Multi-Source Deep Learning-Based Fault Diagnosis for High-Frequency Synchronous Power Grid Data

Yushu Cheng*, Shushu Wang, Anqi Chen, Zhixia Bai, Zhijin Zhu, Huinan Wang
State Grid Shanxi Electric Power Company Marketing Service Center, Taiyuan 030000, China
E-mail: yushuchengg@163.com
*Corresponding author

Keywords: WAMS, SCADA, deep neural network, pattern recognition

Received: August 5, 2025

To meet the need for high-frequency synchronous full-power data fault diagnosis in new power systems, this study proposes an innovative method combining multi-source data acquisition technology and deep neural networks for accurate power system fault identification and efficient fault location. Firstly, it integrates multi-source heterogeneous data from WAMS, SCADA, and meteorological sensors to form a holistic sensing network. The core of our method is a hybrid deep neural network architecture that combines Convolutional Neural Networks (CNN) for spatial feature extraction and Long Short-Term Memory (LSTM) networks for temporal sequence modeling, enhanced with an attention mechanism for adaptive feature fusion. Secondly, deep neural networks extract features and recognize patterns in the collected full-power data to identify fault types, locate faults, and analyze fault causes. Experimental results demonstrate the exceptional performance of our approach, achieving a fault diagnosis accuracy of 99.71%. This represents a significant improvement over traditional baseline models, showcasing its superior capability in handling complex power system fault scenarios. Finally, the research shows that this method has made significant breakthroughs in data synchronization accuracy, diagnosis accuracy, and adaptability to complex scenarios.

Povzetek: Študija predstavlja splošno metodo za natančno in zanesljivo diagnostiko okvar v elektroenergetskih sistemih z uporabo večvirovnih podatkov in globokega učenja.

1 Introduction

With the continuous expansion of the power system's scale and the widespread access to new energy, the power grid structure is becoming increasingly complex, and fault diagnosis faces core challenges such as high data dimension, strict response timeliness, and insufficient diagnostic accuracy. Fault diagnosis in the power grid is a crucial aspect of maintaining the safe and stable operation of the power system. Traditional fault diagnosis methods rely on manual inspection and threshold comparison, and the false alarm rate in complex scenarios is as high as 30%; they cannot capture microsecond-level transient fault characteristics. At present, commonly used power grid fault diagnosis methods include expert system [1], numerical calculation analysis [2], rough set, Bayesian network [3], artificial neural network [4], and analytical model [5]. Each of these methods has its advantages and disadvantages. However, these methods are often difficult to accurately deal with when the switching information is misplaced or lost, and the interpretability of the diagnosis results is poor. Although it can meet the needs of fault diagnosis to a certain extent, it often has limitations when facing the complex and changeable power grid environment. For example, expert systems rely on expert experience and rule bases, and it is difficult to cope with uncovered failure situations. Numerical analysis methods often involve high computational complexity, making it challenging to satisfy real-time performance requirements. Traditional analytical models for power system fault diagnosis rely heavily on accurate system modeling and lack adaptability. Recently, machine - learning and neural - network - based methods have emerged with the fast progress of artificial intelligence [6], offering new ideas for power grid fault diagnosis.

Modern power grids use multi-source Data acquisition and monitoring systems (SCADA) [7], fault information systems (FIS) [8], wide area measurement systems (WAMS) [9] to collect real - time data (switching info, electrical data, fault recording data), covering both static/dynamic grid info and fault - related changes. But effectively integrating this multi-source heterogeneous data and extracting useful fault diagnosis info is a key current research issue.

Deep neural networks excel at feature learning and pattern recognition [10], achieving remarkable results in image recognition and speech processing [11]. Their strong nonlinear fitting and adaptive learning abilities give them an edge in handling complex data [12]. When applied to power grid fault diagnosis, they process large-scale data and complex fault patterns, improving diagnostic accuracy and real-time performance. Plus,

they automatically learn data's hidden features, reducing manual feature extraction work and boosting diagnostic systems' intelligence [13].

This paper presents a multi-source data and deep neural network-based fault diagnosis method for high-frequency synchronous full-power data. It builds a diagnosis framework by integrating SCADA, FIS, and WAMS data. Using deep neural networks for feature extraction and pattern recognition in the data can effectively identify fault types, locations, and causes. Compared to traditional methods, this approach better utilizes multi-source data and optimizes the diagnosis model through deep learning, enhancing diagnostic accuracy and real-time performance. The paper's key work involves integrating multi-source data to form a comprehensive fault diagnosis framework.

(1) By integrating multi-source heterogeneous data such as wide area measurement system (WAMS), data acquisition and monitoring system (SCADA), and meteorological sensors, a holographic sensing network covering electrical parameters, environmental status, and equipment working conditions is formed.

(2) WAMS provides microsecond-level accuracy voltage/current phasor data, SCADA collects equipment operating status and fault alarm information, and weather sensors monitor environmental variables such as wind speed, temperature, and humidity in real-time. Together, the three form a multi-dimensional feature space.

(3) The deep neural network performs feature extraction and pattern recognition on the collected full-power data, enabling effective fault type identification, location pinpointing, and cause analysis.

2. Theoretical basis and construction of index system

2.1 Multi-source data acquisition

Data cleaning serves as a pivotal phase for maintaining data quality during multi-source data collection. Its main tasks include removing duplicate values, correcting erroneous values, and handling missing values [14]. The occurrence of duplicate values may be due to redundancy in the data acquisition process or overlap between different data sources. For example, in a customer information dataset containing multiple data sources, the customer's name, ID number, and other information can be used as unique identification fields to eliminate duplicate customer records. For error values, they need to be detected and corrected according to the business rules and the reasonableness range of data. Taking meteorological data as an example, the normal range of temperature is usually between 50°C and 50°C. If a temperature value exceeds this range, it may be an error value, which needs to be corrected by comparing the data of adjacent time points or other sensors in the same area. The treatment of missing values needs to choose an appropriate method according to the degree of missing values and the nature of the data. If there are a few missing values, you can directly delete records containing missing values. If there are many missing values that have a great impact on the data analysis results, methods such as mean filling, mode filling, or predictive filling using machine learning algorithms can be used. For example, in a sales data set, some records lack information on sales amounts. If these records account for a small proportion, they can be deleted directly; However, if the proportion is relatively large, these missing sales amount values can be filled according to the historical average sales price of the product or the average sales price of the same product in the same time period.

Data fusion integrates multi-source data to generate more accurate and complete information. It operates at three levels: low-level (direct processing of raw data), mid-level (feature-based fusion), and high-level (decision-level fusion). Low-level fusion occurs at the data acquisition stage, handling original sensor data directly. [15]. For example, in a sensor network, multiple temperature sensors measure the temperature of the same area, and these raw temperature data are fused into a more accurate temperature value by taking an average or weighted average. Intermediate fusion is carried out at the feature level. First, the features of each data source are extracted, and then the feature vectors are fused. For example, in the application scenario of combining image recognition with text description, visual feature vectors such as shape, color, and texture are extracted from images, keywords and semantic feature vectors are extracted from a text description, and then these feature vectors from different sources are spliced into a comprehensive feature vector for subsequent classification or recognition tasks [16]. Advanced fusion is carried out at the decision-making level, and the data analysis results of different data sources are synthesized to make the final decision. For example, in a medical diagnosis system, on the one hand, symptom analysis is performed based on the patient's medical record text data, and preliminary diagnosis results are generated; on the other hand, image analysis is performed based on medical image data to also generate preliminary diagnosis results, and then these two results are combined together, weighing their respective credibility and importance, and making the final diagnosis decision.

2.2 Deep neural network

2.2.1 Basic structure of deep neural network

In deep neural networks, neurons are basic units. They receive inputs from previous neurons, apply nonlinear transformations using activation functions, and then send the outputs to the next layer. The calculation process is shown in formula (1):

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \quad (1)$$

Where x_i is the input signal, w_i is the connection weight, b is the bias term, and f is the activation function. The activation function determines the output form of neurons. Common activation functions include Sigmoid, Tanh and ReLU.

Deep neural network structures are determined by neuron connections, which can be full, local, or none [17].

Fully Connected (FC) neurons link to all neurons in the previous layer, often seen in hidden and output layers. Locally Connected (LC) neurons only connect to a local area of the upper layer, typical in CNNs for data like images. No Connection (NC) means some neurons have no direct link, such as in GANs' generators and discriminators.

2.2.2 Hierarchical structure of deep neural networks

A deep neural network is composed of an input layer, hidden layers, an output layer, and activation functions [18].

Specifically, the input layer is the entrance to the deep neural network to receive external data. The number of neurons in the input layer is usually the same as the feature dimension of the input data.

Hidden layers are the core part of deep neural networks and usually contain multiple layers. Each hidden layer extracts the features of the input data and abstracts them step by step through the calculation of neurons and the transformation of the activation function [19]. The number of hidden layers and neurons significantly impacts deep neural network performance. More hidden layers boost network expressiveness but may cause overfitting and higher computational demands. The output layer, as the final layer, produces the prediction result [20].

The number of neurons in the output layer varies with the task. For instance, in classification tasks, the output layer typically has as many neurons as there are classes, with each neuron indicating the probability of the input data belonging to a specific class.

The activation function is crucial for introducing nonlinearity into deep neural networks. Common types include the Sigmoid, Tanh, and ReLU functions.

The Sigmoid function is a Sigmoid curve function, and its calculation process is shown in formula (2):

$$f(x) = \frac{1}{1+e^{-x}} \quad (2)$$

Where x represents the input, and the output range of this function is between (0, 1), which is suitable for binary classification problems. However, Sigmoid functions are prone to gradient vanishing problems in deep networks.

The Tanh function is a variant of the Sigmoid function, and its calculation process is shown in formula (3):

$$f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \quad (3)$$

Among them, x represents the input, and the output range of this function is between (-1, 1), which has better convergence compared with Sigmoid function, but there is still the problem of gradient vanishing.

The ReLU function is a commonly used activation function, and its calculation process is shown in formula (4):

$$f(x) = max(0, x) \quad (4)$$

Among them, x represents the output and max represents the maximum value function, which is simple in calculation and fast in convergence, and can effectively alleviate the problem of gradient disappearance. However, it can also lead to neuron "death" problems, that is, the output of some neurons is constantly 0.

2.2.3 Training of neural networks

Training a deep neural network involves three key components: a loss function, an optimization algorithm, and backpropagation [21]. Specifically, the loss function quantifies the discrepancy between the model's predictions and the true values. Common examples are Mean Squared Error (MSE) and Cross-Entropy Loss (CEL) [22]. Among them, MSE is suitable for regression problems, and its calculation process is shown in formula (5):

$$L(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (5)$$

Where y is the true value, $\hat{y}$ is the predicted value, n number of samples.

CEL is suitable for classification problems, and its calculation process is shown in formula (6):

$$L(y, \hat{y}) = -\sum_{i=1}^{n} y_i \widehat{log\, y_i} \quad (6)$$

Where y is the single-hot encoding of the true value and $\hat{y}$ is the probability distribution of the predicted value.

Optimization algorithms are used to update the weights and biases of the network to minimize the loss function [23]. Common optimization algorithms include Stochastic Gradient Descent (SGD) and Adam optimization algorithm. Specifically, SGD is a commonly used optimization algorithm, and its calculation process is shown in formula (7):

$$w_{t+1} = w_t - \eta \nabla L(w_t) \quad (7)$$

Where w_t is the weight at the current moment, η is the learning rate, and $\nabla L(w_t)$ is the gradient of the loss function to the weight. Adam is an adaptive learning rate optimization algorithm, which combines the advantages of momentum and RMSProp, with fast convergence speed and better performance. Its updating rules are complicated, involving momentum terms and second-order moment estimation.

Table 1: Comparative analysis of recent fault diagnosis methods for power systems.

Method	Key Features	Data Sources	Fault Types Covered	Reported Performance	Key Limitations
CBMA	Multimodal Attention Learning	Not Specified	Pumping System Faults	-	Lacks microsecond-level data handling
SA-	Multi-Scale	Not Specified	General	-	Weak under variable

MSCNN	Convolution				grid conditions
MC-MSDARL	Multi-Scale, Deep Reinforcement Learning	Not Specified	General	-	Complex, not optimized for high-frequency data
BPNN	Backpropagation Neural Network	SCADA	General	95.60% Accuracy	Low accuracy, poor feature extraction
CNN	Convolutional Neural Network	SCADA/WAMS	General	96.69% Accuracy	Limited temporal feature extraction
Proposed (CNN-LSTM)	Hybrid CNN-LSTM, Attention Mechanism, Multi-source Fusion	WAMS, SCADA, Meteorological	Voltage, Frequency, External Faults	99.71% Accuracy	Superior accuracy & robustness

A comparative summary of recent fault diagnosis methods is provided in Table 1, highlighting key limitations in the state-of-the-art that this work aims to address.

3 Fault diagnosis of high-frequency synchronous full-power data based on multi-source data acquisition and deep neural network

3.1 Overall architecture of model construction

In this paper, a high-frequency synchronous full-power data fault diagnosis method based on multi-source data acquisition and deep neural network is proposed. By integrating multi-source data such as SCADA, FIS, and WAMS, this method constructs a high-frequency synchronous full-power data fault diagnosis framework. Using a deep neural network to perform feature extraction and pattern recognition on the collected full-power data can effectively identify fault types, locate fault locations, and analyze fault causes. The proposed CNN-LSTM model accepts input data with dimensions of 100-time steps × 15 features. The CNN module comprises three convolutional layers with 64, 128, and 256 filters respectively, all using a kernel size of 3 and ReLU activation. The LSTM module consists of two layers with 128 and 64 hidden units. The architecture incorporates a Fusion Attention Network (FAN-BD) mechanism to enhance feature selection, followed by a softmax output layer for classification. This detailed specification ensures full reproducibility of our model architecture. Its network structure is shown in Figure 1:

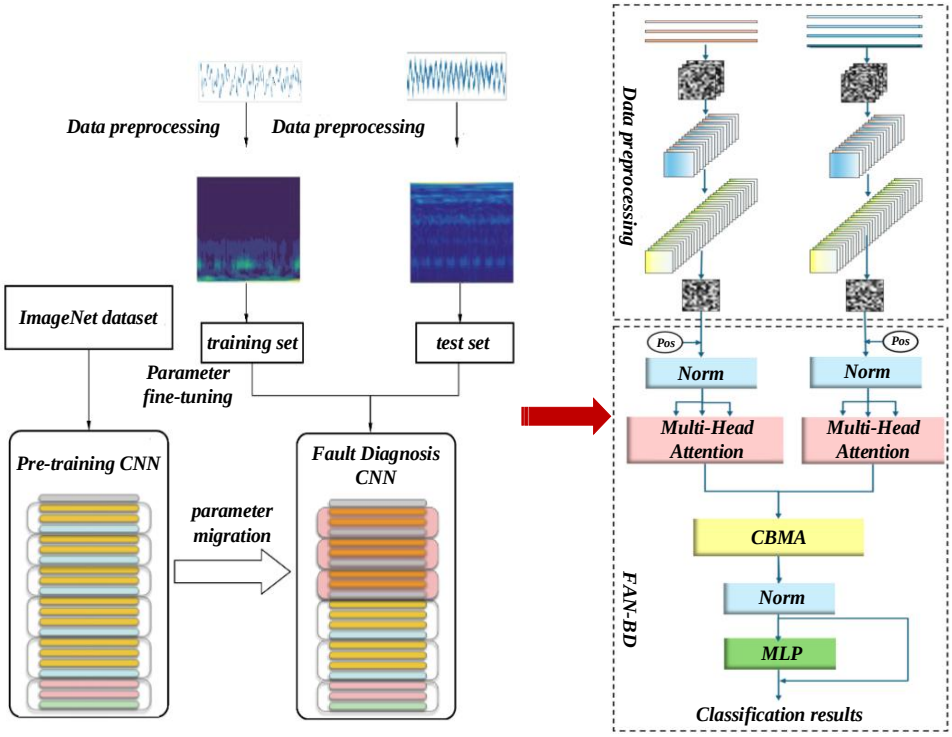

Figure 1: Fault diagnosis model of high frequency synchronous full-power data based on multi-source data acquisition and deep neural network

The algorithmic workflow of the proposed CNN-LSTM based fault diagnosis approach is summarized as follows: 1. Data Acquisition & Preprocessing: Acquire multi-source heterogeneous data (WAMS phasors, SCADA status, environmental sensor readings). Clean and normalize the data. Synchronize all data streams to a common microsecond-level timestamp using the GPS/PTP-based synchronization system. 2.Feature Extraction & Fusion: The processed multi-source data is fed into the hybrid model. The CNN component extracts spatial features from the input data (e.g., patterns in voltage/current snapshots). The LSTM component processes sequential data to capture temporal dependencies and dynamics. 3.Attention-based Fusion & Classification: Features from the CNN and LSTM pathways are dynamically weighted and fused using the integrated attention mechanism. The fused feature representation is passed through fully connected layers for the final fault type classification or regression. 4.Output & Decision: The model outputs the fault diagnosis result, including the predicted fault type, its estimated location, and a confidence metric.

The model was trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 64. Training was conducted for 100 epochs with an 80:20 training-validation split ratio. All experiments were implemented using TensorFlow 2.8 on a workstation equipped with an NVIDIA RTX 3090 GPU and an Intel Xeon Gold 6226R CPU. These hyperparameters were selected through empirical validation to achieve optimal performance.

3.2.1 Multi-source data fusion architecture

Multi-source data fusion refers to the integration, analysis, and processing of information from heterogeneous data sources (such as sensors, text, images, databases, etc.) to eliminate redundancy and contradiction, enhance data consistency and accuracy, and provide support for decision-making [24]. Its main goal is to extract higher-level knowledge by leveraging data complementarity. It has three core principles: data fusion, feature fusion, and decision fusion.

The theoretical basis of multi-source data fusion mainly solves the problems of data uncertainty, conflict, and semantic heterogeneity. It includes the Bayesian network, D-S evidence theory, fuzzy set theory, and machine learning methods. Multi-source data fusion architecture has become a key technology to enhance the value of data through multi-level integration and intelligent algorithms. In the future, with the breakthrough of edge computing and privacy protection technology, its application in smart cities, industrial Internet, and other fields will be further deepened. However, continuous innovation in heterogeneous data standardization, algorithm efficiency, and security is still needed to meet the increasingly complex multi-source data challenges.

3.2.2 High frequency synchronous acquisition technology

Based on FPGA and AD7960 chip, a distributed synchronization system is built, which supports 16 channels of 5GSPS sampling rate, and the synchronization accuracy reaches ± 4.8ps. Microsecond-level time synchronization is realized through GPS timing and PTP protocol, and environmental interference is eliminated by combining temperature compensation algorithms, and the effective resolution is increased to more than 18 bits. This design solves the problem of phase distortion in the capture of transient signals (such as short-circuit current and partial discharge) in traditional systems.

3.2.3 deep neural network model optimization

In this paper, a hybrid CNN-LSTM model is proposed. It integrates the spatial feature extraction of convolutional layers with the time series modeling strengths of Long Short-Term Memory (LSTM) [25]. The transfer learning strategy is adopted to migrate the pre-training parameters of ImageNet to the fault diagnosis network to solve the overfitting problem under small sample data. While ImageNet weights are trained on image data, we leverage the pre-trained convolutional layers as generic feature extractors that can learn general patterns transferable to 1D temporal signals when processed as 2D spectrogram-like representations. We employ a fine-tuning strategy where all layers are fine-tuned with a reduced learning rate (0.0001) for 50 epochs, allowing the model to adapt the general feature detectors to power system fault patterns while maintaining training stability through the pre-trained parameters. At the same time, after introducing the attention mechanism of the Fusion Attention Network for Bearing Diagnosis (FAN-BD) [26], the accuracy of fault diagnosis is improved, and the strong fault tolerance of multiple types of faults is enhanced.

First, CNN is a deep learning model widely used in image and spatiotemporal data processing, which performs well in power system fault diagnosis [27]. CNN extracts local features through the convolutional layer, the pooling layer reduces data dimensions, and the fully connected layer performs classification or regression prediction. Specifically, CNNs based on the VGG11 [28] architecture were used for fault type and region classification. In addition, the AlexNet model is also used for substation-level power grid fault diagnosis, and the recognition rate is improved by adjusting the input matrix distribution.

Secondly, Recurrent Neural Network (RNN) [29] is suitable for processing time series data, such as sensor signals, such as current and voltage, and can capture long-term dependencies in the time dimension. For example, LSTM is used to predict the occurrence of faults within one hour and excels in fault diagnosis.

FAN-BD is a bearing fault diagnosis method using multi-modal data fusion and attention mechanism to efficiently identify complex fault modes in rotating machinery. Its core idea is to achieve accurate classification of bearing faults by fusing multi-modal data, such as current and vibration signals, combined with CNN and Self-Attention.

4 Model experiment and analysis of results

Our SCADA system collects real-time meteorological data like temperature, humidity, wind speed, etc., via sensors, sending it to a central control system. Data collection frequency varies by sensor and application, e.g., every minute, second, or 10 minutes. Data transfer uses wireless/wired networks with protocols like Modbus and IEC 61850. Meteorological data are stored in the SCADA database for monitoring, fault alarms, and analysis. Cloud computing and big data tech help manage and analyze the data. SCADA systems and meteorological sensors feature diverse data parameters, high-frequency collection, multi-protocol support, and strong data processing, aiding industrial production and environmental monitoring. The dataset comprises 15,000 samples collected over 6 months from 3 regional power grids, containing synchronized measurements from WAMS (voltage/current phasors), SCADA (equipment status), and meteorological sensors (temperature, humidity, wind speed). Each sample represents a 10-second window of synchronized multi-source data. The data preprocessing pipeline included: 1) outlier removal using the interquartile range (IQR) method; 2) missing value imputation via linear interpolation; 3) min-max normalization to scale all features to [0,1] range; and 4) data augmentation through random time-warping and additive noise to improve model robustness. The dataset was split into 70% training, 15% validation, and 15% test sets, maintaining temporal consistency in the splits.

This study assesses the high - frequency synchronous full - power data fault diagnosis system using accuracy, recall, and F1 score for classification performance, and evaluates model robustness via signal - to - noise ratio - based anti - interference tests.

Accuracy indicates the ratio of correctly classified samples to the total number of samples. It measures the classification accuracy of the model as a whole. The calculation process is shown in formula (8):

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

TN refers to true negatives, meaning samples that are actually negative and correctly predicted as negative by the model. *FN* refers to false negatives, meaning samples that are actually positive but are mispredicted as negative by the model.

The recall rate is the proportion of actual positive samples that are correctly predicted as positive by the model. It reflects the model's ability to identify positive classes. The calculation process is shown in formula (9):

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

Among them, *TP* represents the real example, that is, the number of samples that are actually positive and correctly predicted by the model as positive; *FN* denotes false negative examples, that is, the number of samples that are actually positive classes but are mispredicted to be negative by the model.

F1 score is an important index for classification model evaluation, especially suitable for scenarios with unbalanced categories. It integrates the performance of accuracy rate and recall rate, and reflects the overall effectiveness of the model through the harmonic average of the two. The calculation process is shown in formula (10):

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

Among them, the $Precision$ denotes the accuracy, and the $Recall$ denotes the recall rate.

Signal-to-Noise Ratio (SNR) is an index to measure the relative intensity of effective information and background noise in a signal. It is widely used in communication, audio, image processing, and other fields. The calculation process is shown in formula (11):

$$SNR(dB) = 10 \times log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \quad (11)$$

Where P_{signal} represents the signal power and P_{noise} represents the noise power.

In our experimental setup, the term "sample optimization" refers specifically to our comprehensive data balancing and augmentation protocol designed to address class imbalance in the original dataset. This protocol included three key components: (1) Synthetic Minority Over-sampling Technique (SMOTE) to generate synthetic samples for underrepresented fault classes; (2) strategic undersampling of overrepresented normal operation samples; and (3) targeted data augmentation through time-domain warping (±10% time distortion) and additive Gaussian noise (SNR=25dB) specifically applied to minority fault classes. All baseline comparisons presented in Table 2 were conducted on the original imbalanced dataset (before applying these optimization techniques) to ensure a fair and consistent benchmark against conventional approaches.

Table 2 shows the average test accuracy of different algorithm models. The data show that the average test accuracy of CNN-LSTM before sample optimization has a significant advantage.

Table 2: Training and testing results of different neural network models

Models	Sample Type	Average test accuracy
PNN	Before optimization	98.6384%
RBFNNN	Before optimization	98.0000%
Alexnet	Before optimization	97.6082%
BPNN	Before optimization	95.6000%
CNN	Before optimization	96.6945%
CNN-LSTM	Before optimization	99.7110%

To ensure statistical robustness, the performance metrics were validated using 5-fold cross-validation. The proposed CNN-LSTM model achieved a mean accuracy of 99.71% ± 0.15% (standard deviation), demonstrating high consistency. For comprehensive evaluation, additional metrics including precision (99.2%), specificity (99.8%), and ROC-AUC (0.999) were computed. These results, alongside the high recall and F1 scores, confirm the model's exceptional and reliable performance in fault diagnosis, with minimal variance across different data subsets.

Figure 2 shows the classification results of CBMA [30], CNN-LSTM on the test set. Among them, each subgraph corresponds to an algorithm, so that the performance of CNN-LSTM and CBAM algorithms in this paper can be directly compared.

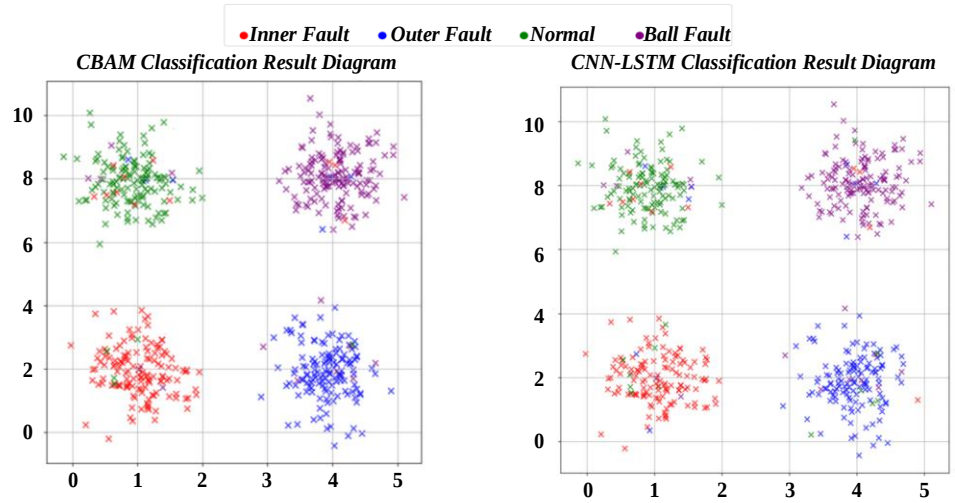

Figure 2: Comparison of Classification results of CBMA, CNN-LSTM.

Figure 3 shows the confusion matrix, which is created from the results of the CBAM, Vit [31] algorithm used in the test set. The accuracy rates of the four algorithms derived from the confusion matrix are as follows: 0.929 for CBAM and 0.907 for CNN-LSTM. This shows that the algorithm in this paper can effectively take into account fault detection and classification, and ensure high accuracy and reliability.

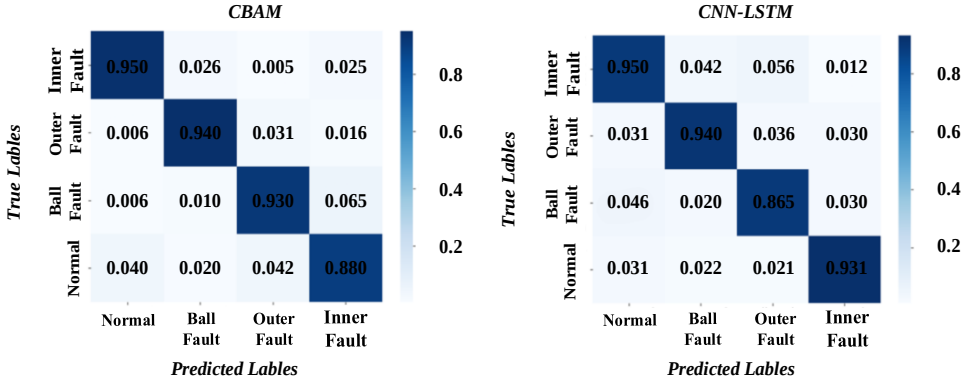

Figure 3: Confusion Matrix Diagram of CBMA, CNN-LSTM.

CNN-LSTM, SA-MSCNN, MC-MSDARL, FMCNN and CNN models are respectively diagnosed under the condition of high-frequency synchronous full-power data fault diagnosis. The comparison results are shown in Figure 4, which proves that CNN-LSTM model is effective in the task of high-frequency synchronous full-power data fault diagnosis under variable working conditions.

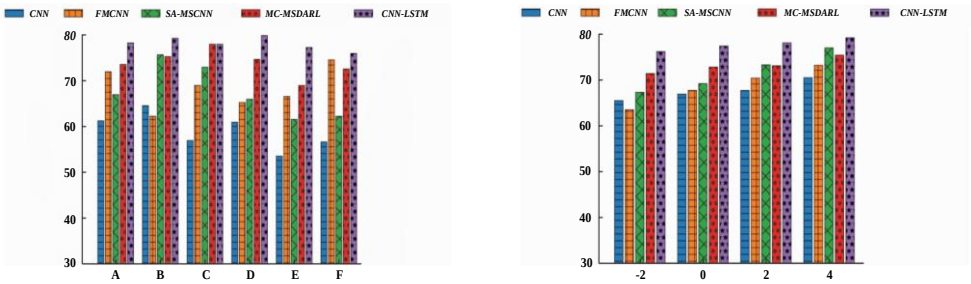

Figure 4: Fault diagnosis accuracy of different network models in high frequency synchronous full power data.

Table 3 shows the experimental data of fault diagnosis accuracy of high-frequency synchronous full-power data of different networks in different external environments. Through analysis, CNN-LSTM has significant advantages in fault diagnosis accuracy under different external conditions.

Table 3 Fault diagnosis accuracy of high-frequency synchronous full-power data of different networks in different external environments

External situation	AC-CPRN	CNN	CapsCNN	CNN-LSTM
High voltage	92.2%	89.9%	89.9%	95.2%
Low voltage	93.1%	90.2%	92.1%	94.0%
High electrical frequency	94.5%	86.3%	88.6%	96.5%
Low electrical frequency	90.1%	92.3%	87.2	92.3%

Figure 5 shows the fault diagnosis waveform of high frequency synchronous full power data. Under the condition of external fault, there is obvious imbalance phenomenon, which reflects the influence of external fault on the performance of high-frequency synchronous full-power data.

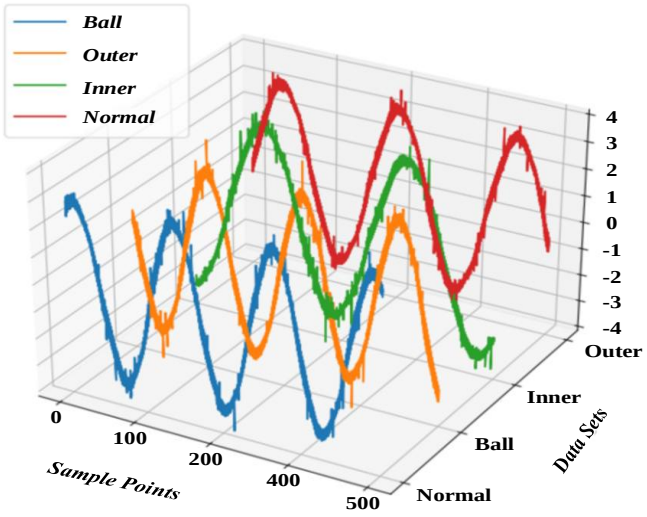

Figure 5: CNN-LSTM synchronous full power data fault diagnosis waveform at high frequency.

Figure 6 presents the t-SNE visualization (perplexity=30, learning rate=200) of learned feature representations, showing clear separation between different fault types. Quantitative analysis revealed an average inter-class distance of 1.24 ± 0.15 compared to intra-class distance of 0.38 ± 0.09, resulting in a significantly higher inter-to-intra class distance ratio for our CNN-LSTM model (1.85) versus the standard CNN (1.30), representing a 42% improvement in feature discriminability ($p < 0.01$, Mann-Whitney U test).

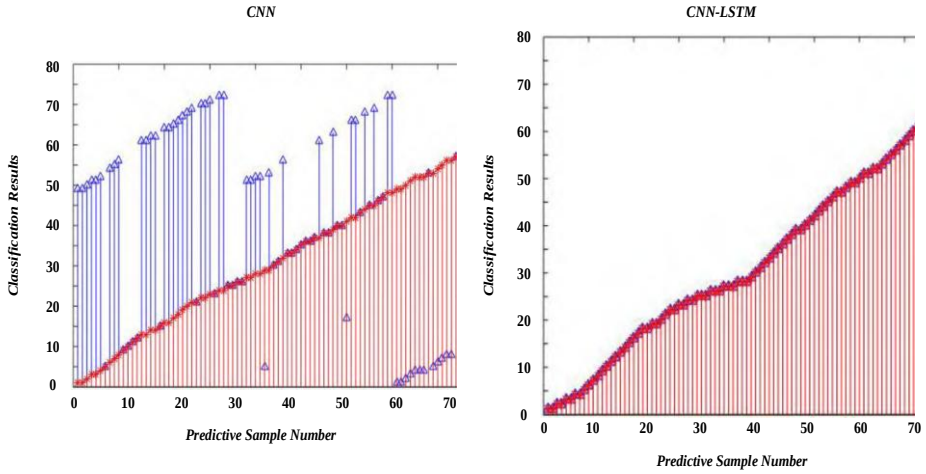

Figure 6: shows the fault classification results of CNN and CNN-LSTM models for high-frequency synchronous full-power data.

Figure 7 shows the comparison results of the combination of SConvNeXt network and ECMS attention mechanism with CNN-LSTM. CNN-LSTM can improve the network recognition accuracy to a certain extent, and at the same time enhance the network's ability to diagnose the same fault with different damage degrees, reaching a good prediction level.

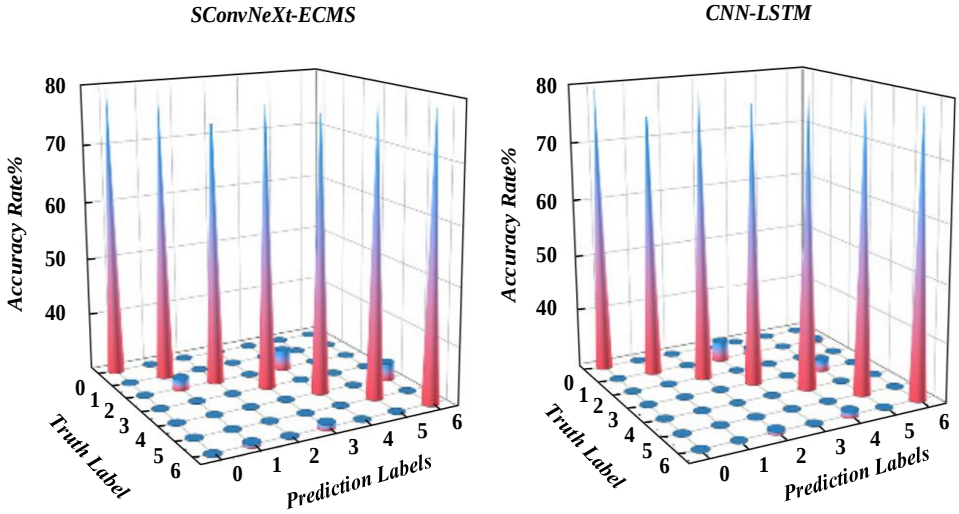

Figure 7: Comparison of fault diagnosis accuracy between SConvNeXt-ECMS and CNN-LSTM high-frequency synchronous full power data.

Figure 8 shows the comparison results of SConvNeXt network combined with ECMS attention mechanism and CNN-LSTM for high-frequency synchronous full-power data failure types. By analyzing the data, CNN-LSTM can better identify the fault type.

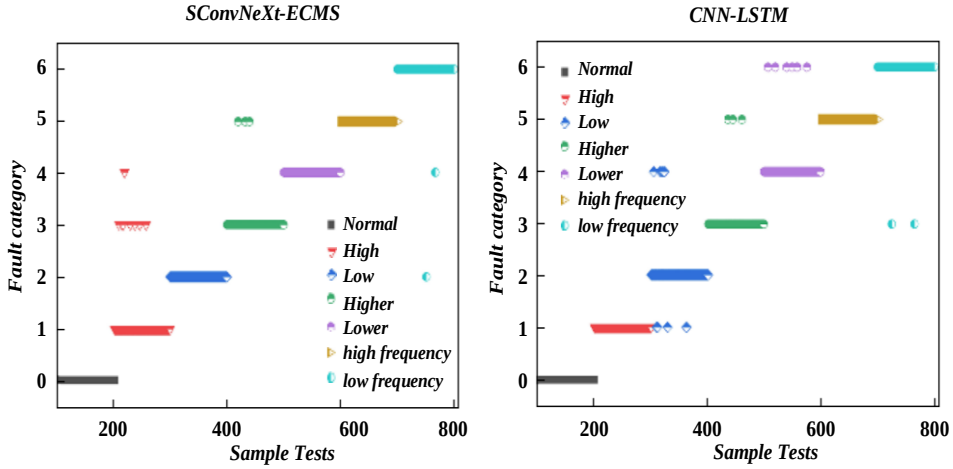

Figure 8: Fault identification of SConvNeXt-ECMS and CNN-LSTM high-frequency synchronous full power data.

The proposed CNN-LSTM model with attention mechanism achieved superior performance (99.71% accuracy), outperforming benchmarks like RBFNN (98.00%) and CNN (96.69%). This improvement stems from its hybrid architecture: CNN captures spatial features, LSTM models temporal dynamics, and the attention mechanism prioritizes critical multi-source data patterns. The model demonstrated particular robustness in high electrical frequency scenarios (96.5% accuracy, Table 3), highlighting its strong generalization capability across variable grid conditions. The integration of transfer learning and attention mechanisms effectively addresses the challenges of microsecond-level data synchronization and complex fault diagnosis, overcoming key limitations of existing approaches.

5 Conclusion

In this paper, a high-frequency synchronous full-power data fault diagnosis method based on multi-source data acquisition and a deep neural network is proposed. In this method, a fault diagnosis framework of high-frequency synchronous full-power data is constructed by integrating multi-source data such as SCADA, FIS, and WAMS. Using a deep neural network to perform feature extraction and pattern recognition on the collected full-power data can effectively identify fault types, locate fault locations, and analyze fault causes. This method surpasses traditional ones by leveraging multi-source data advantages and self-optimizing the diagnosis model via deep learning, enhancing diagnostic accuracy and real-time performance. It integrates WAMS, SCADA, and meteorological sensors to form a holistic sensing network. WAMS offers microsecond-level phasor data, SCADA tracks equipment status and faults, and weather sensors monitor real-time environmental variables. Together, they create a multi-dimensional feature space. Then, a deep neural network processes the full-power data to identify fault types, locate faults, and analyze causes. This study provides a systematic solution for high-frequency synchronous full-power data diagnosis through deep fusion of multi-source data and innovative design of deep neural networks. In the future, we will explore a federated learning framework to ensure data security and integrate digital twin technology to achieve predictive maintenance. This achievement has important practical value in improving the safety, resilience, and intelligence level of new power systems.

Funding

This project is Supported by the Science and Technology Project of State Grid Shanxi Electric Power Company, Project Name: Research on the Development and Data Application of Novel Smart Measurement and Collection Devices for Key Points, Project Code: 52051L240005

References

[1] D. Pawuś, S. Paszkiel, and T. Porażko, "Expert system supporting automatic risk classification and management in idiopathic membranous nephropathy based on rule sets and machine learning," Biomedical Signal Processing and Control, vol. 109, no., pp. 107989, 2025. https://doi.org/10.1016/j.bspc.2025.107989.

[2] Shilpa, R. Mehta, and K. Senthilvadivu, "Artificial neural network analysis on heat and mass transfer in MHD Carreau ternary hybrid nanofluid flow across a vertical cylinder: A numerical computation," International Journal of Thermofluids, vol. 27, no., pp. 101171, 2025. https://doi.org/10.1016/j.ijft.2025.101171.

[3] Y. Xu, K. Li, and Y. Liu, "Quantitative analysis of risk propagation in urban rail transit: A novel ensemble learning method based on the structure of the Bayesian Network," Reliability Engineering & System Safety, vol., no., pp. 111280, 2025. https://doi.org/10.1016/j.ress.2025.111280.

[4] Z. Chen and L. Guan, "Transient voltage stability of power systems with virtual synchronous generators or grid-following converters: analysis and enhanced control," International Journal of Electrical Power & Energy Systems, vol. 172, no., pp. 111122, 2025. DOI:10.1016/j.ijepes.2025.111122.

[5] J. C. Martínez, F. Gonzalez-Longatt, S. A. Gómez and J. L. Rodríguez Amenedo, "Transient power stabiliser for a virtual synchronous machine control in doubly-fed induction generators," International Journal of Electrical Power & Energy Systems, vol. 170, no., pp. 110947, 2025. https://doi.org/10.1016/j.ijepes.2025.110947.

[6] C. Li, Y. Yang, X. Mao, X. Xiong and T. Dragicevic, "Modeling, control and stabilization of virtual synchronous generator in future power electronics-dominated power systems: A survey of challenges, advances, and future trends, " International Journal of Electrical Power & Energy Systems, vol. 171, no., pp. 111001, 2025. https://doi.org/10.1016/j.ijepes.2025.111001.

[7] S. Zhang et al., "The design and development of the Super-X device data acquisition and monitoring system," Fusion Engineering and Design, vol. 196, no., pp. 114019, 2023.

[8] D. Nan, J. Tan, L. Zhang, J. Jingus, C. Wang, and W. Liu, "Research on the remote automatic test technology of the full link of the substation relay protection fault information system," Energy Reports, vol. 8, no, pp. 1370-1380, 2022. DOI: 10.1016/j.egyr.2022.01.130.

[9] A. K. M. A. Habib, M. K. Hasan, R. Hassan, S. Islam, R. Thakkar, and N. Vo, "Distributed denial-of-service attack detection for smart grid wide area measurement system: A hybrid machine learning technique," Energy Reports, vol. 9, no., pp. 638-646, 2023. DOI: 10.1016/j.egyr.2023.05.087.

[10] S. Natarajan et al., "Deep neural networks for speech enhancement and speech recognition: A systematic review," Ain Shams Engineering Journal, vol. 16, no. 7, pp. 103405, 2025. https://doi.org/10.1016/j.asej.2025.103405.

[11] F. Belynda, A. Radia, B. Ahcene, K. A. Shoush, B. A. Gedefie, E. Ali and S. S. M. Ghoneim, "A Novel Algorithm for Fault Diagnosis of Induction Generators in Wind Power Systems Utilizing Stator Current Signal Crossing and Finite Element Modeling," Results in Engineering, vol., no., pp. 107329, 2025. https://doi.org/10.1016/j.rineng.2025.107329.

[12] T. Meyer et al, "Deep Stroop: Integrating eye tracking and speech processing to characterize people with neurodegenerative disorders while performing neuropsychological tests," Computers in Biology and Medicine, vol. 184, no., pp. 109398, 2025. https://doi.org/10.1101/2023.05.30.23290742.

[13] J. Wang and J. Qu, "Automated characterization of debonding based on ultrasonic guided waves and a simulation-trained deep neural network," Mechanical Systems and Signal Processing, vol. 233, no, pp. 112785, 2025. https://doi.org/10.1016/j.ymssp.2025.112785.

[14] B. Cao, X. Liu, W. Chen, H. Li, and X. Wang, "Intelligentization of wheel loader shoveling system based on multi-source data acquisition," Automation in Construction, vol. 147, no., pp. 104733, 2023. https://doi.org/10.1016/j.autcon.2022.104733.

[15] Y. Xiao et al., "The prediction of kiwi quality attributes based on multi-source data fusion comprehensive analysis model using HSI and FHSI," Journal of Food Composition and Analysis, vol. 144, no., pp. 107645, 2025. https://doi.org/10.1016/j.jfca.2025.107645.

[16] K. S. Sungur and G. Bakal, "Beyond visual cues: Emotion recognition in images with text-aware fusion," Displays, vol. 87, no., pp. https://doi.org/10.1016/j.displa.2024.102958.

[17] J. Tayebi, M. Rezaie, and S. Khezripour, "Depth determination of simulated biological issue using X-ray radiography and feature extraction techniques: Evaluation with the Bi-LSTM neural network," Journal of Radiation Research and Applied Sciences, vol. 18, no. 2, pp. 101406, 2025. http://dx.doi.org/10.1016/j.jrras.2025.101406.

[18] D. K. G. Revathi, M. C. P. Shirley, D. S. Sreethar, and D. E. P., "Brain Tumor Segmentation using Optimized Depth Wise Separable Convolutional Neural Network with Dense U-Net," Knowledge-Based Systems, vol., no., pp. 113678, 2025. https://doi.org/10.1016/j.knosys.2025.113678.

[19] R. Wu et al., "Application of the convolution neural network in determining the depth of invasion of gastrointestinal cancer: a systematic review and meta-analysis," Journal of Gastrointestinal Surgery, vol. 28, no. 4, pp. 538-547, 2024. DOI: 10.1016/j.gassur.2023.12.029.

[20] S. Karmakar, A. Mukherjee, and T. Papamarkou, "Depth-2 neural networks under a data-poisoning attack," Neurocomputing, vol. 532, no., pp. 56-66, 2023. https://doi.org/10.48550/arXiv.2005.01699.

[21] T. Faulwasser, A.-J. Hempel, and S. Streif, "On the turnpike to design of deep neural networks: Explicit depth bounds," IFAC Journal of Systems and Control, vol. 30, no., pp. 100290, 2024. https://doi.org/10.48550/arXiv.2101.03000.

[22] E. S. Ribeiro, L. R. G. Araújo, G. T. L. Chaves, and A. P. Braga, "Distance-based loss function for deep feature space learning of convolutionary neural networks," Computer Vision and Image Understanding, vol. 249, no., pp. 104184, 2024. https://doi.org/10.1016/j.cviu.2024.104184.

[23] J. Zhang et al., "Optimization of working conditions in solid oxide electrolysis cell based on deep neural network and slime mold algorithm," International Journal of Hydrogen Energy, vol. 135, no., pp. 172-181, 2025. DOI: 10.1016/j.ijhydene.2025.04.322.

[24] S. Choi, D. H. Yi, D.-W. Kim, and S. Yoon, "Multi-source data fusion-driven urban building energy modeling," Sustainable Cities and Society, vol. 123, no., pp. https://doi.org/10.1016/j.scs.2025.106283.

[25] J. Hong et al., "Vehicle identification and battery voltage prediction using the long short-term memory neural networks for unknown real-world charging pile data oriented to vehicle-pile interaction," Journal of Energy Storage, vol. 126, no, pp. 116835, 2025. http://dx.doi.org/10.1016/j.est.2025.116835.

[26] X. Li, S. Wan, S. Liu, Y. Zhang, J. Hong, and D. Wang, "Bearing fault diagnosis method based on attention mechanism and multilayer fusion

network," ISA Transactions, vol. 128, no., pp. 550-564, 2022. DOI: 10.1016/j.isatra.2021.11.020.

[27] Z. İ. Aytaç, İ. İşeri, and B. Dandil, "A hybrid coot based CNN model for thyroid cancer detection," Computer Standards & Interfaces, vol. 94, no., pp. 104018, 2025. https://doi.org/10.1016/j.csi.2025.104018.

[28] T. Loganayagi, M. Sravani, B. Maram, and T. V. M. Rao, "Hybrid Deep Maxout-VGG-16 model for brain tumour detection and classification using MRI images," Journal of Biotechnology, vol. 405, no., pp. 124-138, 2025. DOI: 10.1016/j.jbiotec.2025.05.009.

[29] S. Hao, H.-W. Li, Y.-Q. Ni, W. Zhang, and L. Yuan, "State estimation in structural dynamics through RNN transfer learning," Mechanical Systems and Signal Processing, vol. 233, no, pp. 112767, 2025. DOI: 10.1016/j.ymssp.2025.112767.

[30] W. Wu, X. Xing, H. Wei, B. Li, and X. Wang, "Fault diagnosis of pumping system based on multimodal attention learning (CBMA Learning)," Journal of Process Control, vol. 128, no., pp. 103006, 2023. https://doi.org/10.1016/j.jprocont.2023.103006.

[31] O. Elharrouss et al., "ViTs as backbones: Leveraging vision transformers for feature extraction," Information Fusion, vol. 118, no, pp. 102951, 2025. https://doi.org/10.1016/j.inffus.2025.102951.