Proceedings of the 3rd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2025.28592

Fault Prediction of 5G Base Station RF Module Based on Machine Learning Algorithms

Mingzan Ning

College of Information Science and Technology & Artificial Intelligence, Nanjing Forestry University, Nanjing, China
15085356999@njfu.edu.cn

Abstract. In this paper, a Transformer optimization framework based on differential evolution algorithm and bidirectional long short-term memory network is proposed, which provides an intelligent solution for fault prediction of RF modules in 5G base stations. Through variable correlation analysis, it can be concluded that modulation efficiency (0.19) has the strongest positive correlation with fault_type, while runtime (-0.08) has the strongest negative correlation with fault_type. After comparing the framework with models such as Decision Tree, Random Forest, CatBoost, XGBoost, LightGBM, and ExtraTrees, it is found that DE BiLSTM Transformer has significantly better comprehensive performance in fault classification prediction of 5G base station RF modules: its accuracy reaches 0.956, which is higher than 0.925 of XGBoost and LightGBM, 0.921 of ExtraTrees, 0.904 of Random Forest, 0.864 of Decision Tree, and 0.798 of CatBoost, indicating a clear lead in overall classification accuracy; The recall rate is 0.956, enabling more comprehensive identification of actual fault samples; The accuracy of 0.958 is only slightly lower than the 0.919 of Random Forest and much higher than the 0.9 of XGBoost and 0.86 of LightGBM, demonstrating excellent ability to accurately identify fault types; F1 values of 0.941 and AUC values of 0.979 were also highest, showing better performance in balancing accuracy and recall, distinguishing faulty from fault-free samples. This advantage stems from the synergistic effect of differential evolution algorithm optimizing model parameters, BiLSTM capturing bidirectional temporal features, and Transformer self-attention mechanism focusing on key features, making it more suitable for complex feature correlation scenarios in RF module fault prediction.

Keywords: Differential Evolution Algorithm, Bidirectional Long Short Term Memory Network, Transformer, 5G Base Station RF Module Fault Prediction

1. Introduction

With the deep penetration of 5G networks into all walks of life, the stable operation of base stations as core infrastructure is directly related to communication quality and service continuity. As the core component of 5G base station signal transmission and reception, RF module undertakes key functions such as signal modulation, power amplification and wireless transmission [1]. However, it is exposed to complex outdoor environments for a long time and is easily affected by factors such as

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.28592

temperature fluctuations, electromagnetic interference, and component aging, resulting in signal attenuation, network disconnection, and even equipment burn-out [2]. According to industry data, RF module failures account for more than 35% of total base station failures [3]. The traditional maintenance mode, which relies on manual inspection or threshold alarm, not only has response lag and high missed detection rate, but also increases the maintenance cost due to the surge of 5G base station deployment density (3-5 times higher than 4G). Therefore, building an accurate and efficient fault prediction system to realize the transformation from "post-repair" to "pre-warning" has become a key issue to ensure the reliability of 5G networks [4].

Machine learning algorithms provide an intelligent solution for predicting failures in the RF module of 5G base stations. By modeling multi-dimensional temporal data such as temperature, voltage, and signal power generated during the operation of the module, the machine learning model can reveal the hidden fault precursor patterns in the data, realizing early prediction of fault type and occurrence probability [5]. For example, traditional support vector machines can capture nonlinear fault features through feature mapping, while random forests can reduce the risk of overfitting through ensemble learning. However, the fault data of 5G RF modules often exhibit strong temporal correlation and multimodal coupling, making it difficult for traditional algorithms to effectively model long-term dependencies, resulting in a significant decrease in prediction accuracy in complex scenarios [6].

Aiming at the above problems, this paper proposes a transformer optimization framework based on differential evolution algorithm and bidirectional long short-term memory network. Transformer model can process long-term data in parallel through self-attention mechanism, but its hyperparameter selection depends on experience, and it is easy to fall into local optimum. Bidirectional long short-term memory networks (BiLSTM) can capture bidirectional time dependencies through forward and backward propagation, compensating for the deficiency of Transformer's sensitivity to local time features. Differential evolution algorithm simulates the mutation, crossover and selection operations in the process of biological evolution, and can adaptively optimize the attention weight of Transformers and the hidden layer dimension of BiLSTM to achieve global optimization across model parameters. The synergistic effect of the three not only retains the advantages of Transformer in handling long sequences, but also enhances the local feature modeling ability through BiLSTM, and finally improves the robustness of the model through differential evolution algorithm, providing a more accurate temporal modeling solution for 5G RF module fault prediction.

2. Data sources

This dataset focuses on predicting faults in the RF module of 5G base stations, with a total of 758 records covering 8 feature variables and 1 classification target variable. The specific characteristic variables are as follows: temperature, which reflects the heat generated by the module during operation; Voltage, reflecting the working power supply of the module; Signal power, representing the strength of the signal transmitted by the module; Noise level, indicating the degree of interference in signal transmission; Run time, record the cumulative working time of the module; Modulation efficiency reflects the effectiveness of signal modulation; Uplink rate, reflecting the transmission speed of data from the terminal to the base station; The downlink rate reflects the transmission speed of data from the base station to the terminal. The target variable is the fault type, which is divided into four categories: 1 represents minor faults, 2 represents moderate faults, and 3 represents normal. Select some data for display, as shown in Table 1.

DOI: 10.54254/2755-2721/2025.28592

Table 1. Selected partial dataset

Temperature	Voltage	Signal power	Noise level	Operating hours	Modulation efficiency	Uplink speed	Downlink speed	Fault type
42.5	4.91	-32.4	-105.4	1612	88.8	135.4	359.7	3
39.3	5.07	-36.8	-105.4	6564	93.7	144.6	326	3
49.3	4.83	-38.5	-84.5	4192	91.6	87.4	301.3	1
44.1	5.03	-40.5	-97.3	5570	93.7	93	283.9	1
41	4.89	-44.4	-101.6	5024	89.3	130.5	383.2	3

3. Correlation analysis

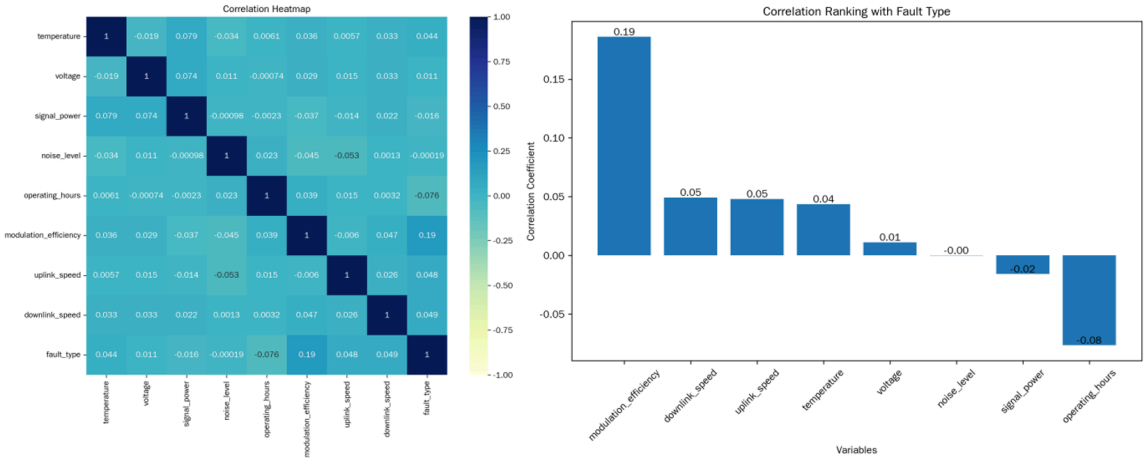


Figure 1. (a) The correlation heatmap. (b) The correlation ranking

Perform correlation analysis on various variables and draw a correlation heatmap, as shown in Figure 1. (a). For each variable's correlation with faultype, we have created a correlation ranking, as shown in Figure 1. (b). From the results, it can be seen that there is a positive correlation between runtime (-0.076), uplink rate (0.048), downlink rate (0.049), and faultype; The positive correlation between modulation efficiency (0.19) and faultype is the strongest, and the negative correlation between operation time (-0.08) and faultype is the strongest.

4. Method

4.1. Differential evolution algorithm

Differential evolution algorithm is a stochastic optimization algorithm based on swarm intelligence, which is mainly used to solve global optimization problems in continuous space [7]. It draws on the idea of "survival of the fittest" in biological evolution, and gradually approaches the optimal solution by simulating the process of population variation, crossover and selection. Compared with traditional evolutionary algorithms such as genetic algorithms, the core advantage of differential evolution lies in using the differential information among individuals within a population to generate new solutions without relying on the gradient information of the problem. Therefore, it has stronger

DOI: 10.54254/2755-2721/2025.28592

adaptability and robustness to nonlinear, multimodal, high-dimensional and other complex optimization problems [8].

The crossover operation aims to increase population diversity by exchanging partial dimensional values with the corresponding parent individuals with a certain probability, producing experimental individuals. The crossover probability determines the proportion of dimensions of mutant individuals in the experimental population, typically ranging from 0 to 1 [9].

4.2. Bidirectional long short term memory network

The Bidirectional Long Short-Term Memory Network (BiLSTM) is an extension of the Long Short-Term Memory Network (LSTM), which focuses on processing sequence data bidirectionally while capturing forward and backward context dependencies. The network structure of BiLSTM is shown in Figure 2. Although the traditional LSTM solves the long-term dependence problem of recurrent neural network (RNN), it can only transmit information from the beginning point to the end point in one direction, and it is difficult to handle the task that needs to combine future information. BiLSTM deploys two LSTMs in opposite directions in parallel, processing data from the beginning to the end and from the end to the end of the sequence, so that the output of each time step fuses past and future information, greatly improving the ability to capture contextual information [10].

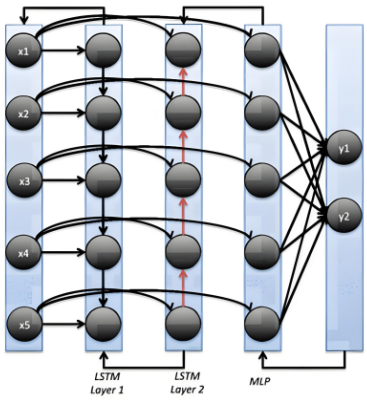


Figure 2. The network structure of BiLSTM [11]

The structure of BiLSTM consists of two independent LSTM units, called forward LSTM and backward LSTM, respectively. The forward LSTM processes the sequence in chronological order, the hidden state of each time step t recording information from the beginning of the sequence to t; The reverse LSTM processes in reverse chronological order, hiding information from the end of the sequence to t. At each time step, the hidden states of the two LSTMs are concatenated into a composite vector, which is used as the final output of that time step. This bidirectional parallel structure allows the model to both "review" historical information and "look ahead" to future information, making it particularly suitable for processing sequential data with bidirectional dependencies.

4.3. Transformer

The Transformer network is a sequence modeling model based on self attention mechanism, which completely eliminates the dependence of recurrent neural network (RNN) on sequence sequential processing, and greatly improves the modeling efficiency of long sequence data through parallel computing. The network structure of Transformer is shown in Figure 3. Unlike models such as RNN

DOI: 10.54254/2755-2721/2025.28592

and LSTM that rely on temporal progression, Transformer only captures the correlation between elements in a sequence through attention mechanism, and combines it with feedforward neural network to achieve end-to-end sequence transformation.

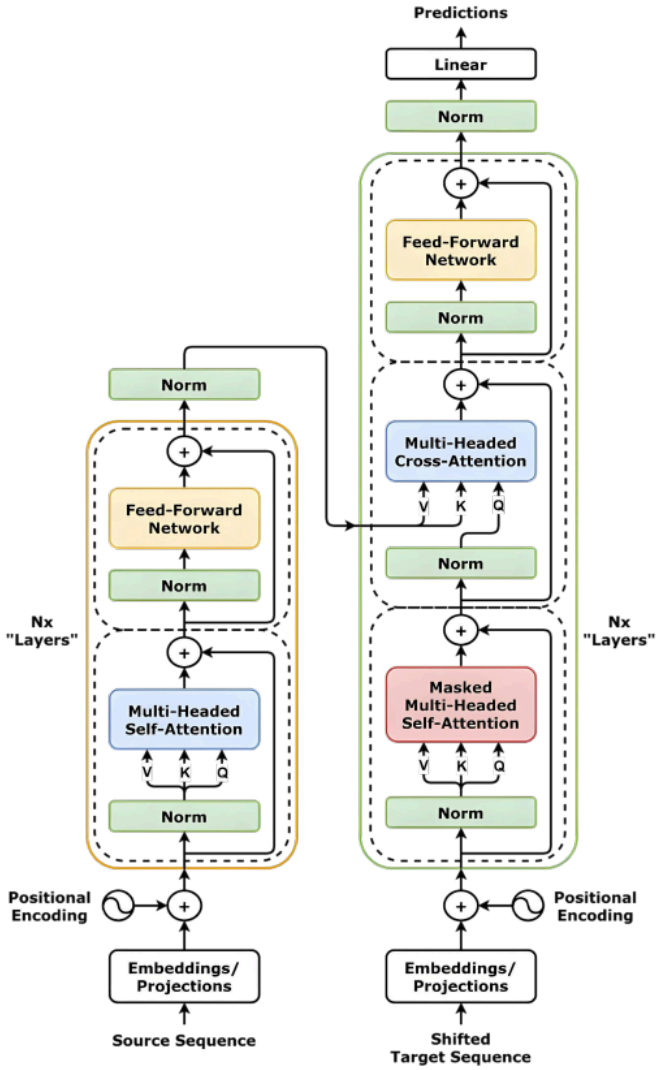


Figure 3. The network structure of Transformer [12]

4.4. DE-BiLSTM-Transformer

The core of differential evolution algorithm and bidirectional long short-term memory network optimization Transformer lies in the leap of temporal data modeling capability through multi-model collaboration and intelligent optimization. Transformer's self-attention mechanism excels at capturing global dependencies of long sequences, but lacks sensitivity to local temporal details; BiLSTM enhances forward and backward local feature extraction through a bidirectional gated structure to form a "global local" complementary architecture. In practical applications, the input sequence is first processed by the Transformer encoder to extract cross-time step dependent features, and then further encoded by the BiLSTM layer to identify adjacent time dependencies. This combination retains the advantages of Transformer for processing long sequences in parallel and complements the temporal continuity modeling capabilities with BiLSTM, making it particularly

DOI: 10.54254/2755-2721/2025.28592

suitable for scenarios such as 5G base station RF module fault prediction where both long-term trends and instantaneous fluctuations need to be considered. Differential evolution algorithm takes on the role of global parameter optimization in this framework, and solves the limitation of traditional manual parameter tuning. It generates candidate parameter combination by initializing population, generates new solution by mutation operation using individual difference information, and retains better parameter combination by crossover and selection operation. This optimization mechanism can dynamically adjust the model structure: on the one hand, it improves the allocation accuracy of Transformer attention weight and reduces the interference of irrelevant features; On the other hand, optimizing the gate parameters of BiLSTM enhances the filtering capability of temporal features. The cooperation of the three makes the model enhance the local time sensitivity while maintaining the global modeling capability, and finally realizes the double improvement of prediction accuracy and convergence efficiency through evolutionary iteration, providing a more robust solution for complex temporal data modeling.

5. Result

The experimental parameters of this project include 8 search agents, 6 maximum iterations, and 3 optimization parameter dimensions; The training parameters include a maximum of 1000 training epochs, a batch size of 256, a learning rate reduction factor of 0.1, a learning rate reduction period of 850, a gradient clipping threshold of 10, an optimizer of Adam, and an execution environment of Auto; The parameters related to Transformer and BiLSTM include 4 self attention heads, 128 key channels, a maximum position encoding of 512, and a dropout ratio of 0.01; The proportion of training set in data partitioning is 0.7, and the output dimension is 1. The software used is Matlab R2024a.

In terms of comparative models, this article selected decision trees, random forests, CatBoost, XGBoost, LightGBM, and ExtraTrees. The results of DE BiLSTM Transformer and various comparative experiments are shown in Table 2, and the bar chart of the comparative results is shown in Figure 4.

Table 2. The results of DE BiLSTM Transformer and various comparative experiments

Model	Accuracy	Recall	Precision	F1	AUC
Decision tree	0.864	0.864	0.856	0.86	0.91
Random Forest	0.904	0.904	0.816	0.858	0.959
CatBoost	0.798	0.798	0.871	0.829	0.931
XGBoost	0.925	0.925	0.9	0.909	0.958
LightGBM	0.925	0.925	0.86	0.892	0.967
ExtraTrees	0.921	0.921	0.919	0.887	0.972
DE-BiLSTM-Transformer	0.956	0.956	0.958	0.941	0.979

In the fault classification prediction of 5G base station RF modules, DE BiLSTM Transformer shows significantly better comprehensive performance than other models. Its accuracy reaches 0.956, higher than the 0.925 of XGBoost and LightGBM, 0.921 of ExtraTrees, 0.904 of Random Forests, 0.864 of Decision Trees, and 0.798 of CatBoost, leading significantly in overall classification accuracy. The recall rate is 0.956, which is also higher than other models, indicating

DOI: 10.54254/2755-2721/2025.28592

that it can more comprehensively identify actual fault samples; The accuracy rate is 0.958, only slightly lower than the 0.919 of random forest, far higher than the 0.9 of XGBoost and 0.86 of LightGBM, and performs excellently in accurately identifying fault types. The F1 value of 0.941 and AUC value of 0.979 are also the highest, indicating that it is more outstanding in balancing accuracy and recall, as well as distinguishing between faulty and non faulty samples. This advantage stems from the optimization of model parameters by differential evolution algorithm, the capture of bidirectional temporal features by BiLSTM, and the focusing of key features by Transformer self attention mechanism.

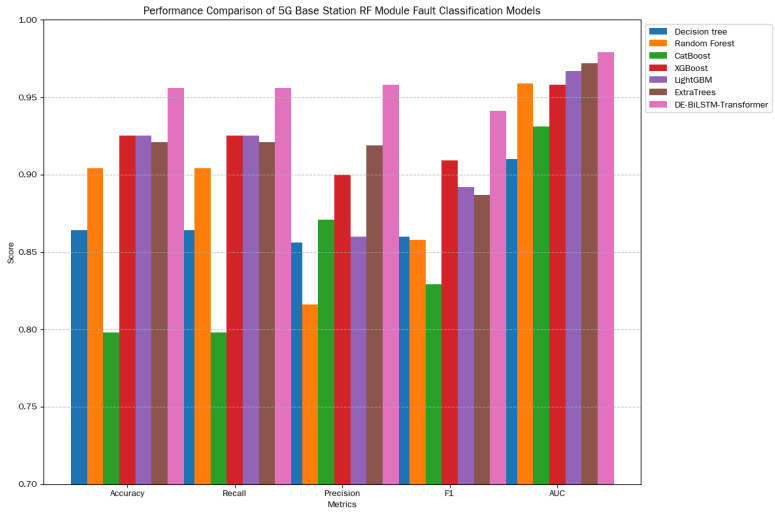


Figure 4. The bar chart of the comparative results

Output the confusion matrix of the test set for DE BiLSTM Transformer, as shown in Figure 5.

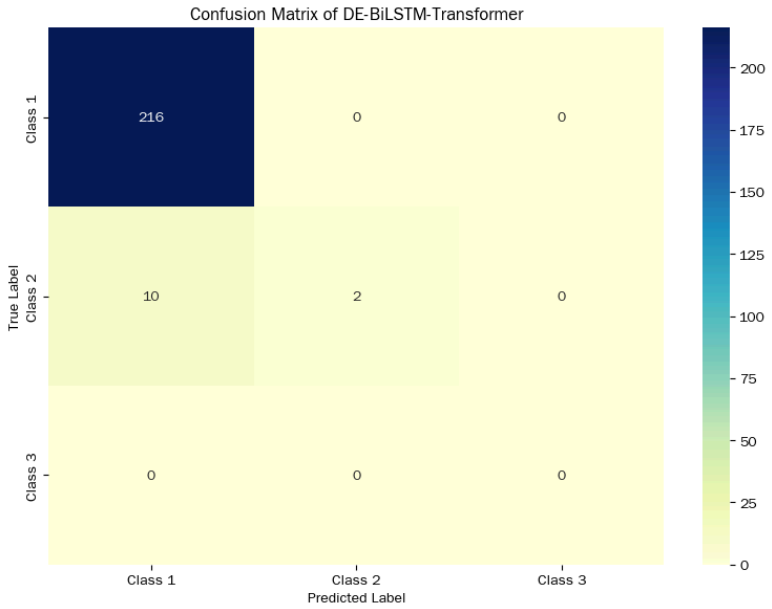


Figure 5. the confusion matrix of the test set for DE BiLSTM Transformer

In summary, aiming at the intelligent demand for fault prediction of RF modules in 5G base stations, this paper proposes an innovative framework based on differential evolution algorithm and

DOI: 10.54254/2755-2721/2025.28592

bidirectional long short-term memory network (BiLSTM) to optimize Transformer. Through multi-dimensional variable correlation analysis, it can be concluded that the positive correlation coefficient between modulation efficiency and fault type is 0.19, which is the strongest feature in the positive correlation; The negative correlation coefficient between runtime and fault type was-0.08, which was the most significant negative correlation. rison experiments show that the proposed DE BiLSTM Transformer exhibits comprehensive advantages compared to classical models such as decision tree, random forest, CatBoost, XGBoost, LightGBM, ExtraTrees, etc. Its accuracy is as high as 0.956, which is clearly ahead of XGBoost and LightGBM's 0.925, ExtraTrees's 0.921, Random Forest's 0.904, and Decision Tree's 0.864 and CatBoost's 0.798, showing a significant advantage in overall classification accuracy; The recall rate reaches 0.956, enabling more comprehensive identification of actual fault samples; The accuracy rate of 0.958 is slightly lower than Random Forest's 0.919, but far better than XGBoost's 0.9 and LightGBM's 0.86, demonstrating excellent ability to accurately identify fault types. Furthermore, the model has the highest F1 value (0.941) and AUC value (0.979) of all comparison models, showing excellent performance in balancing accuracy and recall, and distinguishing faulty from faulty-free samples.

References

[1] Al-Jumaily, A., et al. (2023). Machine learning modeling for radiofrequency electromagnetic fields (RF-EMF) signals from mmWave 5G signals. *IEEE Access, 11*, 79648–79658.
[2] Nakarmi, P. K., Sternby, J., & Ullah, I. (2022). Applying machine learning on RSRP-based features for false base station detection. *Proceedings of the 17th International Conference on Availability, Reliability and Security*.
[3] Park, H., et al. (2025). 5G native network function for false base station detection using machine learning technique. *Information Systems Frontiers*, 1–16.
[4] Islam, M. A., et al. (2022). A deep neural network-based communication failure prediction scheme in 5G RAN. *IEEE Transactions on Network and Service Management, 20*(2), 1140–1152.
[5] Tsipi, L., et al. (2024). Machine learning-based methods for MCS prediction in 5G networks. *Telecommunication Systems, 86*(4), 705–728.
[6] Lin, Y., Wang, H., & Zha, H. (2024). The technology of radio frequency fingerprint identification based on deep learning for 5G application. *Security and Safety, 3*, 2023026.
[7] Porch, J. B., et al. (2020). Machine learning approach for automatic fault detection and diagnosis in cellular networks. *2020 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*. IEEE.
[8] Falkenberg, R., et al. (2018). Machine learning based uplink transmission power prediction for LTE and upcoming 5G networks using passive downlink indicators. *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*. IEEE.
[9] Yuliana, H. (2024). Comparative analysis of machine learning algorithms for 5G coverage prediction: Identification of dominant feature parameters and prediction accuracy. *IEEE Access, 12*, 18939–18956.
[10] Jere, S., et al. (2023). Bayesian inference-assisted machine learning for near real-time jamming detection and classification in 5G new radio (NR). *IEEE Transactions on Wireless Communications, 23*(7), 7043–7059.
[11] Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2019). The performance of LSTM and BiLSTM in forecasting time series. *2019 IEEE International Conference on Big Data (Big Data)*. IEEE.
[12] Han, K., et al. (2021). Transformer in transformer. *Advances in Neural Information Processing Systems, 34*, 15908–15919.