BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

AI-BASED UNDERGROUND CABLE FAULT DETECTION - Fault Classification and Location Using Machine Learning, Arduino and Online Alerts

[1]Tanvi Salunkhe, [2]Pooja Dehankar

[1,2]Department of Artificial Intelligence and Data Science
[1,2]Ajeenkya DY Patil School of Engineering, Pune, India
[1]salunkhe.tanvii06@gmail.com; [2]poojadehankar@dypic.in

Abstract

As soon as there occurs any fault in a buried power cable, it becomes impossible to determine the fault's position, and a significant amount of time gets wasted while trying to find out where it is located. This paper suggests an economical and compact method of solving this problem. The voltage and current values of the faulty cable at the sending end will be generated from the sequence networks by using equations that will be obtained for six states of cables: normal operation, line-to-ground fault, line-to-line fault, double line-to-ground fault, three-phase fault, and open circuit state. The classifier will decide about the faulty cable state, whereas the regressor will predict the fault distance. All these methods are implemented within a Flask service where all events are stored in the SQLite database and where the explanation of conclusions is also provided together with alert notifications by means of an n8n workflow, as well as by an Arduino UNO along with the LCD and relay modules as a local output. Random forest algorithm correctly classified all test events and detected faults with an average error of 0.65 km on a 5 km long line.

Keywords: Underground cables, fault categorization, fault localization, machine learning, random forest, Arduino, predictive maintenance

Introduction

It is worth noting that utilities have been burying their power feeders under the ground for some time now. An underground cable will not be affected by the weather, tree branches or birds and ensures that the streets remain clean. The problem is faced when there is a fault in the system. It should be known that an underground fault will not present any evidence of a fault on the surface, implying that fault identification will take some time since it will involve checking the voltage levels at different points along the cable.

Various techniques have been developed for estimating the fault location based on electrical parameters. The technique that used a distributed-parameter cable model to eliminate the error associated with a lumped impedance method was proposed by Yang et al. (2008). Ali et al. (2014) and Nag et al. (2020) proposed the technique that involved waveform matching and wavelet transformation for the fault location problem in the case of high impedance faults, while optimisation approach was used by Nag et al. (2020) to locate the fault. Subsequently, various learning-based approaches have been developed for solving the fault location

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

problem. These included the extreme learning machine-based fault location approach for cables proposed by Ray and Mishra (2015).

These papers pay much attention to the algorithm itself. They do not say much about the process following a prediction: how people are informed about it, where it is logged, and how a user can make an estimate on whether the decision is reliable enough. The research described in this paper creates a whole chain using cheap hardware and open source software, and evaluates the efficiency of the learning component in the presence of noise, load variation, and fault resistivity.

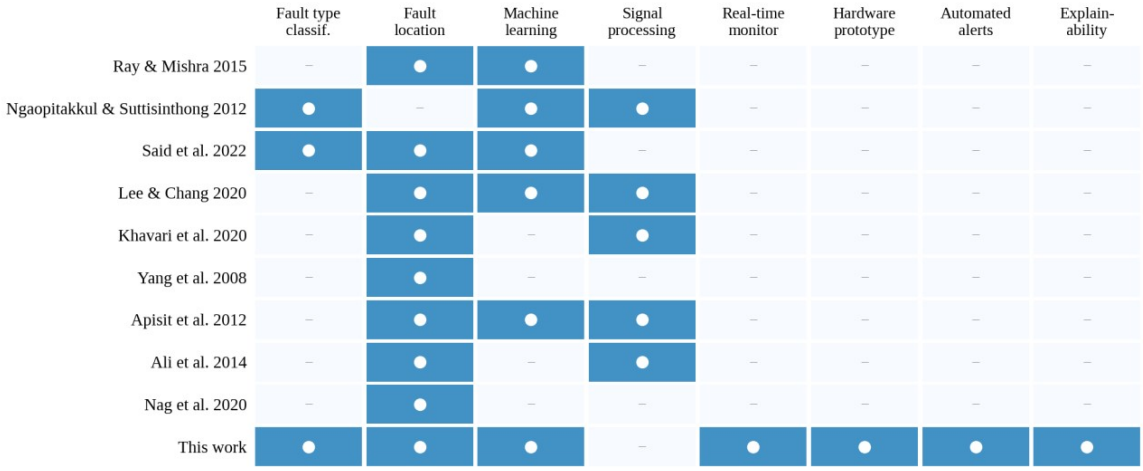


Figure 1: Tasks covered by each reviewed study. A filled cell marks a main element of the study, judged from its primary focus.

In some cases, a physical signature is first detected and then converted into a distance. In the approach suggested by Yang et al. (2008), there is no error connected with the lumped impedance of the long cable that appears in the modeling of the cable through the use of distributed parameters. By applying the matching and wavelet transforms approach by Ali et al. (2014), it is possible to receive a correct assessment of the fault location, even if it is rather poor due to high resistance. The optimization approach by Nag et al. (2020) tries to detect the nearest position to the waveforms received.

Some others use a learner to discover the mapping. In Ray and Mishra (2015), the ELM learns the location mapping. Wavelet-based fault type classification using a probabilistic neural network has been studied in Ngaopitakkul and Suttisinthong (2012), while in Apisit et al. (2012), there is a similar scheme for estimating location. Learning algorithms learn reflectometry measurements from multi-core instrumented cables in Lee and Chang (2020), and in Said et al. (2022), a deep network learns the classification and localization of nuclear power plants' cables. All the learning approaches presented above have been developed using conventional software frameworks, while the approach employed in this paper is based on Scikit-learn (Pedregosa et al., 2011).

Problem Statement

Fault diagnosis for underground cables takes time, thus prolonging downtime and increasing maintenance costs. The problem under study is to develop an automatic system that is capable of diagnosing the type of

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

fault, determining the distance to the fault and alerting the personnel using affordable hardware. The objectives of the study are as follows.

1.Gather the parameters used in fault detection and classification of underground cables.

2.Develop models for fault diagnosis and estimation of fault distance.

3.Integrate the models with Arduino hardware and web-based alarm system.

Provide measures for validation and indicate limitations of the evidence provided.

System Architecture

Figure 2 shows how the chain is wired up. In the prototype design, three switches correspond to faults along the cable sections at 1 km (red), 2 km (yellow), and 3 km (green). They can only have one switch at a time active. Three Arduino UNO digital inputs read the status from the A0 to A2 ports using the pull-ups that come with the chip. It also writes out the color and distance of the faulty cable onto an 16×2 I2C LCD and sends the signal for the corresponding relay channel using pins D8 to D11. Finally, the event travels over USB serial at 9600 baud rate to a Python code where it becomes JSON and gets posted to the /api/fault route of a Flask application.

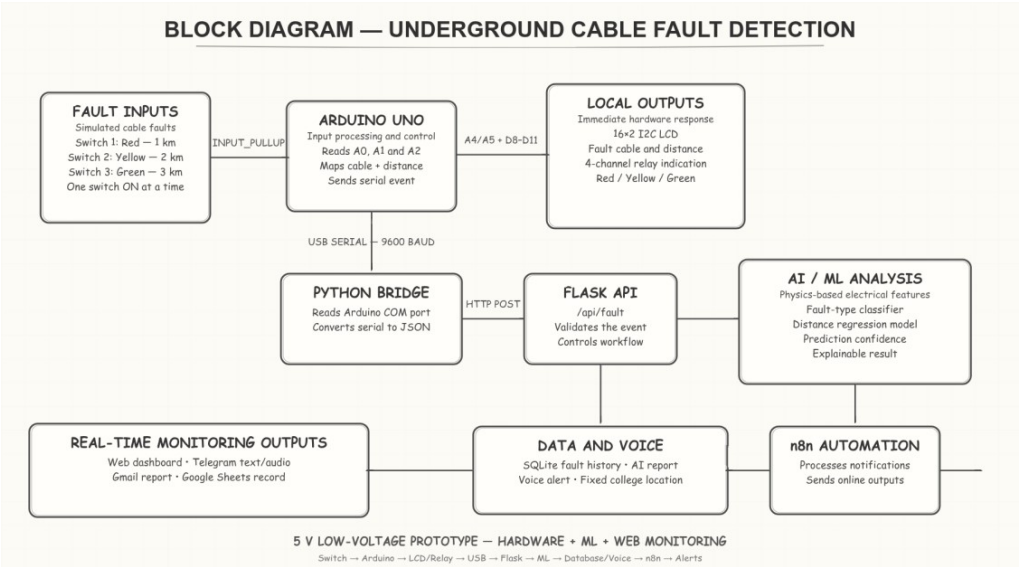


Figure 2: Block diagram of the proposed system, from simulated fault inputs to local outputs and online alerts (5 V prototype).

Flask evaluates the event and performs analysis on it. This analysis creates electrical feature vectors of the detected cable and distance, invokes the classifier and regressor, and provides confidence along with an explanation. The alert record, containing the report in text format and an audio message, gets saved in the SQLite database. An n8n flow sends the result to a web dashboard, Telegram, Gmail, and Google Sheets. Since there is no GPS component in the prototype system, the college location is hard-coded into every alert.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Methodology

Data Generation

Recorded faults for a cable could not be accessed; therefore, labeled data was generated for a radial three-phase feeder with source phase voltage E. The sequence impedance for a fault distance d (in kilometers) from the source is

$$Z1=Z2=Zs1+z1d, \quad Z0=Zs0+z0d \tag{1}$$

where Zs1, Zs0 represent the source impedance, while z1, z0 are cable impedances per kilometer. Using fault resistance Rf, symmetrical components determine the sequence currents for each type of fault as follows:

$$\text{LG: } I1=I2=I0=EZ1+Z2+Z0+3Rf \tag{2}$$

$$\text{LL: } I1=-I2=EZ1+Z2+Rf, \quad I0=0 \tag{3}$$

$$\text{LLG: } I1=EZ1+Z2\|Z0+3Rf, \tag{4}$$

$$\text{LLL: } I1=EZ1+Rf, \quad I2=I0=0 \tag{5}$$

The sequence components convert sequence currents into phase currents, adding them to a randomly generated balanced load current of random size and power factor. Sending end voltages follow by deducting source drop for each sequence network. In an open conductor, the phase with the fault carries only the charging current for the section still in service:

$$Iopen\approx j\omega CdE \tag{6}$$

These values are provided in Table 2. These are indicative values of an 11 kV distribution line and have been chosen arbitrarily for demonstration purposes.

Table 2: Simulation parameters

Parameter	Value
Line-to-line voltage	11 kV
z_1 (positive/negative sequence)	0.16 + j0.11 Ω/km
z_0 (zero sequence)	0.55 + j0.30 Ω/km
Z_{s1}, Z_{s0}	j0.35 Ω, j0.20 Ω
Cable capacitance	0.30 µF/km
Feeder length, fault distance	5 km, 0.2 to 5 km
Fault resistance R_f	0.5 to 60 Ω
Load current, power factor	15 to 100 A, 0.80 to 1.00
Measurement noise	5% Gaussian

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Parameter	Value
Events per class, total	1500, 9000

Features and Models

For each event, there are 22 features at the sending end: normalised values of voltage and current magnitude of all three phases, their respective cosines and sines of the angles between voltages and currents, the residual current |Ia + Ib + Ic|/3, and their respective impedances, resistance and reactances. Residual current indicates any earth-related problem, apparent impedance gives distance information, and angles discriminate resistive and inductive components. Gauss noise is added to all the features to model sensor errors.

The dataset was divided in 80/20 with stratification resulting in 7200 for training and 1800 for testing instances. There were six classification algorithms tested: decision tree, k-nearest neighbors, logistic regression, RBF SVM, gradient boosting and random forest. There were also six regression algorithms compared: linear regression, decision tree, k-NN, RBF SVM, gradient boosting and random forest. Distance models were trained only on faulted instances as there is no fault distance of a healthy cable. The number of trees in forests was 300. For classification, accuracy and macro F1 were used to evaluate performance.

For regression:

$$\text{MAE}=1Ni=1Ndi-di, \quad \text{RMSE}=1Ni=1Ndi-di2 \tag{7}$$

and the coefficient of determination R2. Fivefold cross-validation of the forest classifier proved independence of the test result from luck in splitting.

Explanation and Cable Health

For each prediction, there are the classifier's confidence level and the top contributing factors to this prediction, which are obtained using feature importance from the random forest. A separate health report considers both the current prediction and the past predictions made by the random forest model using SQLite.

Implementation in Hardware

Firstly, the circuit was tested using the Wokwi simulator and then constructed using the breadboard. The switches represent the faulty areas. The required features by the model are derived from the above mentioned simulator for the particular cable and distance, hence the prototype is a demonstration of the working of software and hardware together in a chain at 5V and not on an actual live cable. Table 3 indicates the rough cost of the components used.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Table 3: Approximate hardware cost

Item	Cost (Rs.)
Arduino UNO R3 compatible board	550
Four-channel 5 V relay module	650
16×2 I^2C LCD	190
Rocker switches (3)	105
Breadboard	60
Jumper wires	40
USB cable, hook-up wire, 5 V supply	180
Total	1775

Results

Classification of Faults

Each of the six classification algorithms attained at least 99.7% accuracy. Specifically, the random forest, decision tree, gradient boosting and logistic regression achieved 100%, while the k-NN and SVM obtained 99.78% and 99.94%, respectively. In five-fold cross validation for the random forest, the mean score was 99.99%, and the confusion matrix in Figure 3 shows no off-diagonal elements. This is as expected of an idealised radial machine whose six states generate highly distinguishable waveforms, so even simple classifiers can distinguish between them. This indicates that the attributes contain the relevant information. This result cannot be used to predict the accuracy in practice, given the presence of transients, incipient faults and sensor faults.

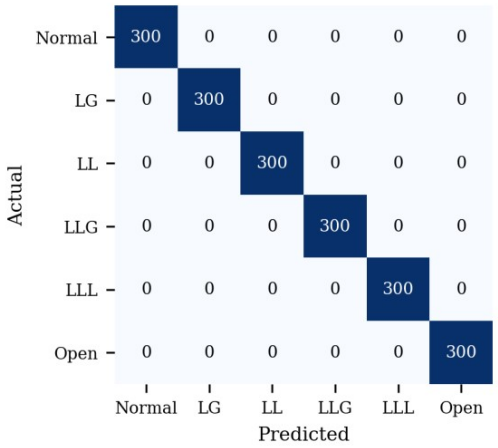


Figure 3: Confusion matrix of the random forest classifier on 1800 test events (300 per class).

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Distance Estimation

Distance estimation is more difficult. As illustrated in Table 4, the random forest is the best regressor with MAE = 0.65 km and R2 = 0.52 on 5 km feeder. This is followed by gradient boosting and the support vector regressor while the linear regression lags far behind, reinforcing the preference for non-linear algorithms. Figure 4 shows prediction vs actual distance graphically. The data points lie close to the ideal straight line

Table 4: Distance estimation on faulted test events

Model	MAE (km)	RMSE (km)	R^2
Linear regression	0.92	1.14	0.32
Decision tree	0.87	1.35	0.06
k-NN	0.85	1.14	0.32
SVR (RBF)	0.70	1.01	0.46
Gradient boosting	0.71	1.00	0.48
Random forest	0.65	0.96	0.52

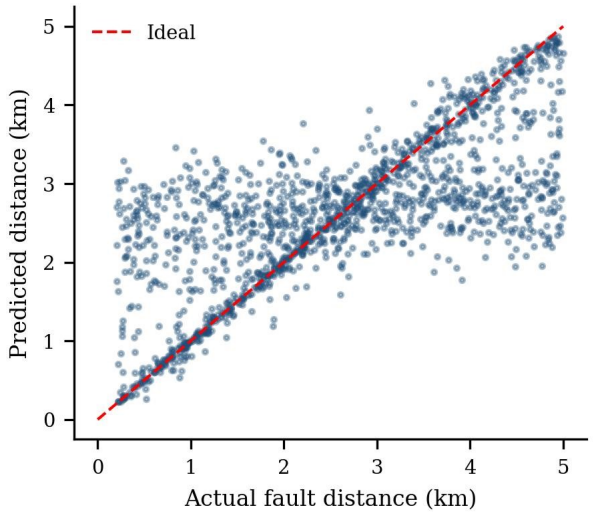


Figure 4: Predicted versus actual fault distance for the random forest regressor.\

Important Quantities

Phase current, apparent impedance, and residual current are the most important parameters for classification purposes because they reflect the magnitude of the short-circuit and earth engagement in terms of physics. When it comes to distance determination, reactance of the fault loop and phase angles are the two important quantities because reactance increases with increase in the length of the faulty cable.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

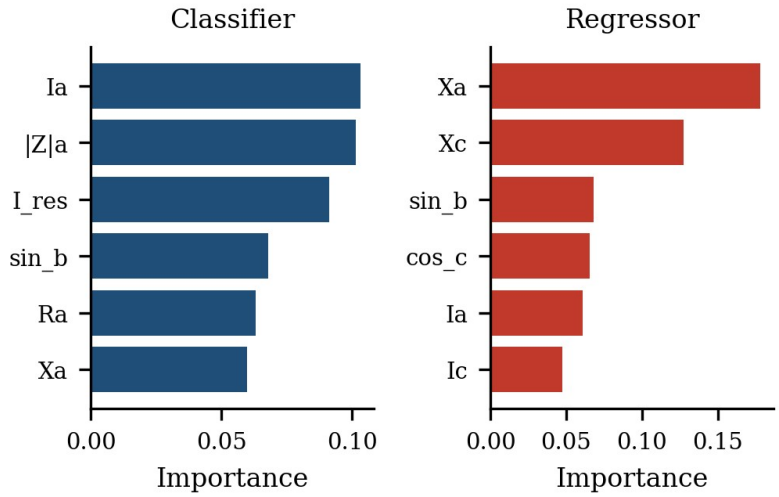


Figure 5: Top feature importances for the random forest classifier and regressor.

Robustness

The classifier was trained with 5% noise and tested with 20% noise maximum. The accuracy of the classifier did not fall below 99.8%. The location error increased from 0.60 km for 0% noise to 0.82 km for 20% noise. Since there is still error when there is no noise, we conclude that this error stems from the uncertainty of load and fault resistance, and not from sensors. The error further grows with fault resistance, from 0.42 km under 5 Ω to 0.90 km above 30 Ω due to the unknown resistive component in case of high fault resistance. In terms of classes, the open-circuit and double line-to-ground faults are located with an accuracy of 0.1 km, while LG, LL, and LLL faults are found with 1 km error.

The predicted means in terms of the three switch prototype settings are 1.78 km, 2.30 km, and 2.84 km with mean absolute errors of 0.81, 0.37, and 0.29 km respectively from a fault simulation of 1000 trials for the 1 km, 2 km, and 3 km switch positions. It maintains the correct sequence of the switch positions, although it is least accurate at positions closest to the origin where the apparent impedance value is low and susceptible to noise.

Discussion

Comparison with Traditional Utility Monitoring

Traditional utility monitoring uses programmable logic controllers, remote terminal units, and SCADA. Such a system has certified protection features, field sensors, and proven reliability; the present prototype does not intend to compete with it. However, it is different in terms of scope and cost. The fault type is determined by a learned classifier and not by predetermined alarm rules, the location is determined by a regression function with switch-distance cross verification, alerts are sent using Telegram, e-mail, and a logbook spreadsheet, and the cost of the hardware used is below Rs. 2000.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Limitations

The signals were taken from an idealised radial model, so getting 100% accuracy in classification is highly idealistic and does not reflect performance in the real world. The idealised model does not incorporate travelling waves, sheath currents, joints and branches, while the open circuit model utilises charging current approximation. The actual hardware tests only use three fixed distances. The lack of knowledge regarding fault resistance and load means that actual loop impedance is unknown and this is why there is error in locating faults near 1 km distance. Incipient faults are not considered.

Conclusion

In this paper, a machine learning pipeline has been developed for cable faults underground where it starts with the extraction of signal features through classification, localization, and explanation of an event, to finally end in local display, storage, and alerting. In simulations, random forest successfully differentiated between six cable states without any error and located faults with an average distance of 0.65 km on a 5 km long feeder line, where most of the remaining errors were found due to fault resistance and load. In future studies, the models will be tested on laboratory or field measurements of cable and traveling wave features will be added to reduce location error.

References

1. Bera, A., Tiwari, S., & Walia, N. K. (2026). Underground cable fault detection and geo-mapping system using real-time GPS integration. *SSRN preprint*.
2. Shao, Q., Fan, S., Zhang, Z., Liu, F., Fu, Z., Lv, P., & Mu, Z. (2025). Artificial intelligence in cable fault detection and localization: Recent advances and research challenges. *Energies, 18*(14), 3662.
3. Ali, H., Qamar, A., Alqahtani, N., Alenezi, A. H., & Ali, F. (2025). Magnetic field sensing for 2D fault localization in underground power cables. *PLOS ONE, 20*(10), e0334515.
4. Fu, H., Qiu, L., Ai, Y., Tu, J., & Yan, Y. (2025). Deep learning-based fault detection and location in underground power cables using resonance frequency analysis. *Electrical Engineering, 107*, 4051–4062.
5. Sah, R. K., Maharjan, R., Maharjan, S., Shrestha, N., Bhaskar, K., & Maharjan, B. S. (2025). AI-based fault detection, classification, and localization in power distribution systems. *2025 Fourth International Conference on Power, Control and Computing Technologies (ICPC2T)*.
6. Chen, H., Li, J., & Chen, P. (2024). Design of fault location algorithm based on online distributed travelling wave for HV power cable. *PLOS ONE, 19*(1), e0296513.
7. Vivek, B., Teja, B. H., Mallala, B., & Srinitha, G. (2024). Electrical fault detection and localization using machine learning. *2024 International Conference on Expert Clouds and Applications (ICOECA)*, 820–825.
8. Mishra, S., Swain, S. C., Biswal, T., & Routray, A. (2023). Tree based fault classification in underground cable. *Procedia Computer Science, 218*, 524–531.
9. Said, A., Hashima, S., Fouda, M. M., & Saad, M. H. (2022). Deep learning-based fault classification and location for underground power cable of nuclear facilities. *IEEE Access, 10*, 70126–70142.
10. Tariq, R., Alhamrouni, I., Rehman, A. U., Tag Eldin, E., Shafiq, M., Ghamry, N. A., & Hamam, H. (2022). An optimized solution for fault detection and location in underground cables based on traveling waves. *Energies, 15*(17), 6468.