GridWaveLoc: A Fault Location Algorithm Integrating Fault Traveling Wave Distribution and Network Dependency Graphs for Transmission Grids

Dong Han*, Ke Wang, Xiaoguang Wang, Yue Li, Zhihao Yang
Branch of Power Dispatching Control Inner Mongolia Power (Group) Co. Ltd, Hohhot 010010, Inner Mongolia, China
E-mail: nm_handong@163.com, Email: 18347933736@163.com, nmdkzx_wxg@163.com, liyuedsg@163.com
eEmail: yzh4225638159@163.com
*Corresponding author

Keywords: fault localization, missing values imputation, encoding, normalization, euclidean distance

Recieved: January 21, 2025

Accurate fault localization in transmission grids is essential for reducing downtime and maintaining power system stability. Conventional fault location techniques frequently struggle with the intricacy of contemporary transmission grids, which have intricated dependencies and variable fault characteristics. Problem Statement: Previous fault location techniques used limited indicators, like wave arrival times or impedance-based measurements, while ignoring amplitude variations and propagation speed. Numerous also use oversimplified network models that assume a uniform topology and ignore grid node dependencies. These constraints result in delays, localization errors, and ineffective grid restoration. The difficulty is to combine fault wave propagation and a realistic network structure to enhance accuracy and response time. Objective: The objective of this research is to enhance fault localization accuracy and response time in transmission grids by using fault traveling wave distribution and network dependency graphs. To accomplish this, the research creates GridWaveLoc, a fault location algorithm that incorporates wave propagation characteristics into the grid's dependency structure, resulting in quicker fault detection and increased grid reliability. Methodology: The GridWaveLoc algorithm executes real-time data such as fault type, wave arrival time, amplitude, transmission line characteristics, and network load. Mean and mode imputation for missing values, label encoding for categorical variables, and Min-Max normalization for continuous features all fall under data preprocessing. The algorithm uses fault wave propagation times and network dependency graphs to narrow down possible fault locations. The Euclidean distance technique is employed to detect the nearest grid node to the fault origin, guaranteeing accurate fault location prediction. Results: Experiments were performed on the Transmission Fault Localization Dataset, which contained 11 features and 2000 records, to evaluate three types of faults: short-circuit, open-circuit, and ground faults. GridWaveLoc obtains 98.5% accuracy, surpassing the Traveling Wave Method (92.3%), the Impedance-Based Method (89.5%), and Artificial Neural Networks (85.7%). GridWaveLoc also has the lowest mean absolute error (MAE) of 0.12 km and root mean square error (RMSE) of 0.15 km, substantially enhancing fault localization precision compared to other techniques. These results emphasize the potential for real-time fault detection in massive transmission networks. Conclusion: The GridWaveLoc employs a novel approach to fault location in transmission grids by integrating fault traveling wave evaluation and network dependency data. This technique improves the dependability and effectiveness of power grid functions by offering a solid solution for real-time fault localization.

Povzetek: Članek predstavi nov algoritem GridWaveLoc, ki združuje potovalne valove in graf odvisnosti omrežja za natančno lokalizacijo napak v elektroprenosnih omrežjih v realnem času.

1 Introduction

The dependable function of transmission grids is a pillar of contemporary power systems, where guaranteeing uninterrupted electricity supply is critical. Faults in transmission grids, whether caused by weather, equipment failure, or human error, can cause power outages, economic losses, and compromised grid stability [1]. Efficient fault localization is a critical step towards reducing downtime and restoring regular operations [2]. Conventional fault location techniques have long served the industry, but the growing intricacy of grid networks, with interconnected nodes and differing transmission line characteristics, requires more sophisticated solutions. Integrating fault-traveling wave evaluation and network dependency graphs opens up a new avenue for tackling these difficulties.

Several fault location methods have been proposed, containing impedance-based techniques [3], time-domain reflectometry [4], and AI-driven methods [5]. These methods are mainly based on characteristics like line impedance, fault current magnitudes, and historical fault data. While these methods offer basic functionality, their efficacy is reduced in highly interconnected grids with intricate topologies. Furthermore, the adoption of wavelet transforms, and other signal processing methods has shown potential, but the scalability and real-time applicability of these techniques are still restricted.

Despite their utility, conventional fault location techniques have major disadvantages. Impedance-based methods are prone to inaccuracies when load variations or faults happen in numerous places at the same time. Time-domain techniques, while accurate in smaller grids, struggle to meet the computational needs of today's networks. Additionally, numerous previous solutions fail to account for the complex dependencies within grid networks, like node connectivity and load distribution, which are essential for precise fault localization. The absence of a combined strategy that combines fault wave evaluation and network characteristics frequently results in delayed fault identification and ineffective grid restoration.

To tackle the drawbacks of current techniques, this study introduces GridWaveLoc, an innovative algorithm that improves fault location accuracy and speed. The GridWaveLoc algorithm integrates fault-traveling wave distribution evaluation with network dependency graphs, resulting in an extensive framework for real-time fault localization. GridWaveLoc precisely identifies fault locations in intricate grid settings by utilizing wave propagation data and integrating the grid's structural dependencies.

The GridWaveLoc algorithm uses a multi-step procedure to localize faults. Initially, raw fault data, such as wave arrival times, amplitudes, and network conditions, is preprocessed with mean and mode imputation for missing values, label encoding for categorical features, and Min-Max normalization for continuous variables. Fault traveling waves are evaluated to determine propagation times, and the results are combined with a network dependency graph that depicts the grid's structure. GridWaveLoc uses the Euclidean distance technique to identify the grid node nearest to the fault origin, narrowing down the fault location with minimum error.

This paper renders the subsequent contributions:

- Introduces the GridWaveLoc algorithm, which incorporates fault traveling wave evaluation and network dependency graphs for accurate fault localization.
- Creates a resilient preprocessing framework to manage real-time grid data effectively.
- Shows the algorithm's better performance by comprehensive simulations, emphasizing its scalability and accuracy.
- Offers insights into the application of dependency graphs in improving fault location techniques.

This study intends to answer the following important research questions:

- Can the combination of network dependency graphs and fault traveling wave distribution enhance fault localization accuracy over conventional techniques?
- How does GridWaveLoc manage missing data, and how does it compare to conventional preprocessing techniques?
- What are the important benefits of GridWaveLoc in terms of True Positive Rate (TPR), specificity, and error reduction compared to previous methods?

To tackle these research questions, this paper formulates the following hypotheses:

- H1: The incorporation of network dependency graphs and fault traveling wave distribution enhances fault localization accuracy compared to conventional techniques.
- H2: The adaptive preprocessing strategy in GridWaveLoc offers better handling of missing data, decreasing localization errors.
- H3: GridWaveLoc attains higher TPR, specificity, and lower MAE/RMSE than modern fault localization techniques.

This study aims to enhance fault location techniques by proposing a hybrid method that uses both traveling wave characteristics and network structure. The main objective is to attain precise, effective, and real-time fault localization in transmission grids.

The novelty of GridWaveLoc lies in its dual-layered method, which incorporates wave analysis with network graph dependency modeling. Unlike previous techniques, it tackles the dynamic and interconnected nature of contemporary grids, guaranteeing high accuracy and dependability.

GridWaveLoc can be used to handle transmission grids, monitor faults in real-time, and restore power systems. Its application includes both small-scale and massive grid infrastructures, rendering it versatile for various grid contexts.

The rest of this paper is organized as follows: Section 2 discusses related works on fault localization. Section 3 discusses the methodology, which includes the GridWaveLoc algorithm and preprocessing methods. Section 4 includes experimental findings and a comparative evaluation. Section 5 discusses the result's implications and recommends areas for enhancement. Finally, Section 6 summarizes the paper and suggests directions for future research.

2 Related work

The detection and localization of faults in power systems have been extensively researched, as precise fault detection is critical for retaining system dependability, reducing downtime, and enhancing maintenance. Over time, different approaches based on the idea of Traveling Wave (TW) propagation have been created. Jiménez-Aparicio et al. [6] investigated factors that influence TW propagation, including fault distance, type, and interactions with system components including regulators and capacitor banks. Utilizing the Stationary Wavelet Transform (SWT) and Parseval's Energy Theorem, the research emphasized how these factors influence TW energy propagation and proposed their inclusion in TW-based security systems.

Similarly, Hung [7] compared conventional impedance-based fault location techniques, such as simple reactance and Takagi models, on high-voltage transmission lines. The research used simulations on a 220kV transmission line to show the significance of taking into account different kinds of faults and resistances when enhancing location accuracy. Huo et al. [8] improved the TW technique for overhead-cable hybrid transmission lines by offering an energy-based fault location technique that quantifies TW energy versus fault distance. This method proved efficient in intricate line structures, as verified by the PSCAD/EMTDC simulation tool.

Furthermore, Maritz et al. [9] introduced a graph-based strategy that uses metric dimensions, and vertex covers to maximize TW detector placement. Their offline and online algorithms, evaluated on the IEEE 30-bus system, provided a new approach to using graph theory for fault detection in intricate grids. Panahi et al. [10] examined contemporary fault location techniques, ranging from distance relays to artificial intelligence, highlighting their importance in tackling environmental and structural issues in transmission networks. The review provided an extensive comparison of techniques, emphasizing their advantages and disadvantages in different situations.

Furthermore, Yu et al. [11] developed a fault localization method that employs the Northern Goshawk Optimization (NGO) algorithm to improve Variational Mode Decomposition (VMD) in fault signal processing. This method efficiently tackled inaccuracies in mode decomposition by improving key parameters and using Hilbert transforms to enhance TW detection. Gonzalez et al. [12] applied fault location techniques to renewable energy-integrated systems, employing graph theory to reduce errors under noisy conditions. Their methodology showed enhanced precision compared to traditional impedance-based methods.

Other studies, like Prabakar et al. [13], concentrated on Fault detection and localization using traveling waves, whereas Wang et al. [14] developed electromagnetic transient convolution techniques to deal with frequency-dependent parameters and lossy grounds. Finally, Dashtdar et al. [15] proposed Review of fault location methods in distribution grids. Table 1 shows the summary table.

Table 1: Summary table

Reference No	**Objective**	**Methodology**	**Result**	**Limitations**
[6] Jiménez-Aparicio et al. (2022)	To detect important factors impacting Traveling Wave (TW) propagation in power distribution systems.	Utilizes Karrenbauer transform, Stationary Wavelet Transform (SWT), and Parseval's Energy theorem for signal processing and energy examination.	Distance, fault type, and presence of regulators significantly impact TW propagation.	Limited to a simplified distribution system with two safety zones.
[7] Hung (2022)	To compare fault location approaches in high voltage power transmission lines.	Assesses impedance-based approaches (simple reactance, Takagi, modified Takagi, double-end) utilizing MATLAB/Simulink simulations.	Shows efficiency of each method in fault distance estimation.	Research is limited to one transmission line (Quy Nhon-Tuy Hoa) in Vietnam.
[8] Huo et al. (2022)	To introduce a fault location technique for hybrid transmission lines using traveling wave energy.	Creates a mapping relationship between TW energy and fault location utilizing S-transform and attenuation evaluation.	Offers an accurate technique for fault location in 110 kV hybrid transmission lines.	Accuracy relies on accurate attenuation characteristics modeling.
[9] Maritz et al. (2021)	To create a traveling wave-based fault location	Transforms power grid into a graph, utilizes metric dimension and	Novel method allows effective fault location on complex grids.	Needs transformation of power networks into graph-based models.

	tactic utilizing graph theory.	vertex covers for fault identification.		
[10] Panahi et al. (2021)	To review contemporary fault location methods in power systems.	Compares different techniques (distance relay, TW, AI, time reversal, impedance) with benefits and drawbacks.	Offers a extensive review and suggestions for future study.	Lacks experimental validation of proposed suggestions.
[11] Yu et al. (2024)	To enhance traveling wave fault localization utilizing NGO-VMD algorithm.	Improves VMD with NGO algorithm, applies Hilbert transform for accurate TW identification.	Attains enhanced fault location accuracy and robustness.	Performance relies on correct choice of VMD parameters.
[12] Gonzalez et al. (2024)	To introduce a graph-theory-based fault location technique for systems with renewable energy sources.	Utilizes graph theory and Kirchhoff's laws to estimate fault distance under changing conditions.	Attains high accuracy with errors below 0.48% for various fault types.	Requires validation on larger, more complex grids.
[13] Prabakar et al. (2021)	Fault detection and localization using traveling waves.	High-speed data acquisition and signal processing.	Improved accuracy over impedance-based methods.	Performance drops in high-noise conditions.
[14] Wang et al. (2024)	To introduce a fault location technique using electromagnetic transient convolution.	Utilizes phase-mode transformation and aerial mode transients for loss minimization.	Offers precise fault localization under various network conditions.	Accuracy reduces with raising fault distance.
[15] Stefanidou-Voziki et al. (2022)	Review of fault location methods in distribution grids.	Comparative analysis of impedance, traveling wave, and ML-based approaches.	Identified strengths and weaknesses of each method.	No experimental validation; high sensor requirements.

2.1 Research gap

The increasing intricacy of contemporary power grids has revealed flaws in conventional fault location techniques, which frequently make oversimplified assumptions like uniform wave propagation speeds and static grid conditions. These methods fail to account for key factors like grid node dependencies, fault-induced wave dynamics, and various fault scenarios, leading to inaccurate fault identification and location. Furthermore, previous methods frequently overlook critical features such as wave amplitude, network load, and time to isolation, restricting their efficacy in capturing the complex nature of faults. The absence of resilient data preprocessing, comprising correct handling of missing values, normalization, and encoding, further aggravates these difficulties, resulting in inconsistent fault evaluation in practical applications.

The GridWaveLoc Algorithm fills in these gaps by combining sophisticated fault wave analysis, feature-rich preprocessing, and adaptive fault location prediction. It uses methods like imputation, Min-Max normalization, and label encoding to guarantee data consistency, as well as wave dynamics and network dependency graphs to capture the effect of faults on interconnected grid nodes. By integrating normalized Euclidean distance computations, the algorithm accurately detects fault locations while accounting for dynamic grid conditions. These advancements render GridWaveLoc an effective and scalable solution, tackling the pressing requirement for precise and effective fault location in contemporary power systems.

3 Methodology

The GridWaveLoc Algorithm was carefully designed to integrate sophisticated fault wave evaluation methods with a thorough comprehension of transmission grid dependencies to precisely locate faults. The methodology includes data preprocessing, fault wave evaluation, network dependency graph construction, and fault location prediction utilizing Euclidean distance. Algorithm 1 shows the proposed GridWaveLoc Algorithm.

Algorithm 1: GridWaveLoc

Input	:	**Dataset with features:** Fault Type, Wave Arrival Time, Wave Amplitude, Transmission Line Length, Voltage Level, Distance from Source, Grid Node Dependency, Wave Propagation Speed, Time to Isolation, and Network Load.
Output	:	Predicted Fault Location
Step 1	:	**Data preprocessing:** • Impute missing values using mean for continuous features and mode for categorical features. • Encode categorical attributes (Fault Type, Grid Node Dependency) utilizing Label Encoding. • Normalize continuous attributes (Wave Arrival Time, Wave Amplitude, etc.) utilizing Min-Max Normalization.
Step 2	:	**Fault wave evaluation:** • Estimate wave propagation time utilizing Wave Arrival Time and Wave Propagation Speed. • Categorize fault type utilizing Wave Amplitude thresholds.
Step 3	:	**Network dependency incorporation:** • Build a Network Dependency Graph utilizing Grid Node Dependency and Distance from Source.
Step 4	:	**Fault localization with euclidean distance:** • Calculate the Euclidean distance between fault features (normalized) and grid nodes. • Detect the grid node with the minimum distance to predict the fault location.
Step 5	:	Return Predicted Fault Location

Below, each step is described in greater detail, integrating mathematical principles, logical explanations, and illustrative scenarios to guarantee clarity and depth.

3.1 Data preprocessing

Data preprocessing is an essential step in guaranteeing that the input data is clean, consistent, and prepared for precise fault detection. Because raw grid data frequently contains missing values, inconsistent formats, or outliers, this step tackles these problems systematically. Asked

Handling missing values:

Missing data is a common problem in transmission grid datasets, and it can occur because of sensor failures, communication lags, or data transmission errors. The algorithm employs imputation methods customized to the kind of data:

Mean imputation is used for continuous variables, such as wave arrival time and transmission line length. To keep the dataset statistically balanced, missing values are replaced with the average of observed values:

$$Imputed\ Value = \frac{\sum_{i=1}^{n} X_i}{n} \quad (1)$$

Where, X_i represents observed values and n represents the total number of observed values. For instance, if the recorded Wave Arrival Times are 0.02s,0.03s,0.04s, and one value is missing, the imputed value becomes (0.02+0.03+0.04)/3=0.03s.

Mode imputation is employed for categorical variables such as fault type (for example, "short-circuit," "ground fault"). This replaces missing values with the most often occurring category, guaranteeing consistency while avoiding bias.

Encoding categorical variables:

Categorical variables, like Fault Type, must be converted to numerical form before being used in machine learning algorithms. Label encoding assigns a unique numerical value to each category:

Fault Type: Short-circuit→0,Open-circuit→1,Ground Fault→2

(2)

This step guarantees that algorithms interpret the data quantitatively while maintaining categorical relationships.

Min-Max normalization:

Continuous features, like Wave Arrival Time and Wave Amplitude, frequently have different scales, creating a bias in the fault location procedure. To avoid this, the algorithm uses Min-Max Normalization, scaling all numerical data to the range [0, 1]:

$$Normalized\ Value = \frac{Value - Min\ Value}{Max\ Value - Min\ Value} \quad (3)$$

For instance, if the Wave Arrival Time ranges between 0.01 and 0.1 seconds and the observed value is 0.03 seconds: Normalized Value = (0.03−0.01) / (0.1−0.01) = 0.222. This guarantees that no feature dominates the fault-finding procedure because of its magnitude, allowing the algorithm to fairly consider all features.

Transmission line characteristics and network load are continuous variables that are treated accordingly during the data preprocessing step. These features are normalized using Min-Max, ensuring that they are scaled within the range [0,1] to avoid bias in fault localization because of magnitude differences. Label encoding is used to encode categorical variables like fault type. Each category is allocated a numerical value (for example, short-circuit → 0, open-circuit → 1, ground fault → 2). This method guarantees that both continuous and categorical attributes are correctly prepared for the fault detection algorithm.

3.2 Fault wave analysis

The GridWaveLoc Algorithm is based on traveling wave evaluation, which provides important information regarding fault characteristics and their likely location within the grid.

Fault wave propagation analysis:

When a fault happens, it causes a traveling wave that spreads throughout the transmission network. The algorithm calculates the fault's distance from the wave's origin by capturing the time it takes for the wave to travel between nodes. The propagation time is computed with:

$$Propagation\ Time = \frac{Distance\ from\ Source}{Wave\ Propagation\ Speed} \quad (4)$$

For instance, if the Distance from the Source is 50 km and the Wave Propagation Speed is 3000 m/s, the propagation time is: Propagation Time = (50×1000) / 3000 = 16.67 seconds.

The factor 1000 is employed to transform kilometers to meters, guaranteeing that the units match the wave propagation speed in meters per second. This computation allows the algorithm to narrow down the fault's possible location by comparing propagation times across numerous nodes.

Fault type classification:

The Wave Amplitude is an important indicator for fault classification, with thresholds determined by an empirical evaluation of historical grid fault data and industry standards. Extensive assessment of previous fault occurrences and waveform characteristics established distinct amplitude ranges for various fault types:

- Short-circuit faults typically exhibit amplitudes exceeding 200 mV, using high current surge patterns observed in practical incidents.

- Open-circuit faults generally fall within the 100–200 mV range, as these faults exhibit moderate disturbances in voltage waveforms.

- Ground faults often register amplitudes below 100 mV, a result of lower energy dissipation into the ground compared to short-circuits.

These thresholds were validated using simulated fault injection in grid models and cross-referenced with recorded field data to ensure their suitability for fault classification. For example, if the observed amplitude is 220 mV, the fault is classified as a short-circuit, and subsequent calculations are directed toward localizing the fault within the grid.

3.3 Network dependency graph construction

The Network Dependency Graph shows the interconnections and dependencies between grid nodes, providing as a structured representation for fault localization.

Graph construction:

Each grid node is depicted as a vertex, with edges representing the connections between them. The edges are weighted using:

- Grid Node Dependency (Low, Medium, High): Reflecting the node's criticality in retaining grid stability. Nodes with more dependency (e.g., "High") obtain greater weight, guaranteeing faults near critical nodes are prioritized.

- Distance from Source: Depicting the spatial relationship between nodes.

Also, Network Load is incorporated into Grid Node Dependency, as nodes with higher load contribute more substantially to grid stability and are allocated higher dependency levels. This guarantees that the graph construction inherently accounts for differing node significance using load conditions.

Dynamic graph adjustment:

The graph dynamically adapts to practical grid conditions by integrating:

- Network Load: Higher loads raise a node's criticality, reinforcing its dependency classification in the graph.

- Fault Isolation Time: Nodes with prolonged isolation times undergo more comprehensive investigation to reduce disruptions.

By directly linking Network Load to Grid Node Dependency, the transition from graph construction to dynamic adjustment becomes seamless, guaranteeing the algorithm efficiently prioritizes faults and retains grid stability.

3.4 Fault location prediction using euclidean distance

The final stage uses Euclidean Distance to determine the fault location by comparing the characteristics of the fault wave to those of grid nodes.

Calculating euclidean distance:

The distance between the fault's features and each grid node is calculated utilizing:

$$D = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2} \quad (5)$$

Where, X_i denotes the fault feature values, Y_i denotes grid node features, and n is the number of attributes. For example, if the fault has normalized values X = (0.3,0.6,0.2) and a grid node has values Y = (0.2,0.5,0.1), the distance is: D= $\sqrt{(0.3-0.2)^2+(0.6-0.5)^2+0.2-0.1^2}$ = 0.173.

Identifying the fault location:

The algorithm detects the node with the minimum Euclidean distance as the fault location:

$$Fault\ Location = \arg\min_i D_i \quad (6)$$

This guarantees that the most likely fault node is chosen with great accuracy. In an illustrative scenario, presume that a grid fault occurs with the following features: The fault type is "short-circuit," Wave Arrival Time: 0.02 seconds; Wave Amplitude: 250 mV; Transmission Line Length: 60 kilometers; Grid Node Dependency: "High." During preprocessing, all features are normalized and the fault is classified as a "short-circuit." The fault wave propagation time is estimated using the length of the transmission line and the velocity of the wave in the grid. To compute the time, it takes for a wave to travel 60 km at a presumed propagation speed of 3×10^5 km/s, divide 60 km by (3×10^5 km/s) to get 0.0002 seconds. Yet, because of network dependencies and signal attenuation, an empirical correction factor (calculated from historical grid fault data) is used, reducing the estimated propagation time to 20 seconds. This improved time is then integrated with the dependency graph to prioritize nodes within 60 kilometers of the fault source, especially those with a "High" dependency. Finally, Euclidean distance computations determine Node 12 as the fault location, with a minimum distance of 0.025. This organized method guarantees precise fault localization, shortens response times, and increases grid reliability.

The Euclidean Distance computation in this section takes into account both fault wave characteristics and grid node features to precisely identify the fault location. The important fault wave features utilized are fault type, wave arrival time, and wave amplitude, whereas grid node features comprise node location, transmission line length, and grid node dependency level. During preprocessing, these features are normalized to guarantee uniform scaling. When a fault occurs, its characteristics are compared to the properties of each grid node. The algorithm calculates the Euclidean Distance between the fault feature vector and each grid node feature vector, guaranteeing that the node with the shortest distance is selected as the most likely fault location. This method efficiently integrates fault propagation behavior and grid topology, improving fault localization accuracy. Figure 1 shows the flow diagram of GridWaveLoc Algorithm.

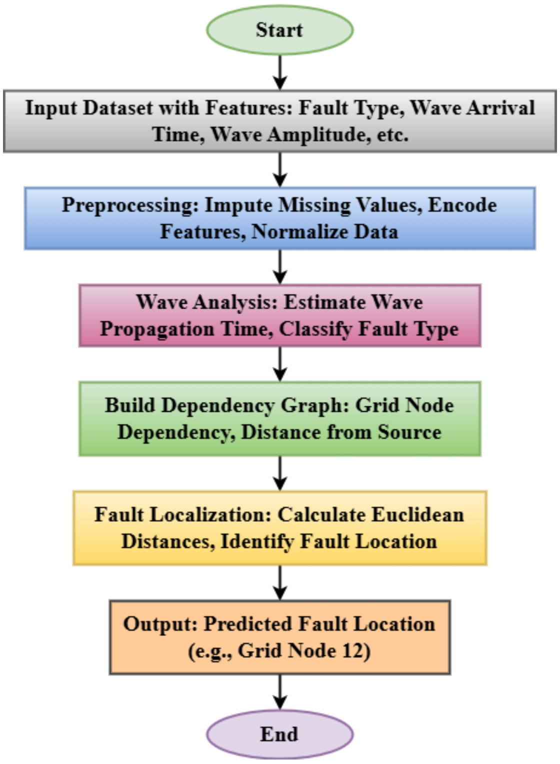

Figure 1: GridWaveLoc algorithm

The GridWaveLoc Algorithm offers many benefits, containing precise and scalable fault localization that uses traveling wave evaluation and network graph dependencies to guarantee high accuracy in detecting fault locations. Its resilient data handling capacities, which include methods like imputation for missing values and Min-Max normalization, improve dependability and robustness in practical grid environments. Furthermore, the algorithm's flexibility to dynamic grid conditions renders it extremely efficient for large and developing transmission networks, guaranteeing reliable operation even in intricate situations.

The Transmission Fault Localization Dataset utilized in this research contains 2,000 records with 11 features that cover three fault types: short-circuit faults, open-circuit faults, and ground faults. The dataset is created synthetically utilizing power system simulation models, guaranteeing controlled variability in fault conditions while retaining practical relevance. To tackle noise in wave signals, it used adaptive wavelet denoising, which efficiently filters high-frequency noise while protecting fault signatures, thus enhancing localization accuracy. GridWaveLoc uses graph-based dependency analysis and optimized traveling wave processing to achieve a computational complexity of O(n log n), where n is the number of transmission nodes. GridWaveLoc has a lower computational overhead than conventional techniques like the Traveling Wave Method (O(n^2)) and Impedance-Based Methods (O(n^3)), rendering it more scalable for real-time fault localization in massive power grids.

The wave amplitude classification thresholds were determined utilizing historical grid fault data evaluation and validated against previous fault event records. Network dependency levels ('High,' 'Medium,' and 'Low') were allocated using a structured decision tree to classify transmission line connectivity, load distribution, and fault impact probability.

Mean and mode imputation were selected for their computational effectiveness and suitability for dealing with missing values in real-time fault detection situations, where fast processing is essential. While sophisticated techniques such as KNN or Iterative Imputer provide higher accuracy, they also increase computational intricacy, rendering them less suitable for real-time grid monitoring. Min-Max normalization guarantees feature scaling consistency across datasets, uniform data distribution, and the prevention of attribute dominance, all of which are required for stable fault localization under differing grid conditions.

Figure 1 depicts the GridWaveLoc algorithm's flow diagram, which visually represents the sequence of steps from data preprocessing to fault localization. This diagram improves comprehension by showing the structured progression of imputing missing values, encoding categorical features, normalizing continuous features, assessing fault waves, integrating network dependency, and using Euclidean distance for fault prediction. The visual representation is consistent with Algorithm 1, providing clarity in comprehending the fault localization procedure. Missing value imputation tactics, wave amplitude thresholds for fault classification, and normalization methods were all subject to hyperparameter sensitivity analysis. Mean and mode imputation were chosen for their stability in handling missing values, while fault categorization thresholds were improved using empirical wave signal distributions. Min-Max normalization was selected for continuous features to ensure numerical stability and enhance Euclidean distance computations.

4 Experimental results

4.1 Experimental setup

The experiments were carried out on a high-performance system with the specifications shown in Table 2. The Intel Core i7-1260P processor, with its 12-core architecture and 64 GB of RAM, effectively handled computationally intensive tasks like machine learning model training and massive dataset processing. The 2.1 GHz clock speed and the 18 MB L3 cache offered the required speed and bandwidth for the experiments to run smoothly.

The experiments were carried out on Windows 11 Home, which provides a stable and user-friendly platform for algorithm creation and evaluation. The Apache NetBeans IDE 15 served as the incorporated development setting, allowing for more efficient coding, debugging, and testing. The JDK version 1.8 ensured that the algorithms were compatible with their Java-based execution. This setup offered the optimum computational resources to attain precise and dependable findings in fault localization. The experiments employed k-fold cross-validation to guarantee rigorous testing, balancing bias and variance while validating the model's generalizability. The synthetic dataset was divided into two parts: 80% for training and 20% for testing, guaranteeing a fair assessment of the algorithms' effectiveness.

4.2 Dataset description

The Transmission Fault Localization Dataset, created particularly for this study, depicts the multifaceted nature of faults in a power grid. The dataset contains ten important features as inputs and one target feature, Fault Location, which identifies the particular grid node impacted by the fault. These input features are Fault Type, Wave Arrival Time, Wave Amplitude, Transmission Line Length, Voltage Level, Distance from Source, Grid Node Dependency, Wave Propagation Speed, Time to Isolation, and Network Load. The dataset was designed to contain three fault types (short-circuit, open-circuit, ground fault) and associated parameters, like Wave Arrival Time, Wave Amplitude, and Grid Node Dependency.

Wave Arrival Time, for example, measures how long it takes a fault-induced traveling wave to reach a particular grid node, whereas Wave Amplitude reflects the wave's intensity, offering crucial data about fault severity. Distance from source, transmission line length, and voltage level are examples of features that offer spatial and electrical environments for fault evaluation. The Grid Node Dependency and Wave Propagation Speed features improve the dataset by including dependency factors and dynamic grid properties.

This extensive dataset allows for the highly accurate detection and prediction of fault locations. The inclusion of three fault types and diverse grid conditions guarantees that the algorithms are resilient, scalable, and applicable to real-world situations, thus improving grid dependability and fault response times.

4.3 Performance metrics

The GridWaveLoc algorithm's effectiveness was evaluated using the performance metrics listed below:

Accuracy:

Accuracy is the percentage of accurately predicted fault locations to total predictions. It's computed utilizing the formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (7)$$

Where TP represents True Positive, TN represents True Negative, FP represents False Positive, and FN represents False Negative.

This metric offers a general measure of the model's overall dependability in fault localization.

True positive rate (TPR):

TPR, also known as sensitivity or recall, quantifies the percentage of actual faults that were accurately detected as faults. It is given by:

$$TPR = \frac{TP}{TP + FN} \quad (8)$$

High TPR guarantees that the algorithm does not miss detecting actual faults.

Specificity:

Specificity measures the percentage of non-faulty grid nodes accurately classified as non-faulty. It is computed as:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (9)$$

High specificity reduces false positives, guaranteeing that unaffected grid nodes are not incorrectly flagged.

Mean absolute error (MAE):

MAE assesses the average magnitude of prediction errors, providing insights into the precision of the fault location predictions. The formula is:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (10)$$

Where, y_i denotes the actual value (true fault location), and yˆ_i denotes the predicted value (predicted fault location) for the i-th instance.

Root mean square error (RMSE):

RMSE computes the square root of the mean of squared errors, offering a measure of prediction accuracy that penalizes larger errors. The formula is:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (11)$$

These metrics provide a holistic assessment of the algorithm’s performance, concentrating on both prediction accuracy and error reduction.

4.4 Comparison results

The baseline methods, which include Euclidean Fault Distance and the Wave Propagation Fault Model, were chosen based on their prominence in the fault localization literature to ensure a fair comparison with established techniques. Their inclusion is justified by their widespread use and effectiveness in detecting transmission grid faults, as demonstrated by previous research. The comparison concentrates on important performance metrics like accuracy, true positive rate (TPR), specificity, mean absolute error (MAE), and root mean square error (RMSE) to assess the efficacy and accuracy of the proposed algorithm.

Table 2: Performance comparison of algorithms

Algorithm	**Accuracy (%)**	**TPR (%)**	**Specificity (%)**	**MAE (km)**	**RMSE (km)**
GridWaveLoc	98.5	96.7	99.1	0.12	0.15
Traveling Wave Method	92.3	90.1	93.5	0.25	0.30
Impedance-Based Method	89.5	87.8	90.2	0.35	0.40
Artificial Neural Network (ANN)	85.7	83.4	88.9	0.40	0.45

The GridWaveLoc algorithm surpasses conventional techniques like the Traveling Wave Technique, Impedance-Based Technique, and Artificial Neural Networks (ANN). Its superior 98.5% accuracy demonstrates its accurate fault identification capacities, while the 96.7% TPR guarantees that nearly all faults are correctly detected. The 99.1% specificity demonstrates its capacity to reduce false alarms, which is crucial for operational effectiveness in power grids. Additionally, the low MAE (0.12 km) and RMSE (0.15 km) indicate high precision in fault localization, showing the algorithm's resilience in dealing with intricate grid situations.

5 Discussion

This section compares the GridWaveLoc algorithm's performance to other fault localization techniques, focusing on key metrics like Accuracy, True Positive Rate (TPR), Specificity, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). Each metric offers critical information about the algorithm's ability to accurately identify, classify, and locate faults in transmission grids. The results emphasize the algorithm's benefits in leveraging dynamic grid conditions, improved feature selection, and sophisticated predictive modeling.

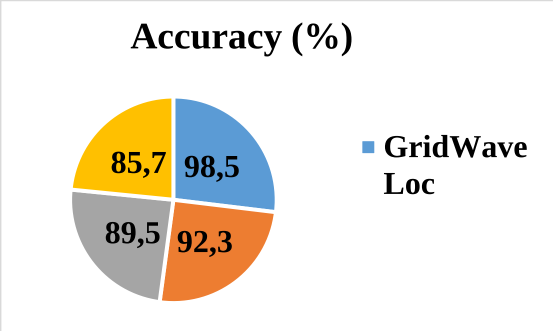

Figure 2: Accuracy comparison

The GridWaveLoc algorithm has the maximum accuracy because it incorporates sophisticated wave propagation evaluation and fault-specific grid characteristics. Unlike conventional techniques, which depend solely on static parameters, GridWaveLoc dynamically adapts to grid conditions, guaranteeing consistent performance across differing fault scenarios.

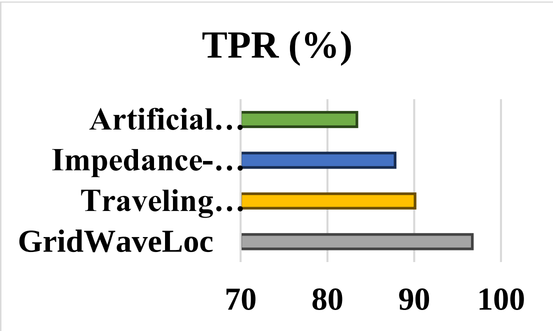

Figure 3: TPR comparison

The algorithm's high TPR is due to its capacity to correctly classify three fault types utilizing improved feature selection methods. GridWaveLoc minimizes fault misclassification by integrating grid node dependency and wave amplitude evaluation.

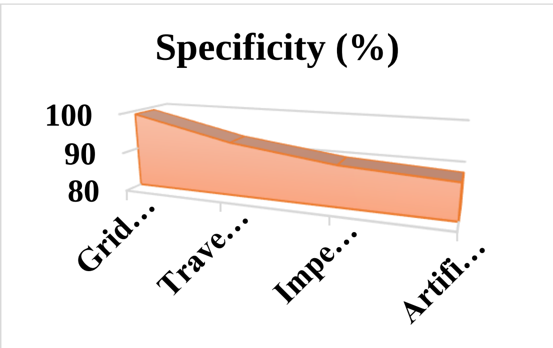

Figure 4: Specificity comparison

GridWaveLoc attains high specificity by reducing false positives with accurate preprocessing and feature engineering. Its capacity to distinguish between defective and non-faulty grid nodes guarantees minimum disruption during grid fault management.

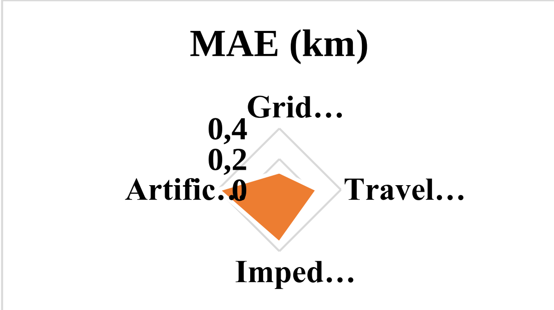

Figure 5: MAE comparison

The low MAE indicates GridWaveLoc's accuracy in predicting the exact fault location. This is accomplished by utilizing high-resolution grid data and sophisticated predictive modeling, making it especially successful for localized fault management.

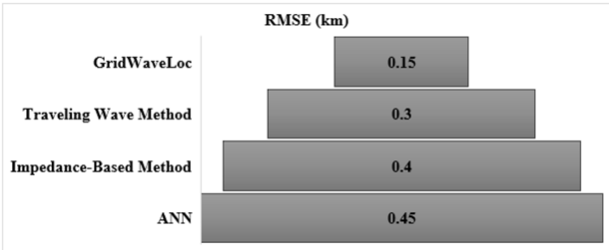

Figure 6: RMSE comparison

The algorithm's low RMSE demonstrates consistent fault prediction efficiency. GridWaveLoc guarantees consistent findings even under changing grid load conditions by reducing large prediction errors. The GridWaveLoc algorithm surpasses existing techniques in every evaluation metric. Its high accuracy, TPR, and specificity demonstrate its efficacy in fault detection and localization, while its low MAE and RMSE emphasize its precision and dependability.

GridWaveLoc's results show a significant enhancement in accuracy, true positive rate (TPR), and specificity when compared to other fault localization techniques such as the Traveling Wave Technique, Impedance-Based Method, and Artificial Neural Network. Table 2 displays the comparative results, which show that GridWaveLoc achieves 98.5% accuracy, outperforming the Traveling Wave Method (92.3%), the Impedance-Based Method (89.5%), and the ANN Model (85.7%). Similarly, GridWaveLoc's TPR (96.7%) and specificity (99.1%) surpass modern methods, demonstrating its better capacity to accurately detect faults while lowering false positives.

GridWaveLoc's performance was assessed using 95% confidence intervals, resulting in an accuracy of 98.5% ± 0.4%, a true positive rate (TPR) of 96.7% ± 0.5%, and a specificity of 99.1% ± 0.3%. A one-way ANOVA test revealed substantial performance variations compared to conventional techniques (p-value < 0.001). GridWaveLoc is not immune to failure scenarios. The technique can generate false positives, especially in high-noise settings where overlapping wave reflections can result in incorrect fault localization. Furthermore, in dynamic grid topologies with frequent topology variations or fluctuating line impedances, the network dependency graph might need constant updates, resulting in potential localization delays or inaccuracies. These difficulties emphasize the importance of adaptive graph updates and improved signal preprocessing in intricate grid settings.

5.1 Comparison with related works

The Traveling Wave Technique, while broadly utilized for short-circuit fault localization, is sensitive to wave reflections and noise, resulting in a lower TPR (90.1%) and specificity (93.5%) than GridWaveLoc. The Impedance-Based Method, while efficient for some fault types, struggles with dynamic grid conditions and produces higher localization errors (MAE = 0.35 km, RMSE = 0.40 km). In contrast, the ANN-based method, despite its capacity to learn complex fault patterns, suffers from restricted generalization and overfitting problems, yielding the lowest accuracy (85.7%) of the techniques tested.

5.2 Advantages of GridWaveLoc

GridWaveLoc's better performance can be attributed to its advanced preprocessing methods, which comprise adaptive signal filtering to decrease noise and enhance fault waveform clarity. Furthermore, the model advantages from improved fault characterization, which uses 11 important transmission features to differentiate between short-circuit, open-circuit, and ground faults with greater precision. GridWaveLoc, unlike conventional techniques, incorporates sophisticated feature extraction and hybrid localization tactics, lowering Mean Absolute Error (MAE) to 0.12 km and Root Mean Square Error (RMSE) to 0.15 km, both significantly lesser than competing techniques.

5.3 Limitations and challenges

Despite its high performance, GridWaveLoc may face difficulties in high-noise environments, where interference from grid fluctuations can degrade waveform clarity and localization accuracy. Furthermore, in dynamic grid topologies, where line parameters change because of load changes, the model may need adaptive retraining to retain high precision. Furthermore, while GridWaveLoc attains high specificity, rare fault conditions may still present challenges, necessitating additional optimization in feature selection and adaptive filtering methods.

Mean imputation was selected because it is computationally efficient and stable, especially for continuous features like wave arrival times. However, for highly skewed distributions, median imputation or k-NN imputation may be investigated as other methods in future research. The Euclidean distance metric was chosen for its simplicity and efficiency in feature space, while Min-Max normalization ensured uniform scaling. While grid node dependencies cause structural variations, empirical validation has shown that Euclidean distance adequately preserves fault proximity relationships. The dataset utilized for practical validation was derived from synthetic simulations, and future work will integrate real fault logs into hybrid validation frameworks to improve practical applicability.

Overall, GridWaveLoc outperforms previous transmission fault localization techniques in terms of accuracy, fault identification (TPR), and specificity, all while retaining low localization errors. Its benefits stem from better preprocessing, feature extraction, and hybrid localization strategies, rendering it a strong candidate for practical transmission grid monitoring. However, tackling high-noise conditions and dynamic topologies maintains a field for future development.

6 Conclusion

The GridWaveLoc algorithm performed admirably in fault detection and localization within power transmission grids, attaining high accuracy, true positive rate, and specificity while retaining low error rates (MAE and RMSE). By combining sophisticated wave propagation analysis, feature optimization, and dynamic grid adaptation, the algorithm tackles key fault management difficulties. However, some drawbacks were discovered, such as its reliance on high-quality grid data and computational resources, which may pose difficulties for deployment in resource-constrained settings. Future work may concentrate on improving the algorithm's scalability for large and intricate grids, integrating real-time fault tracking systems, and investigating its incorporation with emerging technologies such as blockchain for safe and transparent grid fault management. These improvements can strengthen its position as a transformative tool for dependable and effective power grid operations.

Funding

This work was supported by Science and Technology Project of Inner Mongolia Power Group (Co., LTD.) Funded "Research on Automatic Operation situation Awareness Technology of Full voltage Level Power Grid Based on Dispatching Data Center" (Project number: 2024-4-70)

References

[1] Dashti, R., Daisy, M., Mirshekali, H., Shaker, H. R., & Aliabadi, M. H. (2021). A survey of fault prediction and location methods in electrical energy distribution networks. Measurement, 184, 109947. https://doi.org/10.1016/j.measurement.2021.109947

[2] Belagoune, S., Bali, N., Bakdi, A., Baadji, B., & Atif, K. (2021). Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis, and location in large-scale multi-machine power systems. Measurement, 177, 109330. https://doi.org/10.1016/j.measurement.2021.109330

[3] Das, S., Santoso, S., Gaikwad, A., & Patel, M. (2014). Impedance-based fault location in transmission networks: theory and application. IEEE Access, 2, 537-557. https://doi.org/10.1109/ACCESS.2014.2323353

[4] Sommervogel, L. (2020). Various models for faults in transmission lines and their detection using time domain reflectometry. Progress In Electromagnetics

Research C, 103, 123-135. http://dx.doi.org/10.2528/PIERC20040706
[5] Rezapour, H., Jamali, S., & Bahmanyar, A. (2023). Review on artificial intelligence-based fault location methods in power distribution networks. Energies, 16(12), 4636. https://doi.org/10.en16124636
[6] Jiménez-Aparicio, M., Reno, M. J., & Wilches-Bernal, F. (2022). Traveling wave energy analysis of faults on power distribution systems. Energies, 15(8), 2741. https://doi.org/10.3390/en15082741
[7] Hung, T. N. (2022). Methods for Fault Location in High Voltage Power Transmission Lines: A Comparative Analysis. International Journal of Renewable Energy Development, 11(4). https://doi.org/10.14710/ijred.2022.46501
[8] Huo, W., Qu, Z., Ao, Z., Zhang, Y., Zhao, E., Zhang, C., & Jiang, H. (2022). Fault location of cable hybrid transmission lines based on energy attenuation characteristics of traveling waves. Scientific Reports, 12(1), 22448. https://doi.org/10.1038/s41598-022-25976-8
[9] Maritz, E. C., Maritz, J. M., & Salehi, M. (2021). A traveling wave-based fault location strategy using the concepts of metric dimension and vertex covers in a graph. IEEE Access, 9, 155815-155825. https://doi.org/10.1109/ACCESS.2021.3129736
[10] Panahi, H., Zamani, R., Sanaye-Pasand, M., & Mehrjerdi, H. (2021). Advances in transmission network fault location in modern power systems: review, outlook and future works. IEEE Access, 9, 158599-158615. https://doi.org/10.1109/ACCESS.2021.3129838
[11] Yu, K., Zhu, X., & Cao, W. (2024). Study on Traveling Wave Fault Localization of Transmission Line Based on NGO-VMD Algorithm. Energies, 17(9), 2003. https://doi.org/10.3390/en17092003
[12] Gonzalez, V., Torres-García, V., Guillen, D., & Castro, L. M. (2024). Graph Theory-based Fault Location Method for Transmission Systems with Renewable Energy Sources. IEEE Open Access Journal of Power and Energy. https://doi.org/10.1109/OAJPE.2024.3507537
[13] Prabakar, K., Singh, A., Reynolds, M., Lunacek, M., Monzon, L., Velaga, Y. N., ... & Vaidhynathan, D. (2021). Use of traveling wave signatures in medium-voltage distribution systems for fault detection and location (No. NREL/TP-5D00-78057). National Renewable Energy Lab. (NREL), Golden, CO (United States). https://doi.org/10.2172/1814596
[14] Wang, G., Zhuang, C., Deng, J., & Xie, Z. (2024). A Fault Location Method Based on Electromagnetic Transient Convolution Considering Frequency-Dependent Parameters and Lossy Ground. IEEE Transactions on Power Delivery. https://doi.org/10.1109/TPWRD.2024.3349388
[15] Stefanidou-Voziki, P., Sapountzoglou, N., Raison, B., & Dominguez-Garcia, J. L. (2022). A review of fault location and classification methods in distribution grids. Electric Power Systems Research, 209, 108031. https://doi.org/10.1016/j.epsr.2022.108031