Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 524–531

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

Tree Based Fault Classification in Underground Cable

Sanhita Mishra[a], Sarat Chandra Swain[a], Tapaswini Biswal[a], Aurobinda Routray[b]

[a]KIIT Deemed To be University, BBSR,751024, India

[b]IIT Kharagpur, Kharagpur,721302, India

Abstract

This paper presents tree-based machine learning techniques like decision trees, and random forest algorithms to categorize various permanent faults such as a line to ground, line to line, line-line to ground, and line-line-line to ground fault simulated in an underground cable. The simulation has been performed with the variation in fault types, fault location, fault resistance, and fault inception angle and the sending end current signals are presented as input to the classifier for classifying the fault. A relative comparison between the decision tree technique and random forest technique has been performed and the random forest algorithm turns out to be the best for classifying the permanent faults in an underground cable. Both the algorithms are deliberated by precision, F1 score, and recall. The best result is presented by Random Forest Algorithm with an accuracy of 98.8%.

Keywords: Decision Tree; Random Forest Algorithm; Underground cable; Fault Analysis

1. Introduction

The underground(UG) cable is essential in smart cities and provides numerous benefits over overhead transmission networks. Several challenges are faced by power engineers [1] due to the unusual design features of UG cable. Underground distribution network failures are considered to be one of the major reasons for power outages[2]. Hence, faulted segment identification is an essential survey, which not only increases the dependability of the underground distribution system but also minimizes the interruption time. Therefore, to manage these problems, smart methods are needed for fault identification and classification. Wavelet and ANIFIS based methods are presented in [3] for fault localization. Traveling wave techniques are highlighted in[4,5] for both fault identification and fault localization in the transmission network. But in a real scenario, it is challenging to apply the same since the sample rate requirements are considerably greater. Different conventional methods for fault detection in UG cables are mentioned in [6], such as the Murray and Varley Loop method. This method is prone to error due to approximations involved as well as an increase in temperature while testing the cable.

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: sanhita.mishrafel@kiit.ac.in

1877-0509
10.1016/j.procs.2023.01.034

A DWT-ANN based method has been used for detection and fault classification[7]. The detection method is used to segregate the faults mostly occur in the transmission system[7]. Apart, a review of the classification of various faults has been discussed in [8] where, many generic methods like the signal processing approach, and computational intelligence approach have been discussed. In addition, several hybrid techniques like the neuro-fuzzy technique, and wavelet-ANN techniques have also been presented. The techniques like DWT, and FFT helps to select the important attributes from the current signature. The researchers are emphasizing more on machine learning techniques for analyzing fault signals nowadays. There are various applications of machine learning techniques in multiple fields, a few of them are highlighted below. Z.Huang.et.al. have mentioned a novel Hierarchical Clustering Random Forest(HCRF) model to diagnose breast cancer[9]. For optimizing the features variable importance method(VIM) has been used. Both of these methods are tested with two data sets. These methods perform better with less computational time. Further, Ying.Yu.et.al. have proposed a hybrid fuzzy random forest algorithm[10] for classification. In addition, for tuning the parameter a harmony search algorithm is developed which works well even with an unbalanced dataset. Several machine learning techniques like Naive-Bayes, DT method, RF method, and SVM method are discussed in [11] to analyze EEG signals. These mentioned methods have certain advantages and disadvantages. Hence, to get rid of general ML techniques, a few hybrid learning techniques like reinforcement learning techniques are described, which work well in a complex environment. Furthermore, A.Esuli has presented a web-based interactive classification method[12] for text classification. This ICS adapts as well as updates the system and automatically classifies as desired. Apart, Rice leaf diseases are classified using a few classifiers such as the Naive Bayes technique, DT method, GB method, and RF algorithm, in [13] where, RF shows 69% accuracy. Further improvisation has been made with the RF algorithm mentioned in[14]. Here, A deep walk random forest (DWRF) Algorithm is proposed, where Deep Walk is used for feature extraction and the RF method is used for prediction of metabolite-disease association. Apart, there are several applications of ML techniques for fault analysis in transmission and distribution networks. In [15] some modern approaches are mentioned that perform better compared to the generic approaches. P.Ray.et.al. Have developed a support vector machine-based method to categorize and localize the faults generated in a long transmission line[16]. Further, a detailed SVM with certain improvisation is highlighted in [17] for future purposes. Nonetheless, for classifying various causes of fault in a transmission system, a decision tree-based method is implemented in [18]. But there is a loop fall of overfitting issues that mostly occurs in the decision tree. So *D.P.Mishra.et.al.* have projected [19] a random forest algorithm for the analysis of fault signals in a transmission network and the accuracy for classifying the fault is nearly 99.5%.

This paper depicts fault classification using tree-based ML techniques. By varying the types of fault(FT), fault -resistance(FR), inception angle of fault(FIA), and location of the fault(FL), the simulation data have been generated. For fault classification, Various ML techniques such as the DT method, and RF algorithm have been used. The classification report for both algorithms has been calculated and the evaluation shows Random Forest method classifies the faults with better accuracy.

Nomenclature

LG	Line to Ground
LL	Line to Line
LLG	Line Line Ground
LLLG	Line Line Line Ground
RF	Random Forest
DT	Decision Tree
AG	Ph. A - Ground
BG	Ph. B - Ground
CG	Ph. C - Ground
AB	Ph. A - Ph. B
AC	Ph. A -Ph. C
BC	Ph. B - Ph. C
ABG	Ph. A - Ph. B - Ground
BCG	Ph. B - Ph. C - Ground

ACG	Ph. A - Ph. C - Ground
ABCG	Ph. A - Ph. B - Ph. C - Ground
ML	Machine Learning
DT	Decision Tree
RF	Random Forest
SVM	Support Vector Machine

2. System Studied

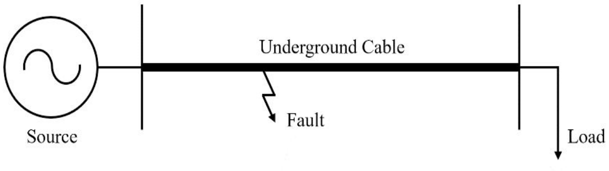

Fig 1. System Studied

A system with 11 kV voltage and 50 Hz frequency is implemented to study various faults[1] in an underground cable. Here Fig.1. is simulated with PSCAD/EMTDC, which is the most accurate and adopts a frequency-dependent model. A 95 mm2, XLPE cable with 5MW and 0.5 MVAR loads is used to simulate the model. The sample data has been generated with 10 different kinds of faults with one healthy condition. Several parameters have been varied as per the requirement for data generation. The data set for classification is represented in Table I. Implementing the above parameters, the data set for fault classification has been generated. The entire data set has been divided into 80% training and 20% testing. The testing samples have been utilized to validate the techniques implemented in this paper. Fig 2 represents the flow chart for fault classification implementing machine learning techniques.

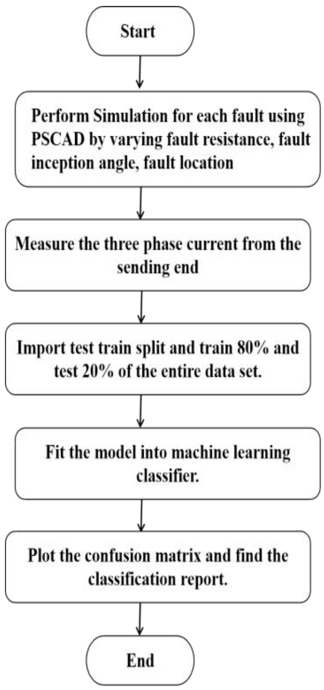

Fig 2: Flow chart for fault classification

Table I: Parameters for Classification

FIA(degrees)	0, 18, 36, 54,⋯..288, 306, 324, 342
FR (Ω)	0.1,1,5,10,15,20
FT	AG, BG, CG, ABG, BCG, ACG, AB, BC, AC, ABCG
FL (km)	2,4,6,8,10,12,14,16,18

3. Methodology

There are several machine learning techniques used for fault classification in underground cables. Here, only the RF algorithm, and DT algorithm are used and those are presented below.

3.1. Decision Tree

The Decision Tree(DT) is a perfect approach for collecting the data for building a classification system that is based upon multiple features. In this technique, a batch is arranged like a branch that majorly creates a tree that has three nodes. Those are root nodes, leaf nodes, and decision nodes. The major advantage of this technique is, this can handle a huge number of data. Constructing multiple decision rules, Decision trees are used to extract knowledge from an immense no.of accessible facts[22]. Fig 3. represents the basic diagram of a DT method.
The primary idea behind a decision tree is to discover the features that hold most of the information about the target feature and then it segregates the data set along with the values of these features to have pure target feature values that are feasible at the resulting nodes. The search for the most informative feature will continue till the leaf node is reached. To measure the informativeness of the features, it is highly essential to understand the entropy and information gain[23]. Entropy measures the randomness in a data set.

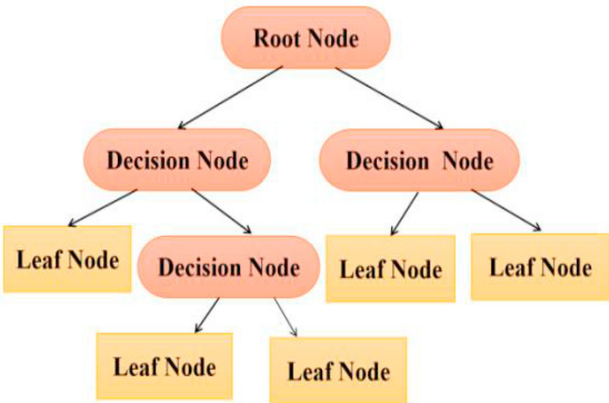

Fig 3: Decision Tree Algorithm

$$\text{Entropy} = -\sum (p(x = l) * log_2^{p(x=l)} \qquad (1)$$

Where $p(x = l)$ represents the probability that a target feature with a value is specified as l. The maximum value of entropy mostly depends upon the number of classes. Information gain is calculated, by calculating the entropy of the data set to determine the best features mostly served as root nodes. Information gain of a feature is mostly calculated as presented in equation (2).

$$\text{information gain(features)} = \text{entropy(dataset)} - \text{entropy(features)} \quad (2)$$

The feature with the largest information gain is used as the root node for initiating the decision tree. The Gini Index is a performance metric that is commonly used in the evaluation of classification systems, particularly those based on decision trees. Assuming the goal variable is a binary variable with values of 0 or 1, the Gini Index (G) is then computed as follows.

$$\mathrm{G} = 1 - \sum (p(x = l))^2 \quad (3)$$

3.2. Random Forest Algorithm

Ensemble learning is a machine learning prototype where multiple inputs are trained to solve a common problem and combined to get improved results. Bagging and boosting are two broadly used ensemble learners. Nonetheless, these two methods may be implemented with various numerical algorithms, the utmost dominant practice has been with DT. But, overfitting is a common problem that mostly occurs in the decision tree. If the maximum depth of the decision tree is not restricted with unlimited flexibility, it will prone to overfitting.

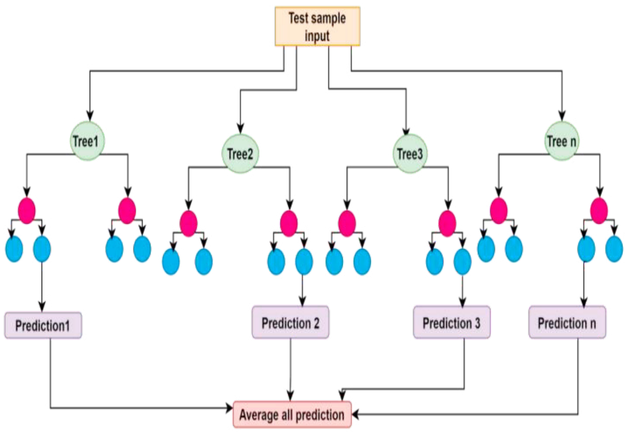

Fig 4: Random Forest Algorithm

The RF technique builds a “forest” consisting of several DTs and is trained using the Bagging method[20,21]. The algorithm executes in several steps, the first step being the selection of random samples from the data set. Then the selection of a decision tree for each sample takes place and finally prediction results from each decision tree are taken. The prediction results are now choices of decision trees for voting.

4. Results and Discussion

The data set has been created with the parameters given in Table I. For each method of classification, three parameters such as precision, recall, and F1 score [18] have been calculated which evaluate the efficacy of the model. The simulation has been performed for various. faults The diagonal elements of the confusion matrix depicted in fig 5 represent the accuracy of the Decision Tree method while classifying these faults. The overall classification accuracy of the DT method is 98.15%. We have a fairly good classification result but still cannot be used for the classification of faults, since the classification is not good for a few classes too.

$$\text{Precision} = \frac{\text{True Posotive}}{\text{true positive} + \text{false positive}} \tag{4}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True positive} + \text{false negative}} \tag{5}$$

$$\text{F}_1\text{Score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{6}$$

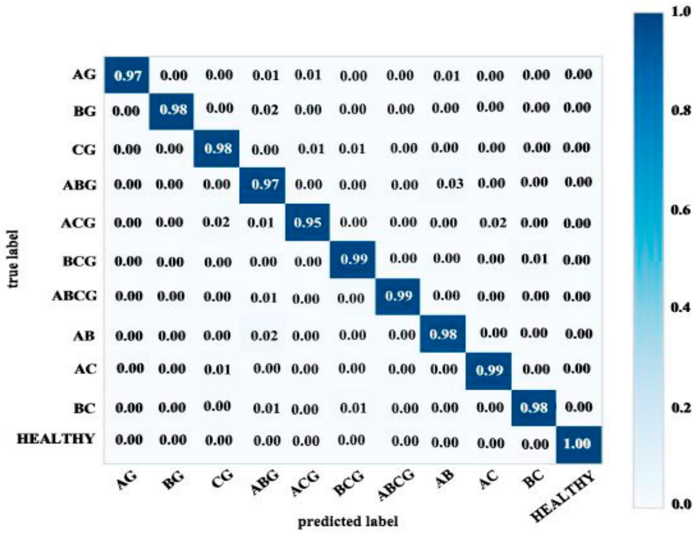

Fig 5: Confusion Matrix for Decision Tree Algorithm

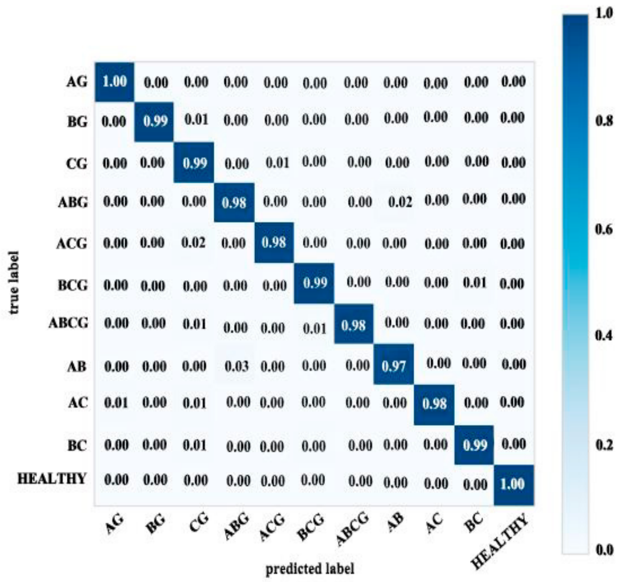

Fig 6: Confusion Matrix for Random Forest Algorithm

Table II represents the precision, recall, and F_1 score of various faults using the DT algorithm. Hence, for further improvisation in the overall accuracy with the same dataset, the Random Forest algorithm has been implemented. After the completion of the voting process, the prediction of having a majority vote becomes the final prediction. Fig 4. depicts the diagram of the RF algorithm.

Table II: Classification Report using Decision Tree Algorithm

Types of fault	Precision	Recall	F1
Healthy	1	1	1
BC	0.989	0.984	0.987
AC	0.979	0.995	0.987
AB	0.958	0.982	0.970
ABCG	0.993	0.986	0.990
BCG	0.979	0.995	0.987
ACG	0.979	0.953	0.966
ABG	0.943	0.973	0.958
CG	0.979	0.979	0.979
BG	1	0.977	0.988
AG	1	0.973	0.986

Table III: Classification Report using RF Algorithm

Fault	Precision	Recall	F1
Healthy	0.995	1	0.998
BC	0.989	0.989	0.989
AC	1	0.989	0.992
AB	0.981	0.970	0.975
ABCG	1	0.979	0.990
BCG	0.984	0.989	0.987
ACG	0.995	0.984	0.990
ABG	0.973	0.984	0.979
CG	0.959	0.990	0.974
BG	1	0.994	0.997
AG	0.995	0.973	0.997

The overall accuracy of the RF model is 98.8%. Table III represents the fault, precision, and recall of all the fault classes using the RF method. In Table III, the F1 score for AG, and BG faults is more in comparison to other faults. As the random forest is a combination of multiple Decision Trees, so it performs better in comparison to the DT method. The precision, recall, and F1 score are better compared to the DT method. Furthermore, in [24] a DWT-based and ANN-based method has been tested with a 735 kV,50Hz, 300 km, series compensated transmission system. Here, the classification accuracy for the ANN method is 65.43% and DWT based method is 97.43%. But in this paper, the proposed RF algorithm is tested with an 11kV,50Hz, 20km underground cable system, and the classification accuracy is quite fair compared to classical AI techniques like ANN.

5. Conclusion

This research represents the algorithm that has been implemented for fault classification. To increase the reliability of the UG cable several researchers have been encouraged to study several power systems challenges such as detecting, classifying, and localizing the faults in an underground cable as well as a hybrid system. During the occurrence of a fault, a switching phenomenon is observed. This happens because nonlinearities exist between fault types and the current signal. Hence, during this period, the structure of the entire equivalent circuit changes w.r.t time. Several ML-based fault classifiers such as DT and RF algorithms have been implemented for classifying the faults in the underground cable. It is observed that the RF algorithm has an accuracy of 98.8% and the DT algorithm has an accuracy of 98.15%. The suggested approach provides a rapid and accurate fault classification evaluation of the fault current. For improving accuracy, hybrid techniques and deep learning techniques can be implemented in the future.

References

[1] R. Swaminathan, S. Mishra, A. Routray, and S. C. Swain, "A CNN- LSTM-based fault classifier and locator for underground cables," *Neural Comput. Appl.*, 2021;33(22):15293–15304

[2] E. Gord, R. Dashti, M. Najafi, and H. R. Shaker, "Real fault section estimation in electrical distribution networks based on the fault frequency component analysis," *Energies*. 2019; 12(6):1145. https://doi.org/10.3390/en12061145

[3] Shimaa Barakat, Magdy B. Eteiba, Wael Ismael Wahba,Fault location in underground cables using ANFIS nets and discrete wavelet transform,Journal of Electrical Systems and Information Technology;2014,1(3):198-211, https://doi.org/10.1016/j.jesit.2014.12.003.

[4] X. Dong, W. Kong, and T. Cui, "Fault classification and faulted phase selection based on the initial current travelling wave," *IEEE Trans. Power Delivery*, 2009;24(2): 552–559

[5] M. N. Et. al., "Fault identification in power cables by using travelling wave-based approach," Turk. J. Comput. Math. Educ. (TURCOMAT),(2021);12(2): 855–865

[6] A. Nag, A. Yadav, A. Y. Abdelaziz and M. Pazoki, "Fault Location in Underground Cable System Using optimization Technique," 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, India;(2020):261-266, doi: 10.1109/ICPC2T48082.2020.9071462

[7] K. M. Silva, B. A. Souza, and N. S. D. Brito, "Fault detection and classification in transmission lines based on wavelet transform and ANN," *IEEE trans. power deliv.*,2006; 21(4): 2058–2063

[8] Avagaddi, Prasad & Edward, Belwin & Ravi, Kuppan. A Review on Fault Classification Methodologies in Power Transmission Systems: Part − I;2015, (5). 10.1016/j.jesit.2017.01.004.

[9] Z. Huang and D. Chen, "A Breast Cancer Diagnosis Method Based on VIM Feature Selection and Hierarchical Clustering Random Forest Algorithm," in IEEE Access, *vol10*, 3284–3293. https://doi.org/10.1109/access.2021.3139595

[10] Y. Yu, F. Yu, Y. He and W. Yang, "A Hybrid Fuzzy Random Forest Algorithm Using Harmonic Search Algorithm for Parameter Tuning," 2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), 2021, pp. 936-939,

[11] M. -P. Hosseini, A. Hosseini and K. Ahi(2021), "A Review on Machine Learning for EEG Signal Processing in Bioengineering," in *IEEE Reviews in Biomedical Engineering*,2021, vol. 14, pp. 204-218, doi: 10.1109/RBME.2020.2969915.

[12] A. Esuli,"ICS: Total Freedom in Manual Text Classification Supported by Unobtrusive Machine Learning," in IEEE Access,2022, vol. 10, 64741-64760, doi: 10.1109/ACCESS.2022.3184009.

[13] P. Mekha and N. Teeyasuksaet, "Image Classification of Rice Leaf Diseases Using Random Forest Algorithm," 2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering,: 165-169, doi: 10.1109/ECTIDAMTNCON51128.2021.9425696.

[14] J. Tie, X. Lei and Y. Pan, "Metabolite-disease association prediction algorithm combining DeepWalk and random forest," in Tsinghua Science and Technology;2022, 27(1); 58-67,doi: 10.26599/TST.2021.9010003.

[15] A. Prasad, J. Belwin Edward, and K. Ravi, "A review on fault classification methodologies in power transmission systems: Part-II,Journal of Electrical Systems and Information Technology;2018,5(1):61-67, https://doi.org/10.1016/j.jesit.2016.10.003.

[16] P. Ray and D. P. Mishra, "Support vector machine based fault classification and location of a long transmission line," *Eng. Sci. Technol. Int. J.*, 2016,19(3):1368–1380, .https://doi.org/10.1016/j.jestch.2016.04.001.

[17] M. Awad and R. Khanna, "Support Vector Machines for Classification," in *Efficient Learning Machines*, Berkeley, CA: Apress, 2015, pp. 39–66. https://doi.org/10.1007/978-1-4302-5990-9_3

[18] S. H. Asman, N. F. Ab Aziz, U. A. Ungku Amirulddin, and M. Z. A. Ab Kadir, "Decision tree method for fault causes classification based on RMS-DWT analysis in 275 kV transmission lines network," *Appl. Sci. (Basel)*;2021,11(9):4031, https://doi.org/10.3390/app11094031

[19] D. P. Mishra, P. Ray, and S. R. Arya, "Comparative analysis of classification and location of faults in a long transmission line using random forest technique," in *2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP)*, 2019, 1–5.

[20] M. Pal ,"Random forest classifier for remote sensing classification, International Journal of Remote Sensing";2005 ,26(1): 217-222. DOI: 10.1080/01431160412331269698

[21] J. Kaewmanee, T. Indrasindhu, T. Menaneatra and T. Tosukolvan, "Underground Cable Fault Location via Random Forest Algorithm," 2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia), 2019, pp. 270-273, doi: 10.1109/GTDAsia.2019.8715921.

[22] S. S. Gosavi, "Machine Learning Methods for Fault Classification," Institute of Computer Architecture and Computer Engineering, University of Stuttgart, 2014.https://d-nb.info/1055944281/34

[23] Njoku, Obinna Chilezie, "Decision Trees and Their Application for Classification and Regression Problems" (2019). MSU GraduateTheses.3406.https://bearworks.missouristate.edu/theses/3406

[24] P. Ray, D. P. Mishra, K. Dey and P. Mishra, "Fault Detection and Classification of a Transmission Line Using Discrete Wavelet Transform & Artificial Neural Network," 2017 International Conference on Information Technology (ICIT), 2017;178-183, doi: 10.1109/ICIT.2017.24.