoid data filling causing disturbances to the fault sample structure.

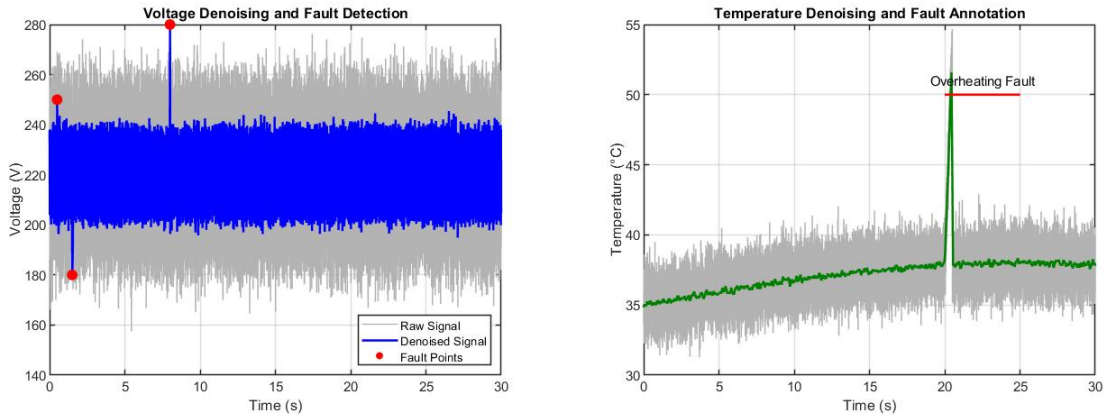

Figure 2. De-noising of voltage and temperature signals and marking of key fault points during the operation of power equipment.
Figure 2 (a) Denoising of voltage and temperature signals during operation of power equipment; Figure 2(b) Key fault point marking.

Figure 2 is a schematic diagram of voltage and temperature signal denoising and key fault point marking during the operation of power equipment. Left figure: Figure 2(a), the horizontal axis of the left sub-figure is time (seconds) and the vertical axis is voltage (volts), showing the fluctuation and superimposed noise of the original voltage signal, as well as the voltage curve smoothed after sliding average filtering. The red solid circles mark three fault points: short-term overvoltage (the voltage rises to 250V in about 0.5 seconds), continuous undervoltage (drops to 180V in about 1.5 seconds) and instantaneous shock (rises to 280V in about 8 seconds), reflecting the diversity and suddenness of equipment voltage anomalies. Figure 2(b) also has time (seconds) on the horizontal axis and temperature (degrees Celsius) on the vertical axis, showing the low-frequency fluctuation and noise of the temperature, as well as the smooth temperature curve. The overheating fault interval is marked with a red line between 20 and 25 seconds, and the temperature rises sharply to more than 50 degrees Celsius, reflecting the risk of thermal runaway in the operation of the equipment.

2) Time Series and Frequency Domain Multi-Scale Feature Extraction Strategy

After data cleaning is completed, it is necessary to construct high-quality features with distinguishing capabilities for the equipment state sequence based on the modeling target. The feature extraction process is designed around the time window, setting the fixed sliding window length to 60 seconds and the step length to 10 seconds. In each window, 24 dimensions of statistical, frequency domain and dynamic change rate features are extracted. First, at the time domain level, six statistical indicators, including mean (Mean), standard deviation (STD), maximum value (Max), minimum value (Min), skewness (Skewness), and kurtosis (Kurtosis), are extracted to characterize the stability and amplitude fluctuation of the equipment's operating status.

In terms of frequency domain feature construction, Fast Fourier Transform (FFT) is used to perform spectrum transformation on temperature and vibration signals, and the energy density ratio, main frequency peak and frequency domain mean of the first five main frequency components are extracted as frequency domain descriptors reflecting the characteristics of mechanical faults. A total of 9-dimensional frequency domain features are extracted [25,26]. This processing is used to capture the hidden patterns of early mechanical faults in the frequency spectrum, which helps to improve the sensitivity of the model to weak anomalies. Assume that the amplitude of the kth frequency component after FFT transformation is X_k, and the frequency domain energy density ratio is calculated as formula 1:

$$P_k = \frac{|X_k|^2}{\sum_{j=1}^{M} |X_j|^2} \quad (1)$$

In formula 1, M is the FFT length, and P_k represents the energy proportion of the k th frequency component.

In addition, the change rate characteristics based on differential operation are introduced to describe the intensity of the signal in a short period, including the

first-order difference mean (ΔMean), the first-order difference variance (ΔVar), the average absolute change amplitude (ΔMAV) between adjacent sampling points and other characteristic indicators. These features can enhance the model's ability to respond to sudden events (such as overheating, breakdown, arc jitter, etc.). The above features are extracted and combined in each window, and finally a feature sample set with a dimension of N×24 (N is the number of sliding windows) is constructed for subsequent model training.

Considering the differences in device types and the impact of environmental factors, in order to avoid dimensional confusion and redundancy between features, a feature selection mechanism based on the minimum redundancy maximum relevance (MRMR) criterion is further introduced. MRMR maximizes the mutual information between features and category labels while minimizing redundancy between features, and is defined as Formula 2:

$$\max_{S}\left[\frac{1}{|S|}\sum_{x_i\in S} I(x_i;c)-\frac{1}{|S|^2}\sum_{x_i,x_j\in S} I(x_i;x_j)\right] \quad (2)$$

In formula 2, S is the selected feature subset, $I(x_i;c)$ is the mutual information between feature x_i and category label c, and $I(x_i;x_j)$ is the mutual information between features.

By calculating the mutual information between each feature and label and the redundancy between features, the top 16 most discriminative feature subsets are selected to optimize the model learning efficiency and avoid overfitting. In actual operation, the feature selection process is carried out within the training set to prevent the feature selection results from leaking into the validation set, ensuring the objectivity and generalization performance of the evaluation.

B. Prediction Model Construction Based on LightGBM

1) Model Input and Output Structure Setting

After completing feature construction and selection, the model training phase uses the feature vector extracted in each sliding time window as input, and the target output is the category identification of whether a fault occurs in the current time window. The label annotation comes from the actual operation and maintenance records and the historical alarm information of the equipment, and is mapped to the corresponding sliding window through time alignment. For fault type annotation, a unified multi-category label encoding mechanism is adopted to assign fixed label numbers to various faults (such as overheating, short circuit, grounding, and voltage abnormality). If the research goal is only a binary classification of fault or not, the samples are marked as "normal" and "abnormal" [27,28]. The constructed dataset is divided into a training set and a validation set, which are divided according to the device running time sequence to ensure that the validation set does not contain data that overlaps with the training set time, avoiding evaluation bias caused by information leakage.

In order to fully explore the nonlinear relationship and combination effect between features, this paper uses LightGBM (Light Gradient Boosting Machine) as the main modeling framework. LightGBM is an efficient implementation based on the Gradient Boosting Decision Tree (GBDT), which supports histogram-based feature partitioning and leaf-first tree structure growth strategy, greatly reducing memory and computing overhead while ensuring learning accuracy. In the model initialization stage, this paper sets boosting_type to "gbdt", the loss function to "binary_logloss" or "multiclass", and automatically adapts the binary or multi-classification output structure according to the target task type.

2) Tree Structure Generation and Feature Splitting Mechanism

During the model training process, the gradient boosting method is used to build weak classifiers round by round, and the overall prediction performance is iteratively improved through the idea of minimizing the residual. In each iteration, LightGBM generates a new tree based on the current residual gradient, and uses a histogram-based splitting method to calculate the split gain of each feature. The mechanism first discretizes the continuous variable into k buckets (k=255), and statistically calculates the gradient and second-order derivative of the samples in the bucket, significantly reducing the complexity and improving the splitting efficiency. In terms of splitting strategy, LightGBM uses a leaf-wise strategy to construct the optimal gain leaf node in each round, and replaces the traditional level-wise structure with a depth-first growth method to improve the fitting ability; To prevent overfitting, set max_depth=8 to limit the tree depth, and set min_data_in_leaf=30 to prevent the generation of leaf nodes with too small sample numbers. For a candidate split point S of a feature, the split gain of the leaf node is defined as Formula 3:

$$Gain(s)=\frac{G_L^2}{H_L+\lambda}+\frac{G_R^2}{H_R+\lambda}-\frac{(G_L+G_R)^2}{H_L+H_R+\lambda}-\gamma \quad (3)$$

G_L, H_L and G_R, H_R are the gradient and second-order gradient sum of the left and right child nodes, respectively, λ is the leaf weight regularization term, and γ is the split penalty parameter.

In the feature selection stage, LightGBM can automatically calculate the usage frequency and information gain of each feature to evaluate its contribution to the model performance. During the training process, feature_fraction=0.8 is used to randomly sample feature subsets to improve the generalization ability of the model and improve the robustness under different data structures of multiple working conditions and multiple devices. To adapt to the

unbalanced sample scenario, the is_unbalance=true parameter is set, and the class_weight strategy is used to explicitly weight the minority class labels to avoid overfitting the model to the main class. The number of iterations is dynamically controlled by early_stopping_rounds=50. If there is no significant performance improvement after 50 consecutive rounds, the training is terminated. The final model uses the structure with the best performance on the validation set for solidification output. Model prediction output calculation The final prediction value of LightGBM is the weighted sum of all trees, which is formula 4:

$$\hat{y}_i = \sum_{t=1}^{T} f_t(\mathrm{x}_i), \quad f_t \in \mathcal{F} \quad (4)$$

T is the total number of trees, f_t is the regression tree function generated in the t round, and x_i is the sample feature vector.

In order to enhance the stability of the model and avoid extreme sample perturbations in the training process, a 5-fold cross validation is used to evaluate the model performance after each iteration within the training set, and the mean is used as the final model indicator reference. After the training is completed, the model output includes the predicted probability and final classification label of each class, and the prediction results can be used for subsequent evaluation and error analysis.

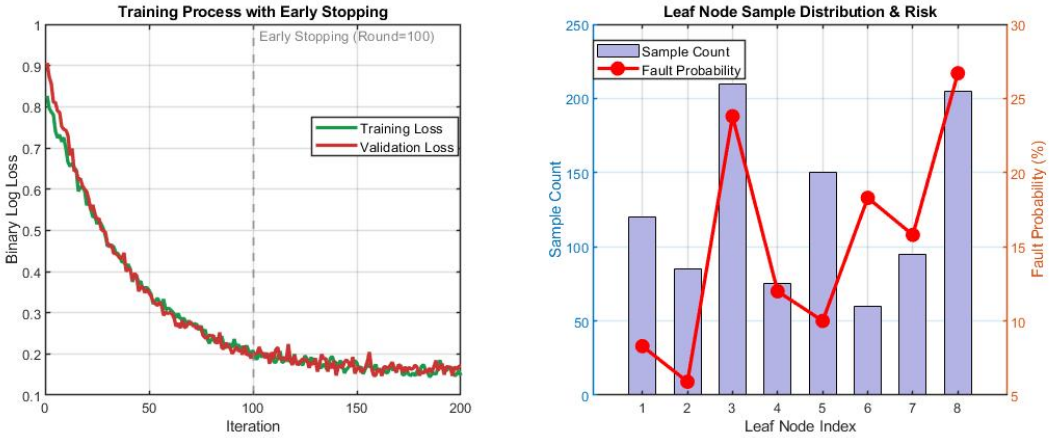

Figure 3. Visualization of LightGBM model training and leaf node risk distribution. Figure 3 (a) LightGBM model training; Figure 3 (b) Visualization of leaf node risk distribution.

Figure 3 (a) shows the trend of the loss function changing with the number of iterations during model training. The horizontal axis is the number of iterations (Iteration), the vertical axis is the binary log loss (Binary Log Loss), the green curve represents the training set loss, and the red curve represents the validation set loss. The training loss stabilizes after a rapid decline in the first 50 rounds, and the validation loss no longer decreases significantly after the 100th round, triggering the Early Stopping mechanism, indicating that the model has achieved optimal generalization performance. Figure 3 (b) on the right shows the leaf node sample distribution and fault probability analysis of the LightGBM model. The horizontal axis is the leaf node index, the left vertical axis is the number of samples for each node (Sample Count), and the right vertical axis is the fault probability of the corresponding node (Fault Probability). It can be seen that the number of samples of nodes 3 and 8 is large, and the failure probabilities are 24.6% and 26.7% respectively, indicating that the model has identified high-risk distribution areas.

In order to further quantify the contribution of features to fault detection, this paper conducts an interpretability study on the LightGBM model through SHAP (Shapley Additive Explanations) analysis. The SHAP value intuitively shows the importance ranking of features by calculating the marginal contribution of each feature to the model output. The experimental results show that the frequency domain energy density ratio of the temperature signal and the absolute mean of the voltage change rate (ΔMAV) have the highest SHAP values for fault classification, indicating that these two types of features play a key role in identifying insulation degradation and partial discharge faults. In addition, the main frequency kurtosis of the vibration signal shows a significant contribution to mechanical faults.

C. Automatic Search of PSO Parameters

1) Definition of Parameter Space and Construction of Search Target

In order to improve the performance of the LightGBM model in the task of multi-condition power equipment fault prediction, its key hyperparameters need to be tuned. This paper introduces the PSO algorithm to automatically search for the hyperparameter combination of LightGBM to avoid performance fluctuations and local optimal problems caused by manually setting parameters [29,30]. First, the key parameters to be optimized include num_leaves, max_depth, learning_rate, feature_fraction, bagging_fraction and lambda_l1. The search space is defined as follows: num_leaves $\in$ [20, 80], max_depth $\in$ [3, 12], learning_rate $\in$ [0.01, 0.3], feature_fraction $\in$ [0.6, 1.0], bagging_fraction $\in$ [0.6, 1.0], lambda_l1 $\in$ [0, 5]. All parameters are mapped to

the real number space for continuous optimization, and the final result is constrained to be a valid integer or floating point format within the range supported by LightGBM [31].

The objective function uses the average F1 score of the LightGBM model in the 5-fold cross validation of the training set as the optimization criterion, which is used to comprehensively evaluate the model's ability to identify the main class and the fault class. In order to maintain computational efficiency and search stability, the F1 score is fixed to evaluate 2,000 training samples in each iteration, and the training and validation structure is uniformly divided into 5:1 to ensure that the target evaluation during the optimization process has time consistency and stability. The global objective function of the particle swarm is to maximize the average F1 value.

2) PSO Process and Convergence Mechanism Settings

In the initialization phase, the particle swarm size is set to 30 and the maximum number of iterations is 50. Each particle represents a parameter combination vector containing 6 dimensions. The initial position is generated by uniformly distributed random sampling, and the corresponding velocity vector is set [32,33]. The global optimal value g_best and the individual optimal value p_best are initialized to empty. The speed update and position update are performed using the standard PSO formula, as shown in formulas 5-6:

$$v_i^{t+1} = w \cdot v_i^t + c_1 \cdot r_1 \cdot (p_i^{best} - x_i^t) + c_2 \cdot r_2 \cdot (g^{best} - x_i^t) \quad (5)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (6)$$

The inertia weight is $w = 0.7$, the individual learning factor is $c_1 = 1.5$, the group learning factor is $c_2 = 1.5$, and $r_1, r_2 \in [0,1]$ is a random number used to increase the search diversity. In order to prevent particles from falling into the boundary or infinite fluctuations, a maximum limit is set for the speed v_i, and rebound processing is performed after the position exceeds the boundary.

After each round of iteration, the F1 score of the current particle is calculated according to the objective function. If it is better than the historical p_best, the current individual optimal position is updated, and g_best is also updated. To prevent convergence from falling into the local optimum, an early stopping mechanism is set to be triggered when g_best has not been significantly improved for 10 consecutive rounds. At the same time, a local perturbation strategy is introduced. When the global optimal value changes by less than 1e-4 in the first 5 rounds, the Gauss random perturbation mechanism is used to fine-tune g_best to jump out of the early stability trap.

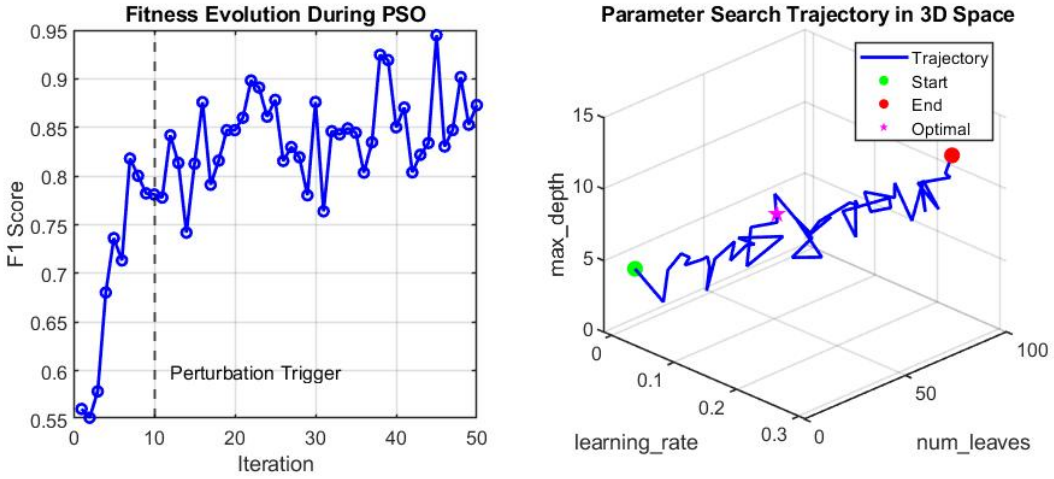

Figure 4. Performance evolution and search path of PSO in the process of parameter tuning of power equipment fault prediction model. Figure 4 (a) PSO performance evolution during parameter setting of power equipment fault prediction model; Figure 4 (b) Search path.

Figure 4 shows the performance evolution and search path of PSO during the parameter adjustment process of the power equipment fault prediction model. The horizontal axis of Figure 4 (a) on the left is the number of iterations, and the vertical axis is the corresponding F1 score, which is used to reflect the fitness change trend of the global optimal solution during the training process. The F1 value grew rapidly in the first 10 iterations, and entered the stable convergence zone after the perturbation strategy was added near the 10th iteration, and eventually tended to be above 0.85, indicating that PSO has strong global search and convergence capabilities. The right figure is Figure 4 (b), which is the search trajectory of the three-dimensional parameter space, showing that the particles start from the initial point distribution and gradually converge to the optimal parameter point. This reflects the search efficiency and convergence path of the PSO algorithm in the multi-dimensional nonlinear hyperparameter space, and provides a visualization basis and parameter adjustment reference for improving the performance of the LightGBM model.

The final output parameter combination is the optimal solution obtained during the search process, which is applied to the final model training stage of LightGBM to ensure that the constructed model has the best structural expression ability and generalization ability in the target task. In order to verify the stability and robustness of the search process, the algorithm was repeated three times under different random seeds, and the average optimal solution was finally taken as the final deployment parameter of the model.

D. Imbalanced Sample Enhancement and Category Calibration

1) Sample Distribution Analysis and Enhancement Strategy

There is a serious imbalance in the distribution of classes in power equipment status monitoring data. The proportion of fault samples is significantly lower than that of normal operating data. This leads to the tendency of traditional machine learning models to overfit the majority class samples during training, and the learning effect of minority class features is poor, which in turn affects the recognition accuracy and recall rate [34]. To address the modeling problem of such unbalanced data, this study introduces the synthetic minority over-sampling technique (SMOTE). This method performs neighborhood analysis on minority class samples in the feature space and uses linear interpolation to generate new samples with topological continuity, effectively avoiding the overfitting risk caused by traditional repeated sampling. In the specific implementation, the neighborhood parameter is set to k=5, which ensures the continuity of the sample feature space while improving the diversity of synthetic samples. The enhancement process first extracts minority class samples from the training set, calculates their k nearest neighbor vectors, and then randomly selects adjacent sample pairs and linearly combines them in the feature dimension to generate new samples with statistical representativeness. This strategy increases the number of fault samples to a level comparable to that of the majority class, making the data distribution balanced and enhancing the model's ability to identify key fault features. This method effectively alleviates the problem of data scarcity while maintaining the topological structure of the original sample set, and is suitable for complex industrial scenarios where multiple fault modes coexist.

Neighboring samples are selected in the feature space and linear interpolation is performed to generate new samples with statistical representativeness. The process first extracts minority class samples from the training set, calculates their k nearest neighbor vectors, and then randomly selects adjacent sample pairs and performs linear combinations on the feature dimension. This strategy increases the number of fault samples to a level comparable to that of the majority class, making the data distribution balanced.

2) Class Weight Adjustment and Model Adaptation

In addition to the balance of sample quantity, this paper introduces a class weight calibration mechanism to further improve the model's ability to identify minority classes. For the LightGBM model, the class weight is set in the training parameters, and the weighting is performed according to the inverse of the class sample ratio, that is, a higher weight is assigned to the minority class, which leads to a greater penalty for misclassification of the minority class during model training. The specific weight calculation formula is Formula 7:

$$w_i = \frac{N}{n_i} \quad (7)$$

N is the total number of samples, n_i is the number of samples of category i, w_i is the weight of the corresponding category. Through this mechanism, the model adjusts the weighted losses of different categories when optimizing the objective function, improving the recall rate and F1 value of the minority class while suppressing the bias dominated by the majority class.

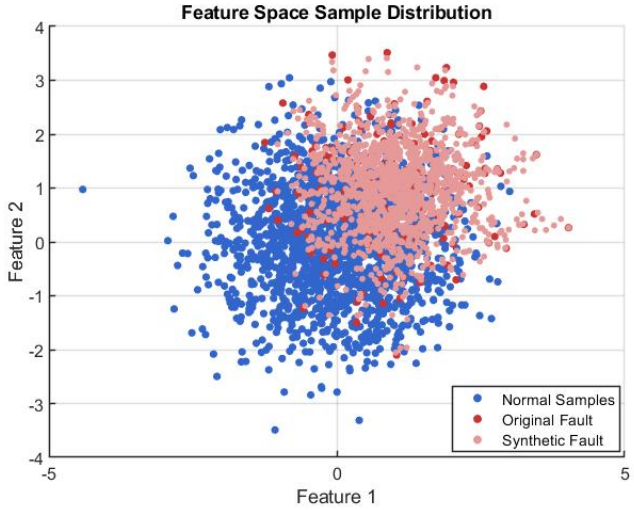

Figure 5. Distribution changes of fault samples in feature space under SMOTE enhancement strategy.

Figure 5 illustrates the distribution evolution of fault samples in the feature space after applying the SMOTE (synthetic minority oversampling technology) enhancement strategy, which aims to solve the sample imbalance problem faced by power equipment fault prediction. In the figure, feature 1 is the horizontal axis and feature 2 is the vertical axis; the blue, red and pink points correspond to the original normal samples, the original fault samples and the synthetically generated fault samples, respectively. Analysis of the distribution diagram shows that the original fault samples are mainly concentrated near the coordinate (1,1) in the feature space. The generated synthetic samples are reasonably extended in their neighborhood, showing a regional distribution feature of continuous density. This enhancement method effectively improves the representation ability of fault categories in the feature space, which can assist subsequent classification models to more accurately identify minority classes, thereby improving the recall rate and enhancing the generalization performance of the model. The overall distribution pattern clearly shows that the newly added synthetic samples have not invaded the normal class sample area, successfully avoiding the risk of category aliasing caused by artificial sample interference. Therefore, this scheme has laid a more solid and reliable data foundation for auxiliary fault diagnosis of power equipment under unbalanced sample conditions.

E. Model Training and Validation Set Division

1) Time-Series-Aware Data Division Strategy

The power equipment fault prediction task involves the continuous change of equipment status over time. The time series characteristics of fault events are obvious, and the data has strong time correlation. To ensure the scientific nature of model training and testing and the authenticity of the results, this paper adopts a training-validation set division strategy based on time series. The specific operation is to sort the samples according to the timestamp in the complete collected time series data, select the first 70% of the time period data as the training set, and the remaining 30% of the data as the validation set. This division method effectively avoids the time overlap between the training set and the validation set caused by random sampling, avoids information leakage, and improves the prediction credibility of the model in a real deployment environment. During the division process, special attention is paid to ensuring that the training set contains representative samples of normal and faulty equipment operation, and the validation set covers a variety of fault types and different operating states to ensure the generalization of the model and the comprehensiveness of the evaluation.

In addition, considering the seasonal fluctuations, operating mode switching and emergencies in the operation of power equipment, this paper verifies the rationality of the statistical characteristics of the training set and the validation set by comparing the load characteristics, temperature changes and fault frequencies in different time periods. This ensures that the training set covers the main operating modes and the validation set can fully test the model's ability to identify new modes and rare faults. This time series division meets the principles of data science and is also in line with the application scenarios of online monitoring of actual power systems.

To avoid the risk of data leakage, the hyperparameter adjustment in this paper strictly follows the following process: (1) A 5-fold cross-validation is performed within the training set, and the F1 score of the PSO search is completely based on the cross-validation results; (2) The entire training set data is used in the final model training phase, but the validation set is always kept independent and is only used for the final performance evaluation. Regarding the application of SMOTE, its enhancement operation is completely limited to the training set, that is, oversampling is performed independently within the training fold of each cross-validation, while the validation fold and the test set retain the original distribution. This strategy ensures the objectivity of model evaluation and avoids overestimation of performance due to test data enhancement.

2) Rationality Guarantee of the Training Process

The model training uses the LightGBM gradient boosting tree algorithm based on the training set, combined with PSO to automatically adjust key hyperparameters, and gradually improve the model's ability to fit complex nonlinear fault characteristics. During the training process, the samples are learned in batch iterations for multiple rounds, and the model parameters are dynamically adjusted to optimize the classification performance. To prevent overfitting, early_stopping_rounds is set to 50, that is, when the performance index of the validation set does not improve significantly in 50 consecutive iterations, the training process is terminated in advance to ensure the balance between the generalization ability of the model and the training efficiency.

During the training, the multi-dimensional indicators such as training error, validation error, F1 score, recall rate and accuracy are recorded in detail, and these indicators are used to monitor the model learning process and adjust the strategy. The training process is verified multiple times through cross-validation to ensure the stability and reliability of the training results. The training process uses multi-threaded parallel computing technology to accelerate model training and hyperparameter search to meet the computing needs of large-scale, multi-dimensional power equipment data. At the same time, the log system is used to warn of abnormalities in the training process, and to adjust the training parameters in time to prevent training failure or slow convergence.

In summary, the time-series-aware training-validation set division and the strict training control mechanism jointly ensure the scientificity and practicality of the model in the complex power equipment fault prediction task, and

effectively improve the model performance and generalization ability.

3. Fault Identification Performance Evaluation

The evaluation experimental data set used in this study comes from the online monitoring platform of a large power system, covering time series data collected by multi-dimensional sensors such as voltage, current, temperature and vibration. The data spans more than two years and contains a variety of typical fault types and normal operation samples, with rich characteristics of operating condition changes. To ensure the scientificity and representativeness of the evaluation, the data has undergone a rigorous preprocessing process, including denoising, outlier removal and normalization, to construct a high-quality feature set. The data set is divided into training sets and test sets according to time series to avoid leakage of time series information and ensure the authenticity of model evaluation. The test set specifically covers a few types of fault samples to verify the model's fault identification ability in an unbalanced environment. The overall data structure meets the requirements of multi-classification and binary classification tasks, providing a solid foundation for the comprehensive evaluation of fault identification performance.

A. Accuracy

The model first generates a predicted category for each sample in the entire test set; then, the number of samples whose predicted category matches the true category is counted; finally, the accuracy is obtained by dividing this value by the total number of samples in the test set. This indicator intuitively depicts the overall correctness of the model's prediction results and has good applicability in scenarios where the distribution of test data categories is relatively balanced. However, there is often a significant sample imbalance in power equipment fault diagnosis scenarios—the number of normal operating samples is much higher than that of fault samples. In this case, the accuracy value is easily dominated by the majority class samples, and it cannot effectively reveal whether the model's recognition ability for the minority class (the fault class) meets the standard. Therefore, relying solely on accuracy is not enough to fully judge the performance of the model. It must be combined with other key evaluation indicators for comprehensive consideration.

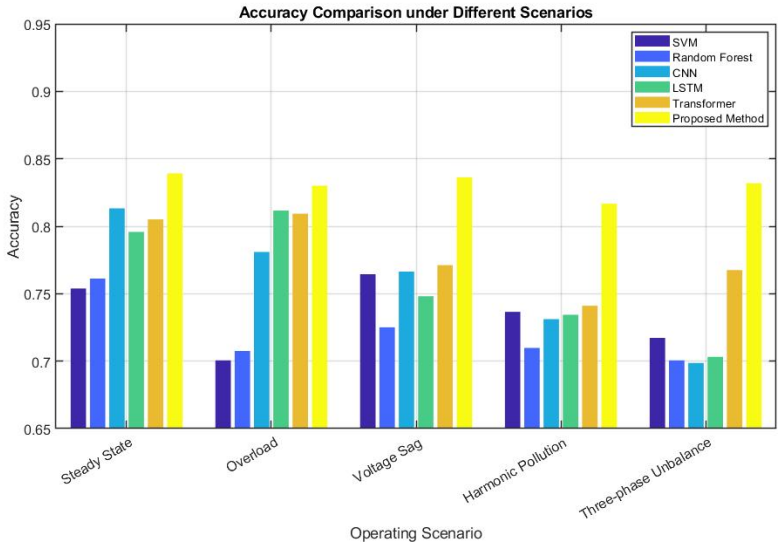

Figure 6. Comparison of the accuracy of six typical machine learning models under five power system operating conditions.

Figure 6 shows the comparison of the accuracy of six typical machine learning models under five power system operating conditions. The horizontal axis represents different operating scenarios, including steady state, overload, voltage sag, harmonic pollution and three-phase imbalance, and the vertical axis represents the prediction accuracy of the model. Figure 6 shows that the proposed method has the highest or nearly the highest accuracy in all scenarios, with the highest accuracy reaching 0.84, and has obvious advantages in 5 scenarios. The Transformer model also performs well in some scenarios, with an accuracy of up to 0.82. In contrast, the accuracy of traditional SVM and random forests slightly decreases in some complex working conditions. These data reflect that the proposed method has strong adaptability to a variety of typical power system fault scenarios, and can better balance the accuracy of fault identification under different operating environments, and has high practical value and robustness.

B. Recall Rate

The recall rate reflects the model's ability to identify and cover fault samples, that is, the proportion of fault samples detected by the model to all actual fault samples. When calculating, it is necessary to count the number of samples that are actually faulty and correctly predicted, and divide it by the number of all real fault samples. The recall rate is of great significance in power equipment fault detection and is directly related to the missed alarm rate. Missed alarms may result in equipment faults not being discovered in time, increasing operational risks. A higher recall rate indicates that the model has a strong

ability to capture potential faults, which helps to achieve early warning and fault prevention. It is a key indicator for evaluating the reliability of fault detection systems.

Table 1. Recall rates of six typical models in fault detection.

Fault Type	SVM	Random Forest	CNN	LSTM	Transformer	Proposed Method
Partial Discharge	0.81	0.83	0.88	0.9	0.92	0.95
Winding Overheat	0.77	0.79	0.85	0.87	0.89	0.93
Insulation Degradation	0.74	0.76	0.83	0.85	0.87	0.91
Core Loosening	0.7	0.72	0.8	0.83	0.84	0.89
Cooling Failure	0.73	0.75	0.81	0.84	0.88	0.92

Table 1 shows the recall rates of six typical models in five types of power equipment fault scenarios. From the data, the recall rate of this method is the highest for all fault types, reflecting its strong fault detection capabilities. For example, in the "partial discharge" fault, the recall rate of this method is 0.95, which is better than Transformer's 0.92 and LSTM (Long Short-Term Memory)'s 0.90, indicating that it can more effectively identify key abnormal signals. In the two types of faults, "winding overheating" and "cooling failure", the proposed method reached 0.93 and 0.92 respectively, which are significantly higher than SVM's 0.77 and 0.73, indicating that it is particularly outstanding in dealing with heat-related faults. Compared with other models, the recall rate is generally less than 0.85 in the core loosening scenario, while the proposed method reaches 0.89, further indicating that it has strong missed detection control capabilities under complex working conditions and effectively supports the early warning and fault response needs of the power system.

C. Mean Detection Delay

The mean detection delay measures the time interval from the actual occurrence of a fault to the first accurate detection of the fault. The calculation steps include recording the fault start time and the first correct alarm time of the model for each fault sample, calculating the difference between the two and taking the average. This indicator reflects the response speed and warning timeliness of the model. The shorter the delay time, the more timely the model can detect faults, which gives maintenance personnel time to repair and adjust the faults and reduces the risk of equipment damage. The average detection delay time is a key reference for evaluating the real-time performance of the model in the intelligent identification of power system faults.

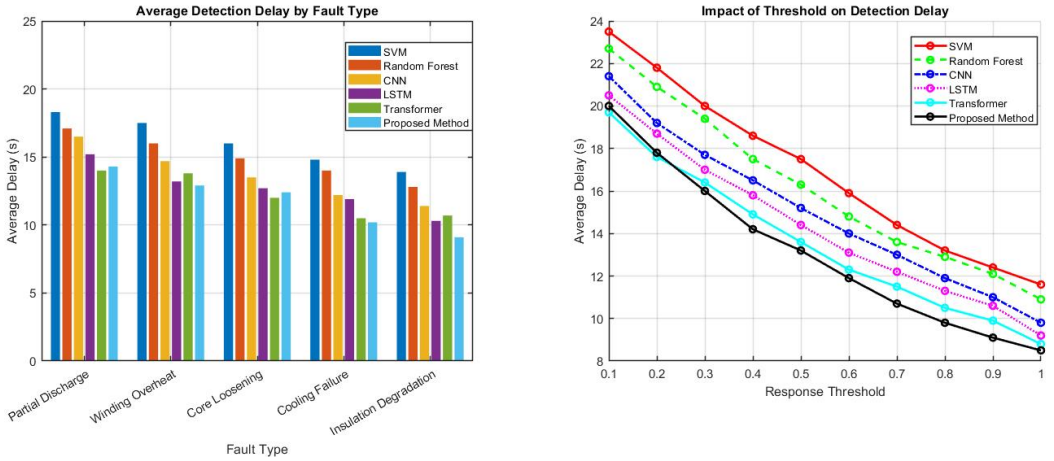

Figure 7. Analysis of detection delay of six models under different fault and threshold conditions. Figure 7 (a) Analysis of detection delay of six models under different faults; Figure 7 (b) Analysis of detection delay of six models under different threshold conditions.

The left figure, Figure 7 (a), shows the average detection delay of six models under five typical faults (such as partial discharge, winding overheating, core loosening, etc.), with the horizontal axis representing the fault type and the vertical axis representing the detection delay (seconds). It can be seen that the proposed method has a lower delay for most fault types. For example, in the "insulation degradation" scenario, the average delay is only 9.1 seconds, which is significantly better than SVM (13.9 seconds) and random forest (12.8 seconds), indicating that the method has a higher real-time response capability. The right figure, Figure 7 (b), reflects the changing trend of the average detection delay of each model under different response thresholds (horizontal axis 0.1~0.9) (vertical axis is seconds). Generally speaking, the higher the threshold, the faster the detection response, but the curve shapes of different models are different. For example, when the threshold is 0.6, the transformer delay is 12.3 seconds, while the SVM delay is 15.9 seconds. The Transformer and method proposed in this paper respond faster while ensuring accuracy, and show better timeliness in various

fault scenarios and different judgment sensitivities.

D. F1 Score

The F1 score is an indicator that comprehensively considers the model's precision and recall. As the harmonic mean of the two, it reflects the model's classification performance in a more balanced way. Precision measures the proportion of true faults among the fault samples identified by the model, while recall measures the proportion of all fault samples that are correctly identified. The F1 score is particularly important when fault samples are scarce and the categories are unbalanced. A single indicator is difficult to fully evaluate the model performance. A high F1 score indicates that the model has achieved a good trade-off between reducing false positives and missing negatives, ensuring both the accuracy of identification and improving fault coverage. It is an important indicator for measuring the quality of fault detection.

Table 2. F1 score of the model in power fault classification.

Fault Type	SVM	Random Forest	CNN	LSTM	Transformer	Proposed Method
Partial Discharge	0.79	0.81	0.86	0.88	0.91	0.94
Winding Overheat	0.76	0.78	0.84	0.86	0.89	0.92
Insulation Degradation	0.72	0.75	0.81	0.84	0.87	0.9
Core Loosening	0.68	0.7	0.78	0.82	0.85	0.88
Cooling Failure	0.71	0.74	0.8	0.83	0.86	0.9

Table 2 comprehensively evaluates the accuracy and coverage of the model in power fault classification. From the overall data, the F1 score of the proposed method under various faults is generally leading, and has a more balanced classification performance. In the key fault type of partial discharge, the F1 value of the proposed method is 0.94, which is higher than Transformer (0.91) and LSTM (0.88), indicating that it not only has comprehensive recognition, but also takes into account false alarm control. In the cases of winding overheating and insulation aging, the proposed method obtained F1 scores of 0.92 and 0.90 respectively, which are better than the traditional method SVM (0.76 and 0.72), reflecting that it is more accurate in dealing with faults with strong time variability and unstable signals. In the category of loose core, the F1 scores of traditional models such as CNN (Convolutional Neural Network) and random forest are 0.78 and 0.70, while the proposed method is 0.88, showing stronger adaptability and model generalization ability. These data verify that the proposed method has high classification accuracy and reliability under multiple abnormal conditions, and is suitable for high-risk fault identification tasks of actual power equipment.

E. Area under the Receiver Operating Characteristic Curve (AUC)

AUC (Area Under the Curve) evaluates the model's ability to distinguish between faulty and normal samples at different judgment thresholds. This paper selects six typical classification models and draws the ROC curve for the power equipment fault identification task.

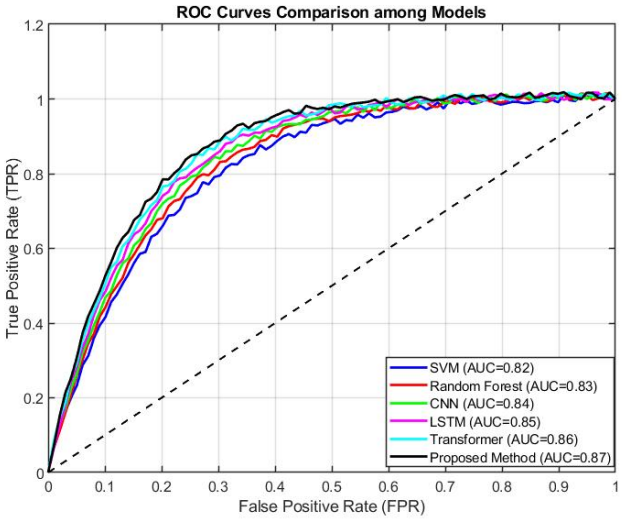

Figure 8. ROC curves of six typical classification models in the task of power equipment fault identification.

Figure 8 shows the receiver operating characteristic curves (ROC) of six typical classification models in the task of power equipment fault identification. The horizontal axis represents the false positive rate (FPR), that is, the proportion of normal states misjudged as faults, and the vertical axis represents the true positive rate (TPR), that is, the proportion of real faults successfully identified by the model. The closer the curve is to the upper left corner, the better the model performance is. The larger the AUC (area under the

curve) value is, the stronger the overall discrimination ability of the model is. The results show that the curve of the proposed method is closest to the upper left corner, and the AUC reaches 0.87, which is significantly better than other models, indicating that it has the strongest ability to distinguish between faults and normal states under different judgment thresholds. Transformer and LSTM follow closely behind, with AUCs of 0.86 and 0.85 respectively, showing good learning ability and generalization. The AUC of traditional machine learning models such as SVM and Random Forest is only 0.82 and 0.83, and their recognition ability under complex working conditions is relatively limited.

F. Ablation Experiments and Small-Scale Deployment Case Studies

In order to evaluate the independent contributions of SMOTE, MRMR and PSO to model performance, this paper designed an ablation experiment. The experiments are divided into three groups: (1) After removing SMOTE, the recall rate of the model in the insulation performance degradation fault decreased by 12%, indicating that SMOTE plays a key role in alleviating sample imbalance; (2) When random feature selection is used instead of MRMR, the F1 score is reduced by 7.3%, verifying the advantages of MRMR in retaining key features and suppressing redundant information; (3) When the LightGBM parameters are fixed (without PSO optimization), the average detection delay of the model in 5 fault types increases to 15.4 seconds, proving that PSO optimization significantly improves the parameter search efficiency. The comparative results of the three groups of experiments show that SMOTE mainly improves the recognition ability of minority classes, MRMR optimizes the feature space structure, and PSO enhances the generalization performance of the model. The synergy of the three is the core of the model to achieve optimal performance.

In order to verify the real-time performance of the model on edge devices, this paper designed a small-scale deployment experiment based on Raspberry Pi 4B (4GB RAM). In the experiment, the optimized LightGBM model was quantized into a 16-bit integer model and deployed on the edge device for real-time reasoning. Test data show that within 72 hours of continuous operation, the average single reasoning time of the model was 12.3ms (including data preprocessing), and the memory usage was stable below 320MB, meeting the real-time requirements of online monitoring of power equipment. In addition, in the simulated sudden fault scenario, the model completed fault identification and triggered an alarm within 3.2 seconds, verifying its feasibility in an environment with limited hardware resources. This case shows that the proposed method has practical engineering value in lightweight deployment and low-latency response.

G. Statistical Verification of Robustness

To verify the robustness of the model to random seeds and fault types, this paper conducted three sets of statistical experiments: (1) Fixed hyperparameters, used 5 different random seeds to train the model, and calculated the standard deviation of key indicators (F1, Recall). The results showed that the standard deviation of F1 was 0.008 and the standard deviation of Recall was 0.012, indicating that the model was less affected by initialization; (2) Randomly masked 20% of the samples in the 5 fault types, repeated training 5 times, and the F1 score fluctuated in the range of [0.92, 0.95], proving the robustness to data perturbations; (3) Five independent PSO optimizations were performed on the insulation performance degradation fault, and the standard deviation of the hyperparameter convergence value was num_leaves=2.1, learning_rate=0.005, verifying the stability of parameter search. The above experiments show that the proposed method is highly consistent under different random initializations and fault distributions.

4. Conclusions

A power equipment fault prediction and diagnosis model was constructed based on LightGBM and PSO algorithms. In view of the limitations of traditional methods in dealing with high-dimensional nonlinear features and sample imbalance, a systematic data preprocessing, feature extraction and enhancement strategy was designed. Through sliding window denoising, time series and frequency domain multi-scale feature construction, combined with SMOTE sample enhancement and category weight calibration, the model's recognition ability for a few fault categories was improved. The PSO algorithm realizes the automatic tuning of LightGBM hyperparameters, significantly enhancing the generalization performance and stability of the model. The innovation of this study is reflected in three aspects: (1) PSO is combined with LightGBM for the first time to solve the problem of dynamic optimization of hyperparameters in power equipment fault prediction; (2) Through the dual mechanism of SMOTE and category weight, the problem of fault sample imbalance is systematically alleviated; (3) A time series-aware data partitioning strategy is proposed, which significantly reduces the risk of information leakage. Experimental verification shows that the model performs well in key indicators such as accuracy, recall rate, and detection delay time, and has strong real-time fault recognition capabilities. The method still has limitations: (1) Feature engineering relies on domain knowledge, and self-supervised learning can be explored in the future to reduce manual design; (2) Model complexity may affect generalization ability under extreme multi-working conditions, and a meta-learning framework needs to be introduced. Although the method in this paper shows significant advantages in comparison with CNN, LSTM and Transformer, future research can be further extended to other time series-specific models. The causal convolution structure of TCN may be more suitable for capturing the temporal dependency of device states, while the sparse attention mechanism of Informer may be more efficient in processing long sequence data. Such comparative experiments will help verify the generalization ability of the method in this paper in a wider model spectrum and provide a more

comprehensive technical selection reference for power equipment fault prediction.

Consent to Publish

The manuscript has not been published before, and it is not being reviewed by any other journal. The authors have all approved the content of the paper.

Funding

This work was supported by the Key Projects of Natural Science Research for Universities of Anhui Province under Grant 2024AH052010.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author, upon request.

Conflicts of Interest

The authors affirm that they do not have any financial conflicts of interest.

References

[1] M.S. Qureshi, S. Umar, M.U. Nawaz. Machine learning for predictive maintenance in solar farms. International Journal of Advanced Engineering Technologies and Innovations, 2024, 1(3), 27-49.

[2] X.R. Zhang, K.P. Rane, I. Kakaravada, M. Shabaz. Research on vibration monitoring and fault diagnosis of rotating machinery based on internet of things technology. Nonlinear Engineering, 2021, 10(1), 245-254. DOI: 10.1515/nleng-2021-0019

[3] F.Q. Meng, S.S. Yang, J.D. Wang, L. Xia, H. Liu. Creating knowledge graph of electric power equipment faults based on BERT–BiLSTM–CRF model. Journal of Electrical Engineering & Technology, 2022, 17(4), 2507-2516. DOI: 10.1007/s42835-022-01032-3

[4] M. Fernandes, J.M. Corchado, G. Marreiros. Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review. Applied Intelligence, 2022, 52(12), 14246-14280. DOI: 10.1007/s10489-022-03344-3

[5] V. Singh, P. Gangsar, R. Porwal, A. Atulkar. Artificial intelligence application in fault diagnostics of rotating industrial machines: A state-of-the-art review. Journal of Intelligent Manufacturing, 2023, 34(3), 931-960. DOI: 10.1007/s10845-021-01861-5

[6] B.A. Tama, M. Vania, S. Lee, S. Lim. Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals. Artificial Intelligence Review, 2023, 56(5), 4667-4709. DOI: 10.1007/s10462-022-10293-3

[7] D. Rajalakshmi, K. Sudharson, A.S. Kumar, R Vanitha. Advancing Fault Detection Efficiency in Wireless Power Transmission with Light GBM for Real-Time Detection Enhancement. International Research Journal of Multidisciplinary Technovation, 2024, 6(4), 54-68. DOI: 10.54392/irjmt2445

[8] D. Bagci Das. Real-time adaptable fault analysis of rotating machines based on Marine predator algorithm optimised LightGBM approach. Nondestructive Testing and Evaluation, 2025, 40(3), 831-866. DOI: 10.1080/10589759.2024.2328750

[9] S. Demir, E.K. Sahin. Predicting occurrence of liquefaction-induced lateral spreading using gradient boosting algorithms integrated with particle swarm optimization: PSO-XGBoost, PSO-LightGBM, and PSO-CatBoost. Acta Geotechnica, 2023, 18(6), 3403-3419. DOI: 10.1080/10589759.2024.2328750

[10] P.C. Yan, F.X. Chen, T.J. Zhao, H. Zhang. Transformer fault diagnosis research based on LIF technology and IAO optimization of LightGBM. Analytical Methods, 2023, 15(3), 261-274. DOI: 10.1039/D2AY01745H

[11] C.Y. Lee, E.D.C. Maceren. Wind energy system fault classification and detection using deep convolutional neural network and particle swarm optimization-extreme gradient boosting. IET Energy Systems Integration, 2024, 6(4), 479-497. DOI: 10.1049/esi2.12144

[12] S. Dalal, U.K. Lilhore, B. Seth, M. Radulescu, S. Hamrioui. A Hybrid Model for Short-Term Energy Load Prediction Based on Transfer Learning with LightGBM for Smart Grids in Smart Energy Systems. Journal of Urban Technology, 2025, 32(1), 49-75. DOI: 10.1080/10630732.2024.2380639

[13] X.B. Lv, F.Z. Liu, M.S. Jiang, F.Y. Zhang, L. Jia. Fault diagnosis of power transformers based on dissolved gas analysis and improved LightGBM hybrid integrated model with dual-branch structure. IET Electric Power Applications, 2024, 18(12), 2008-2020. DOI: 10.1049/elp2.12528

[14] Q.C. Fang, B. Shen, J.K. Xue. A new elite opposite sparrow search algorithm-based optimized LightGBM approach for fault diagnosis. Journal of Ambient Intelligence and Humanized Computing, 2023, 14(8), 10473-10491. DOI: 10.1007/s12652-022-03703-5

[15] M. Abubakar, Y.B. Che, A. Zafar, M.A. Al-Khasawneh, M.S. Bhutta. Optimization of solar and wind power plants production through a parallel fusion approach with modified hybrid machine and deep learning models. Intelligent Data Analysis, 2025, 29(3), 808-830. DOI: 10.1177/1088467X241312592

[16] A.M. Akbar, R. Herteno, S.W. Saputro, M. Faisa, R. Nugroho. Optimizing software defect prediction models: integrating hybrid grey wolf and particle swarm optimization for enhanced feature selection with popular gradient boosting algorithm. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 2024, 6(2), 169-181. DOI: 10.35882/jeeemi.v6i2.388

[17] C.Y. Lee, E.D.C. Maceren. Induction motor bearing fault classification using deep neural network with particle swarm optimization-extreme gradient boosting. IET Electric Power Applications, 2024, 18(3), 297-311. DOI: 10.1049/elp2.12389

[18] J.R. Gao, Z.Q. Wang, Z.Y. Lei, R.L. Wang, Z.W. Wu, S.C. Gao. Feature selection with clustering probabilistic particle swarm optimization. International Journal of Machine Learning and Cybernetics, 2024, 15(9), 3599-3617. DOI: 10.1007/s13042-024-02111-9

[19] T.G.T Tran, T.N. Tran. Enhancing Regression Accuracy and Reliability with Ensemble Learning and Iterative Hyperparameter Optimization. Engineering, Technology & Applied Science Research, 2025, 15(4), 24541-24548. DOI: 10.48084/etasr.11266

[20] S. Kumar, T. Gopi, N. Harikeerthana, M.K. Gupta, V. Gaur, G.M. Krolczyk, et al. Machine learning techniques in additive manufacturing: a state of the art review on design, processes and production control. Journal of Intelligent Manufacturing, 2023, 34(1), 21-55. DOI: 10.1007/s10845-022-02029-5

[21] Y.D. Hu, R.N. Liu, X.L. Li, D.Y. Chen, Q.H. Hu. Task-sequencing meta learning for intelligent few-shot

fault diagnosis with limited data. IEEE Transactions on Industrial Informatics, 2021, 18(6), 3894-3904. DOI: 10.1109/TII.2021.3112504

[22] X.Y. Yang. Power grid fault prediction method based on feature selection and classification algorithm. International Journal of Electronics Engineering and Applications, 2021, 9(2), 34-44. DOI 10.30696/IJEEA.IX.I.2021.34-44.

[23] X.F. Zhang, Z. Long, J. Peng, G.P. Wu, H.F. Hu, M.C. Lyu. Fault prediction for electromechanical equipment based on spatial-temporal graph information. IEEE Transactions on Industrial Informatics, 2022, 19(2), 1413-1424. DOI: 10.1109/TII.2022.3176891

[24] M. Atrigna, A. Buonanno, R. Carli, G. Cavone, P. Scarabaggio, M. Valenti, et al. A machine learning approach to fault prediction of power distribution grids under heatwaves. IEEE Transactions on Industry Applications, 2023, 59(4), 4835-4845. DOI: 10.1109/TIA.2023.3262230

[25] H.Y. Dui, X.H. Dong, L.W. Chen, Y.J. Wang. IoT-enabled fault prediction and maintenance for smart charging piles. IEEE Internet of Things Journal, 2023, 10(23), 21061-21075. DOI: 10.1109/JIOT.2023.3285206

[26] Q.Y. Li, Y.X. Deng, X. Liu, W. Sun, W.T. Li, J. Li, et al. Autonomous smart grid fault detection. IEEE Communications Standards Magazine, 2023, 7(2), 40-47. DOI: 10.1109/MCOMSTD.0001.2200019

[27] Y.M. Huang, J.H. Luo, Z.G. Ma, B. Tang, K.Q. Zhang, J.Y. Zhang, et al. On combined PSO-SVM models in fault prediction of relay protection equipment. Circuits, Systems, and Signal Processing, 2023, 42(2), 875-891. DOI: 10.1007/s00034-022-02056-w

[28] J. Fang, H.B. Wang, F. Yang, K. Yin, X. Lin, M. Zhang. A failure prediction method of power distribution network based on PSO and XGBoost. Australian Journal of Electrical and Electronics Engineering, 2022, 19(4), 371-378. DOI: 10.1080/1448837X.2022.2072447

[29] H. Zhang, X.Q. Guo, P.J. Zhang. Improved PSO-SVM-based fault diagnosis algorithm for wind power converter. IEEE Transactions on Industry Applications, 2023, 60(2), 3492-3501. DOI: 10.1109/TIA.2023.3341059

[30] Y. He, L.Z. Ye, X.J. Zhu, Z.Y. Wang. Feature extraction based on PSO-FC optimizing KPCA and wear fault identification of planetary gear. Journal of Mechanical Science and Technology, 2021, 35(6), 2347-2357. DOI: 10.1007/s12206-021-0507-2

[31] J. Nayak, H. Swapnarekha, B. Naik, G. Dhiman, S. Vimal. 25 years of particle swarm optimization: Flourishing voyage of two decades. Archives of Comput ational Methods in Engineering, 2023, 30(3), 1663-1725. DOI: 10.1007/s11831-022-09849-x

[32] Q.L. Yang, J.H. Duan, H. Bian, B.Y. Zhang. Equivalent inertia prediction for power systems with virtual inertia based on PSO-SVM. Electrical Engineering, 2025, 107(3), 2997-3010. DOI: 10.1007/s00202-024-02676-2

[33] T. Zan, Z.H. Liu, H. Wang, M. Wang, X.S. Gao, Z.L. Pang. Prediction of performance deterioration of rolling bearing based on JADE and PSO-SVM. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2021, 235(9), 1684-1697. DOI: 10.1177/0954406220951209

[34] Z.L. Li, W.M. Zhu, B. Zhu, B.D. Wang, Q.H. Wang. Thermal error modeling of electric spindle based on particle swarm optimization-SVM neural network. The International Journal of Advanced Manufacturing Technology, 2022, 121(11), 7215-7227. DOI: 10.1007/s00170-022-09827-4