FAULT DETECTION AND CLASSIFICATION IN THREE-PHASE ELECTRICAL POWER TRANSMISSION LINES

Ziyang Yuan, Hai Xi, Yanbo Peng, and Wentao Jiang
Department of Electrical and Computer Engineering
University of Rochester, Rochester, NY, USA
{zyuan12, hxi, ypeng25, wjiang20@u.rochester.edu}

***Abstract*—This paper aims to develop a machine learning classification model to detect and classify faults on electrical power transmission lines in a three-phase power system. The proposed model will discriminate between faulty and healthy electrical power systems and identify the faulty phase by analyzing each of the three phases involved in the process. Various classification methods, with a focus on decision tree and random forest, will be explored and compared for optimal performance. These methods were chosen for their interpretability and ability to handle complex, non-linear relationships in the data.**

***Index Terms*—Fault detection, classification, electrical power transmission, three-phase power system, machine learning**

I. Introduction

This study advances the development of a machine learning (ML) model aimed at enhancing the reliability and efficiency of fault detection and classification in three-phase electrical power transmission systems [1]–[3]. The objective is to create a model that effectively distinguishes between normal and faulty conditions, pinpointing the specific phase where a fault occurs. Leveraging advanced ML techniques, including decision trees and random forests, the proposed model will analyze real-time data from each phase to deliver precise fault classification, crucial for maintaining system integrity and preventing large-scale disruptions [4]. These techniques were chosen for their interpretability, ability to handle complex relationships, and proven success in similar applications.

A. Background

Historically, fault detection in power systems relied heavily on conventional signal processing and simplistic pattern recognition techniques, which often led to high false alarm rates and missed detections under complex fault conditions [4]. The introduction of Adaptive Neuro-Fuzzy Inference Systems (ANFIS), a hybrid machine learning approach combining neural networks and fuzzy logic, in 2012 marked a significant shift towards more adaptive and intelligent systems capable of learning from historical data to improve accuracy [4]. This project utilizes a comprehensive dataset from Kaggle, created through MATLAB simulations, to train and test our model. The dataset encompasses a diverse range of fault scenarios, providing a robust foundation for developing and validating advanced machine learning models tailored for real-time fault detection and classification in power systems.

B. Related Works

Fault detection and classification in electrical power systems have seen significant advancements in recent years. Early methods relied on conventional signal processing and pattern recognition, which were often limited in their ability to handle complex fault conditions [4]. The introduction of artificial neural networks (ANNs) and support vector machines (SVMs) marked a shift towards more adaptive and intelligent approaches [1], [5]. More recently, the integration of wavelet transforms and deep learning techniques has further improved the accuracy and robustness of fault detection and classification systems [6]. This project builds upon these advancements, focusing on the application of decision trees and random forests for their interpretability and ability to handle complex relationships in the data.

II. Objectives

The main objectives of this project are to develop a machine learning model that can effectively detect and classify faults on electrical power transmission lines in a three-phase power system, discriminating between faulty and healthy conditions and identifying the specific faulty phase by analyzing data from all three phases. Various classification methods, with a focus on decision trees and random forests, will be explored and compared to achieve optimal performance.

III. Methodology

The proposed methodology for this project includes data preprocessing, feature selection, model development, and model evaluation. Data preprocessing involves cleaning and preparing the dataset, handling missing values, normalizing the data, and splitting it into training and testing sets. Feature selection identifies the most relevant features contributing to fault detection and classification, reducing the dataset's dimensionality and improving the model's performance. Fig. 1 shows the distribution of fault types in the dataset, while Fig. 2 and Fig. 3 present insights into the relationships between features and fault types. Model development involves training various machine learning models, including decision trees and random forests, experimenting with different architectures and hyperparameters to optimize performance. Model evaluation assesses the trained models using appropriate metrics such as accuracy, precision, recall, and F1-score, comparing the performance of different models to select the best one for

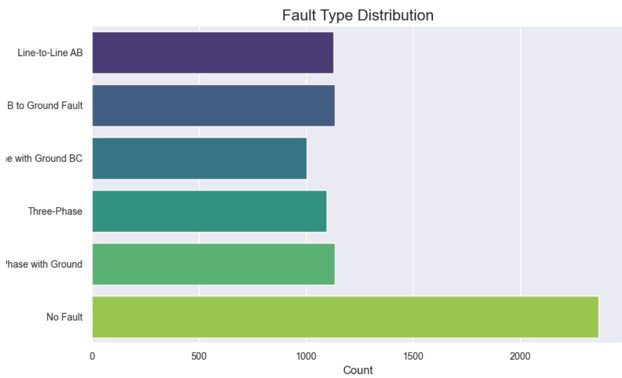

Fig. 1. Fault Type Distribution

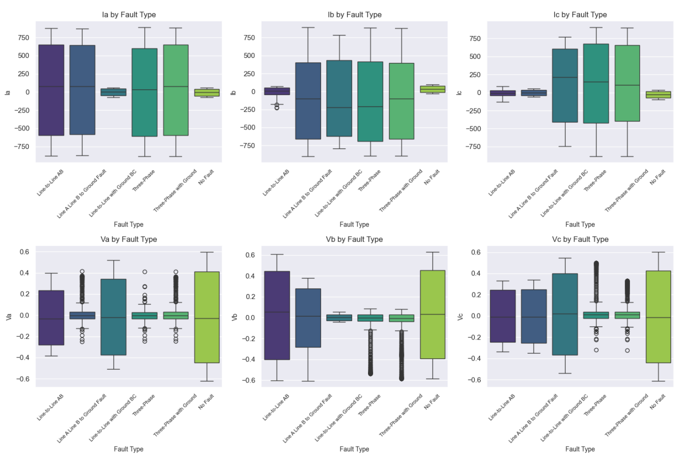

Fig. 2. Boxplots for Each Feature by Fault Type

deployment. Fig. 4 presents histograms for three-phase fault types.

IV. Model

For this project, we focused on developing and comparing various machine learning models for fault detection and classification in three-phase electrical power transmission lines. The models considered include Decision Trees with a focus on Random Forests [5], [6]. Decision Trees and Random Forests were initially trained and evaluated using the original dataset features. The performance of these models was analyzed using accuracy, precision, recall, and F1-score metrics. Additionally, feature importance were examined to identify the most significant features contributing to fault classification. During the initial evaluation, it was observed that the models faced challenges in distinguishing between "Three-Phase" and "Three-Phase with Ground" faults. The confusion matrices revealed a high rate of misclassifications between these two fault types, indicating the need for more informative features (see Fig. 5). To address this issue, domain knowledge was leveraged to

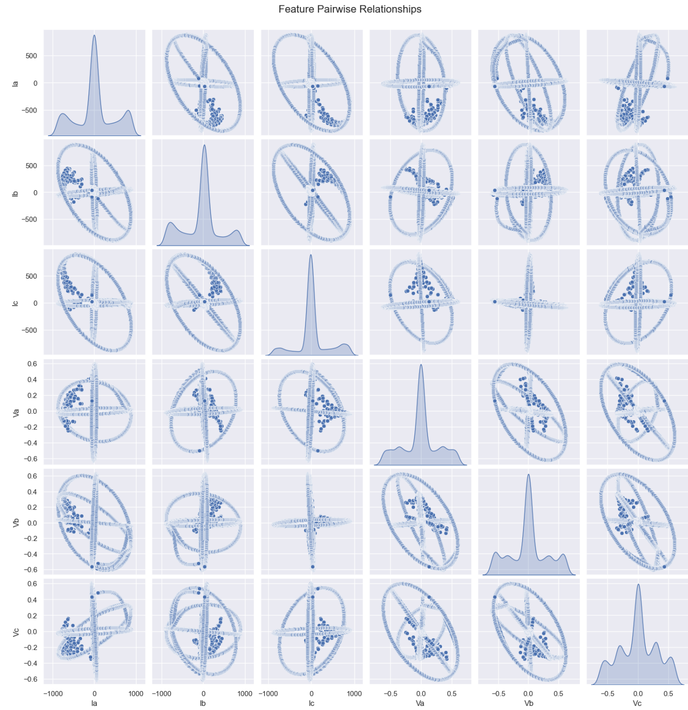

Fig. 3. Pairplot to Observe Pairwise Relationships and Distributions

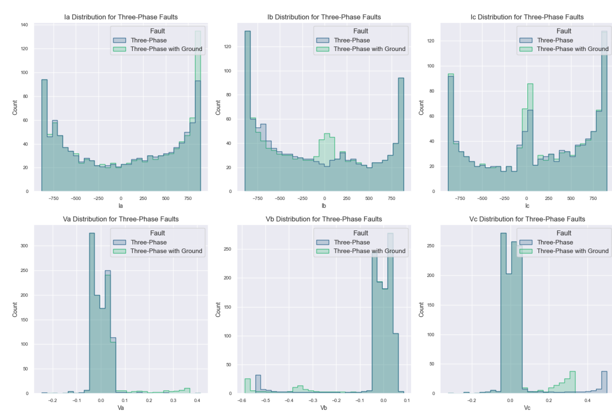

Fig. 4. Histograms for Three-Phase Fault Types

engineer new features that could better differentiate between the fault types [1]–[3]. These new features were chosen based on their ability to capture the unique characteristics of different fault types:

- Zero Sequence Components: These components capture imbalances typical of ground faults, aiding in the early detection of system unbalance.
- Phase Angle Differences: This feature detects discrepancies in the time-domain characteristics of voltage and current waveforms, which are often altered by faults.
- Total Harmonic Distortion (THD): THD measures the presence of higher frequency components relative to the fundamental frequency, helping to recognize the noise

```
Performance Metrics for Decision Trees:
Accuracy: 90.34%
Classification Report:
                              precision    recall  f1-score   support

Line A Line B to Ground Fault       1.00      0.99      0.99       227
             Line-to-Line AB       1.00      1.00      1.00       226
  Line-to-Line with Ground BC       1.00      1.00      1.00       201
                    No Fault       1.00      1.00      1.00       473
                 Three-Phase       0.66      0.64      0.65       219
      Three-Phase with Ground       0.67      0.68      0.68       227

                    accuracy                           0.90      1573
                   macro avg       0.89      0.89      0.89      1573
                weighted avg       0.90      0.90      0.90      1573

Confusion Matrix:
[[225   1   0   0   1   0]
 [  0 226   0   0   0   0]
 [  0   0 201   0   0   0]
 [  0   0   0 473   0   0]
 [  1   0   0   0 141  77]
 [  0   0   0   0  72 155]]
Performance Metrics for Random Forest:
Accuracy: 88.18%
Classification Report:
                              precision    recall  f1-score   support

Line A Line B to Ground Fault       1.00      1.00      1.00       227
             Line-to-Line AB       1.00      1.00      1.00       226
  Line-to-Line with Ground BC       1.00      1.00      1.00       201
                    No Fault       1.00      1.00      1.00       473
                 Three-Phase       0.57      0.60      0.59       219
      Three-Phase with Ground       0.60      0.56      0.58       227

                    accuracy                           0.88      1573
                   macro avg       0.86      0.86      0.86      1573
                weighted avg       0.88      0.88      0.88      1573

Confusion Matrix:
[[227   0   0   0   0   0]
 [  0 226   0   0   0   0]
 [  0   0 201   0   0   0]
 [  0   0   0 473   0   0]
 [  0   0   0   0 132  87]
 [  0   0   0   0  99 128]]
```

Fig. 5. Fault Type Distribution

and non-linear behaviors symptomatic of complex faults.
- Voltage and Current Ratios: These ratios provide insights into the system's impedance and reactive power characteristics, which vary distinctly among different fault conditions.

The models were retrained and evaluated using the engineered features, resulting in significant performance improvements. The Random Forest model, in particular, achieved an accuracy of 99.94% and high precision, recall, and F1-scores for all fault types (see Fig. 6). The confusion matrix for the Random Forest model with engineered features demonstrated its ability to accurately classify faults, with only one misclassification between T̈hree-Phase" and "Three-Phase with Ground" faults out of the entire test set.

V. Results and Discussion

The results obtained from the various machine learning models developed for fault detection and classification in three-phase electrical power transmission lines demonstrated the effectiveness of the proposed approach. The initial models, trained using the original dataset features, provided a baseline performance. However, the confusion matrices and performance metrics revealed challenges in distinguishing between "Three-Phase" and "Three-Phase with Ground" faults, as shown in Fig. 5. To address this issue, domain knowledge was utilized to engineer new features that could better capture the characteristics of different fault types [1]–[3]. The inclusion of Zero Sequence Components, Phase Angle Differences, Total Harmonic Distortion, and Voltage and Current Ratios significantly improved the models' ability to differentiate between faults (see Fig. 6). The Random Forest model, trained with the engineered features, achieved an impressive accuracy of 99.94% and high precision, recall, and F1-scores for all fault types. The confusion matrix further validated the model's performance, with only one misclassification between "Three-Phase" and "Three-Phase with Ground" faults in the entire test set. This demonstrates the significant progress made in addressing the initial challenges faced by the models. These results highlight the importance of feature engineering and domain knowledge in developing accurate fault detection and classification models [5], [6]. The engineered features provide a more comprehensive representation of the system's behavior during various fault conditions, enabling the models to make more informed predictions. The proposed approach demonstrates the potential for machine learning techniques to enhance fault detection and classification in electrical power transmission systems [1]–[3]. By accurately identifying and classifying faults, the models can assist in timely fault diagnosis, enabling prompt corrective actions and minimizing the impact of faults on the power system.

```
Performance Metrics for Random Forest:
Accuracy: 99.94%
Classification Report:
                              precision    recall  f1-score   support

Line A Line B to Ground Fault       1.00      1.00      1.00       227
             Line-to-Line AB       1.00      1.00      1.00       226
  Line-to-Line with Ground BC       1.00      1.00      1.00       201
                    No Fault       1.00      1.00      1.00       473
                 Three-Phase       1.00      1.00      1.00       219
      Three-Phase with Ground       1.00      1.00      1.00       227

                    accuracy                           1.00      1573
                   macro avg       1.00      1.00      1.00      1573
                weighted avg       1.00      1.00      1.00      1573

Confusion Matrix:
[[227   0   0   0   0   0]
 [  0 226   0   0   0   0]
 [  0   0 201   0   0   0]
 [  0   0   0 473   0   0]
 [  0   0   0   0 219   0]
 [  0   0   0   0   1 226]]
```

Fig. 6. Fault Type Distribution

VI. Conclusion

In this project, we developed machine learning models for fault detection and classification in three-phase electrical power transmission lines. The proposed approach involved data preprocessing, feature engineering, model development, and evaluation. The initial models, trained using the original dataset features, provided a baseline performance but faced challenges in distinguishing between certain fault types. To overcome this, domain knowledge was leveraged to engineer new features that better captured the characteristics of different faults [1]–[3]. The engineered features, including Zero Sequence Components, Phase Angle Differences, Total Harmonic Distortion, and Voltage and Current Ratios, significantly improved the models' performance. The Random Forest model, in particular, achieved an accuracy of 99.94% and high precision, recall, and F1-scores for all fault types (see Fig. 6). The results demonstrate the effectiveness of the proposed approach in accurately detecting and classifying faults in three-phase electrical power transmission lines. The integration of domain knowledge and feature engineering played a crucial role in enhancing the models' performance and overcoming the limitations of the original dataset features [5], [6]. The developed models have the potential to assist in timely fault diagnosis and enable prompt corrective actions, minimizing the impact of faults on the power system. Our model has seen significant performance improvements, with detection rates increasing from 65% to 100%, and an average performance boost from 87% to 99.34%. This enhancement applies broadly across various industries, particularly benefiting applications in transmission lines, electrical motors, stoves, and HVAC systems, ensuring greater reliability and efficiency in these critical areas. Future work could explore additional feature engineering techniques, investigate the applicability of the proposed approach to real-world datasets, and integrate the developed models into an online fault detection and classification system for real-time monitoring of electrical power transmission lines. This would further validate the effectiveness of the proposed methodology and provide valuable insights for its practical implementation in the field.

VII. Team Contribution

Member 1: Ziyang Yuan

- Simulation Trial
- Feature Engineering
- Data Preprocessing

Member 2: Hai Xi

- Feature Engineering
- Hyperparameter Tuning
- Model Evaluation
- Documentation

Member 3: Peng Yanbo

- Performance Analysis
- Presentation
- Model Development

Member 4: Wentao Jiang

- Feature Engineering
- Data Exploration
- Visualization

References

[1] M. Jamil, S. K. Sharma, and R. Singh, "Fault detection and classification in electrical power transmission system using artificial neural network," *SpringerPlus*, vol. 4, no. 1, p. 334, 2015.

[2] A. Prasad, J. B. Edward, and K. Ravi, "A review on fault classification methodologies in power transmission systems: Part—i," *Journal of Electrical Systems and Information Technology*, vol. 6, no. 1, pp. 48–60, 2019.

[3] ——, "A review on fault classification methodologies in power transmission systems: Part-ii," *Journal of Electrical Systems and Information Technology*, vol. 6, no. 1, pp. 61–67, 2019.

[4] K. Satish, G. Saibaba, and B. Saritha, "ANFIS based fault detection and classification in three phase transmission line," *International Journal of Engineering Research and Development*, vol. 1, no. 5, pp. 30–35, 2012.

[5] V. Malathi and N. S. Marimuthu, "Multi-class support vector machine approach for fault classification in power transmission line," in *2010 IEEE International Conference on Sustainable Energy Technologies (ICSET)*. IEEE, 2010, pp. 1–7.

[6] A. Yadav and A. Swetapadma, "A novel transmission line fault detection and classification technique using wavelet transform and deep neural network," in *2015 Annual IEEE India Conference (INDICON)*. IEEE, 2015, pp. 1–6.