A DISSERTATION

ON

AN INTELLIGENT APPROACH TO TRANSMISSION LINE FAULT ANALYSIS

Submitted in Partial Fulfillment for the Requirements for the Award of the Degree of

MASTERS of TECHNOLOGY in ELECTRICAL ENGINEERING

(with Specialization in POWER SYSTEM ENGINEERING)

Under

ASSAM SCIENCE AND TECHNOLOGY UNIVERSITY

Submitted By:

MRIDUSMITA BORUAH

M.Tech 4th Semester

Roll No.: PG-E-22-10

ASTU Roll No.: 220620064008

Under the guidance of

Dr. Barnali Goswami (Prof.) & Mrs. Ritu Nazneen Ara Begum (Asst. Prof.)

DEPARTMENT OF ELECTRICAL ENGINEERING

ASSAM ENGINEERING COLLEGE

JALUKBARI, GUWAHATI -781013

SESSION: 2022-2024

DECLARATION

I hereby certify that the work contained in this thesis is original and has been done by me under the guidance of my supervisor Prof. Dr. Barnali Goswami and Asst. Prof. Mrs. Ritu Nazneen Ara Begum. The work has not been submitted to any other institute for any degree or diploma. I have followed the guidelines provided by Assam Science and Technology University, Guwahati in preparing the report. I have conformed to the norms and guidelines given in the Ethical Code of Conduct in the University. Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources. I have given due credit to them by citing them in the text of the report and giving their details in the references.

Date:

Place: Guwahati

MRIDUSMITA BORUAH

M. Tech 4th Semester

Roll No.: PG-E-22-10

ASTU Roll No.: 220620064008

Department of Electrical Engineering

Assam Engineering College

Jalukbari, Guwahati-781013

CERTIFICATE FOR APPROVAL FROM THE SUPERVISOR

This is to clarify that the project entitled "AN INTELLIGENT APPROACH TO TRANSMISSION LINE FAULT ANALYSIS" has been carried out and presented by:

MRIDUSMITA BORUAH

ASTU Roll No: 220620064008

College Roll No: PG/E/22/10

Session: 2022-2024

student of M.Tech 4th Semester (Electrical Engineering), Assam Engineering College, under my supervision and guidance in a manner satisfactory to warrant its acceptance as a prerequisite for the award of the degree of Masters of Technology in Power System Engineering of Electrical Engineering, Department of Assam Engineering College, Jalukbari under Assam Science and Technology University.

Further, the report has not been submitted/reproduced in any form for the award of any other degree/diploma.

Date:

Place: Guwahati

DR. BARNALI GOSWAMI

(Professor)

Department of Electrical Engineering

Assam Engineering College

Jalukbari, Guwahati-781013

CERTIFICATE FOR APPROVAL FROM THE Co-SUPERVISOR

This is to clarify that the project entitled "AN INTELLIGENT APPROACH TO TRANSMISSION LINE FAULT ANALYSIS" has been carried out and presented by:

MRIDUSMITA BORUAH

ASTU Roll No: 220620064008

College Roll No: PG/E/22/10

Session: 2022-2024

student of M.Tech 4th Semester (Electrical Engineering), Assam Engineering College, under my supervision and guidance in a manner satisfactory to warrant its acceptance as a prerequisite for the award of the degree of Masters of Technology in Power System Engineering of Electrical Engineering, Department of Assam Engineering College, Jalukbari under Assam Science and Technology University.

Further, the report has not been submitted/reproduced in any form for the award of any other degree/diploma.

Date:

Place: Guwahati

MRS. RITU NAZNEEN ARA BEGUM

(Asst. Professor)
Department of Electrical Engineering

Assam Engineering College

Jalukbari, Guwahati-781013

CERTIFICATE OF ACCEPTANCE FROM THE HEAD OF DEPARTMENT

This is to certify that the project entitled "AN INTELLIGENT APPROACH TO TRANSMISSION LINE FAULT ANALYSIS" has been submitted by:

MRIDUSMITA BORUAH

ASTU Roll No: 220620064008

College Roll No: PG/E/22/10

Session: 2022-2024

in partial fulfillment of the requirements for the award of the degree of Masters of Technology in Power System Engineering of Electrical Engineering Department of Assam Engineering College, Jalukbari under Assam Science and Technology University.

Date:

Place: Guwahati

DR. RUNUMI SARMA
(Professor & Head of the Department)
Department of Electrical Engineering
Assam Engineering College
Jalukbari, Guwahati-781013

CERTIFICATE FOR APPROVAL FROM THE EXTERNAL

This is to clarify that the project entitled "AN INTELLIGENT APPROACH TO TRANSMISSION LINE FAULT ANALYSIS" has been carried out and presented by:

MRIDUSMITA BORUAH

ASTU Roll No: 220620064008

College Roll No: PG/E/22/10

Session: 2022-2024

student of M.Tech 4th Semester (Electrical Engineering), Assam Engineering College, submitted the dissertation report in a manner satisfactory to warrant its acceptance as a prerequisite for the award of the degree of Masters of Technology in Power System Engineering of Electrical Engineering, Department of Assam Engineering College, Jalukbari under Assam Science and Technology University.

Further, the report has not been submitted/reproduced in any form for the award of any other degree/diploma.

Date:

Place: Guwahati

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

I take this opportunity to express my sincere gratitude and heartiest thanks to my supervisor and co-supervisor Dr. Barnali Goswami and Mrs. Ritu Nazeen Ara Begum respectively, Department of Electrical Engineering, Assam Engineering College, for their valuable guidance, cooperation, and helpful suggestions throughout my project. Their keen interest in this topic has been a constant source of encouragement to me.

I would also like to express my heartiest thanks to our respected Head of the Department Dr. Runumi Sarma for valuable advice and help during the entire course of our project.

I am indebted to all the faculties of the Department of Electrical Engineering for the deep insights and discernment given through the various courses they taught. I also owe it to my family and all my friends who have supported me during my studies.

Date:

Place:

MRIDUSMITA BORUAH

ASTU Roll No: 220620064008

College Roll No: PG/E/22/10

Session: 2022-2024

CONTENTS

LIST OF FIGURES

ABBREVIATION

(1) ML-Machine Learning.

(2) AI-Artificial Intelligence

(3) DL-Deep Learning

(4) CM-Confusion Matrixes

ABSTRACT

This research work focuses on a spontaneous fault detection (FD) and fault classification (FC) system based on a machine learning approach. Upon normalizing the data, a machine learning algorithm was used as the classifier. In the contemporary world, machine learning (ML) is extensively used in every aspect of life.

In this work, primarily two different machine learning models Random Forest and Naive Bayes have been proposed for fault detection and fault classification purposes through training and by comparing their accuracies. The method achieved fault detection accuracies of 85.7% and 80.61% for the Random Forest and Naive Bayes respectively. Random Forest and Naive Bayes are two widely used algorithms with distinct characteristics. By comparing, it has beenobserved the Random Forest model demonstrates a shorter processing time and reduced computational complexity.

Transmission line faults present significant challenges to the reliability and stability of power distribution systems. In this abstract, we explore the application of the Random Forestalgorithm for fault analysis in transmission lines such as 1L-G, 2L-G, and 3L-G faults. RandomForest, a powerful ensemble learning technique, offers a data-driven fault detection and classification approach. By leveraging historical fault data and system parameters, Random Forest demonstrates robust performance in accurately identifying various faults and providinginsights into fault patterns and relationships. Through analysis, this abstract highlights the efficacy of Random Forest in enhancing fault analysis accuracy and efficiency, therebycontributing to the improvement of power transmission reliability and resilience. The waveform of different kinds of faults obtained from the Random Forest algorithm was observedand analyzed.

CHAPTER-1

INTRODUCTION

1.1 GENERAL

To provide stability and a continuous supply of power, the detection and classification of faults in the transmission lines (TLs) are important in this modern age. It is required to remove a faulty section from a healthy section to provide safety and minimize power loss due to the fault. Faults affect the reliability of the power system network. The frequency and degree of faults determine the downtime (outage time) of a power system network. The more frequent fault occurs, the less reliable the power system is. To ensure a high degree of reliability and to provide quality service at a reduced cost to consumers, protection, and control of fault are required. Successful protection and regulation of faults in the power system demand adequate knowledge and analysis of faults.

In the current era of industry, the electrical power demand is increasing continuously. To fulfill the demand, the number of power generation units is also increasing. All these units are connected through a complex power system network (PSN), which has three major components: power generation, transmission, and distribution. Power is transferred from one place to another place through the transmission lines. Reliable and stable operation of the power system is essential to minimize its impacts on industry, business, transportation, and domestic sectors. Any fault in the power system can cause major disruption causing significant financial loss. Hence, uninterrupted and secured power transmission is vital to ensure a country's economic activities.

Power outages occur due to different reasons such as transmission line insulation breaks, thunderstorms, equipment failure, human intervention, animal interference, fallen trees, and so on. Transmission lines are short-circuited due to these reasons and make the power system unstable. A huge amount of power is lost during a short-circuit fault. To provide stability, it is necessary to detect the fault type and its physical location accurately. Afterward, the faulty section must be removed from the healthy section to guarantee a smooth flow of power.

In real-life scenarios, a relay and circuit breaker perform this operation. However, actuating the relay is time-consuming and this operation can be made faster using machine learning

(ML). Using available fault data such as fault voltage and current the ML model can be trained to identify and classify the faulty cases. Hence, in the area of power systems, researchers have focused on detecting and classifying faults using different Deep Learning (DL) and ML algorithms.

The transmission line is the most crucial part of the power system. The requirement of power and its allegiance has grown exponentially over the modern era, and the prominent role of a transmission line is to transmit electric power from the source area to the distribution network. The electrical power system consists of so many complex dynamic and interacting elements that are always prone to disturbance or an electrical fault.

Normally, a power system operates under balanced conditions. When the system becomes unbalanced due to the failures of insulation at any point or due to the contact of live wires, a short–circuit or fault, is said to occur in the line. Faults may occur in the power system due to several reasons like natural disturbances (lightning, high-speed winds, earthquakes), insulation breakdown, falling off a tree, bird shorting, etc.

1.2 THE PROBLEMS OF THE TRADITIONAL POWER SYSTEM

In traditional power systems, the stable operation of the power grid almost entirely relies on reports provided by personnel. When unexpected faults occur in the power system, this method cannot efficiently and quickly repair the faults, nor does it provide analysis and summary of the accident causes. Additionally, it fails to accurately locate the fault positions. Therefore, it can be seen that there are problems in the traditional operation of the power system identified as:

I. The low perception rate of faults;
II. The lack of correct understanding of local distribution network faults;
III. Slow location analysis and a lack of comprehensive analysis of related power outages and faults;
IV. Ultimately, low maintenance efficiency, difficulty in locating faults, slow repair progress, and high human resource consumption.

1.3 LITERATURE REVIEW

Bikash et. al [7] used Wavelet Packet Entropy (WPE) and a Radial Basis Function Neural Network (RBFNN) for the detection, classification, and estimation of fault location at any random position on a transmission line for both types of low and high fault impedances, RBFNN's output layer outputs for classification and estimation of fault location, while its input layer has 12 inputs. The activation function has been proposed to be the Gaussian radial basis function. About 98% of faults were correctly classified. However, one of the single line(AG) faults had an accuracy of roughly 93%, and for other faults as well, the accuracy varied from fault to fault.

P. Balakrishnan et. al [6] introduced a DWT-based algorithm for overhead lines. DWT was employed with "db6" as the mother wavelet. A ground threshold value served as the basis for the classification procedure. Issues can be discovered by receiving fault information along the entire transmission system, from the regional terminal end to the initial terminal end. All eleven categories of TL faults were detected, classified, and located using DWT to identify the signal, extract the detail coefficient, and then locate the faults. The threshold value used for classification varies with different systems.

Shahriar et. al 8] offered an unsupervised framework based on a Capsule Network (CN) for identifying and categorizing TL defects. It was done using a sparse filtering extension to CN. By actively learning the critical defect characteristics, the capsule network with sparse filtering (CNSF) improves model performance without needing a substantial amount of information. The proposed technique acquires cycle post-fault three-phase data and decodes it into a single image, which is the feed of the considered CNSF model. Four distinct topologies were used to support the proposed CNSF model's efficacy. However, it was not consistent in the analysis because of the diversity in transmission line topologies, system parameters, and operating conditions.

Daniel et.al [9] studied a fault selection system for double-circuit transmission lines using various learning techniques. The suggested method preprocesses the transmission line's raw data using the Discrete Fourier Transform (DFT) before feeding it to the learning algorithm, which uses a training period to identify and categorize any faults. Then, using simulations, the effectiveness of various machine learning algorithms was numerically compared. In the comparison, an accuracy of 98.47% was found to be achieved by an artificial neural network

(ANN). The ANN method's shortcomings include its inability to produce results that can be explained and its lack of robustness to noisy measurements.

Pathomthat et. al [10] analyzed faults in a transmission line and high-voltage capacitor banks using DWT. The findings showed that when compared to failures occurring in a capacitor bank, the features of system parameters in the event of transmission line faults are distinct. DWT was also used to resolve the disagreement between system characteristics in cases when failures occurred in both a single capacitor bank and two capacitor banks linked in a back-to-back topology. However, the method cannot be applied to complex networks.

Nguyen et. al [11] created a hybrid approach based on machine learning (ML) techniques to recognize, categorize, and find electrical defects on transmission lines. First, characteristics from the current or voltage signals were extracted using the WT approach. Eleven coefficients were created by decomposing the extracted signals. The data of various fault kinds were transformed into an RGB image, and these coefficients were calculated according to the energy level. Second, the fault is classified using the Google Net model, and the fault's location is suggested using the Convolutional Neural Network (CNN) method.

Yann Qi Chen et. al [3] used a Summation-Wavelet Extreme Learning Machine (SW-ELM) is an ML method that incorporates feature extraction in the learning process to offer an integrated framework integrating fault classification and location. Additionally, the summation-Gaussian extreme learning machine (SG-ELM), which was proposed and successfully applied to transmission line fault diagnosis, was developed as an extension of the SW-ELM. Due to its comprehensive self-learning capabilities and lack of ad hoc feature extraction requirements, SG-ELM may be deployed with the least amount of expert subjectivity. However, it was unaffected by changes in the fault inception angle.

Mou Fa Guo et. al [5] used the HHT band-pass filter to create the time-frequency energy matrix from the recorded fault waveform. For fault classification, a CNN-based technique for image similarity identification is utilized. The nonstationary and nonlinear signal can be analyzed using the HHT. The Hilbert transform and the empirical mode decomposition (EMD) are its two component phases. The original fault voltage and current signals are divided into a number of Intrinsic Mode Functions (IMF) using the EMD. Then, each IMF is subjected to the Hilbert transform, yielding the time-frequency plot of the fault signals. However, accuracy is poor for B-G faults in noisy environments. Additionally, the distribution generator access affects the accuracy rate for two-phase short circuits.

Fezan et. al [12] suggested a technique that uses Long Short Term Memory (LSTM) units acting directly on operational information rather than characteristics. The method employs the temporal sequence of the operational information of the power network to build an "end-to-end" model. End-to-end learning eliminates the requirement for time-consuming feature extraction by learning directly from the labeled datasets. This quickens decision-making.

Ji Han et. al [14] suggested a unique diagnosis model for power systems reduce the need for constant model modification effort when the system topology changes. The gradient similarities among the multichannel electrical signals were first converted to the visible similarity pictures, which were then given to the neural network. This data preprocessing method uses gradient computation and similarity evaluation. Then the CNN used Spatial Pyramid Pooling (SPP) and Hashing Classifier (HC). Even when there are topological changes in the power systems, the fault-diagnosis model's structure can be kept constant with the help of the SPP and HC approaches. But the accuracy seems to drop with noise.

Yanhai Wang et.al [15] used a quality-aware fine-grained-based image classification for transmission line fault detection. This method was used to detect the fault zone. The technique uses wavelet-based support vector machines and quality-based discriminative feature extraction to extract the characteristics of line currents by leveraging Fast R-CNN-based image samples decomposition, where the quality module is used to select the most discriminative regions. The retrieved features are then used to train an SVM to identify the issue. The suggested method didn't work well for the L-G fault.

Arash et. al [16] used Frequency Response Analysis (FRA) to assess the effects of impedance and to identify the fault location when a fault occurs. Since the interpretation of FRA is weak, Convolutional Long short-term Memory was proposed to extract the features of frequency response curves for each fault.

Yanhui Xi et. al [17] developed a fault classification method based on SA-MobileNetV3. The method is similar to image recognition. The three-phase current and voltage signals are transformed into two-dimensional images based on Continuous Wavelet Transform (CWT). Then the proposed method is used in classifying the faults.

Praveen et. al. [13] customized CNN for fault classification in the distribution networks with DGs. The developed model doesn't need any preprocessing, which makes it efficient during the testing period.

1.4 BENEFITS OF AI-BASED FAULT DETECTION

Recent developments have involved combining the relative powers of AI techniques to solve power system problems. Because of the nature of various types of power system problems different types of solutions may be required. The real-world power system problems may neither fit the assumption of a single AI technique nor be effectively solved by the strengths and capabilities of a single AI technique. By leveraging AI algorithms and machine learning techniques, historical data and patterns can be analyzed to forecast potential faults before they occur. While fault detection is crucial, fault prediction takes it a step further by enabling proactive measures to prevent failures.

Fault prediction provides the following benefits:

1. Minimizing Downtime: Predicting faults allows for scheduled maintenance and repairs, minimizing unplanned downtime and its associated costs. By identifying emerging issues in advance, necessary actions can be taken to prevent system failures, ensuring continuous operation

2. Optimal Resource Allocation: Fault prediction enables effective resource planning. By anticipating potential faults, maintenance personnel, spare parts, and tools can be allocated efficiently, reducing unnecessary expenses and optimizing resource utilization.

3. Cost Reduction: Early fault prediction helps in reducing maintenance costs. Rather than performing routine or reactive maintenance, resources can be directed specifically toward the areas that are most likely to experience faults. This targeted approach saves time, effort, and expenses associated with unnecessary inspections and repairs.

4. Enhanced Safety: Fault prediction contributes to enhanced safety in technical systems. By identifying potential risks and faults in advance, appropriate measures can be taken to mitigate these risks, ensuring the safety of personnel and minimizing the chances of accidents or hazardous situations

5. Improved Performance: Proactively addressing faults based on prediction models improves system performance. By preventing failures, the system can operate optimally, meeting operational requirements and delivering consistent results.

6. Increased Reliability: AI-driven fault prediction helps organizations maintain high-quality operations, leading to enhanced reliability and customer satisfaction.

7. Data-Driven Decision-Making: AI systems analyze large volumes of data to provide actionable insights, enabling informed decision-making and long-term planning.

1.5 CHALLENGES AND ISSUES

Some of these challenges and issues associated with fault detection and prediction are:

1. Data quality: AI algorithms rely heavily on data to learn and make predictions. It is important to ensure that the data used is accurate, complete, and representative of the system being monitored.

2. Model complexity: AI models can be very complex, which can make it difficult to understand how they are making predictions. This can be a challenge when trying to diagnose faults or understand why a particular prediction was made.

3. Model training: AI models require training to learn from data. This can be a time-consuming and resource-intensive process, particularly if the data is complex or the model is large.

4. Model validation: Once an AI model has been trained, it is important to validate its performance on new data. This can be challenging, as it requires a large amount of data that is representative of the system being monitored.

5. Interpretability: AI models can be difficult to interpret, which can make it challenging to understand why a particular prediction was made. This can be a problem when trying to diagnose faults or understand how the system is behaving.

6. False positives and false negatives: AI models can produce false positives (predicting a fault when there is none) or false negatives (failing to predict a fault when there is one). This can be a problem if it leads to unnecessary maintenance or missed faults.

Therefore, this literature review highlights machine learning-based of Fault detection (FD) and fault classification (FC) methods. In the work, Random Forest and Naïve Bayes have been proposed for training to identify various types of faults and provide insights into fault patterns and, relationships.

1.6. PROJECT OBJECTIVES

1. To develop an AI-based model for fault detection in the transmission line
2. To compare the efficacy between the selected Random Forest and Naive Bayes model.
3. To observe and analyze different kinds of faults observed from the proposed algorithm.

CHAPTER-2

MACHINE LEARNING FOR FAULT DETECTION

Deriving from the importance of determining the fault, we illustrate an artificial intelligence method to analyze and solve the problem. Recent studies have demonstrated the effectiveness of artificial intelligence in many fields, including but not limited to marketing, banking, power systems, health care, and so forth. Among the well-known methods in artificial intelligence is machine learning.

Machine Learning:

Machine learning refers to the use of artificial intelligence that offers systems the capacity to robotically learn and advance from experience devoid of being overtly programmed. More specifically, machine learning focuses on the advancement of computer programs that can obtain data and use these data to learn in a self-reliant way. The aim of machine learning is to comprehend the structure of the data and use them to construct models that can be comprehended and used by humans. While machine learning is a subdivision of computer science, it differs from conventional computational strategies. In conventional computing, a programmer sets specific algorithms of clearly programmed instructions used by computers to solve a problem. Instead, machine learning has algorithms that permit computers to learn from data inputs and to use statistical analysis to produce values within a particular range. For that reason, machine learning enables computers to develop models from sample data and to make decisions based on the obtained data inputs. In the present-day world, machine learning has many and varied practical applications. For instance, machine learning is applied in the facial recognition technology used in social media sites to assist users in tagging themselves and their friends on 45 photos. Moreover, the optical character recognition (OCR) system enables the conversion of text images into movable types. Furthermore, machine learning technology is also used in the navigation of self-driving automobiles to navigate the roads. In fact, due to the changes that require higher efficacy and efficiency in manufacturing, custom execution of algorithms is normally needed for production systems. Firms are continually looking for systems that are faster, better, and require less effort to operate and have lower costs of production. Using machine learning tools helps businesses achieve higher revenue. Importantly, executing algorithms helps to enhance the skills necessary to find these solutions.

Machine learning technology is also used in electrical power systems–more specifically, in power transmission, generation, and maintenance. Accordingly, power firms widely use the statistical and discovery methods of machine learning for pre-emptive maintenance. Among other applications, machine learning systems and methods are used to convert historical data from electrical data into predictive models. Furthermore, machine learning can be used to generate transformer rankings, cable, and feeder failure rankings, as well as to compute the mean time between failure estimations. Machine learning also has interfaces for business management that allow for a direct incorporation of prediction ability into decision support and corporate planning. Machine learning is also beneficial in the maintenance operations of power companies. Interestingly, it assists in fixing a problem proactively, instead of fixing an issue when it has already occurred. Said differently, machine learning makes it possible to prevent failures, rather than to cope with their consequences, such as cascading failures, fires, and expensive emergency repairs. A major requirement for a machine learning algorithm is data analysis. In fact, data analysis is the prerequisite for beginning a machine learning algorithm. Data analysis is a process of data collection, cleaning, aggregating, visualizing, and exploring. All these processes help in making appropriate predictions and acquiring data from flat files, spreadsheets, and databases, conducting exploratory data analysis (EDA), data reshaping, and data visualization. Furthermore, data exploration involves pursuing correlations, determining the missing content, and visualizing. The building of models also includes visualization of the results, development of model diagnostics, and residual diagnostics. The machine learning algorithms can use the models to predict the future. Machine learning algorithms also require an understanding of Python codes and R codes and how to operate them. To this end, Panda's library, which is useful for reshaping and aggregating the data, and Matplotlib library, which is important for data visualization, are frequently used. Similarly, the Seaborn library can be used for advanced analytical processes. Several basic data visualization techniques include bar charts, histograms, heat maps, and scatterplots. At this stage, the selection of the algorithm is implemented. A researcher should be specific in the selection of the type and class of algorithm, as well as in the description of the system to execute. The next step of selecting a set of problems to validate and test the execution of the algorithm is available. Finally, the results of the performance of the built algorithm are evaluated based on several parameters, such as precision, F1-score, and recall for further detail.

2.1 RANDOM FOREST CLASSIFIER

Random Forest is an ensemble learning method for classification, regression, and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the Random Forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random Forest is used on the job by data scientists in many industries including banking, stock trading, medicine, and e-commerce. It’s used to predict the things that help these industries run efficiently, such as customer activity, patient history, and safety Random Forest is used in banking to detect customers who are more likely to repay their debt on time. It’s also used to predict who will use a bank’s services more frequently. They even use it to detect fraud.

Stock traders use Random Forests to predict a stock’s future behavior. It’s used by retail companies to recommend products and predict customer satisfaction as well. Scientists in China used Random Forest to study the spontaneous combustion patterns of coal to reduce safety risks in coal mines. In healthcare, Random Forest can be used to analyze a patient’s medical history to identify diseases. Pharmaceutical scientists use Random Forest to identify the correct combination of components in a medication or predict drug sensitivity. Sometimes Random Forest is even used for computational biology and the study of genetics.

The most convenient benefit of using Random Forest is its default ability to correct decision trees’ habit of overfitting to their training set. Using the bagging method and random feature selection when executing this algorithm almost completely resolves the problem of overfitting which is great because overfitting leads to inaccurate outcomes. Plus, even if some data is missing, Random Forest usually maintains its accuracy. Random Forest is much more efficient than a single decision tree while performing analysis on a large database. On the other hand, a Random Forest is less efficient than a neural network. A neural network, sometimes just called a neural net, is a series of algorithms that reveal the underlying relationship within a dataset by mimicking the way that a human brain thinks. A neural network is more complicated than Random Forests but generates the best possible results by adapting to changing inputs. Unlike neural nets, Random Forest is set up in a way that allows for quick development with minimal hyper-parameters (high-level architectural guidelines), which makes for less set-up time.

There aren’t many downsides to Random Forest, but every tool has its flaws. Because Random Forest uses many decision trees, it can require a lot of memory on larger projects. This can make it slower than some other, more efficient algorithms. Sometimes, because this is a decision tree-based method and decision trees often suffer from overfitting, this problem can affect the overall forest. This problem is usually prevented by Random Forest by default because it uses random subsets of the features and builds smaller trees with those subsets. This can slow down processing speed but increase accuracy.

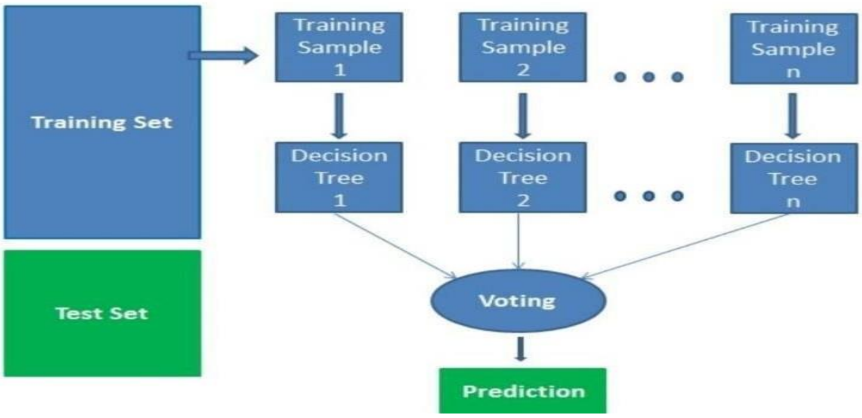


Fig:1. Random Forest Classifier

2.1.1 ADVANTAGES OF RANDOM FOREST CLASSIFIER:

- Reduced risk of overfitting: Decision trees run the risk of overfitting as they tend to tightly fit all the samples within training data. However, when there’s a robust number of decision trees in a Random Forest, the classifier won’t overfit the model since the averaging of uncorrelated trees lowers the overall variance and prediction error.
- Provides flexibility: Since Random Forest can handle both regression and classification tasks with a high degree of accuracy, it is a popular method among data scientists. Feature bagging also makes the Random Forest classifier an effective tool for estimating missing values as it maintains accuracy when a portion of the data is missing.
- Easy to determine feature importance: Random Forest makes it easy to evaluate variable importance, or contribution, to the model. There are a few ways to evaluate feature importance. Gini importance and mean decrease in impurity (MDI) are usually used to measure how much the model’s accuracy decreases when a given variable is excluded. However, permutation importance, also known as mean decrease accuracy (MDA), is

another important measure. MDA identifies the average decrease in accuracy by randomly permutating the feature values in samples.

2.1.2 DISADVANTAGES OF RANDOM FOREST CLASSIFIER:

- Time-consuming process: Since Random Forest algorithms can handle large data sets, they can provide more accurate predictions, but can be slow to process data as they are computing data for each individual decision tree.
- Requires more resources: Since Random Forests process larger data sets, they'll require more resources to store that data.
- More complex: The prediction of a single decision tree is easier to interpret when compared to a forest of them.

2.2 NAIVE BAYES MODEL:

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.To start with, let us consider a dataset. One of the most simple and effective classification algorithms, the Naïve Bayes classifier aids in the rapid development of machine learning models with rapid prediction capabilities.

Assumption of Naive Bayes:

The fundamental Naive Bayes assumption is that each feature makes an:

- Feature independence: The features of the data are conditionally independent of each other, given the class label.
- Continuous features are normally distributed: If a feature is continuous, then it is assumed to be normally distributed within each class.
- Discrete features have multinomial distributions: If a feature is discrete, then it is assumed to have a multinomial distribution within each class.
- Features are equally important: All features are assumed to contribute equally to the prediction of the class label.
- No missing data: The data should not contain any missing values.

2.2.1 TYPES OF NAIVE BAYES MODEL:

There are three types of Naive Bayes Model:

a. Gaussian Naive Bayes:

In Gaussian Naive Bayes, continuous values associated with each feature are assumed to be distributed according to a Gaussian distribution. A Gaussian distribution is also called a Normal distribution When plotted, it gives a bell-shaped curve that is symmetric about the mean of the feature values.

b. Multinominal Naive Bayes:

Feature vectors represent the frequencies with which certain events have been generated by a multinomial distribution. This is the event model typically used for document classification.

c. Bernoullie Naive Bayes:

In the multivariate Bernoulli event model, features are independent Booleans (binary variables) describing inputs. Like the multinomial model, this model is popular for document classification tasks, where binary term occurrence (i.e. a word occurs in a document or not) features are used rather than term frequencies (i.e. frequency of a word in the document).

2.2.2 ADVANTAGES OF NAIVE BAYES CLASSIFIER

- Easy to implement and computationally efficient.
- Effective in cases with a large number of features.
- Performs well even with limited training data.

2.2.3 DISADVANTAGES OF NAIVE BAYES CLASSIFIER

- Assumes that features are independent, which may not always hold in real-world data.
- Can be influenced by irrelevant attributes.
- May assign zero probability to unseen events, leading to poor generalization.

CHAPTER-3

METHODOLOGY

3.1 PROCESS

Data readily available from platforms like Kaggle involves choosing appropriate data from past substation records to predict various types of faults. This dataset comprises information on three-phase current and voltage. Subsequently, the dataset is split into two subsets: one for training the model and another for testing its performance. Typically, the majority of the data is allocated for training, with a smaller portion reserved for testing. The trained model's outcomes are then utilized to predict faults using an AI-driven Random Forest model.

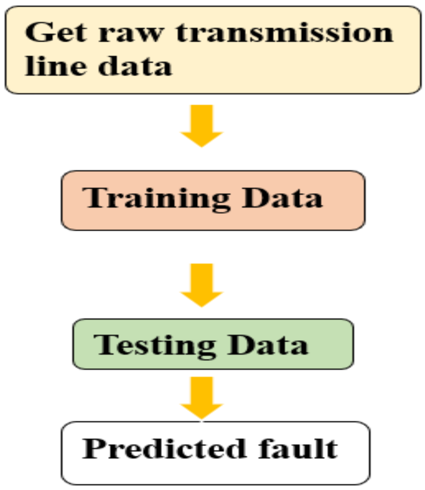


Fig:2 System Flow Diagram

3.2 EXPERIMENTATION

The outputs comprise four variables A, B, C, and G. Of these, a value close to unity for any of the first three variables corresponds to the appropriate a, b, or c phases being faulty and a near unity value of G signifies that ground is involved in a fault. Example:

Inputs - [Ia, Ib, Ic, Va, Vb, Vc]

Ia = Current in line A

Ib = Current in line B

Ic = Current in line C

Va = Voltage in line A

Vb = Voltage in line B

Vc = Voltage in line C

- [G C B A] - Outputs
- [0 0 0 0] - No-Fault
- [1 0 0 0] - Ground Fault
- [0 0 0 1] - Fault in Line A
- [0 0 1 0] - Fault in Line B
- [0 1 0 0] - Fault in Line C
- [1 0 0 1] - LG fault (Between Phase A and Ground)
- [1 0 1 0] - LG fault (Between Phase B and Ground)
- [1 1 0 0] - LG fault (Between Phase C and Ground)
- [0 0 1 1] - LL fault (Between Phase B and Phase A)
- [0 1 1 0] - LL fault (Between Phase C and Phase B)
- [0 1 0 1] - LL fault (Between Phase C and Phase A)
- [1 1 0 0] - LG fault (Between Phase C and Ground)
- [1 0 1 1] - LG fault (Between Phase B, C and Ground)
- [1 1 0 1] - LG fault (Between Phase A, C and Ground)
- [1 1 0 1] - LLG Fault (Between Phases A, B and Ground)
- [0 1 1 1] - L Fault (Between all three phases)
- [1 1 1 1] - LLLG fault (Three phase symmetrical fault

3.3 STEPS PERFORMED DURING EXECUTION:

- Step 1: Import the Libraries
- Step 2: Getting to Visualizing the line fault prediction data
- Step 3: Assigning proper labels to each fault type based on the combination values of the values of G, C, B, A

  0000 = No fault

  1001 = Line A to ground fault

  0110= Line B to line C fault

  1011= Line A line B to ground fault

  0111= Line A line B line C fault

  1111= Line A line B line C to ground fault

- Step 4: Each label is mapped to a number.

  E.g. Line A line B line C fault is represented by :0

  Line A line B line C to ground fault :1

  Line A line B to ground fault:2

  Line A to ground fault :3

  Line B to line C fault: 4

  No fault:5

- Step 5: Set two variables X and Y. Eg X for input values (Va, Vb, Vc, Ia, Ib, Ic) and Y for the target variable.
- Step 6: Creating a Training and a Test dataset for fault Prediction in the transmission line.
- Step 7: Construct the Random Forest model for fault prediction.
- Step 8: Construct the Naive Bayes model for fault prediction.
- Step 9: Comparing the accuracy of these two models.
- Step 10: Final fault prediction using Random Forest algorithm.

3.4 DATASET USED FOR FAULT DETECTION

The power department provided the raw, real-time data. There are some examples of 5 sets of three-phase voltage and current parameters which are input for the trained model to obtain corresponding output results as demonstrated in Table 1.

	G	C	B	A	Ia	Ib	Ic	Va	Vb	Vc
0	1	0	0	1	-151.291812	-9.677452	85.800162	0.400750	-0.132935	-0.267815
1	1	0	0	1	-336.186183	-76.283262	18.328897	0.312732	-0.123633	-0.189099
2	1	0	0	1	-502.891583	-174.648023	-80.924663	0.265728	-0.114301	-0.151428
3	1	0	0	1	-593.941905	-217.703359	-124.891924	0.235511	-0.104940	-0.130570
4	1	0	0	1	-643.663617	-224.159427	-132.282815	0.209537	-0.095554	-0.113983

Table 1: Data for various kinds of fault

CHAPTER-4

RESULTS AND DISCUSSION

To identify faults, training, and testing have been carried out in python with data sets for various fault levels. In the Random Forest model, training accuracy is 85.7 % and Naïve Bayes model accuracy is 80.61 %. The results indicate that Random Forest gives a higher level of accuracy. Table 2 & Table 3 show the accuracy of the proposed algorithm at different levels of fault. The results indicate a high accuracy of more than 80 % using the Random Forest model. Therefore, in the present work, the Random Forest model has been used for fault detection and classification in transmission lines.

4.1 RESULT WITH RANDOM FOREST METHOD:

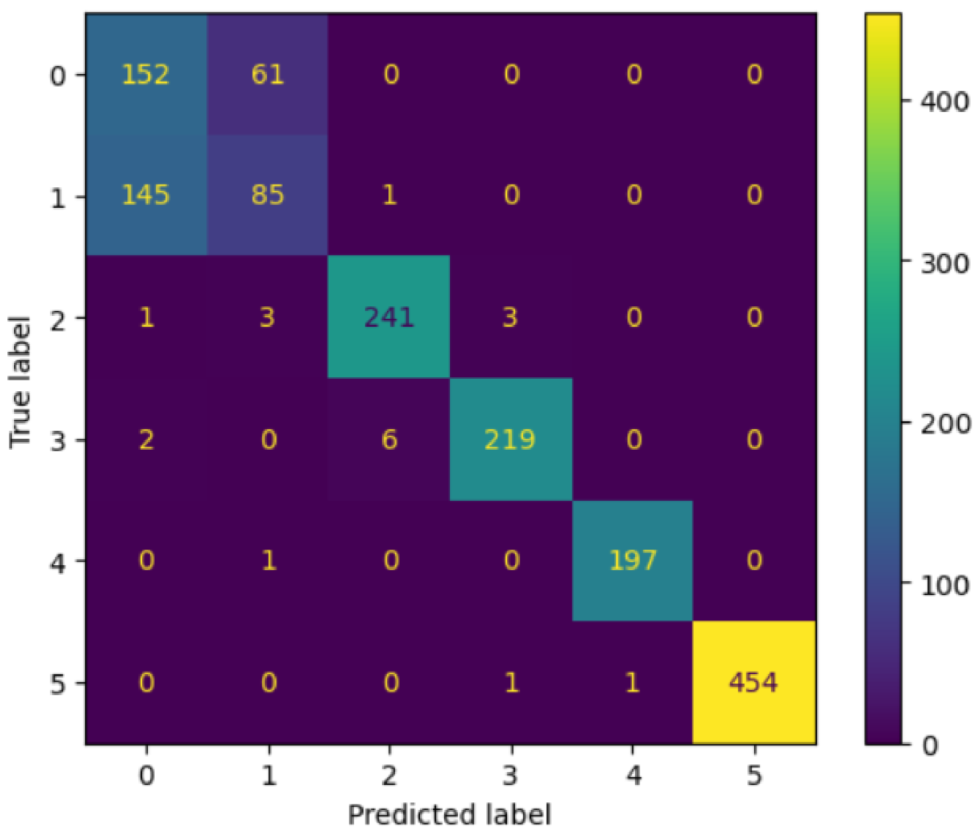


Fig:3. Confusion matrix for Random Forest model

Classification_Report:

	precision	recall	f1-score	support
0	0.51	0.71	0.59	213
1	0.57	0.37	0.45	231
2	0.97	0.97	0.97	248
3	0.98	0.96	0.97	227
4	0.99	0.99	0.99	198
5	1.00	1.00	1.00	456
accuracy			0.86	1573
macro avg	0.84	0.83	0.83	1573
weighted avg	0.86	0.86	0.85	1573

Table 2: Classification report for Random Forest model

4 .2 RESULT WITH NAIVE BAYES METHOD:

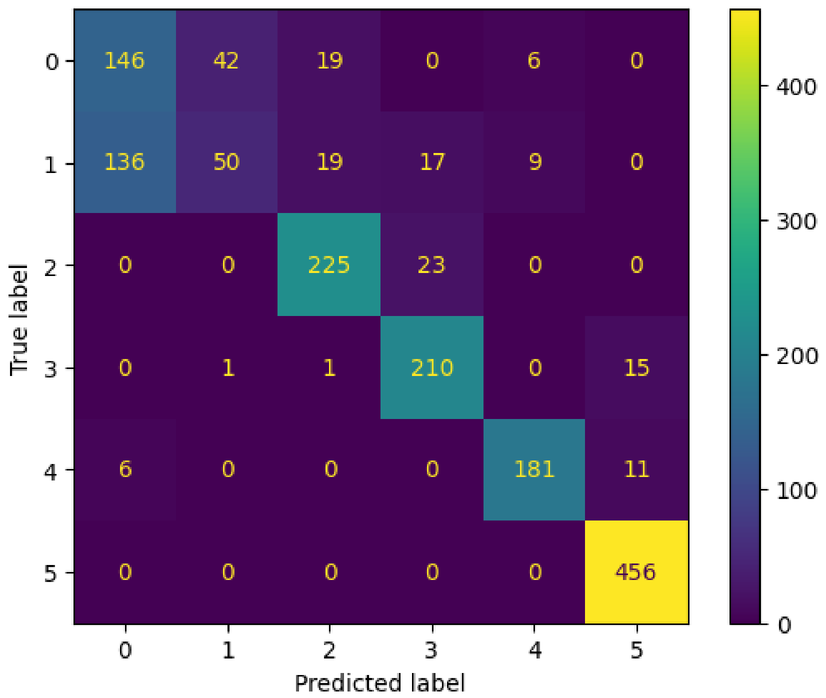


Fig:4. Confusion matrix for Naive Bayes model

Classification_Report:

	precision	recall	f1-score	support
0	0.51	0.69	0.58	213
1	0.54	0.22	0.31	231
2	0.85	0.91	0.88	248
3	0.84	0.93	0.88	227
4	0.92	0.91	0.92	198
5	0.95	1.00	0.97	456
accuracy			0.81	1573
macro avg	0.77	0.77	0.76	1573
weighted avg	0.79	0.81	0.79	1573

Table 3. Classification report for Naive Bayes model

4.3 RESULT OUTPUT OF VOLTAGE AND CURRENT DURING VARIOUS FAULT

Figures (5-18) show an example of the current and voltage waveforms when the system experiences various fault connections, where blue, red, and green lines represent each of the phases. This shows that during a fault event, the current is very large, and the voltage becomes smaller for a short amount of time. The fault categorization technique is separated into three steps. First, the signal analysis method is used to process the voltage and current defect signals. Second, the properties of the processed signals are retrieved. The trained model receives the characteristics, and the classification outcomes are then obtained. To analyze voltage and current fault signals, the Random Forest algorithm is used.

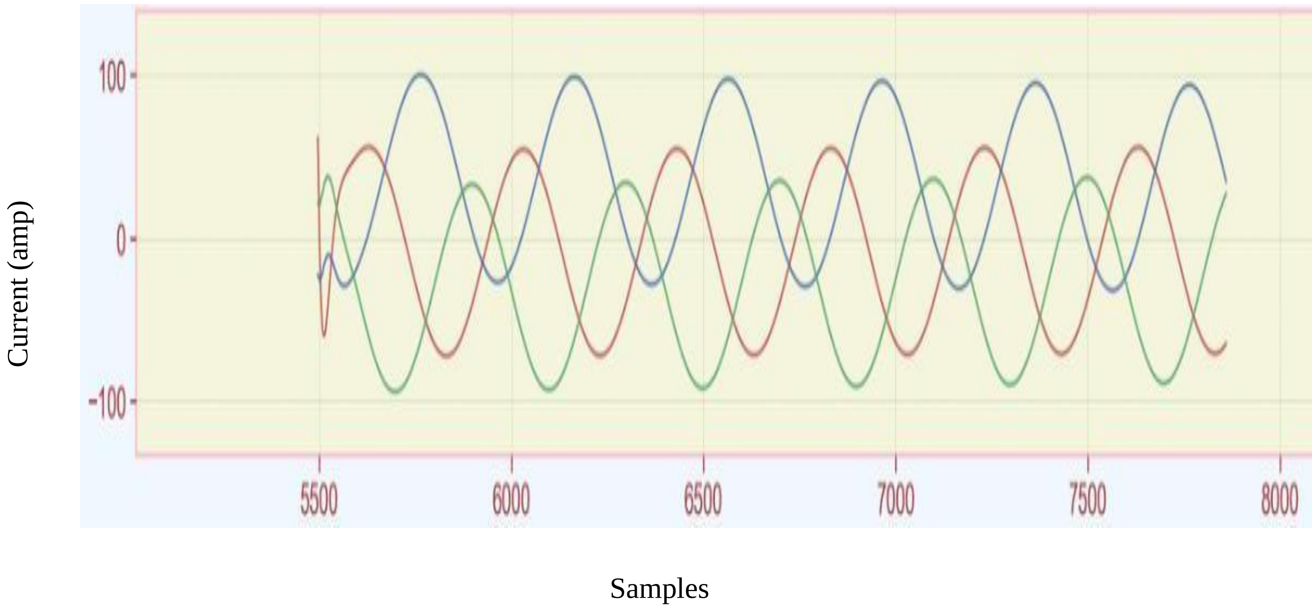


Fig.5: Current waveform when fault has not occurred on either phase A, B, or C.

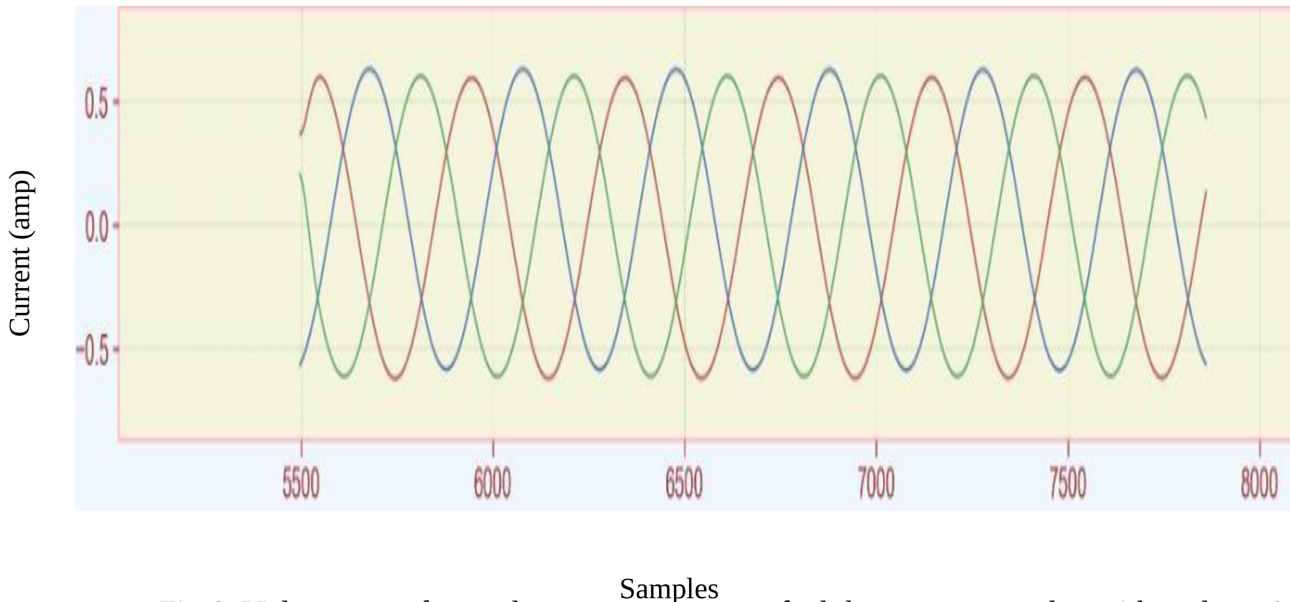
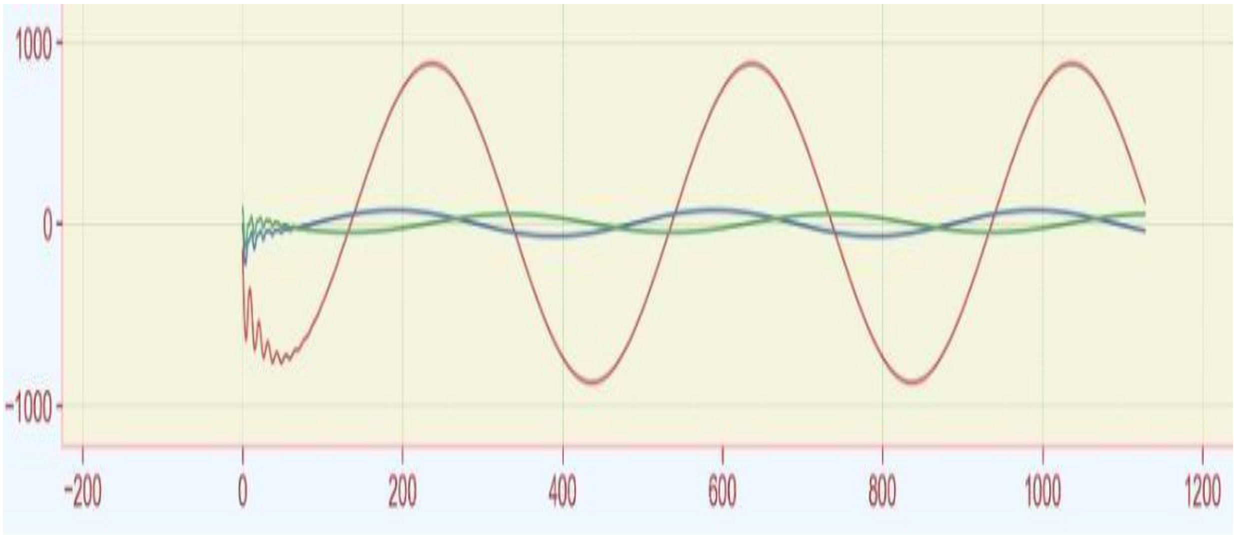


Fig.6: Voltage waveform when fault has not occurred on either phase A, B, or C.

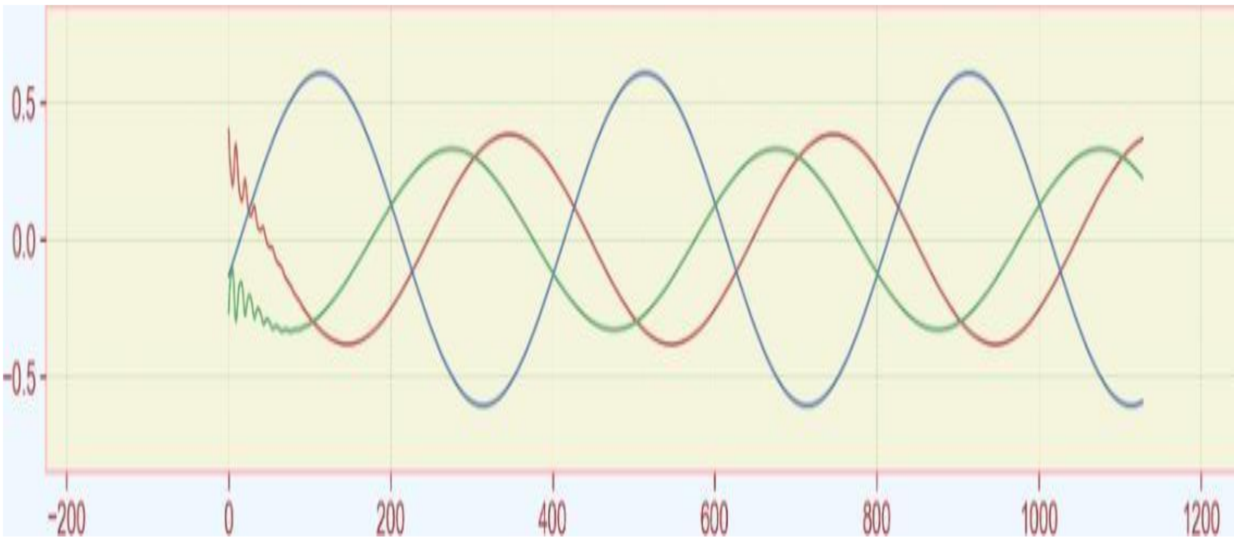
Figure (5) and Figure (6) show the waveform of voltage and current when there is no fault on the transmission line. This indicates that fault has not occurred on either phase of the transmission line. The magnitude of the voltage was maintained at approximately 0.5 p.u. The magnitude of the current was maintained at approximately 100 amp.



Current (amp)

Samples

Fig.7: Line to ground fault current



Current (amp)

Samples

Fig.8: Line to ground fault voltage

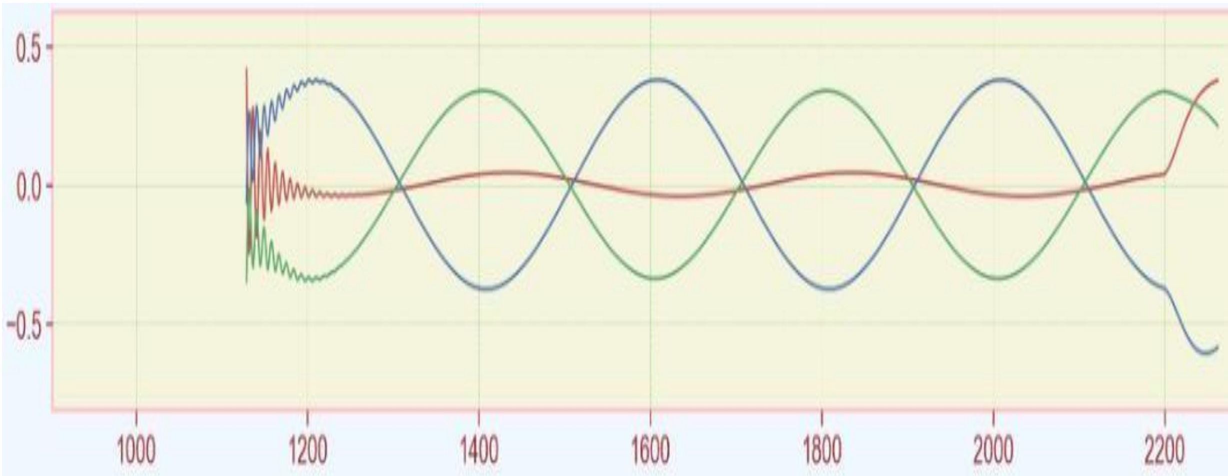
Figure (7) and Figure (8) show the waveform of current and voltage when a fault occurs between Phase A and the ground. During this period of fault occurrence, there was a distortion in the waveform of the voltage and current, as can be seen in Fig (7) and Fig (8). The magnitude of the voltage is very small in phase A which is approximately less than 0.5 p.u. The magnitude of the current is very high in phase A which is approximately 1000 amp.

Current (amp)

Samples

Fig.9: Line-Line to ground fault current

Current (amp)

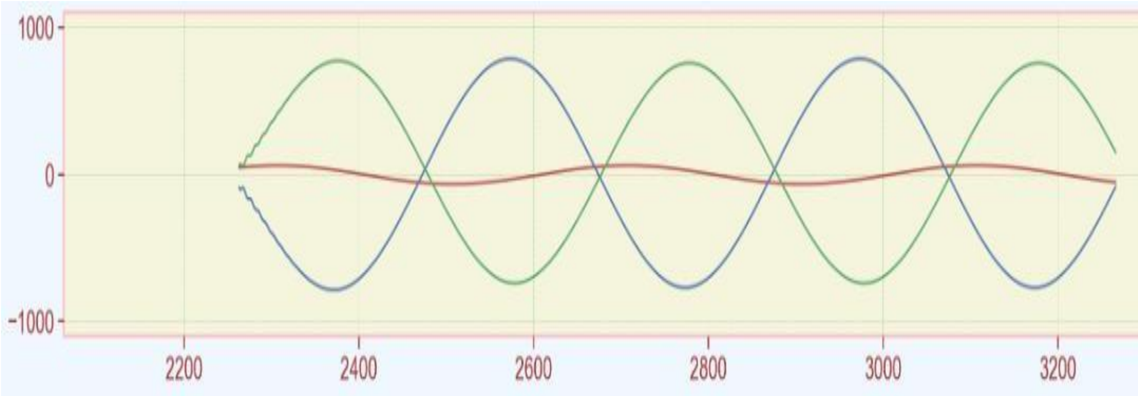


Samples

Fig.10: Line-Line to ground fault voltage

Figure (9) and Figure (10) show the waveform of current and voltage when faults occur between Phase A, phase B, and the ground. During this period of fault occurrence, there was a distortion in the waveform of the voltage and current, as can be seen in Fig (9) and Fig (10). The magnitude of the voltage is very small in phase A and phase B which is approximately less than 0.5 p.u. The magnitude of the current is very high in phase A and phase B which is approximately 1000 amp.

Current (amp)

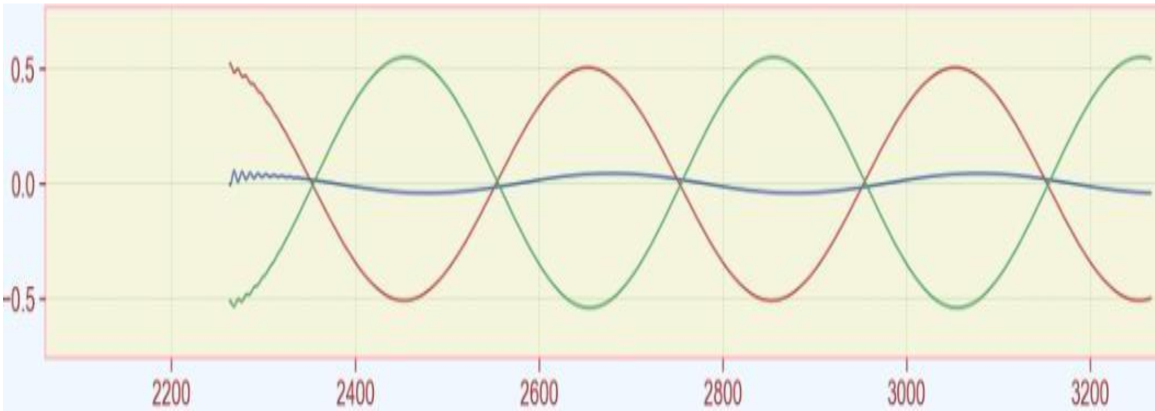


Samples

Fig.11: Line to Line fault current

Figure (11) and Figure (12) show the waveform of current and voltage when faults occur between Phase A and Phase B. During this period of fault occurrence, there was a distortion in the waveform of the voltage and current, as it can be seen in fig (11) and fig (12). The magnitude of the voltage is very small in phase A and phase B which is approximately less than 0.5 p.u. The magnitude of the current is very high in phase A and phase B which is approximately 1000 amp.

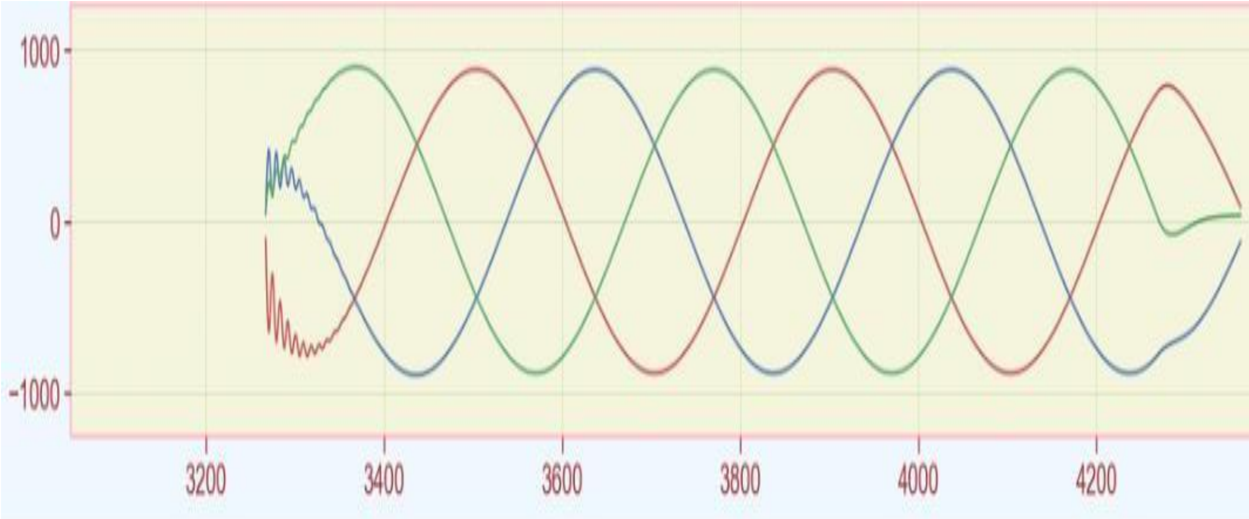
Current (amp)



Samples

Fig.12: Line to Line fault voltage

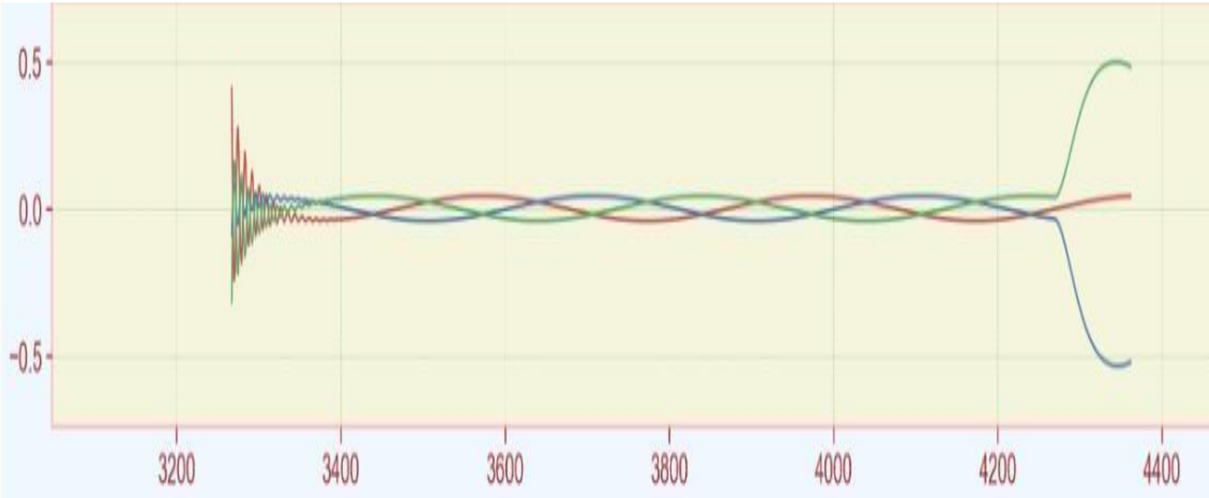
Current (amp)



Samples

Fig.13: Line-Line-Line fault current

Current (amp)



Samples

Fig.14: Line-Line-Line fault voltage

Figure (13) and Figure (14) show the waveform of current and voltage when faults occur between Phase A, phase B, and Phase C. During this period of fault occurrence, there was a distortion in the waveform of the voltage and current, as can be seen in Fig (13) and Fig (14). The magnitude of the voltage is very small in phase A, phase B, and phase C which is approximately less than 0.5 p.u. The magnitude of the current is very high in phase A and phase B which is approximately 1000 amp.

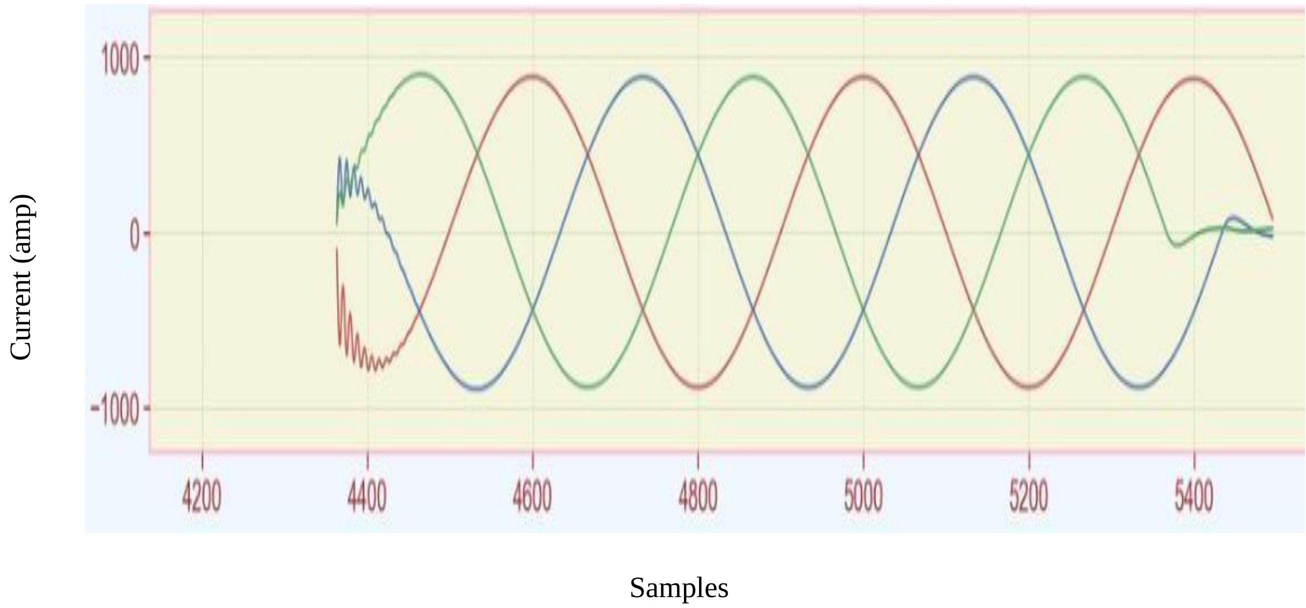


Fig.15: Line-Line-Line ground fault current

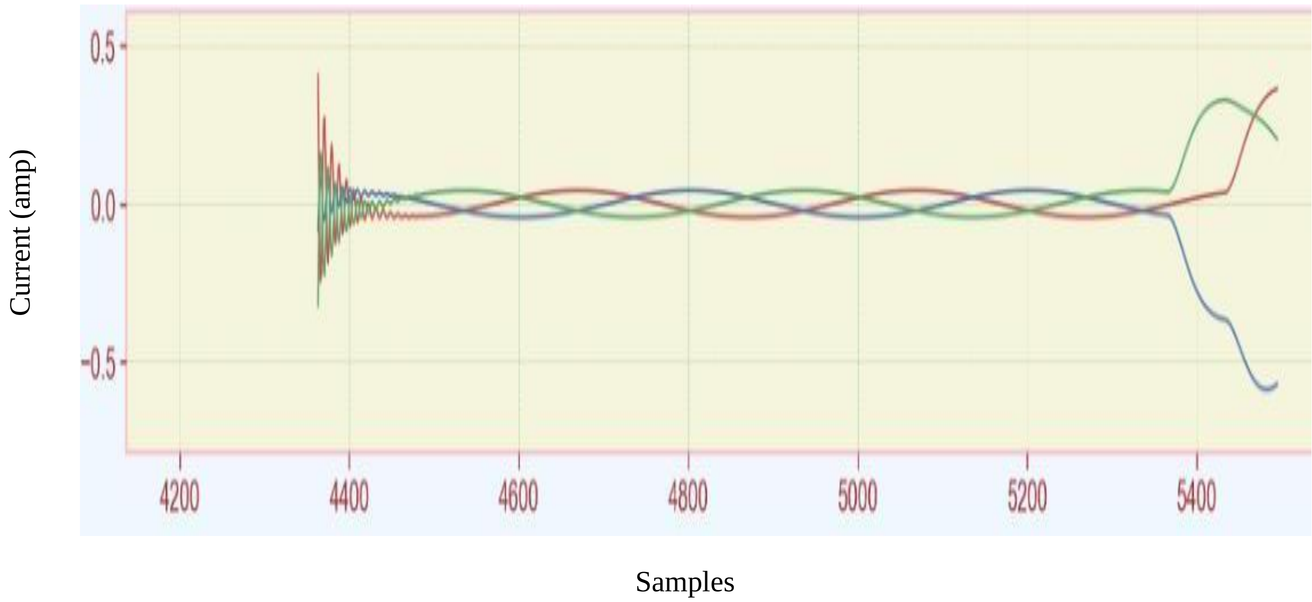


Fig.16: Line-Line-Line ground fault voltage

Figure (15) and Figure (16) show the waveform of current and voltage when faults occur between Phase A, phase B, phase C, and the ground. During this period of fault occurrence, there was a distortion in the waveform of the voltage and current, as can be seen in Fig (15) and Fig (16). The magnitude of the voltage is very small in phase A phase B and phase C which is approximately less than 0.5 p.u. The magnitude of the current is very high in phase A phase B and phase C which is approximately 1000 amp.

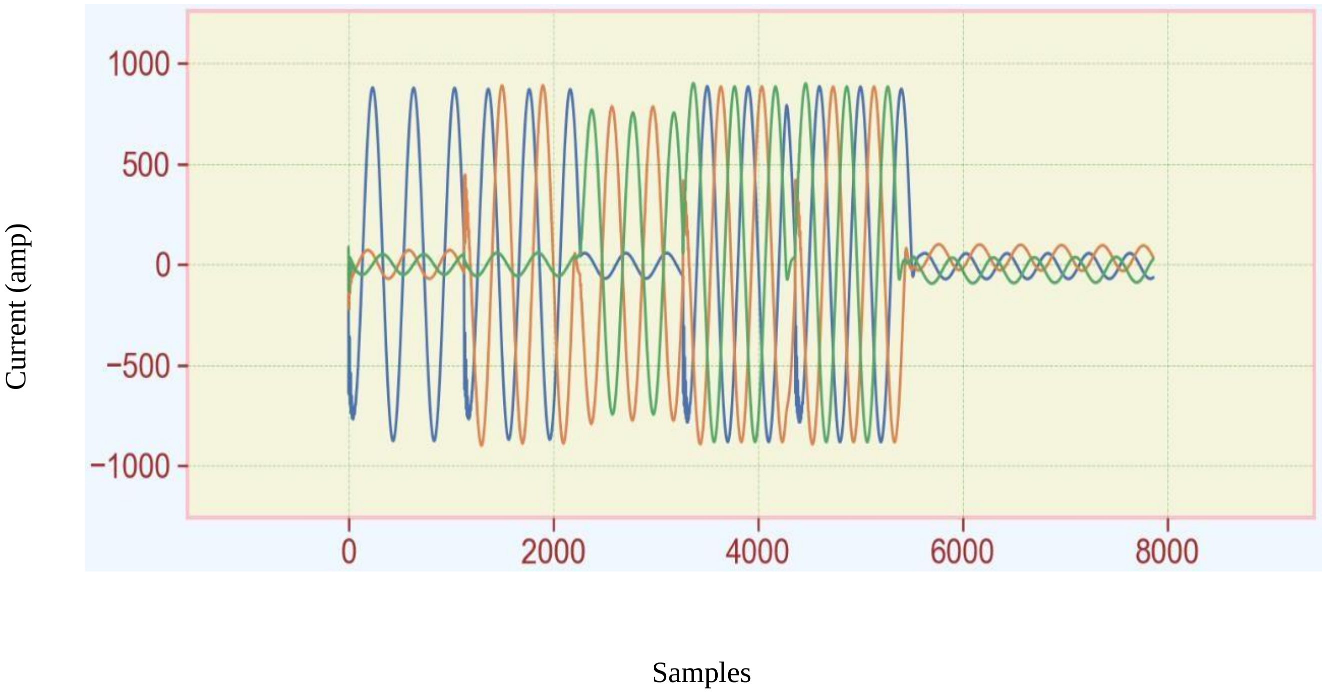


Fig.17: Total fault current

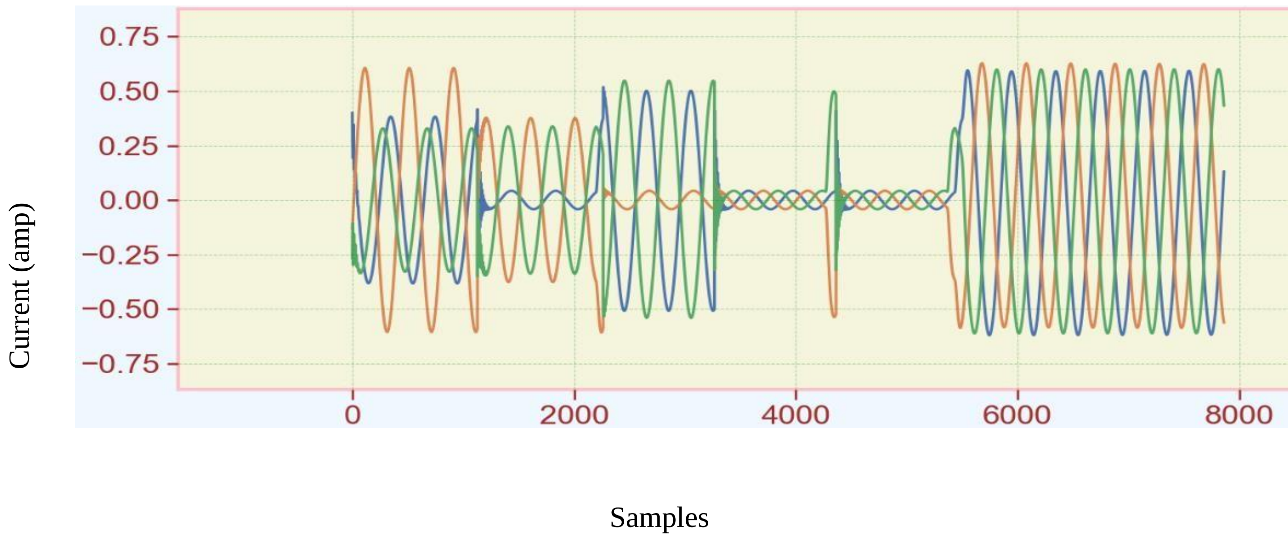


Fig.18: Total fault voltage

Figure (17) and Figure (18) show the waveform of current and voltage when various kinds of faults occur between Phases. During this period of fault occurrence, there was a distortion in the waveform of the voltage and current, as can be seen in Figure (17) and Figure (18). The magnitude of the voltage is very small and the magnitude of the current is very high in phases when a fault occurs in the transmission line.

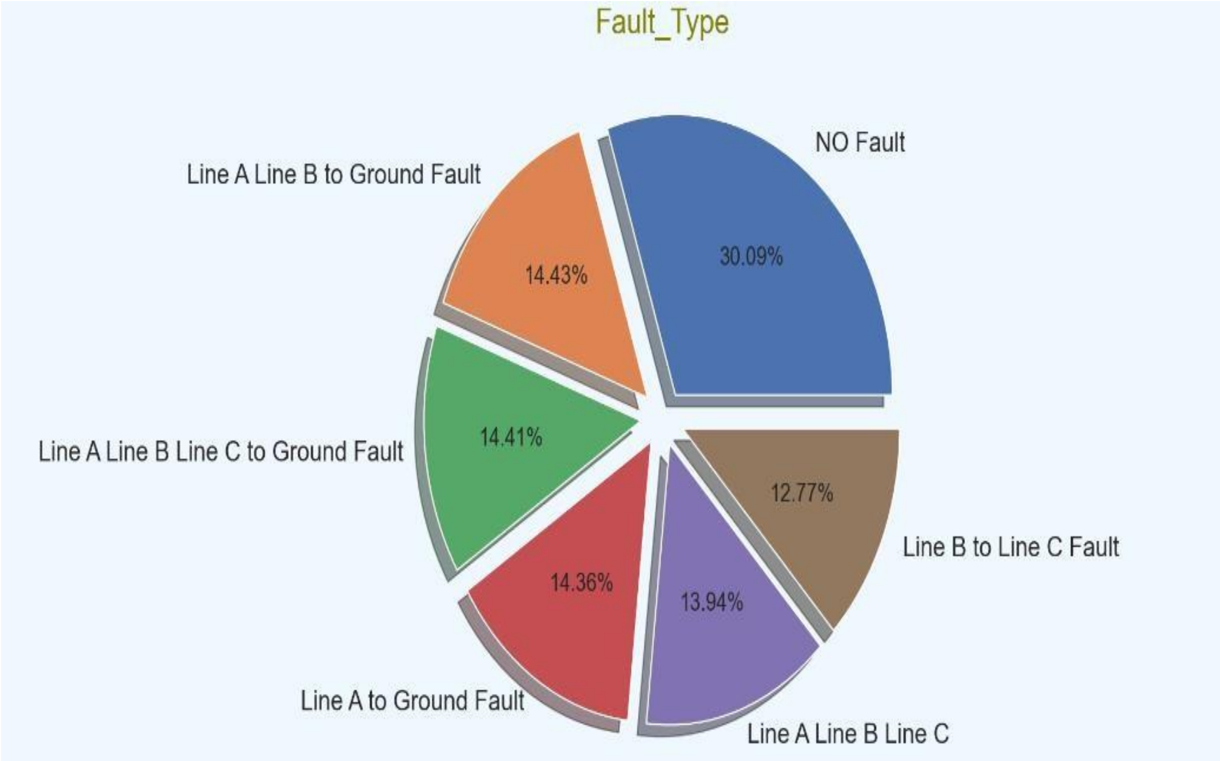


Fig.19: Pie chart for all fault

The above figure shows the pie chart demonstration of various kinds of faults in transmission lines obtained from the previous data set.

CHAPTER-5

CONCLUSION

Based on the rapidly developing artificial intelligence technology, this thesis conducts real-time analysis of substation operation data, efficiently and quickly identifying the types of faults in the distribution network. Artificial Intelligence technology can improve the efficiency of grid analysis of real-time data, and at the same time more easily determine the fault type of the distribution network compared with the traditional method.

The performance of a transmission line can be taken to a higher level. In terms of efficiency, stability, and fault analysis, our results demonstrated that the Random Forest shows a better performance than the Naive Bayes. Random Forest gives a satisfactory result with 86 percent accuracy in fault detection than the Naive Bayes model.

The waveform of different kinds of faults obtained from Random Forest algorithm was observed and is discussed in this thesis such as L-G, L-L, L-L-G, L-L-L, and L-L-L-G fault. When a fault occurs in the transmission line the magnitude of the voltage is decreased and the magnitude of the current is increased as shown in the previous waveform. During this period of fault occurrence, there is a distortion in the waveform of the voltage and current. Also, a pie chart was represented to show different fault types using Random Forest.

Additionally, a Python program was run to utilize a Random Forest classifier to detect faults. The program takes three-phase current and three-phase voltage as inputs, with the output interpreted through Python. By employing these inputs, the program demonstrates the ability to identify various fault types and classify them using a Random Forest classifier in Python.

BIBLIOGRAPHY

---

[1] M. A. Taha and M. E. El-Hawary, "Fault Detection and Classification in Power Systems Using the Wavelet Transform and Random Forest," *in IEEE Transactions on Power Delivery*, vol. 28, no. 1, pp. 52-60, Jan. 2013. doi: 10.1109/TPWRD.2012.2226196

[2] R. F. Abdel-Rahman and M. A. Abido, "A Hybrid Machine Learning Technique for Transmission Line Fault Classification," *in IEEE Transactions on Power Delivery*, vol. 28, no. 4, pp. 2401-2409, Oct. 2013. doi: 10.1109/TPWRD.2013nai.2265404

[3] Y. Q. Chen, O. Fink and G. Sansavini, “Combined Fault Location and Classification for Power Transmission Lines Fault Diagnosis With Integrated Feature Extraction,” *in IEEE Transactions on Industrial Electronics*, vol. 65, no. 1, pp. 561-569, Jan. 2018, doi: 10.1109/TIE.2017.2721922.

[4] K. Chen, J. Hu, and J. He, “Detection and Classification of Transmission Line Faults Based on Unsupervised Feature Learning and Convolutional Sparse Autoencoder,” *in IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1748-1758, May 2018, doi: 10.1109/TSG.2016.2598881.

[5] M. F. Guo, N. C. Yang and W. -F. Chen, “Deep-Learning-Based Fault Classification Using Hilbert–Huang Transform and Convolutional Neural Network in Power Distribution Systems,” *in IEEE Sensors Journal*, vol. 19, no. 16, pp. 6905-6913, 15 Aug.15, 2019, doi: 10.1109/JSEN.2019.2913006.

[6] P. Balakrishnan, “Transmission line fault detection, classification, and location using wavelet transform”. (2019). International Journal of Engineering and Advanced Technology, 8(6S3), 1770-1775.

[7] Patel.B, Bera.P, Saha.B, “Wavelet Packet Entropy and RBFNN Based Fault Detection, Classification and Localization on HVAC Transmission Line, Electric Power Components and Systems”, 46:1, 15-26, (2020) DOI: 10.1080/15325008.2018.143181

[8] Sarker.R, Muyeen, Das.S, Kamwa.K, I. “A deep learning-based intelligent approach in the detection and classification of transmission line faults*”. International Journal of Electrical Power Energy Systems*, (2021) 133, 107102.

[9] Rojas, D., Christou, I. T., Dantas, D., Narayanan, A., Nardelli, P. H., Yang, Y. “Performance evaluation of machine learning for fault selection in power transmission lines”. Knowledge and Information Systems, 64(3), 859-883.

[10]P.Chiradeja and A. Ngaopitakkul, “Analyzing the Characteristics of Faults in a Transmission Line and High Voltage Capacitor Banks in a 115-kV-Power System Using Discrete Wavelet Transform,” in IEEE Access, vol. 10, pp. 86765-86781, 2022, doi: 10.1109/ACCESS.2022.3199362

[11] Bon, N. N., Dai, L. V. (2022). “Fault identification, classification, and location on transmission lines using combined machine learning methods”. International Journal of Engineering and Technology Innovation, 12(2), 91-109.

[12] Rafique.F, Fu.L, Mai.R, “End-to-end machine learning for fault detection and classification in power transmission lines”, Electric Power Systems Research, Volume 199, 2021, 107430, ISSN 0378-7796, https://doi.org/10.1016/j.epsr.2021.107430

[13] Rai.P, Londhe.N, Raj.R, “Fault classification in power system distribution network integrated with distributed generators using CNN”, *Electric Power Systems Research, Volume* 192, 2021, 106914, ISSN 0378-7796,

[14] J. Han, S. Miao, Y. Li, W. Yang, and H. Yin, “Fault Diagnosis of Power Systems Using Visualized Similarity Images and Improved Convolution Neural Networks,” in IEEE Systems Journal, vol. 16, no. 1, pp. 185-196, March 2022, doi: 10.1109/JSYST.2021.3056536.

[15] Wang.Y, Li.Q, Chen.B, “Image classification towards transmission line fault detection via learning deep quality-aware fine-grained categorization, Journal of Visual Communication and Image Representation”, Volume 64, 2019, 102647, ISSN 1047-3203,

[16] Moradzadeh.A, Teimourzadeh.H, Mohammadi.B,Pourhossein.K, “Hybrid CNN-LSTM approaches for identification of type and locations of transmission line faults”, *International Journal of Electrical Power Energy Systems,* Volume 135, 2022, 107563, ISSN 0142-0615.

[17] Yanhui Xi, Weijie Zhang, Feng Zhou, Xin Tang, Zewen Li, Xiangjun Zeng, Pinghua Zhang, “Transmission line fault detection and classification based on SA-MobileNetV3”, Energy Reports, Volume 9, 2023, Pages 955-968, ISSN 2352-4847,

[18] Y. A. I. Mohamed, K. Y. Lee, and S. A. Muyeen, "A New Approach for Automatic Fault Detection and Classification of Transmission Lines Using Random Forest Classifier," *in IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 982-991, March 2015. doi: 10.1109/TSG.2014.2346868

[19] Ahmed, N., Ahsan, S., Hussain, S., Alam, M. M., & Haque, M. E. “Automatic detection and classification of faults on a transmission line using artificial neural networks”. 2017 4th *International Conference on Advances in Electrical Engineering (ICAEE), Dhaka*, 336-340. doi: 10.1109/ICAEE.2017.8212269

[20] Le, H., Islam, M. N., & Liao, L. J. “Enhanced Fault Detection and Classification in Transmission Lines Using Wavelet Transform and Random Forest Algorithm”. *IEEE Access,* 8, 108243-108253. (2020) doi: 10.1109/ACCESS.2020.3005783

[21] B. M. Albrecht and G. J. Rogers, "Power System Fault Detection and Classification Using Random Forest," *2014 IEEE PES Transmission & Distribution Conference and Exposition, Chicago,* IL, 2014, pp. 1-5. doi: 10.1109/TDC.2014.6863374

[22] S. Pal and S. K. Goswami, "Fault Detection and Classification of Transmission Line Using Random Forest," *2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai*, 2017, pp. 40-45. doi: 10.1109/ICCCT2.2017.7975433

[23] Sun, X., Lu, Y., Li, Y., & Xiang, L. "Intelligent fault diagnosis method for transmission lines based on gradient boosting decision tree algorithm". *IEEE Access*, 9, 37049-37058. (2021) doi: 10.1109/ACCESS.2021.3064186

[24] Luo, S., & Wang, J. "An intelligent diagnosis method for fault identification of transmission lines based on a convolutional neural network. Energies", 12(7), 1257. (2019) doi: 10.3390/en12071257

[25] Zhang, L., Zhang, S., Guo, H., & Gao, X. "An intelligent fault diagnosis method for transmission lines based on an extreme learning machine optimized by bat algorithm. Energies", 13(5), 1277. (2020) doi: 10.3390/en13051277

[26] Mahmood, A., & Hussain, S. "An intelligent technique for fault location identification in high voltage transmission lines using data analytics". IEEE Access, 8, 85851-85865. (2020) doi: 10.1109/ACCESS.2020.2995122

[27] B. Vahidi, & A. Saghafinia. "A novel intelligent approach for fault diagnosis in transmission lines using deep learning" *International Journal of Electrical Power & Energy Systems*, 111, 267-275. (2019), doi: 10.1016/j.ijepes.2019.04.018