Research on Power Time Series Fault Detection Based on CNN - LSTM Model

Shengyuan Hao [1,*], Yuxuan Yang [2]

[1] School of Mathematics and Statistics, Hainan University, Haikou, China, 570228

[2] Department of Mathematics and Physics, University College London, London, UK, WC1E 6BT

* Corresponding Author Email: hnuhsy8663@163.com

Abstract. With the accelerated construction of new power systems, the scale and complexity of transmission systems are increasing, while transmission line fault diagnosis algorithms for complex power timing data urgently need to be studied and tested.In this paper, a transmission line fault diagnosis model fusing convolutional neural network (CNN) and long short-term memory network (LSTM) is proposed. The model realizes automatic extraction of local spatio-temporal features of fault signals through convolutional neural network, captures the long-term dependence of power system operation state by using long and short-term memory network, and introduces an attention mechanism to enhance the characterization of key fault features, which effectively improves the identification accuracy of the fault types of the transmission line under the complex working conditions. It adapts to the requirements of smart grid real-time monitoring scenarios on algorithm robustness. Also, it considers the use of the synthetic few over-sampling technique (SMOTE) to realize the data enhancement to improve the imbalance of data distribution. The results show that compared with the traditional connection neural network, back propagation neural network, and LSTM model, the CNN-LSTM model is only 0.0497 in the loss value, which is reduced by about 70%, and the fault identification accuracy and precision rate are more than 97%, which is improved by about 2%. It provides theoretical support for solving the problem of real-time fault diagnosis of large-scale transmission systems, and has significant engineering practice value for reducing the cost of grid operation and maintenance and improving the reliability of power supply.

Keywords: Power System, Line Fault Detection, Deep Learning, CNN-LSTM Model, SMOTE Algorithm.

1. Introduction

With the rapid development of social and economic development and the acceleration of industrialization, the stable operation of the power system has gradually become an important issue related to the livelihood of the country. However, due to various reasons, power system failures still occur from time to time [1], of which transmission line failures are particularly noteworthy. As an important part of the power system, the safety and stability of transmission lines are directly related to the reliability of power supply. But due to multiple factors such as long lines, wide coverage, and complex operating environments, transmission line failures are difficult to avoid completely. Therefore, in-depth study of the causes and laws of transmission line faults is of great practical significance to improve the stability and reliability of the power system.

The traditional fault survey mainly relies on manual inspection. However, it is limited by the visual ability of inspectors and geographic constraints, which often leads to problems such as misdetection and omission. So, it is difficult to meet the needs of smart grid development [2]. With the development of artificial intelligence, machine learning-based target detection techniques show significant potential in transmission line fault detection. Jie Yang et al[3]constructed the set of feature vectors to determine the type of line faults by performing phase-to-mode conversion of three-phase current fault components and calculating the Euclidean distance and Pearson's correlation coefficient between the two modes of currents. Wanyu Shi [4] used adaptive t-distribution to improve the sparrow search algorithm to optimize the support vector machine (SVM) parameters, screen key features and build a fault diagnosis model. Majid Jamil et al [5] carried out fault analysis using feed forward neural network and back propagation algorithm based on artificial neural network (ANN). Lin Jikang et al

[6] constructed a deep learning neural network (DNN) based on logistic regression to establish regression, and utilized the BP method to complete the model training. Fezan Rafique et al [7]constructed an “end-to-end” model, which does not require a manually designed feature extraction process, and directly processes operational data through Long Short-Term Memory (LSTM) units and realizes fault detection with the help of Recurrent Neural Networks (RNN), which has the advantages of high convergence accuracy and good robustness. The advantages are high convergence accuracy, good robustness and so on.

In the current fault detection research, the model has significant limitations in feature extraction, long-range dependence modeling and key factor identification, and it is difficult to effectively deal with the higher-order nonlinearities, long-range dependence, and complex coupling characteristics of power timing data, and lacks a systematic model comparison and applicability exploration for complex power timing data. To this end, this paper proposes a fault detection model that integrates CNN, GRU and attention mechanism, with the help of CNN to extract local features, GRU to memorize the temporal dynamics, and attention mechanism to focus on the key information, and at the same time introduces an unbalanced data testing mechanism to effectively solve the problem of unbalanced data samples, which provides theoretical support and a basis for model selection for the modeling of complex features and fault detection of transmission lines, and helps to promote the development of engineering deployment of new fault detection methods in power grids. The technology roadmap for this paper is shown in Figure 1 below.

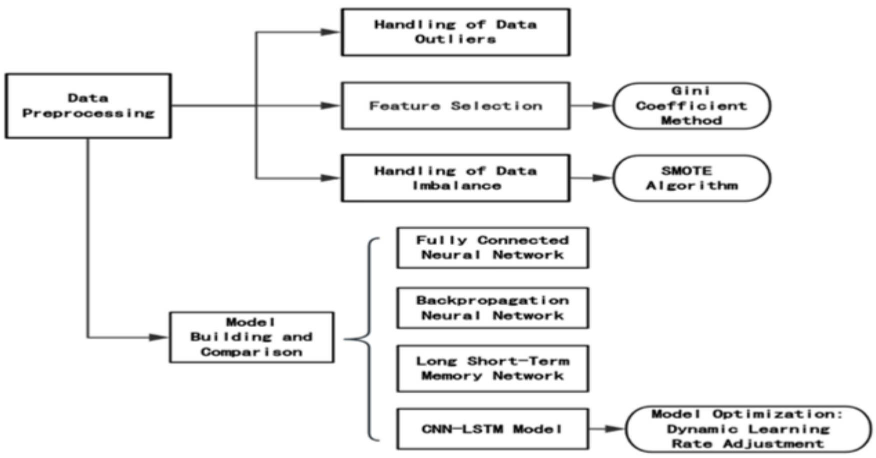


Figure 1. Technology roadmap.

2. Principles of the algorithm

2.1. Fully connected neural network

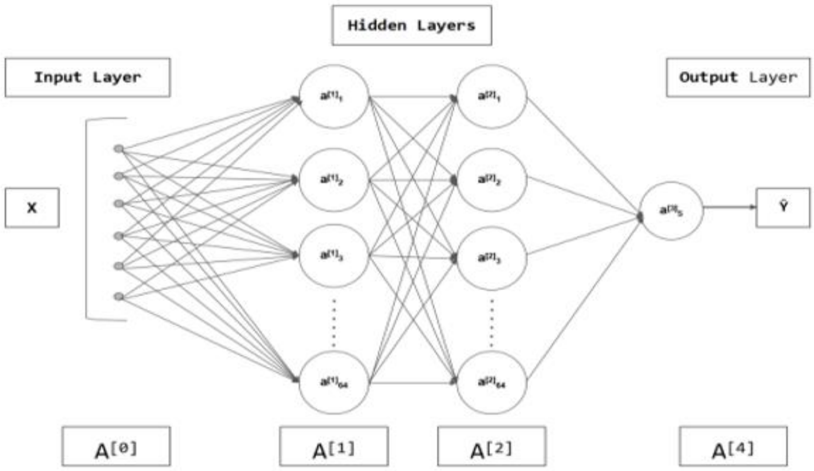
FCNN as a benchmark model [8] uses a three-layer fully-connected layer structure with input layer dimensions consistent with the number of features, a hidden layer using ReLU activation function, and an output layer that realizes multi-classification through linear transformations. The structure is shown in Figure 2(a).

2.2. Back propagation neural network

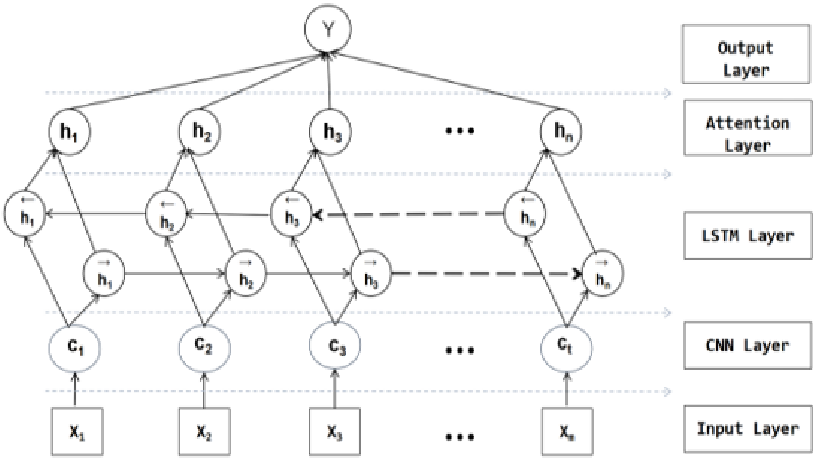
The BP model is similar in structure to FCNN[9], and both belong to the traditional feed-forward neural network, relying on the back-propagation algorithm to optimize the parameters. The difference is that the BP model emphasizes more on the gradient back propagation mechanism in the training process. The structure is shown in Figure 2(a).

2.3. Convolutional long and short-term memory network

The CNN-LSTM model uses a one-dimensional convolutional neural network (1D-CNN) as the basic feature extraction module to give full play to its efficient processing advantages for time-series data [10]. The convolutional kernel of 1D-CNN slides along the time dimension to accurately capture local features in transmission line fault data, which can effectively circumvent the computational redundancy problem due to dimensional expansion compared to two-dimensional convolutional neural network (2D-CNN). The model firstly adapts the dimension of the input data to match the 1D-CNN convolutional layer, and then extracts the local spatio-temporal features of the data; then mines the long-term dependency of the sequences through the LSTM layer to obtain the global features; then enhances the weight of the key time steps by using the multi-attention mechanism, and highlights the important features; and finally maps the features into the classification space with the help of the fully connected layer to achieve the accurate identification and classification of the transmission line faults. accurate identification and classification. The structure is shown in Figure 2(b).



(a) Schematic diagram of FCNN and BP model structure.



(b) Schematic diagram of CNN-LSTM model structure.

Figure 2. Schematic diagram of the model structure.

2.4. Principles of the SMOTE algorithm

Transmission lines have a low probability of fault occurrence in engineering practice, and the distribution of fault types is extremely uneven. In this paper, the unbalanced data testing mechanism is introduced in the model evaluation, and the SMOTE algorithm is utilized to realize oversampling, which enhances the adaptability and robustness of the model in real application scenarios, and avoids the high-precision artifacts derived from relying on the ideal balanced datasets only, and the specific process is as follows.

For each minority class sample, calculate its k nearest neighbors in the feature space. The Euclidean distance is usually used to measure the similarity between the samples. For two points $X = (x_1, x_2, ..., x_n)$ $Y= (y_1, y_2, ..., y_n)$ in an n-dimensional space, the Euclidean distance is defined as:

$$d(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \quad (1)$$

A new synthetic sample is generated randomly by choosing a nearest neighbor from the k nearest neighbors of each minority class sample and generating a new synthetic sample randomly on the connecting line between that minority class sample and its selected nearest neighbor in the following manner:

$$X_{new}=X_i+\beta\times(X_j-X_i) \quad (2)$$

Where X_{new} is a newly generated synthetic sample, X_i is the original minority class sample, X_j is a nearest neighbor randomly selected from the k nearest neighbors of X_i, and β is a random number between 0 and 1. By varying the value of β, samples can be generated at different locations on the connecting line between X_i and X_j. The above steps are repeated until a sufficient number of synthetic samples are generated such that the number of minority class samples increases to a similar number as the number of majority class samples.

3. Case analysis

3.1. Data presentation

The dataset used in this paper is obtained from the Kaggle platform (http://www.kaggle.com), and the simulation model of the three-phase transmission line system is constructed based on Simulink for MATLAB and its SimPowerSystems toolbox. The power system consists of four generators with a rated voltage of 11×10^3 V, which are distributed at both ends of the transmission line and arranged in pairs at each end. The simulation scenarios cover both normal operation and various fault conditions, and the line voltage and current information at the output of the system is recorded.

The datasets contain seven features: I_a, I_b, I_c, V_a, V_b, V_c and Fault_Type, where I_a, I_b, I_c denote three-phase currents, V_a, V_b, V_c denote three-phase voltages and Fault_Type denotes the fault type. The fault categories involved include: No Fault, Line A to Ground Fault, Line B to Line C Fault, Line A Line B to Ground Fault and Line A Line B Line C Fault 5 categories. The total number of samples in the datasets is 7861 and some examples are shown in Table 1.

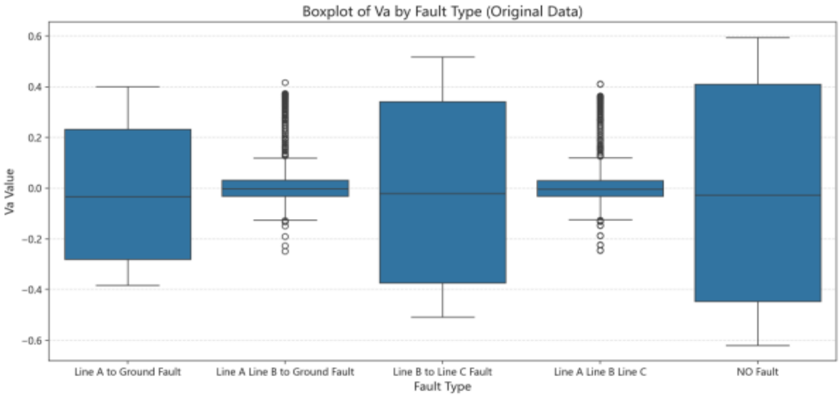
Table 1. Datasets presentation.

ID	I_a	I_b	I_c	V_a	V_b	V_c	Fault_Type
1	-151.29	-9.67	85.80	0.40	-0.13	-151.29	Line A to Ground Fault
2	-336.18	-76.28	18.32	0.31	-0.12	-336.18	Line A to Ground Fault
3	-502.89	-174.64	-80.92	0.26	-0.11	-502.89	Line A to Ground Fault
...	...	...	...	...	...	...	...
7859	-65.44	36.47	26.10	0.11	-0.55	0.44	No Fault
7860	-65.02	35.47	26.68	0.12	-0.56	0.43	No Fault
7861	-64.59	34.48	27.25	0.13	-0.56	0.43	No Fault

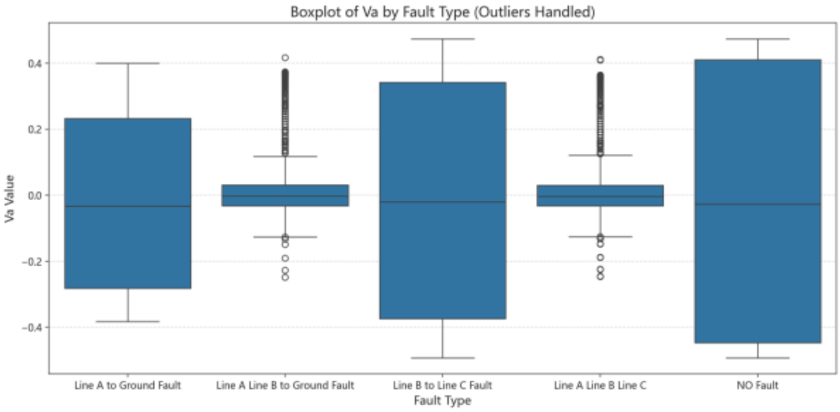
3.2. Feature engineering

3.2.1. Outlier handling

In this paper, observations using box plots revealed the presence of certain outliers. Therefore, the quartile method is used to identify and process the outliers in the datasets [11]. Firstly, the interquartile spacing IQR, which is the difference between the third quartile Q_3 and the first quartile Q_1, is calculated to determine the boundaries for the determination of the outliers: the lower limit is Q_1-1.5*IQR, and the upper limit is Q_3-1.5*IQR. as shown in Figure 3(a), the outliers that significantly deviate from this range are present in the original data. In order to balance the data quality, the truncation method is used to deal with the outliers, and all the values smaller than the lower limit are adjusted to the lower limit value, and the values larger than the upper limit are adjusted to the upper limit value. The results of the adjustment are shown in Figure 3(b) for the V_a feature.



(a) Box plots of V_a outliers before treatment.



(b) Box plots of V_a outliers after treatment.

Figure 3. Comparison of box plots before and after treatment of V_a outliers

3.2.2. Feature selection

Deep learning often faces the problem of overfitting, so eliminating irrelevant or redundant features and reducing the number of features can improve model accuracy and reduce runtime. Feature selection is mainly categorized into three main groups: Filter, Wrapper, and Embedded [12]. The embedded method which is suitable for the real-time diagnostic task of the industrial fault datacan avoid the defects of the filtering method's disconnection between evaluation and application and the complexity of the packing method's computation because it can deeply couple the feature selection in model training, adaptively capture the nonlinear association between features and fault types, and possess the automated filtering mechanism without the need to manually adjust the parameter. Therefore, this paper adopts the Random Forest algorithm derived from the Random Forest algorithm, which is based on the guinness impurity of the node splitting of the decision tree. Reduction quantifies the contribution of the feature to fault classification, the higher the score indicates that the feature has a more significant impact on fault type identification. The relevant comparison results can be seen in Figure 4

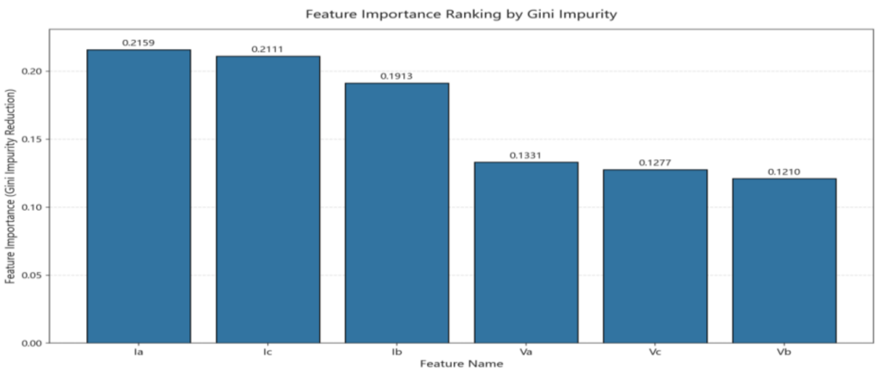
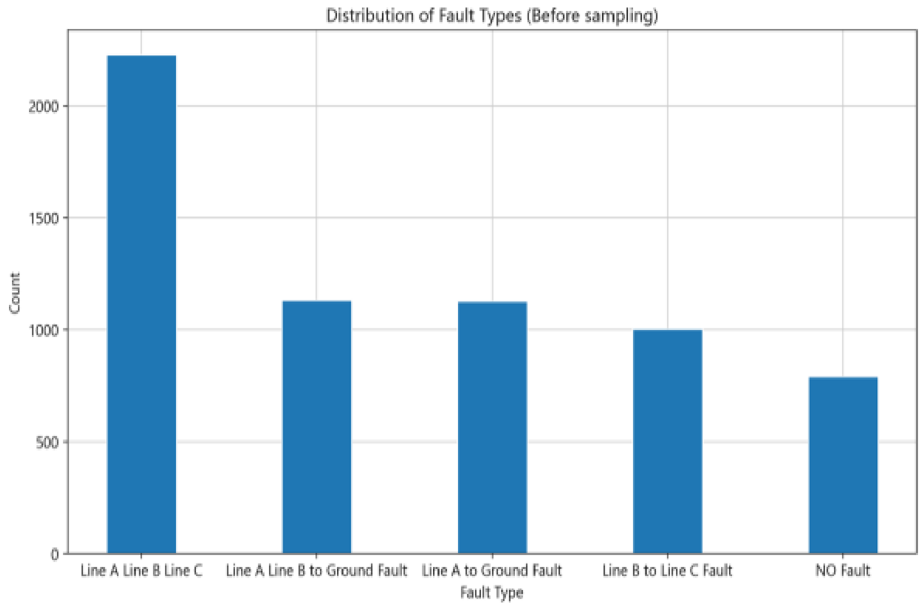


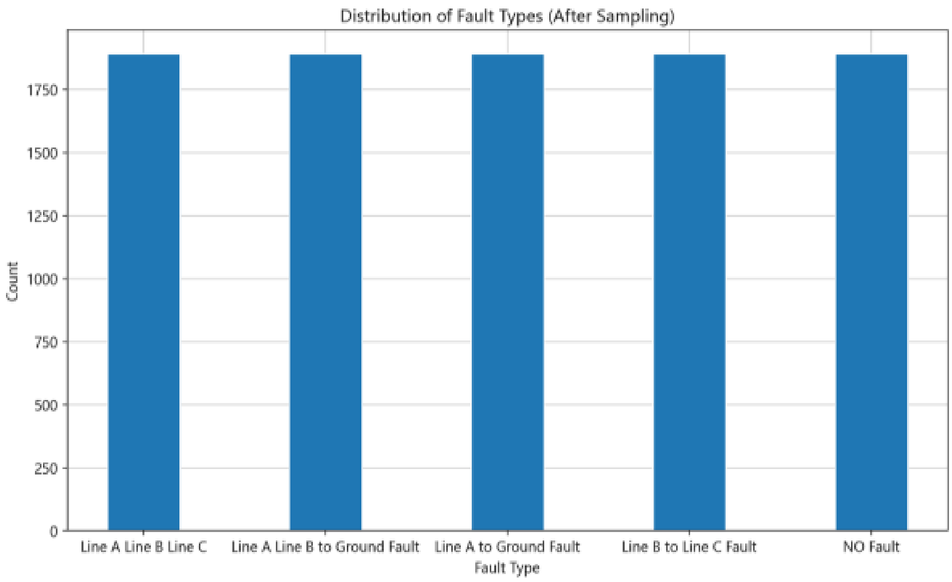
Figure 4. Feature importance ranking.

3.2.3. Data imbalance treatment

In this paper, the first 80% data of the datasets is selected as the training set, and the second 20% data is selected as the test set. Statistically, the number of samples of each category in the training set is shown in Figure 5(a) below, and there is an imbalance of samples for the classification problem. The traditional sampling methods for dealing with unbalanced datasetss are mainly random undersampling and random oversampling. Random undersampling refers to randomly deleting samples from the majority class samples, but this method is prone to lose important information, leading to a decline in classifier performance. Random oversampling refers to adding new samples by randomly sampling from the samples of a few classes, but it increases the possibility of model overfitting [13]. Compared to the traditional random oversampling method, the SMOTE method balances the distribution of samples from each class by adding artificially synthesized samples from a few classes to the data, which reduces the possibility of overfitting and improves the accuracy of the prediction of the few classes[14]. The distribution of each fault category before and after using the SMOTE algorithm in the training set is shown in Figure 5 below.



(a) Before SMOTE treatment.



(b) After SMOTE treatment

Figure 5. Before and after treatment using the SMOTE method.

3.3. Modeling

3.3.1. Parameter setting and training process

The four types of models, namely, fully connected neural network, BP neural network, LSTM, and CNN-LSTM, all use the cross-entropy loss function and Adam optimizer, with a set learning rate of 0.001 and a training period of 200 epochs. In the data preprocessing stage, the input features are standardized and then divided into a training set and a test set, and the labels are processed by unique thermal encoding. During the training process, the loss value, accuracy, precision, recall and F_1 score are recorded for model performance comparison and convergence analysis.

The detailed parameters of one of the CNN-LSTM models are shown in Table 2 below:

Table 2. Table of CNN-LSTM model parameters.

Parameter	Description	Value
Conv1	First convolutional layer with 3x1 kernel and padding	Input channel: 1 Output channels: 32
Conv2	Second convolutional layer with 2x1 kernel and padding	Input channels: 32 Output channels: 32
ReLU	Activation function applied after convolution layers	ReLU activation
LSTM	LSTM layer with an input size based on the flattened convolution output	Input size: 32 Hidden size: 64
Attention	Multihead attention mechanism with 4 attention heads	Number of heads: 4 Hidden size: 64
Fully Connected Layer	Final linear transformation to the output dimension	Output dimension: 5

3.3.2. Comparison of model evaluation indicators

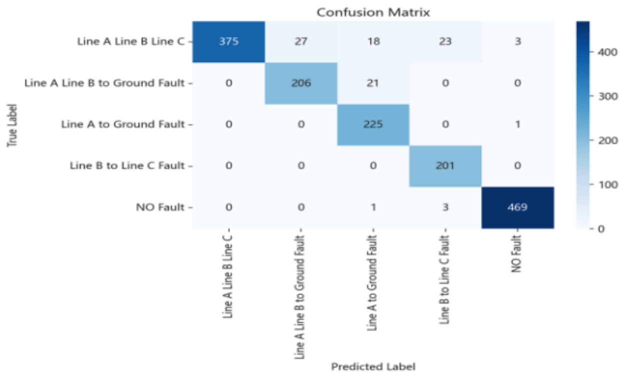
In this paper, the models are comprehensively evaluated from multiple dimensions and perspectives, including loss value, accuracy, precision, recall and F_1 value, and running time, and the comparison of the evaluation indexes of each model is shown in Table 3 below:

Table 3. Comparison of evaluation indicators of each model.

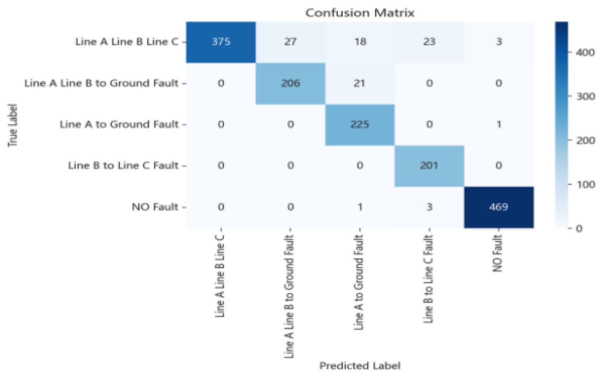
	FCNN	BP	LSTM	CNN - LSTM
Train Loss	0.3935	0.2034	1.0573	0.0497
Val Loss	0.2674	0.2200	1.0654	0.0627
Train Accuracy	85.72%	96.82%	89.60%	98.45%
Val Accuracy	93.83%	95.49%	88.30%	97.71%
Train Precision	85.86%	96.91%	90.20%	98.48%
Val Precision	94.44%	95.70%	89.84%	97.81%
Train Recall	85.72%	96.82%	89.60%	98.45%
Val Recall	93.83%	95.49%	88.30%	97.71%
Train F_1	85.71%	96.81%	89.36%	98.44%
Val F_1	93.81%	95.50%	88.21%	97.73%
Train Time(s)	3.6342	2.6204	19.8580	21.1543

Observing the above Table 3, it is found that CNN-LSTM has the lowest training and validation loss, the highest accuracy, precision, recall and F_1 value, strong generalization ability, optimal performance, but the longest training time, training efficiency needs to be improved; BP model, although its performance is slightly inferior to that of CNN-LSTM, but the indicators are good, and the training time is only 2.6204s, which is a better performance and training time balance; FCNN and LSTM models have higher training and validation losses, relatively low performance indicators and poor performance.

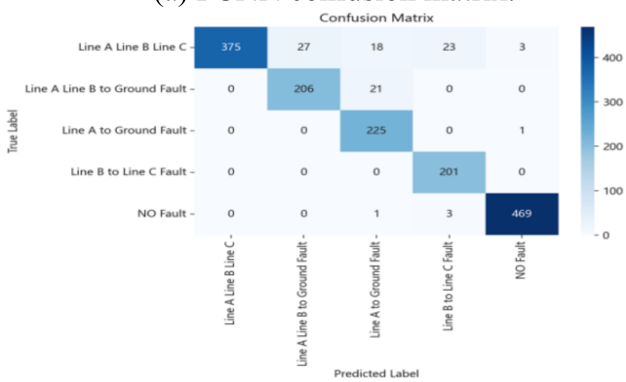
3.3.3. Comparison of Confusion Matrices across Models



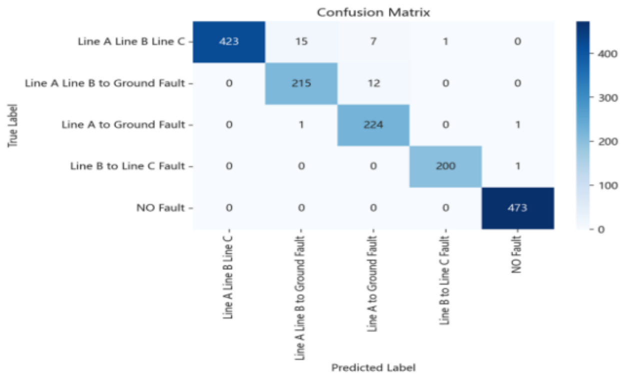
(a) FCNN confusion matrix.



(b) BP confusion matrix.



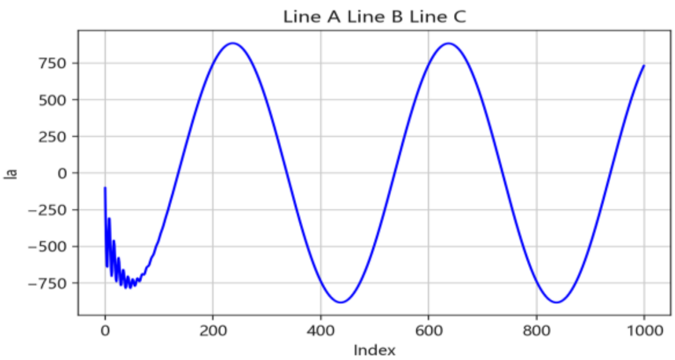
(c) LSTM confusion matrix.



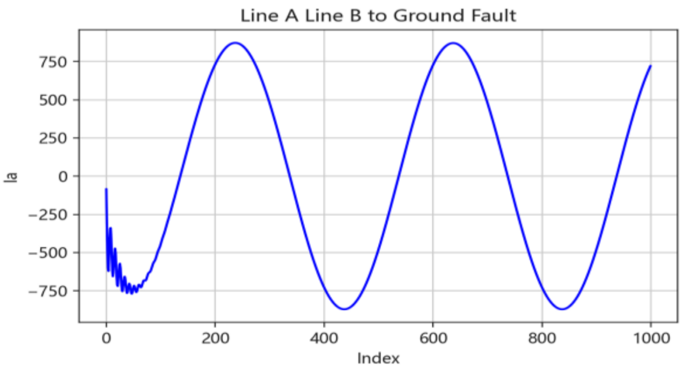
(d) CNN-LSTM confusion matrix.

Figure 6. Comparison of four model confusion matrices.

Observing the above Figure 6, it is found that CNN-LSTM has the best prediction performance with the highest correct prediction rate, while the LSTM model has the lowest correct prediction rate.



(a) Line A Line B Line C Failure Ia Change Laws.



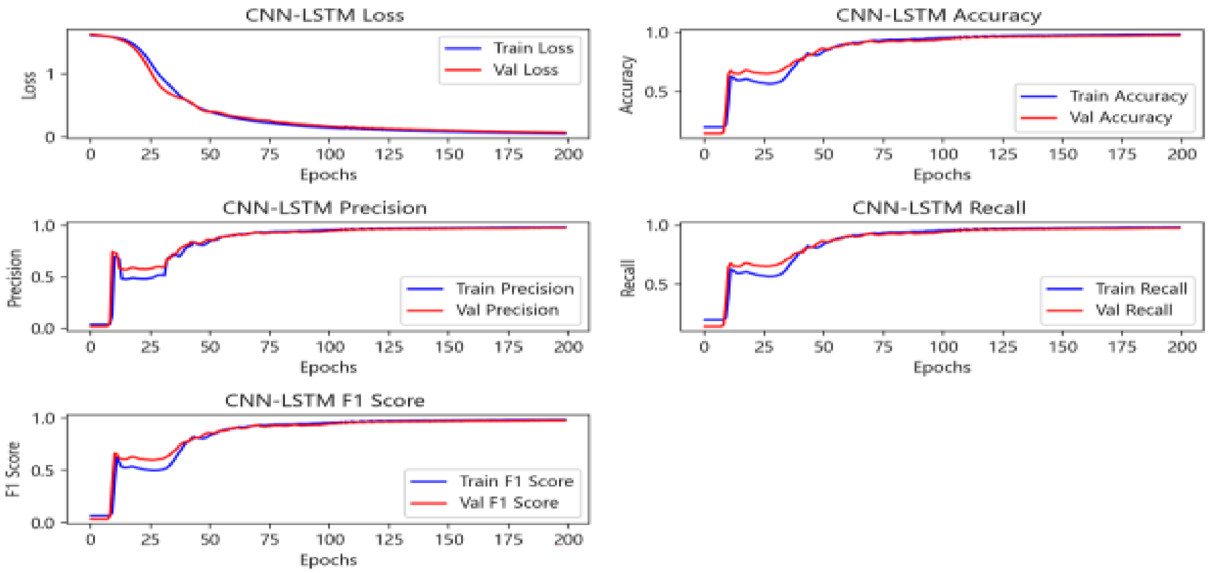
(b) Line A Line B to Ground FaultFailure Ia Change Laws.

Figure 7. Comparison of the change rule of two fault types I_a

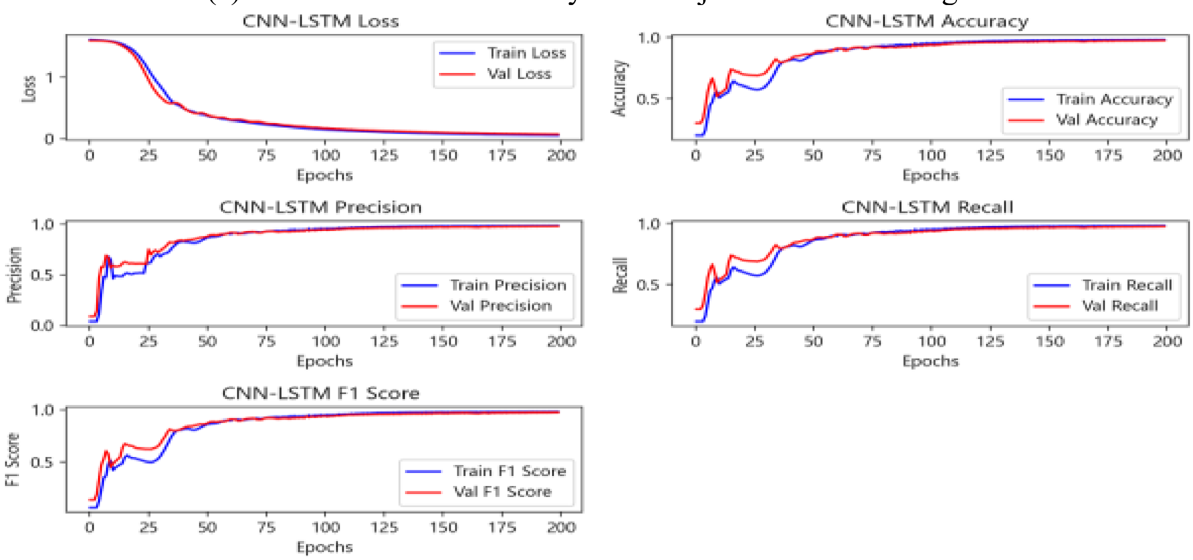
In Figure 6 above, for the prediction of Line A Line B Line C faults, all four types of models show a high tendency to misjudge. This may be attributed to the fact that when the two types of faults occur, the current in the circuit will produce a significant sudden change, accompanied by a different degree of voltage drop, such as Figure 7 above in the two categories of the I_a rule of change, which leads to a high degree of similarity between the two in the electrical characteristics of the electrical characteristics of the two, thereby increasing the difficulty of the model to accurately distinguish between them.

3.4. Model optimization

In order to balance the convergence speed and stability of the model, the learning rate strategy is dynamically adjusted: a larger learning rate is used at the beginning of training to accelerate convergence and avoid the risk of parameter oscillations; the learning rate is reduced at the later stage to finely tune the parameters, jump out of the local optimum and prevent overfitting [15]. Through the ReduceLROnPlateau scheduler based on the validation loss adaptive adjustment, when the loss of consecutive rounds of epoch did not improve automatically attenuate the learning rate. The comparison of each evaluation index before and after adding dynamically adjusted learning rate is shown in Figure 8.



(a) Before the inclusion of dynamic adjustment of learning rates.



(b) After the inclusion of dynamic adjustment of learning rates

Figure 8. Before and after adding dynamically adjusted learning rates.

Observing the above Figure 8, it is found that the loss decreases faster after using this method, and eventually converges to a lower loss value, and there is a significant improvement trend in indicators such as training set and validation set accuracy, and the training time is shortened from 21.1543s to 18.3431s, which indicates that the performance of the model in the aspects of training stability, fault detection efficiency and accuracy is significantly improved by adding the dynamically adjusted learning rate.

4. Conclusions

Transmission line faults can cause serious losses in environment, production, life and other aspects; in order to improve the accuracy of fault detection and reduce the detection cost, it is of great practical significance to apply deep learning algorithms to line fault detection. In this paper, CNN-LSTM neural network is constructed for fault detection: firstly, outlier processing, feature selection and data imbalance processing are carried out on the datasets, and four types of models are constructed, namely FCNN, BP, LSTM, and CNN-LSTM, and by comparing the indexes of the models, it is found that the loss of the CNN-LSTM model is reduced by about 70% compared with the other models, and the accuracy, precision, and recall are improved by about 2%, which is sufficiently high. and about 2% in accuracy, precision, recall, etc., fully reflecting its superiority. The model effectively integrates the spatial feature extraction capability of convolutional neural network and the temporal modeling advantage of long and short-term memory network, which provides an innovative solution for accurate and fast detection of transmission line faults, and helps to promote the development of transmission line fault monitoring technology in smart grid.

Future research can focus on the CNN-LSTM model, explore the fusion strategy of multimodal data such as current, voltage, environmental factors, etc., and continue to improve the robustness and adaptability of the model in complex operating environments, so as to provide a more solid technical support for guaranteeing the stable operation of transmission lines and improving the reliability of power systems.

References

[1] Chen GW. Research on fault diagnosis and repair strategy of power equipment based on artificial intelligence[J]. Network Security and Informatization,2023, (12):75-77.

[2] Qi Zhijia. Research on transmission line component identification and fault detection based on convolutional neural network[D]. Anhui University of Technology,2024.

[3] Yang, J., Wu, H., Wang, S., & Dong, X. A New Method of Fault Phase Selection for Transmission Line Based on Composite Feature of Fault Current and Support Vector Machine[J]. Journal of Vibration Testing and SystemDynamics,2022,6(2) :215--234.

[4] SHI Wanyu, WEI Junqiang, ZHAO Yunhao. Transmission line fault diagnosis based on improved sparrow algorithm-support vector machine[J]. Zhejiang Electric Power,2021,40(11):38-45.

[5] Jamil, M., Sharma, S.K. & Singh, R. Fault detection and classification in electrical power transmission system using artificial neural network[J]. SpringerPlus,2015, 4: 334.

[6] LIN Ji-Kang, REN Yi-Rui, SHAN Xin, et al. Probabilistic diagnosis of power system faults based on logistic regression deep neural network[J]. Journal of Tianjin University (Natural Science and Engineering Technology Edition),2021,54(02):186-195.

[7] Rafique, F., Fu, L., & Mai, R. End to End Machine Learning for Fault Detection and Classification in Power Transmission Lines[J]. Electric Power Systems Research, 2021,199(10):107430.

[8] YANG Yuanpeng, WANG Wenzhuo, ZHENG Shengquan, et al. FCNN to enhance the speed of RCS computation in the optimal design of electrical large-size objectives[J]. Journal of Terahertz Science and Electronic Information,2025,23(03):272-277.

[9] H.F. Wu. Research on the improvement of pattern recognition accuracy of high-dimensional data mining based on BP neural network optimization algorithm[J]. Information System Engineering,2025, (04):67-70.

[10] Huang Kangpeng, Feng Feng. Sequential recommendation algorithm based on one-dimensional convolutional neural network[J]. Computer Technology and Development,2025,35(03):172-178.

[11] GAO Yadong, LI Guan, ZHAO Lifeng, et al. A study on the application of a GM (1,1) model optimized by quartile method in the monitoring of high-risk slopes[J]. Urban Surveying,2023, (03):171-175.

[12] LI Maolin, LI Xingcheng, LIU Haidong, et al. Study on the effectiveness of feature selection method for influencing factors of gas outflow[J]. Coal Technology,2025,44(05):147-151.

[13] SHI Hongbo, CHEN Yuwen, CHEN Xin.A review of research on SMOTE oversampling and its improvement algorithm[J]. Journal of Intelligent Systems,2019,14(06):1073-1083.

[14] Zhou Ji. Fatigue driving detection method based on GRU-Attention network and SMOTE algorithm[J]. Automobile Practical Technology,2025,50(09):46-51+66.

[15] MENG Qianqian, LIU Tengfei, XIE Jongna. Privacy-preserving scheme for learning rate tailoring gradient optimization for federated learning[J]. Journal of Beijing Institute of Electronic Science and Technology,2024,32(04):46-54.