Fault Detection and Classification in Transmission Line by using KNN and DT Technique

P. P. Wasnik[1], Dr. N. J. Phadkule[2], Dr. K. D. Thakur[3]

[1]PG Scholar, Government College of Engineering, Amravati, India.
[2]Assistant Professor, Government College of Engineering, Amravati, India.
[3]Associate Professor, Government College of Engineering, Amravati, India.

---***---

Abstract - *The electrical power system consists of many complex interconnected and interacting elements, which are always prone to disturbances or electrical faults. An important and more difficult task is the detection of faults and classification of their fundamental causes. The paper presents a semi-supervised machine learning approach based K Nearest Neighbor (KNN) and Decision Tree (DT) classifiers designed for fault detection and classification on transmission line. This work uses semi-supervised machine learning approach to process both the labeled and unlabeled power system data. The Discrete Wavelet Transform (DWT) is applied for feature extraction of fault current and voltage signal. The operation of KNN is based on the K^{th} nearest data point and the DT operation is based on the decisions for constructing optimal classification tree. The performance of the proposed system is examined on three phase series compensated transmission line network in MATLAB environment and the results of the two classifiers are compared based on the accuracy and time requirements, to obtain most suitable technique of fault analysis.*

***Key Words*: Fault detection, Fault classification, DWT, Machine Learning, KNN, DT**

1. INTRODUCTION

The power generation, transmission and distribution systems are the key enablers of the modern economy and sophisticated daily lives. A vital attribute of electrical power network is the continuity of service with a high level reliability. The lack of situational awareness on the network might be the primary cause of the sequence of events leading to abnormal operation or blackout. Power system protection is a focal point in the research area since the establishment of electricity. The detection as well as classification of fault is a crucial task encourages the advanced technologies over the year to analyse the power system events. The primary requirement of the present interconnected power system is reliable, sensitive and fast operating protection scheme in order to reduce the damage to transmission line due to disturbances. Determining the location of disturbance is also important in order to reduce the system outage time.

Many efforts have been made that defines various methods of protection of power system. The study [1] depends on the zero sequence and negative sequence values of the currents and voltages, to classify the single phase to ground and two phases to ground fault. The Multi-resolution analysis (MRA) is used in [2], to extract the current signal features and they are used for classification of fault. The location of fault and type of fault were determined in [3] using combination of the principle and sequence component analysis.

While expert knowledge based techniques are limited due to the unavailability of the real data thus, Machine Learning Algorithm are applied to deal with this, where knowledge is automatically extracted from data. The Machine Learning (ML) is the new advanced emerging technique for various applications in power system [4]. The ML algorithm can be used as problem solver for big data analysis, which provides the mean to handle the large power system data. It pays attention on the process of building programs which can learn from past experiences. The paper [5] introduces the application of ML for fault analysis in power system.

Supervised Machine Learning, one of the types of ML, has been widely used for fault classification in power system. In this approach, all the data must be labeled but large numbers of power system events are being recorded without label. Presence of unlabeled data would be problematic to the Supervised ML which may lead to fault misclassification. The ample amount of unlabeled power system data can be handled effectively with Semi-supervised ML approach. The previous work [7] addresses the application of semi-supervised ML based KNN for the fault identification-classification on the small transmission system. The amplitude (magnitude) of the fault current is the feature used for detecting the fault and also as the input to the classifier for classification. The results states that the classifier works well for small system with less attributes.

The present work focuses on the new technology where very small amount of data is labeled and is used to label the unlabeled data using appropriate classifiers. The Discrete wavelet transform (DWT) is implemented to extract the feature contained in three phase current and voltage signals. WT is used to unveil the hidden information in the fault current and fault voltage and the frequency components are derived which carries the fault information, as in case of [8]. The decomposed current signal and voltage signal are used as feature vectors for class prediction. The K Nearest Neighbor (KNN) and Decision Tree (DT) classifiers are used in this work to detect and classify the transmission line faults and are compared to obtain the most suitable classifier to detect and classify the transmission line faults.

1.1. Discrete Wavelet Transform

Representation of the signal into another form while keeping all the information contains in the signal is known as transformation of signal. Wavelet transform is systematic tool of signal processing to represent signal in time and frequency domain.

For any function, DWT can be calculated as,

$$DWT\,(m,n) = \frac{1}{\sqrt{2^m}} \sum_k f(k)\ \phi \left[\frac{n-k2^m}{2^m}\right] \quad (1)$$

Where,
f(k) is input signal,
$\phi \left[\frac{n-k2^m}{2^m}\right]$ is the mother wavelet,
n is the integer refers to the a particular sample number in an input signal.

The time scale representation of a discrete signal in DWT is procured using digital filtering (technique). The signal which is to be analyzed is send through series of (multiple) filters having different scales and frequencies to obtain low and high frequency content of the original signal. The DWT is derived by consecutive high pass (g) and low pass (h) filtering of discrete (time) signal, as shown in fig 1.

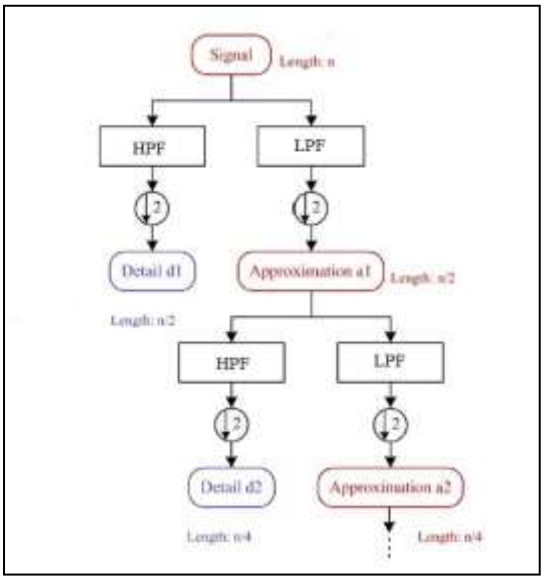


Fig-1: Decomposition Tree

In the first level of decomposition, signal is break down into D1 and A1, with the frequency band of (fs/2-fs/4) and (0-fs/4) respectively. In the second level of decomposition, the output of low pass filter from first level, A1 is break down into D2 and A2 with frequency band (fs/4-fs/8) and (0-fs/8) respectively [9], and the process continues, where, fs is the frequency of original signal. Similarly, features are extracted from the reconstructed detail coefficients in [10].

In the present work, the fault current and fault voltage measured at Bus 2 are decomposed into smaller and smaller components using four level DWT decomposition with 'db1' as mother wavelet. The output of the fourth decomposition level forms the feature vectors which are used as the input to both the classifiers.

1.2. K Nearest Neighbor

One of the most suitable methods of classification is k nearest neighbor algorithm as it a type of simple logical pattern-recognition method. Generally, it depends on the samples in the vicinity to predict the label. The KNN works on the simple assumption that, the things closer to each other belongs to the same group. In the problem of overlapping areas, such as in big data, KNN is most suitable. In the case of transmission line having many category restrictions, the KNN results may deviate as it only underlines the closest sample. KNN is one of the instance based lazy learning algorithms in which, the function is approximated regionally only and computations are put off till classification. The nearest points in the case of KNN are the entities that are previously classified. The method works as class classification as explained in [7], [11] and [12]. The output of the KNN is the class label to the new sample, which is predicted based on one or more nearest samples.

The Euclidean distance formula between any two points p and q,

$$d(p,q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} \quad (2)$$

$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \quad (3)$$

Where, i=1, 2, 3,……,n. are the number of samples.

The operation of KNN is based on the category of the most nearest one or a few samples to determine the classification of test samples. The KNN along with the DWT, for detection and classification of transmission line faults is discussed in [13]-[16] for different power system model but the idea behind the operation is same, which is Euclidean distance between the samples.

1.3. Decision Tree

The decision tree is the data classification process of checking the similarities in a dataset that classify them into different classes. Decision trees are widely used for the classification based on the choices of an attribute for data division. These attributes are split into several branches at nodes recursively, until the classification is reached, as shown in [17] and [18]. Decision tree is a non- parametric learning technique, which is able to create classifier for a given problem that can evaluate new, unseen circumstances and unveil the mechanisms driving the problem. The building process of DT depends on the learning dataset, start at the top most node of tree with entire learning dataset. The tree progresses by repetitively (iteratively) creating nodes by splitting the learning dataset into subsets with more classification purity. The process of splitting of data is stopped when the decision is made [19].

A decision tree graphically describes the decisions to be made, the events that may occur and the outcomes related with combinations of events and decisions. The Gini index is used to determine the node impurity as in [14]. The index defines better path in the tree to split the attributes. The Gini impurity can be obtained by summing the probability p_i of an sample with label i being chosen times the probability $\sum_{k\neq i} p_k = 1 - p_i$ of a mistake in categorizing that item. It reaches its minimum when all cases in the node fall into a single category.

$$G = \sum_{i=1}^{c} p_i * (1 - p_i) \quad (4)$$

The DT is used for the analysis of various problems such as power quality disturbance, faults on solar PV system, faults on double –circuit- transmission line are discussed in [20], [21] and [23].

2. TEST SYSTEM DESCRIPTION

A test system consist of six 3 phase, 350 MVA, 735kV, 60 Hz generators are delivering power to variable load through 300 km transmission line. The base quantities are listed in table. Generator side bus is named as bus B1 and that of load side is bus B2. The line circuit breakers at two buses B1 and B2 are CB1 and CB2 respectively. Currents and voltages are measured on bus B1 as well as bus B2. The test system is implemented in MATLAB environment [7]. Figure 2 shows the single line diagram of three phase test system used for fault detection and classification. The desired fault at desired location on the transmission line is created using fault breaker block. The fault situations are simulated in MATLAB environment and the results or fault data is stored in Workspace. The specifications of the transmission system model are listed in table 1.

The fault current and fault voltage measured at bus B2 is used for fault identification and classification of the line between B1 and B2. The data of all types of faults created on the test system is picked up from workspace for training the classifiers. The amplitude of the voltage and current signal changes with fault location and during each type of fault the change is different, thus sufficient to detect and classify the fault situation.

Table -1: Model Specification

Parameters	Value
Base Voltage	735 kV
Base Power	100 MW
Transmission Line	735 kV, 300 km, 60 HZ

3. FAULT DETECTION AND CLASSIFICATION

The KNN and the DT classifiers serve the role of classifiers to detect and classify the fault. Assume A, B and C be the three phases with G as ground. The table 2 shows the types of faults and the normal operation with their binary representation and their respective classes. The '1' means high represents the faulty line and '0' means low represents the absence of line in fault. Each fault class is represented by binary number like Phase AB-ground fault is represented as '1101' with the fault class as four.

Table-2: Fault types and classes

Fault type	Binary Value				Class
	A	B	C	G	
Phase A-Ground	1	0	0	1	1
Phase B-Ground	0	1	0	1	2
Phase C-Ground	0	0	1	1	3
Phase AB-Ground	1	1	0	1	4
Phase BC-Ground	0	1	1	1	5
Phase AC-Ground	1	0	1	1	6
Phase AB	1	1	0	0	7
Phase BC	0	1	1	0	8
Phase AC	1	0	1	0	9
Phase ABC	1	1	1	0	10
Phase ABC-Ground	1	1	1	1	11
Normal Operation	0	0	0	0	12

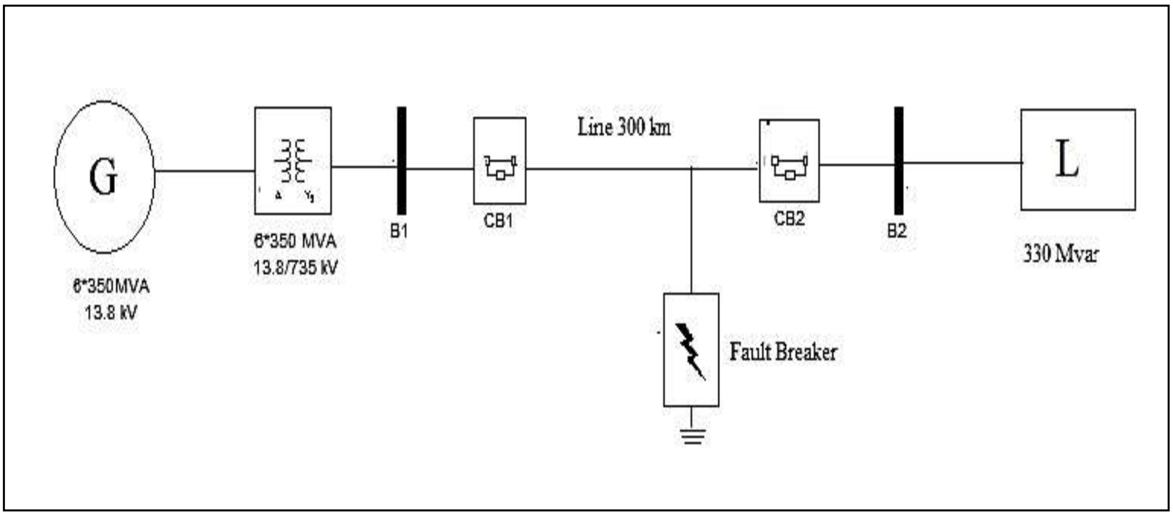


Fig.-2: Three Phase Transmission Line Network for fault detection and classification

The detection of the fault is the process of discriminating between the healthy operation or fault situation. The classification of fault is the process of determining the type or class to which the fault belongs. After detection for the abnormal condition, the fault is categorized into one of the 11 classes using the classifiers.

4. RESULT AND DISCUSSION

The total 24 cases are considered in the work, out of which two cases are of normal operation and others are being the faults of different nature occurred at same or different location created using the fault breaker block in MATLAB. The table 3 shows the set of faults generated at various parameters and conditions. The time required for training KNN classifier is lesser than DT classifier, while DT classifier is able to test the fault case in lesser time as compare to KNN classifier, shown in table 4 and table 5. It is observed that the algorithms proposed in the work are capable of detecting and classifying the faults successfully and independently for all the cases and the classification accuracy can be evaluated by the confusion matrix shown in fig.3. The correctly classified cases are shown by green blocks; incorrectly classified cases are shown by red blocks and the blue box showing the classifier accuracy. Thus taking into account the above factors, the results shows that the DT outperforms the KNN operation with lesser testing time offering 100% accuracy and confirming the effectiveness of the proposed work in the real-time implementation.

Table-3: Experimental conditions

Fault type - ABCG, ABC, AB, BC, AC, AG, BG, CG, ABG, BCG and ACG
Fault Locations- 280 and 300 Km
Fault Resistance- 0.001 and 0.1 Ω

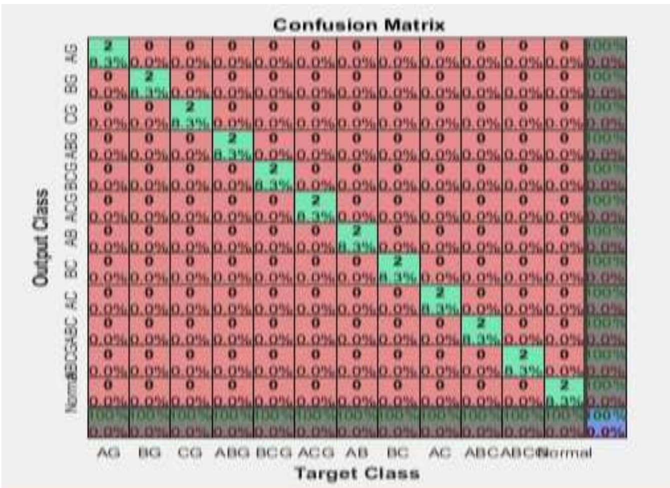
Table -4: Average Training Time

Classifiers	Time (Second)
KNN	6.90
DT	7.53

Table-5: Average Testing Time (second)

Fault	KNN	DT
Phase A to Ground	7.25	5.43
Phase A-B to Ground	4.85	4.57
Phase A to B	5.23	4.49

Fig-3: Confusion Matrix



The comparison of different ML techniques for the applications in power system, including KNN and DT is carried out in [23], [24] and [25], shows that different techniques have their own merits and limitations and hence different techniques are suitable for different applications.

5. CONCLUSION

The paper presents the semi-supervised machine learning approach for detection and classification of faults on transmission line. The aim is to compare the two machine learning algorithms, KNN and DT, to obtain the most suitable technique of fault analysis. The present approach is tested on Three Phase Series-Compensated Network model in MATLAB environment for all 11 types of faults. The DWT is used to extract the fault current and fault current features. The four-level decomposition in DWT is employed in order to obtain the feature vectors. The KNN and DT algorithms are used as classifiers to discriminate between the normal and fault situations and also to categorize these faults according to their nature. It is observed from the results that DT outperforms the operation of KNN with lesser testing time offering same classification accuracy. It is noticed that once the classifiers are trained, it is possible to test any fault case that lies within the classifier operation criteria.

REFERENCES

[1] A. Rahmati and R. Adhami, "A fault detection and classification technique based on sequential components," IEEE Trans. Ind. Appl., vol. 50, DOI 10.1109/TIA.2014.2313652, no. 6, pp. 4202-4209, Dec. 2014.

[2] M.J.B. Reddy, D.V. Rajesh and D.K. Mohanta, “Robust transmission line fault classification using wavelet multi-resolution analysis,” Computers & Electrical Engineering, vol. 39, DOI 10.1016/j.compeleceng.2013.02.013, no. 4, pp. 1219-1247, May 2013.

[3] Q. Alsafasfeh, I. Abdel-Qader and A. Harb, "Fault classification and Localization in Power Systems Using Fault Signatures and Principal Components Analysis", Energy and Power Engineering, vol. 4, DOI 10.4236/epe. 2012.46064, no. 6, pp. 506-522, Nov. 2012.

[4] Seyed Mahdi Miraftabzadeh, Federica Foiadelli, Michela Longo, Marco Pasetti, “A Survey of Machine Learning Applications for Power System Analytics”, IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), DOI: 10.1109/EEEIC.2019.8783340, 2019.

[5] Halil Alper Tokel, Rana Al Halaseh, Gholamreza Alirezaei, and Rudolf Mathar, “A New Approach for Machine Learning-Based Fault Detection and Classification in Power Systems”, IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), DOI: 10.1109/ ISGT.2018. 8403343, 2018.

[6] Maitreyee Dey, Soumya Prakash Rana, Sandra Dudley, “Semi-Supervised Learning Techniques for Automated Fault Detection and Diagnosis of HVAC Systems”, 30th International Conference on Tools with Artificial Intelligence (ICTAI), DOI: 10.1109/ICTAI.2018.00136, August 2018.

[7] Prerana P. Wasnik, Dr. N. J. Phadkule, Dr. K. D. Thakur, “Fault Detection And Classification Based On Semisupervised Machine Learning Using KNN”, in Proc, International Conference on Innovative Trends and Advances in Engineering and Technology (ICITAET), 2019.

[8] Suman Devi, Nagendra K. Swarnkar, Sheesh Ram Ola and Om Prakash Mahela, “Detection of Transmission Line Faults Using Discrete Wavelet Transform”, Conference on Advances in Signal Processing (CASP), DOI: 10.1109/ CASP. 2016. 7746152 , 2016.

[9] M. Gowrishankar, P. Nagaveni and P. Balakrishnan, “Transmission Line Fault Detection and Classification Using Discrete Wavelet Transform and Artificial Neural Network”, Middle-East Journal of Scientific Research, DOI: 10.5829/ idosi.mejsr.2016. 24.04.23063, 2016.

[10] Papia Ray, B. K. Panigrahi and N. Senroy, “Extreme Learning Machine based Fault Classification in a Series Compensated Transmission Line”, IEEE International Conference on Power Electronics, Drives and Energy Systems, DOI: 10. 1109/ PEDES. 2012. 6484297, 2012.

[11] Aida Asadi Majd, Haidar Samet and Teymoor Ghanbari, “k-NN based fault detection and classification methods for power transmission systems”, Protection and Control of Modern Power Systems, DOI 10.1186/s41601-017-0063-z, 2017.

[12] Yue Zhang, Jianxia Chen, Qin Fang, Zhiwei Ye, “Fault Analysis and Prediction of Transmission Line Based on Fuzzy K-Nearest Neighbor Algorithm”, 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), DOI: 10.1109/FSKD.2016.7603296, Aug. 2016.

[13] Anamika Yadav, Aleena Swetapadma “Fault analysis in three phase transmission lines using k-nearest neighbor algorithm”, in Proc. Advances in Electronics, Computer and Communication (ICAECC),DOI10.1109/ ICAECC. 2014. 7002474, pp. 1-5, oct . 2014.

[14] Tamer S. Abdelgayed, Student Member, Walid G. Morsi and Tarlochan S. Sidhu, Fellow Member, IEEE“ Fault detection and classification based on co-training of semi-supervised machine learning”, IEEE Trans. on Industrial, DOI 10.1109/THE 2017.2726961.

[15] Bruno. G. Costa, Jean C. Arouche Freire, Hamilton S. Cavalcante, Marcia Homci, Adriana R. G. Castro, Raimundo Viegas, Bianchi S. Meiguins and Je_erson M. Morais , "Fault Classi_cation on Transmission Lines using KNN-DTW", Conference Paper in Computer Science, DOI: 10.1007/978-3-319-62392-4_13, 2017.

[16] M.Pavani, S.Raghunath Sagar, G.Ganesh, "Fault Classification in Transmission Lines Using Wavelet Transform and K-Nearest Neighbors", International Journal of Education and Applied Research (IJEAR), Vol. 9, Dec 2019.

[17] Subhra Jana and Abhinandan De, "Transmission Line Fault Pattern Recognition using Decision Tree based Smart Fault Classifier in a Large Power Network",IEEE Calcutta Conference (CALCON), DOI 978-1-5386-3745-6/17, 2017.

[18] Chengxi Liua, Zakir Hussain Rathera,b, Zhe Chena, Claus Leth Bak, " An overview of decision tree applied to power systems", International Journal of Smart Grid and Clean Energy, vol. 2, DOI: 10.12720/sgce.2.3.413-419, 2013.

[19] S. M. Shahrtash, A. Jamehbozorg, "A Decision Tree Based Method for Fault Classification in Transmission Lines", 2008 IEEE/PES Transmission and Distribution Conference and Exposition, DOI: 10.1109/TDC.2008.4517258, April 2008.

[20] Prakash K. Ray, Soumya R. Mohanty, Nand Kishor, and João P. S. Catalão, "Optimal Feature and Decision Tree-Based Classification of Power Quality Disturbances in Distributed Generation Systems", Ieee Transactions On Sustainable Energy, DOI 10.1109/TSTE. 2013.2278865, 2013.

[21] A. Jamehbozorg and S. M. Shahrtash ,"A Decision Tree-Based Method for Fault Classification in Double-Circuit Transmission Lines", IEEE Transactions on Power Delivery, Volume 25, DOI: 10.1109/ TPWRD.2010. 2050911, Oct. 2010.

[22] Ye Zhao, Ling Yang, Brad Lehman , Jean-François de Palma, Jerry Mosesian, Robert Lyons, "Decision Tree-Based Fault Detection and Classification in Solar Photovoltaic Arrays", Twenty-Seventh Annual IEEE Applied Power Electronics Conference and Exposition (APEC), DOI- 10. 1109/APEC.2012. 6165803, Feb. 2012.

[23] Comp knn and dt- Sayali D. Jadhav1, H. P. Channe, " Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques", International Journal of Science and Research (IJSR), Paper ID: NOV153131 ,Volume 5 Issue 1, January 2016.

[24] Knn and dtt- Amr E. Mohamed, "Comparative Study of Four Supervised Machine Learning Techniques for Classification", International Journal of Applied Science and Technology, Vol. 7, No. 2, June 2017.

[25] Coexistence- Caroline El Fiorenza J, Deepeli Sikerwar, Jayasyam Reddy, Nitisha k, Muskan Joshi, "Co-Exercise of DT And KNN Classifiers for Fault Detection and Classification of Semi Supervised Machine Learning", International Conference on Physics and Photonics Processes in Nano Sciences, doi:10.1088/1742- 6596/ 1362/1/012058, 2019.