Int. J. Simul. Multidisci. Des. Optim. **15**, 15 (2024)

https://doi.org/10.1051/smdo/2024010

IJSMDO
Available online at:
https://www.ijsmdo.org

RESEARCH ARTICLE OPEN ACCESS

Enhanced fault localization in multi-terminal transmission lines using novel machine learning

Yingyun Wang, Xiaoxia Qi and Yang Chen*

School of Big Data and Artificial Intelligence, Anhui Xinhua University, Anhui, 230088, China

Received: 6 March 2024 / Accepted: 14 May 2024

Abstract. Accurate fault location on transmission lines is paramount for ensuring the reliable and efficient operation of the electricity grid, which underpins every aspect of modern society. Existing fault localization methods for transmission lines often face shortcomings, particularly in scenarios involving multi-terminal transmission lines, where complexities arise due to dispersed generations and intricate network configurations. Traditional approaches may struggle to provide accurate fault localization, impacting the reliability and efficiency of the electricity grid. Research provide a unique fault localization technique in this article depends on Phasor Measurement Units (PMUs) and Bidirectional Gradient Boost Random Forest (BDGB-RF) machine learning technique to address these challenges. The proposed method offers several advantages over traditional methods, including enhanced accuracy and efficiency. By leveraging PMU data and BDGB-RF, the method provides a two-phase fault localization strategy, incorporating fault line selection based on nodal current imbalance and subsequent fault distance determination. Simulation results demonstrate the effectiveness of the proposed approach, achieving up to 97% accuracy in the majority of studied scenarios, even under different tapping configurations. The adoption of this approach could significantly impact the practices of experts in the field, facilitating more reliable fault detection and localization in complex transmission line networks. This, in turn, can contribute to the resilience and stability of power systems, ultimately improving grid reliability and minimizing downtime.

Keywords: Fault location / transmission line / tapped lines / detection system / phasor measurement unit

1 Introduction

One-terminal and multi-terminal-based fault location methods are the most common types. One-terminal methods are often straightforward to implement. These calculations always assume the source impedance, fault resistance, loading, and other parameters, as well as the network's topology at the time of the failure [1]. Furthermore, considering multi-terminal network topologies presents a challenge for one-terminal-based methodologies to produce reliable findings. In multi-terminal-based techniques, phasors are measured at both ends of the network to cut down on mistakes made when finding faults. This is in contrast to one-terminal-based approaches, which make assumptions for the sake of convenience. This enables the attainment of more accurate results. Furthermore, by considering multi-terminal network topologies, we can accurately determine fault sites using data from all nodes [2]. In the past few years, the Taipower system has included transmission lines tapped by an individual producing a plant or load with relatively short transmission lines due to the difficulty of obtaining additional rights of way for transmission lines in Taiwan. Once these connections have been completed, the system will expand to include three more terminals. We need to set up additional infrastructure to build fault locators using these methods for the tapped lines [3]. New fault locator installations, on the other hand, are unnecessary and expensive because the tapped leg is always short and intermediate. To estimate the impact of the tapped leg while utilizing the two-terminal-based method to determine the fault site, a model of the tapped leg must be built. However, the tapped leg model is not always straightforward to create, especially in the case of nonlinear load models. As automation becomes more commonplace in manufacturing, so does the need for reliable electricity [4].

These operations need a certain amount of voltage and power frequency. Considering the repair time in the event of permanent faults, which significantly impacts the system's dependability and causes economic losses by halting industrial activities, pinpoint fault detection in transmission lines is a pressing need. Quick identification and repair of a transmission line problem can expedite the

* e-mail: chenyang@axhu.edu.cn

restoration of electricity, minimize outages, and enhance the overall dependability of the power system [5]. So far, one- and two-terminal approaches have been the most popular. One-terminal methods, which only use one-terminal voltage and current phasors, are often less accurate at finding faults because of the remote-terminal network resistance and fault resistance. The goal of developing two-terminal algorithms is to improve the reliability of fault localization [6]. To transmit electricity, a tapped four-route transmission line uses four parallel wires laid out in a square pattern. Two sets of the four wires each link to a separate voltage source. Both sets of conductors provide taps to generate step-up and step-down voltage levels. When compared to a standard two-route transmission line, the increased power transfer capacity and improved voltage control provided by the four-route architecture are clear advantages. The taps also allow for localized voltage regulation at various places along the transmission line [7].

High-voltage AC and DC transmission systems often use taps in four-route transmission lines to facilitate long-distance power delivery. Distribution networks and substations also use them to convert and distribute power. However, faults can occur on any of the four conductors or taps, the tapped four-route layout can make fault detection and protection more difficult [8]. PMU fault location is one example of a specialized protection strategy and fault localization approach utilized to overcome these obstacles. Scientists estimate the fault location based on differences in the arrival timing of fault-generated waves at various PMU sites [9]. To calculate how far away a fault is, scientists measure how long it takes for waves created by the fault to reach various PMUs. These waves move at almost the speed of light. Once an estimate of the distance has been made, the line specifications and tap settings may be used to locate the transmission line on a map. To swiftly discover and isolate problems and restore power to consumers, power system operators need a fast and precise method, and PMU fault localization is a perfect fit [10].

This study introduces a novel fault location algorithm that can be retrofitted into current fault locators, eliminating the need for either new installations of fault locators or the tapped leg model. For multi-terminal transmission lines, we suggest using Bidirectional Gradient Boost Random Forest (BDGBRF) to identify the location of faults. The proposed approach relies only on phasor measurements taken from two ends of the original transmission lines that have been synced using PMUS. Since the suggested novel method is suitable for the temporarily tapped transmission lines, there is no requirement for a new installation to implement it. Furthermore, the fault localization technique may be readily expanded to various sorts of the tapped leg, including a generator or a load, since the model of the tapped leg is not employed in the proposed algorithm.

Parallel transmission lines are used using the fault localization strategy outlined in [11]. Measurements of voltage and current at only one terminal are all that are required since the method depends on the distributed parameter model of a transmission line. The approach uses sequence networks, where the sequence voltages at the fault spot are computed based on voltage and current measurements and the unknown resistance of the fault. It is possible that this method's fault-finding mistakes may grow in the presence of untransposed transmission lines and unknown tapping loads. The approach presented in [12] used propagating waves as its foundation. A matrix is constructed using the ratio of the fault distance to the length of the branch produced by the pair of terminals studied as its coefficients. The fault's local and post-fault distances are calculated by the author using this matrix. The fact that it can only be used with single-circuit transmission lines and requires measurements at each terminal is a major drawback. An adaptive technique for detecting faults based on coordinated readings from PMUs was presented in [13], which does not need measurements supplied by power companies. Only untapped loads on single-circuit transmission lines may use this technique. The technique for solving the identification of power lines using double and multiple circuits with tap-connected loads is described in [14]. The technique uses voltage and current phasors to pinpoint the malfunction. The approach relies on two distinct algorithms, one for data accessible on two terminals and another for data available from a single terminal, to pinpoint the source of the malfunction. The description of a matching degree index includes a two-phase fault identification optimization method [15]. Large transmission networks may benefit from the technique. The related approach for PMU installation is also provided. High-impedance faults may throw off the findings of a technique that relies only on voltages for computations.

Single-circuit transmission lines may be analyzed using the approach given in [16]. These phasors are employed by some terminal fault location techniques to compute the current and voltage phasors at the tap sites. Since the approach cannot predict the impedance of the loads, it can only be used with single-circuit transmission lines. A two-terminal fault-localization method was expanded to incorporate the N-terminal transfer lines [17]. The design functions across any multi-terminal connection setup. However, the computational overhead is increased since (N-1) two-terminal indices must be determined. In [18], fault diagnosis based on the fluctuating zero-sequence current's power across the specified range of frequencies (SFB) is described, and this requires synchronized observations in distribution networks. The resilience and exceptional precision of the suggested approach remain unaffected by the presence of strong background noise, different fault locations, different types of loads, different grounding resistances, different initial angles, different locations, and different power supplies for DGs. However, the proposed method can only account for faults that originate and are limited by the D-PMU units' configuration towards a neutral region. In addition to information from security systems and financial institutions, the proposed method for fault detection also makes use of data from smart meters, current and voltage meter measurements, and computer-based models [19]. In [20], the author described a method for locating faults that makes use of network topology data and travelling waves that generate reclosures. The amount of time that has

passed between the points at which the fault reclosed and the point at which the reflected travelling wave arrived at the fault site is used in the process of fault localization. The study examined object-based defect localization techniques, enhancing the Swin Transformer algorithm from the viewpoint of detection and feature extraction networks [21]. The best object detection algorithm for finding subsurface road faults in GPR pictures will be determined by evaluating the model with indicators like recall, accuracy, and average accuracy [22].

2 Materials and methods

In nature, the challenge of identifying faults could be considered a classification of multiple classes problems. The multi-category classification in this work is accomplished via the use of a Bi-Directional gradient boost Random Forest. Following the extraction procedure based on feature significance using BDGBRF, the model is generated in the RF algorithm.

2.1 Data collection

The model predicts where four-phase-to-ground problems will happen in the distribution network by measuring the voltage level (voltage sag) at the bus locations. The data was made with Open DSS software by simulating how power flows during a fault. There are 500 specimens for each fault area in the Chinese location. There are two parts to the data: feature information and label information. The voltage readings of all of the system's bus stages are included in the feature data.

2.2 Data pre-processing

Data cleansing, transformation, and analytical priming are all aspects of pre-processing. The goal of this process is to enhance the level of the data so that it can be better understood and analyzed. Normalization is a pre-processing step that involves scaling, mapping, or other similar procedures. Where we may expand upon already established ranges. It has great potential for use in forecasting and predicting the future. As is well-known, there is a wide range of methods that may be used for making predictions and forecasts. Therefore, the Normalization method is essential to bring them closer together to keep the big variety of prediction and forecasting.

Normalization of Min-Max is a method that linearly transforms the initial spectrum of data. Min-Mix is a method that preserves the connections between the original data. Min-Max normalization is a straightforward method that precisely fits the data inside a specified limit. The values for each characteristic may be anywhere from 0 to 1, where 0 is the minimum and 1 is the maximum. The process of normalizing may be expressed as an equation.

$$M_{norm} = \frac{M_r - M_{min}}{M_{min} - M_{\max}}. \tag{1}$$

When dealing with data in batches, the lowest and maximum values are denoted by M_{min} and M_{max} respectively, where M_r is a data point. Using a weighted ranking, Recursive Feature Elimination (RFE) determines the importance of each feature and delivers a number based on that. The size of feature subsets might be reduced if unnecessary information was eliminated. To determine a feature's rating, it may be sorted from most essential to least important.

3 Bidirectional gradient boosts random forest

Bidirectional gradient boosts random forest is a widely used algorithm in the field of boosting learning since it not only satisfies the needs of the simulation but also shortens the amount of time needed to train it. To be useful for feature extraction, BDGBRF may learn the significance of each feature throughout the modeling process. Since the random forest is often used as the base classifier in both XGBoost and, it's not surprising that they have certain similarities. XGBoost is an upgraded technique of Bidirectional Gradient Boosting random forest (BDGBDT). In this research, BDGBDT is first introduced, and then the fundamentals of the random forest-based XGBoost algorithm are outlined.

The BDGBRF is an efficient technique that finds a happy medium between how well a model performs and how long it takes to run. It is an enhanced version of BDGBRF, a boosting method that uses decision trees and incorporates regularization into its optimized objective function. To help the increase in trees attain the rapid computation and great performance that are the objectives of the engineering, the BDGBRF algorithm is developed.

CART is also used as the foundational classifier in the XGBoost method. Let's assume that we have a multi-classification issue with n samples $(W_j, z_j)_{j=1}^{m}$, m features $W_j = (y_j1,\ y_j2,...,\ y_j,\ n)$, and C classes $z_j \in \{0,\ 1,\ 2,\ 3,\ ...,\ D-1\}$. As demonstrated in equation (2), the task aims to minimize the value of the function with objectives by reducing both the training loss and the tree complexity. To accomplish multi-classification, we choose the Cross-Entropy Loss as the loss function and combine it with softmax.

$$obj = \sum_{j=1}^{m} Loss(zj, \widehat{z}j) + \sum_{l=1}^{L} \Omega(hl), \tag{2}$$

$$Loss(zj, \widehat{z}j) = -\sum_{j=1}^{m} z_j log(\widehat{z}_j). \tag{3}$$

$$Soft\ max\colon Z_i = \frac{f^{yi}}{\sum_i f^{yi}}. \tag{4}$$

The random forest $F = \{s_1,\ s_2,\ ...,\ s_s\}$ is an ensemble method consisting of multiple unrelated decision trees. These independent trees provide a little chance into model F's decision trees, helping it make a more precise generalization. The bagging method, which incorporates both scaling and aggregation, is used to get these broad

conclusions. Assume that the input feature space XRD includes parameters like voltage (v), phase angle (Ø), current (i), and frequency (f), and that the training set $T = \{W^n, Z^n\}^n (N = 1)$ where $W \subset Q^c$ has these values. The position of the fault and its associated time, both denoted by the symbol Z, are part of a continuous space with the dimensions $Z \subset Q^{c'}$ Bootstrap is a subset St of the whole training set S, where each instance is randomly picked from the distribution with or without replacement, and M is the number of specimens. The number of specimens in the bootstrap sample is the same as the number of samples in the original data set S, but about one-third of the samples are duplicates, and approximately one-third of the occurrences are eliminated. To generate bootstraps for each tree, the input data is processed via several iterations. When using gain data for training and evaluation, the predictions from many trees are combined into a single score.

The regression outcome may be represented as long as output variables have a multivariate Gaussian distribution with mean and covariance.

$$O(z|y, O_s) = M_s(z|\mu_s, \Sigma_s). \tag{5}$$

O_s is a multimodal Gaussian with mean μ_s and covariance Σ_s, predicted in what comes out of space Y from the selections of the training dataset, and O_s is is the partition constructed by the random tree *tt*. The goal of training the trees is to lessen the guesswork involved in using the multivariate Gaussian model, and this is particularly true when choosing a suitable splitting function *f* to divide the training set into subsets T_k. Each incoming node *Nl* in the tree *tt* performs these computations to mitigate "splitting"-related prediction inaccuracy.

The Gini index and knowledge gain are two examples of function *f*. An ideal function for calculating information gain in a regression job is the free-of-weight differentially entropy function, which is a continuous variant of Shannon's entropy (SE). The SE function was chosen because it provided results that were good in terms of prediction error, which is,

$$e(T_1) = f(z \in Z) \sum_{j=1}^{m} O(z|T_k)/log(O(z|T_k))cZ, \tag{6}$$

Where *j* is the instance being used as input and *z* is the final result that takes into account both the time and location of the issue. Since the posterior is modeled using a multivariate Gaussian, we may reformulate *f* to read:

$$e(T_1) = \frac{1}{2} log((\pi \exp)^{c'} |\Sigma^{(T_s)}|). \tag{7}$$

The calculated covariance matrix $|\Sigma^{(T_s)}|)$ for subset T_k is denoted by (T_k). The informational gain is determined by first dividing the subset T_k at node *Nl* into separate subsets, T_k^{right} and T_k^{left}, using the function *f*.

$$\Delta = e(T_k) - x_k e(T_k^{left}) - x_q e(T_k^{right}). \tag{8}$$

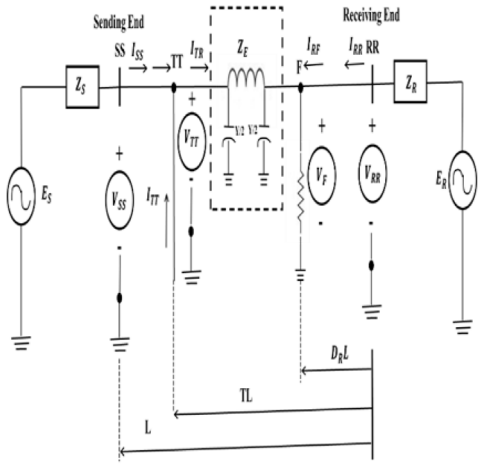

Fig. 1. Transmission line for ground faulted.

After the instruction phase is over, the new instances are sent through the trees of the forest, and the accommodate of all the trees is calculated with the help of the following equation:

$$O(z|w) = \frac{1}{s} \sum_{x=1}^{s} O(z|w, O_s), \tag{9}$$

where O_s is the partitioning brought about by tt and S is the total amount of trees in the forest. The model's ability to forecast fault duration and location for a given new occurrence is based on the maximization of a posterior:

$$\widehat{z} = arg\max_{z \in Z} O(Z|w). \tag{10}$$

Thus, the bidirectional gradient boosts random forest is determined.

The following premises are used to develop the fault location technique:

A resistive fault impedance.

Initially, one must specify the fault type,

Literature addressing the fault localization problem often makes the preceding assumptions.

4 Fault location algorithm

This study uses the three-phase transposition transmission line with tapped legs illustrated in Figure 2 to demonstrate the concepts of the subsystem behind the tapped leg is a generator, a load, or a hybrid of the two. Tapped leg connections are made at T [p.u.] from the far end of the transmission lines and are denoted by the symbol TT. The transfer line's two ends are designated by the symbols SS and RR. Figure 1 shows that the values at both ends are phase voltage and current vectors. Calculations of the EMTP and the creation of fault-location techniques rely on the long-distance decentralized transmission line model.

Partial differential calculations determine the voltages and currents x kilometers distant from the receiver.

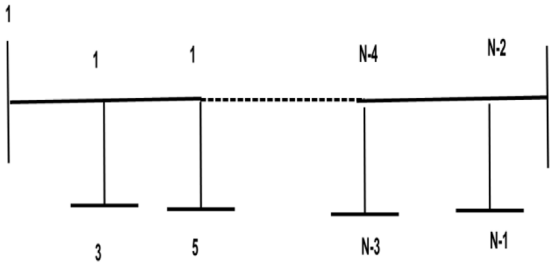

Fig. 2. Transmission line system of n-terminal.

Algorithm 1: process of BDGB-RF

PartialDifferentialCalculations(distance_x):
Initialize Variables:
receiver_voltage = Vector of phase voltages at the receiver end
receiver_current = Vector of phase currents at the receiver end
transmission_line_model = Decentralized transmission line model
voltage_x = Vector of phase voltages at distance x
current_x = Vector of phase currents at distance x
For each Phase k:
For each Element in the Transmission Line Model:
Calculate Voltage Drop due to Element:
voltage_drop = CalculateVoltageDrop(element_impedance, element_current)
Update Voltage and Current at Distance x:
voltage_x[k] += voltage_drop
current_x[k] += CalculateCurrent(element_impedance, element_current)
Return voltage_x, current_x

The vulnerabilities in BDGB-RF protection systems need to be located using a precise and rapid method. Machine learning (ML) is utilized to obtain a rapid response for fault detection in this work. The trained network is prepared for error detection after the learning phase of the ML approach, which is a time-consuming process. When it comes to finding bugs in tests, ML is light years ahead of logic techniques.

5 Fault location index

Consider the n-terminal line of the transmission shown in Figure 2. Every node was separated into two groups: such as terminating network p and tapped network q. We consider the generally used presumption that a PMU is installed on each terminal node. Thus, all terminals have access to the synchronized voltage and current accommodate.

The positive sequencing evaluations are used in this article because they are the only network sequence that can account for all forms of defects. If not otherwise noted, all of the quantities pertain to positive sequence quantities. The lines for transmission under consideration are thought to be in the wrong order.

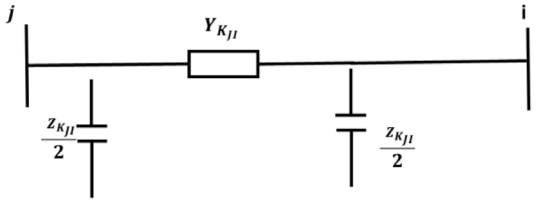

Fig. 3. An equivalent model of i–j.

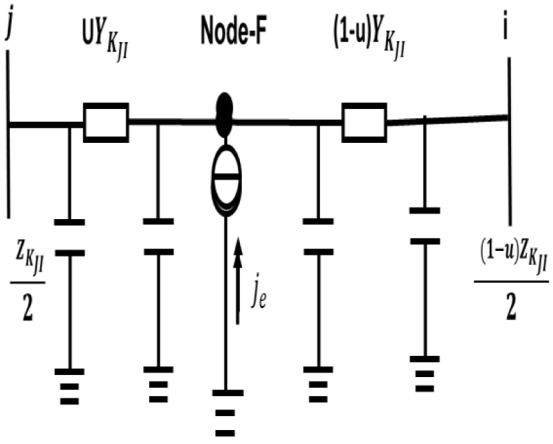

Fig. 4. Fault occurs on line segment i–j.

The node susceptibility equation depicts the fault-free condition of the n-terminal structure in Figure 3.

$$X_{m\times m}W^0_{m\times 1} = J^0_{m\times 1}, \tag{11}$$

Where $W^0_{m\times 1}$ represents the pre-fault network voltage format $J^0_{m\times 1}$ represents the pre-fault network current injection format and $X_{m\times m}$ represents the pre-fault bus admittance matrix. Figure 4 depicts the analogous model of line segment i–j utilized in this work.

The corresponding impedance of line i–j in this model is denoted by Y_{kJI}. Admittance Z_{kJI} accounts for the influence of shunt capacitances. In Figure 4, we see a defect on the line segment i–j, which serves as an example to implement the fault location index.

Because the fault network- F may be viewed as a fictional node, the line can be represented by Pair of models. The proportion of the line length represents a description of the unknown component x that fault distance from network i to fault node F represents. Clearly, x ranges from 0 to 1, which will be solved in the article.

$$X^{'}_{jj} = X_{jj} - \frac{x_{kji}}{2} - \frac{1}{D_{kji}} + \frac{yx_{kji}}{2} + \frac{1}{yD_{kji}}, \tag{12}$$

$$X^{'}_{j(m+1)} = X^{'}_{(m+1)j} = -\frac{1}{yD_{kji}}, \tag{13}$$

$$X^{'}_{j(m+1)(m+1)} = X^{'}_{(m+1)j} = -\frac{1}{(1-y)D_{kji}}, \tag{14}$$

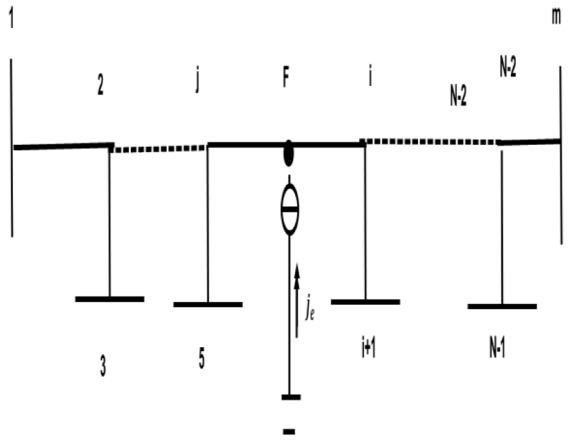

Fig. 5. Faults occur on line segment *i–j*.

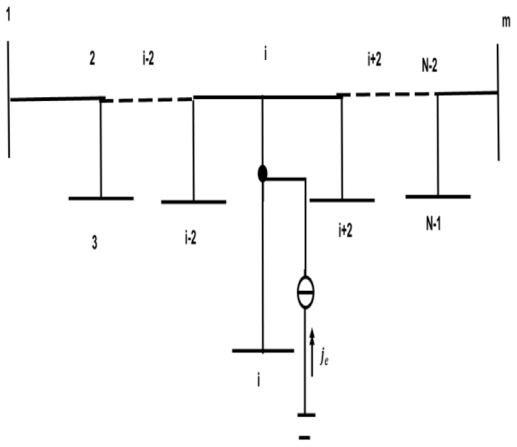

Fig. 6. Faults occur on tapped lines *i–j*.

$$X'_{(m+1)} = \frac{x_{kji}}{2} + \frac{1}{yD_{kji}} + \frac{1}{(1-y)D_{kji}}. \quad (15)$$

As a result, if a defect develops on line *i–J*, the equation that follows may be formed:

$$X_{new}\left[W_{m\times 1_{W_c}}\right] = \left[J_{m\times 1_{J_c}}\right], \quad (16)$$

Where W_c,J_c represent the wireless current and voltage at the issue's location F, Correspondingly; $W_{m\times 1}$ represents post-fault power field in the system; Where $J_{m\times 1}$ represents the post-fault network current field.

6 The defect on the primary line segment

Lines 1–2, 2–4, 4–6, etc., on the main line segment, are where the fault is located, as indicated in Figure 5. In this instance, the presumption is correct. That leaves the tap node voltages in the dark.

There will only be two nonzero items in the calculated fault location index n1 I if a fault arises on the main line segment *i–j*.

Specifically, in Figure 3, the principal line segments consist of lines$12-and\ k-(k+2)(k-\in-2,\ 4...m-2)$, so $\Delta J_{m\times 1}$ can be described as follows:

$$\Delta J_{m\times 1} = \begin{cases} [\Delta J_1, \Delta J_2, 0, \ldots 0]^s \, fault\ on\ line\ k-2 \\ [0,\ldots,0\Delta J_1, 0, \Delta J_{k+2}, 0, \ldots, 0]^s \\ \qquad fault\ on\ line\ k-(l+2)\,. \end{cases} \quad (17)$$

7 The broken part of the tapping line

This assumption breaks down if a fault develops along the tapped line segment (lines 2–3, lines 4–5, lines 6–7, etc.) in Figure 6. Both the fault current $f\ l$ and the voltage 'Ui at the tap node i are ignored. This means that the organize' Ui does not correspond to the actual voltage Ui at tap node i ('U U i i). Here, n1 I exhibits a unique property.

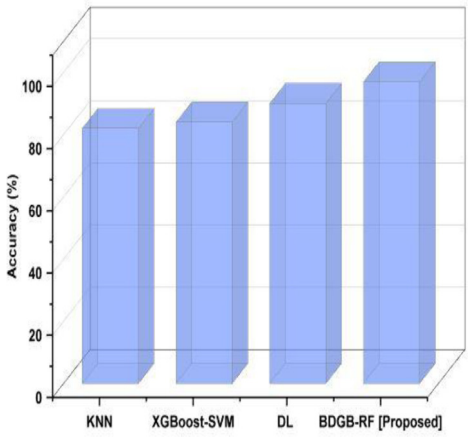

Fig. 7. Accuracy.

In Figure 7, the calculated voltage X'_j at tap node j is calculated,

$$X'_j = (J_i - X_{ji}X_i)/Z_{ij}. \quad (18)$$

Voltage 'X'_j was not identical with the current that exists X'_j at access network J_e because of the influence of fault current j. The following applies here:

$$X_{m+1} = [X_1 \ldots X'_j .. X_1 .. X_1]^s, \quad (19)$$

$\Delta J_{m\times 1}$ can also be obtained as follows:

$$\Delta J_{m\times 1} = Z_{m\times m}X_{m\times 1} - J_{m\times 1}, \quad (20)$$

we will evaluate all non-zero measures of $\Delta J_{m\times 1}$ in this case

For node j we have

$$\Delta J_j = Z_{j1}X_1 + \ldots + Z_{jj}X'_j + \ldots + Z_{ji}X_i + .. + Z_{in}X_m - J_j, \quad (21)$$

$$\Delta J_j = (Z_{jj}X'_j - Z_{jj}X_j) + (Z_{jj} - Z'_j)X_j - Z'_{j(m+1)}X_j. \quad (22)$$

For node i we have:

$$\Delta J_j = Z_{jj}X_1 + \ldots + Z_{ij}X_j^{'}\ X_j' + \ldots + \ldots + Z_{ij}X_j + Z_{ij}X_j - J_i. \tag{23}$$

Because the i^{th} row in $Z_{m\times m}$ has only two non-zero elements: Z_{ij} and Z_{ij} so (30) can be simplified as below:

$$\Delta J_j = Z_{jj}X_j^{'} + Z_{jj}X_j - J_i, \tag{24}$$

$$\Delta J_j = Z_{jj}(J_i - Z_{ij}X_j)/Z_{ij} + Z_{ij}X_j - J_i = 0. \tag{25}$$

For any other node $l(lj \neq and\ l \neq i)$, we have

$$\Delta J_j = Z_{jj}X_1 + \ldots + Z_{lj}X_j^{'} + \ldots + Z_{ij}X_j + \ldots + Z_{lm}X_m + J_k, \tag{26}$$

$$\Delta J_k = Z_{ij}(X_j^{'} - X_j).$$

In Figure 7, only node (j–2) and node (j–2) are directly connected with node, so $Z_{(j+2)j} \neq 0$ for any Z_{lj} other node l ($l \neq j-2\ and\ l \neq j+2$). Therefore, we can conclude based on

$$\{\Delta J_k = 0(l \neq j-2, j, j+2), \tag{27}$$

when the plugged connection possesses an issue j–i shown in Figure 5, $\Delta J_{\mathrm{m}\times 1}$ I has three non-zero elements $\Delta J_{\mathrm{m}\times 1}$has three non-zero elements $\Delta J_{j-2}\Delta J_j$ and ΔJ_{j-2}

$$\Delta J_{m\times 1} = [0, \ldots, \Delta J_{j-2}, 0, \Delta J_j, 0, \Delta J_{j+2}, \ldots 0]^s. \tag{28}$$

The two linked nodes in the event of a failure on the plugged connection 2–3 would become 1 and 4. Hence, $\Delta J_{m\times 1}$I require to be expressed in the following manner:

$$\Delta J_{m\times 1} = [\Delta J_1, \Delta J_2, \Delta J_4, 0, \ldots, 0]^s. \tag{29}$$

In conclusion,

$$\Delta J_{m\times 1} = \begin{cases} [\Delta J_1, \Delta J_2, 0, \ldots 0]^s\ fault\ on\ line\ 2-3 \\ [0, \ldots, 0\Delta J_{j-2}, 0, \Delta J_{j+2}, 0, \ldots 0]^s\ fault\ on\ line\ j-i \end{cases}. \tag{30}$$

8 Result and discussion

We recommend employing Bidirectional Gradient Boost Random Forest (BDGB-RF) to identify defects along multi-terminal transmission lines. The effectiveness of a suggested technique is measured against that of established methods like Kernel closest neighbor (KNN) [23], Extreme Gradient Boost – Support Vector Machines (XGBoost-SVM) [24], and Deep Learning (DL) [25]. Multiple metrics, including accuracy, precision, recall, f1 score, and computation time, are used to evaluate these methods in comparison to their predecessors.

A variety of parameters, including the number and placement of PMUs, the precision of PMU measurements, the complexity of the power system, and the fault type and location, contribute to the reliability of PMU fault detection. PMU fault location is often regarded as a very precise method. The number of PMUs and where they are located in the power grid both affect how precise the data is equation (31) may be used to determine the accuracy. The accuracy of the proposed and current systems is shown in Figure 7. To find faults in transmission lines with multi-terminals, we advise using BDGB-RF. Compared to KNN (82%), XGBoost-SVM (84%), and DL (90%), the suggested approach achieved 97% accuracy. It is proved that the proposed approach is superior to the status order. In addition to this, Table 1 represents the initial sample data proportion and its detection accuracy with different proposed models.

Table 1. Initial data with different models.

Methods	Testing data proportion	Detection accuracy
DL	0.85	98.88
XGBoost-SVM	0.75	98.91
KNN	0.65	98.96
BDGB-RF [Proposed]	0.55	99.8

$$Accuracy(\%) = \frac{TP + TN}{TP + FP + FN + TN}(\%). \tag{31}$$

Power system protection relies on the accurate detection and isolation of faults, and PMU fault finding is a crucial duty in this endeavor. How precisely a PMU can pinpoint the site of a problem on a transmission or distribution line is what is meant by fault location precision. Equation (32) allows for the determination of precision. The precision of the proposed and current systems is shown in Figure 8. To find faults in transmission lines with multi-terminals, while compared to KNN (88%), XGBoost-SVM (76%), and DL (87%), the suggested approach achieved 95% precision. These results demonstrate the suggested method performs better than the previous methods. Table 2 shows the relative precision scores.

$$Precision = \frac{TP}{TP + FP}. \tag{32}$$

The recall of the measurements and the algorithms used to detect faults may also have an impact on the recall of PMU fault localization, as can the location and quantity of PMUs. Having a large enough number of PMUs strategically located around the power system is essential for improving recall in PMU problem locations. Equation (33) allows for the determination of recall. The recall of the proposed and current systems is shown in Figure 9. To find faults in transmission lines with multi-terminals, we advise using BDGB-RF. While Compared to KNN (70%), XGBoost-SVM (74%), and DL (82%), the suggested approach achieved 89% of recall. These results

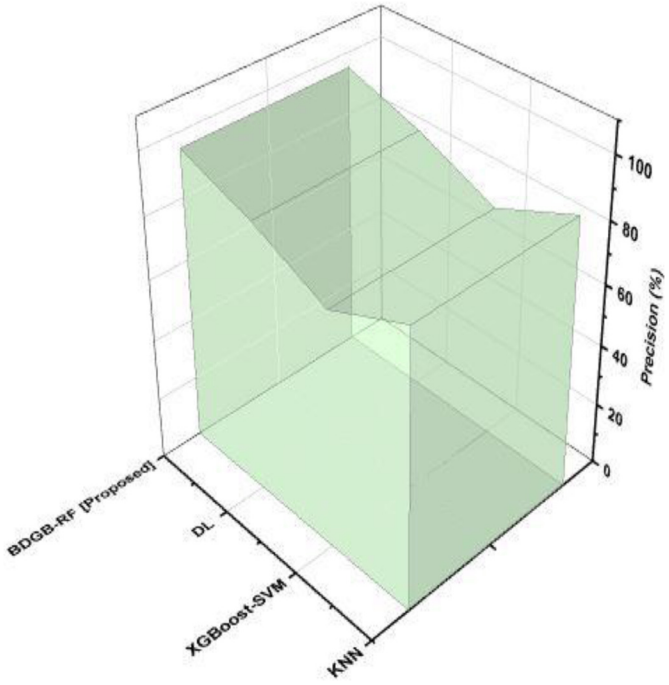

Fig. 8. Precision.

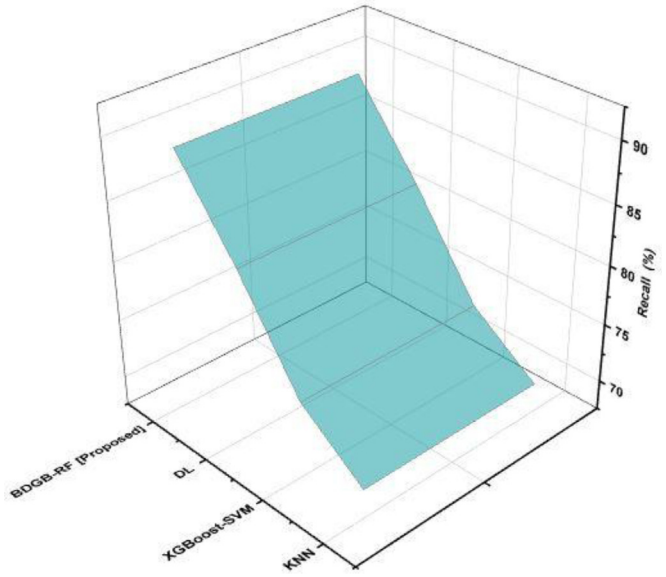

Fig. 9. Recall.

Table 2. Precision.

Precision (%)	
KNN	90
XGBoost-SVM	76
DL	87
BDGB-RF [Proposed]	95

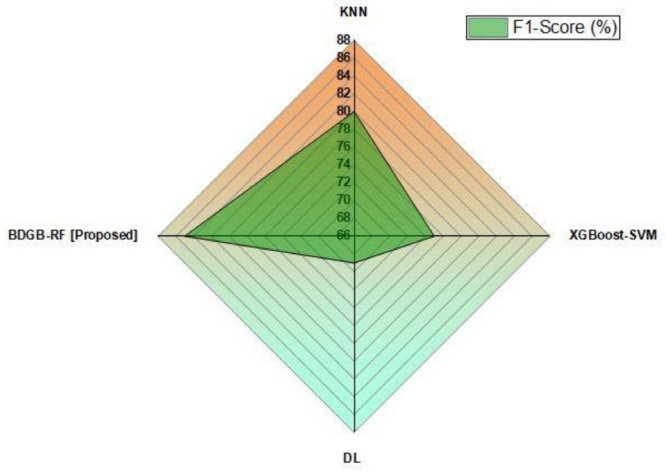

Fig. 10. F1-Score.

demonstrate the suggested method performs better than the previous methods. Table 2 shows the relative Recall scores.

$$Recall = \frac{TP}{TP + FN}. \tag{33}$$

The F1 score evaluates a classification system's efficacy by considering both its accuracy and the number of instances for which it was used. A PMU fault locating system's ability to detect and diagnose problems on transmission and distribution lines may be measured using the F1-score. Equation (34) allows for the determination of the f1-score. The f1-score of the proposed and current systems is shown in Figure 10. To find faults in transmission lines with multi-terminals, we advise using BDGB-RF. When Compared to KNN (80%), XGBoost-SVM (75%), and DL (69%), the suggested approach achieved an 85% f1-score. These results demonstrate the superiority of the suggested method over the current status line.

$$F1 - score = 2\left(\frac{Precision \times Recall}{Precision + Recall}\right). \tag{34}$$

The timely detection of potential problems is critical to ensuring the reliability and stability of the power system and the isolation of faults, making computation time a key concern in PMU fault localization. The time it takes to compute the position of a PMU failure depends on a variety of variables, such as the number and placement of PMUs, the size and complexity of the power system, and the methods used to identify and pinpoint the problem. The computational time of the proposed and current systems is shown in Figure 11. To find faults in transmission lines with multi-terminals, we advise using BDGB-RF. When

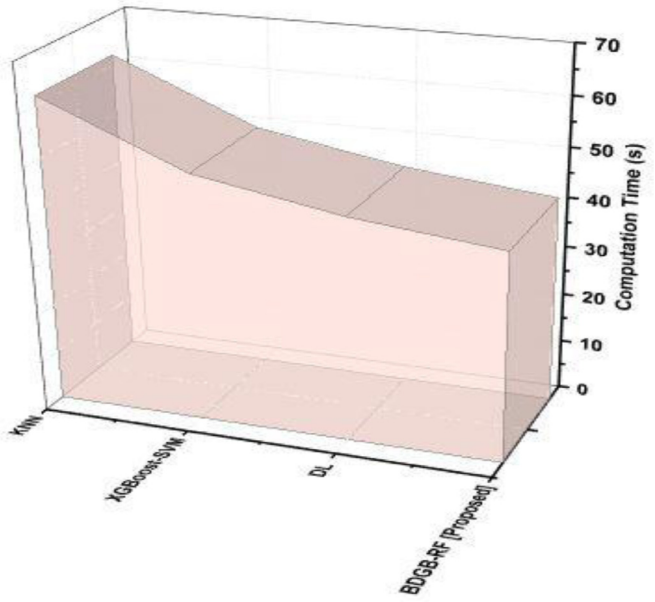

Fig. 11. Computational time.

Compared to KNN (62%), XGBoost-SVM (50%), and DL (45%), the suggested approach achieved 42% computational time. These results demonstrate the superiority of the suggested method performs over the current status line.

9 Discussion

Accuracy of the fault location heavily depends on the quality and quantity of available data for training the K-nearest neighbors (KNN) [23] algorithm. KNN face limitations such as sensitivity to noise, dependency on accurate data, and potential challenges in handling complex network configurations, impacting its robustness and applicability in certain scenarios. The XGBOOST-SVM [24] model may pose challenges, particularly in real-time applications where rapid fault detection and location are crucial. XGBOOST-SVM faces limitations in scalability due to computational demands, potential overfitting concerns, and reliance on accurate fault data for training, which may be challenging to obtain in real-world scenarios. Deep Learning (DL) [25] faces challenges such as limited labeled data, potential model overfitting, and the need for robustness in diverse fault scenarios, impacting the algorithm's accuracy and generalization capabilities.

The proposed BDGB-RF system introduces a novel approach to enhance the performance of tapped four-route transmission lines. This innovation aims to optimize signal transmission and improve overall efficiency in complex network configurations.

10 Conclusion

In this study, we introduced a novel approach; Bidirectional Gradient Boost Random Forest (BDGB-RF) algorithm demonstrates a superior ability to handle complex scenarios in four-route transmission lines with taps, offering a more robust solution compared to traditional fault location methods. The experimental results showed Accuracy (97%), Precision (95%), Recall (89%), F1-Score (85%) and Computation Time (42%). The outcomes of the evaluations showed that the suggested strategy was more effective. The transmission lines are tapped with a temporary and provisional communication range, four-terminal fault identification methods are inefficient Instead of economical. Modifications in the characteristics of input information could have an impact the fault location strategy efficiency. The initial dataset's lack of diversity or the absence of specific factors addressed throughout retraining could have impact on the model's ability to generalize. Future research should concentrate on developing strategies to deploy the fault location model in real-time on power system control platforms. This involves addressing computational challenges and ensuring the model's responsiveness to provide timely fault location information.

Funding

This research study is sponsored by Anhui Province quality engineering project (2020xsxxkc214 and 2016sjjd041).

Conflicts of interest

The authors declare no conflicts of interest.

Data availability statement

The data used to support the findings of this study are available from the corresponding author upon request.

Author contribution statement

The research was devised by Yingyun Wang, and the three of them also wrote the paper. Yang Chen coordinated the effort, and Xiaoxia Qi had a hand in both the design and analysis of the research. Yingyun Wang collected such information as investigation, study and evaluation. The final text was reviewed and approved by all writers.

References

1. M. Dashtdar, M. Dashtdar, Fault location in the distribution network based on phasor measurement units (PMU), Sci. Bull. Electr. Eng. Faculty **19**, 38–43 (2019)
2. A. Saber, PMU-based fault location method for three-terminal transmission lines with an outage of one line branch, Electric Power Syst. Res. **182**, 106224 (2020)
3. Y.J. Lee, T.C. Lin, C.W. Liu, Multi-terminal nonhomogeneous transmission line fault location utilizing synchronized data, IEEE Trans. Power Delivery, **34**, 1030–1038 (2019)
4. T.C. Lin, Z. Xum, F.B. Ouedraogo, Y.J. Lee, A new fault location technique for three-terminal transmission grids using unsynchronized sampling, Int. J. Electr. Power Energy Syst. **123**, 106229 (2020)

5. T.C. Lin, J.Z. Yang, C.S. Yu, C.W. Liu, Development of a transmission network fault location platform based on cloud computing and synchrophasors, IEEE Trans. Power Delivery **35**, 84–94 (2019)
6. H. Mirshekali, R. Dashti, A. Keshavarz, A.J. Torabi, H.R. Shaker, A novel fault location methodology for smart distribution networks, IEEE Trans. Smart Grid **12**, 1277–1288 (2020)
7. B. Mallikarjun, G. Pathirikkat, D. Sinha Roy, J.B.R. Maddikara, A real-time synchronized harmonic phasor measurements-based fault location method for transmission lines, J. Control Autom. Electr. Syst. **30**, 1082–1093 (2019)
8. Y. Jia, Y. Liu, B. Wang, D. Lu, Y. Lin, Power network fault location with exact distributed parameter line model and sparse estimation, Electric Power Syst. Res. **212**, 108137 (2022)
9. V.K. Gaur, B.R. Bhalja, Synchrophasor based fault distance estimation method for tapped transmission line. In *2019 International Conference on Smart Grid Synchronized Measurements and Analytics* 2019, May (SGSMA), IEEE, pp. 1–5
10. M. Azeroual, Y. Boujoudar, K. Bhagat, L. El Iysaouy, A. Aljarbouh, A. Knyazkov, M. Fayaz, Fault location and detection techniques in power distribution systems with distributed generation: Kenitra City (Morocco) as a case study, Electric Power Syst. Res. **209**, 108026 (2022)
11. M. Jamei, R. Ramakrishna, T. Tesfay, R. Gentz, C. Roberts, A. Scaglione, S. Peisert, Phasor measurement units optimal placement and performance limits for fault localization, IEEE J. Sel. Areas Commun. **38**, 180–192 (2019)
12. M. Dashtdar, A. Hussain, H.Z. Al Garni, A.A. Mas'ud, W. Haider, Fault location in distribution network by solving the optimization problem based on power system status estimation using the PMU, Machines, **11**, 109 (2023)
13. A. Khan, A.Q. Khan, M. Sarwar, M. Abubakar, N. Iqbal, An accurate fault location algorithm for meshed power networks utilizing hybrid sparse voltage and current measurements, arXiv preprint 2019. arXiv:1906.12113
14. H. Manoharan, Y. Teekaraman, R. Kuppusamy, A. Radhakrishnan, Application of solar cells and wireless system for detecting faults in phasor measurement units using non-linear optimization, Energ. Explor. Exploit. **41**, 210–223 (2023)
15. F.L. Grando, A.E. Lazzaretti, M. Moreto, H.S. Lopes, Fault classification in power distribution systems using pmu data and machine learning. In *2019 20th International Conference on Intelligent System Application to Power Systems (ISAP)*, 2019, December, IEEE, pp. 1–6
16. G. Paramo, A. Bretas, S. Meyn, High-impedance non-linear fault detection via eigenvalue analysis with low pmu sampling rates, 2023. arXiv preprint arXiv:2301.04123
17. M.Z.C. Wanik, A. Sanfilippo, N. Singh, A. Jabbar, Z. Cen, PMU analytics for power fault awareness and prediction. In *2019 International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA)*, 2019, May, IEEE, pp. 1–8
18. K. Likhitha, O.D. Naidu, Setting free fault location for three-terminal hybrid transmission lines connected with conventional and renewable resources, IEEE Access **11**, 23839–23856 (2023)
19. R. Patekar, P. Yadav, S.R. Tambay, Effect of frequency variation on fault location in transmission line using synchrophasor technology. In *2019 3rd International Conference on Recent Developments in Control, Automation & Power Engineering (RDCAPE)*, 2019, October, IEEE, pp. 70–75
20. B.P. Bandi, S.R. Tambay, I. Srivastava, Fault location for large transmission networks using synchronized voltage measurements. In *2020 IEEE First International Conference on Smart Technologies for Power, Energy, and Control (STPEC)*, 2020, September, IEEE, pp. 1–6
21. J. An, L. Yang, Z. Hao, G. Chen, L. Li, Investigation on road underground defect classification and localization based on ground penetrating radar and Swin transformer, Int. J. Simul. Multidiscip. Des. Optim. **15**, 7 (2024)
22. E.S. Maputi, R. Arora, Design optimization of a three-stage transmission using advanced optimization techniques, Int. J. Simul. Multidiscip. Des. Optim. **10**, A8 (2019)
23. D.O. Amoateng, R. Yan, M. Mosadeghy, T.K. Saha, Topology detection in power distribution networks: a PMU based deep learning approach, IEEE Trans. Power Syst. **37**, 2771−2782 (2022)
24. A. Swetapadma, A. Yadav, Comput. Electr. Eng. **69**, 41–53 (2021)
25. K. Liu, T. Kang, X. Ye, M. Bai, Y. Fan, A fault location method of a distribution network based on XGBoost and SVM algorithm, IET Cyber-Phys. Syst.: Theor. Appl. **6**, 254–264 (2021)

Cite this article as: Yingyun Wang, Xiaoxia Qi, Yang Chen, Enhanced fault localization in multi-terminal transmission lines using novel machine learning, Int. J. Simul. Multidisci. Des. Optim. **15**, 15 (2024)