

# Ensemble Learning-based Transmission Line Fault Classification for Smart Grid

Paul Michael Custodio, Made Adi Paramartha Putra, Dong-Seong Kim, Jae-Min Lee

Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea  
(pmbcustodio, mdparamartha95, dskim, ljmpaul)@kumoh.ac.kr

**Abstract**—Fault occurrences in smart grids cause disruptions in power system services. The most affected elements among power systems are overhead transmission lines; as such maintaining system reliability is of high priority. This paper aims to deploy a transmission line fault classification system. In which, the proposed model is able to classify the fault utilizing an ensemble model from multiple machine learning classifiers. The performance of the ensemble model for classification on accuracy, precision, recall, and F-1 score are 99.57%, 99.58%, 99.57%, and 99.58% respectively. Future works of this paper include detection of fault location as well as deployment of model to a prototype real-time system, consider higher reliability contingency cases, and employ the use of federated learning for increased security of data.

**Index Terms**—Ensemble learning, fault classification, power system, smart grid

## I. INTRODUCTION

In current smart grid (SG) systems, system reliability is one of the main issues being encountered. System reliability refers to the probability of a system to work at normal condition [1]. Among problems in SGs, faults experienced by transmission lines (TLs) failure are the most occurring. TL faults may be classified into two main types: symmetrical and unsymmetrical. For symmetrical fault, it is the type of fault wherein the system experiences balanced disturbance among the three phases, while unsymmetrical faults experience imbalance within the transmission lines. Hence, early detection and classification of fault in power system allows for faster system operator response [2].

The contributions of this study are as follows:

- 1) Generate labeled TL dataset for fault classification;
- 2) Create an ensemble model for TL fault classification using different multi-class classifiers.

## II. SYSTEM METHODOLOGY

This section discusses the methodology adopted in this study. It consists of (i) Dataset Generation and (ii) Ensemble Learning Classification.

### A. Dataset Generation

In order to simulate the target system, MATLAB Simulink was used to model the IEEE 5-Bus System, which is made up of the following elements: two (2) generators, three (3) load buses, and seven (7) TLs (Line 1-2, Line 1-3, Line 2-3, Line 2-4, Line 2-5, Line 3-4, Line 4-5); the Simulink model may

be seen in Fig. 1. Making use of the modeled system, the following faults have been induced in each TL individually: three-phase-to-ground (3Ph), double line-to-ground (DLG), single line-to-ground (SLG), by use of Simulink's 'Three-Phase Fault' Block. The induced fault was set to occur at

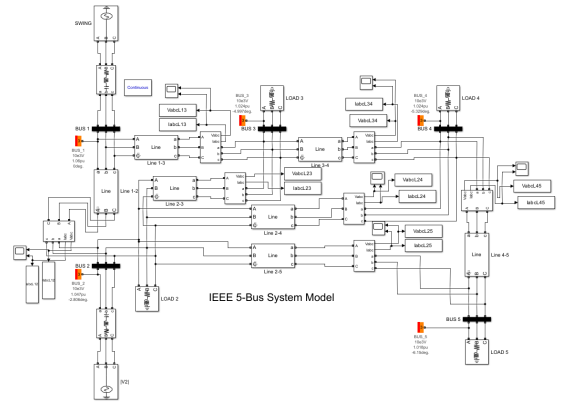


Fig. 1. IEEE 5-Bus Model

one-sixtieth (1/60) of a second until five-sixtieths (5/60) of a second, while the extracted data which consists of the phase voltages ( $V_a$ ,  $V_b$ ,  $V_c$ ) and line currents ( $I_a$ ,  $I_b$ ,  $I_c$ ) of each TL, was gathered using Simulink's Three-Phase V-I Measurement Block. Multiple instances have been extracted and labeled according to the induced fault experienced by the system. Labeling of the dataset created are named as 3Ph, DLG, SLG, and No Fault. Total instances showed the following amount as seen in Table I.

TABLE I  
DATASET INSTANCES

Type of Fault	Instances
3-Phase	824
Double Line-to-Ground	776
Single Line-to-Ground	1235
No Fault	39497

### B. Ensemble Learning Classification

From the dataset generated, labeled faults were determined each having forty-two (42) features, six (6) features for each

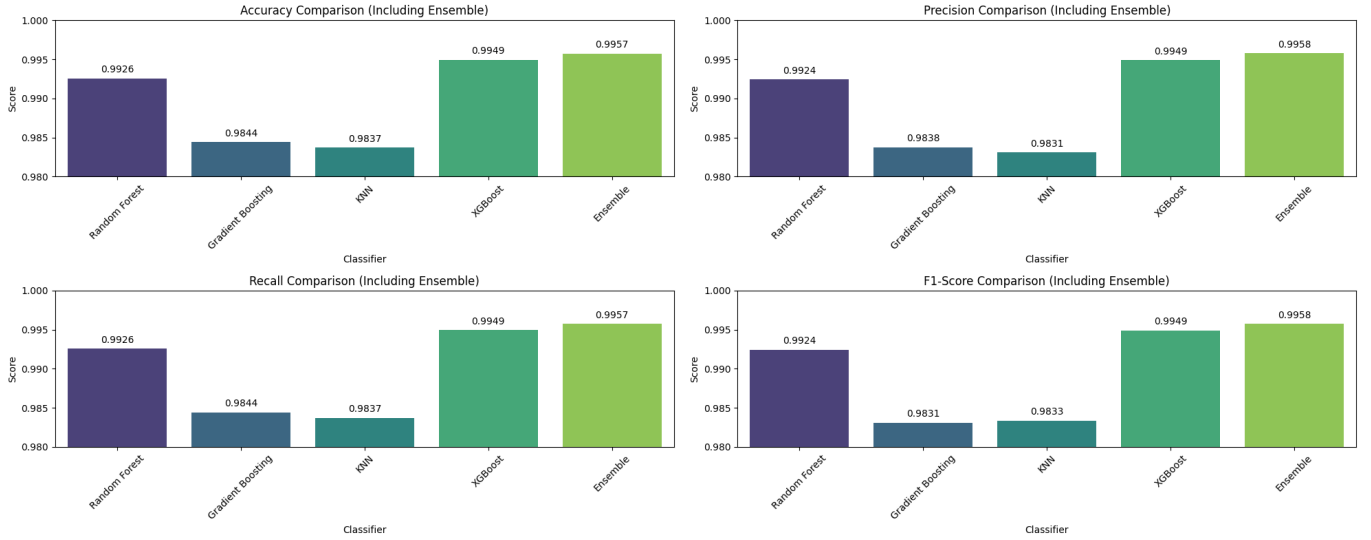


Fig. 2. Performance Metrics Results

TL. The machine learning (ML) classifiers used on the dataset were Random Forest (RF), Gradient Boosting (GB), K-Nearest Neighbor (kNN), and Extreme Gradient Boosting (XGB), different ML methods were used for fault classification [3] [4]. These ML classifiers were then trained and tested on the dataset, after which is then compiled for ensemble learning, as this method allows for higher performance for predictions [5]. The model of the system made use of the hard voting ensemble wherein each individual model in the ensemble makes a prediction, and the final prediction for the ensemble is determined by selecting the class that receives the most votes among the individual models' predictions.

### III. RESULTS ANALYSIS AND EVALUATION

Since the model is a classifier, its performance tested on the these statistical metrics: accuracy, precision, recall, and F1-score. Accuracy quantifies the proportion of correctly classified instances out of the total number of instances in the dataset. The ensemble model gave the highest accuracy of 99.57% as seen in Fig. 2a against 99.26%, 98.44%, 98.37%, and 99.49% for RF, GB, kNN, XGB, respectively.

Precision evaluates the model's ability to correctly identify positive instances among the predicted positive results. It is calculated as the ratio of true positives to the sum of true positives and false positives. The ensemble model gave the highest precision of 99.58% as seen in Fig. 2b against 99.24%, 98.28%, 98.31%, and 99.49% for RF, GB, kNN, XGB, respectively.

Recall, often referred to as sensitivity or true positive rate, assesses the model's ability to identify all positive instances in the dataset. The model returned the highest recall at 99.57% as seen in Fig. 2c against 99.26%, 98.44%, 98.37%, and 99.49% for RF, GB, kNN, XGB, respectively.

Finally, the F1-score is the harmonic mean of precision and recall and offers a balanced performance metric, particularly

when dealing with imbalanced datasets. It strikes a balance between precision and recall, giving equal importance to both metrics; in which the ensemble model gave an F1-score of 99.58% as seen in Fig. 2d against 99.24%, 98.31%, 98.33%, and 99.49% for RF, GB, kNN, XGB, respectively.

### IV. CONCLUSION

Classification of TL faults were achieved. The simulation model of the smart grid system was completed and generation of labeled TL fault dataset was obtained using MATLAB Simulink, and classes were split into four (4) namely: three-phase fault, double line-to-ground fault, single line-to-ground fault, and no fault. The data was then used to train and test the following machine-learning models: RF, GB, kNN, and XGB; which were then used to generate the proposed ensemble model. The ensemble model gave results of 99.57%, 99.58%, 99.57%, and 99.58% for accuracy, precision, recall, and F-1 score, respectively; which shows the effectiveness of the model. Future works of this study include: fault location detection, inclusion of higher level contingency and simultaneous fault occurrences, implementation of federated learning, and prototype deployment in real-time systems.

### ACKNOWLEDGMENT

This research was supported by Priority Research Centers Program through the NRF funded by the MEST (2018R1A6A1A03024003), and by MSIT under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2023-2020-0-01612) and by the MSIT, Korea, under the ICT Challenge and Advanced Network of HRD program(IITP-2023-RS-2022-00156394) both supervised by the IITP.

## REFERENCES

- [1] A. Abdukhakimov, S. Bhardwaj, G. Gashema, and D. -S. Kim "Reliability Analysis in Smart Grid Networks Considering Distributed Energy Resources and Storage Devices," *International Journal of Electrical and Electronic Engineering & Telecommunications*, Vol. 8, No. 5, pp. 233-237, September 2019. Doi: 10.18178/ijeetc.8.5.233-237
- [2] K. Al Kharusi, A. El Haffar, and M. Mesbah, "Fault Detection and Classification in Transmission Lines Connected to Inverter-Based Generators Using Machine Learning," *Energies*, vol. 15, no. 15, p. 5475, Jul. 2022, doi: 10.3390/en15155475.
- [3] H. Li, G. Hu, J. Li and M. Zhou, "Intelligent Fault Diagnosis for Large-Scale Rotating Machines Using Binarized Deep Neural Networks and Random Forests," in *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 2, pp. 1109-1119, April 2022, doi: 10.1109/TASE.2020.3048056.
- [4] Md. O. F. Goni et al., "Fast and Accurate Fault Detection and Classification in TLs using Extreme Learning Machine," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 3. Elsevier BV, p. 100107, Mar. 2023. doi: 10.1016/j.prime.2023.100107.
- [5] G. O. Anyanwu, C. I. Nwakanma, J. -M. Lee and D. -S. Kim, "Falsification Detection System for IoV Using Randomized Search Optimization Ensemble Algorithm," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 4, pp. 4158-4172, April 2023, doi: 10.1109/TITS.2022.3233536.