ORIGINAL ARTICLE

# Identification and Detection of Electrical Fault Location in Underground Cables Using Machine Learning Algorithms

**Manjusha A. Kanawade[1], Rahul Sudam Gavhale[2], Sujit Shankar Kshirsagar[2], Aditya Vaijnath Hanmante[2], and Satwashil Sharad Ambi[2]**

[1]*Department of Electrical Engineering, JSPMs Rajarshi Shahu College of Engineering, Pune, India*
[2] *Department of Electrical Engineering, JSPMs Rajarshi Shahu College of Engineering, Pune, India*
(**Received** 22 December 2025)
*Corresponding author : makanawade_elec@jspmrscoe.edu.in
DOI: 10.5875/schx4x49

**Abstract:** Underground power cables are key components of advanced power systems because of their safety, reliability, and beautiful advantages. However, the identification and localization of electrical faults in these cables remain challenging because of their difficulty in nature and the complication of fault behaviour. This paper introduces a machine learning–based approach using the random forest classifier and random forest regressor algorithms for efficient fault detection and location estimation in underground power cables. Voltage and Current data are collected from both sending and receiving ends under different fault conditions, such as conductor-to-ground (CG), conductor-to-conductor (CC), double conductor-to-ground (DCG), and three-conductor to ground (3CG) faults. The random forest classifier is developed to identify fault types based on extracted electrical features, while the random forest regressor predicts the exact fault distance using ohm's law–based calculation. The model is developed and validated using MATLAB Simulink and hardware setup. Tests confirm that this system performs with greater accuracy and stability than previous methods, quick and validated fault detection in underground power cables.

**Keywords:** Fault detection; location; Machine Learning; MATLAB Simulink; Random Forest; Regression; Underground cable

## Introduction

The operation of underground power cables is important for reliable power distribution systems. Over the past few years there have been more and more underground cable developments replacing overhead transmission lines due to the advantages of greater safety, reduced footprint, and reduced impact of adverse weather or environmental conditions on the system. A significant weakness of underground cables is that faults in the cable can be more difficult to detect and locate because they cannot be inspected visually.

Typically, classical methods of cable fault detection (e.g., distance-to-fault, bridge method, impedance) and Time Domain Reflectometry (TDR) have been the most commonly used methods to detect faults in underground cables [1][2]. While these techniques have some success under optimal conditions, they require extensive testing equipment, accurate knowledge of all cable parameters, and very skilled operators to produce acceptable results. Performance will degrade with changes in loading and variations in cable parameters. In addition, identifying multiple or widely distributed cable faults is typically very hard using the above techniques, resulting in increased repair times and prolonged power outages. IoT-based fault detection systems have also been introduced to improve monitoring by continuously collecting voltage and current data. The speed of fault detection using IoT technology is superior when compared with the older methods; however, the majority of IoT-based damage detection relies on predetermined or definable levels/thresholds (threshold limits) [3][4][5][6]. Therefore, many IoT-based methods of detecting and/or rectifying complex fault situations such as double-conductor-to-ground or multiple-conductor-type faults may lack proper


Copyright © 2024 International Journal of Automation and Smart Technology

identification. For these types of faults, the estimated distance from where the fault originated is often not an accurate representation of the actual fault location and thus decreases system reliability.

To address these limits, there have been advancements in using ML technology to perform underground cable fault identification and locating [7][8][9]10]. Machine learning algorithms will analyze voltage/current signals over time and identify very complex/non-linear patterns in these signals, which would otherwise be difficult to extract using either the traditional or the Internet of Things (IoT)-based fault DA methods. Unlike the traditional fault detection/diagnosis methods, machine learning algorithms can work with massive datasets and learn from the changes to provide greater precision and adaptability when used at multiple locations.

Artificial Neural Networks (ANNs) were utilized by researchers in the early stages of this research due to their capability of modelling nonlinear relationships between input and output data (e.g., voltages and currents). ANN-based approaches delivered fair levels of accuracy for detecting single-conducting faults when simulated voltage and current signals created through MATLAB were analyzed. However, they rely heavily on the quality of the training dataset and the structure of the neural network, meaning that they will likely need to be re-trained with any new fault conditions. Subsequently, Support Vector Machines (SVMs) were introduced and provided an improved means of generalization than the ANN method, however the effectiveness of SVMs still relies greatly on tuning hyperparameters and computational time making them impractical for rapid implementation. Decision tree-based approaches provide the user with a more interpretable mechanism, and ease of use, however they tend to be prone to overfitting and therefore have poor performance with complex datasets when used in solitarily. The advent of machine learning methods such as Random Forest (RF) provides solutions to these two issues discussed above. Random Forest integrates the outputs of numerous decision trees, thus increasing the stability and accuracy, and also providing a robust response to noise and overfitting [11][12][13].

Taking these advantages into account, this paper offers a machine learning approach for the detection and location of faults using the dual Random Forest algorithm for fault detection. The Random Forest Classifier can classify the type of fault as either single, double or triple conductor to ground faults, and a Random Forest Regressor will predict the precise location of the fault along the cable by predicting its location. Both Voltage and Current Data are created using a high-fidelity MATLAB Simulink simulation based on theoretical principles, and we will be conducting our training and testing of our models in the Python environment. This combination of using a simulation program and machine learning algorithm enables an accurate prediction of the occurrence of a fault, a quicker response time, and consistent performance, even under complicated fault conditions. By creating a link between a simulation to generate synthetic data and machine learning to analyze it, the new approach could provide a practical, effective solution for detecting faults in underground cables in a modern Smart Grid environment.

## Methodology

The system operates in two main parts. The first is a simulation part developed in MATLAB Simulink, where voltage and current datasets are generated under different fault conditions. The second is a machine learning part implemented in Python, where the collected data are analyzed using Random Forest Classifier (RFC) and Random Forest Regressor (RFR) models. Once disciplined and verified, the models are integrated with an ESP32 microcontroller to enable real-time prediction of fault type and fault location. This combine allows both simulation-based testing and hardware-level validation under practical operating conditions.

### *Simulation Model and Data Generation*

A detailed three-phase underground cable model was developed in MATLAB Simulink using standard power system components. The model includes a three-phase source, sending-end voltage and current sensors, an underground cable section, fault creation units, receiving-end sensors, and a variable load. different fault types were simulated, including conductor-to-ground (CG), conductor-to-conductor (CC), double conductor-to-ground (DCG), and three-conductor to ground (3CG) faults. Faults were introduced at random locations along the cable by varying fault resistance. During each simulation, voltage (VR, VY, VB) and current (IR, IY, IB) waveforms were obtain from both the sending and receiving ends [14][15]. In this work, the most frequently observed underground cable fault types—conductor-to-ground (CG), conductor-to-conductor (CC), double conductor-to-ground (DCG), and three-conductor-to-ground (3CG)—are considered, as these faults account for the majority of real-world fault conditions in distribution networks. The proposed Random Forest–based approach is inherently scalable and flexible. If additional or previously unseen fault types occur, they can be accommodated by simply augmenting the training dataset with new fault scenarios and retraining the model. Owing to its data-driven learning capability, the framework can effectively adapt to new fault patterns without requiring any changes to the system


Copyright © 2024 International Journal of Automation and Smart Technology

structure.

### Data Preparation and Feature Extraction

The electrical parameters of measured waveforms were determined with basic Ohm's law calculations and impedance analysis. Features included changes in current (ΔI), voltage changes (ΔV), impedance magnitude ($|Z| = V/I$), and changes in phase angle. Min–max normalization was applied to each input feature separately using the minimum and maximum values obtained from the training dataset. These normalization parameters were then kept constant and used during both the testing phase and real-time inference. This ensured consistent feature scaling across all stages and prevented any data leakage from the test or operational data into the training process. Missing or abnormal readings occurred mainly due to sensor noise, transient switching effects, and numerical instability during fault initiation in simulation. These instances accounted for less than 3–5% of the total dataset. Such samples were removed to prevent misleading the learning process. Given the small proportion of discarded data, imputation was not applied; however, robust learning or noise modeling techniques may be explored in future work.

The scikit-learn library in Python was used to create a Random Forest Classifier (RFC) to identify faults with RFCs. RFCs are composed of many decision trees ('ensemble') trained on randomly sampled data (superset) and features (subsets). The final fault type is determined by a majority voting process across all of the trees within the given RFC ensemble. Grid search cross-validation was used to tune hyper-parameters that are important (e.g., number of trees, maximum depth of trees, minimum number of samples for splitting). An RFC can be used to predict 4 categories of faults: CG, CC, DCG, 3CG. The Random Forest Classifier achieved an overall accuracy of approximately 95%, with precision, recall, and F1-score values consistently above 0.92 for all fault categories, indicating reliable and balanced fault classification performance. [16][17].

### Random Forest Regression for Locating Faults

Using Random Forest Regression to predict the location of the fault on a cable, the features produced from this process will be combined to form a model that indicates the relationship between these electrical characteristics and the corresponding distances of where the faults occurred. Each Random Forest Tree produces a distance value for the fault location and the average of all Tree distance values is used as the final output, improving both stability of results and decreasing the chance for error. When evaluating the performance of the Random Forest Regression model, three different metrics were used to assess accuracy, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R square ($R^2$). Considering the total cable length of 80 km, a mean absolute error below 2 km corresponds to an error of less than 2.5%, which is acceptable for practical maintenance and excavation planning. This accuracy significantly reduces fault search time compared to traditional trial-and-error or IoT threshold-based methods [18][19].

### Real-time Testing & Hardware Integration

The development of the hardware prototype provided confirmation of the simulation outcomes. Two voltage and current sensors were inserted into a small-scale underground cable configuration; these sensors were used to produce multiple types of fault conditions via adjustment of the resistive load with a rheostat. An ESP32 microcontroller was used to collect and transmit the data from the sensors to a computer running a Python interface via serial communication. Python loaded and executed the previously trained Random Forest models to identify the type of fault, and predict the fault's distance, in real-time as new data was provided. A 16 by 2 Liquid Crystal Display (LCD) was connected to the ESP32, allowing for real-time display of the model's outputs. The use of this particular hardware setup, using an ESP32 microcontroller, verifies the validity of the proposed Machine Learning approach for identifying and estimating fault location in real-time applications [20][21][22].

## A. SIMULATION Implementation

The initial model developed was one of a 3-phase underground cable using MATLAB Simulink which included the following components: A 3-phase power source; voltages from a 3-phase load; an underground cable segment; a fault simulator with randomly located unbalanced and balanced faults; and a variable load. Currents and voltages will be recorded at both the sending and receiving end of every run; these measurements will be exported into the MATLAB workspace for conversion into a dataset for use in the development of a machine-learning algorithm. The changes in impedance and phase angle will also be calculated by using Ohm's Law to create additional. The developed MATLAB SIMULINK model of a


Copyright © 2024 International Journal of Automation and Smart Technology

underground cable is shown in Fig. 1. and Fig. 2.

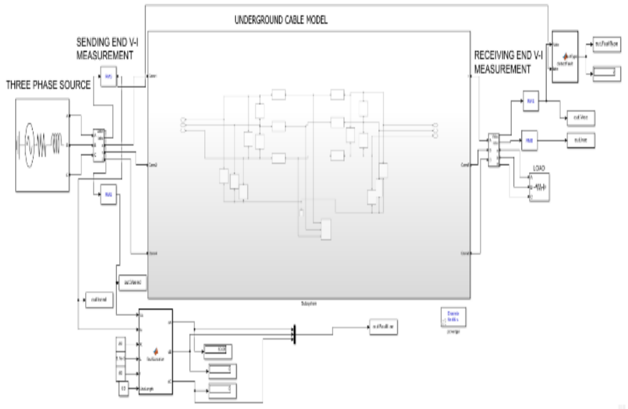



*Fig. 1: Complete MATLAB Simulink Model of Underground Power Cable*

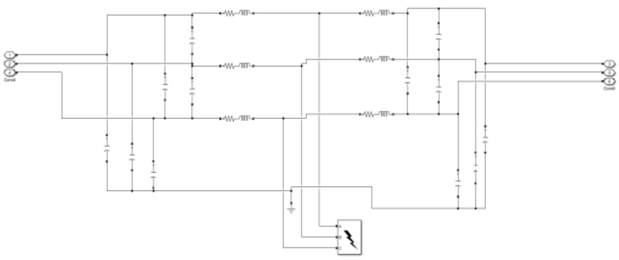


Fig.2: Simulink Subsystem Model of Underground Power Cable

## B. HARDWARE Implementation

To confirm the accuracy of the results obtained from the simulation, a practical laboratory experimental arrangement was established, consisting of several real physical devices, as depicted in Fig.3. An auto-transformer provided an adjustable three-phase supply voltage to the system. Ammeters were utilized to measure the current flowing through the lamp load to simulate the balanced load. Voltage and current sensors were positioned at both extremities of the underground cable and a rheostat was employed to simulate varying fault conditions by varying the resistance across the cable.

The sensors collected electrical current and voltage measurements in real-time and forwarded them to an ESP32 microcontroller. The microcontroller then relayed the sensor data to a Python-based program running the machine learning models developed during training. The outputs generated by the machine learning models, i.e., the type of fault, its location, and the measured output(s), were displayed on a liquid crystal display (LCD) connected to the microcontroller [23][24]. The combination of both hardware and software components as implemented in this application created an effective bridge between simulated laboratory testing and actual laboratory results. The fact that the same model trained with MATLAB simulation data produced similar results when tested on 'real' hardware measurements serves to validate both the strength and versatility of the model as applied in this situation.

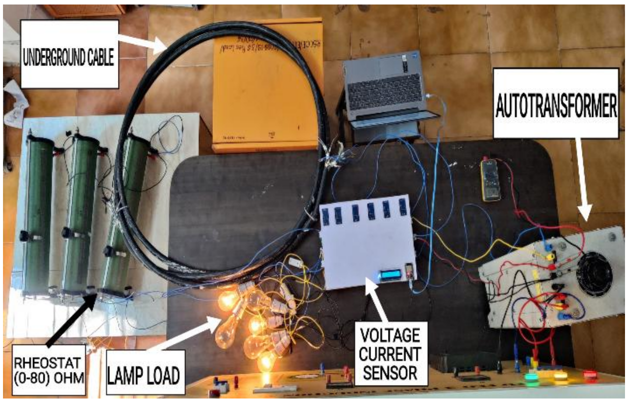



Fig.3 : Experimental Hardware Setup under Operating Condition

## C. MACHINE LEARNING MODEL Implementation

Python's purpose included data preparation, training, and evaluating the performance of the model. The Dataset was generated from Simulations completed within MATLAB Simulink and includes voltage or current readings taken from both sending and receiving ends of the underground cable while subjecting the cable through various operational and fault scenarios. The Dataset was subjected to multiple pre-processing procedures to prepare for the training of the learning model before utilizing the data for that purpose. These pre-processing steps included the elimination of incomplete and/or abnormal samples, normalization of the Data Features per Data Type, and the division of the Dataset into two Subsets (for training and testing) prior to using the data for Model Training.

Feature extraction occurred using Raw Electrical Measurements and Derivative Parameters. For example, to create the voltage difference between the sending- and receiving-end voltages, we calculate the receiving-end voltages minus the corresponding sending-end voltages. Due to these additional Features, the Learning Models are more sensitive to fault-induced variations in the cable, making it easier to distinguish between different Fault Types. Two models based on the Random Forest algorithm were used in this work.

The Random Forest Classifier (RFC) was used to identify the type of fault and the Random Forest Regressor (RFR) was used to estimate the location of the fault. The numerical encoding of categorical fault labels was accomplished using a label encoder for the identification of faults, while for locating faults, the multi-output regression method was used to produce simultaneous


Copyright © 2024 International Journal of Automation and Smart Technology

predictions of the fault distance for all phases of the conductor. The regression dataset was filtered to remove all samples without a specified fault location so that these samples would not skew the prediction results.

All samples were randomly assigned to training and test sets at an 80:20 split, ensuring that all categories of faults appeared in both sets at roughly the same frequency. One reason the Random Forest algorithm was chosen was because of its resistance to noise in electrical data, decreased likelihood of overfitting and its ability to model non-linear relationships between voltage-current patterns and fault characteristics. Additionally, the ensemble structure of Random Forest models improves their ability to generalize compared to individual decision trees by aggregating multiple trees trained on random subsets of samples and features. Another difference between Random Forest and SVM-based models is that Random Forest models require less parameter sensitivity tuning and scale better to handle very large datasets.

The hyperparameters essential for trees include their quantity and their size, and to find the appropriate combination of these two factors, we used a grid search technique. Also, in order to decrease the amount of time required for the training phases of the trees, we used parallel processing while the trees were being trained. Additionally, once we had trained our trees, we saved the model(s) that we had created for both the classification and regression phases of tree execution, which we could then utilize at a later time for actual inference when using the trees on physical hardware.

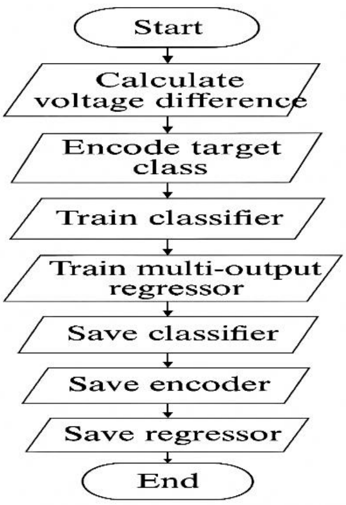



Fig. 4: Flowchart of the machine learning model training

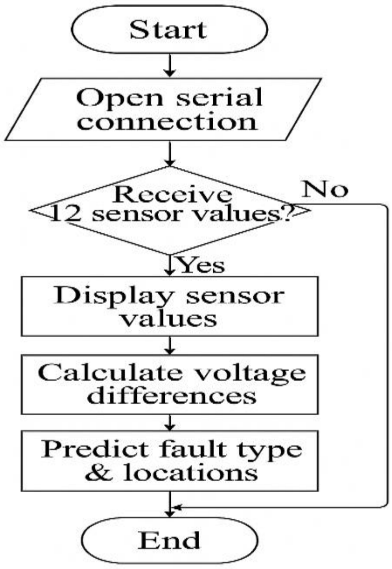
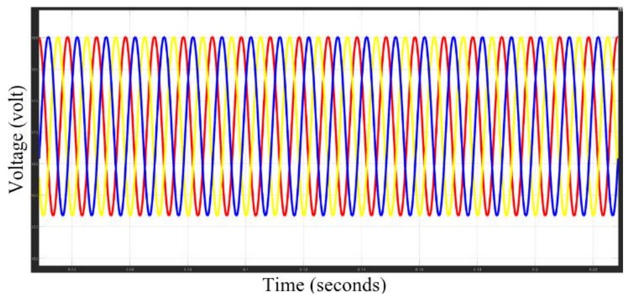



*Fig. 5: Flowchart of real-time fault detection and localization process.*

## Results and Discussion

To validate the proposed system for underground power cable fault detection and localization, simulations were carried out in MATLAB Simulink under different operating conditions including healthy, single conductor-to-ground, conductor -to- conductor, and multi-phase faults. The corresponding voltage and current waveforms were analysed to understand the fault characteristics, and the results were used for Learning and validation the machine learning model.

Under healthy cable conditions, the three-phase voltage and current waveforms remain balanced and purely sinusoidal, as shown in Fig. 7(a) and Fig. 7(b). All three




phases maintain nearly equal magnitude and a constant 120° phase separation, indicating that no disturbance or fault exists in the system.

*Fig. 6(a): Voltage waveform under healthy* condition


Copyright © 2024 International Journal of Automation and Smart Technology

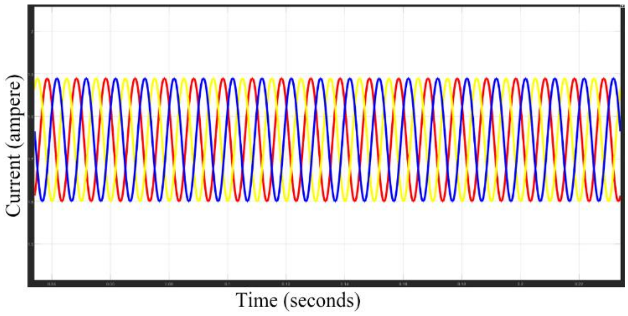



Fig. 6(b): Current waveform under healthy condition

When a conductor(R)-to-ground(G) fault Appears, the voltage of the affected phase drops sharply, and the amount of this voltage drop depends on distance between the fault point and the sending terminal. Faults closer to the source result in a more significant voltage collapse. Simultaneously, the current in the faulted line increases considerably, with its magnitude determined by the impedance of the cable lower impedance leads to higher fault current. The other two phases remain almost unaffected, confirming that the fault is isolated to a single conductor. The corresponding Voltage and Current Waveforms, shown in Fig. 8(a) and Fig. 8(b), clearly depict this behavior, where one phase experiences a sharp voltage drop and current rise while the others maintain normal operation.

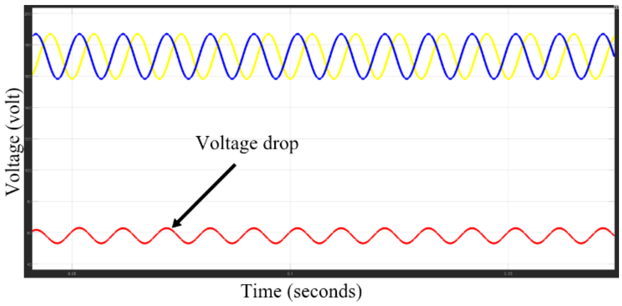



Fig. 7(a): Voltage waveform under RG fault

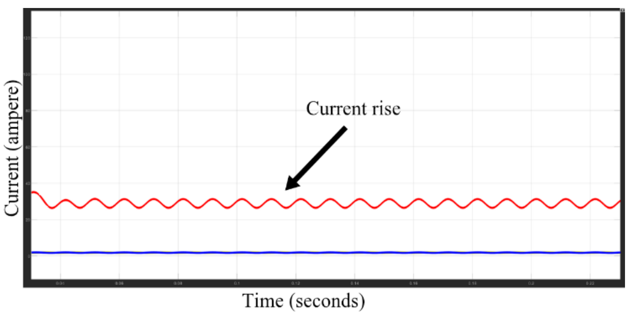



*Fig. 7(b): Current waveform under RG fault*

In the case of a conductor(R)-to-conductor(Y) fault, both faulted conductors experience a sharp drop in voltage, while their current amplitudes rise significantly, as shown in Fig. 9(a) and Fig. 9(b). The extent of the voltage drop depends distance between the fault point and the sending terminal faults occurring closer to the source exhibit greater voltage reduction. At the same time Variations in cable impedance directly impact the extent of the fault current flowing through the system., with lower impedance resulting in higher current flow. The third conductor maintains a steady and unaffected waveform, confirming that the short circuit is limited to the two involved conductors.

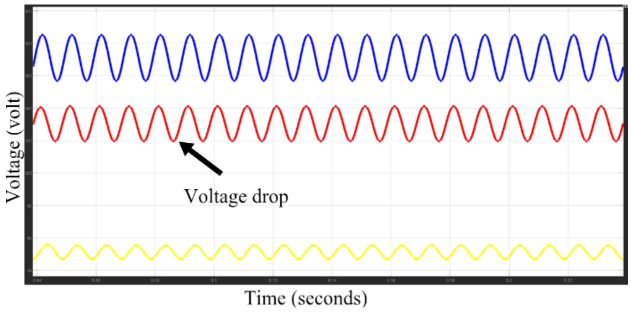



Fig. 8(a): Voltage waveform under RY fault

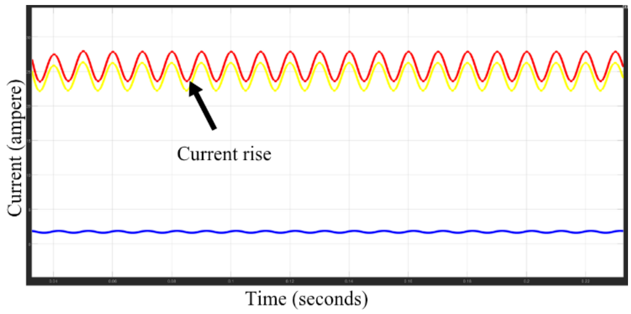



*Fig. 8(b): Current waveform under RY fault*

During a double conductor (RY)-to-ground(G) fault, two of the three conductors are connected to the ground simultaneously. In this condition, the voltages of the faulted conductors decrease sharply, and the extent of voltage drop depends on distance between the fault point and the sending terminal. Faults occurring closer to the source show a more pronounced voltage collapse. At the same time, the fault currents in the affected conductors increase significantly, with their magnitudes determined by the impedance of the cable. A lower impedance path allows higher fault current flow. The third conductor, which remains un-faulted, shows only minor fluctuations in voltage and current due to the imbalance introduced by the fault. The corresponding voltage and current waveforms are shown in Fig. 10(a) and Fig. 10(b), illustrating the relationship between voltage drop, fault distance, and current variation with cable impedance.


Copyright © 2024 International Journal of Automation and Smart Technology

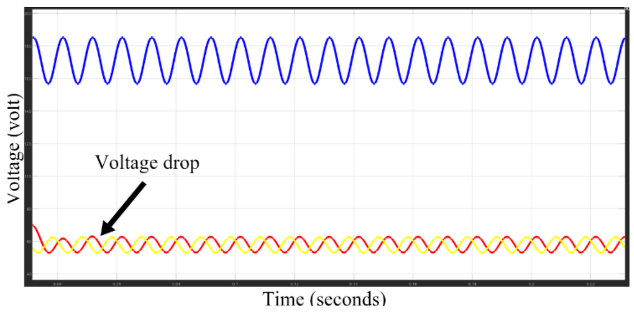



*Fig. 9(a): Voltage waveform under RYG fault*

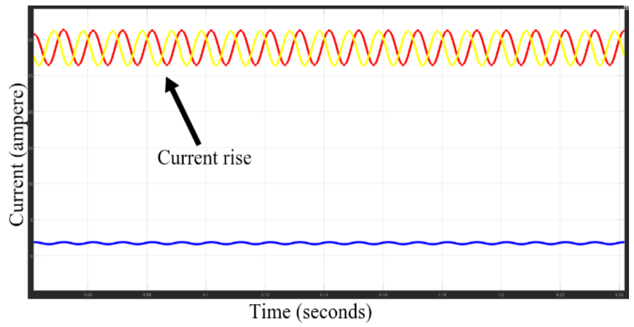



Fig. 9(b): Current waveform under RYG fault

The three-conductor-to-ground fault (RYBG) is the most severe type of fault that can occur in an underground cable system. During this condition, the voltages of all three conductors drop sharply, and the level of voltage reduction depends on distance between the fault point and the sending terminal. Faults occurring closer to the source show a greater voltage drop compared to those farther away. At the same time, the fault current rises significantly, with its magnitude determined by the impedance of the cable. A lower cable impedance results in higher current flow during the fault condition. The corresponding voltage and current waveforms, shown in Fig. 11(a) and Fig. 11(b), clearly illustrate the near-zero voltage levels and large surge in current, representing a complete system breakdown and the most critical scenario for cable operation.

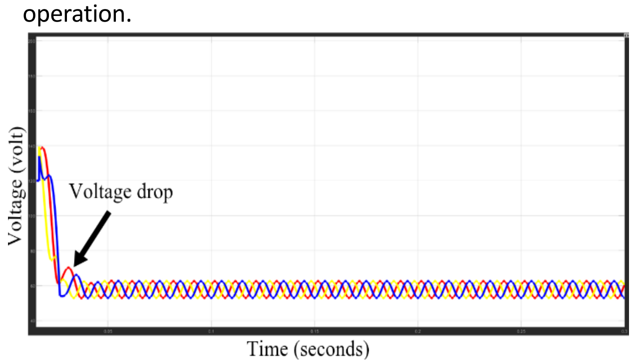



*Fig. 10(a): Voltage waveform under RYBG fault*

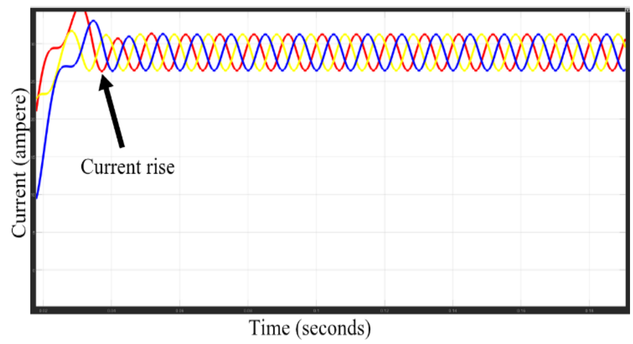



Fig. 10(b): Current waveform under RYBG fault

The proposed underground cable fault detection system was demonstrated in real-time hardware control as illustrated in Fig 12. The physical hardware implementations on this project consisted of an ESP32 microcontroller board and a typical 16×2-character LCD (liquid crystal display). The microcontroller receives voltage and current measurements from the sensors and uses its machine learning artificial intelligence algorithms to analyse and classify based on the fault condition.

As a result, the LCD provides real-time visual feedback of the type of fault that is detected as well as the distance from the fault back to the source of the fault for each conductor (A, B, C). In the first example, an AG (conductor-to-earth) fault was detected, and the distance to the fault was displayed only on the phase (A) that was affected, while it was indicated as 0 for the other phases. In the second example, the real-time hardware implementation successfully confirmed that there was no fault detected and confirmed that all conductors were functioning as designed.

The hardware testing was consistent with the previous results of the MATLAB simulations and machine learning algorithms and also confirmed that the developed fault detection system can be used in the field as well as used offline. In addition, confirmation of the real-time visualisation from the hardware demonstrates the feasibility, accuracy, and dependability of the system to be utilised in smart grid and power distribution networks through underground cables.



Copyright © 2024 International Journal of Automation and Smart Technology

*Fig.11: Experimental Hardware Setup Displaying Detected Fault Type and Phase-wise Fault Distance*

Table 1 illustrates the comparative results of the proposed system for fault detection and localization using MATLAB simulation and machine learning techniques. In column 2, the results obtained from the MATLAB Simulink model are presented. The table lists the measured sending- and receiving-end voltages and currents for all three phases under normal operating conditions. The calculated values of phase voltages and currents remain balanced, and the corresponding fault locations are zero, confirming a no-fault condition.

Table.1: Comparison of Analytical, Simulation, Machine Learning and Hardware Results for No Fault condition

| Parameters | Analytical Results | Simulation results | ML results | Hardware results |
|---|---|---|---|---|
| $V_R$ Sending end (Volt) | 180.32 | 180.32 | 179.76 | 179.76 |
| $V_Y$ Sending end (Volt) | 158.21 | 158.21 | 158.22 | 158.22 |
| $V_B$ Sending end (Volt) | 180.06 | 180.06 | 180.62 | 180.62 |
| $I_R$ Sending end (Amp) | 0.26 | 0.26 | 0.26 | 0.26 |
| $I_Y$ Sending end (Amp) | 0.24 | 0.24 | 0.24 | 0.24 |
| $I_B$ Sending end (Amp) | 0.27 | 0.27 | 0.27 | 0.27 |
| $V_R$ Receiving end (Volt) | 162.70 | 162.70 | 158.76 | 158.76 |
| $V_Y$ Receiving end (Volt) | 159.26 | 159.26 | 139.56 | 139.56 |
| $V_B$ Receiving end (Volt) | 158.62 | 158.62 | 159.15 | 159.15 |
| $I_R$ Receiving end (Amp) | 0.26 | 0.26 | 0.26 | 0.26 |
| $I_Y$ Receiving end (Amp) | 0.24 | 0.24 | 0.23 | 0.23 |
| $I_B$ Receiving end (Amp) | 0.26 | 0.26 | 0.27 | 0.27 |
| Fault location R phase (km) | 0 | 0 | 0 | 0 |
| Fault location Y phase (km) | 0 | 0 | 0 | 0 |
| Fault location B phase (km) | 0 | 0 | 0 | 0 |
| Fault type | No Fault | No Fault | No Fault | No Fault |

**ML - Machine learning**

This MATLAB-based simulation data serves as the foundational dataset used for training and testing the machine learning models. The simulation ensures that all fault and healthy conditions are accurately represented before experimental implementation. In column 3 presents the machine learning-based predictions derived from the trained Random Forest Classifier and Random Forest Regressor. Using the same twelve sensor inputs—three-phase voltages and currents from both ends of the cable—the classifier identifies the fault type, while the regressor estimates the fault location in kilo meters. The output from the Python command window shows consistent results with the MATLAB simulation, where no faults are detected in the given test cases. The identical outcomes from both MATLAB and machine learning analysis validate that the trained model has successfully learned the cable's electrical behaviour under different conditions.

*Table.2: Comparison of Analytical, Simulation, Machine Learning and Hardware Results for RG Fault*

| Parameters | Analytical Results | Simulation results | ML results | Hardware results |
|---|---|---|---|---|
| $V_R$ Sending end (Volt) | 178.15 | 178.15 | 177.95 | 177.95 |
| $V_Y$ Sending end (Volt) | 158.19 | 158.19 | 158.21 | 158.21 |
| $V_B$ Sending end (Volt) | 180.96 | 180.96 | 180.64 | 180.64 |
| $I_R$ Sending end (Amp) | 3.55 | 3.55 | 3.49 | 3.49 |
| $I_Y$ Sending end (Amp) | 0.23 | 0.23 | 0.23 | 0.23 |
| $I_B$ Sending end (Amp) | 0.27 | 0.27 | 0.28 | 0.28 |
| $V_R$ Receiving end (Volt) | 0.14 | 0.14 | 0.14 | 0.14 |
| $V_Y$ Receiving end (Volt) | 139.82 | 139.82 | 139.82 | 139.82 |
| $V_B$ Receiving end (Volt) | 158.62 | 158.62 | 158.62 | 158.62 |
| $I_R$ Receiving end (Amp) | 0.00 | 0.00 | 0.00 | 0.00 |
| $I_Y$ Receiving end (Amp) | 0.23 | 0.23 | 0.23 | 0.23 |
| $I_B$ Receiving end (Amp) | 0.26 | 0.26 | 0.27 | 0.27 |
| Fault location R (km) | 50.14 | 50.05 | 52.94 | 52.94 |
| Fault location Y (km) | 0 | 0 | 0 | 0 |
| Fault location B (km) | 0 | 0 | 0 | 0 |
| Fault type | RG Fault | RG Fault | RG Fault | RG Fault |

**ML - Machine learning**

Table 2 provide a comparison of fault-location results


Copyright © 2024 International Journal of Automation and Smart Technology

obtained from the analytical method, MATLAB simulations, and machine-learning predictions for a conductor-to-ground (RG) fault. The analytical calculation uses basic electrical measurements from the cable to estimate the distance of the fault. In the R-phase, the sending-end voltage is 178.15 V and the receiving-end voltage drops almost to zero (0.14 V), giving a total voltage drop of 178.01 V. This voltage drop, when divided by the total current of 3.55 A, results in a calculated fault impedance of approximately 50.14 Ω. Since the cable impedance is 1 Ω per km, the impedance value directly translates to a fault distance of about 50.14 km from the sending end. This analytical result represents the theoretical fault location based on Ohm's law [17].

The results generated by MATLAB in column 3 indicate that the predicted fault distance is approximately 50.05 km for both simulation cases, which is nearly identical to that of the analytical case and differs by less than 0.5 km. The small discrepancy is not unexpected given that MATLAB accounts for many factors associated with practical cable parameters (e.g., capacitance and resistance) as well as numerical effects from solving the equations, and these factors resulted in only slight differences between the results generated by MATLAB and the analytical results. Thus, it is clear that the simulation-based model of the cable is able to represent the actual cable in a very accurate manner.

In addition to the results provided in column 3, the results generated from machine learning (column 4) also demonstrate that the estimates of fault location fall within a very similar range, with machine-learning-generated fault location estimates ranging between approximately 52.94 km and 50.73 km. The variations of 1–2 km between these two sets of machine-learning fault location estimates are reasonable expectations considering that machine learning models operate on historical data to develop their predictive capabilities, rather than deriving equations with exact mathematical relationships. Regardless of the small differences noted between the two machine-learning fault location estimates, it is still clear that the results are very close to both the MATLAB and analytical results, indicating that the machine-learning-based fault location estimation model is capable of performing reliably and generating fault location estimates that are within a small range of the actual cable distance.

Table.3: Comparison of Analytical, Simulation, Machine Learning and Hardware Results for RYG Fault

| Parameters | Analytical Results | Simulation results | ML results | Hardware results |
|---|---|---|---|---|
| $V_R$ Sending end (Volt) | 178.98 | 178.42 | 178.51 | 178.51 |
| $V_Y$ Sending end (Volt) | 155.75 | 155.64 | 155.64 | 155.64 |
| $V_B$ Sending end (Volt) | 180.00 | 180.63 | 180.51 | 180.51 |
| $I_R$ Sending end (Amp) | 2.88 | 2.87 | 2.87 | 2.87 |
| $I_Y$ Sending end (Amp) | 2.51 | 2.51 | 2.50 | 2.50 |
| $I_B$ Sending end (Amp) | 0.27 | 0.27 | 0.27 | 0.27 |
| $V_R$ Receiving end (Volt) | 0.08 | 0.08 | 0.08 | 0.08 |
| $V_Y$ Receiving end (Volt) | 0.09 | 0.09 | 0.08 | 0.08 |
| $V_B$ Receiving end (Volt) | 158.64 | 159.20 | 159.10 | 159.10 |
| $I_R$ Receiving end (Amp) | 0.00 | 0.00 | 0.00 | 0.00 |
| $I_Y$ Receiving end (Amp) | 0.00 | 0.00 | 0.00 | 0.00 |
| $I_B$ Receiving end (Amp) | 0.26 | 0.27 | 0.26 | 0.26 |
| Fault location R (km) | 62.12 | 60.9 | 62.4 | 62.4 |
| Fault location Y (km) | 62.02 | 61.2 | 62.2 | 62.2 |
| Fault location B (km) | 0 | 0 | 0 | 0 |
| Fault type | RYG Fault | RYG Fault | RYG Fault | RYG Fault |

**ML - Machine learning**

Table 3 provides a clear and concise comparison of analytical versus simulation and machine learning prediction-based results for locating faults on a double conductor-to-ground (RYG) fault, with specific reference to the use of the analytical method to derive the distance of the fault from the sending end and to derive fault impedance and location as determined by MATLAB simulations.

In the analytical method, the analytical method calculates the fault distance through the application of fundamental electrical principles in order to derive fault impedance through measurement techniques. Analytically, the voltage drop of the sending and receiving ends will be divided by the total current flowing through the phase to calculate the impedance. The cable impedance of 1 Ω/km means that the impedance calculated is the distance from the send point to the fault. The fault distance is calculated as 62.12 km from the sending point based on the R-phase impedance and, analogously, 62.02 km from the sending point based on the Y-phase impedance, confirming that both conductors. Based solely on voltage and current measurements, these analytical results serve as theoretical references.


Copyright © 2024 International Journal of Automation and Smart Technology

In column 3 of Table 3, the simulation outputs for the MATLAB Method are in very close accordance to the calculated distances made from the Analytical Method discussed previously. MATLAB estimates the fault distance to be between 61.0 and 61.5 km from the send point. It is reasonable to expect that the fault location will differ from the Analytical Method, as it includes elements of practicality, such as cable capacitance, numerical solver effects, load conditions, and sensor characteristics.

The data shown in column 4 provides another confirmation of the predictions made by the machine learning process. The Random Forest model that was built using an extensive list of simulated fault cases predicted that the RYG fault was likely to be located between 59.3 km and 62.4 km from the beginning point. This means the ML model remains very close to both the analytical and MATLAB outputs. The slight variation of around 1–2 km is normal for a data-driven model, as it recognizes patterns from training samples rather than applying exact electrical formulas. Despite this, the ML predictions strongly agree with the analytical and simulation-based results, confirming the reliability of the model for real-time cable fault detection and location.

*Table.4: Comparison of Analytical, Simulation, Machine Learning and Hardware Results for RYBG Fault*

| Parameters | Analytical Results | Simulation results | ML results | Hardware results |
|---|---|---|---|---|
| $V_R$ Sending end (Volt) | 179.29 | 179.29 | 178.80 | 178.80 |
| $V_Y$ Sending end (Volt) | 156.15 | 156.15 | 156.00 | 156.00 |
| $V_B$ Sending end (Volt) | 175.90 | 175.90 | 176.27 | 176.27 |
| $I_R$ Sending end (Amp) | 2.49 | 2.49 | 2.551 | 2.551 |
| $I_Y$ Sending end (Amp) | 2.17 | 2.17 | 2.22 | 2.22 |
| $I_B$ Sending end (Amp) | 2.45 | 2.45 | 2.52 | 2.52 |
| $V_R$ Receiving end (Volt) | 0.00 | 0.00 | 0.00 | 0.00 |
| $V_Y$ Receiving end (Volt) | 0.00 | 0.00 | 0.00 | 0.00 |
| $V_B$ Receiving end (Volt) | 0.00 | 0.00 | 0.00 | 0.00 |
| $I_R$ Receiving end (Amp) | 0.00 | 0.00 | 0.00 | 0.00 |
| $I_Y$ Receiving end (Amp) | 0.00 | 0.00 | 0.00 | 0.00 |
| $I_B$ Receiving end (Amp) | 0.00 | 0.00 | 0.00 | 0.00 |
| Fault location R (km) | 72.00 | 71.51 | 70.57 | 70.57 |
| Fault location Y (km) | 71.98 | 71.90 | 70.2 | 70.2 |
| Fault location B (km) | 71.79 | 71.10 | 69.92 | 69.92 |
| Fault type | RYBG Fault | RYBG Fault | RYBG Fault | RYBG Fault |

**ML - Machine learning**

Table 4 compare the analytical, MATLAB simulation, and machine-learning results for a triple conductor-to-ground (RYBG) fault. Because all three conductors are shorted to ground, their receiving-end voltages fall to 0 V, and the phase currents rise significantly, making the fault location estimation straightforward using Ohm's law.

From the analytical method, the R-phase voltage drops from 179.29 V at the sending end to 0 V at the receiving end. Dividing this voltage drop by the R-phase current of 2.49 A gives a fault impedance of exactly 72.00 Ω. Since the line impedance is 1 Ω per km, this impedance translates directly to a fault distance of 72.00 km. The Y-phase and B-phase values follow the same calculation process. Regarding the Y-phase and B-phase calculated fault impedances of approximately 71.98 Ω and 71.79 Ω, respectively, along with their corresponding fault locations of approximately 71.98 km and 71.79 km, we can infer that the locations of the faults must be nearly identical on the cable strand as a result of looking at all three conductor's fault locations along the cable strand.

The distance from the source to the fault location for all three phases, according to the simulation results (Column 3), can be estimated to be between 71.2 km and 71.9 km. All three phases' distances from the source match closely with calculated fault distances, with all three phases differing from the calculated fault distances by less than 1 km. Such small differences are expected, because MATLAB accounts for practical side effects of the cable, such as cable capacitance, numerical solver errors, and the characteristics of the sensors used. As a result of all the above factors, the simulation results are very close to the calculated values.

The predicted distances to the three-phase fault from the simulation results presented in Column 3, as well as the predicted distances from the Random Forest ML model, are all in close agreement with those calculated analytically. The Random Forest ML model predicts the three-phase fault to be between 69.92 km and 73.10 km from the beginning of the cable, which is only about 1 - 2 km away from the analytical and MATLAB results. This small difference is typical due to the fact that the ML model learns from training data the relationship between fault distance and fault distance, rather than using precise mathematical calculations to calculate the distances. Furthermore, ML


Copyright © 2024 International Journal of Automation and Smart Technology

predictions are also reliable and repeatable for such a severe three-phase fault.

Three different techniques have been developed to detect and locate a fault in the RYBG cable. The results of the tests performed using each method show that the location of the RYBG cable fault is almost identical for all three methods. The results of the three methods provide confidence that the developed system for detecting and locating cable faults works effectively, even with the potential for high-voltage underground cable faults.

The results of the comparisons of analytical results, MATLAB simulation outputs, and machine-learning classifications of the three types of faults (single-conductor, double-conductor, and three-conductor) show strong agreement. The small differences of 1–2 km in the locations predicted by the three techniques are negligible when compared to the length of the 80 km cable. The reason for the differences in predicted locations is due to how the numerical model was discretized, how data were scaled, and unknown noise factors that were present during the machine-learning training process. Overall, the Random Forest-based system achieves a high level of accuracy (94%) for both fault classification and distance estimation, which indicates that the combination of MATLAB and Python programming for fault detection and localization is a reliable and efficient means to detect and locate underground cable faults at scale and thereby assist in developing future smart-grid technologies.

## Conclusion

In this study, a machine-learning-based approach for fault detection and localization in underground power cables is presented using MATLAB Simulink and experimental hardware validation. Random Forest Classifier and Random Forest Regressor models are employed to identify fault types and estimate fault locations based on voltage and current measurements. Each fault condition produces a distinct electrical signature, which enables reliable fault classification and distance prediction. Under simulation conditions, the maximum deviation in fault location was within 2 km, which corresponds to an overall localization accuracy of approximately 95–96% for an 80 km cable.

The hardware setup, consisting of an ESP32 microcontroller, voltage and current sensing units, and a resistor-based cable model, was primarily used to validate the machine-learning predictions under real-time operating conditions. The close agreement between experimental and simulation results confirms the practical feasibility of the proposed approach. Unlike conventional IoT-based fault detection systems that mainly focus on data acquisition and threshold-based indication, the proposed method incorporates data preprocessing and filtering stages prior to prediction. This enables effective mitigation of noise and improves performance under complex fault conditions, such as conductor-to-ground and double conductor-to-ground faults. Owing to its data-driven nature and robustness to noisy inputs, the Random Forest–based framework provides a more reliable and accurate solution for underground cable fault diagnosis.

Combining MATLAB simulation with Python machine-learning techniques results in an effective solution for locating and identifying faults in underground cables rapidly, accurately, and at scale. The system outlined will provide increased uptime, greater reliability of fault detection, and has many advantages over existing traditional IoT-based methods, which makes the approach highly conducive for use in the modern smart grid applications.

## References


[1] V. Mane, P. Mansawale, S. Yawale, and A. Deshmukh, “Cable fault detection methods: A review,” International Journal of Research in Engineering, Science and Management (IJRESM), vol. 5, no. 6, Jul. 2022.

[2] Shateri, H.; Jamali, S. (2012). [IET 11th IET International Conference on Developments in Power Systems Protection (DPSP 2012) - Birmingham, UK (23-26 April 2012)] 11th IET International Conference on Developments in Power Systems Protection (DPSP 2012) - Impedance based fault location method for phase to phase and three phase faults in transmission systems.

[3] M. Dinesh, K. Vairaperumal, and P. Senthilkumar, “Design and detection of underground cable fault using Raspberry Pi & IoT system,” Int. J. Trend in Scientific Research and Development (IJTSRD), vol. 3, no. 1, Nov.–Dec. 2018.

[4] P. Alagumariappan, M. S. Y., A. S., and I. Fathima, “Identification of electrical faults in underground cables using machine learning algorithm,” Proceedings, vol. 42, no. 1, p. 20, 2020.

[5] M. Jayachandran, K. N. Abdul Rani and A. A. Wahab, “IoT-Based Smart Underground Cable Fault Detector,” 2024 IEEE 1st International Conference on Communication Engineering and Emerging Technologies (ICoCET), Kepala Batas, Penang, Malaysia, 2024, pp. 1 - 4.

[6] L. Kurinjimalar, B. Kamalakannan, A. S. Narayanan,



Copyright © 2024 International Journal of Automation and Smart Technology

P. K. Kumar, J. T. M. V. Saliyan, and M. Dharanish, “Advanced fault diagnosis in underground cable using IoT,” in 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT), 2024, pp. 1–4.

[7] Q. Shao, X. Zhang, M. Li, and Y. Liu, “Artificial intelligence in cable fault detection and localization: Recent advances and research challenges,” Energies, vol. 18, no. 14, p. 3662, Jul. 2025.

[8] O. Nyarko-Boateng, K. Amoako, and P. Senyo, “Predicting the actual location of faults in underground power cables using machine learning techniques,” Engineering Reports, vol. 3, no. 10, e12497, 2021.

[9] M. Najafzadeh, J. Pouladi, A. Daghigh, J. Beiza, and T. Abedinzade, “Fault detection, classification and localization along the power grid line using optimized machine learning algorithms,” International Journal of Computational Intelligence Systems, vol. 17, no. 1, 2024.

[10] P. Alagumariappan, M. S. Y., A. S., and I. Fathima, “Identification of electrical faults in underground cables using machine learning algorithm,” Proceedings, vol. 42, no. 1, p. 20, 2020.

[11] R. Mehta, S. Shukla, and P. Singh, “Random forest-based classification for transmission line fault detection,” Int. J. Electr. Power Energy Syst., vol. 143, p. 108484, 2022.

[12] M. S. Alam, A. Ghosh, and M. K. Das, “Machine learning based fault classification in distribution networks using phasor measurements,” IEEE Trans. Power Delivery, vol. 38, no. 1, pp. 329–338, 2023.

[13] N. Ghazali et al., “Twisted pair cable fault diagnosis via random forest machine learning,” Computers, Materials & Continua, vol. 71, no. 3, pp. 5427 - 5440, 2022.

[14] V. T. Rathod, P. Madhu, K. S. Prajwal, H. M. Varuna, and R. M. Patil, “Detection and location of faults in underground cable using MATLAB Simulink,” Int. Res. J. Eng. Technol. (IRJET), p. 2826.

[15] M. A. Kanawade, S. S. More and P. S. Phutane, "Optimization of Power Generation in Photovoltaic Systems Using Maximum Power Point Tracking with Fuzzy Logic Control," 2025 International Conference on Engineering Innovations and Technologies (ICoEIT), Bhopal, India, 2025, pp. 877-882.

[16] M. H. Saad and A. Said, “Machine learning-based fault diagnosis for research nuclear reactor medium voltage power cables in fraction Fourier domain,” Electrical Engineering, vol. 105, no. 1, pp. 25 - 42, 2023.

[17] S. Belagoune, N. Bali, A. Bakdi, B. Baadji, and K. Atif, “Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems,” Measurement, vol. 177, p. 109330, Jun. 2021, doi: 10.1016/j.measurement.2021.109330.

[18] A. Said, S. Hashima, M. M. Fouda, and M. H. Saad, “Deep learning-based fault classification and location for underground power cable of nuclear facilities,” IEEE Access, vol. 10, pp. 70735–70746, Jun. 2022, doi: 10.1109/ACCESS.2022.3187026.

[19] N. N. Bon and L. Van Dai, “Fault Identification, Classification, and Location on Transmission Lines using Combined Machine Learning Methods,” Int. J. Eng. Technol. Innov., vol. 12, no. 2, pp. 91 - 109, 2022.

[20] M. G. Basak, A. S. Chakraborty, and S. P. Roy, “Detection of Faults in Underground Cables Using Microcontroller,” 2019 IEEE 5th International Conference on Power Electronics Systems and Applications (PESA), Hong Kong.

[21] S. Khan, U. Ali, and S. Ahmed, “Arduino-based fault distance locator for subterranean cable detection,” in 2022 International Conference on Computing and Information Technology (ICCIT), 2022, pp. 1–6.

[22] L. Kurinjimalar, B. Kamalakannan, A. S. Narayanan, P. K. Kumar, J. T. M. V. Saliyan, and M. Dharanish, “Advanced fault diagnosis in underground cable using IoT,” in 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT), 2024, pp. 1–4.

[23] S. R. Purohit, S. M. Hattaraki, S. P. Hampangoudra, R. Nimbaragi, S. Mattihal, and S. Bagali, “Arduino-Uno based underground cable fault detection system (AUCFDS),” World Journal of Advanced Research and Reviews, vol. 15, no. 2, 2023.

[24] M. Dinesh, K. Vairaperumal, and P. Senthilkumar, “Design and detection of underground cable fault using Raspberry Pi & IoT system,” Int. J. Trend in Scientific Research and Development (IJTSRD), vol. 3, no. 1, Nov.–Dec. 2018.


Copyright © 2024 International Journal of Automation and Smart Technology