An Efficient Technique to Improve Fault Categorization in Transmission Lines

Shradha Umathe
Department of Electrical Engineering, G H Raisoni University, Amravati, India
shradhatul07@gmail.com (corresponding author)

Prema Daigavane
Department of Electrical Engineering, G H Raisoni University, Amravati, India
pdaigavane@gmail.com

Manoj Daigavane
Technical Education Regional Office, Nagpur, India
mbdaigavane07@gmail.com

Received: 1 December 2024 | Revised: 3 January 2025 and 22 January 2025 | Accepted: 24 January 2025

 DOI: https://doi.org/10.48084/etasr.9793

ABSTRACT

Machine Learning (ML) has become an essential tool for solving complex problems in electrical engineering. A major application of ML algorithms in this field is the fault categorization in transmission lines. ML models take into account the presence of faulty voltage or current while the fault is occurring in order to identify and categorize it. This research confirms the efficiency of ML algorithms by utilizing a faulted transmission line that was simulated in the MATLAB/Simulink environment. The main fault classification techniques implemented in this study are Decision Tree (DT) and Random Forest (RF). The Receiver Operating Characteristic (ROC) curve, the Precision-Recall (PR) curve, and the confusion matrix demonstrate the efficiency of the proposed techniques, which are able to optimize the fault categorization and increase both precision and effectiveness by accurately detecting faults within the transmission lines.

Keywords-fault classification; machine learning; random forest; transmission lines

I. INTRODUCTION

The continuous and reliable supply of electricity is critical today, and transmission lines, which form the backbone for distributing electricity over large geographical areas, are considered a critical part of power systems [1-3]. The stability and reliability of transmission lines can be threatened by a number of factors, with faults being the primary concern [4]. Transmission line faults cover a wide range of conditions, including short circuits, open circuits, and line-to-ground faults [5]. When a fault occurs, it interrupts the flow of electricity, potentially causing voltage sags, power outages, and even damage to equipment. It is necessary to detect and mitigate these faults in a timely manner to ensure continuity of power supply and prevent cascading failures in the network [6, 7]. Traditionally, the identification and categorization of faults in transmission lines has relied on rule-based systems and statistical techniques. These methods often use predefined thresholds or expert-based decision rules, knowledge, or historical data to identify and classify faults [8, 9]. While these approaches are effective in certain scenarios, they have significant limitations, especially when dealing with the complexity and uncertainties inherent in real power systems. Standard fault detection and classification approaches cannot capture the full range of fault states or changes in system behavior [10]. In addition, these methods have difficulty adapting to changes in power conditions or to the different characteristics of different types of faults [7]. As a result, there has been a growing interest in exploring alternative approaches that can overcome these limitations and improve the accuracy of fault categorization [9].

Traditional methods for fault classification in transmission lines, such as impedance-based methods, wavelet transform-based methods [6, 11, 12], and artificial intelligence-based methods [13-15], are mainly based on mathematical modeling and signal processing techniques [6, 16]. While effective in certain scenarios, they have inherent limitations. In particular, those that rely on signal processing techniques can be vulnerable to noise and disturbances in the power system [17, 23]. Recent advances in Machine Learning (ML) based fault classification techniques have changed this dramatically and offer significant advantages over traditional methods [18]. Some of the most prominent techniques are Decision Trees (DT), which partition data into branches based on feature conditions, and Random Forests (RF), which combine multiple decision trees to improve accuracy [12, 13, 19]. These techniques can handle noisy data and still produce accurate results. Traditional methods may struggle to accurately classify complex fault scenarios, such as multiphase faults or faults with high fault tolerance [20]. Therefore, there is a growing interest in using advanced techniques such as DT and RF to improve the accuracy and reliability of fault categorization in transmission lines [17, 18, 21, 22].

Motivated by the aforementioned literature, this paper proposes an ML algorithm to improve the fault classification accuracy of transmission lines. The main contributions of this paper are:

1. ML is used to perform an efficient classification in transmission line faults.
2. The performance of the proposed method is compared with previously proposed methods.
3. It is concluded that the proposed method provides better accuracy than other methods.

II. PROPOSED METHOD

A. Studied System

The MATLAB Simulink model of the proposed system is shown in Figure 1. The model consists of a single generating unit and a single RLC load of 100 MW and 10KW, respectively. Various faults on the transmission line, such as Line to Ground (LG), Line to Line to Ground (LLG), and Line to Line to Line (LLL), are created by using a fault block. Specifically, the dataset for training and testing the system is obtained by introducing the faults line-to-line (AB, BC, CA), line-to-ground (AG, BG, CG), double line to ground, three-phase and no fault on the transmission line and collecting the current and voltage values.

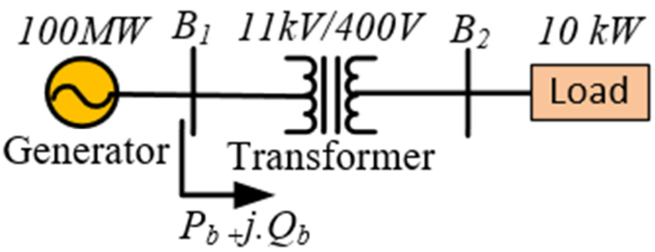

Fig. 1. Single line diagram of the proposed system.

B. Model Simulation and Data Generation

To perform effective fault classification, the training and testing data must contain enough samples, which requires generating large amounts of data covering different system conditions and faults. These data are generated using multiple runs of MATLAB software. The task is to classify the electrical faults in the system using current and voltage measurements from three phases A, B, and C. At the output side of the power system, we then collect and store the observed phase voltages and currents. The dataset contains features as voltages (Va, Vb, Vc) and currents (Ia, Ib, Ic) for the three phases. We have collected 2000 data points, and then labeled the data.

C. Normalization

This process includes cleaning, missing value handling, and normalization/scaling. Descriptive statistics and visualizations (e.g., histograms, correlation matrices) are used to understand the data distributions and the relationships between fault types and features. The process involves the steps of loading data, checking for missing values, defining different failure types, normalization, model training, and testing. Specific analysis includes current and voltage distributions for different fault types, highlighting overlaps and unique patterns, and using various ML algorithms for classification.

D. Random Forest

The RF training uses bootstrap sampling to create multiple subsets of the training data. Each split considers a subset of random features to reduce overfitting and each decision tree predicts fault types using the Gini Impurity given by [4].

$$\text{Gini Impurity} = 1 - \sum_{i=1}^{k} p_i^{\,2} \quad (1)$$

where k is the number of classes in the dataset, p_i is the proportion of instances in class i. The Gini Impurity measures the "impurity" of a node, with 0 indicating perfect classification. The Final Prediction in RF is given by [4]:

$$\text{Final Prediction} = \frac{1}{N} - \sum_{i=1}^{N} \hat{y}_i \quad (2)$$

where $\hat{y}_i$ is the prediction from the i-th tree and N is the number of trees in the RF. The splits are chosen to minimize impurity in the fault classification nodes. Ensemble learning predictions from all trees are aggregated using majority voting. Out-of-Bag (OOB) error samples are not used during training are used to estimate the performance, which is evaluated using metrics such as Recall, Precision, Accuracy, and confusion matrices. Classification accuracy refers to the number of correct predictions made relative to the total number of samples that were submitted, given as [4]:

$$\text{Accuracy} = \frac{N_{cp}}{T_p} \quad (3)$$

where N_{cp} is the number of correct predictions and T_p is the total number of predictions. Precision is calculated by dividing the total number of successful positive results by the total number of positive results that were predicted by the classifier [4, 8]:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (4)$$

Recall is the total number of successful positive results divided by the total number of applicable samples [4].

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (5)$$

The $\text{F1} - \text{score}$ is a measure of how accurately a classifier is able to identify and Recall specific instances. It is an important measure of robustness of the classifier. Although high Precision and low Recall are ideal for classification, they can miss many difficult cases. The higher $\text{F1} - \text{score}$, the better the performance of the model [4].

$$\text{F1} - \text{score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

This approach allows for robust, accurate, and real-time fault classification [4].

III. RESULTS AND ANALYSIS

The proposed model was analyzed based on the fault data generated from the simulated system. The data obtained from the simulation were used for training logistic regression, DT, K-Nearest Neighbors (KNN), and RF. Among them, DT and RF performed well and were therefore considered for testing the model. An Accuracy of approximately 87.52% was obtained on the validation data.

The Receiver Operating Characteristic (ROC) curve with the DT model, which is the gateway to understanding the classification performance of a model, is shown in Figure 2. The ROC curve is considered a powerful tool for evaluating binary classifiers. It visually represents the trade-off between the true positive rate and the false positive rate at various classification thresholds. The observations from the ROC curve graph are as follows:

- Line-to-line with ground AB class (area = 0.96): This curve hugs the upper left corner, indicating excellent performance. The high Area Under Curve (AUC) indicates that the model effectively distinguishes this fault type from others.
- Line-to-line with ground AC class (area = 1.00): A perfect AUC of 1.00. The model successfully detects this type of fault with no false positives or false negatives.
- Line-to-line with ground BC class (area = 1.00): Another perfect AUC, the model is accurate in identifying this fault type.
- No fault class (area = 0.96): A solid AUC, indicating reliable performance in detecting the normal state.
- Macro-average ROC curve (area = 0.98): Overall, the model has exceptional performance in every class.

Figure 3 illustrates the Precision-Recall (PR) curve, an important tool for evaluating the performance of the classification model, especially when dealing with unequal class sizes. The PR curve shows the trade-off between Precision and Recall. It should be noted that Precision is important when false positives are costly, and Recall is important when false negatives are critical. The observations from this curve are:

- Line-to-line with ground AB class (AP = 0.80): This curve demonstrates a good balance between Precision and Recall. An Average Precision (AP) of 0.80 suggests reliable performance in identifying this fault type.
- Line-to-line with ground AC class (AP = 0.99): The model excels in Precision and Recall for this fault type with an impressive AP of 0.99.
- Line-to-line with ground BC class (AP = 1.00): Another perfect score that balances Precision and Recall.
- No fault class (AP = 0.92): Solid performance in discriminating normal conditions.
- Macro-average PR curve (AP = 0.92): The model maintains a high AP across all classes.

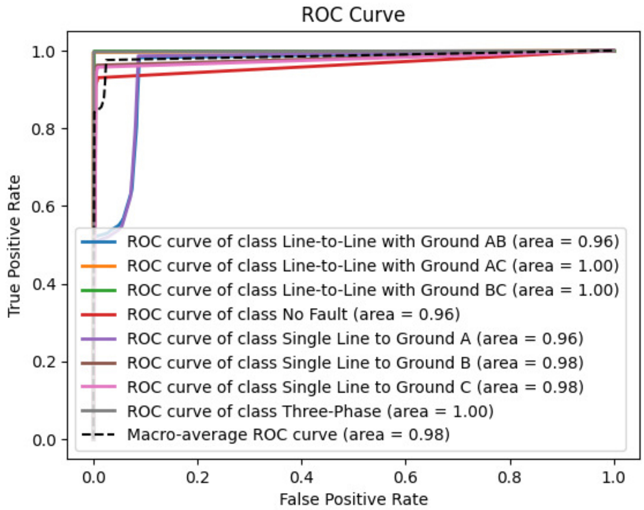

Fig. 2. ROC curve with DT model.

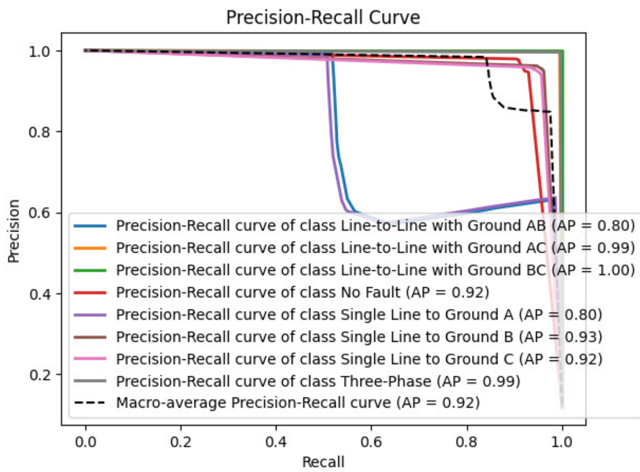

Fig. 3. Precision-Recall curve with DT model.

We can observe that with a small number of training examples the model works really well and the training score is high. As we add more examples, the training score gradually decreases. This is expected because the model becomes more generalized and less overfit to the training set. An effective way to determine how well a classification model works across multiple classes is the confusion matrix. The confusion matrix is the dominant and most popular technique for evaluating the correctness and accuracy of a classification. It also helps to identify the regions where the classifier has performed poorly. In addition, it helps determine whether the classification model is correct and what errors it produces. The confusion matrix with the DT model is shown in Figure 4.

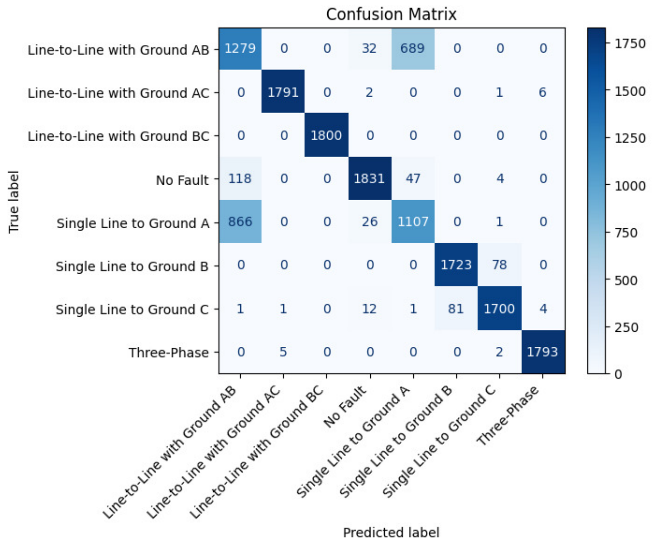

Fig. 4. Confusion matrix with DT model.

The ROC curve with the RF model shown in Figure 5 visualizes how well the model discriminates between different fault types. The observations are:

- Line-to-line with ground AB class (AUC ≈ 1.00): This curve hugs the upper left corner, indicating excellent performance.
- No fault class (AUC ≈ 1.00): The model excels in detecting normal conditions with a perfect AUC.
- Line-to-line with ground AC class (AUC ≈ 1.00): Another perfect result for the model.
- Line-to-line with ground BC (AUC ≈ 0.99): Near perfect score, the model effectively discriminates this type of fault.
- Macro-average ROC curve (AUC ≈ 0.99): Overall, the model performs admirably across all classes.

Figure 6 illustrates the PR curve for the RF model, which provides important information about how well our model balances Recall (true positive rate) and Precision (positive predictive value) for different types of faults. Here are the observations:

- Line-to-line with ground AB class (AP = 0.78): The curve starts strong in Precision, but dips slightly as Recall increases. This trade-off is common, as we prioritize Recall (capturing more true positives), Precision drops.
- Line-to-line with ground AC class (AP = 1.00): A perfect AP score. The model excels in Precision and Recall for this fault type.
- Line-to-line with ground BC class (AP = 1.00): Another perfect score. The model effectively discriminates this fault type.
- No Fault class (AP = 1.00): Again, a perfect score. The model performs flawlessly in detecting normal conditions.
- Macro-average PR curve (AP = 0.94): Overall, our model maintains a high AP across all classes.

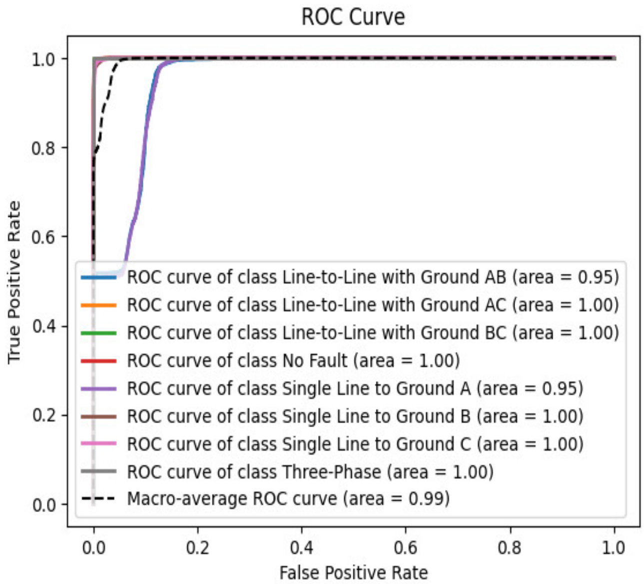

Fig. 5. ROC curve with RF model.

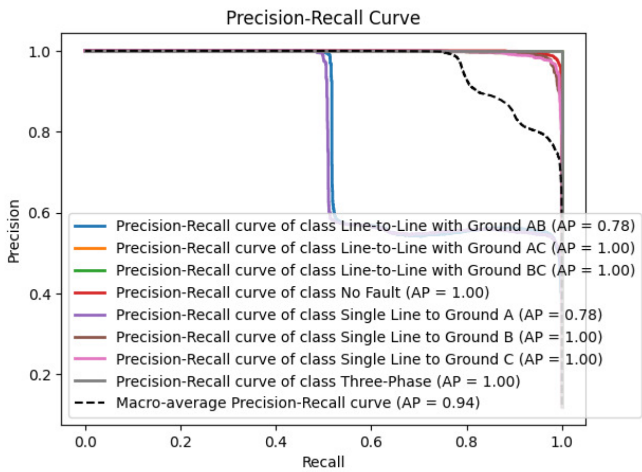

Fig. 6. Precision-Recall curve with RF model.

Figure 7 shows the confusion matrix with RF model, which is a detailed breakdown of predictions versus actual classifications for each fault type. The diagonal represents correct predictions (true positives and true negatives), whereas off-diagonal cells indicate misclassifications. Table I demonstrates the performance analysis for the KNN, DT, RF and logistic regression models. The RF model has the highest Precision, Recall, Accuracy, and F1 − score, making it the best model in this comparison. Logistic regression performs poorly in all metrics except for Precision. Therefore, it would be of little use for this type of dataset, and further tuning may be needed.

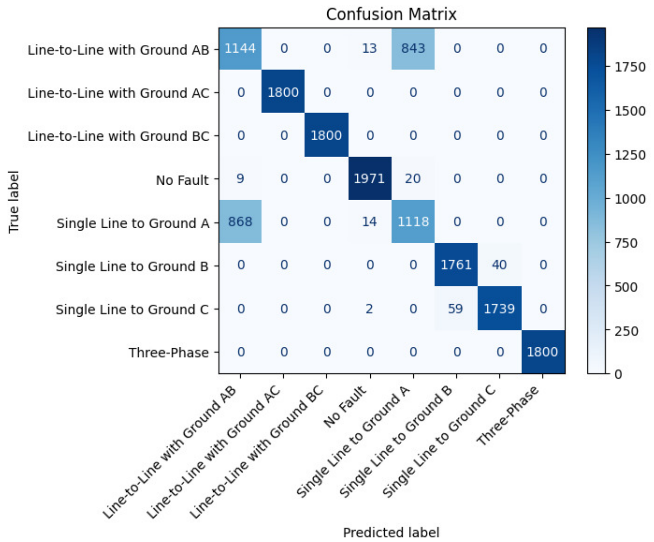

Fig. 7. Confusion matrix with RF model.

TABLE I. PERFORMANCE ANALYSIS

Model	Accuracy	Precision	Recall	F1 − score
KNN	0.830478	0.841939	0.830478	0.832308
DT	0.860076	0.862903	0.860076	0.861160
RF	0.875275	0.875070	0.875275	0.875150
LR	0.517273	0.700223	0.517232	0.515705

Figure 8 shows the comparison of different models on the basis of different matrices. The Accuracy obtained by KNN is 83%, by DT is 86%, by RF is 87% and by LR is 51%. The Precision values are 84%, 86%, 87% and 70% respectively. The Recall values are 83%, 86%, 87% and 51%. The F1 − score values are 83%, 86%, 87% and 51%. The overall performance of RF with respect to the different matrices is high and therefore it is the preferred method.

A. *Comparison of RF with Existing Models*

Existing fault classification models were compared with the proposed RF model to highlight how well the proposed method performs, as shown in Table II. The most widely used method to date for fault classification in transmission lines is the use of Artificial Neural Networks (ANNs). In [13], the amount of data was small and the number of classes considered was 3, with an Accuracy of 84.40%. In [19], ANNs were also used and the number of classes considered was 11. The Accuracy obtained was only 70%. The proposed system was tested on 2000 data points and 7 classes were considered. The Accuracy obtained was 87.52% which is better than all the others.

TABLE II. COMPARISON OF RF WITH OTHER MODELS

Reference	Algorithm	Data (training & testing)	No. of classes considered	Accuracy (%)
Fault Classification (Multiclass)				
[13]	ANN	208 & 44	3	84.40
[19]	ANN	---	11	70.00
Proposed	RF	1400 & 600	7	87.52
Proposed	DT	1400 & 600	7	86.00

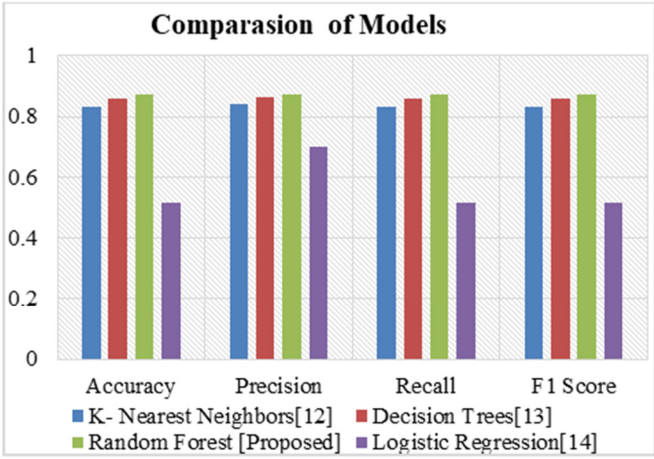

Fig. 8. Comparison of various models on different matrices.

IV. CONCLUSION

In this research, a machine learning algorithm for fault classification in transmission lines was developed, using real-time current and voltage data. The transmission line currents and voltages, acquired using MATLAB/Simulink, were used for training and testing the investigated models. The transmission line currents and voltages were used to classify the faults as a three-phase ground fault, a double line-to-ground fault, and a line-to-ground fault. The classifiers, such as logistic regression, K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF) were considered for training the model and their performance was evaluated using metrics such as Accuracy, F1 − score, Recall and Precision. The results revealed that the RF model performed better than the others. The RF algorithm is computationally efficient, reliable, and improves the classification accuracy compared to other existing methods. MATLAB efficiently handled the data, which allowed preprocessing for both training and testing the model in a simulated environment.

REFERENCES

[1] Md. O. F. Goni *et al.*, "Fast and Accurate Fault Detection and Classification in Transmission Lines using Extreme Learning Machine," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 3, Mar. 2023, Art. no. 100107, https://doi.org/10.1016/j.prime.2023.100107.

[2] H. Rezapour, S. Jamali, and A. Bahmanyar, "Review on Artificial Intelligence-Based Fault Location Methods in Power Distribution Networks," *Energies*, vol. 16, no. 12, Jun. 2023, Art. no. 4636, https://doi.org/10.3390/en16124636.

[3] A. Maheshwari and D. S. Sharma, "Classification of fault identification techniques in power system transmission lines," *Journal of Emerging Technologies and Innovative Research*, vol. 5, no. 1, pp. 89–98, Jan. 2018, http://doi.one/10.1729/IJCRT.17659.

[4] S. Umathe, P. Daigavane, and M. Daigavane, "Improving Fault Classification Accuracy Using Wavelet Transform and Random Forest with STATCOM Integration," *EAI Endorsed Transactions on Energy Web*, vol. 12, Nov. 2025, https://doi.org/10.4108/ew.5950.

[5] M. R. Shadi, M.-T. Ameli, and S. Azad, "A real-time hierarchical framework for fault detection, classification, and location in power systems using PMUs data and deep learning," *International Journal of Electrical Power & Energy Systems*, vol. 134, Jan. 2022, Art. no. 107399, https://doi.org/10.1016/j.ijepes.2021.107399.

[6] B. K. Ponukumati, P. Sinha, M. K. Maharana, A. V. P. Kumar, and A. Karthik, "An Intelligent Fault Detection and Classification Scheme for Distribution Lines Using Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8972–8977, Aug. 2022, https://doi.org/10.48084/etasr.5107.

[7] P. Ilius, M. Almuhaini, M. Javaid, and M. Abido, "A Machine Learning–Based Approach for Fault Detection in Power Systems," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11216–11221, Aug. 2023, https://doi.org/10.48084/etasr.5995.

[8] Md. R. Bishal, S. Ahmed, N. M. Molla, K. M. Mamun, A. Rahman, and Md. A. A. Hysam, "ANN Based Fault Detection & Classification in Power System Transmission line," in *2021 International Conference on Science & Contemporary Technologies*, Dhaka, Bangladesh, 2021, pp. 1–4, https://doi.org/10.1109/ICSCT53883.2021.9642410.

[9] E. Aker, M. L. Othman, V. Veerasamy, I. bin Aris, N. I. A. Wahab, and H. Hizam, "Fault Detection and Classification of Shunt Compensated Transmission Line Using Discrete Wavelet Transform and Naive Bayes Classifier," *Energies*, vol. 13, no. 1, Jan. 2020, Art. no. 243, https://doi.org/10.3390/en13010243.

[10] H. Panahi, R. Zamani, M. Sanaye-Pasand, and H. Mehrjerdi, "Advances in Transmission Network Fault Location in Modern Power Systems: Review, Outlook and Future Works," *IEEE Access*, vol. 9, pp. 158599–158615, 2021, https://doi.org/10.1109/ACCESS.2021.3129838.

[11] P. Jafarian and M. Sanaye-Pasand, "A Traveling-Wave-Based Protection Technique Using Wavelet/PCA Analysis," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 588–599, Apr. 2010, https://doi.org/10.1109/TPWRD.2009.2037819.

[12] M. H. H. Musa, Z. He, L. Fu, and Y. Deng, "A correlation coefficient-based algorithm for fault detection and classification in a power transmission line," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 13, no. 10, pp. 1394–1403, Oct. 2018, https://doi.org/10.1002/tee.22705.

[13] S. R. Fahim, Y. Sarker, O. K. Islam, S. K. Sarker, Md. F. Ishraque, and S. K. Das, "An Intelligent Approach of Fault Classification and Localization of a Power Transmission Line," in *2019 IEEE International Conference on Power, Electrical, and Electronics and Industrial Applications*, Dhaka, Bangladesh, 2019, pp. 53–56, https://doi.org/10.1109/PEEIACON48840.2019.9071925.

[14] P. Pan, R. K. Mandal, and Md. M. Rahman Redoy Akanda, "Fault Classification with Convolutional Neural Networks for Microgrid Systems," *International Transactions on Electrical Energy Systems*, vol. 2022, no. 1, Apr. 2022, Art. no. 8431450, https://doi.org/10.1155/2022/8431450.

[15] P. S. Tembhurne, P. Daigavane, S. Umathe, and K. Pathe, "Fault Classification in Transmission Line Using Soft Computing Tool," in *2024 International Conference on Innovations and Challenges in Emerging Technologies*, Nagpur, India, 2024, pp. 1–6, https://doi.org/10.1109/ICICET59348.2024.10616290.

[16] M. Paul and S. Debnath, "Fault Detection and Classification Scheme for Transmission Lines Connecting Windfarm Using Single end Impedance," *IETE Journal of Research*, vol. 69, no. 4, pp. 2057–2069, May 2023, https://doi.org/10.1080/03772063.2021.1886601.

[17] K. Chen, C. Huang, and J. He, "Fault detection, classification and location for transmission lines and distribution systems: a review on the methods," *High Voltage*, vol. 1, no. 1, pp. 25–33, Apr. 2016, https://doi.org/10.1049/hve.2016.0005.

[18] 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, Jan. 2022, Art. no. 5475, https://doi.org/10.3390/en15155475.

[19] N. A. M. Leh, F. M. Zain, Z. Muhammad, S. A. Hamid, and A. D. Rosli, "Fault Detection Method Using ANN for Power Transmission Line," in *2020 10th IEEE International Conference on Control System, Computing and Engineering*, Penang, Malaysia, 2020, pp. 79–84, https://doi.org/10.1109/ICCSCE50387.2020.9204921.

[20] A. Harish, A. Prince, and M. V. Jayan, "Fault Detection and Classification for Wide Area Backup Protection of Power Transmission Lines Using Weighted Extreme Learning Machine," *IEEE Access*, vol. 10, pp. 82407–82417, 2022, https://doi.org/10.1109/ACCESS.2022.3196769.

[21] D. U. Sarkar, T. Prakash, and S. N. Singh, "Fractional Order PID-PSS Design Using Hybrid Deep Learning Approach for Damping Power System Oscillations," *IEEE Transactions on Power Systems*, vol. 40, no. 1, pp. 543–555, Jan. 2025, https://doi.org/10.1109/TPWRS.2024.3416753.

[22] D. U. Sarkar and T. Prakash, "A Novel Design of Power System Stabilizer via GWO-tuned Radial-basis Function Neural Network for Damping Power Oscillations," in *2022 IEEE 10th Power India International Conference*, New Delhi, India, 2022, pp. 1–6, https://doi.org/10.1109/PIICON56320.2022.10045259.

[23] M. Singh, B. K. Panigrahi and R. P. Maheshwari, "Transmission line fault detection and classification," *2011 International Conference on Emerging Trends in Electrical and Computer Technology*, Nagercoil, India, 2011, pp. 15-22, https://doi.org/10.1109/ICETECT.2011.5760084.