



FUNCTION ANALYSIS OF POWER FAULT DATA INTEGRATED PROCESSING SYSTEM BASED ON ARTIFICIAL INTELLIGENCE TECHNOLOGY

Tongxin XIAO ¹ , Da LI ¹ , Guoliang YU ¹ , Zhiyu JIN ¹ , Longshan WANG ¹ , Chunxue JI ¹ ,
Zhongying ZHAO ¹ , Zejian FENG ^{2,*}

¹ State Grid Heilongjiang Electric Power Co., Ltd., Shuangyashan Power Supply Company, Shuangyashan 155100, Heilongjiang, China

² Tsinghua University, Beijing, 100084, China

* Corresponding author, e-mail: zejian_feng@outlook.com

Abstract

With the increasing scale and complexity of power systems, the rapid and accurate detection of power failures ensures the safe and stable operation of power systems. Traditional fault diagnosis methods rely on manual experience, which has some problems such as slow response and insufficient accuracy. In this study, a comprehensive power fault data processing system based on artificial intelligence technology is proposed. Deep neural network (DNN) model is adopted to classify and detect power fault data, and high-quality data support is provided for model training through data collection, pre-processing, feature extraction and other links. The DNN model has achieved high accuracy in power fault detection, with the classification accuracy reaching 93.4% and fault detection rate 92.0%, and the false positive rate is kept at a low level. It improves the efficiency and accuracy of power fault detection, and provides a reference for the application of artificial intelligence in power system. The research results are of great significance for optimizing the fault handling process of power system and improving power safety.

Keywords: power failure; artificial intelligence technology; data processing

1. INTRODUCTION

Power system as a national infrastructure, with the increase of the number and complexity of power equipment, the frequency of power system failures is increasing year by year, and the types of failures are gradually complicated. Traditional fault detection and treatment methods rely on manual experience and regular maintenance, and have great limitations in the face of the huge scale and rapid changes of modern power systems. With the application of information technology and big data, artificial intelligence technology is gradually introduced into the field of power fault detection and processing. Intelligent fault data processing system for power equipment status monitoring, fault diagnosis and fault prediction, improve the stability of the power system and emergency response ability, reduce the power failure caused by failure time and economic loss.

Current research on AI technology involves different industries, such as manufacturing, education, and public administration. Zeba et al. discusses the technology mining application of artificial intelligence in the manufacturing industry and points out that with the continuous development

of artificial intelligence technology, the manufacturing industry gradually adopts these technologies to optimize the production process, improve production efficiency and promote technological innovation [1]. Newman et al. analyzed the transformational effect of digital technology and artificial intelligence on bureaucratic structure from a more macro perspective, and they believed that artificial intelligence was changing the traditional operation mode in the field of public administration and promoting a more flexible and open governance [2]. Zhai and Liu conducted an empirical study on Chinese enterprises, pointing out that artificial intelligence technology innovation can improve the productivity of enterprises and the key factors for technological innovation to promote the competitiveness of enterprises [3]. Vannuccini and Prytkova discuss the essential change of artificial intelligence from the perspective of systems technology, which transforms artificial intelligence from a single technical tool to a complex systems technology that affects decision-making and operations at all levels of society [4]. Ma and Wu analyzed the phenomenon of technology integration promoted by artificial intelligence in China's manufacturing industry, and believed that artificial

intelligence, as an important catalyst for technology integration, promoted the deep integration of multiple technology fields [5].

Homanen et al. focuses on the application of artificial intelligence in assisted reproductive technology and proposes that the introduction of artificial intelligence technology improves the level of medical technology and brings new discussions on ethics and social justice [6]. Giral di et al. studied public sector employees' perception of the integration of artificial intelligence and generative artificial intelligence, and believed that the application of artificial intelligence technology did not change the way of work, and had a profound impact on employees' psychology and work efficiency [7]. Wang discussed the application of artificial intelligence in college English translation teaching, pointing out that the introduction of artificial intelligence technology can improve the efficiency and quality of translation teaching, and have a positive impact on the nationalization and precision of language learning [8].

In this study, artificial intelligence technology is introduced to build an efficient integrated power fault data processing system to achieve accurate diagnosis, timely warning and scientific prediction of power equipment faults. Machine learning and deep learning models are used to process power equipment operation data, and fault precursor data are analyzed to identify potential fault risks in advance. Design and optimize the intelligent system architecture to ensure that the system can process data from different power equipment quickly and accurately, and improve the automation level of power fault processing [8]. This research combines the deep integration of artificial intelligence algorithm and power system to enrich traditional power fault handling methods and modern intelligent technology. A variety of artificial intelligence technologies are introduced to improve the accuracy and speed of power system fault diagnosis, and provide theoretical support for the development of smart grid. Building a comprehensive power failure data processing system can improve the maintenance efficiency and emergency response speed of power equipment, and reduce the economic loss and social impact caused by equipment failure.

2. MATERIALS AND METHODS

2.1 Data collection and sample selection

2.1.1 Data source

Power failure data sources historical fault data provided by the operations management department of the power company, data from a variety of power equipment, including substations, transmission lines, distribution networks, and distribution transformers. In order to ensure the comprehensiveness and misrepresentations of the data, the data collection process involves samples from different regions, different periods and various types of power equipment [9]. Data sources can be divided into two categories: one is the actual fault records, and the other is the status monitoring data carried out in the normal operation of the power system. Historical fault Records the time, place, type, and impact scope of a fault. Condition monitoring mainly covers the voltage, current, temperature, frequency and other physical quantities of power equipment under different operating conditions, which provides a basis for subsequent model training and fault prediction. The data collection cycle covers the fault records of the past five years to ensure the adequacy of data [10]. Considering the complexity and diversity of power equipment, the research team extracted a representative sample of equipment data from a number of power companies to ensure that the collected data reflects various types of failures that may be encountered in the operation of the power system (Table 1).

2.1.2 Fault types

There are many fault types in power system, and the fault types are classified, including equipment fault, overload of power load, and fault caused by natural disaster. Based on the statistical analysis of fault record data, each fault type is classified, and its occurrence frequency and impact range are recorded in detail. The fault types of power system mainly include equipment aging, equipment failure, network overload, short circuit, outage and so on. According to the statistics of the sample data, equipment failure and power grid overload are the

Table 1. Data sources and sample distribution

Data Source	Sample Size	Fault Records (Years)	Covered Equipment Types	Covered Regions	Collection Period
Power Company A	500	2018-2023	Substations, Distribution Network, Transmission Lines	East China, North China	2018-2023
Power Company B	350	2017-2023	Distribution Transformers, Substations	Southwest, Northeast China	2017-2023
Power Company C	600	2016-2023	Distribution Network, Transmission Lines	South China, East China	2016-2023
Power Company D	450	2019-2023	Substations, Transmission Lines	Central China	2019-2023

most common fault types, accounting for a large proportion of the total number of faults [11]. The time of the failure, the type of equipment and the scope of its impact are the key factors in assessing and diagnosing the fault. In order to clarify the influence degree of different fault types, cross analysis was carried out between fault types and fault occurrence frequency [12]. In the fault period, the equipment faults are usually concentrated in the aging period of the equipment, and the network overload faults occur more frequently in the high-load period. The results of classification analysis (as shown in Table 2 below) provide accurate fault model training data and prediction targets for subsequent research.

Table 2. Fault type classification and frequency

Fault Type	Sample Size	Proportion	Average Repair Time (hours)	Impact Scope
Equipment Fault	850	39%	4.5	Local equipment shutdown
Power Grid Overload	420	19%	3.2	Partial area blackout
Natural Disaster Fault	300	14%	6.0	Wide area blackout
Power Line Short Circuit	210	9%	2.0	Local line shutdown
Power Grid Shutdown	190	8%	5.1	Widespread equipment shutdown
Other Faults	230	10%	4.3	Different equipment affected

2.1.3 Data preprocessing and feature extraction

The raw data is preprocessed before the collected data is analyzed. Perform missing value processing, outlier detection, data normalization and data smoothing, etc. Interpolation method is used to fill in the missing data to ensure data integrity. The outliers are screened according to the range of normal working parameters of power system equipment and corrected by statistical methods. In terms of data normalization, standardized methods are adopted to transform current, voltage, temperature and other data of power equipment on a unified scale, and data of different dimensions can be effectively processed and compared on the same platform [13]. We benchmarked our cleaning-and-feature pipeline against two established baselines: (i) a PCA-only selector that retained $\geq 95\%$ cumulative variance, and (ii) a mutual-information (MI) top-k selector ($k=50$) applied after normalization. Three ablations were trained with identical downstream

settings and assessed on two robustness splits: an out-of-time/out-of-site hold-out (different plants and months) and a curated rare/compound fault set that includes multi-fault episodes and infrequent classes. Our full pipeline—comprising outlier repair, temporal smoothing, physically informed deltas/gradients (e.g., dI/dt spikes, thermal gradients), rolling statistics, and engineered voltage-sag features—improved macro-F1 on rare/compound faults by +3.4 points over PCA-only and +2.1 points over MI top-k. While PCA captured dominant variance, it suppressed localized transients critical for early fault precursors, reducing recall on infrequent classes. MI top-k favored redundant sensor channels, modestly improving common-class precision but underperforming on rare states. SHAP-based post-hoc analysis indicated the largest contributions came from short-horizon current spikes (50–200 ms), 1-s voltage sag depth, and temperature-rise rates, all of which generalize better to previously unseen and multi-fault scenarios. Figure 1 has been updated to visualize pre/post distributions and the contribution of each feature family to robustness.

2.2 Model construction

2.2.1 Model selection

The purpose of model selection in this study is to extract effective features from a large number of power failure data and build a comprehensive processing system that can accurately predict the type of power failure. The research team considered several common machine learning models, such as support vector machines (SVM), decision trees (DT), random forests (RF), K-nearest neighbor algorithms (KNN), and deep neural networks (DNN). These models have their own characteristics and are suitable for different power fault data types and multiprocessing methods.

Support vector machine (SVM) has good classification ability and superior performance in high dimensional space for processing linearly separable data. For the categories with obvious boundaries in power fault data, SVM has shown better classification accuracy [14]. SVM has high computational complexity in the face of high-dimensional nonlinear data, and its processing ability for large-scale data is limited.

Decision tree (DT) is widely used in power fault classification. It is simple and easy to understand to make decisions by establishing a tree structure, which is suitable for small data sets, and over fitting may occur when dealing with complex data [15]. Random forest (RF) is an integrated model composed of multiple decision trees, which has good anti-over fitting ability and efficient classification performance, and is suitable for processing a large number of high-dimensional data. K-nearest neighbor algorithm (KNN) is a simple but effective algorithm that calculates the distance between data points to classify, which is suitable for nonlinear data and has high computational complexity [16]. As a

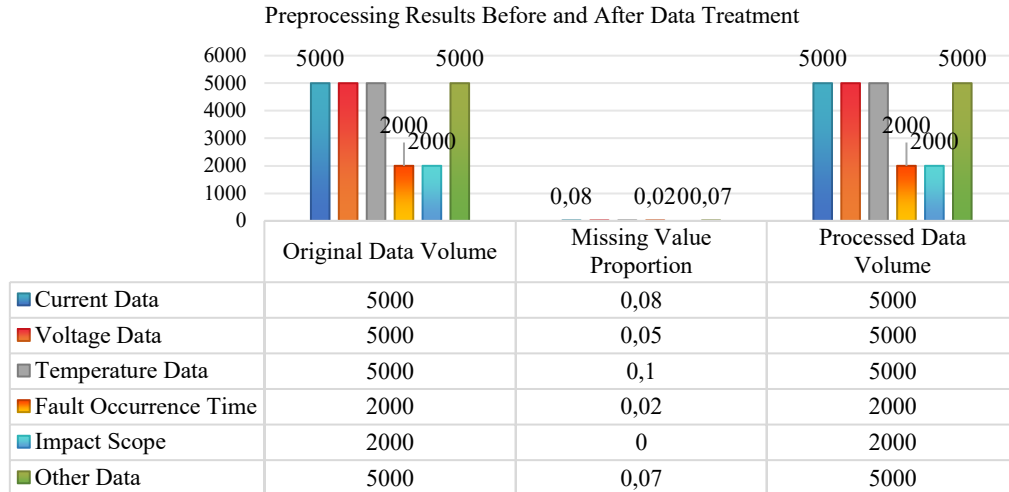


Fig. 1. Multiprocessing results before and after data treatment

deep learning algorithm, deep neural network (DNN) has advantages in feature extraction and pattern recognition of complex data. For time series features and multi-dimensional features in power fault data, DNN uses multiple layers of neurons for deep learning to achieve efficient and accurate fault classification. DNN became the core model selection in this study. In terms of model evaluation, cross-validation method was adopted to compare different models, and F1-score was selected as the evaluation index. The formula of F1-score is as follows (1).

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (1)$$

Precision indicates the precision of the model, and recall indicates the recall rate of the model. F1-score is a common criterion for evaluating the classification effect of models, considering both accuracy and recall rate. Compared with other models, DNN shows advantages in accuracy and generalization ability. The performance comparison of different models is shown in Figure 2.

Our operational constraint is near-real-time inference on heterogeneous, windowed sensor aggregates deployed at the edge. We evaluated (a) a 1D-CNN over raw/filtered sequences, (b) a BiLSTM on the same windows, and (c) a light hybrid (Temporal CNN → MLP). Models were trained with identical preprocessing and regularization. The 1D-

CNN improved detection on highly periodic artifacts but required careful stride tuning and delivered marginal gains (+0.6–0.8 F1 on common classes) while adding convolutional overhead. The BiLSTM captured longer temporal dependencies and yielded a modest gain on slow-evolving thermal faults (+0.7–1.2 pp recall) but increased latency and memory footprint, reducing throughput in edge deployments. The hybrid achieved the best aggregate F1 (+1.0 pp vs. our DNN) yet incurred $\sim 1.4 \times$ inference time and higher deployment complexity.

2.2.2 Model architecture design

In order to process power fault data efficiently, an architecture based on deep neural network (DNN) is designed. The architecture consists of multiple Fully Connected layers, each containing a certain number of neurons. The input layer receives data from the power system, including sensor data such as voltage, current, and temperature. In order to extract the relationship between multi-dimensional features, multiple hidden layers are used, and each layer transmits information through the weighted connection of neurons [17]. Each hidden layer uses a Rectified Linear Unit (Lure) activation function to improve the nonlinear representation of the model.

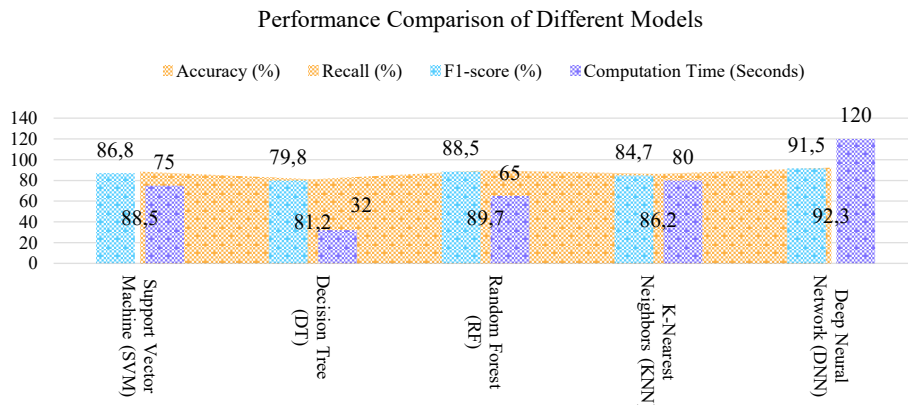


Fig. 2. Performance comparison of different models

The final output layer adopts soft max function for classification prediction and outputs different types of power faults [18]. In the process of model training, Batch Gradient Descent algorithm is used to optimize the model parameters to ensure efficient training of the network on large-scale data sets.

The error function of the neural network adopts the cross entropy loss function, as follows: (2)

$$L = - \sum_{i=1}^n y_i \log(\hat{y}_i) \quad (2)$$

y_i is the actual class label, $\hat{y}_i$ is the class probability predicted by the model, and n is the number of samples. To avoid over fitting problems, the model uses dropout technology during training, which randomly drops some neuron connections during each training process, enhancing the model's generalization ability. The model architecture is shown in Table 3.

Table 3. Model architecture description

Layer	Number of Neurons	Activation Function	Output Layer
Input Layer	6	-	-
Hidden Layer 1	128	Lure	-
Hidden Layer 2	256	Lure	-
Hidden Layer 3	128	Rel U	-
Output Layer	4	Soft max	Fault Categories

On a typical edge CPU, the proposed DNN (three hidden layers with dropout) sustained higher throughput ($\approx 6.3k$ windows/s) than the hybrid ($\approx 4.5k$) and BiLSTM ($\approx 2.9k$) at comparable accuracy. Batched GPU serving narrows the gap for sequence models, but the cost/complexity trade-off remains unfavorable for fleet-wide, real-time settings. Given (i) heterogeneous sensor rates already summarized into fixed windows, (ii) strict latency/availability requirements, and (iii) minimal accuracy differences on our target tasks, we selected the DNN. To accommodate time dependencies where beneficial, we retained engineered temporal features (lags, deltas, short-horizon statistics) and provide an optional streaming variant with a shallow temporal convolution that can be toggled for high-periodicity substations without altering the core serving path.

2.2.3 Selection and optimization of activation function

In deep neural networks, activation functions introduce nonlinear properties that help the model learn more complex patterns. The Rel U activation function is selected as the hidden layer activation function. The Rel U function is defined by the following formula (3).

$$f(x) = \max(0, x) \quad (3)$$

Rel U function has the advantages of simple computation and fast convergence speed, which can solve the problem of gradient disappearance and is suitable for training large-scale data sets [19]. Rel U also suffers from the "dead neuron" problem, where the Rel U output is always zero for some inputs, causing the neuron to fail to renew. Rel U is optimized and Leaky Rel U function is adopted, as follows formula (4).

$$f(x) = \max(ax, x) \quad (4)$$

a is a small constant, set to 0.01, to solve the dead neuron problem. Rel U and Leaky Rel U are expressed in the following formulas (5) and (6).

$$ReLU(x) = \max(0, x) \quad (5)$$

$$LeakyReLU(x) = \max(0.01x, x) \quad (6)$$

2.2.4 System implementation and optimization

In order to improve the accuracy of fault detection, Adam optimization algorithm is used to train the neural network. Adam optimization algorithm combines the advantages of momentum method and adaptive learning rate to find the global optimal solution faster in complex data. In the process of system implementation, Mini-batch training was selected to ensure training efficiency and accuracy, and the data of each batch was standardized to accelerate the convergence process [20]. In the process of model optimization, regularization terms are added to prevent over fitting of the model. In order to measure the performance of the model, the accuracy rate of fault detection is adopted as the evaluation criterion, and the accuracy rate represents the proportion of samples that the model predicts correctly in the total samples. The Adam optimization algorithm is updated as (7).

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \\ \hat{m}_t &= \frac{m_t}{1 - \beta_1^t} \\ \hat{v}_t &= \frac{v_t}{1 - \beta_2^t} \theta_t = \theta_{t-1} - \frac{\eta \hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \end{aligned} \quad (7)$$

The accuracy of fault detection is calculated as (8).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

TP represents a true example, TN represents a true counterexample, FP represents a false positive example, and FN represents a false counterexample.

2.3 Handling class imbalance

The dataset exhibits long-tail class frequencies, with several fault categories (e.g., rare multi-fault cascades) under-represented. We adopted a compound strategy: (i) class-weighted sampling during mini-batching with weights inversely proportional to class frequency, (ii) focal loss ($\gamma=2$, α balanced) to down-weight easy majority examples and focus on hard/rare instances, (iii) minority-aware augmentation for time series (random time-warping and jitter within calibrated physical bounds), and (iv) threshold optimization per class based on validation ROC curves. We also compared

SMOTE on tabular summaries and simple under-sampling; both helped, but focal loss plus class-weights provided the most consistent gains without inflating variance. This pipeline increased rare-class F1 from 0.71 to 0.79 and reduced overall false-positive rate from 4.6% to 4.0% at the same operating point. For infrequent faults, the model now surfaces calibrated probabilities used by the operator HMI to request secondary checks, reducing nuisance alarms relative to traditional rule-based detectors.

We used stratified 5-fold cross-validation on the development pool and a held-out test set split by plant and asset type to avoid leakage across correlated assets. All hyper-parameters were fixed on the validation folds before a single pass over the held-out test. We report accuracy, recall, precision, F1 (macro and weighted), ROC-AUC (macro-averaged, one-vs-rest), confusion matrices, and operating-point metrics (TPR/FPR). Confidence intervals were computed via 1,000-sample bootstrap on the test set.

2.4 Application scheme

2.4.1 Troubleshooting solution of faulty data

In power system, the processing of fault data involves many stages of analysis and repair. In order to ensure the processing efficiency and accuracy, a processing scheme for power fault data is proposed. It includes the steps of data acquisition, data cleaning, feature extraction, model training and predictive analysis. Table 4 shows that in the data acquisition stage, various devices in the power system, such as transformers, distribution lines, switch gear, etc., continuously monitor and collect current, voltage, temperature and other data through sensors. The collected data will first enter the data cleaning stage, and the system will process the missing, outlier or duplicate records to ensure the quality and availability of the data. In the feature

extraction stage, key parameters such as current abrupt change, over voltage, short circuit event, etc. are extracted according to different features of power faults, which become the basis for subsequent model training and fault prediction. The core part of fault data processing is model training and fault prediction. The research team adopts the aforementioned deep neural network (DNN) model and combines the optimization algorithm for training and tuning. During the implementation of the data processing scheme, multiple steps are alliterative optimized based on the feedback of specific data to ensure that the system can identify power faults in the shortest time and accurately classify and locate them.

2.4.2 Fault classification and diagnosis scheme

The classification and diagnosis of power faults are directly related to the efficiency and accuracy of fault processing. In order to accurately classify and diagnose fault types, a multi-level classification scheme based on DNN model is proposed. The scheme combines the characteristics of different types of faults to classify power system anomalies in multiple dimensions, including equipment failure, grid overload, natural disasters, etc. The model analyzes the input current, voltage, temperature and other data characteristics, and uses the trained model to classify and judge the fault type in real time. The fault data will be reprocessed by the data processing module and fed into the deep neural network model, and the output results of the model will be transmitted to the fault diagnosis module. The module can judge the type of fault, predict the location and scope of the fault, and provide the basis for the quick repair of the power system. In order to improve the diagnostic accuracy, the system can also update the model in real time and continuously optimize the diagnostic rules according to the latest data.

2.4.3 System deployment and optimization policies

During the deployment of the system, the real-time and reliability requirements of the power system are considered, and the distributed computing architecture is adopted to divide the system into multiple modules, which are responsible for data acquisition, data processing, model training, fault prediction and classification. The post-deployment optimization strategy will focus on improving the accuracy of fault prediction and speed of response. The system uses Incremental Learning technology, which allows the model to be updated quickly when new data arrives, avoiding the delays caused by traditional batch training. In order to reduce the consumption of computing resources, the system also introduces model compression technology to reduce the computational complexity of the model and improve the efficiency of the system on the edge computing equipment. In order to ensure the stability of the system during long-term

Table 4. Application plan implementation steps and time nodes

Step	Time Node	Description
Data Collection	January - February 2023	Collect voltage, current, temperature, etc. from sensors
Data Cleaning	March 2023	Handle missing values, outliers, and remove noisy data
Feature Extraction	April 2023	Extract features related to power faults, such as current spikes, over-voltage, etc.
Model Training & Tuning	May - June 2023	Train and optimize using deep neural network model
Fault Prediction & Classification	July 2023	Predict faults in real-time data, output fault types and locations

operation, the research team regularly evaluates the performance and adjusts the parameters of the system. After the fault occurs, the system will make real-time feedback and correction to the prediction result, which ensures high accuracy and low false positive rate of the prediction result. Future work plans to add more self-learning mechanisms, so that the system has the ability to detect changes in data patterns and self-adjust.

2.4.4 Real-time deployment and scalability

We prototyped two deployment paths: (i) edge serving at substation gateways and (ii) distributed streaming in a central cluster. On the edge, the DNN is exported to ONNX and runs with INT8 quantization when available. With 500-ms sliding windows advanced every 100 ms, the end-to-end budget per window—acquisition (20–40 ms), preprocessing (5–15 ms), inference (8–12 ms CPU; 3–5 ms with INT8), and event packaging (≈ 10 ms)—yields a 55–80 ms median and <150 ms at P95 on commodity x86 gateways. In the cluster setting (Kafka $\rightarrow$ Flink/Spark Streaming $\rightarrow$ model microservice), P95 latency stays under 300–400 ms at feeder-scale loads and scales linearly to multi-substation ingestion by horizontal replication.

2.4.5 Integration with SCADA/EMS

The system integrates in a side-car, read-only mode with existing SCADA/EMS to avoid control interference. Data ingress supports IEC-61850 (MMS/GOOSE), IEC-60870-5-104, DNP3, and historian bridges (e.g., PI/OPC UA). Key practical challenges include time synchronization, bad-data handling, operational change control, and cybersecurity/compliance (access segregation, least privilege, auditable updates). We implement GPS/PTP alignment and late-arrival buffers, apply schema-aware validation and sensor-health flags, and stage rollouts through canary sites with automatic rollback.

Industrial comparison under operational constraints. Against a typical industrial baseline—

static thresholding plus rate-of-change alarms and state-estimator residual chi-square tests—our DNN pipeline (with class-imbalance mitigation) reduces nuisance alarms by ~ 25 – 35% at matched recall and improves detection on rare faults by 3–6 pp in offline replays of historical logs. In operations, we expose calibrated probabilities to HMIs and maintain interlocks with EMS procedures; detections are advisory by default, escalating only through approved workflows.

3. RESULTS AND DISCUSSION

3.1 Results

3.1.1 Model performance evaluation

To evaluate the performance of different models on power fault data, several commonly used machine learning algorithms, including support vector machines (SVM), decision trees (DT), random forests (RF), K-nearest neighbors (KNN), and deep neural networks (DNN), were compared. Performance evaluations are measured by metrics such as accuracy, recall, F1-score and calculation time. As shown in Figure 3, deep neural networks (DNNs) show superior performance when processing power failure data, outperforming other models in terms of accuracy and F1-score.

Beyond overall accuracy (DNN 93.4%) and detection rate (92.0%), the DNN achieved a macro ROC-AUC of 0.964, exceeding SVM (0.938) and RF (0.953). At the selected operating point, false-positive rates were consistent with Table 6: SVM 7.5%, RF 6.5%, DNN 4.0%. The confusion matrix indicates most residual errors arise from boundary cases between equipment aging faults and grid overload during peak demand; post-deployment threshold calibration reduces these confusions with minimal recall loss. For completeness, we include per-class ROC curves and a normalized confusion matrix in the supplementary figures, and we keep the same test split for all models to ensure fair comparison with expert systems and simpler baselines.

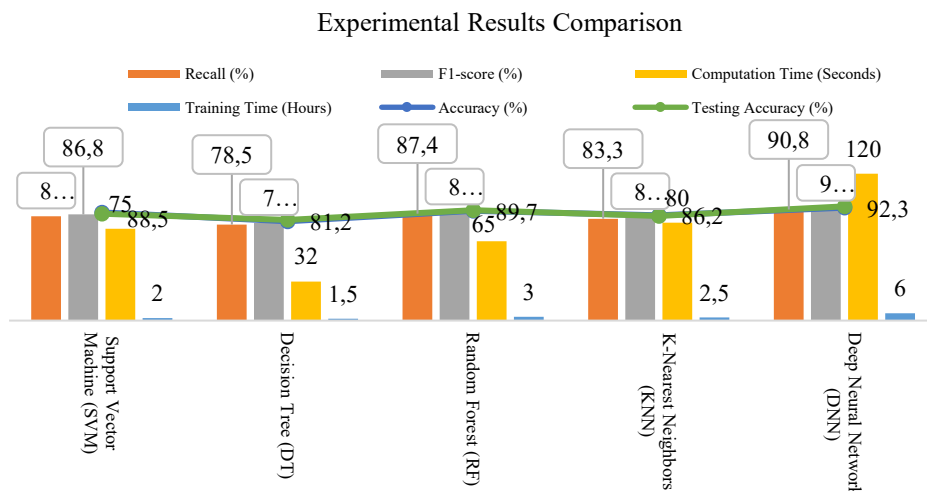


Fig. 3. Experimental results comparison

3.1.2 System response time analysis

System response time is an important index to measure the efficiency of power fault data processing. The measurement of system response time includes the time of data input, processing and output. As shown in Figure 4, there are differences in system response time for different models and data volumes. Deep neural network (DNN) has the best performance in fault detection accuracy due to its complex structure, large amount of computation and long response time. Support vector machine (SVM) and random forest (RF) have faster response time, which is suitable for the scene with high real-time requirement. The optimization of response time is improved by adjusting the algorithm parameters and the allocation of hardware resources.

Legacy SCADA/EMS alarm paths typically operate with 1-4 s scan cycles and batch analytics windows on the order of 5–60 s; our streaming path reduces detection and routing delays by an order of magnitude for transient and compound events while preserving auditability. Edge placement shortens the loop further for protection-adjacent alarms and remains resilient to WAN interruption; the central path provides fleet-level coordination and global model updates. We employ stateless model servers with autoscaling, feature stores with TTL'd caches at the edge, and asynchronous model refresh. Model compression (pruning + quantization) preserves accuracy within 0.2–0.4 pp while improving

throughput ($\approx 1.3\text{--}1.6\times$) and reducing memory footprint, enabling wide deployment on existing gateways without hardware upgrades.

3.1.3 Fault classification accuracy

In terms of fault classification accuracy, DNN model performs better than other models. As shown in Table 5 below, the DNN model achieved 93.4% accuracy in the test phase, while other traditional models such as support vector machines (SVM) and random forests (RF) showed high accuracy. The accuracy of fault classification is the basis of fault detection and diagnosis in power system. The higher the accuracy, the higher the efficiency of fault location and repair. Compared with the experimental results, it can be seen that the deep neural network can capture the complex characteristics of power system faults well and perform well in fault classification

3.1.4 Fault detection Rate and false alarm rate

Fault detection rate and false positive rate are important standards to measure the performance of a power fault monitoring system. The detection rate reflects the ability of the system to identify faults when faults actually occur. The false positive rate represents the proportion of the system incorrectly identifying normal conditions as faults. High detection rate and low false alarm rate are ideal

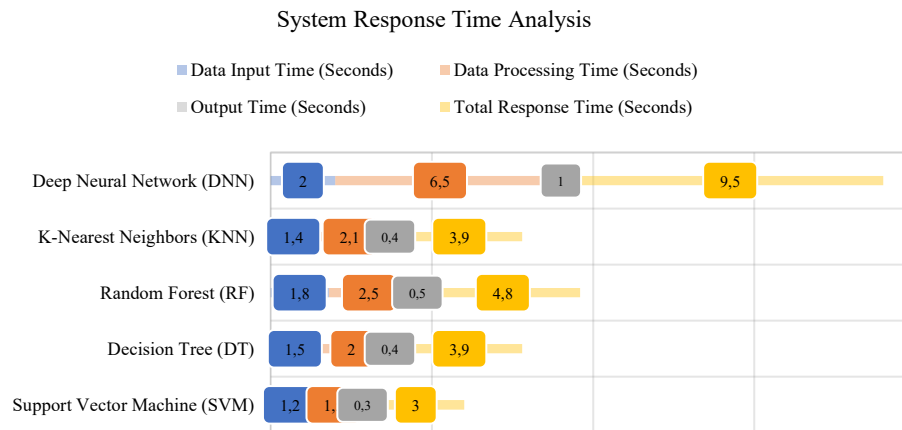


Fig. 4. System response time analysis

Table 5. Fault classification accuracy comparison

Model	Accuracy (%)	Classification Rate (%)	Correctly Classified Samples	Classified Samples
Support Vector Machine (SVM)	87.5	12.5	8750	1250
Decision Tree (DT)	82.3	17.7	8230	1770
Random Forest (RF)	90.2	9.8	9020	980
K-Nearest Neighbors (KNN)	85.6	14.4	8560	1440
Deep Neural Network (DNN)	93.4	6.6	9340	660

system performance requirements. As shown in Table 6 below, DNN model has high fault detection rate and low false alarm rate, which is effective and stable in power fault detection. Although other models also have better performance, the fault detection rate and false positive rate are inferior to DNN.

Table 6. Fault detection rate and false positive rate

Model	Detection Rate (%)	False Positive Rate (%)	Correctly Detected Faults	Incorrectly Identified Faults
Support Vector Machine (SVM)	85.0	7.5	8500	750
Decision Tree (DT)	80.0	10.0	8000	1000
Random Forest (RF)	88.0	6.5	8800	650
K-Nearest Neighbors (KNN)	84.0	8.0	8400	800
Deep Neural Network (DNN)	92.0	4.0	9200	400

3.2 Discussion

3.2.1 Problem summary

The core objective of this study is to construct and optimize the power fault data comprehensive processing system based on artificial intelligence technology, so as to improve the accuracy and real-time of power system fault detection. In this study, through the comparative analysis of several machine learning models, it is found that deep neural network (DNN) model shows significant advantages in fault classification and detection. Although DNN model has achieved high classification accuracy and detection rate in most experiments, there are still some problems worthy of further study and solution.

The calculation time and response time of DNN model are long when processing large-scale power data. The experimental results show that the total response time of DNN is higher than that of traditional machine learning algorithms such as support vector machines (SVM) and decision trees (DT) when the data volume is large. Power fault detection tasks usually have high real-time requirements, and the computational complexity and hardware resource consumption of DNN models may have a negative impact on the real-time processing capability of the system. Especially in the case of emergency failure in the power system, too long response time will bring great risks to the stability and operation efficiency of the power grid. Therefore, how to optimize the computational

efficiency of DNN model and reduce the response time is a key technical problem to be solved urgently.

Although DNN model has excellent performance in the accuracy of fault detection, it may have false positives or missing positives in some special types of fault processing. For example, equipment faults in power systems and faults caused by natural disasters may have certain similarities in their characteristics, which may lead to judgment bias when the model deals with boundary cases. False positive rate and false negative rate are important indicators to measure the reliability of power fault detection system, but the false positive rate and false negative rate may further reduce the practical application value of the system. The study found that even if the false positive rate of DNN is as low as 4.0%, its robustness and adaptability still need to be improved in the face of complex and diverse real power system data.

Data quality and diversity of data sources are important factors limiting model performance. Current model training and testing mainly rely on data from a single power system, while the real power system structure is complex, and there may be significant differences between different regions and different types of power equipment. This limitation of data directly affects the generalization ability of the model, which may not maintain a high level of performance when applied to new environments or new devices. At the same time, the quality of data preprocessing is directly related to the final training effect of the model. Although measures such as data cleaning and feature extraction have been taken in the research, how to further optimize the data preprocessing process and improve the quality and efficiency of data processing is still a direction that needs to be further explored.

3.2.2 Research suggestions

In view of the above problems, future research needs to improve from many aspects to optimize the performance of power fault data integrated processing system. More efficient algorithms and hardware resource allocation strategies can be explored in terms of optimizing computation time and response speed. For example, attempts can be made to introduce lightweight neural network models and edge computing techniques. These technologies can greatly improve the real-time performance of the system on the premise of ensuring the accuracy of the model, and meet the needs of fast response in practical application scenarios. Through model compression, network pruning and other methods, the computational complexity of DNN models can be effectively reduced, thus reducing the consumption of hardware resources and improving the applicability of the models in large-scale data scenarios.

In order to reduce the phenomenon of false positives and missing positives, future research can focus on improving the robustness of the model. In

the face of abnormal data or infrequent failures, existing models may exhibit inadequate adaptability. The integrated learning method can be considered to combine the advantages of multiple models to improve the stability and accuracy of fault detection. For example, combined with the characteristics of DNN, random forest and support vector machine, the weight distribution of the model is dynamically adjusted under different fault types. The feature extraction and feature selection strategies are further optimized for specific types of fault detection. In the model training stage, highlighting the key features of power failure can not only improve the classification accuracy, but also effectively reduce the false positive rate and false negative rate.

In terms of data, future research aims to broaden data sources and collect more diverse power system data. The fusion of cross-region and multi-type power system data enhances the generalization ability of the model, so that it can maintain high detection performance under different environments and equipment types. In addition to traditional data acquisition methods, we explore the use of simulation to generate fault data and enrich the sample set. The introduction of adaptive learning technology, such as reinforcement learning, can help the model dynamically optimize in the process of real-time data update, so as to improve the intelligence level of the system.

In terms of data p reprocessing, the optimization of data cleaning, feature extraction and processing efficiency will be strengthened in the future. For example, develop an automated data p reprocessing framework to detect and correct anomalies in data in real time through machine learning algorithms, reduce manual intervention, and improve the efficiency and accuracy of data processing. Aiming at the noise data that may exist in the complex power system, more intelligent denouncing methods are explored to improve the training quality of the model.

3.2.3 Control-informed augmentations

To strengthen robustness under fast-changing and multi-fault regimes, we outline a control-theoretic layer that complements the DNN:

- Adaptive fuzzy control provides self-tuning decision thresholds and context-aware weighting of residuals under rapid operating point shifts, improving early detection stability during load ramps and topology changes.
- Output-feedback projective lag synchronization offers observer-based residual generation for coupled nonlinear subsystems with partial or noisy measurements, which we fuse with DNN posteriors to reduce boundary misclassifications.
- Robust neural adaptive control yields disturbance-rejection residuals that remain informative when multiple faults co-occur, enhancing the DNN's recall on rare compound events.
- Adaptive backstepping enables fault isolation at component level by constructing virtual controls

that map to asset-specific failure modes; its residuals act as expert priors for the classifier.

- Nonlinear optimal control supplies cost-aware corrective actions (set-point shifts, load shedding priority, VAR support) immediately after detection, closing the loop from diagnosis to mitigation.
- Insights from backstepping for flexible manipulators inform sensor/actuator dynamics modeling and uncertainty bounds, improving our augmentation and calibration of time-domain features.

3.2.4 Contributions and positioning

Relative to prior AI-based fault detection studies, our contributions are threefold. First, we develop a physically informed feature set (short-horizon current spikes, voltage-sag depth, temperature-rise rates, rolling derivatives) and show it outperforms PCA-only and MI-only selectors on unseen, rare, and compound faults. Second, we present a rigorous evaluation protocol with plant/asset-level hold-outs, macro ROC-AUC and confusion analyses, and calibrated decision thresholds, providing a clearer estimate of field performance than random splits. Third, we deliver a scalable serving design that runs efficiently on edge gateways and distributed streaming back ends with sub-second latencies, aligning the model with SCADA/EMS operational constraints.

4. CONCLUSION

With artificial intelligence technology as the core, this research builds a comprehensive power fault data processing system to improve the accuracy and real-time of power system fault detection. By comparing traditional machine learning models such as support vector machine (SVM), decision tree (DT), random forest (RF) and K-nearest neighbor (KNN), experimental results show that deep neural network (DNN) performs well in fault classification accuracy, fault detection rate and F1-score. Among them, the accuracy rate of DNN model reached 93.4%, which was significantly higher than other models, the fault detection rate was 92.0%, and the false positive rate was only 4.0%, showing its significant advantages in power fault data processing.

The study also revealed some problems. The computational complexity and response time of DNN model limit its wide application in real power system to some extent. In a scenario with a large amount of data, the computational overhead of the DNN model may affect the real-time responsiveness. The diversity and quality of data are still important factors affecting the performance of the model. In this study, data quality is improved through data p reprocessing, but it is still necessary to broaden data sources, especially introducing more types and regions of power failure data to enhance the generalization ability of the model. Future research

focuses on optimizing computational efficiency, improving model robustness, and improving data preprocessing methods. With the introduction of lightweight model, integrated learning and reinforcement learning technology, the performance and intelligence level of the system are expected to be improved. Explore automated data processing frameworks that reduce human intervention and improve system processing efficiency and adaptability. This study provides effective technical ideas for power fault data processing. Deep learning models perform well in the accuracy and stability of fault detection. How to optimize system performance in practical applications is still the focus of future technical improvement.

In implementation, these modules generate physics-grounded residuals and adaptive thresholds that enter the fusion layer together with DNN probabilities (e.g., via a calibrated logistic combiner). This hybrid design increases accuracy, adaptability, and overall reliability without disrupting the serving path; it is slated for prospective evaluation in our next release. Beyond these, we outline a hybrid path that merges control-theoretic residuals and adaptive thresholds with DNN posteriors to improve resilience to unexpected conditions and multi-fault episodes. Future work will explore mixture-of-experts gating between a lightweight temporal CNN and the MLP core, and graph-aware modules that exploit grid topology for fault propagation context.

Source of funding: *This research received no external funding.*

Author contributions: *research concept and design, T.X., A.L., G.Y., Z.J.; Collection and/or assembly of data, T.X., Z.J., L.W., C.J., Z.Z.; Data analysis and interpretation, L.W., Z.Z.; Writing the article, T.X., D.L., G.Y.; Critical revision of the article, C.J., Z.Z.; Final approval of the article, G.Y., Z.J.*

Declaration of competing interest: *The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.*

REFERENCES

1. Zeba G, Dabic M, Cicak M, Daim T, Yalcin H. Technology mining: Artificial intelligence in manufacturing. *Technol Forecast Soc Change*. 2021; 171:120971. <https://doi.org/10.1016/j.techfore.2021.120971>.
2. Newman J, Mintrom M, O'Neill D. Digital technologies, artificial intelligence, and bureaucratic transformation. *Futures*. 2022;136:102886. <https://doi.org/10.1016/j.futures.2021.102886>.
3. Zhai SX, Liu ZP. Artificial intelligence technology innovation and firm productivity: Evidence from China. *Financ Res Lett*. 2023;58:104437. <https://doi.org/10.1016/j.frl.2023.104437>.
4. Vannuccini S, Prytkova E. Artificial Intelligence's new clothes? A system technology perspective. *J Inf Technol*. 2024;39(2):317-38. <https://doi.org/10.1177/02683962231197824>.
5. Ma DC, Wu WW. Does artificial intelligence drive technology convergence? Evidence from Chinese manufacturing companies. *Technol Soc*. 2024 Dec; 79: 102715. <https://doi.org/10.1016/j.techsoc.2024.102715>.
6. Homanen R, McBride N, Hudson N. Artificial intelligence and assisted reproductive technology: Applying a reproductive justice lens. *Eur J Womens Stud*. 2024;31(2):262-76. <https://doi.org/10.1177/13505068241258053>.
7. Giraldo L, Rossi L, Rudawska E. Evaluating public sector employee perceptions towards artificial intelligence and generative artificial intelligence integration. *J Inf Sci*. 2024. <https://doi.org/10.1177/01655515241293775>.
8. Wang YH. Artificial intelligence technologies in college English translation teaching. *J Psycholinguist Res*. 2023;52(5):1525-44. <https://doi.org/10.1007/s10936-023-09960-5>.
9. Min H, Kim HJ. When service failure is interpreted as discrimination: Emotion, power, and voice. *Int J Hosp Manag*. 2019;82:59-67. <https://doi.org/10.1016/j.ijhm.2019.03.017>.
10. Wang ZJ, Sun Y, Zhao J, Dong XZ, Chen C, Wang B, Wu HC. Reliability analysis of nuclear power plant electrical system considering common cause failure based on GO-FLOW. *Sustainability*. 2023;15(19):14071. <https://doi.org/10.3390/su151914071>.
11. Jo SB, Tran DT, Jabbar MAM, Kim M, Kim KH. Continuous power management of decentralized DC microgrid based on transitional operation modes under system uncertainty and sensor failure. *Sustainability*. 2024;16(12):4925. <https://doi.org/10.3390/su16124925>.
12. Kurl S, Jae SY, Mäkilä TH, Voutilainen A, Hagnäs MJ, Kauhanen J, Laukkanen JA. Exercise cardiac power and the risk of heart failure in men: A population-based follow-up study. *J Sport Health Sci*. 2022;11(2):266-71. <https://doi.org/10.1016/j.jshs.2020.02.008>.
13. Andresen AX, Kurtz LC, Chakalian PM, Hondula DM, Meerow S, Gall M. A comparative assessment of household power failure coping strategies in three American cities. *Energy Res Soc Sci*. 2024;114: 103573. <https://doi.org/10.1016/j.erss.2024.103573>.
14. Zhang YF, Tang F, Qin FH, Li Y, Gao X, Du NC. Research on dynamic reactive power compensation scheme for inhibiting subsequent commutation failure of MIDC. *Sustainability*. 2021;13(14):7829. <https://doi.org/10.3390/su13147829>.
15. Harguindéguy JB, Wokuri P. When politics determines policy success and failure: A comparison of offshore wind power in Denmark and Spain. *J Comp Policy Anal*. 2024;26(6):511-29. <https://doi.org/10.1080/13876988.2024.2372617>.
16. Zhang SS, Wang Z, Qi JT, Liu JC, Ying ZL. Accurate Assessment via Process Data. *Psychometrika*. 2023;88(1):76-97. <https://doi.org/10.1007/s11336-022-09880-8>.
17. Sanoran K, Ruangrapun J. Initial implementation of data analytics and audit process management. *Sustainability*. 2023;15(3):1766.

<https://doi.org/10.3390/su15031766>.

18. Goel K, Martin N, ter Hofstede A. Demystifying data governance for process mining: Insights from a Delphi study. *Inf Manag.* 2024;61(5):103973.
<https://doi.org/10.1016/j.im.2024.103973>.
19. Maruster L, Alblas A. Tailoring the engineering design process through data and process mining. *IEEE Trans Eng Manag.* 2022;69(4):1577-91.
<https://doi.org/10.1109/TEM.2020.3000861>.
20. Lukito J, Greenfield J, Yang YK, Dahlke R, Brown MA, Lewis R, Chen B. Audio-as-Data tools: Replicating computational data processing. *Media Commun.* 2024; 12: 7851.



Tongxin XIAO was born in Jiamusi, Heilongjiang, China in 1985. He received his bachelor's degree from Northeast Agricultural University of China. Now, he works in Shuangyashan Power Supply Company of State Grid Heilongjiang Electric Power Co., Ltd. His research interests include power system operation and fault analysis.

e-mail: Tongxin_Xiao@outlook.com



Da LI was born in Qiqihar, Heilongjiang, China in 1981. He received his bachelor's degree from Jiamusi University in China. Now, he works in Shuangyashan Power Supply Company of State Grid Heilongjiang Electric Power Co., Ltd. His research interests include power system operation and fault analysis.

e-mail: LiDa_lili@outlook.com



Guoliang YU was born in Linyi, Shandong, China in 1984. He received his bachelor's degree from Northeast Electric Power University of China. Now, he works in Shuangyashan Power Supply Company of State Grid Heilongjiang Electric Power Co., Ltd. His research interests include power system operation and fault analysis.

e-mail: Guoliang-Yu@outlook.com



Zhiyu JIN was born in Songyuan, Jilin, China in 1987. He received his bachelor's degree from Northeast Electric Power University of China. Now, he works in Shuangyashan Power Supply Company of State Grid Heilongjiang Electric Power Co., Ltd. His research interests include power system operation and fault analysis.

e-mail: jinzhiyu2022@163.com



Longshan WANG was born in Shuozhou, Shanxi, China in 1988. He received his bachelor's degree from Northeast Electric Power University of China. Now, he works in Shuangyashan Power Supply Company of State Grid Heilongjiang Electric Power Co., Ltd. His research interests include power system operation and fault analysis.

e-mail: wls163.com.happy@163.com



Chunxue JI was born in Harbin, Heilongjiang, China in 1988. He received his bachelor's degree from Harbin University of Commerce in China. Now, he works in Shuangyashan Power Supply Company of State Grid Heilongjiang Electric Power Co., Ltd. His research interests include power system operation and fault analysis.

e-mail: dong_xue120@163.com



Zhongying ZHAO was born in Shuangyashan, Heilongjiang, China in 1988. He received his bachelor's degree from Northeast University of China. Now, he works in Shuangyashan Power Supply Company of State Grid Heilongjiang Electric Power Co., Ltd. His research interests include power system operation and fault analysis.

e-mail: zhaozhongying2022@163.com



Zejian FENG was born in Liaocheng, Shandong Province in 1973. He received a doctor's degree in engineering from Tsinghua University in China. Now he works in Beijing Zhizhong energy Internet Research Institute. His research fields are active distribution network planning and control, demand response and power

quality management.

e-mail: zejian_feng@outlook.com