Detection, Classification, and Localization of Transmission Line Faults in Power Systems

Saketha Vasu Deva Charya G
Dept. of CSE, IIIT Nagpur
Nagpur, India
sakethavasudev@gmail.com

Dhruva Narayana K
Dept. of CSE, IIIT Nagpur
Nagpur, India
dhruvakodiadka@gmail.com

Akshith Vermisetti
Dept. of CSE, IIIT Nagpur
Nagpur, India
akshithvermisetti@gmail.com

***Abstract*—This paper addresses the critical issue of transmission line faults that affect the stability and reliability of power systems. These faults often lead to power outages and equipment damage, necessitating quick and accurate detection and classification to minimize system downtime. Traditional techniques, such as impedance-based methods, are commonly used but struggle with noise sensitivity and accuracy in complex fault conditions. Recently, fuzzy logic-based classification methods and machine learning approaches, including artificial neural networks (ANNs) and long short-term memory networks (LSTMs), have been explored to enhance fault detection. These methods offer improved performance but require large datasets for training. This paper presents a real-time power system simulation based on the IEEE 30-bus system using the pandapower library. The simulation incorporates a hybrid fault detection approach combining KNN for initial fault detection and Polynomial Regression for fault classification. The system monitors key parameters such as current, voltage, and power flow, and includes features like real-time load variations, fault injection, and performance tracking. This hybrid model provides an efficient solution for fault detection and classification, supporting practical power grid applications.infrastructures.**

***Index Terms*—K-Nearest Neighbors (KNN), Polynomial Regression, real-time simulation, power grid, IEEE 30-bus system, machine learning, fault localization, artificial neural networks (ANN), long short-term memory (LSTM), fuzzy logic.**

I. Introduction

The protection of transmission lines has always been a critical concern in the field of electrical engineering, as these lines are the backbone of any power system. They are responsible for transferring large amounts of electrical energy from generating stations to load centers, often across vast geographical distances. Being continuously exposed to environmental conditions, these lines are susceptible to faults that can result in serious disruptions to the power supply. A fault on a transmission line not only threatens the reliability of power delivery but also poses a risk to equipment and public safety. With rising demand for uninterrupted and high-quality power, it is essential to detect, classify, and locate faults accurately and promptly to ensure minimal downtime and prevent cascading failures in the grid.

Faults in transmission lines are an inevitable aspect of power systems. These faults may arise from a variety of causes such as lightning strikes, tree contact, conductor sagging, insulation failure, equipment malfunction, or human interference. Some of the most devastating power outages in history, like the 2012 India blackout affecting over 400 million people, or the 2003 North American blackout which disrupted life for nearly 50 million residents, were attributed to transmission faults that went undetected or uncontained. These events underscore the pressing need for reliable and intelligent fault management systems in modern power grids. Traditional systems using protective relays and distance relays can isolate faults, but their accuracy can be compromised by varying operating conditions, fault resistances, and transient disturbances.

From a technical perspective, electrical faults in transmission lines are generally classified into symmetrical and unsymmetrical faults. Symmetrical faults, though rare, involve all three phases equally such as three-phase-to-ground (LLL-G) or three-phase faults (LLL) and result in balanced system disturbance. These are usually severe but easier to analyze. On the other hand, unsymmetrical faults—including single line-to-ground, line-to-line, and double line-to-ground faults are more common and complex to analyze due to the imbalance they introduce in the system. These faults can lead to considerable electrical and mechanical stress, endangering both equipment and human lives. Accurate fault classification is essential for proper relay operation and fault isolation

Advancements in signal processing have introduced powerful techniques such as the wavelet transform, which excels in detecting transient events like faults by analyzing both the time and frequency components of voltage and current signals. This allows for more precise detection and localization, particularly during the brief and dynamic disturbance window. Complementing these methods, fuzzy logic has emerged as a flexible decision-making tool that can handle uncertainties in input signals. The integration of wavelet transforms with fuzzy logic systems has led to intelligent fault classifiers capable of operating under diverse and unpredictable conditions.

The advent of communication technologies like Global System for Mobile Communication (GSM) and Wireless Sensor Networks (WSNs) has revolutionized fault detection strategies. These technologies enable real-time transmission of system health data and geographic fault location through GPS. WSNs use distributed sensors along transmission corridors to track electrical anomalies, and the time differences between detection events are used to triangulate the fault location accurately. This is especially helpful in challenging terrains or remote regions where manual inspection is time-consuming or impractical.

In recent years, artificial intelligence (AI) and machine learning (ML) have played an increasingly vital role in power system monitoring and automation. These technologies offer advanced capabilities such as pattern recognition, anomaly detection, and predictive maintenance. AI models can learn from historical fault data and improve fault detection accuracy over time. Applications span from demand forecasting and energy optimization to smart microgrids and real-time grid management. Traditional methods often fall short in complex, high-impedance fault scenarios or evolving grid topologies. ML-based methods, by contrast, offer faster response times, adaptability, and the ability to work under noisy or uncertain data environments, making them indispensable tools in the smart grid era.

Another essential component in fault management is the fault locator. Fault locators are vital for quickly identifying the fault point, especially in systems where physical inspection is impractical or inefficient. Their importance is amplified in rugged terrains, vast transmission corridors, or systems involving multiple jurisdictions. Even when the system is automatically restored, fault locators help in pinpointing the origin be it from bushfires, vandalism, or hardware failure thus enabling maintenance teams to arrive at the site before the evidence is lost and take corrective measures efficiently.

Ensuring the protection of transmission lines through early fault detection, accurate classification, and precise localization is fundamental to maintaining the safety, reliability, and economic operation of power systems. As the demand for smarter and more resilient power infrastructure grows, the integration of AI, advanced signal processing, and communication technologies is poised to transform fault analysis into a proactive, data-driven, and intelligent process.

II. Literature Review

A. Conventional Approach

1) Single-Terminal Impedance-Based Fault Location Methods: Single-terminal impedance-based methods are among the most widely adopted conventional techniques for fault location in transmission lines. These methods are preferred in practice due to their low cost, simple implementation, and independence from communication infrastructure. According to [3], these techniques estimate the fault location by calculating the apparent impedance seen from one end of the line and comparing it with the known impedance per unit length of the transmission line. The relationship is given by:

$$d = \frac{Z_f}{Z_{\text{line}}} \tag{1}$$

where Z_f is the measured impedance during the fault and Z_{line} is the known line impedance per unit length. Despite their simplicity, these methods are prone to several inaccuracies. One major source of error arises from the fault resistance, particularly in ground faults, which causes the measured impedance to deviate from the actual value, leading to significant location errors. Additionally, the load current flowing during the fault affects the accuracy of the estimate due to the "reactance effect." Other critical factors influencing the error include incorrect fault type identification, uncertainty in zero-sequence impedance values, and modeling inaccuracies, such as neglecting mutual coupling or representing untransposed lines as transposed. For instance, a 20% deviation in the zero-sequence impedance can result in up to a 15% error in the calculated fault location. The estimation error is generally expressed as a percentage difference between the estimated and actual fault distances:

$$\text{Error } (\%) = \frac{D_{\text{estimated}} - D_{\text{actual}}}{D} \times 100 \tag{2}$$

where $D_{\text{estimated}}$ represents the estimated distance to the fault, D_{actual} denotes the actual distance to the fault, and D is the total length of the transmission line.

2) Takagi Method: The Takagi method is a single-ended impedance-based fault location technique that enhances accuracy by minimizing the effect of fault resistance, which commonly introduces errors in traditional methods. It uses both the pre-fault and fault voltage and current waveforms to distinguish between load current and fault current components. By doing so, it isolates the fault-generated component and computes a more accurate estimate of the fault distance. This approach is particularly effective for high-resistance faults and is widely implemented in modern numerical relays. However, its accuracy can still be influenced by line modeling inaccuracies and the quality of fault type identification.

3) Travelling Wave-Based Method: The travelling wave-based fault location method relies on detecting the high-frequency transient waves generated at the point of fault. These waves propagate along the transmission line and are captured at one or both ends of the line using high-speed recording devices. The time difference between the arrival of incident and reflected waves is used to estimate the location of the fault. This technique offers high precision and fast fault location, making it suitable for long-distance and extra-high voltage transmission lines. However, it requires fast data acquisition systems and precise time synchronization, such as GPS, which can increase implementation costs.

B. Fuzzy Logic-Based Fault Classification

Fuzzy Logic (FL) is based on the principles of fuzzy set theory and the concept of fuzzy membership, which is commonly found in natural (e.g., spoken) language. FL uses imprecision to provide robust solutions to problems, allowing for more flexible decision-making under uncertainty. A fuzzy set represents a collection of elements, each with a degree of membership, indicated by the membership function μ, that defines the extent to which an element belongs to the set.

The notation for fuzzy sets is as follows: for a member x of a discrete set, with membership μ, we write μ/x, meaning that x is a member of the set to a degree of μ.

For example, a fuzzy set A could be expressed as:

$$A = \mu_1/x_1 + \mu_2/x_2 + \mu_3/x_3 + \cdots + \mu_n/x_n \quad (3)$$

Where $x_1, x_2, x_3, \ldots, x_n$ are the elements of the set and $\mu_1, \mu_2, \mu_3, \ldots, \mu_n$ are their corresponding membership values.

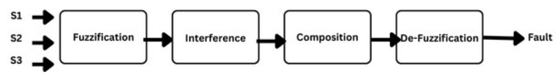

Fig. 1. Fuzzy System.

C. Deep Learning and MIL Techniques

1) Artificial Neural Network (ANN): Artificial Neural Networks (ANNs) have emerged as a powerful tool in the domain of fault detection and localization in power systems due to their capability to model complex nonlinear relationships and learn from data. Inspired by the functioning of biological neurons, ANNs consist of interconnected processing units (neurons) that receive, process, and transmit information using weighted inputs and activation functions.

Each neuron processes the input vector $\mathbf{I} = [i_1, i_2, \ldots, i_n]$ by multiplying it with corresponding weights $w_1, w_2, \ldots, w_n$ and computing a weighted sum S. This summation is then passed through a nonlinear activation function (AF) such as sigmoid, tanh, or ReLU, yielding the neuron's output Y.

However, the primary challenge with ANN-based approaches lies in the selection of optimal network parameters and weights. Improper configuration can lead to issues like overfitting, underfitting, or poor generalization. Additionally, the requirement for substantial labeled data for training adds to the complexity of the method.

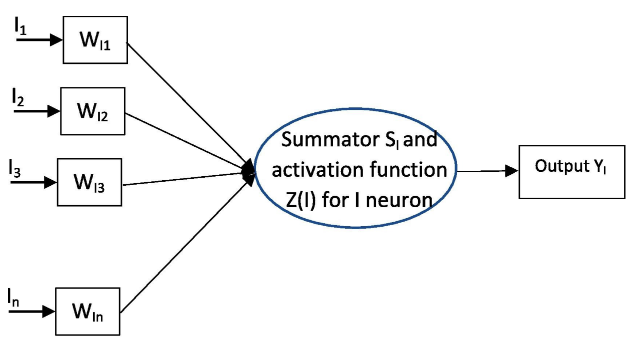

Fig. 2. Deep learning model used for fault detection and localization.

2) Long Short-Term Memory (LSTM): The system structure, known as the Recurrent Neural Network (RNN), is applied to analyze time series data. Its unique feature is that the preceding output information, along with the current input, influences the present output. Although this network has certain benefits, gradient vanishing or explosion is more likely to occur when the dependencies in the data are far apart from the current prediction target.

A variation of the RNN is the Long Short-Term Memory (LSTM) network. Through a mechanism known as a gate unit, it can perform self-regulating weight adjustments, successfully resolving issues related to gradient vanishing and explosion. As a result, LSTM is more suited for handling time series data effectively.

LSTM networks are widely used for dynamic spatial forecasting problems. They learn through multiple layers, which helps in recognizing complex temporal patterns for improved performance. LSTM has the ability to retain important information for extended periods and discard it when it is no longer relevant for prediction.

There are three types of LSTM architectures:

1) **Forward Pass**
2) **Backward Pass**
3) **Bidirectional**

III. Model Comparison and Performance Evaluation

To evaluate the effectiveness of various machine learning algorithms for transmission line fault detection and classification, several models were implemented and compared. The models tested include:

- Linear Regression
- Logistic Regression
- Polynomial Regression
- Naive Bayes
- Decision Tree
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)

The primary objective of this evaluation is to determine whether a fault has occurred in the transmission line system. Each model was trained using a preprocessed dataset tailored for the classification of fault and non-fault scenarios.

TABLE I
Detection Task: Model Accuracy and Error

Model	Accuracy (%)	Error (%)
Linear Regression	1.79	24.38
Logistic Regression	73.84	26.16
Polynomial Regression	86.12	3.45
Naive Bayes	98.06	1.94
Decision Tree	99.47	0.53
SVM	98.21	1.79
KNN	99.24	0.76

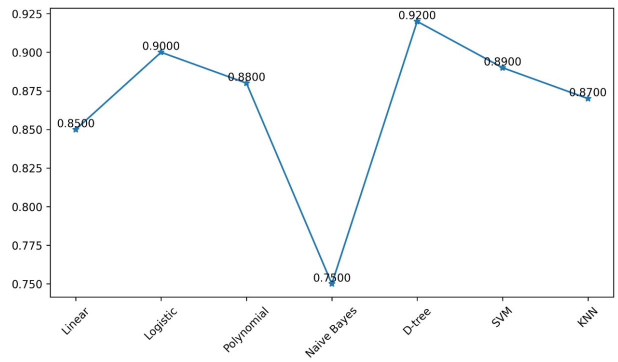

Fig. 3. Detection Task: Model Comparison Based on Accuracy

TABLE II
CLASSIFICATION TASK: MODEL ACCURACY AND ERROR

Model	Accuracy (%)	Error (%)
Linear Regression	3.35	17.30
Logistic Regression	32.52	42.66
Polynomial Regression	94.51	0.01
Naive Bayes	79.69	2.11
Decision Tree	86.63	0.31
SVM	31.68	41.22
KNN	82.47	0.98

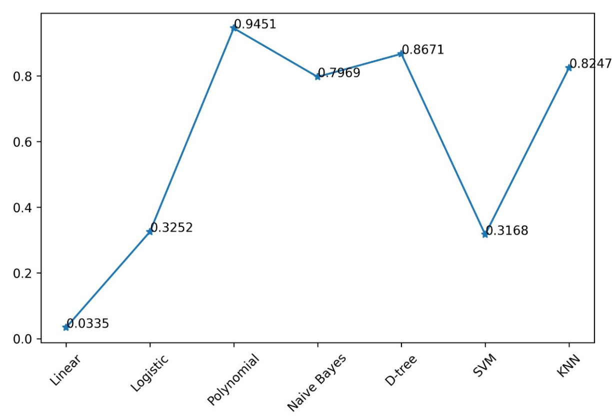

Fig. 4. Classification Task: Model Comparison Based on Accuracy

IV. Observation

Among the various machine learning models tested for fault detection and classification in a three-phase transmission line system, the K-Nearest Neighbors (KNN) model demonstrated the highest performance in fault detection, achieving an accuracy of 99.24% with minimal error.

For fault classification, the Polynomial Regression model outperformed the others with an impressive accuracy of 94.51%. These results indicate that different models excel in different aspects of the fault analysis pipeline.

The success of KNN in fault detection can be attributed to its ability to capture local patterns in the data, whereas Polynomial Regression effectively models the complex non-linear relationships required for accurate classification. These models are therefore well-suited for practical deployment in intelligent power system fault monitoring applications.

V. Fault Detection and Classification Framework

The proposed methodology integrates a real-time, data-driven framework for fault detection and classification in power transmission systems. The framework consists of a two-stage machine learning pipeline, wherein the first stage performs binary fault detection using the K-Nearest Neighbors (KNN) algorithm, followed by multiclass fault classification via Polynomial Logistic Regression. This design ensures a robust and accurate mechanism for intelligent fault monitoring in a dynamic grid environment.

A. Fault Detection Using K-Nearest Neighbors (KNN)

Fault detection is formulated as a binary classification problem, where the KNN algorithm is employed to distinguish between faulty and non-faulty states. The input feature vector is composed of real-time electrical measurements, including phase-wise line currents (I_1, I_2, I_3) and bus voltages (V_1, V_2, V_3), collected from the power system simulation environment.

The KNN classifier operates by evaluating the Euclidean distance between the real-time measurement vector and a labeled training dataset. Based on the k closest neighbors, the model determines the class label—either "Fault" or "No Fault"—through majority voting. This algorithm is chosen for its non-parametric nature, low training overhead, and effectiveness in detecting anomalous patterns in high-dimensional measurement data.

B. Fault Classification Using Polynomial Logistic Regression

Upon detection of a fault, the classification module is invoked to identify the specific type of fault. This stage employs a supervised learning approach using a multinomial Logistic Regression model, enhanced through polynomial feature expansion to capture nonlinear relationships between system parameters.

The feature vector for classification retains the same structure: I_1, I_2, I_3, V_1, V_2, V_3. A polynomial transformation of degree two is applied, introducing interaction and squared terms (e.g., I_1^2, $V_2 \cdot I_3$, V_3^2), thereby enriching the feature space to model complex fault dynamics. The expanded feature set is input to a multinomial logistic regression classifier configured with the softmax function to compute posterior probabilities across multiple fault categories.

The classifier is trained on a labeled dataset representing various fault scenarios, including:

- Single Line-to-Ground (LG)
- Line-to-Line (LL)
- Double Line-to-Ground (LLG)
- Three-Phase (LLL)
- No Fault (NF)

A model pipeline is constructed to incorporate both preprocessing and classification, ensuring consistency during training and inference. Once trained, the model is serialized and deployed for real-time prediction.

C. Real-Time Integration and Execution Flow

During real-time simulation, measurement data is continuously streamed from a pandapower-based power system model. The KNN-based detection module evaluates each timestep's data and flags potential faults. Upon detection, the data is immediately passed to the classification module for precise fault type identification.

The classification output includes both the predicted fault class and a confidence score, aiding in probabilistic decision-making and fault severity assessment. This modular architecture ensures high interpretability and scalability, and it allows independent updates to detection and classification modules without affecting the overall pipeline.

References

[1] A. Q. Khan, Q. Ullah, M. Sarwar, S. T. Gul, and N. Iqbal, "Transmission line fault detection and identification in an interconnected power network using phasor measurement units," *IFAC-PapersOnLine*, vol. 51, no. 24, pp. 1356–1363, 2018.

[2] P. K. Shukla and K. Deepa, "Deep learning techniques for transmission line fault classification – A comparative study," *Ain Shams Engineering Journal*, vol. 15, no. 2, p. 102427, 2024. doi:10.1016/j.asej.2023.102427.

[3] A. Jadhavar, S. Gavali, A. Chavan, and A. P. Sayyad, "Transmission line fault detection system," *International Journal of Engineering Applied Sciences and Technology*, vol. 8, no. 1, pp. 289–292, 2023. [Online]. Available: http://www.ijeast.com

[4] N. Q. Minh, N. T. Khiem, and V. H. Giang, "Fault classification and localization in power transmission line based on machine learning and combined CNN-LSTM models," *Energy Reports*, vol. 12, pp. 5610–5622, 2024. doi:10.1016/j.egyr.2024.11.061.

[5] M. M. I. Hashim, H. W. Ping, and V. K. Ramachandaramurthy, "Impedance-based fault location techniques for transmission lines," in *TENCON 2009 - 2009 IEEE Region 10 Conference*, 2009. doi:10.1109/TENCON.2009.5396253.

[6] P. Choudhary, M. K. Bhaskar, and M. Parihar, "Fuzzy logic-based fault classification for transmission line analysis," *International Research Journal of Engineering and Technology (IRJET)*, vol. 10, no. 10, p. 287, 2023. [Online]. Available: https://www.irjet.net

[7] H. H. Goh, S. Y. Sim, A. Shaykh, M. H. Kabir, C. W. Ling, Q. S. Chua, and K. C. Goh, "Transmission line fault detection: A review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 8, no. 1, pp. 199–205, Oct. 2017. [Online]. Available: https://doi.org/10.11591/ijeecs.v8.i1.pp199-205

[8] N. Saravanan and A. Rathinam, "A comparative study on ANN based fault location and classification technique for double circuit transmission line," *2012 Fourth International Conference on Computational Intelligence and Communication Networks*, pp. 824–830, 2012. [Online]. Available: https://doi.org/10.1109/CICN.2012.15

[9] P. Avagaddi, B. Edward, and K. Ravi, "A review on fault classification methodologies in power transmission systems: Part II," *Journal of Electrical Systems and Information Technology*, vol. 5, 2018. [Online]. Available: https://doi.org/10.1016/j.jesit.2016.10.003

[10] 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*, pp. 15–22, 2011. [Online]. Available: https://doi.org/10.1109/ICETECT.2011.5760084

[11] M. Egan, J. Thapa and M. Benidris, "Machine Learning Using High Precision Data for Fault Location," *2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pp. 1–5, 2022. [Online]. Available: https://doi.org/10.1109/PMAPS53380.2022.9810580.

[12] A. Jamehbozorg and S. M. Shahrtash, "A Decision Tree-Based Method for Fault Classification in Double-Circuit Transmission Lines," *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2184–2189, Oct. 2010. [Online]. Available: https://doi.org/10.1109/TPWRD.2010.2050911.

[13] M. Jayabharata Reddy and D. K. Mohanta, "A wavelet-fuzzy combined approach for classification and location of transmission line faults," *International Journal of Electrical Power and Energy Systems*, vol. 29, no. 9, 2007. [Online]. Available: https://doi.org/10.1016/j.ijepes.2007.05.001.