International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Transmission Line Fault Detection by Using Machine Learning Algorithms

Koyi Kotaiah Chowdary[1], K. Rushitha[2], K. Rajev Venkat[3], A. Vijaya Lakshmi[4], G. Vijay Naga Prasad[5], G. Harsha[6]

Submitted:22/10/2024 **Revised:**03/12/2024 **Accepted:**10/12/2024

Abstract : we cannot image the world without electricity; it is playing a major role in the daily life of the human. The world development is completely depending on electricity. Such electrical network facing the several issues of electrical faults. So detecting faults and rectifying is crucial task in power system of electrical network. In this study machine learning algorithms are proposed to detect the fault in the transmission lines. I used a MATLAB Simulink transmission line model to develop a data set which contain electrical source and electrical lads with a single 3-phase transmission line of 25km it replaces the performance of real transmission line. The data set can be used to train and test the proposed machine algorithms (Linear Regression, Support Vector Classifier, Decision Tree, K-Nearest Neighbor) among all these the K-Nearest Neighbor (KNN) algorithms shows well accuracy and good performance.

***Key words**: Transmission lines, Fault detection, MATLAB-Simulink, Machine learning algorithm.*

1. Introduction:

An electrical network is the huge system which contains generation, transmission and distribution systems. The distribution system is a long network used to interconnect the generation and distribution system. It contains the transmission lines of hundreds and thousands of kilometers to carry the electricity from one place to other place. But these transmission lines faces the fault issues due to the lightning strikes, weather conditions[1], tree or branch contact, animal interference, equipment failure, mechanical stress, vibration, human error, faulty insulation etc…. the transmission line faults creates a serious impact on electrical network like power shortage, damage to equipment, voltage instability, reduced system reliability, increased maintained cost, safety hazards, power quality issues, grid instability and cascading failures, increased load on other lines, financial loss[5]. Addressing these issues typically involve monitoring, fault detection systems and effective maintained practices. So fault detection and clearing is the one of the major task in transmission lines maintenance. The machine learning algorithms are very help full to detect the transmission line faults quickly [6].

1ORC ID: 0009-0004-3865-8932
Sir C R Reddy College of Engineering-Eluru Andhra Pradesh, India
mail id: kkotaiahchowdary@gmail.com

Machine learning is the latest technology, it train the models without explicit programming. Now days the machine learning and deep learning techniques are used in all the applications of industrial, commercial, business analytics, bio medical, agriculture sectors. So the electrical engineering sectors of power generation, transmission and distributions are no exceptions to adopt latest technology[4]. The machine learning models are high accuracy and very quick response as compared to the traditional fault detection models in transmission lines. The ocean of machine learning subject can contain various learning techniques such as supervise learning, unsupervised learning, semi supervised learning, reinforcement learning etc… again each learning contains number of algorithms. Among all these I only prefer supervised learning techniques because we are tainting the transmission line fault detection model with the labeled data collected form MATLAB similink model. In this paper I only concentrated on four supervised machine learning algorithms which are logistic regression, support vector classifier, decision tree and K-Nearest neighbor algorithms [7,8], because they are well designed for classification problems.

2. Electrical faults and classification:

A transmission line fault is an abnormal condition that disturbs the normal flow of electricity in the lines. These transmission line faults are mainly categories in to two ways symmetrical and unsymmetrical faults. The symmetrical faults affect the all faces in the transmission line equally and simultaneously. 2-5% of transmission line faults are symmetrical faults. The unsymmetrical faults are most common faults and they effect unequally in the all phases of transmission line. 75 -80 % faults are unsymmetrical faults in the transmission system[10].

Table 1: Fault Classification

S.no	Fault classification	Fault name	% of appearance
1	Symmetric fault	RYB	5-10 %
		RYBG	
2	Unsymmetrical fault	RG	75-80%
		YG	
		BG	
		RY	
		YB	
		BR	
		RYG	
		YBG	
		BRG	

*RYB-Triple line fault

*RYBG-triple line to ground fault

*RG-Line to ground fault

*YG-Line to ground fault

*BG-Line to ground fault

*RY-Double line fault

*YB-Double line fault

*BR-Double line fault

*RYG-Double to ground fault

*YBG-Double to ground fault

*BRG-Double to ground fault

Table 1 shows the classification various faults under symmetrical and unsymmetrical faults. The RYB & RYBG are comes in to the category of symmetrical faults they develop symmetrical components of fault voltages and currents parameters in to the all three phases of the transmission line uniformly. RG, YG, BG, RY, YB, BR, RYG, YBG and BRG are unsymmetrical faults which regularly appear in the system they develop non uniform components if fault voltages and currents in the transmission line but they may not affect the all three phases uniformly, some phases may effect and some phases may not effect which may depend on the type of the fault.

3. **Proposed methodology**:

Here we are proposing machine learning models to detect the faults in the transmission line.

The block diagram in figure 1 shows the working of the proposed methodology which contains different blocks like MATLAB-Simulik model block, data set, splitting data set, normalization, machine leaning model, performance evaluation. The matlab simulink block contain the transition line model which works as a real transmission line system which is help full to develop a data set[17].

The extracted data variables under different fault conditions are sent to a data generation block to cumulate the data and form a required data set. The data set con spilt into training data and testing data which can be help full to train and test the machine learning models like linear regression, SVC, Decision tress, KNN algorithms.

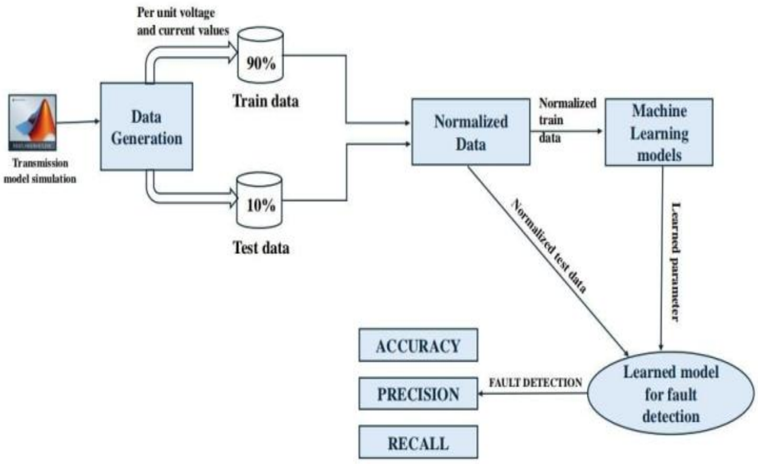

Fig 1: block diagram of fault detection in transmission line by using machine learning algorithms

3.1 MATLAB-Simulink model block: a transmission line model is designed in a matlab simulink software is as shown in figure 2 which replaces a working of the original transmission line. Matlab simulnk software is well popular to develop a real world electrical and electronics circuits due to the availability of all components and devices. The MATLAB Simulink model designed with a three phase generation unit, transmission line of 25km length, load center and a three phase fault system. The three phase fault system block can develop the required faults at desired instead of times in the transmission line to analyze the fault parameters.

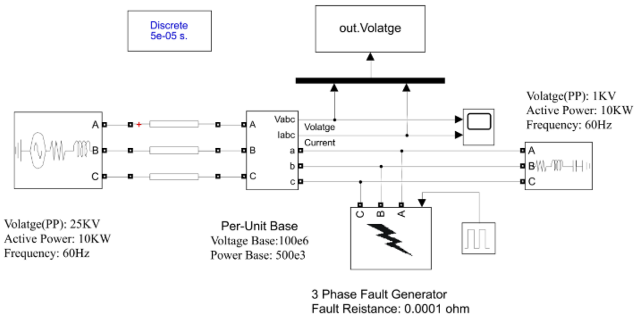

Fig 2: MATLAB Simulink model for transmission line to get data set

3.2 **Data set**: a data set can be developed by collecting the voltage and current parameters under different faults conditions like: LG, LL, LLG, LLL, LLLG, no fault in the matlab simulink transmission line model[3]. In this model we developed 39320 insights of data set with va, va, vb, vc, Ia, Ib, Ic variables under fault and no fault conditions in a CSV format.

3.3 Splitting data set: split the data set into training data and testing data. The training data is help full to rain the machine learning model, and the testing data will be help full to check the performance of the model. Here we dividing the data set into 90:10 which means 90% data 35388 is used to train the machine learning models, 10% data 3932 is used to test the performance of the model[14].

3.4 Normalization: after splitting data is send to normalization block. Normalization is a process of converting abnormal data insights in the data set into a normal form because they may create the errors in training and testing of machine learning models. Normalization is also called scaling to normalize over range insights in to normal rage in data cleaning process[22,23].

3.5 Machine learning models: in this proposed methodology we are using linear regression, SVC, DT and KNN algorithms to train the model and analyze the performance parameters of fault detecting task.

3.6 Preference Evaluation: the performance of the machine learning models can be evaluated with different performances performance metrics, including accuracy, precision, recall, F1 score, and AUC scores [18-21], as outlined below.

Accuracy: Accuracy is a metric used to evaluate the model's performance. The ratio of the total number of right instances to the total number of instances is provided [18].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: refers to the degree of accuracy exhibited by a model's positive predictions. The term "accuracy" refers to the proportion of accurate positive predictions by the model in relation to the overall number of positive predictions [19].

$$Precision = \frac{TP + TN}{TP + FP}$$

Recall: The concept of recall pertains to the evaluation of a classification model's ability to accurately identify all pertinent instances within a given dataset. The measure in question is the proportion of cases that are classified as true positive (TP) to the combined total of instances classified as true positive and false negative (FN) [20]. It is the ratio of true positive predictions (TP) to overall positive predictions ($TP + FP$)

$$Recall = \frac{TP}{TP + FP}$$

The F1 score: It is a metric employed to assess the overall efficacy of a classification module. The harmonic mean constitutes the average of precision and recall [21].

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

AUC Score: The AUC score is a metric that represents the Area under the Curve, which is derived from the ROC curve. The binary classification model's overall performance is assessed. Given that both TPR and FPR fall within the range of 0 to 1, it follows that the area will consistently be confined within this range. A higher value of AUC indicates superior performance of the model. The primary objective is to optimize the utilization of this region with the aim of achieving the maximum TPR and lowest FPR within the specified threshold. The Area Under the Curve (AUC) quantifies the likelihood that the model will assign a positive instance, selected at random, a greater predicted probability in comparison to a negative instance, also selected at random [22,23].

4. Results and Discussion:

In this work we trained the machine learning models logistic regression, Support vector classifiers, decision trees and K-Nearest neighbor algorithms with the proposed data set. And they shown their respective performances in the form of confusion matrix and ROC cures from this we can calculate the different evaluation parameters: accuracy, precision, recall, F1 Score and ROC values which can individually discussed in following session.

Initially we are going to test over machine learning model with logistic regression[16], it is a supervised machine learning algorithm is used for binary classification problems. Here the application is also detecting weather the transmission line subjected to fault or not (no-fault), so I used logistic regression model in this application due to its simplicity, easy to setup and easy to rain the machine learning models. The performance of the logistic regression model can be assed from their confusion matrix and ROC Plots as show in figure 3 and 4.

.

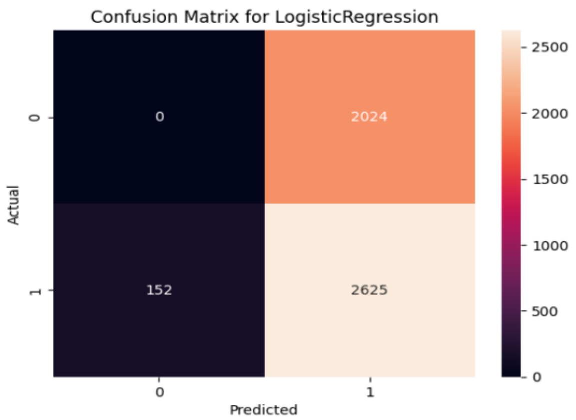

Fig 3 confusion matrix for Logistic regression

From the confusion matrix in figure 3 we can found that the true positive instants are 2625 its means the model correctly predict the fault 2625 times, true negative instants are 0 means that the model does not predict the no-fault instants, false positive instants are 2024 means the model predict the no fault as a fault in 2024 instants, and false negative instants are 152 means the model predict fault as a no-fault in 152 instants. From this the measured accuracy is 0.5467, precision is 0.5646, recall is 0.9452 and F1 score is 0.53. Figure 4 indicated the Receiver operating characteristics of Logistic Regression model, it show a cure how the performance changing for true positive rate verses false positive rate. Here the area under cure (AUC) is 0.91 but the cure is non uniform appearance indicated with orange color.

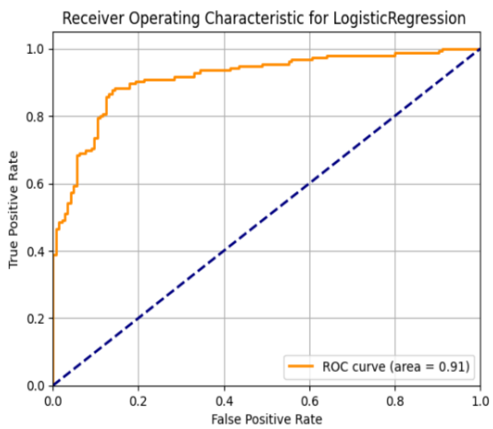

fig 4 ROC for logistic regression

The second technique used to train our machine learning model of transmission line fault detection is support vector classifier. It is a supervised machine learning algorithm used for both classification and regression tasks. It is a powerful tool used to increase the margin between different classes of dataset. The advantages of using support vector machine classifier are it can handle over fitting problems for linear and non-linear data. The following figures 5 and 6 shows the confusion matrix and roc plots of support vector classifier to analyze the performance parameters.

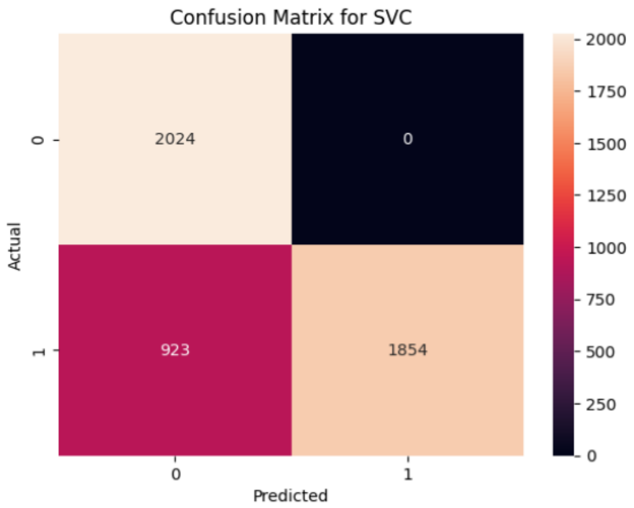

Fig 5 confusion matrix for Support vector classifier

The confusion matrix of support vector classifier in figure 5 shows the true positive instants are 1854 it means the SVC model predict the fault correctly 1854 times. True negative is 2024 it means the SVC model predict the no fault correctly 2024 times. The false positive is 0 it means the SVC model predict zero times the no fault as fault. The false negative is 923 instants it means the SVC classifier founds faults as no faults in 923 instants; it has a high miss classification of fault as a no-fault and may crate several issues in transmission line results a great damage to systems equipment. From this data we can find the accuracy as 0.8077, precision is 1.00, recall 0.6676, F1 score is 0.81. the figure 6 represents the receiver operating characteristics of support vector classifier which show area under curve (AUC) is 0.91 but the shape of the curve is non uniform indicated with the orange color in figure 6.

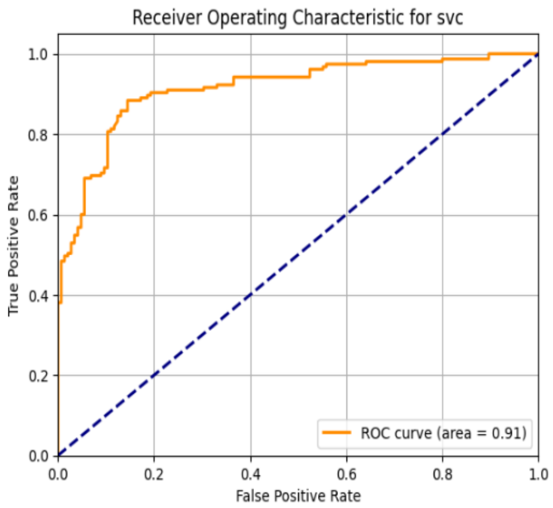

fig 6 ROC for Support Vector Classifier

The third technique which is used to train the machine learning model is Decision tree algorithm due to its ability of understanding the non linear relationships in the data variables. This is a supervised machine learning technique used for both classification and regression tasks[9]. The figures 7 and 8 show the confusion matrix and ROC of decision tree classifier to evaluate the performance parameters.

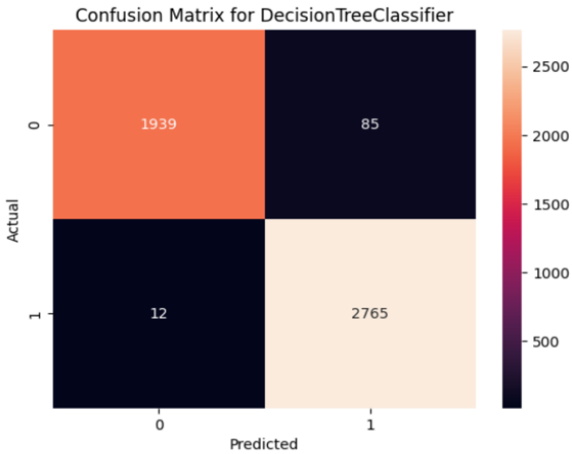

Fig 7 confusion matrix of Decision Tree classifier

From the confusion matrix of decision tree classifier in figure 7 the true positive instants are 2765 means the decision tree model predict fault correctly 2765 times. The true negative instants are1939 which means the decision tree classifier predict no fault correctly 1939 times. The false positive instants are 85 means the decision tree classifier predict no faults as a fault in 85 times. The false negative instants are 12 shows the decision tree classifier finds fault as no fault in 12 times. The accuracy of decision tree classifier is 0.9797, precision is 0.9701, recall is 0.9956 and F1 score is 0.98. The receiver operating characteristics of decision tree classifier is shown in fig 8 from which we can found the area under cure (AUC) is 0.86 which is good and we can observe that the roc cure indicated with orange color is also changing uniformly.

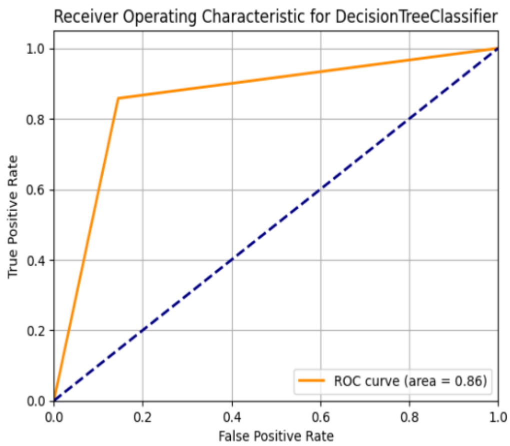

fig 8 ROC of Decision Tree classifier

The final technique which is used to detect the transmission line fault in this work is K Nearest Neighbor algorithm. It is also a simple supervised machine learning algorithm used for both classification and regression tasks. The various advantages of KNN algorithm are it is very simple, easy to understand and enough ability of learning non liner boundaries due to these advantages it becomes special and well performance. The following figures 9 and 10 show the confusion matrix and ROC of KNN algorithems.

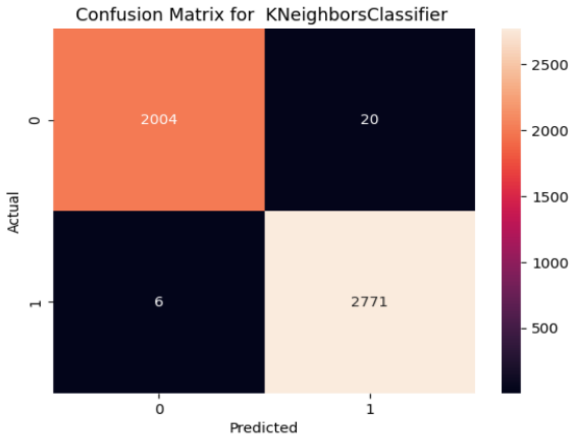

Fig.9 confusion matrix for KNN

From the confusion matrix of KNN Algorithm in figure 9 we can found that the true positive instants are 2771 which indicates the KNN algorithm predict fault correctly 2771 instants, the true negative instants are 2004 which means the KNN algorithms founds no faults correctly in 2004 instants, the false positive is 20 instants which means the KNN algorithm finds no faults as a faults in 20 instants. The false negative is 6 instants which mean the KNN algorithm found faults as no-faults in 6 instants only it is very low error value. Form these values we can found the accuracy as 0.9945 , precision is 0.9928, recall value is0.9978, F1 score is 0.98. form the figure 10 ROC of KNN algorithms we can understand the AUC value is 0.87, the cure is also

Good liner variation as shown with orange color.

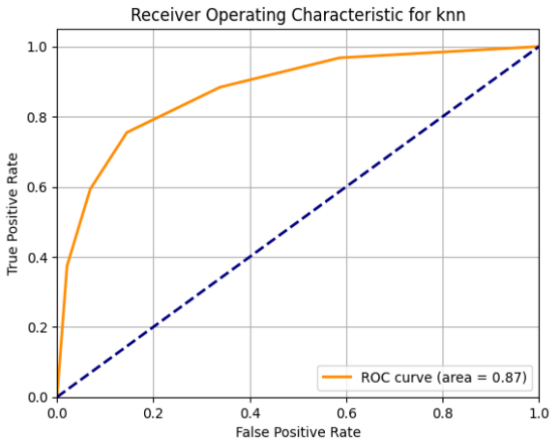

Fig.10 ROC for KNN

The performance evaluation parameters of fault detection in transmission line with different machine learning algorithms are as show in table 2. From this table 2 we can observe how the accuracy, precision, recall, F1 score and AUC values vary from one model to other model among all these the K- Nearest Neighbors algorithm perform well.

Table.2: performance comparison of different machines learning algorithms

Machine Learning Model	Accuracy	Precision	Recall	F1 score	AUC
Logistic regression	0.5467	0.5646	0.9452	0.53	0.91
Support Vector Classifier (SVC)	0.8077	1.00	0.6676	0.81	0.91
Decision tree Classifier	0.9797	0.9701	0.9956	0.98	0.86
K- Nearest Neighbors Classifier (KNN)	0.9945	0.9928	0.9978	0.99	0.87

Conclusion: The four machine learning algorithms logistic regression, support vector classifier, decision tree classifier and K-Nearest Neighbor classifier shows various levels of responses depending on their confusion matrix and Receiver operating characteristics. The logistic regression model has poor accuracy and precision with high AUC value. The logistic regression model is completely fails to detect the faults in transmission line due to their low accuracy and precision. The support vector classifier has sufficient accuracy and precision with good AUC Value, but it struggles with false negatives it misclassify 923 instants of no faults. The Decision tree model is good accuracy and precession with AUC of 0.86 it performance is satisfactory but not up to the mark because slight higher false positive rate. The K-Nearest Neighbors algorithms shows a solid performance of accuracy 0.9945, precision 0.9928, AUC value 0.87 so the KNN algorithm is best fit for transmission line fault detection task.

References

[1] Judy Lhyn Porlaje Sarmiento ., Jam Cyrex De Villa Delfino and Edwin Romeroso Arboleda “Machine learning advances in transmission line fault detection: A literature review” International Journal of Science and Research Archive, 2024, 12(01), 2880–2887, https://doi.org/10.30574/ijsra.2024.12.1.1150

[2] Nandhini K, Prajith C A., “Review on Fault Detection and Classification in Transmission Line using Machine Learning Methods” 2023 International Conference on Control, Communication and Computing (ICCC) 19-21 May 2023, 979-8-3503-3412-8/23/$31.00 ©2023 IEEE

[3] Md.OmaerF. Goni et all., “Fast and Accurate Fault Detection and Classification in Transmission Lines using Extreme Learning Machine” e-Prime - Advances in Electrical Engineering, Electronics and Energy 3 (2023), https://doi.org/10.1016/j.prime.2023.100107

[4] J.C.A.Freire et all., “Transmission Line Fault Classification Using Hidden Markov Models”. IEEE VOLUME 7, 2019, http://creativecommons.org/licenses/by/4.0/

[5] S.R. Fahim et al. “Adeep learning based intelligent approach in detection and classification of transmission line faults”. April 2021 0142-0615/© 2021. https://doi.org/10.1016/j.ijepes.2021.107102

[6] T. T. Ağır, "Using Machine Learning Algorithms for Classifying Transmission Line Faults," Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, vol. 13, no. 2, pp. 227–234, Jun. 2022.

[7] P. P. Wasnik, N. J. Phadkule, and K. D. Thakur, "Fault Detection and Classification in Transmission Line by using KNN and DT

Technique," International Research Journal of Engineering and Technology (IRJET), vol. 7, no. 4, pp. 2395–0072, Apr. 2020.

[8] M. J. B. Reddy, D. K. Mohanta, and S. S. R. Depuru, "K-NN based fault detection and classification methods for power transmission systems," Protection and Control of Modern Power Systems, vol. 2, no. 1, p. 27, Dec. 2017.

[9] S. K. Shukla, E. Koley, and S. Ghosh, "DC Offset Estimation-Based Fault Detection in Transmission Line During Power Swing Using Ensemble of Decision Tree," IET Science, Measurement & Technology, vol. 13, no. 2, pp. 212–222, 2018.

[10] A. K. Sinha and K. K. Chowdoju, "Power System Fault Detection Classification Based on PCA and PNN," in International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT), 2011, pp. 111–115.

[11] C. K. Y. Zerahny et al., "Fault Classification and Location in Three-Phase Transmission Lines Using Wavelet-Based Machine Learning," in 8th International Conference on Intelligent and Advanced Systems (ICIAS), 2021.

[12] S. Singh, D. Vishwakarma, and A. Kumar, "A Novel Methodology for Fault Detection, Classification and Location in Transmission System Based on DWT & ANFIS," Journal of Information and Optimization Sciences, vol. 38, no. 6, pp. 791–801, 2017.

[13] S. K. Singh, D. Vishwakarma, and R. Saket, "An Intelligent Scheme for Categorising Fault Events in Compensated Power Network Using K-Nearest Neighbour Technique," International Journal of Power and Energy Conversion, vol. 11, no. 4, pp. 352–368, 2020.

[14] "Big Trouble in Little Data. Tips for Dealing with Small Data in… | by Brandon Cosley | Towards Data Science." https://towardsdatascience.com/big-trouble-in-little-data-7a1b02bfdc39 (accessed Oct. 19, 2021).

[15] M. Geiger, et al., Scaling description of generalization with number of parameters in deep learning, iopscience.iop.org (2019). Accessed: Oct. 21, 2022. [Online]. Available, https://iopscience.iop.org/article/10.1088/1742-5468/ab633c/meta.

[16] M.H.H. Musa, Z. He, L. Fu, Y. Deng, Linear regression index-based method for fault detection and classification in power transmission line, IEEJ Trans. Electr. Electron. Eng. 13 (7) (2018) 979–987.

[17] A. Saber, A. Emam, H. Elghazaly, A backup protection technique for three-terminal multisection compound transmission lines, IEEE Trans. Smart Grid 9 (6) (2017) 5653–5663.

[18] K. Riehl, M. Neunteufel, and M. Hemberg, "Hierarchical confusion matrix for classification performance evaluation," *arXiv preprint arXiv:2306.09461*, Jun. 2023. [Online]. Available: https://arxiv.org/abs/2306.09461.

[19] S. Sathyanarayanan and B. R. Tantri, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, vol. 27, no. 4S, Nov. 2024. [Online]. Available: https://africanjournalofbiomedicalresearch.com/index.php/AJBR/article/view/4345.

[20] I. Markoulidakis and G. Markoulidakis, "Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis," *Technologies*, vol. 12, no. 7, p. 113, Jul. 2024. https://www.mdpi.com/2227-7080/12/7/113.

[21] K. W. De Bock and D. Van den Poel, "Fair evaluation of classifier predictive performance based on binary confusion matrix," *Computational Statistics*, vol. 37, pp. 1341–1369, 2022. [Online]. Available: https://link.springer.com/article/10.1007/s00180-022-01301-9.

[22] M. S. Mohosheu, F. A. Shams, M. A. A. Noman, S. R. Abir, and Al-Amin, "ROC Based Performance Evaluation of Machine Learning Classifiers for Multiclass Imbalanced Intrusion Detection Dataset," in *Proceedings of the 2023 IEEE 8th International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, Kuala Lumpur, Malaysia, Dec. 2023, pp. 338–343.

[23] C.-I. Chang, "An Effective Evaluation Tool for Hyperspectral Target Detection: 3D Receiver Operating Characteristic Curve Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 6, pp. 5131–5153, Jun. 2021