International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Advancements in Machine Learning Algorithms for Enhanced Fault Analysis and Categorization in Power Systems

Eena Agrawal[1], Dr. Shoyab Ali[2]

[1]Research Student, Electrical Engineering, Regional College for Education Research Technology, Rajasthan, India
[2]Professor, Electrical Engineering, Regional College for Education Research Technology, Rajasthan, India

Abstract
With the growing number of sources, complexity, and dynamic elements of power systems that are prone to disturbance and electrical faults, conventional methods of fault detection are lagging in power system protection. Major faults are exposed to transmission lines due to several factors, such as climate change, heat, lightning strokes, sudden spikes of current and voltage etc. Protecting these lines from faults is cumbersome using traditional methods. However, with the increasing complexity of transmission lines in power systems, advanced technologies such as machine learning come into play for prior fault detection and identification. ML is becoming a popular and intelligent technique for monitoring and predicting health and diagnosing faults in transmission lines of power systems. Since ML is evolving at a faster pace and given the need for and growth of advanced technologies like ML and artificial intelligence, this paper emphasizes comparative analysis of various ML algorithms for transmission line fault analysis. Comparisons for various line faults (LL, LLG, LG, LLL, LLLG, no fault) in power systems using machine learning techniques have been done in terms of accuracy, precision, and other performance measures. The results of the accuracy of various algorithms are represented as tabulated facts in this paper. The Python programming language has been utilized over a dataset to calculate results. For each algorithm, the confusion matrix shows the accuracy of predicting faults over the testing data after training the model or algorithm on the training data. Upon completion of analysis and comparison of MLT's like LR, SVM, DT, RF, K-NN, and gradient descent for analysis of transmission line fault detection, RF and DT machine learning algorithms provided the best accuracy and results. Finally, this paper gives a brief overview of various line faults, ML algorithms accuracy for each transmission fault in a confusion matrix and tabular format and concluded the best machine learning algorithms for power system transmission line fault identification.

Keywords: Power System (PS), Machine Learning (ML), fault detection, SVM (Support vector machine), K-NN (k- nearest neighbors), gradient descent, logistic regression (LR), DT (Decision Tree), random forest (RF), MLT’s (Machine Learning Techniques).

1. Introduction

To ensure continuous and reliable electrical power supply to consumers, it is necessary to maintain the safety of the electrical transmission line and conduct fault analysis over a period, which helps minimize damage, downtime of the power system, and costs associated with faults. Classifying faults is critical to locating the problematic phase and quickly cutting off power to it. This successfully reduces needless power consumption and protects workers and related equipment. In real-life scenarios, four main types of defects appear in three-phase transmission lines [1]. Many fault types are frequently seen in power systems, such as double-line faults (DL), line-to-line (LL), and single-line-to-ground (SLG) faults. DL faults are denoted by LL or DL, while SLG faults are shortened to SLG. Furthermore, triple-line faults are commonly referred to as LLL or 3L faults, while double-line-to-ground problems are designated as LLG or DLG faults. Usually, there are two ways that faults appear: transient or permanent. Transmission lines are prone to temporary faults, which can be caused by unexpected occurrences like falling leaves or huge tree branches, bad weather such as storms, drizzle, or ice and snow, or even creatures like birds flying between the lines and abruptly shorting them out. Although they are frequent, transient faults are less dangerous for power system investigation and protection. The transmission line faults cause sudden spikes in voltages and currents in the line, which sometimes lead to life in danger and create the instability of power substations. Due to transmission line faults power outages happened, which majorly affected commercial, industrial, and residential consumers. and due to these outages, production came to a halt, which caused financial losses for our economy. There are different types of line faults, i.e., LG, LL, LLG, LLL, and LLLG, among which the most severe is the LG fault [11]. So, effective fault detection in transmission lines is required to prevent disaster and maintain the reliability and integrity of power systems. As times increase, handling power system operations and controlling the system are becoming more challenging for the power system operators, and manual monitoring of the system and manual maintenance are difficult for large, complex power systems these days [18]. Previously, power system operators started to detect faults when customers filed complaints. However, the complex integration of various equipment and monitoring devices in the power system has increased the time for detecting faults. The power system operators manually analyze the power system fault location using traditional methods such as impedance and wavelet-based methods. Though these techniques were widely used previously, fault detection using them is very cumbersome and time-consuming and heavily relies on mathematical calculations. Therefore, the investigation of fault identification, categorization, and position-finding techniques in power transmission networks has been the subject of much research.

Hence, it is necessary for early fault detection, classification, and localization to avoid interruptions of continuous power supply, sudden failures, and outages to ensure the reliable operation of the power system. So, a concentrated effort needs to be made to create an automated safety framework that can accurately identify, categorize, and locate faults. With the quick identification and detection of fault locations, the interruption of power supply can be minimized, and it is possible to restore the stable state of the power system, which is helpful to the economy and reduces the cost of manual laboring to find out the faults [4]. These can be done using artificial intelligence techniques. Recently, machine learning techniques have been widely used for fault identification and analysis of transmission lines in power systems.

For fault detection, classification, and localization using machine learning techniques, the major relies on the historical data of the power system, which consists of fault or no-fault data, and trains the machine learning models for automatic decision-making so that the power system can adjust and automate the process to maintain the reliability of the system [3]. Training the machine learning model using Python programming eliminates the dependency on manual mathematical calculations and speeds up the process of making decisions to identify the type of fault.

Due to the increasing concerns about detecting and identifying the types of faults, this research paper discussed how to find out the types of faults automatically using machine learning techniques [10]. This paper consists of a comparative analysis of different transmission line faults such as LL (line-to-line fault), LG (line-to-ground fault), LLG (double line to ground), LLL (triple line fault), and LLLG (triple line to ground fault) using machine learning algorithms such as logistic regression (LR), gradient descent algorithm, K-nearest neighbors (k-NN), SVM (support vector machine), decision tree algorithm (DT), and random forest (RF) and focuses on finding one robust machine learning model by comparison of each model accuracy on testing data points, which helps to find the most severe fault, which is line-to-ground fault. However, each machine learning algorithm majorly gives accurate results, but it is necessary to find out one robust ML technique that can be helpful to fault diagnosis and identification quickly and with the best decision-making, help reduce transmission line faults, and maintain the stability or reliability of the power system [8]. This paper has contributed to fault identification using machine learning models by comparing the various models and calculating the performance of each model through accuracy and result assessments, with the goal of efficiently detecting and classifying problems in electric power lines [16].

In this research paper, static data from the power system has been utilized, and from that huge static data, training data and testing data have been separated, on which analysis has been done using python programming. The training data utilized by machine learning models to train these models has been tested on testing data, and performance has been calculated for each ML model for various kinds of transmission line faults and compared to the accuracy of the models [7]. For each ML algo, a confusion matrix has been created that shows the accuracy of predicting faults in comparison to true values, and the comparison clearly depicts which ML algo is robust to use for fault detection and identification. Improving fault classification or identification methods in transmission line systems for electricity is essential to raising system efficiency and reducing serious damage, given the wide variety of electrical systems and how they are utilized [13].

Lastly, this paper covers an overview of various types of transmission line faults identification and detection, and following that, a description of each machine learning algorithm has been given with a confusion matrix of testing data, which clearly depicts the reliability or robustness of each ML model. All the results, facts, performance, and success rate for each ML model utilized for line faults are represented in a tabular manner in the result section, and the results are concluded in conclusion section.

2. Transmission Line Fault Analysis and Classification

For the stability and reliability of the transmission line of a power system, it is necessary to utilize advanced artificial intelligence, such as machine learning models, for quick fault detection and classification. This section explains the application of machine learning algorithms in classifying various types of faults, including LL (line-to-line fault), LG (line-to-ground fault), LLG (double line to ground), LLL (triple line fault), and LLLG (triple line to ground) [6]. The research paper is based on the utilization of power system transmission line data of current and voltages for three-phase lines, which consist of fault and no-fault data.

Power system transmission line data consist of current and voltage data of three phase lines, which is considered input; Ia, Ib, Ic, Va, Vb, and Vc are considered input features, which are current and voltage data of three phase lines a, b, and c (in general, R, Y, and B); and output taken as matrix data for ground and all three phases, which is shown in matrix via binary number (0 or 1) for different types of faults. The matrix representation of output is shown in Table 1. where G represents ground and A, B, or C represent 3-phase transmission lines. Here, 0 is treated as a no, and 1 is yes for G, A, B, and C faults. Each combination for [G, C, B, A] shows the specific types of faults. For the easy classification of faults, each fault in the transmission lines of a power system is encoded in a numeric number, which is also represented in Table 1.

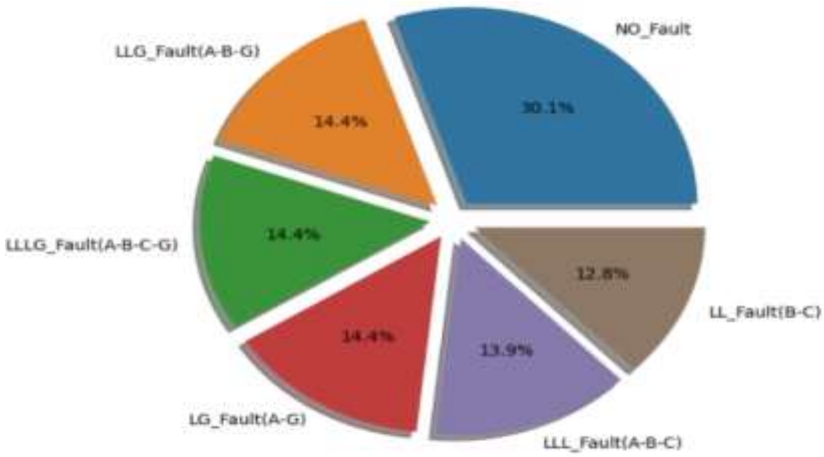
Furthermore, on analysis of power system data, it was found that almost 70% of the of the data is faulty data and 30% is no fault data.

In addition, this data is categorized and illustrates that almost 45% of ground faults happen in power systems. In Fig. 1, the percentage of different types of faults can be observed and analyzed.

Table 1. Representation of fault Matrix with encoding

Ground (G)	Phase C	Phase B	Phase A	Type of Fault	Encoded Number
0	0	0	0	No Fault	5
1	0	0	1	LG	0
0	0	1	1	LL	4
1	0	1	1	LLG	1
0	1	1	1	LLL	3
1	1	1	1	LLLG	2

FIGURE 1. Percentages of Transmission Line Fault in Dataset



The distribution of features, i.e., voltage and current, of three-phase transmission lines has been done, which is crucial for the machine learning model to understand, learn, and differentiate among various transmission line faults in the power system. Each input feature, such as Ia, Ib, Ic, Va, Vb, and Vc, has been represented by a box plot against various faults, which shows how current and voltage values vary with respect to the different fault types or conditions shown in Figure 2.

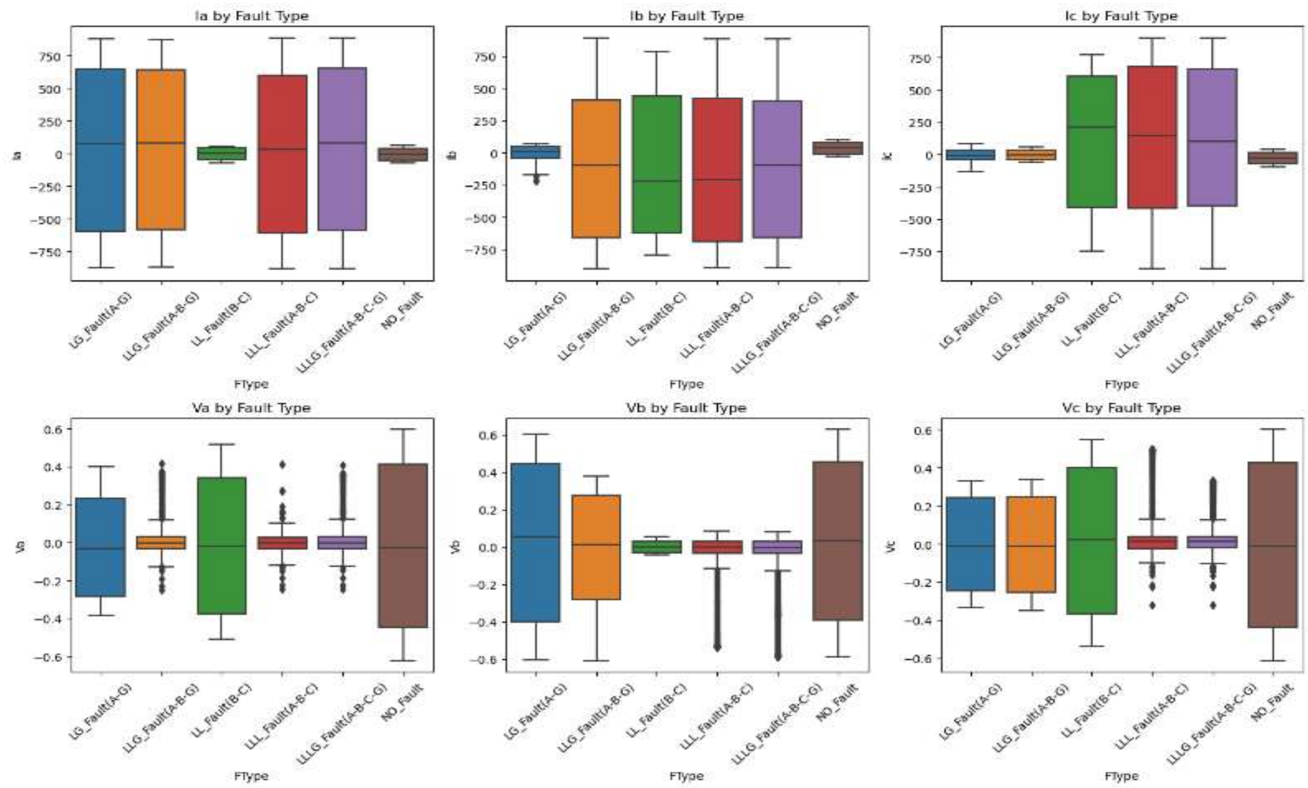


FIGURE 2. Box Plot of Input Features v/s fault types

Later, it represents all three-phase current and voltages over a single graph separately where large fluctuations occur, but it is quite tedious for those graphs to determine which types of faults have which fluctuations. Moving on, for each fault, a different value of three-phase currents and voltages is observed, which is shown below in Fig. 3 to Figure 14.

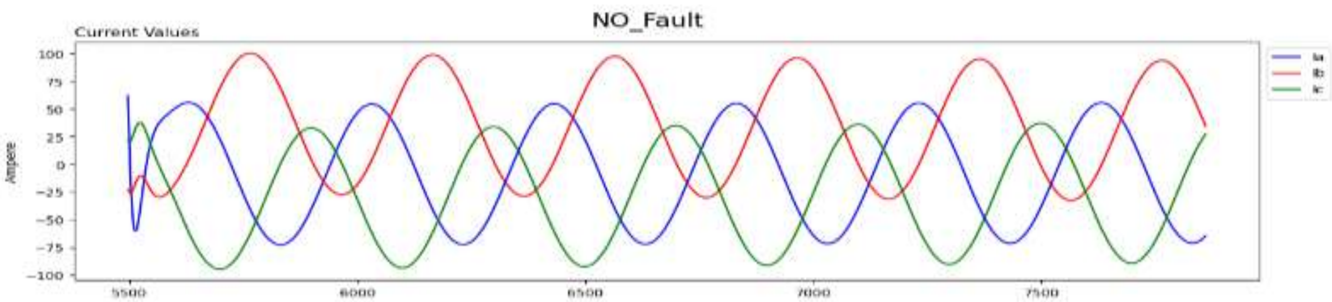


FIGURE 3. Three phase current distribution for no-fault

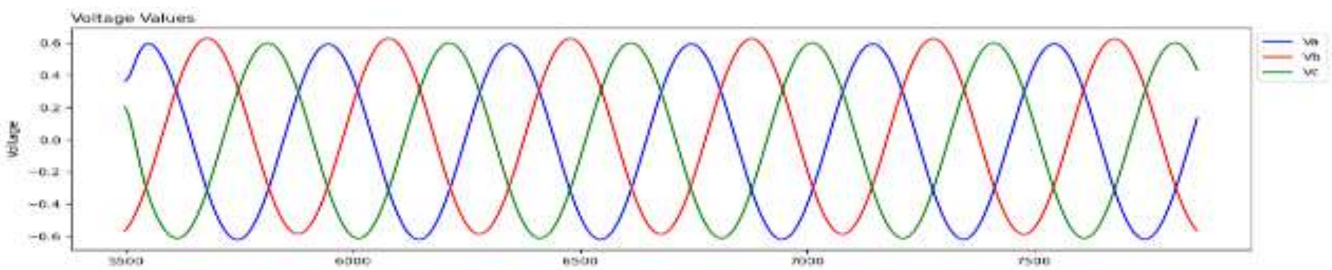


FIGURE 4. Three phase Voltage distribution for no-fault

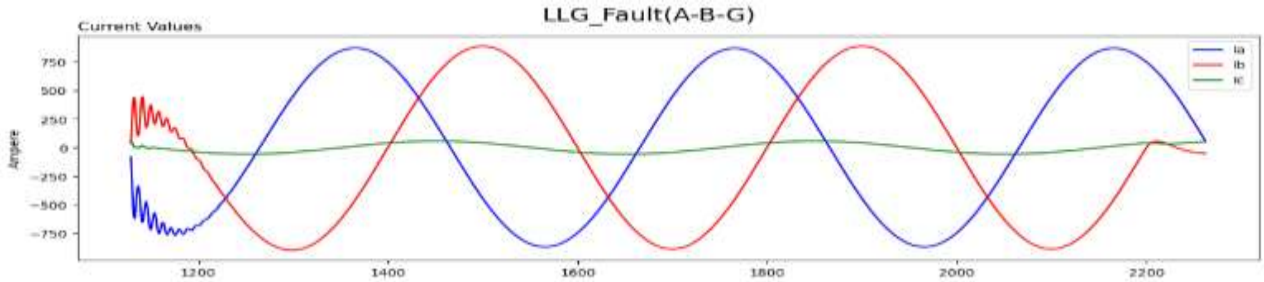


FIGURE 5. Three phase current distribution for LLG fault

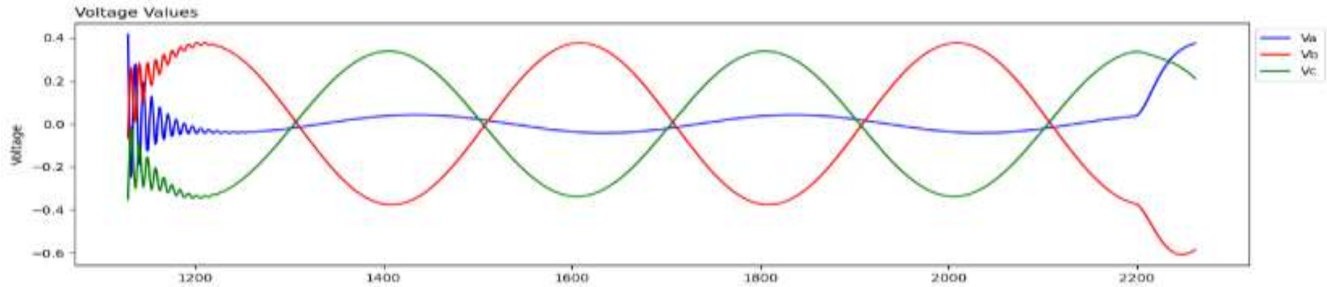


FIGURE 6. Three phase Voltage distribution for LLG fault

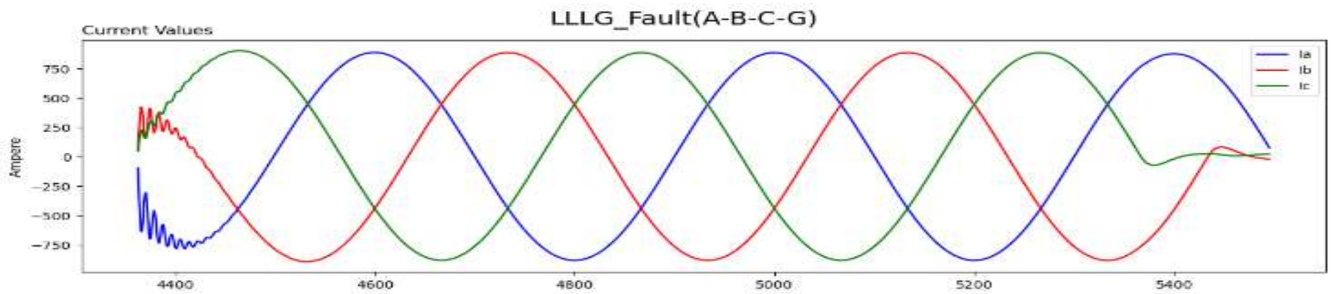


FIGURE 7. Three phase current distribution for LLLG fault

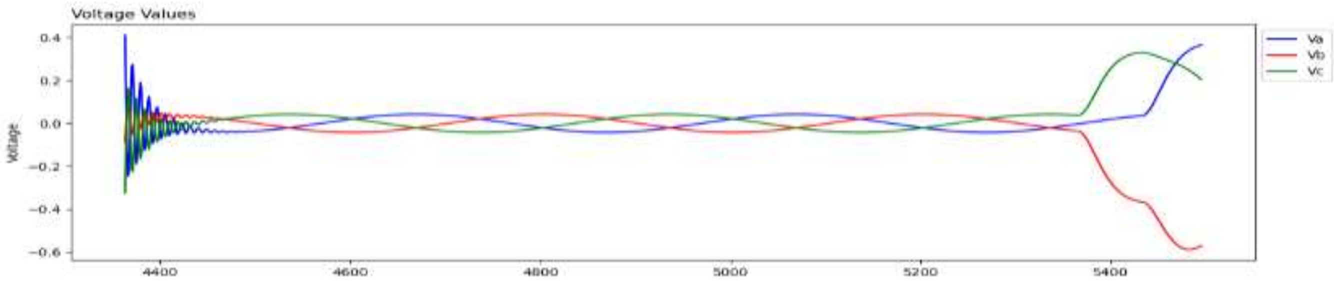


FIGURE 8. Three phase Voltage distribution for LLLG fault

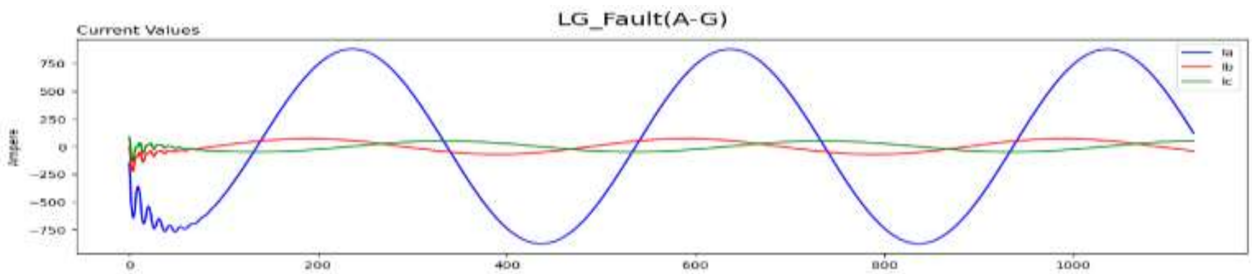


FIGURE 9. Three phase current distribution for LG Fault

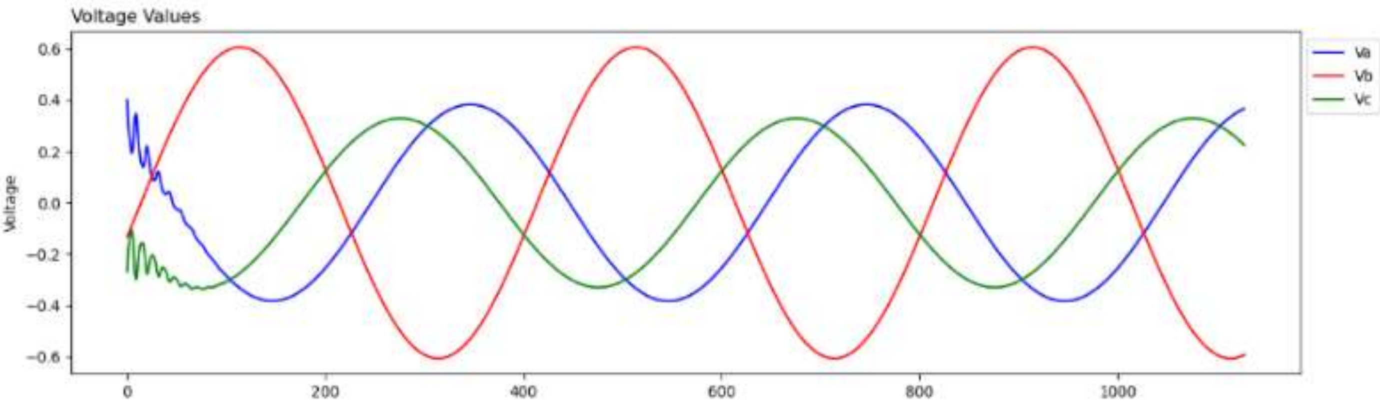


FIGURE 10. Three phase Voltage distribution for LG fault

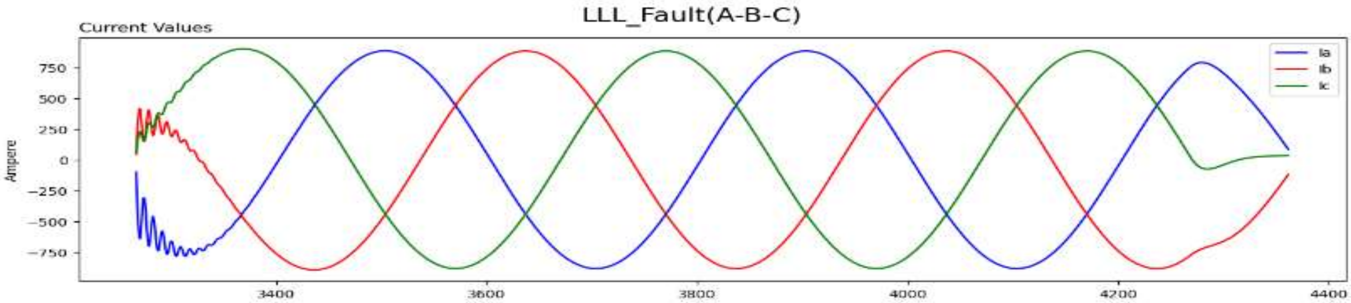


FIGURE 11. Three phase current distribution for LLL Fault

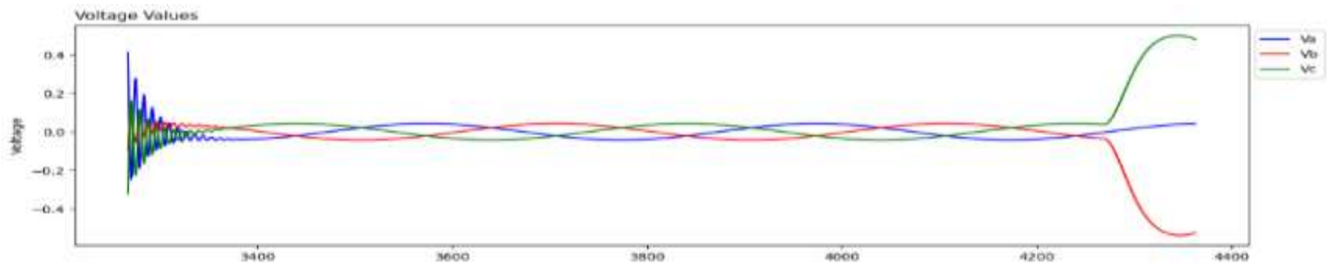


FIGURE 12. Three phase Voltage distribution for LLL Fault

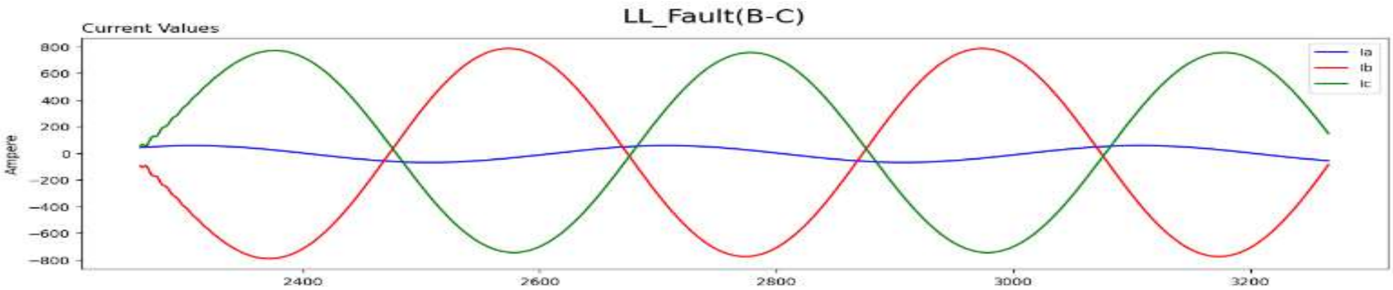


FIGURE 13. Three phase current distribution for LL Fault

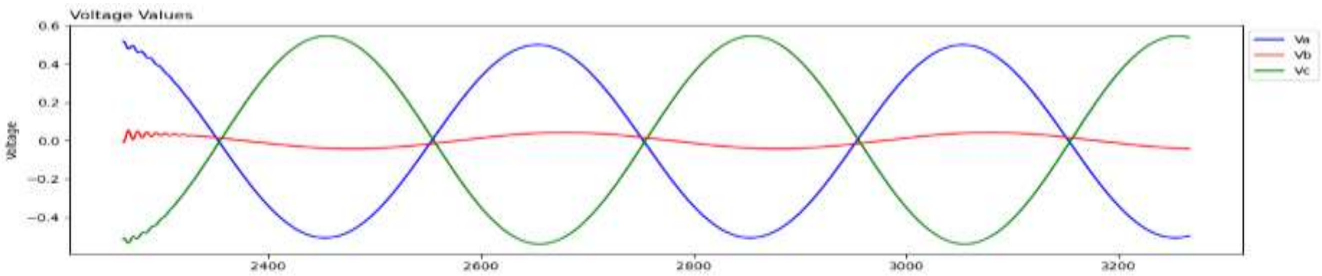


FIGURE 14. Three phase Voltage distribution for LL Fault

As can be observed from Fig. 3-13, there are symmetrical curves for no fault in three-phase current and voltages, but for various types of faults, disruptions always occur in three-phase voltages and currents. After analysis of data and three-phase current and voltage behavior against various faults, various machine learning models have been trained on 80% of the transmission line voltage and current data of the power system. For each machine learning model, the best tuning parameter values have been found with the help of Python programming. Each model has been fed to the best parameters and trained or utilized to predict the faults on testing data, which is approximately 20%. Various machine learning models as logistic regression (LR), gradient descent algorithm, K-nearest neighbors (k-NN), SVM (support vector machine), decision tree algorithm (DT), and random forest (RF), are utilized for training and predicting the type of faults, and for each machine learning algorithm, accuracy is also calculated. A classification report has been generated for each machine learning algorithm, which depicts the model's accuracy and shows how accurately faults in transmission lines are predicted for testing data. Similarly, for each machine learning technique, a confusion matrix has been generated, which shows the accuracy of the prediction of the various kinds of faults [9]. Fault analysis and classification have been done with different machine learning models, and results are computed in terms of the accuracy of predictions over testing data. To classify the transmission line faults of power systems, a matrix of various machine learning models over testing data has been given in the next section, along with a brief description of each algo.

Our focus here is to use the best and optimal machine learning model to identify the fault type in the transmission line of an electrical power system by using the best optimal parameters found through an iterative process with different parameters and calculating the results.

3. MACHINE LEARNING TECHNIQUES FOR FAULT CLASSIFICATION

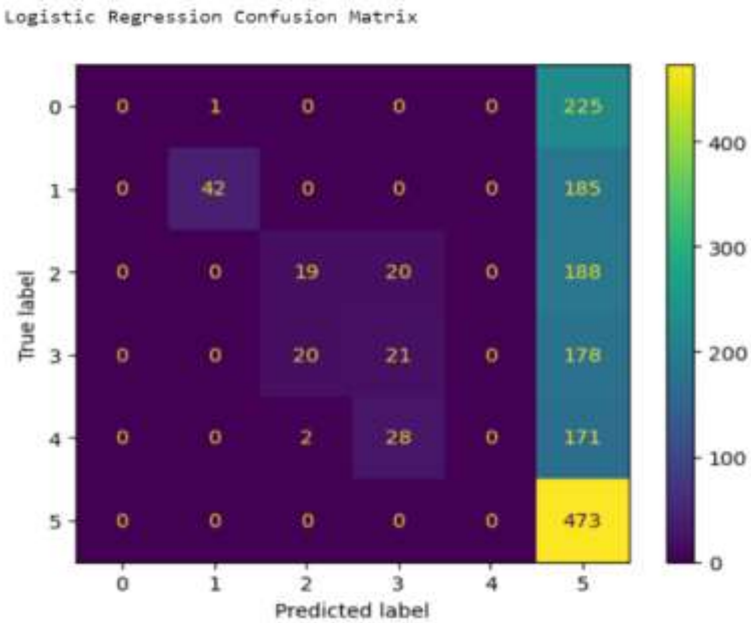
To maintain the reliability and stability of the power system, in this section of the research paper, transmission line fault analysis and detection should be efficiently done with the help of machine learning algorithms [17], as they are used with the help of Python programming and capable of solving any linear or non-linear problems easily. As electrical power systems are dynamic in nature and keep facing fluctuation in current and voltages of transmission lines, to analyze the fault properly and efficiently, this research paper utilized MLT's to cope with the dynamic changes in power systems. This research utilized different machine learning models such as logistic regression, support vector machine, decision tree, random forest, k-NN, and gradient descent as fault classification methods. Machine learning models determine the patterns in changing the current or voltage of the transmission line or try to identify the types of faults in those current and voltage values. Each machine learning model is trained on six input features, which are three-phase transmission line current and voltage values, i.e., Ia, Ib, Ic, Va, Vb, and Vc, and the output of these datasets is in the form of a matrix, which is phase A, B, and C, or ground G,

where values of A, B, C and G are in binary format, either 0 or 1, which shows fault is there or not. Each matrix for [G C B A] consists of a pattern, and each pattern has been assigned as one of the types of faults. Similarly, for each type of fault, there is a numerical number assigned, which makes the process of classifying the fault easier. For example, the [G C B A] value is [1 0 0 1], which is considered a LG fault, and it is encoded as '0'. Similarly, this applies to other types of faults, as shown in Table 1.
This section of research consists of a brief description of each machine learning algorithm utilized for fault detection and classification, along with the confusion matrix of testing data for each machine learning model, which shows how accurately types of faults are predicted by machine learning models.

Logistic Regression Classifier

The logistic regression algorithm is utilized for classification purposes, which is why in this research paper it is used for classification of faults for a set of input features, which are transmission line current and voltages of three phases. In this paper, the LR model is used for multi-class classification because of its adaptability to learn and predict multinomial probability. The benefit of multinomial classification in the LR model is to predict the class labels (types of faults) over the testing dataset. The prediction of faults has been done by Python programming using the predict function. The logistic regression algorithm is tuned for multiclassification of faults using important hyperparameters, i.e., the best parameters, which are the penalty term, the classifier solver, and the value of the weighted coefficient (C), which is the regularization parameter. The best hyperparameter value chosen as 'l2' as penalty, c is 10.0, and 'liblinear' solver.

FIGURE 15. Confusion Matrix of LR Model

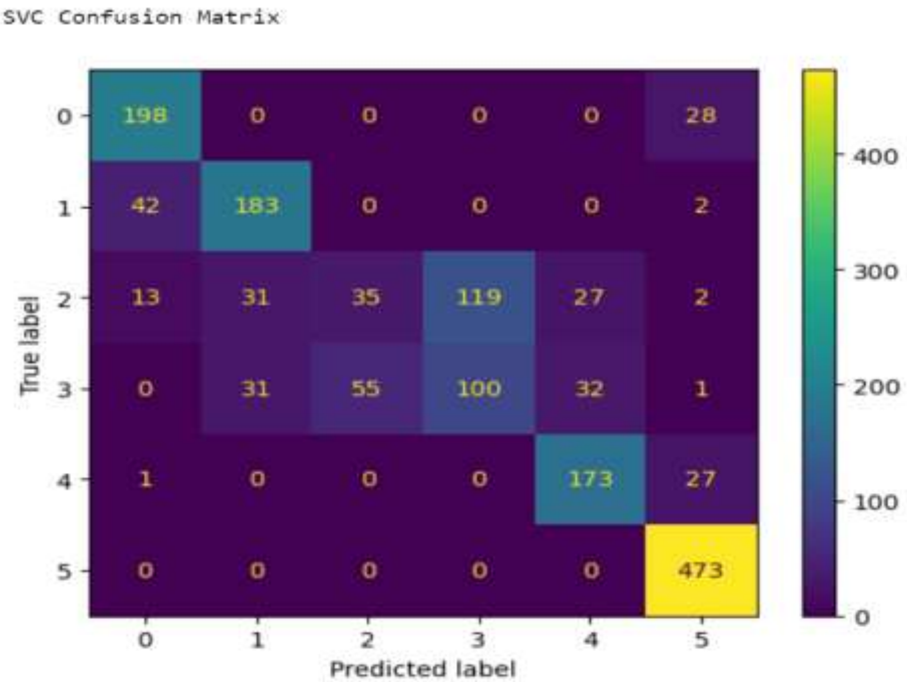


Once the best fit parameter is evaluated, a logistic regression machine learning model is trained on the best fit parameter using the Python fit function. Afterwards, types of faults have been predicted over testing data. The prediction of transmission line faults, i.e., LG, LL, LLG, LLL, and LLLG, over testing data has been shown as a confusion matrix in figure 15. Model accuracy against each transmission line fault is provided in the result section of this paper.

Support Vector Machine Classifier

A support vector machine classifier is a powerful algorithm that is utilized for classification purposes and even for outlier detection. In this research paper, the SVC machine learning technique is used for the classification of transmission line faults in power systems [2]. The main focus of the support vector machine algorithm is to find the optimal hyperplane, which is a cable for dividing data points into various classes, i.e., types of faults in multidimensional space. Support vector machine classifiers consist of efficient memory to keep a subset of training data points and utilize them for the decision-making of fault classification. By collecting the current and voltage data from different sensors for monitoring purposes used in the transmission line of the electrical network of a power system, fault classification is easier and more effective with the help of the SVM model. The aim of the supporting vector classifier is to increase the distance between the hyperplane and the closest data points for each fault class [15]. The model's accuracy or prediction over testing data about the type of fault is shown by the confusion matrix as shown in figure 16. From training the model with the help of the fit function of the classifier, to using the predict function for predicting the fault over testing data, python programming is utilized.

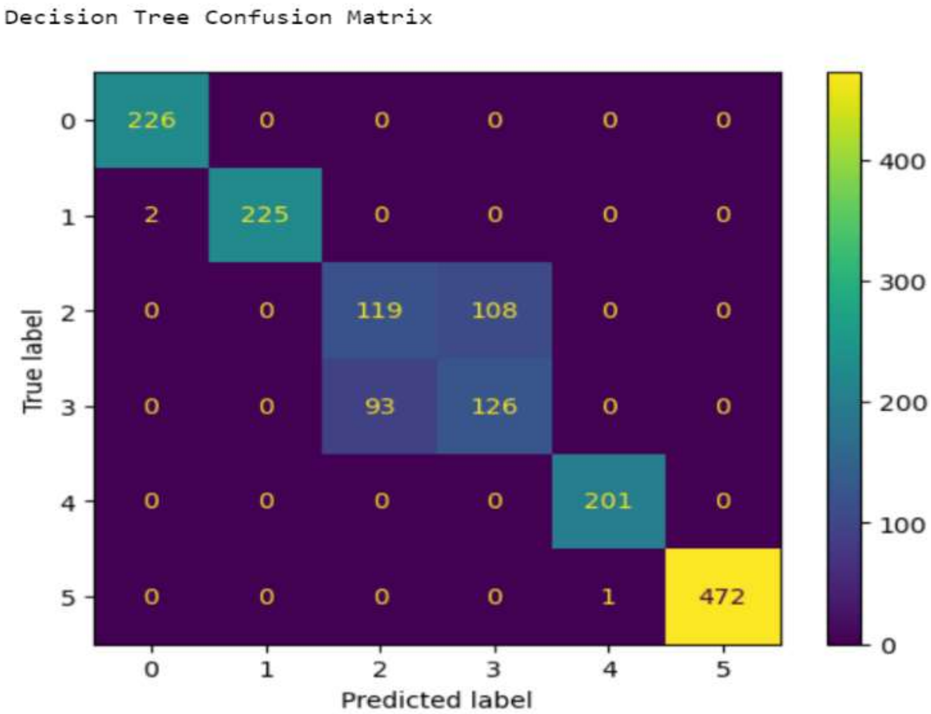
FIGURE 16. Confusion Matrix for SVM Classifier



Decision Tree Classifier [DT]

The decision tree algorithm is a supervised learning algorithm that is used for classification problems. The DT machine learning technique utilizes a tree-like structure where internal nodes are represented as input features, which are voltage and current of the transmission line of the power system dataset; branches of the tree define how data is divided further; and end nodes consist of output, i.e., fault types. Based on decision-making, which is either yes or no, a tree-like structure has been created and utilized by the decision tree algorithm for fault classification of transmission lines [14]. In this research paper, the DT machine learning technique is utilized for fault classification where a model is trained or a graphical representation has been created, and once the prediction starts over testing data, the algorithm utilizes a tree-like structure where for predicting fault for testing data points of current and voltages of line, it first goes to the root node and algo compares the root node with the data point, follows the branch, takes a decision, and moves to the next node of the tree. The process is repeated until the algorithm reaches the end node, which is the output, i.e., classifying the type of fault. The best parameters have been considered by python programming, and a model is trained on those parameters, and a root node for sub nodes, i.e., a decision tree, is created. Fault prediction over testing data has been shown as a confusion matrix in Fig. 17, which shows the accuracy of predicting the type of transmission line fault.

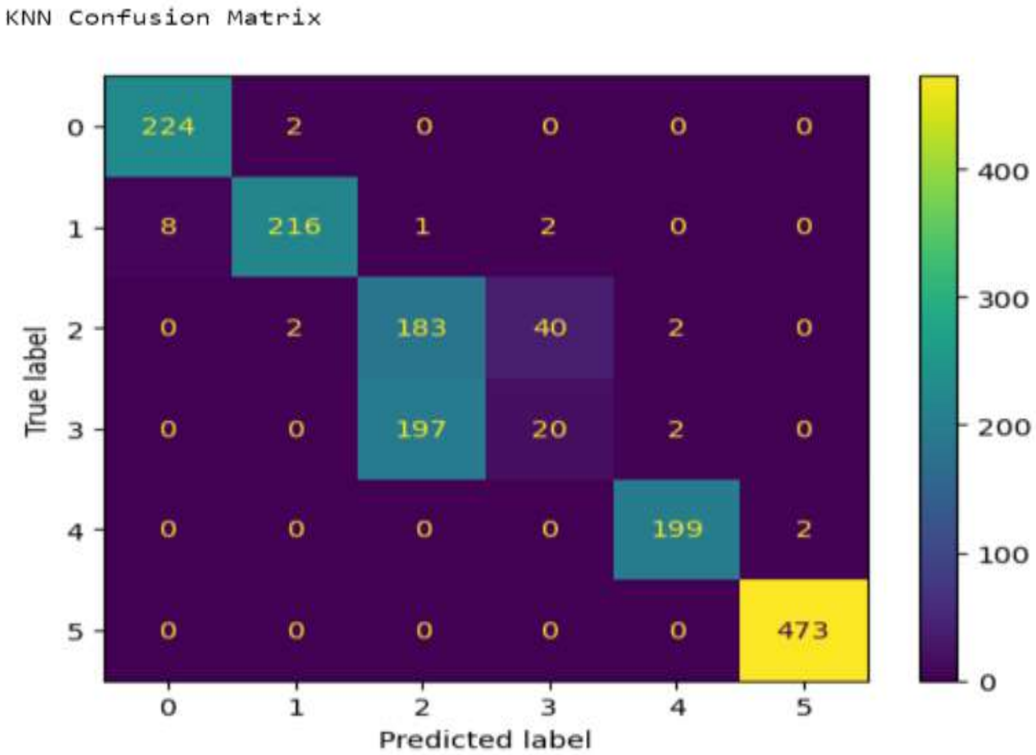
FIGURE 17. Confusion Matrix for Decision Tree



K-NN Classifier

For fault classification, the KNN supervised machine learning technique has been used. The KNN machine learning algorithm works by making assumptions on the basis of similarity between the available data points and new testing data and utilizing the available data points to find out the output for new data points [12]. The KNN machine learning algorithm predicts the faults for testing data, which are the current and voltages of the transmission line, by using the available data stored. So, it is easier to classify the fault types, as while training the model, it will store the training data, and when testing data is given for classification, it provides results, i.e., types of faults, on the basis of the nearest training data point available to the testing data points.

FIGURE 18. Confusion Matrix for KNN

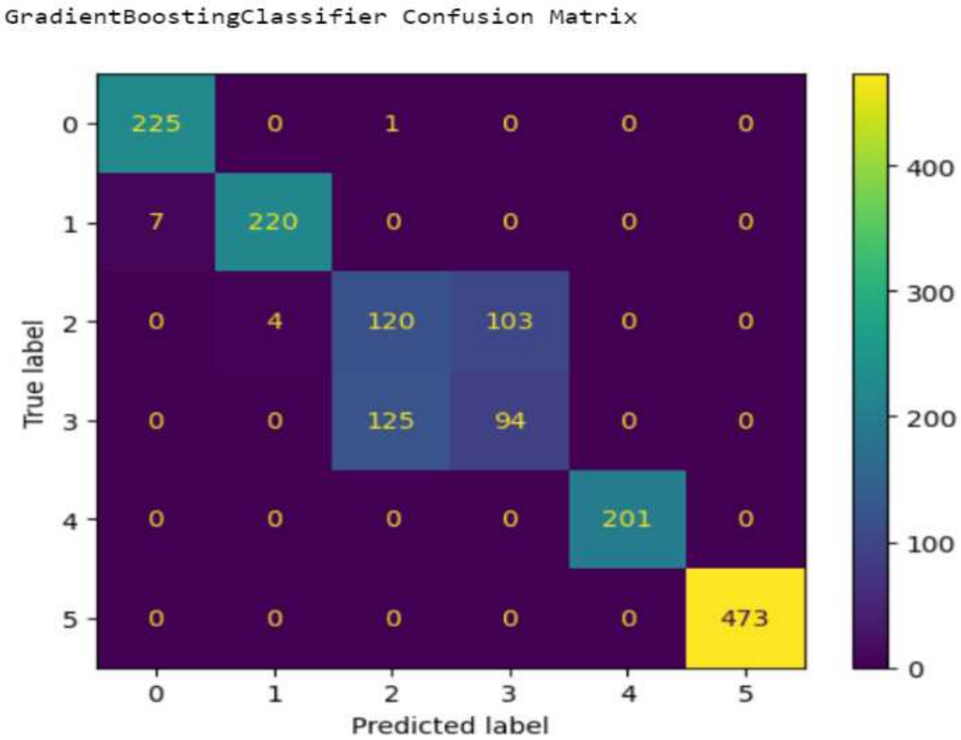


KNN machine learning technique used Euclidean distance to find out the nearest point between testing data and training data points, and the value of 'k' is defined on the basis of the nearest points collected and output, i.e., type of fault, determined on the basis of majority points, and used to predict or classify the faults for testing data. But it has a limitation that its efficiency gradually decreases, when data is increasing, even with this drawback we find this model helpful to identify the fault types in electrical power system [17]. KNN algorithm provide good accuracy to the test data where distribution is not normal event if knn is not a more used for classification purpose. The accuracy of the model has been given as a confusion matrix of testing data against each fault, which shows how accurately each fault is evaluated, as shown in Fig. 18.

Gradient descent Classifier

Gradient descent is one of the powerful optimization machine learning techniques used for the classification of types of faults in the transmission line of a power system. With the help of the gradient descent algorithm, the error between predicted and actual results can be minimized. During the training of the model on the voltage and currents of transmission line data points, the internal cost function plays a crucial role as a parameter, which tries to increase or improve the accuracy of the machine learning model by updating the parameter value with each iteration. In the gradient descent machine learning technique, iteration is continuous until the best parameter value is found to minimize the error or cost. In this research, the best parameter to train the model has been found using a number of iterations with the help of Python programming. And once the model is trained over training data points, i.e., the voltage and current of a three-phase transmission line, it is later utilized to predict the type of faults at each testing data point. The fault prediction accuracy over testing data has been shown in the confusion matrix in Fig. 19, which shows how the model accurately predicts each fault.

FIGURE 19. Confusion Matrix for Gradient Descent Classifier

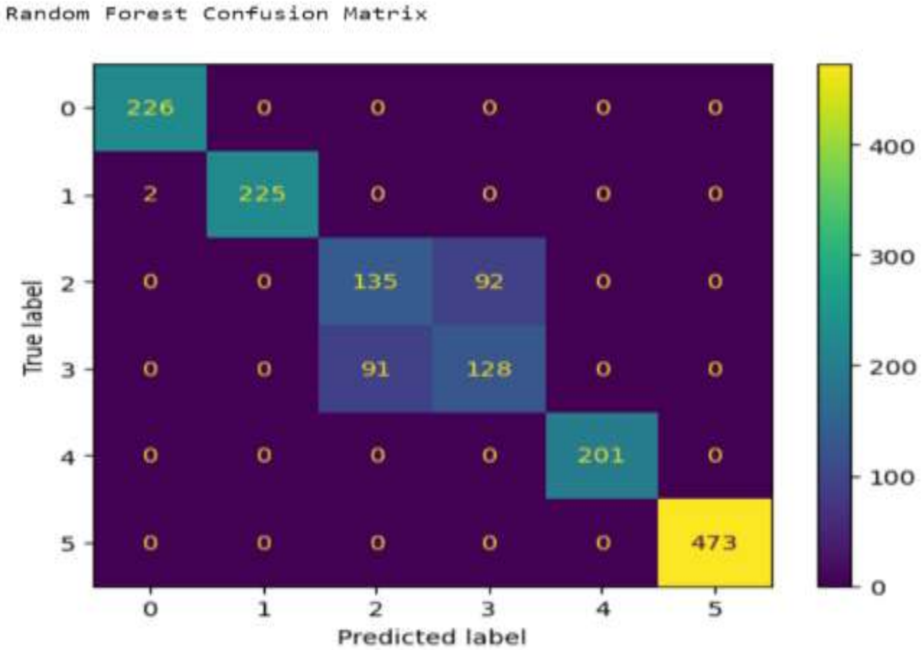


Random Forest Classifier

To identify the fault in the transmission line we can use a most effective supervised machine learning algorithm which is Random Forest. It is based on ensemble learning, which combines all the multiple classifiers for classification problems. In this paper, various decision trees are used and combined to identify transmission line faults and improve the performance and efficiency of the machine learning model. Random forest techniques utilize several decision trees and help predict the type of fault by taking the average outcome of each decision tree used. As the number of decision trees increases, random forest model classifier model efficiency and performance increases.

This model has been used for fault detection because it's highly accurate, efficiently works on larger datasets, and gives high accuracy even if larger data points are missing. In this paper, Python is used to find out the best parameters to train the model, in which parameter estimators represent the number of decision trees taken as 300, and over the best parameters, the model is trained. As part of the evaluation, a confusion matrix is created by comparing the predicted and real fault types to provide an overview of the model's classification accuracy, shown in figure 20.

FIGURE 20. Confusion Matrix for Random Forest Classifier



In this confusion matrix, the diagonal elements represent the number of correctly classified instances for each class (True Positives). Off-diagonal elements represent misclassifications.

4. Results

In this research paper, different supervised machine learning techniques have been utilized for the transmission line fault classification of power systems. For each machine learning model used, a confusion matrix has been created, which shows how the model predicted accurate results for each fault shown in Figs. 15–20. The model accuracy of each type of transmission line fault, i.e., LL (line-to-line fault), LG (line-to-ground fault), LLG (double line to ground), LLL (triple line fault), and LLLG (triple line to ground fault), has been calculated by the confusion matrix, which is created over testing data of voltages and current of three-phase transmission lines of a power system. In this paper, accuracy represents the closeness of the predicted output to the actual output, calculated by the formula given below.

$$\text{Accuracy of ML Models} = \frac{TP}{TP+TN}$$

Where, TP = True Positive

TN = True Negative

Here, if the true value of the fault is the LG fault and the model predicts the LG fault, it comes under TP, or if the predicted value is a different fault, then it goes under TN. In this research paper, for the logistic regression machine learning model, the accuracies for each fault, i.e., LG (line-to-ground fault), LLG (double line to ground), LLLG (triple line to ground fault), LLL (triple line fault), and LL (line-to-line fault), are 0%, 19%, 8%, 10%, and 0%, respectively, as shown in Table 2.

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Table 2. Representation of Accuracy of Fault Prediction of ML Models

Model	Success Rate	LG_Fault	LLG_Fault	LLLG_Fault	LLL_Fault	LL_Fault	NO_Fault
Logistic Regression	35%	0%	19%	8%	10%	0%	100%
Support Vector Classifier	74%	88%	81%	15%	46%	86%	100%
Gradient Boosting	85%	100%	97%	53%	43%	100%	100%
K-Nearest Neighbors	84%	99%	95%	81%	9%	99%	100%
Decision Tree	87%	100%	99%	52%	58%	100%	100%
Random Forest	88%	100%	99%	59%	58%	100%	100%

Even after training the model on the best parameters, the logistic regression model's accuracy is only 35%; the logistic regression model used in this paper correctly predicted no faults majorly. It might be the fact that most of the observations are no fault, almost 30% given in Fig. 1, due to which the model is trained in one way. The second model used in this paper is a support vector machine classifier, whose success rate for fault prediction is 74%. The accuracy of predicting faults such as LG (line-to- ground fault), LLG (double line to ground), LLLG (triple line to ground fault), LLL (triple line fault), and LL (line-to-line fault) are 88%, 81%, 15%, 46%, and 86%, respectively, as shown in table 2. The SVC machine learning model has higher accuracy than the logistic regression model. The third model used in this paper is a decision tree machine learning model, whose success rate for fault prediction is 87%. The accuracy of predicting faults such as LG (line-to-ground fault), LLG (double line to ground), LLLG (triple line to ground fault), LLL (triple line fault), and LL (line-to-line fault) are 100%, 99%, 52%, 58%, and 100%, respectively, as shown in table 2. The DT machine learning model has higher accuracy than the logistic regression and SVM model. The fourth model used in this paper is a K-NN machine learning model, whose success rate for fault prediction is 84%. The accuracy of predicting faults such as LG (line-to-ground fault), LLG (double line to ground), LLLG (triple line to ground fault), LLL (triple line fault), and LL (line-to-line fault) are 99%, 95%, 81%, 9%, and 99%, respectively, as shown in table 2. By the K-NN model, each fault prediction percentage is higher except LLL fault, model predicting LLL fault as LLLG fault majorly, which can be seen by the confusion matrix of K-NN machine learning model shown in fig 18. The fifth model utilized in this paper is a gradient descent machine learning model, whose fault prediction rate is 85%. The accuracy of predicting faults such as LG (line-to-ground fault), LLG (double line to ground), LLLG (triple line to ground fault), LLL (triple line fault), and LL (line-to-line fault) are 100%, 97%, 53%, 43%, and 100%, respectively, as shown in table 2. The gradient descent algorithm's accuracy is higher than that of logistic regression, support vector machines, and k-NN machine learning models. The gradient descent machine learning model almost predicted each fault accurately, except for triple lines. Lastly, the random forest machine learning model is utilized for transmission line fault prediction, and its success rate is 88%. The accuracy of predicting by RF model for each transmission line fault such as LG (line-to-ground fault), LLG (double line to ground), LLLG (triple line to ground fault), LLL (triple line fault), and LL (line-to-line fault) is 100%, 99%, 59%, 58%, and 100%, respectively, as shown in table 2. The random forest model has the highest accuracy for predicting transmission line faults in power systems. However, the logistic regression machine learning approach is not utilized to predict and compute complex designs or data of transmission line faults.

5. Future Scope

There are several facilities provided that can be used in the future for research and evolution. Firstly, we can increase the functionality of the present algorithms that we are using, which are for complex, faulty systems. In order to increase the accuracy of prediction, we have to use complex techniques in the future that have knowledge of the particular domain, and we can also use algorithms or techniques that are advanced in nature. Besides that, there are lots of possibilities to find new approaches to identifying fault location and analysis, as we can use the neural network MLT, which handles or tackles complex designs and computes the complexity of data. In addition, the development of current time evaluation and input techniques into machine learning techniques may improve their resilience and flexibility in flexible applications, providing quick detection and management. Furthermore, broadening the focus of this study to include a range of datasets and application domains would result in a greater understanding of the efficiency of models in a wide range of situations. The research area of machine learning detection of defects and forecasting continues to advance due to the work of a team of experts, business professionals, and specialists in the field. Additionally, in this research paper, static data has been used for fault detection and classification; in the future, streaming data can be utilized for fault analysis using streaming sources like Kafka, Events Hub, etc. All things considered, we may clear the path for more dependable and effective machine learning algorithms that can handle tricky failure situations in a variety of areas by tackling these issues and utilizing the latest technology.

6. Research Gaps

A significant research gap is made clear, in spite the reality that the long list of comparisons gives a solid basis for recognizing the use of algorithms for machine learning, such as Gradient Boosting, Decision Tree, Random Forest, KNN, SVC, and Logistic Regression, in electrical transmission line fault analyzing. Although a lot of research has been done, there are still not enough detailed comparisons which directly compare how well various techniques are effective for fault classification and identification in electric power systems. The available research mainly focuses on the efficiency of steps or provides a restricted number of comparisons among multiple methods. To determine the different benefits and drawbacks of each strategy for handling the complexities of a power system fault study, an in-depth examination of all known techniques is required. Moreover, whereas some studies can provide helpful views on the accuracy of certain algorithms, variations concerning datasets, fault categories, and measurement techniques generally limit the ability to establish comparative analyses. A more detailed understanding of these parameters' importance and effectiveness in real-life situations would be possible with a thorough investigation that defines these characteristics across different machine learning methodologies. To motivate developments in the area and encourage students with the ability to make accurate decisions on algorithm selection as well as execution methods it is essential to address this research gap. By taking consideration of performance measures for example accuracy, computational efficiency, and stability in a wide range of fault conditions, a comparative study will significantly improve advances in machine learning-driven the electric system transmission line identification of faults.

Conclusion

The transmission line fault detection or classification using various machine learning techniques has been reviewed in this paper, along with a description of each machine learning model used to classify the different types of line faults. Machine learning model success rate has been calculated by confusion matrices of ML models utilized in section-III of this paper. It has been concluded that random forest, decision tree, and gradient descent booster classifiers provide higher accuracy (88%, 87%, and 84%, respectively) for predicting faults in transmission lines. However, this research paper shows the limitations of a few algorithms, especially logistic regression, which tends to be difficult for fault analysis. This paper shows how it is important to correct the selection of machine learning techniques for fault analysis. In addition to that, making a comparison of machine learning techniques' effectiveness, complexity, and time of training can improve the effectiveness of algorithms for the classification of transmission line faults in power systems.

Reference

1. Fahim Muntasir, Wasif Al Wazed Wasi, Mahdee Nafis, M. Mofazzal Hossain, “A Fast Response Fault Detection Technique for HVAC Transmission Line Using Machine Learning,” Proceedings of 10th IEEE International Conference on Power Systems (ICPS), December 2023.
2. M. Tanveer · T. Rajani · R. Rastogi · Y.H. Shao, “Comprehensive Review on Twin Support Vector Machines,” Proceeding in Research Gate, May 2021.
3. Baskar. D, Dr. Selvam. P, “Machine Learning Framework for Power System Fault Detection and Classification,” Proceeding in International Journal of Scientific & Technology Research Volume 9, Issue 02, February 2020.
4. Rachna Vaish a, U.D. Dwivedi a, Saurabh Tewari a, S.M. Tripathi b, “Machine learning applications in power system fault diagnosis: Research advancements and perspectives,” Proceeding of Elsevier, Engineering Applications of Artificial Intelligence, Volume 106, November 2021.
5. Nandhini K Prof, Prajith C A, "Review on Fault Detection and Classification in Transmission Line using Machine Learning Methods," Proceedings of 2023 International Conference on Control, Communication and Computing (ICCC) 19-21 May 2023.
6. Daniel Gutierrez-Rojas, Ioannis T. Christou, Daniel Dantas, Arun Narayanan, Pedro H. J. Nardelli & Yongheng Yang, “Performance evaluation of machine learning for fault selection in power transmission lines,” Proceedings of Springer, Volume 64, pages 859–883, 2022.
7. Qian He, Li Chen, Wei Tang, Gong He, Shan Tan, Ya Shu, Li Juan Jiang, “Application and Comparative Analysis of Traditional Machine Learning and Deep Learning in Transmission Line Fault Classification,” Proceedings of IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), December 2022.
8. Jawad Ul Hassan, Imran Fareed Nizami, “Machine Learning Algorithm Analysis for Detecting and Classification Faults in Power Transmission System,” Proceeding in 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2), May 2022.

9. Oishi Jyoti, Md. Munawar Hossain, Emama Nahid, Md. Abu Ismail Siddique, "Comparative Analysis of Machine Learning Algorithms for Transmission Line Fault Detection," Proceeding of 10th IEEE International Conference on Power Systems (ICPS), December 2023.
10. Jian Xie, Inalvis Alvarez-Fernandez, and Wei Sun, "A Review of Machine Learning Applications in Power System Resilience," Proceedings of the IEEE Xplore, 2020.
11. Muhammad Uzair, Li Li, Jian Guo Zhu, "Identifying line-to-ground faulted phase in low and medium voltage AC microgrid using principal component analysis and supervised machine-learning," Proceeding of Australasian Universities Power Engineering Conference (AUPEC), November 2018.
12. Prerana P. Wasnik, N. J. Phadkule, K. D. Thakur, "Fault detection and classification based on semi-supervised machine learning using KNN," Proceeding of International Conference on Innovative Trends and Advances in Engineering and Technology (ICITAET), December 2019.
13. Muhammed Essa Hameed, Mesut Cevik, "Diagnose and Analyze Power System Faults Using Machine Learning," Proceeding in 4th International Conference on Current Research in Engineering and Science Applications (ICCRESA), December 2022.
14. S. R. Samantaray, "Decision tree-based fault zone identification and fault classification in flexible AC transmissions-based transmission line," Proceeding in IET Generation, Transmission and Distribution, vol. 3, no. 5, pp. 425-436, 2009, doi: 10.1049/iet-gtd.2008.0316.
15. A. Naresh kumar, M. Suresh Kumar, M. Ramesha, Bharathi Gururaj, A. Srikanth, "Support vector machine-based fault section identification and fault classification scheme in six phase transmission line," Proceeding of IAES International Journal of Artificial Intelligence (IJ-AI), Vol. 10, No. 4, December 202.
16. Dr. Suvarna Pawar, Sahil Gannarpwar, Aman War, Rahul Khandvi, Aditya Talokar, "An Intelligent Approach of Faults Detection and Location Methods for Power Transmission Line," Proceeding of Interantional Journal Of Scientific Research In Engineering And Management, November 2023.
17. Nguyen Nhan, Le Van Dai, "Fault Identification, Classification, and Location on Transmission Lines Using Combined Machine Learning Methods," Proceeding of International Journal of Engineering and Technology Innovation, February 2022.
18. Oyeniyi Akeem Alimi, (Member, IEEE), Khmaies Ouahada, (Senior Member, IEEE), And Adnan M. Abu-Mahfouz (Senior Member, IEEE), "A Review of Machine Learning Approaches to Power System Security and Stability," Proceeding of IEEE Access, June2020.