Fault Analysis in Electrical System using Machine Learning Algorithms

by

Md. Salman Sakib -190021206
Maleha Mahmud -190021212
Md. Safin Mahmood Shomyo -190021217

A Thesis Submitted to the Academic Faculty in Partial Fulfillment of the Requirements for the Degree of

BACHELOR OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

Department of Electrical and Electronic Engineering
Islamic University of Technology (IUT)
Gazipur, Bangladesh

June 2024

CERTIFICATE OF APPROVAL

The thesis titled **"Fault Analysis in Electrical System using Machine Learning Algorithms"** submitted by Md. Salman Sakib (190021206), Maleha Mahmud (190021212) and Md. Safin Mahmood Shomyo (190021217) has been found as satisfactory and accepted as partial fulfillment of the requirement for the Degree BACHELOR OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

Thesis approved on 25th June, 2024.

Approved by:

Ashraful Islam Mridha
Supervisor and Lecturer,
Department of Electrical and Electronic Engineering,
Islamic University of Technology (IUT),
Boardbazar, Gazipur-1704.

DECLARATION OF CANDIDATE

It is hereby declared that this thesis titled **“Fault Analysis in Electrical System using Machine Learning Algorithms”** is the result of diligent research conducted under the guidance of Ashraful Islam Mridha. This thesis or any part of it has not been submitted elsewhere for the award of any Degree or Diploma.

Md. Salman Sakib
Student No: 190021206

Maleha Mahmud
Student No: 190021212

Md. Safin Mahmood Shomyo
Student No: 190021217

ACKNOWLEDGEMNT

” In the Name of Allah, the Most Beneficent, the Most Merciful. All the praises and thanksbe to Allah, the Lord of the ’Alamin (mankind, jinns and all that exists).”

We extend our gratitude to Allah, the Lord of all creation, for granting us the health, strength, and capability to conduct research and complete this dissertation.

We deeply appreciate and acknowledge our supervisor, Ashraful Islam Mridha, for his invaluable guidance, precise instructions, and support throughout. His wealth of experience, mentorship, and significant contributions were indispensable to the success of this thesis. We are especially thankful for his assistance in conceptual exploration, diligent efforts, and the drafting of this thesis.

Our heartfelt thanks go to our friends and families, particularly our parents, whose unwavering support and dedication have been pivotal in our journey. Their constant encouragement has shaped us into the individuals we are today, and without their steadfast belief in our abilities, our accomplishments would not have been possible."

Table of contents

Table of Figures

List of Tables

List of Abbreviations

k-NN: k -Nearest Neighbors

LR: Linear Regression

SVM: Support Vector Machine

DT: Decision Trees

MLP: Multi-Layer Precision

SLG: Single-Line-to-Ground

LLG: Line-to-Line-Ground

DLG: Double-Line-to-Ground

ABSTRACT

This research provides an overview of the prediction accuracy of different classification-based machine learning algorithms from a 'Simulink based 3-phase Electrical System Model'. The model consists of a three-phase source and two relay bus bars connecting the ends of two subsystems. The data generation for faults was embedded by connecting a three-phase fault block between two subsystems. Later, the whole simulation was explained by interchanging the parameters among phase A, B, C and Ground. Each of the simulation placed between 0 to 1 second. The data table exhibits more than 16,000 samples across voltages ranging from -0.1023 V to 1.9776 V and a current ranging from -0.13813 A to 12.53622 for each of the faulty phase. More than 70% of the data was used for model training later and rest of the 30% raw data was planned for prediction. For classification; KNN, SVM, LR, DT, Gradient Boosting, Random Forest and MLP Classifiers algorithms were used for comparison. After several Preprocessing, 'Random Forest' algorithm comes with the highest accuracy of 0.9477 or 94.77%. Thus, the prediction can help the electrical engineers to automate the process of fault detection with the help of machine learning.

CHAPTER 1

Chapter 1: Introduction and Background

Accurate identification and categorization of shortcomings within power systems are fundamental to ensuring network reliability and reducing downtime. Traditional fault detection methods often rely on intricate rule-based systems or overly simplistic threshold-based approaches, which may not fully capture the nuanced nature of fault patterns and can consequently lead to inaccuracies or false alerts. In contrast, machine learning techniques present a promising avenue for developing robust predictive models capable of automatically detecting and classifying power system issues.

We present a forward-looking approach for the categorization of power system deficiencies by employing machine learning algorithms in this study. The demonstration utilizes recorded data obtained from sensors positioned across the control lattice in order to generate and enhance classification calculations. An examination is conducted on and comparison is made of the efficacy of various machine learning algorithms, including arbitrary forests, back vector machines, neural systems, and decision tree methods, in the context of fault classification.

Our findings demonstrate that the proposed predictive model is capable of accurately categorizing power system defects, surpassing traditional approaches in terms of precision and efficacy. The developed programmed possesses the capacity to enhance fault location functionalities in power systems, thereby advancing framework reliability and reducing maintenance expenses. Further, the proposed method is highly compatible with pre-existing frameworks for observing control networks due to its adaptability and versatility.

This research is motivated by the necessity to enhance the efficiency and accuracy of fault analysis in electrical systems. Conventional methods have limitations, prompting the development of more advanced, automated, and reliable techniques. Machine learning (ML) algorithms, with their ability to process extensive data and identify intricate patterns, provides an exclusive solution to these challenges. By integrating with ML, it shows potential to significantly improve the detection, classification, and diagnosis of electrical faults, ultimately leading to stronger and more resilient power systems.

This research aims to enhance fault analysis in electrical systems using machine learning algorithms to improve accuracy, efficiency, and reliability. Objectives include identifying common fault types, compiling comprehensive datasets, developing and evaluating machine learning models, and exploring real-time deployment and scalability in power systems. The goal is to automate fault detection and diagnosis, ensuring reliable and stable power distribution networks.

1.1 Basic Functionalities of Power System Faults

A few essential processes are typically included in a vision show for the machine learning-based classification of power system fault issues, compiling data from many sources, including meters, sensors, and official records, within the power system. This data emphasized important features related to fault categorization, such as voltage levels, current levels, power factor, frequency, etc. Also, the data gathering process encompassed standard working conditions as well as instances of documented malfunctions. This also includes managing missing numbers, exceptions, and disturbances in the data. Lastly, generating new underused highlights or modify current ones to extract additional important information for fault categorization. These methods may involve using Fourier transforms for analyzing patterns of recurrence or wavelet transforms for analyzing patterns of time-frequencies, helping to elect an appropriate machine learning algorithm for fault categorization.

Typical calculations used for classification tasks include k-nearest neighbors (KNN), support vector machines (SVM), decision tree, random forest and neural networks. Thus, calculation can fluctuate depending on several factors, i.e. the magnitude and complexity of the dataset, computational resources required, and the interpretability of the model.

1.2 Different Aspects of Power System Faults

The study explores the use of foresight modelling to classify fault in power systems, which has managed to gain attention gradually due to its ability to improve the dependability and effectiveness of electrical grids. In this case, Machine learning algorithms provide powerful tools to develop predictive models for fault categorization. This study explores many viewpoints of predictive modelling for power system fault classification using machine learning, including feature selection, algorithm selection, data preprocessing, model evaluation, and interpretability.

A few aspects of our solution on which we will be focusing are given below

1. Diagram of power system flaws and their centrality.
2. Significance of prescient modeling in blame classification.
3. Part of machine learning in creating prescient models. Information collection and preprocessing methods.
4. Dealing with lost values and exceptions.
5. Information normalization and standardization
6. Recap of diverse perspectives talked about within the paper.
7. Prospects and bearings for inquire about in prescient modeling for power system blame classification utilizing machine learning.

1.3 Over Head Transmission Line

Overhead transmission lines play a critical role in electrical power systems, efficiently transmitting large amounts of electricity. Understanding their behavior during faults, such as three-phase faults, involves analyzing fault currents and ensuring proper protective measures are in place to maintain system reliability and safety.

Components of an Overhead Transmission Line:

- **Conductors:** Aluminum or steel wires that carry electricity.
- **Insulators:** Glass or porcelain to support and insulate conductors.
- **Supporting Structures:** Towers or poles to maintain conductor spacing and height.

Three-Phase Fault Scenario:

In electrical power systems, a three-phase fault occurs when all three phases (typically labeled as phases A, B, and C) come into direct contact with each other or with ground simultaneously. This results in a short circuit that can lead to significant currents flowing through the transmission line.

Behavior and Equations:

1. **Fault Current (I_{fault}):** During a three-phase fault, the fault current I_{fault} can be calculated using Ohm's Law:

$$I_{fault} = \frac{V_{line}}{Z}$$

Where:

V_{line} is the line-to-line voltage of the transmission line.

Z is the total impedance of the transmission line.

The fault current can also be expressed as:

$$I_{fault} = \frac{\sqrt{3}V_{phase}}{Z}$$

Where V_{phase} is the phase voltage.

2. **Impedance of the Transmission Line (Z):** The impedance Z consists of resistive (R) and inductive (X) components:

$$Z = R + jX$$

R is the resistance of the conductors.

X is the inductive reactance due to the magnetic field around the conductors.

3. **Voltage Drop (V_{drop}):** The voltage drop across the transmission line during the fault can be significant and is given by:

$$V_{drop} = I_{fault} . Z$$

Impact and Protection:

- **Equipment Stress:** High fault currents can stress equipment and components, potentially causing damage.
- **Protection Coordination:** Protective devices such as circuit breakers and relays are installed along the transmission line to detect faults and isolate the affected section promptly.
- **System Stability:** Quick fault detection and isolation help maintain system stability and prevent widespread outages.

Figure 1.1: Overhead transmission line

Chapter 2

Chapter 2: Overview of Power System Faults

2.1 Introduction

Electrical control demand is rising worldwide. Transmission lines in a power system are crucial to meeting this growing demand. Transmission framework has its own problems controlling warm and soundness edges. Consequently, this circumstance causes an unstable framework. The electricity system is adjusted under normal conditions and becomes unequal if there are problems. Thus. Voltage deviations are blamed and current in common power systems that indicate unsettling effect.

1. Blame occurs anytime, anywhere, with varying blame impedance. The Blame is a sink point; therefore, voltage tends to zero. All focuses with more potential than the defective point transmits current to it, raising the blame level many times higher than usual.

2. Unsymmetrical blames comprise one or more stages, but symmetrical blames include all three. Damage to the power system is common with any failure. It's important to treat and analyze this common power system issue to maintain constant conditions and system operation. The poor framework voltage and current are decreed during blame conditions so the terrible framework can be identified and remedied. Blame inquiry evaluates power system voltages and streams under various situations. Power systems have thousands of buses, making it difficult to calculate essential data without a computer programmed.

3. Deterministic computations and well-characterized control lines lead to customary procedures and errors in delayed discovery. Traditional transfers consider. blame control swing. The system allows the control gearbox line to be detached if damage occurs to protect the gear.

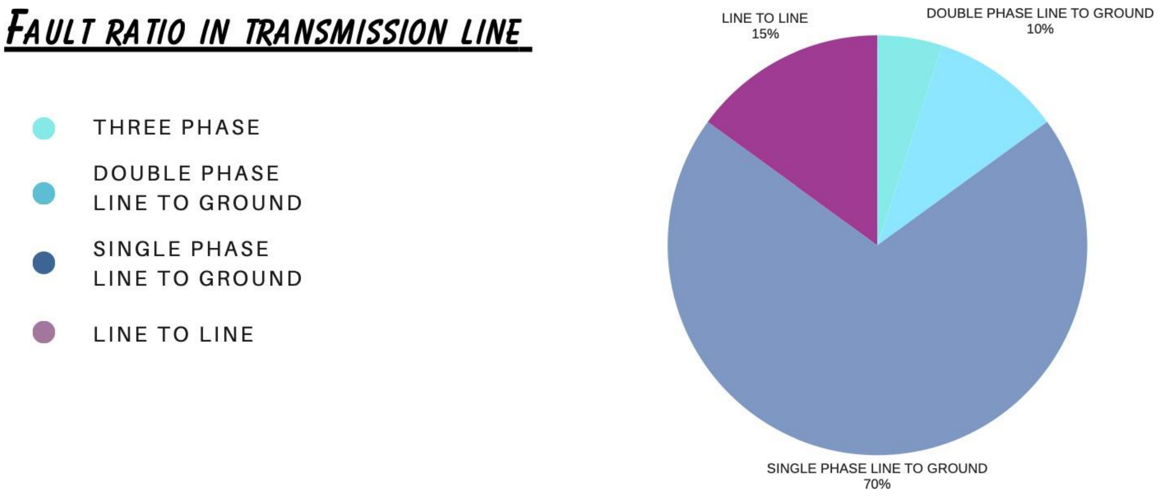


Figure 2.1: Fault ratio in transmission line

Fault Classification in Overhead Transmission Lines:

1. **Single Line-to-Ground Fault (SLG Fault)**

- Description: This happens when a conductor makes direct contact with the ground or an object

- Causes: Lightning strikes, tree branches falling on the line, insulator failure, and contamination such as salt spray.

- Effects: This fault can lead to significant voltage drops, causing instability and possible damage to equipment. It is the most common type of fault in overhead lines.

2. Line-to-Line Fault (LLG Fault)

- Description: This happens when two conductors make contact with each other.

- Causes: High winds causing lines to sway and touch, broken conductors, or foreign objects like tree branches bridging the gap between lines.

- Effects: Line-to-line faults can cause severe arcing and heating, potentially damaging conductors and other equipment.

3. Double Line-to-Ground Fault (DLG Fault)

- Description: This happens when two conductors simultaneously meet the ground or a grounded object.

- Causes: Severe weather conditions, fallen trees, or failure of multiple insulators.

- Effects: This type of fault can lead to large current flows, causing extensive damage to equipment and posing a high risk of fire.

4. Three-Phase Fault (Symmetrical Fault):

- Description: This involves all three conductors simultaneously and is the most severe type of fault.

- Causes: Catastrophic equipment failures, such as transformer explosions or severe lightning strikes.

- Effects: Three-phase faults result in large current flows and significant voltage drops, potentially causing widespread outages and extensive damage.

5. Open Circuit Fault

- Description: This occurs when one or more conductors break or are disconnected.

- Causes: Mechanical damage due to storms, high winds, falling trees, or equipment failure.

- Effects: Open circuit faults lead to unbalanced loads, which can cause overloading of other parts of the network and instability.

Causes of Faults:

Electrical systems, machinery, and components frequently experience various faults during operation. These faults can cause impedance values of machines to deviate from their original levels until the fault is cleared. Faults can be caused by structural collapses, failures in conductor insulation, path disruptions, and natural disasters, all of which have the potential to result in short circuits and significant system damage. During typical operation, electric apparatus within power grids operates at specified voltage and current levels. However, in the event of a fault, voltage and current levels may deviate from their anticipated values.

1. Environmental Factors:

- Weather Conditions: Lightning strikes, high winds, ice storms, and heavy snow can cause physical damage to transmission lines.
- Vegetation: Trees and branches can fall on lines, especially during storms, causing short circuits or grounding.
- Wildlife: Animals such as birds, squirrels, and rodents can meet electrical components, causing faults.

2. Equipment Failures:

- Insulator Failure: Insulators can degrade over time due to environmental exposure, leading to arcing and faults.
- Conductor Failures: Mechanical stress, corrosion, and wear and tear can cause conductors to break or sag excessively.
- Transformer Failures: Transformers and other substation equipment can fail, leading to faults that propagate through the transmission system.

3. Human Errors

- Operational Mistakes: Errors during maintenance or operation, such as incorrect settings on protection devices or mishandling of equipment.
- Construction Activities: Accidental damage to lines during construction or excavation activities.

4. System Overloads

- Overloading: Excessive demand can cause conductors and other components to overheat and fail.
- Power Quality Issues: Harmonics, voltage sags, and swells can stress equipment, leading to faults over time.

Effects of Faults:

-Power Outages: Faults can lead to interruptions in the power supply, affecting both residential and industrial customers.

-Equipment Damage: Faults can cause severe damage to conductors, insulators, transformers, and other components, leading to expensive repairs and replacements.

-Safety Hazards: Faults can pose significant risks to safety, including fire hazards and electrocution risks for maintenance personnel and the public.

-System Instability: Faults can cause voltage drops and instability in the power system, potentially leading to cascading failures and widespread outages.

Understanding and addressing the causes of faults in overhead transmission lines is crucial for ensuring a reliable and safe power supply. Through effective monitoring, maintenance, and advanced fault detection technologies, utilities can significantly reduce the incidence and impact of these faults.

2.2 Network Architecture

Let's first consider a Simulink based Model. This model includes all blocks that make up a power system transmission line. This allows for effective analysis and identification of fault conditions in the system.

In this Simulink based Model, there are several blocks serving different purposes to the system:

1. Source Block: With a basic Y-setup, a three-phase balanced voltage source is connected with ground alongside a small internal impedance. By choosing particular values for Resistance and Inductance, or the source including the resistance and capacitance internally could be adjusted by varying its Impedance/Resistance ratio with short-circuit capabilities.

2. Impedance Block: It computes a linear circuit's frequency-dependent impedance between two nodes. Here, a voltage measurement (Vz) is placed in parallel with the current source and a current source (Iz) is linked to the block's inputs 1 and 2.

3. Power Lines: This block uses the lumped-parameter pi-line model to simulate a three-phase transmission line. Phase resistance, mutual inductance between phases, self-inductance and phase resistance, line-to-ground and line-to-line capacitance are all included.

4. Bus Bar Block: A Bus Bar Block is a type of power system block that is not a physical component but rather is designed to symbolize an important location for measurement. Direct connections are made between many physical signal lines without the need of a bus bar.

5. Fault Block: An external signal from Simulink or an internal control timer can be used to operate the three-phase circuit breaker known as the fault block. The 3-Phase Fault block makes use of breakers. The device can be configured to detect faults of either between two phase (P-P), phase-to-ground (P-G), or a mixed of both kinds. There is an individual on/off switch for each of the system

6. Data Block: While the Simulink is operating, data creation from the failure block: This block records the voltages and currents observed during the execution of Simulink, specifically under various fault scenarios. These values significantly enhance our ability to analyze the fault that happened.

The voltages and currents collected from the three-phase system are transformed into Root Mean Square (RMS) values to improve accuracy. The timer establishes the duration during which the fault data is recorded, ranging from 5/50 seconds to 15/50 seconds. In the article, the line length is specifically specified to 90.01 km. This distance is considered a necessary threshold for fault consideration, as depicted in the block diagram displayed in Figure 2.1. The data block stores the data produced in a tabular format and stores it in the MATLAB workspace. There are two Bus Bars in the diagram:

Relay Bus 1: In the center of the diagram, it serves as a link between Subsystems 1 and 2

Relay Bus 2: Connecting Subsystem 2 to the Three-phase load which is situated in the center of the diagram.

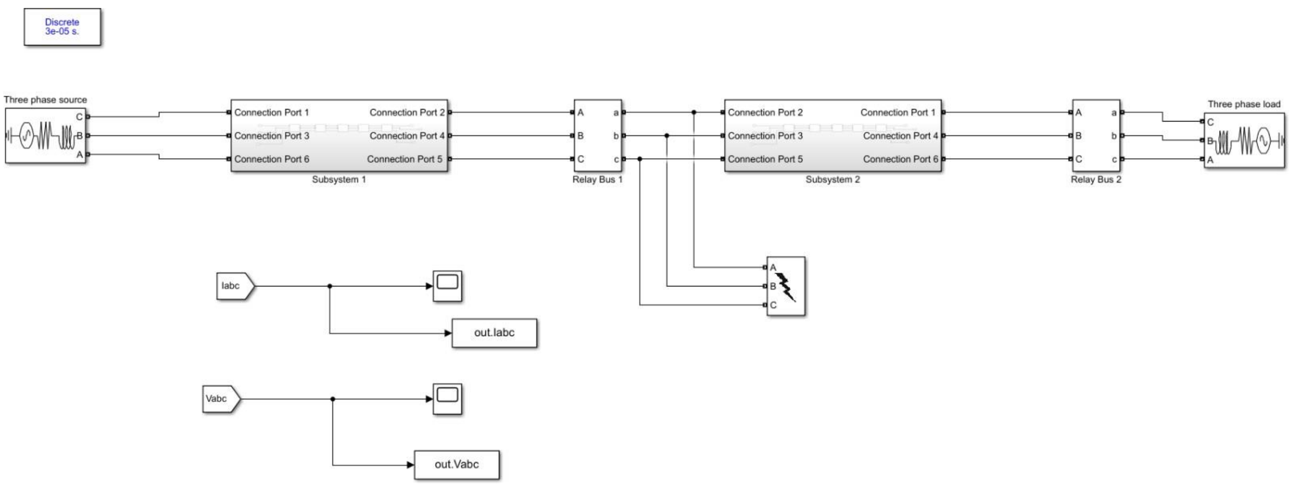


Figure 3: Tentative Simulink based block model

Output Flowchart:

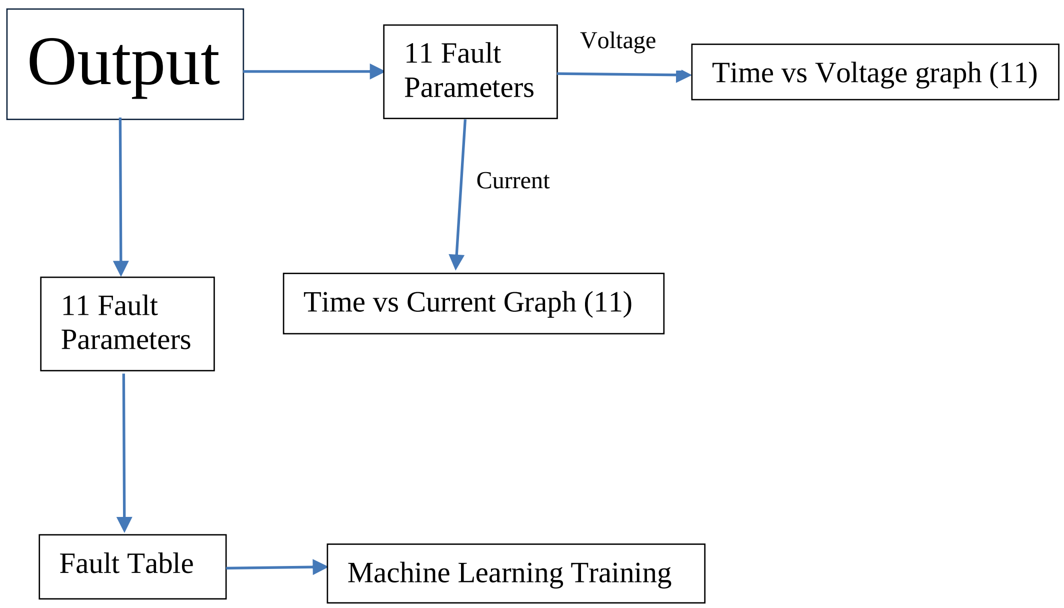


Figure 4: Output Flowchart

2.3 Output Voltage Graph:

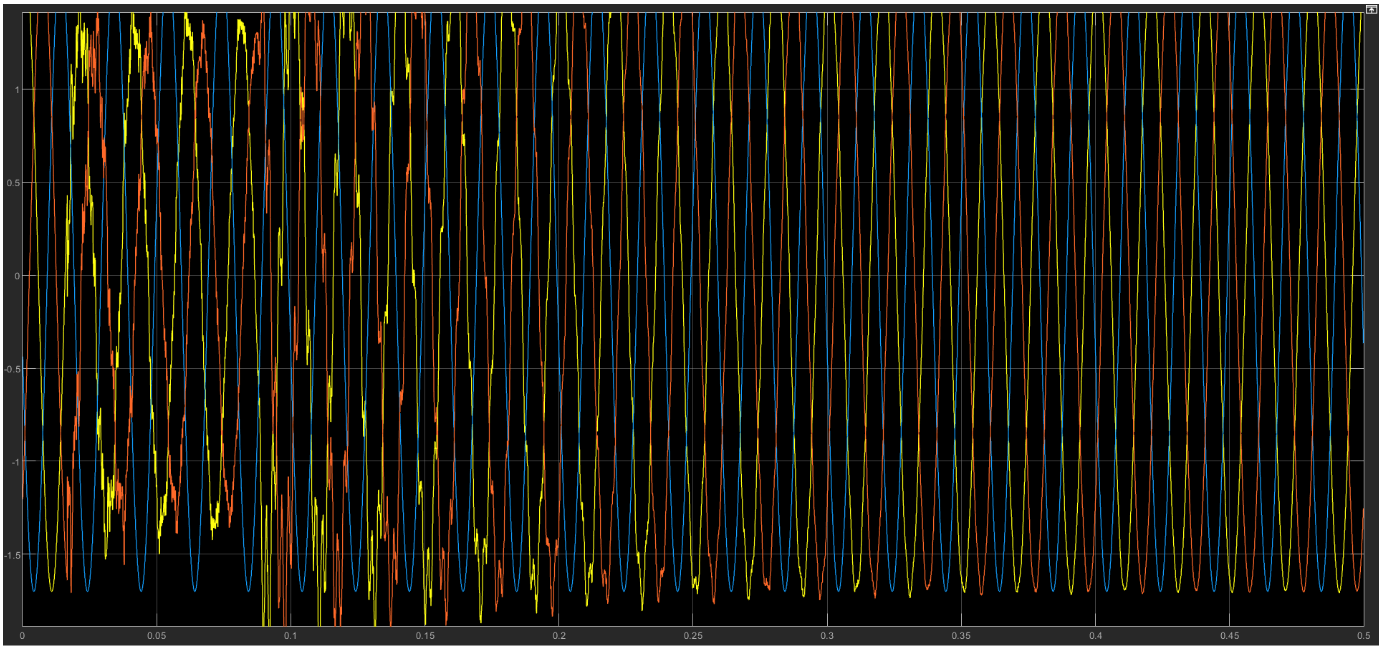


Figure 5: Output graph: voltage between Phase A and Ground

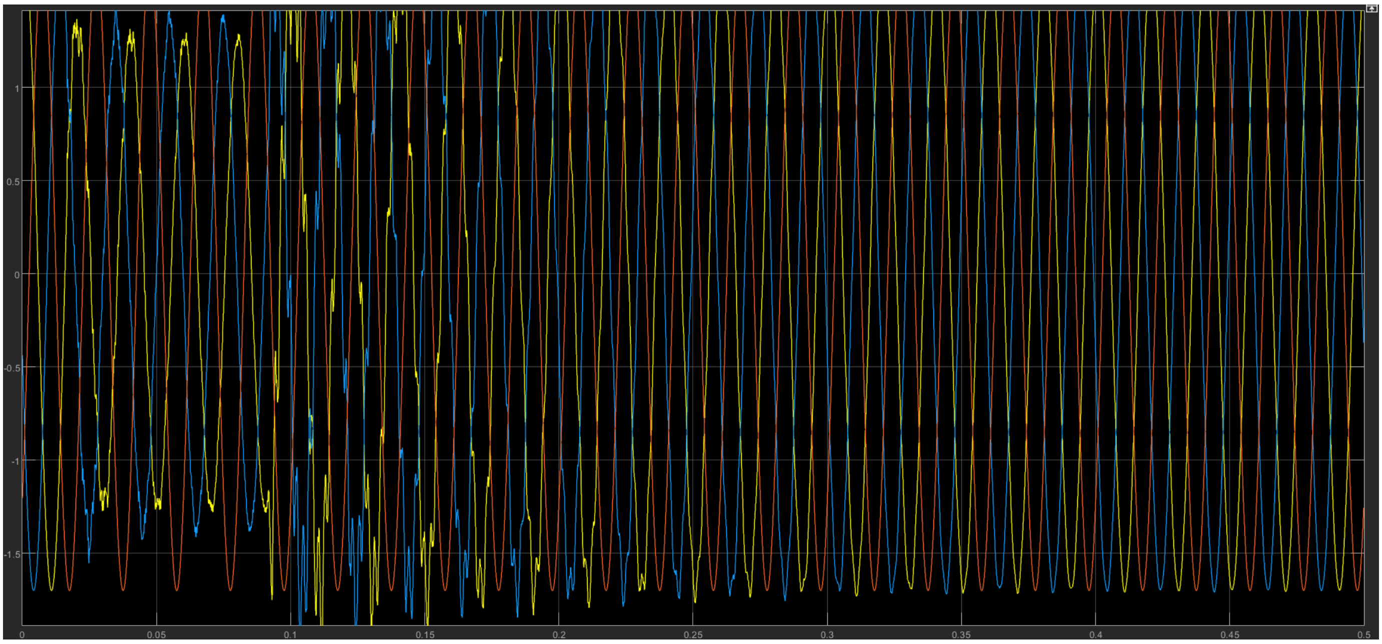


Figure 6: Output graph: voltage between Phase B and Ground

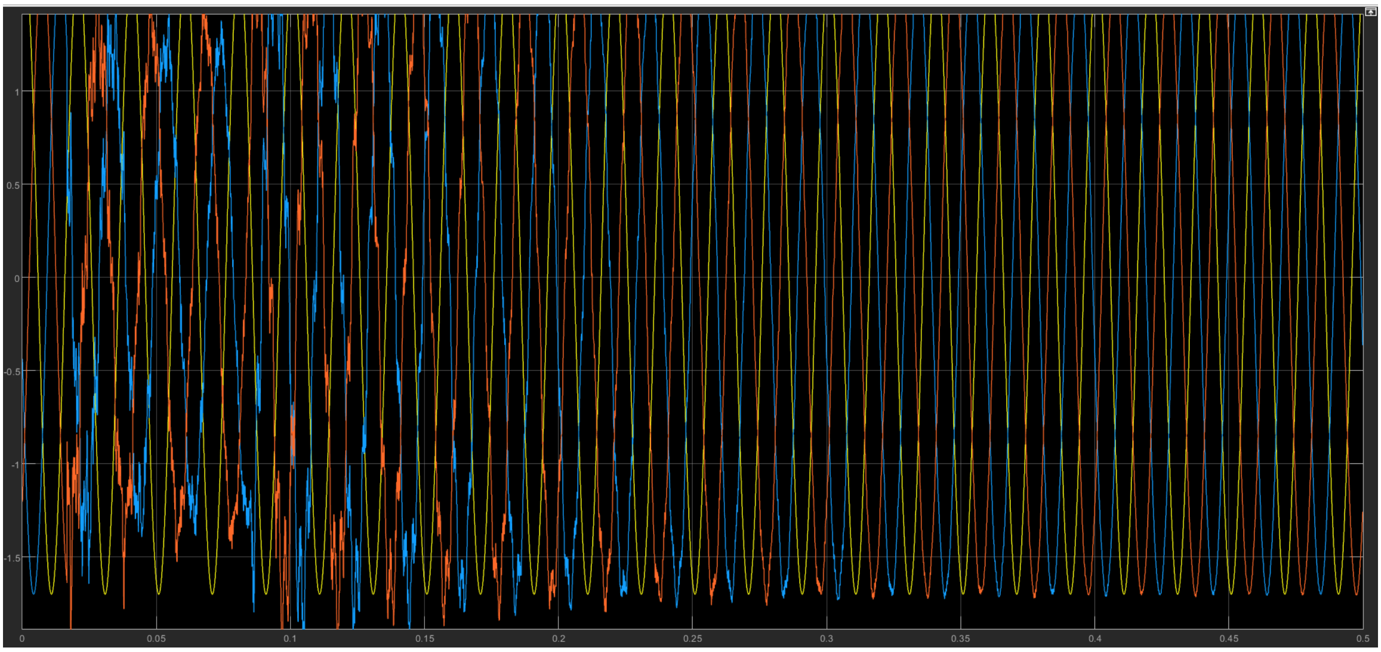


Figure 7: Output graph: voltage between Phase C and Ground

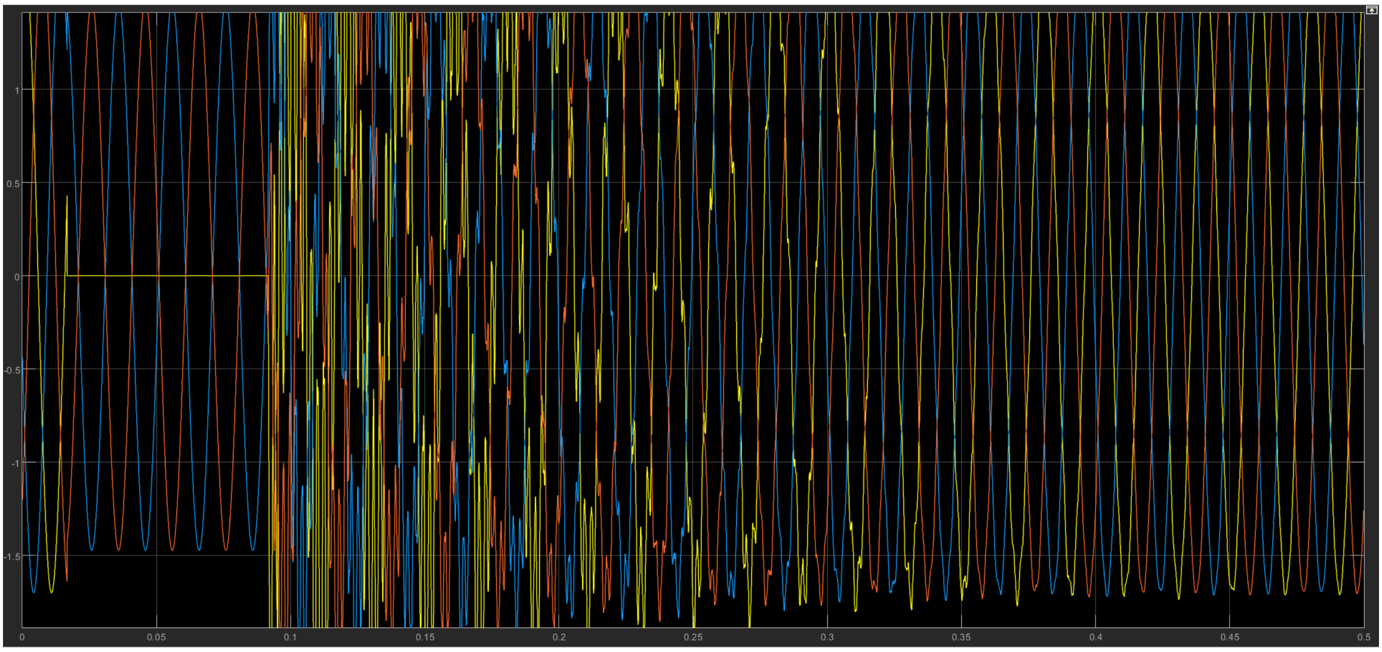


Figure 8: Output graph: voltage between Phase A, B

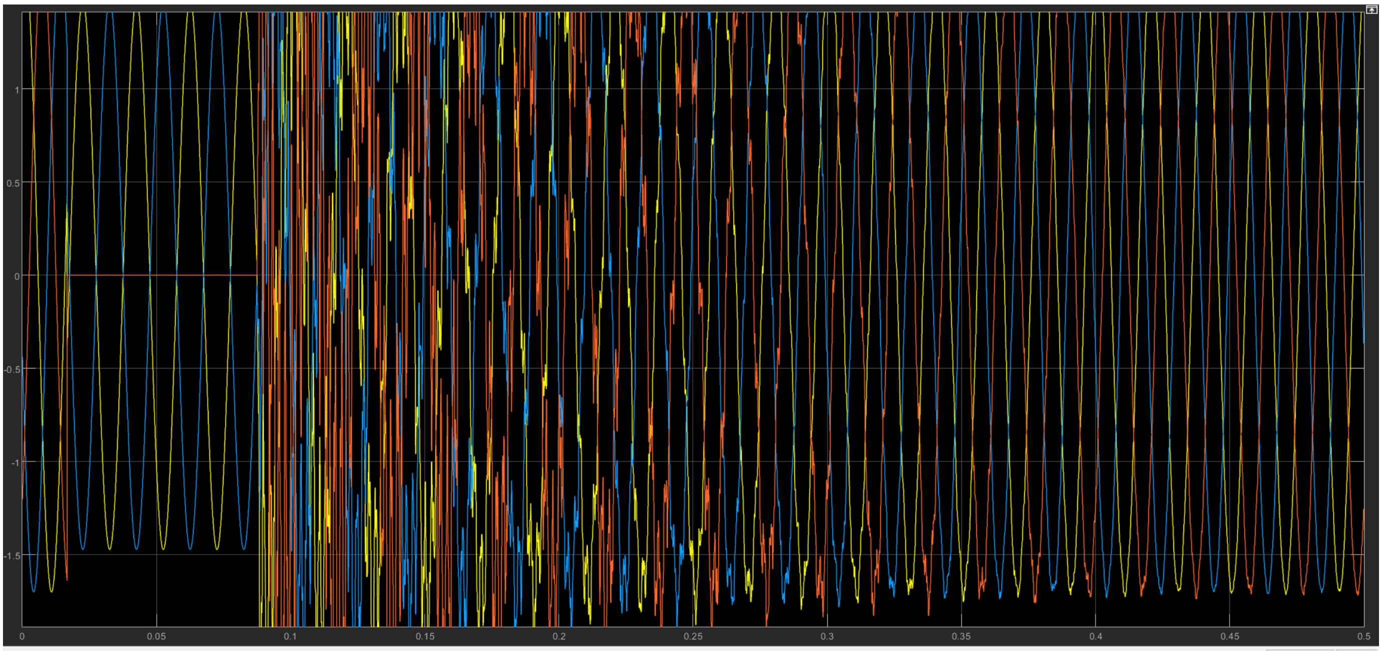


Figure 9: Output graph: voltage between Phase A, C

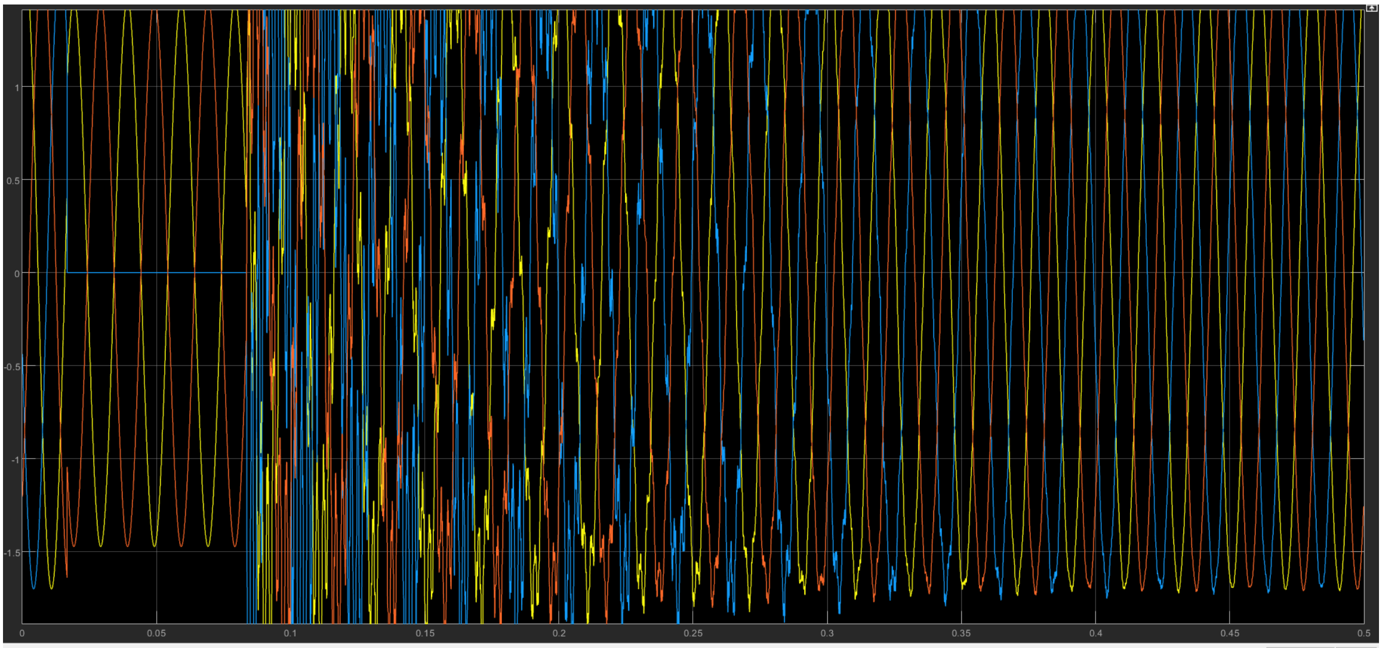


Figure 10: Output graph: voltage between Phase B, C

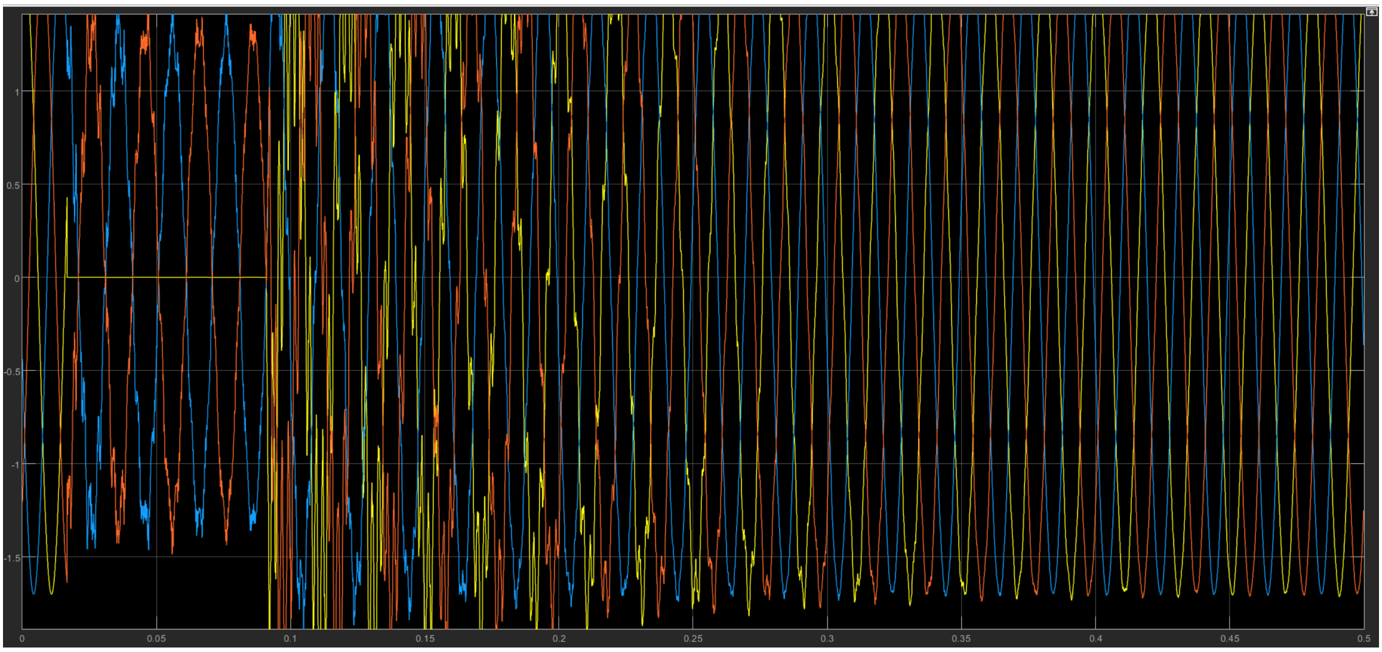


Figure 11: Output graph: voltage among Phase A, B and Ground

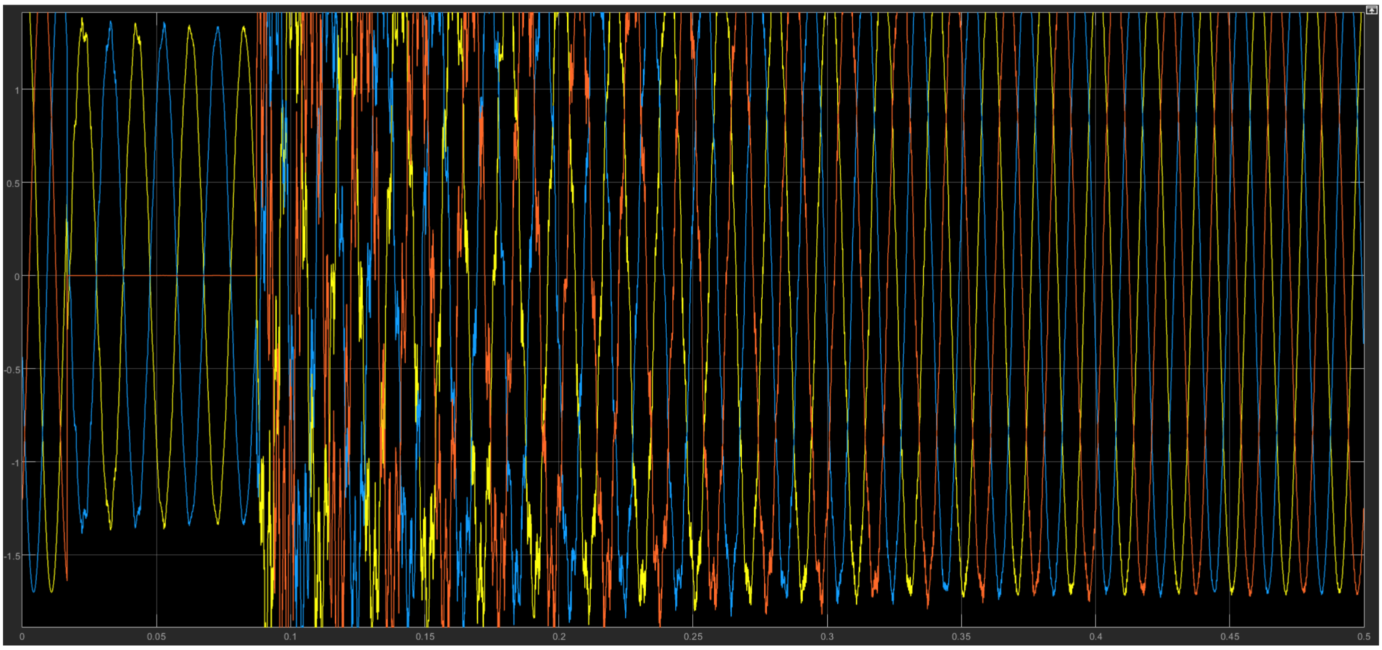


Figure 12: Output graph: voltage among Phase A, C and Ground

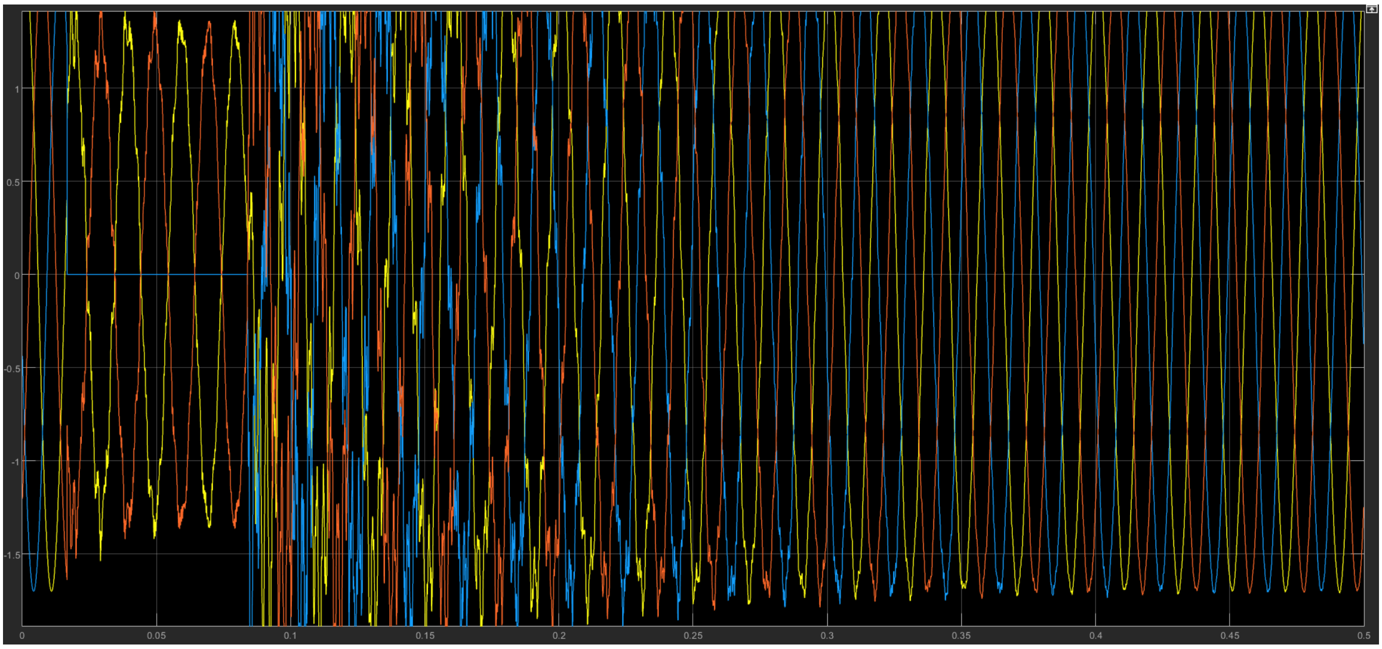


Figure 13: Output graph: voltage among Phase B, C and Ground

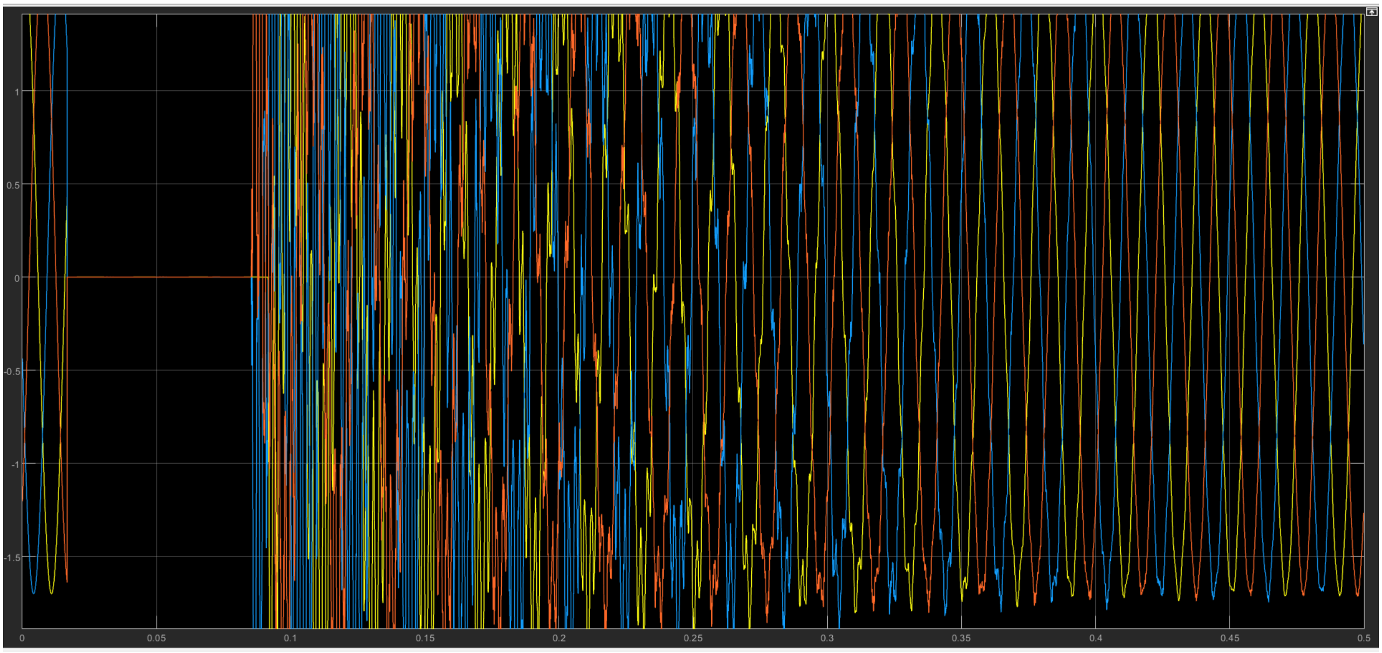


Figure 14: Output graph: voltage among Phase A, B, C and Ground

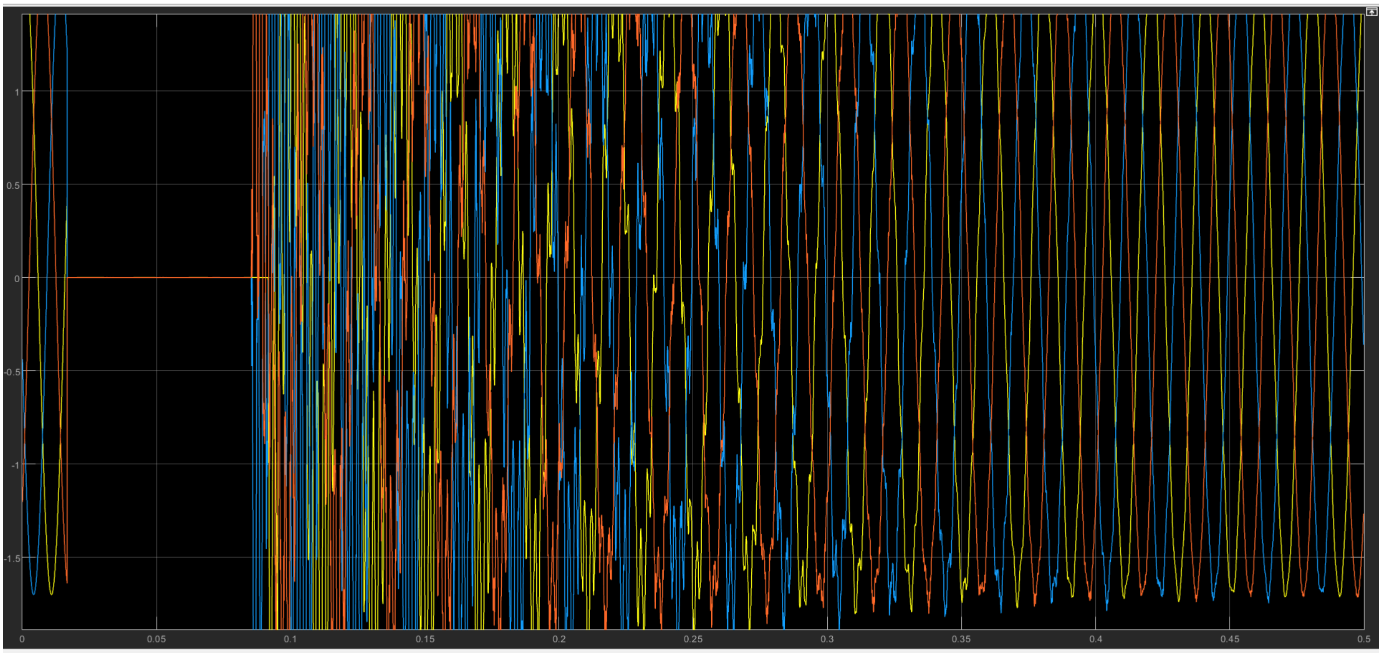


Figure 15: Output graph: voltage among Phase A, B and C

2.4 Output Current Graph:

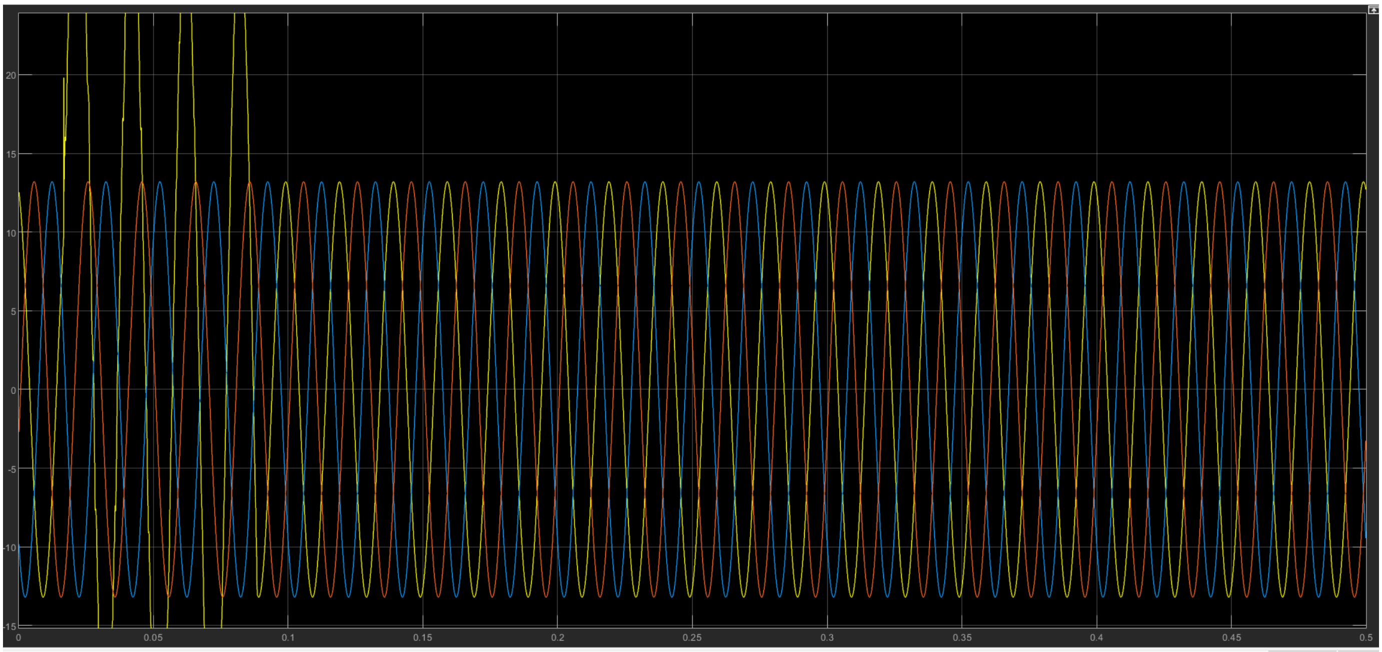


Figure 16: Output graph: current between Phase A and Ground

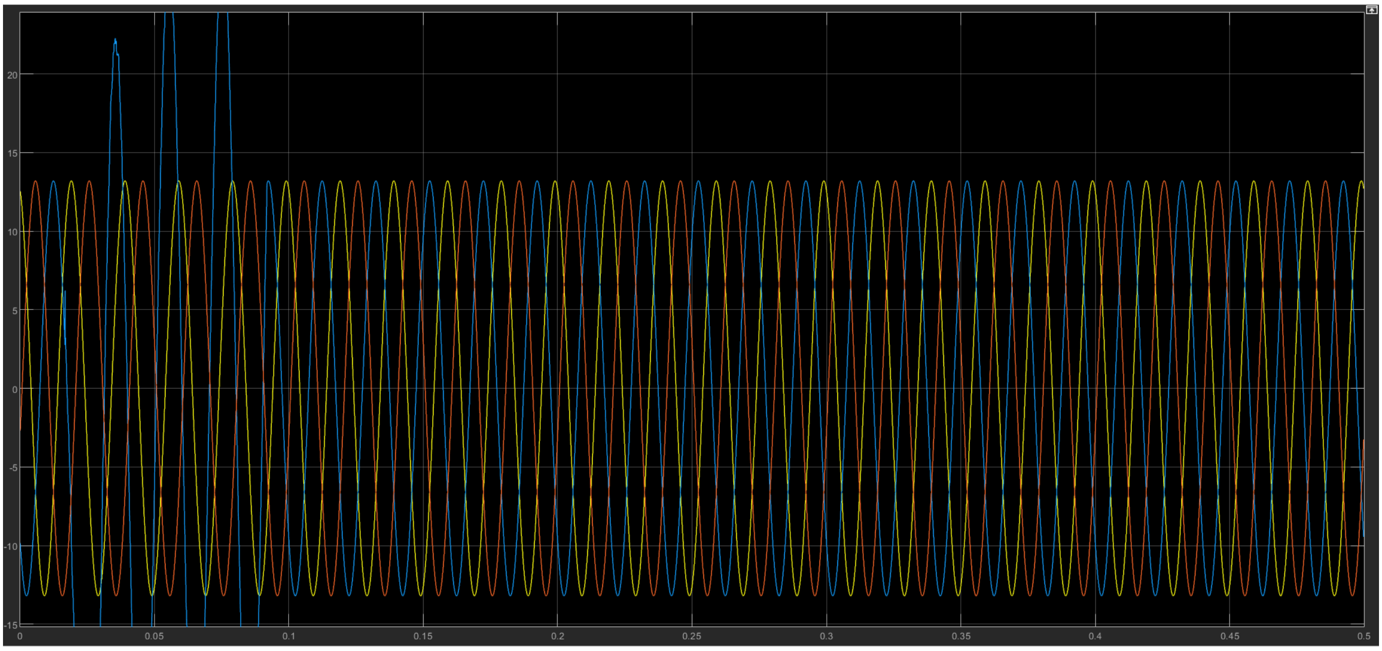


Figure 17: Output graph: current between Phase B and Ground

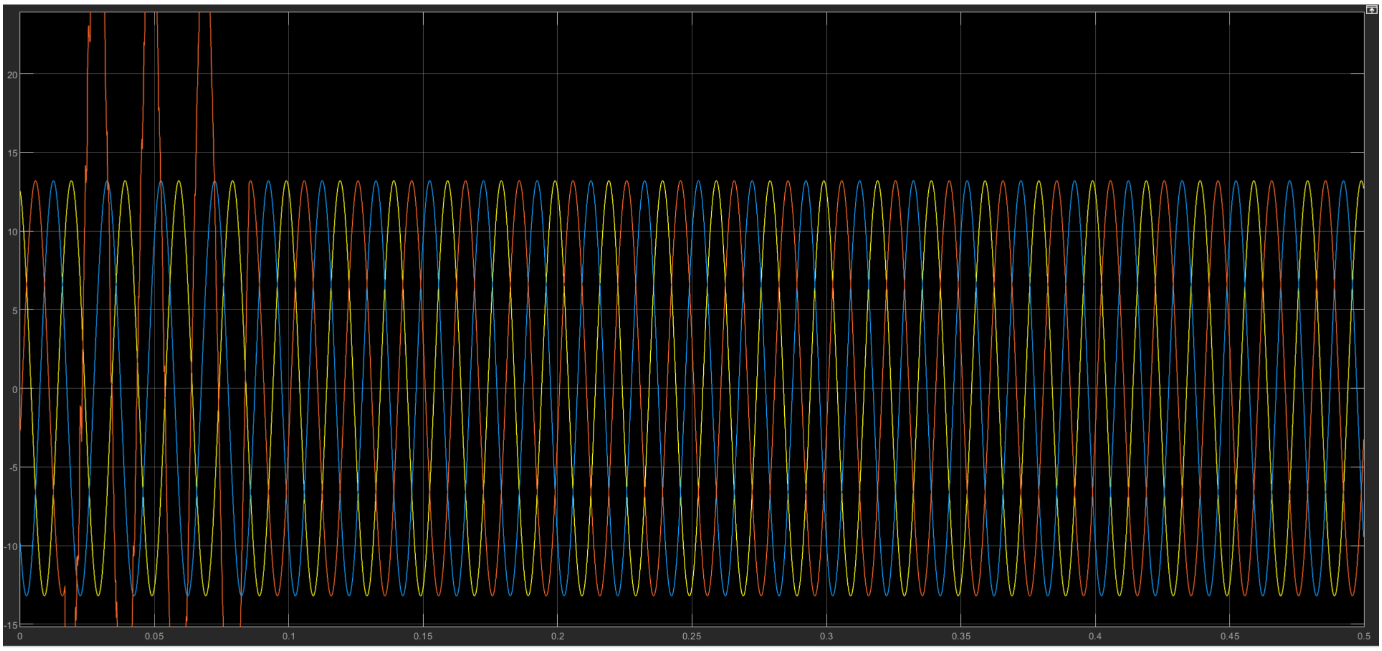


Figure 18: Output graph: current between Phase C and Ground

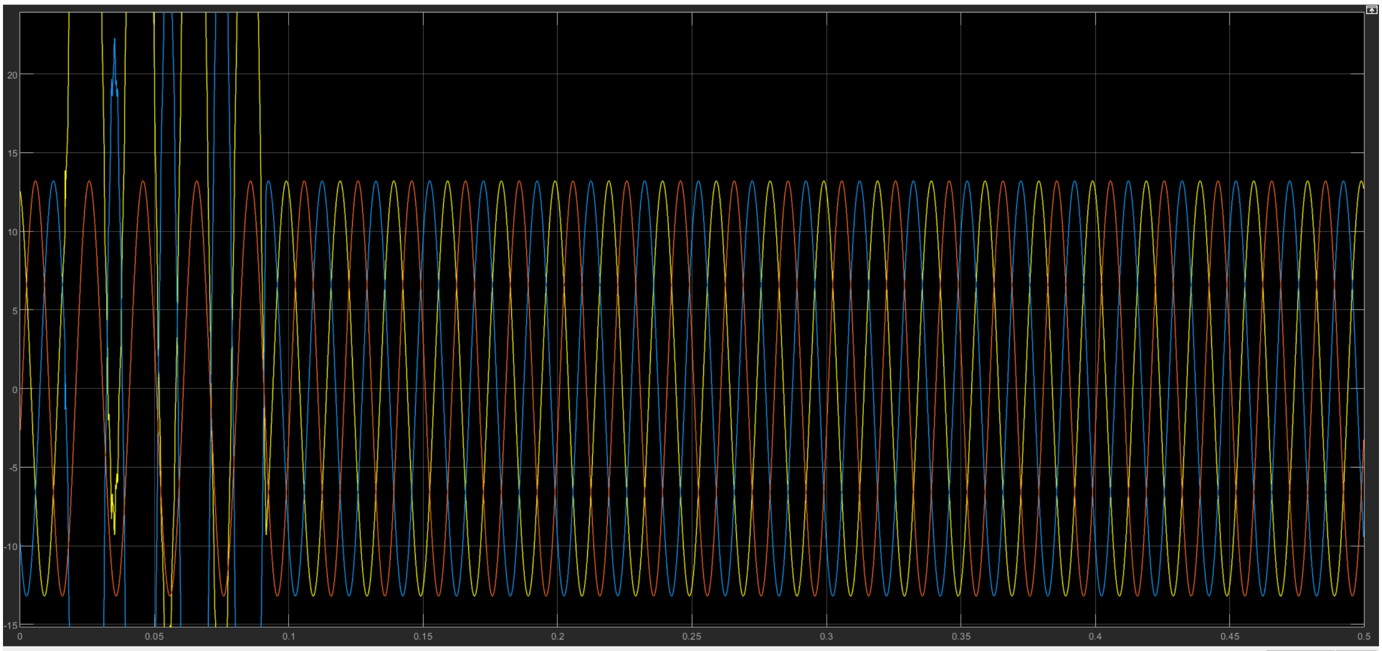


Figure 19: Output graph: current between Phase A and B

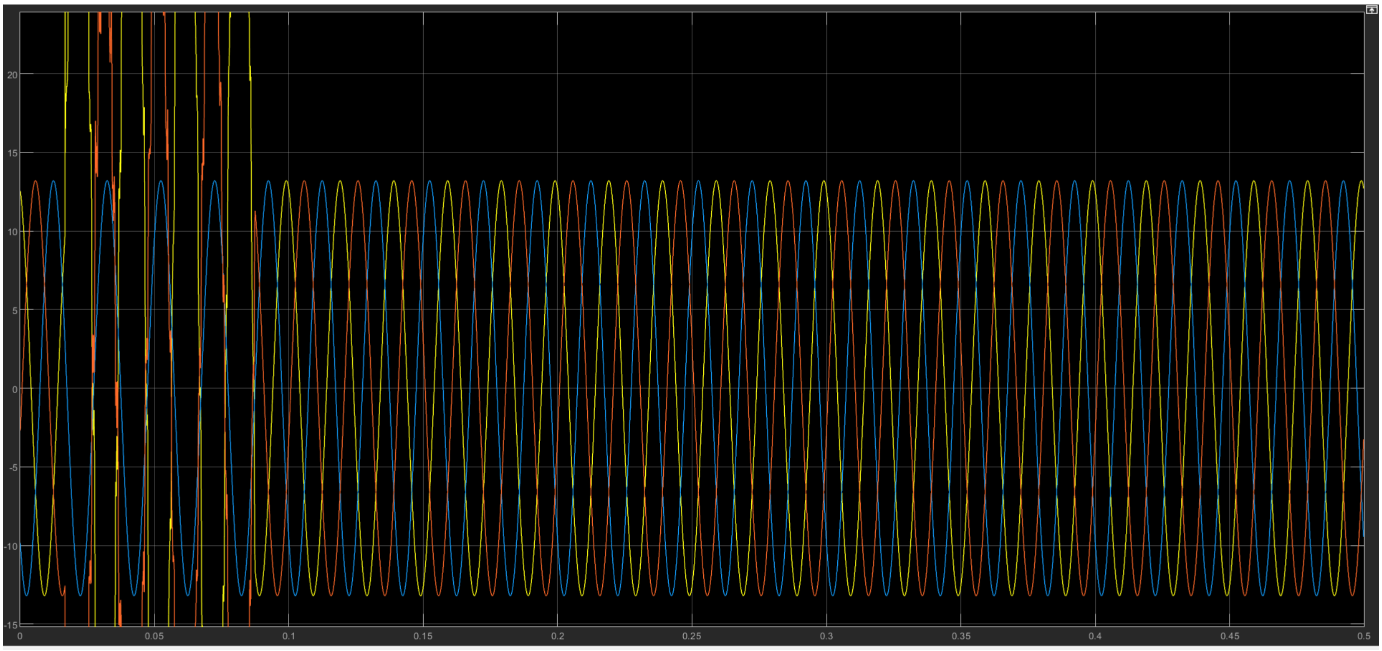


Figure 20: Output graph: current between Phase A and C

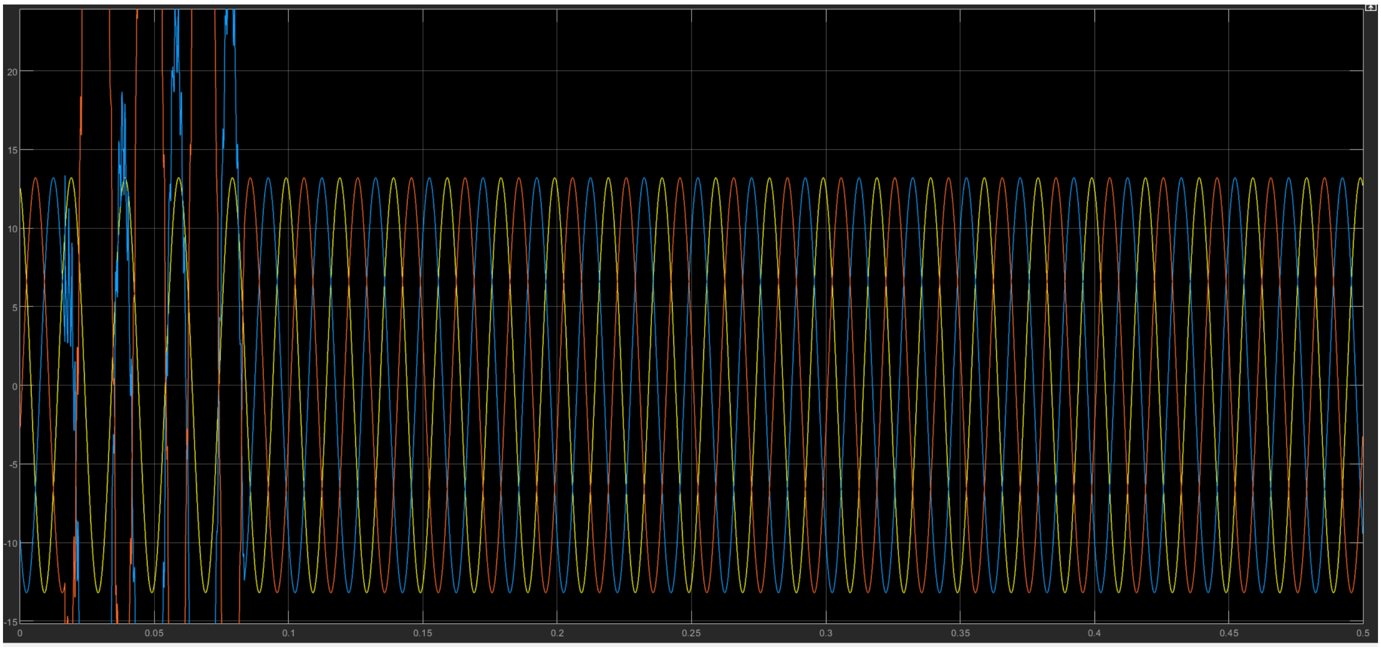


Figure 21: Output graph: current between Phase B and C

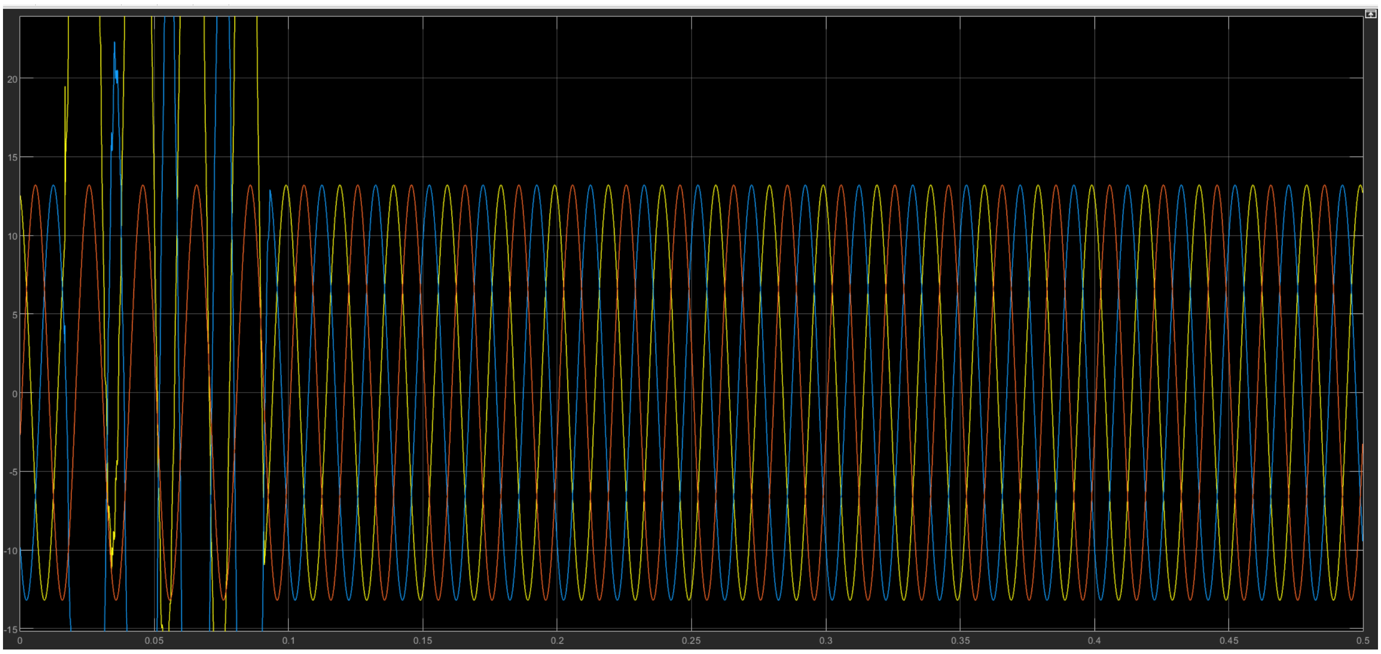


Figure 22: Output graph: current among Phase A, B and Ground

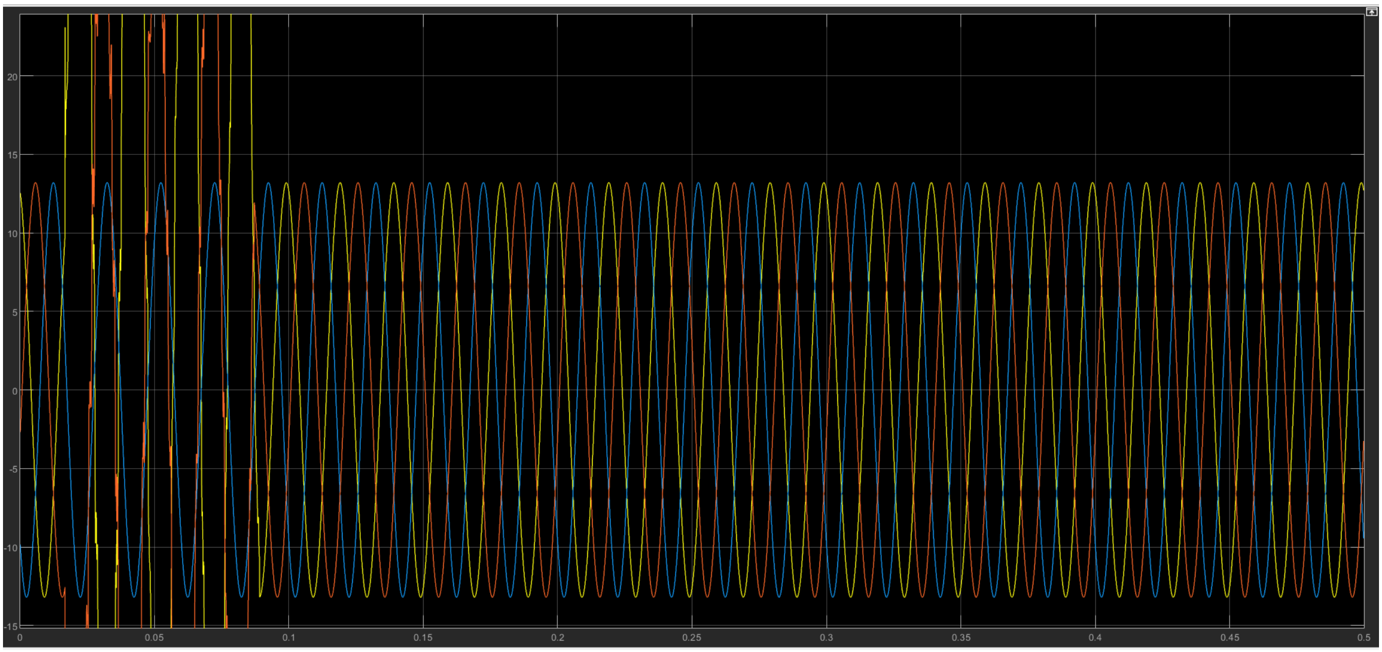


Figure 23: Output graph: current among Phase A, C and Ground

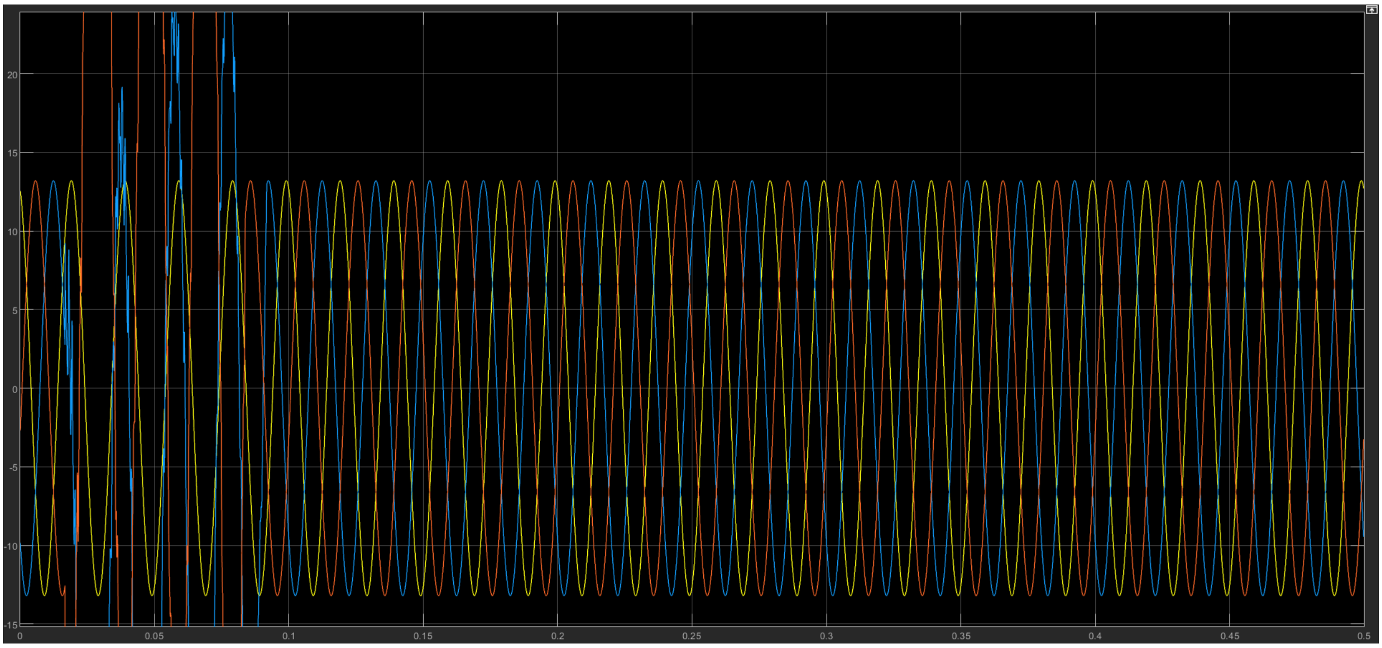


Figure 24: Output graph: current among Phase B, C and Ground

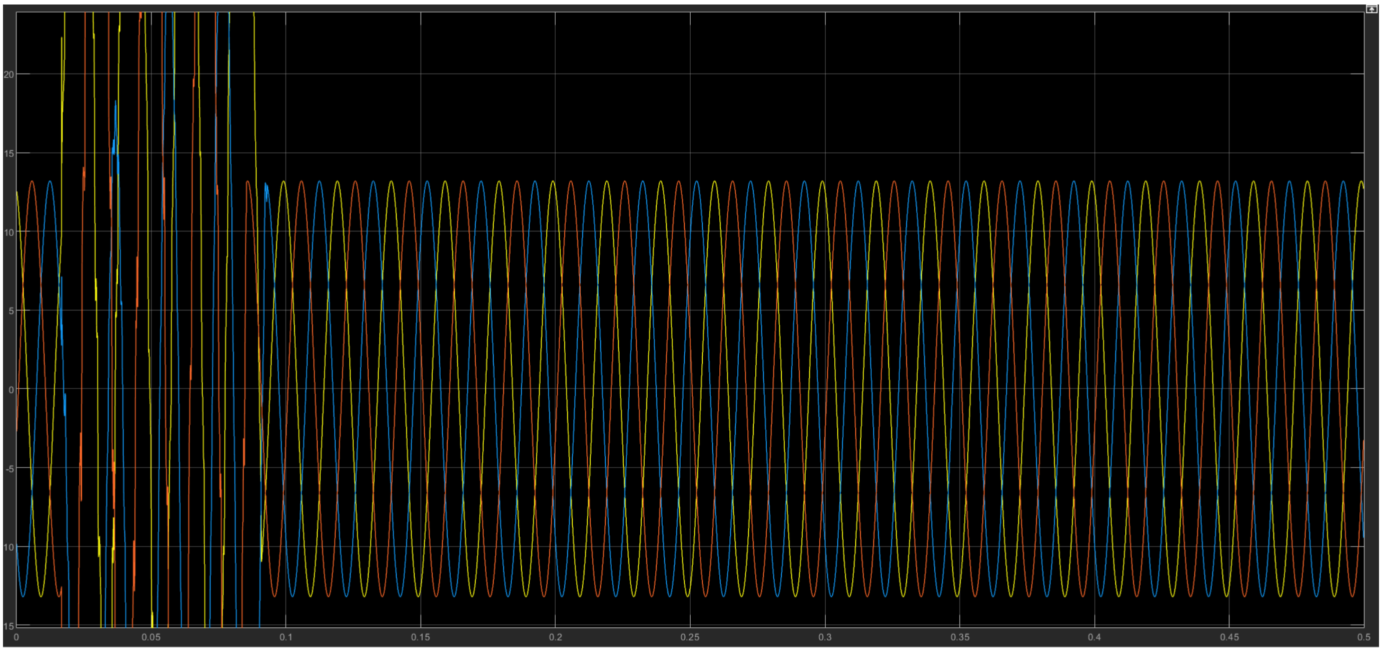


Figure 25: Output graph: current among Phase A, B, C and Ground

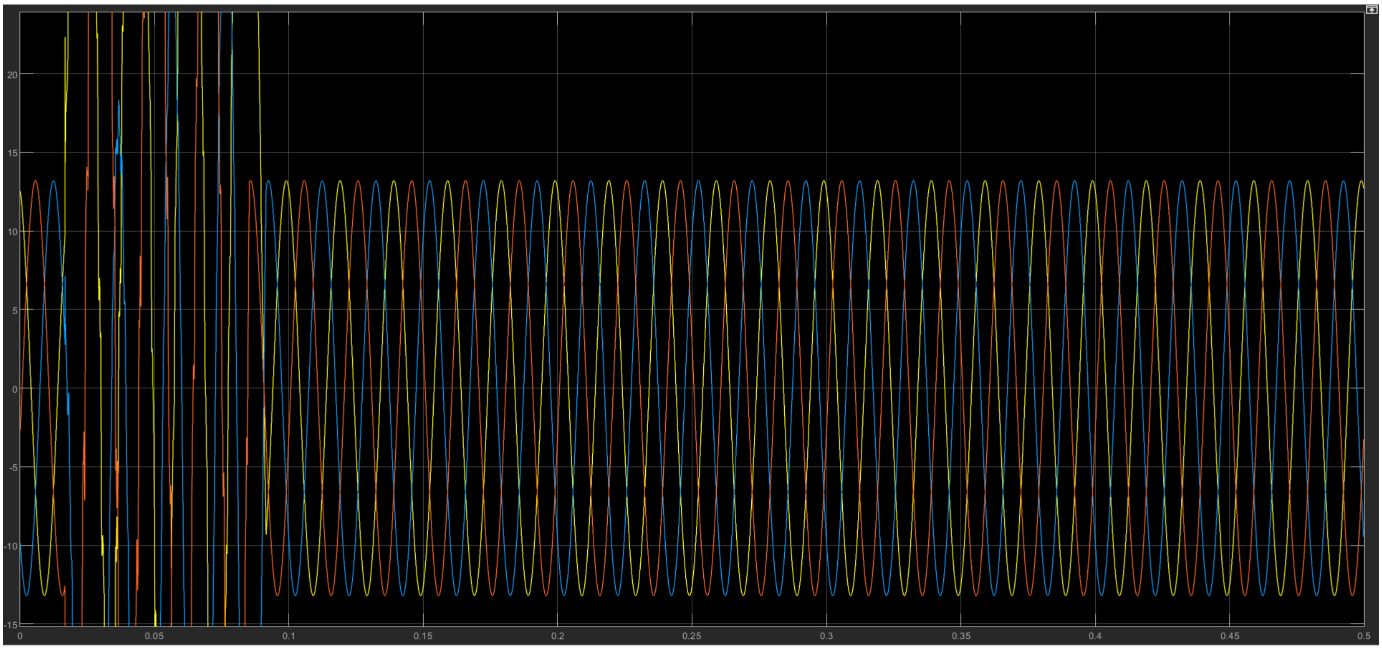


Figure 26: Output graph: current among Phase A, B and C

Chapter 3

Chapter 3: Methodology

Classifying the Electrical Fault → Classification based ML Algorithm

Let's break down how classification models can differentiate between symmetrical and unsymmetrical faults and aid in rapid response:

3.1. Training Data Preparation:

To train a classification model, a dataset generated from *Simulink* will need to be prepared.

- ➔ Need to check whether it got any missing value or not, if yes, then we can remove the row from the spreadsheet
- ➔ In X-axis, we got only 6 parameters; voltage and current for each of the three individual phases
- ➔ We would create another column named ‘Classifier’ which will be our ‘Target Column’, it will only carry the existing fault that we already predetermined through our 70% data for each fault

3.2 Feature Extraction and Selection:

Before training the model, relevant features that characterize different fault types need to be extracted from the dataset. In this case, feature selection methods will be used to figure out the most relevant features for fault classification purposes.

3.3 Model Training:

Following the preparation of the dataset and the extraction of its features, the machine learning model is trained through the use of supervised learning techniques. DT, SVM, k-NN,and Ensemble methods are some of the classification algorithms that are employed in this job. The labeled examples in the dataset are used to teach the model how to translate input features to corresponding fault kinds during training.

3.4 Fault Type Identification:

After the model training, it will be sent to identify and classify faults in real-time operation. When new data from the power system is fed into the model, it applies the learned patterns to predict the type of fault present. For example, if the features extracted from the data indicate symmetrical variations in voltage and current, the model may classify it as a symmetrical fault. Conversely, if asymmetrical patterns are detected, the model may classify it as an unsymmetrical fault.

3.5 Response and Action:

Once the fault type is identified, appropriate response and action will be initiated by protection and control systems. For example, protective relays can trip circuit breakers to isolate the faulty section of the network and prevent further damage. Rapid and accurate fault identification facilitated by machine learning models enables swift response, minimizing downtime and enhancing system reliability.

Overall, machine learning-based fault type identification offers a data-driven approach that will gradually improve the efficiency as well as effectiveness of fault detection and response in power systems. By leveraging the power of advanced algorithms and real-time data analysis, utilities can enhance the resilience and stability of their electrical networks.

3.6 Classification based Machine Learning Algorithms:

1. Logistic Regression:

- Binary classification tasks, like fault/non-fault classification, are best performed using the linear model known as logistic regression
- Through the logistic function, which associates the properties of the input with a probability value within 0 to 1, it predicts the neighboring inputs of a given input is a member of a specific class
- The classification of new instances based on feature values is made possible by logistic

regression, which fits probabilities to create a decision border between the two groups

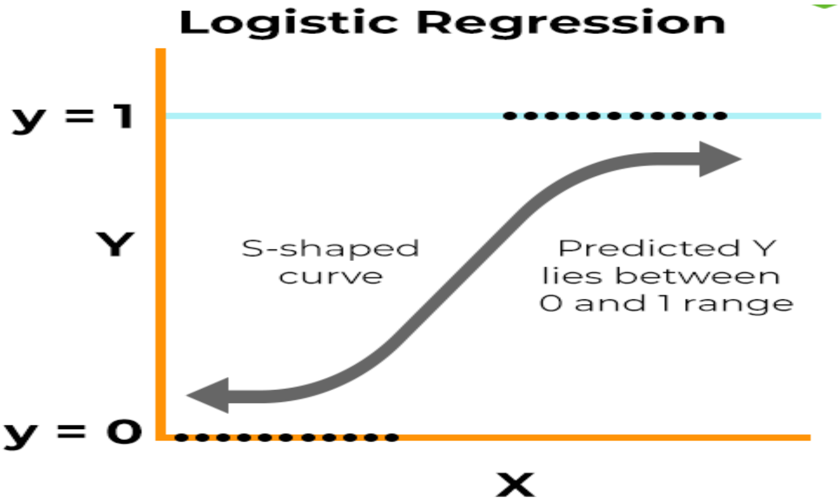


Figure 27: Logistic Regression

2. Decision Trees:

- To classify instances, decision trees employ a hierarchical framework of binary judgements depending on input features
- They divide the feature space into regions, where a class label is represented by a leaf node and an interior node denotes a choice based on feature
- Decision trees work well for fault classification problems when feature correlations are nonlinear because they are interpretable and capable of capturing intricate interactions between features

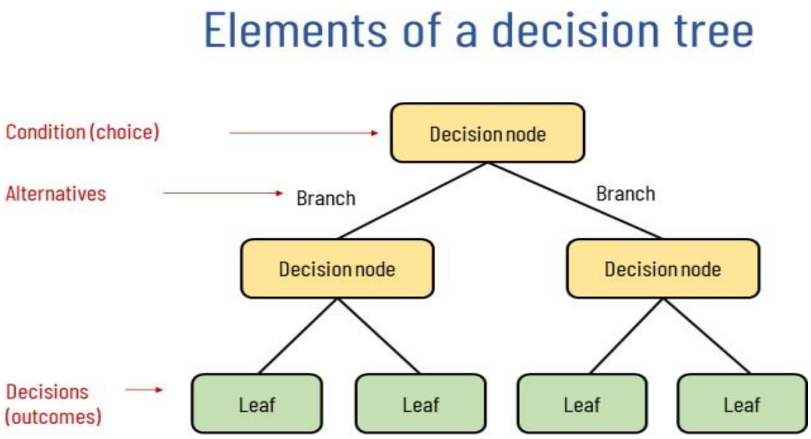


Figure 28: Decision Trees

3. Random Forest:

- Random multiple decision tress considered/simulated simultaneously in random forest algorithms and this is an ensemble learning technique, to increase classification resilience and accuracy

- It creates a forest of decision trees by bootstrapping the training data and using feature randomness to grow each tree independently.

- Random Forest reduces the chances of overfitting and increases the overall performance by aggregating the predictions of multiple tree data, making it effective for fault classification tasks with high-dimensional feature spaces.

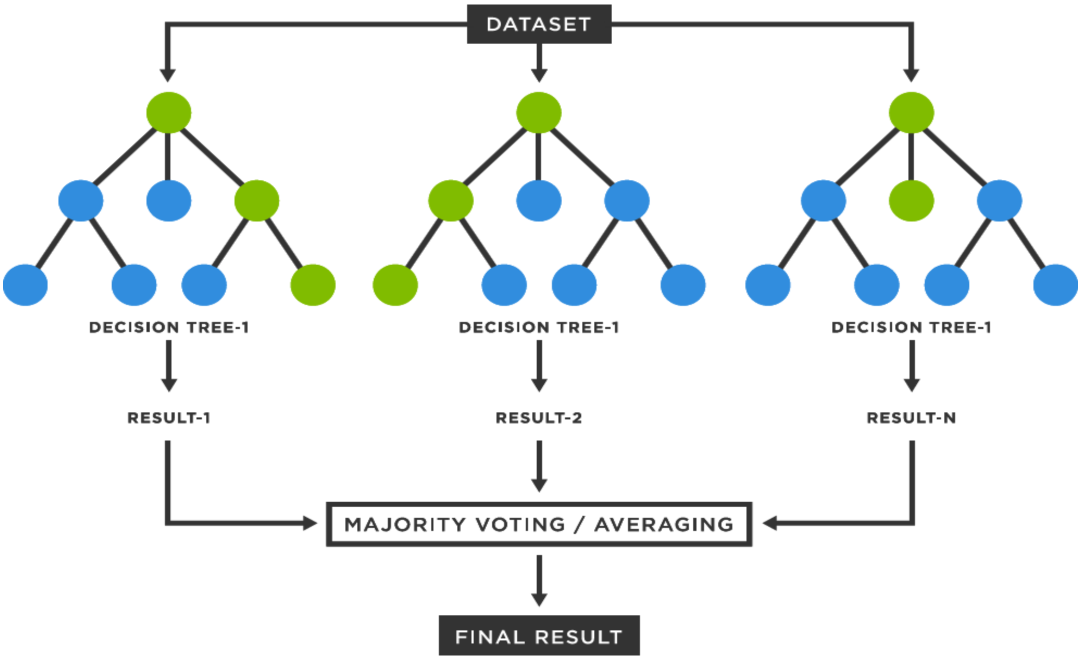


Figure 29: Random Forest

4. Support Vector Machines (SVM):

- SVM creates a multidimensional hyperplane to divide classes with the greatest margin
- Employing a kernel function; ensures to get the accurate map input information which will be later used into a higher-dimensional space. Also, it performs well on both linear and non-linear classification problem
- As SVM handles high-dimensional feature boundaries and nonlinear decision boundaries, they are useful for fault classification in power systems

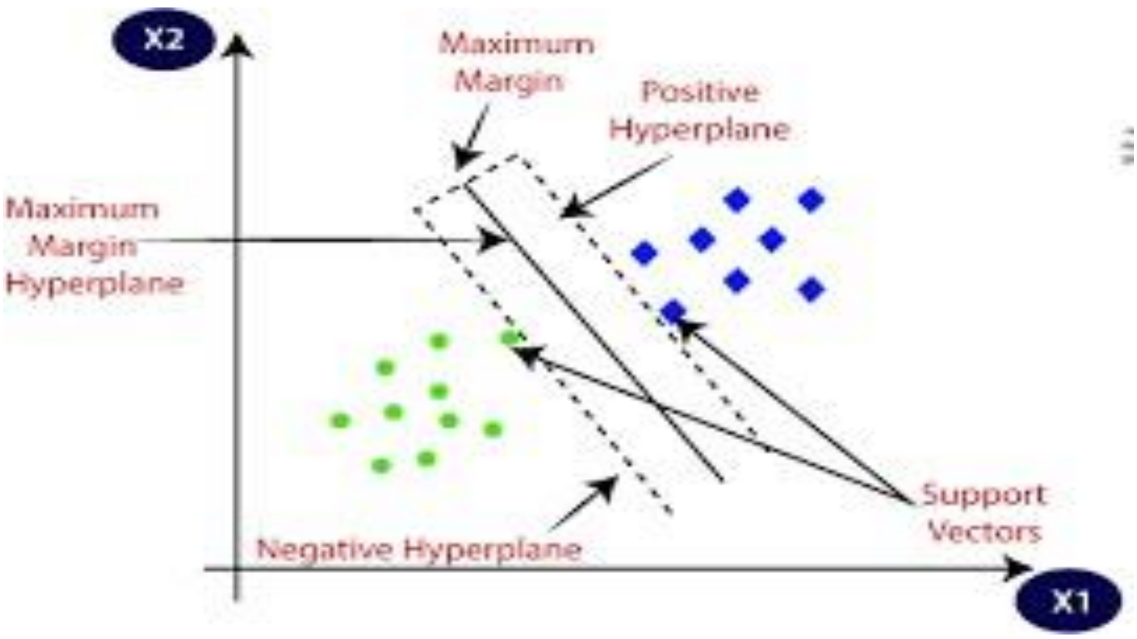


Figure 30: Support Vector Machines (SVM)

5. k-Nearest Neighbors (k-NN):

- Clustered points are classified according to the most frequent class of their k-nearest neighbors within their feature boundary
- It is a method where it's unable to make sheer assumptions about the dispersion of underlying data
- k-NN is easy-going and intuitive, making it suitable for fault classification tasks where the decision boundary is primarily linear or unclear.

Figure 31: k-Nearest Neighbors (k-NN)

6. Naive Bayes:

- Using the Bayes' theorem, Naive Bayes predicts class probabilities under the assumptions that features are independent
- Naive Bayes can function effectively in practice despite its simplifying assumption, particularly for text classification and other high-dimensional data.
- It is appropriate for fault classification tasks with a certain number of training samples since it’s computationally efficient and requires little training data

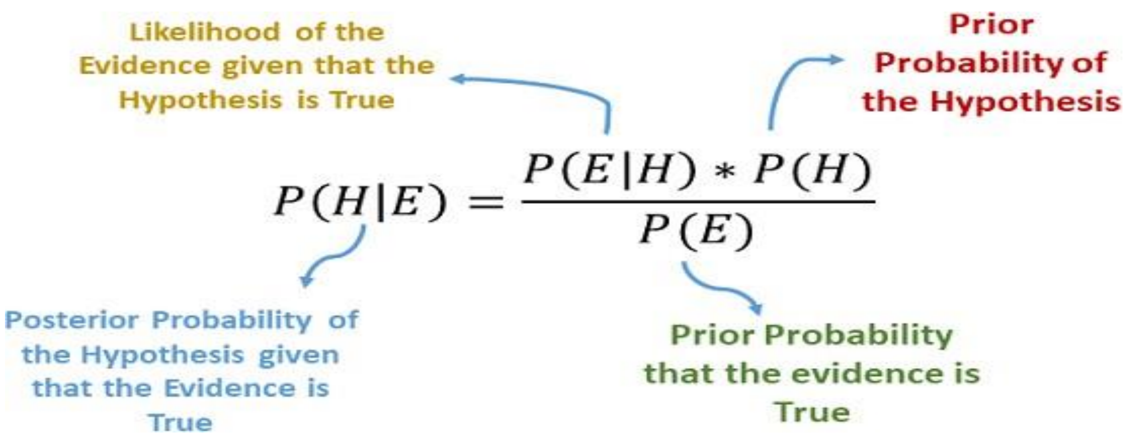


Figure 32: Naive Bayes

7. Neural Networks:

- Through backpropagation, neural networks-which are multi-layered networks-learn intricate correlations between characteristics and target classes
- Their effectiveness in fault classification tasks involving complicated feature interactions stems from their chance to detect intricate patterns and non-linear connections in the data
- Neural networks can overfit, though, and require a large amount of data and processing capability to train up

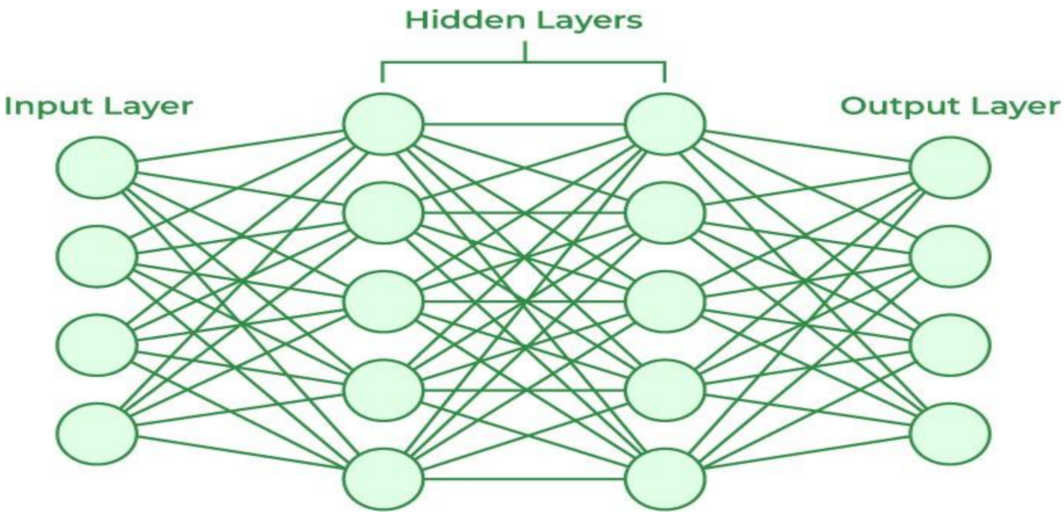
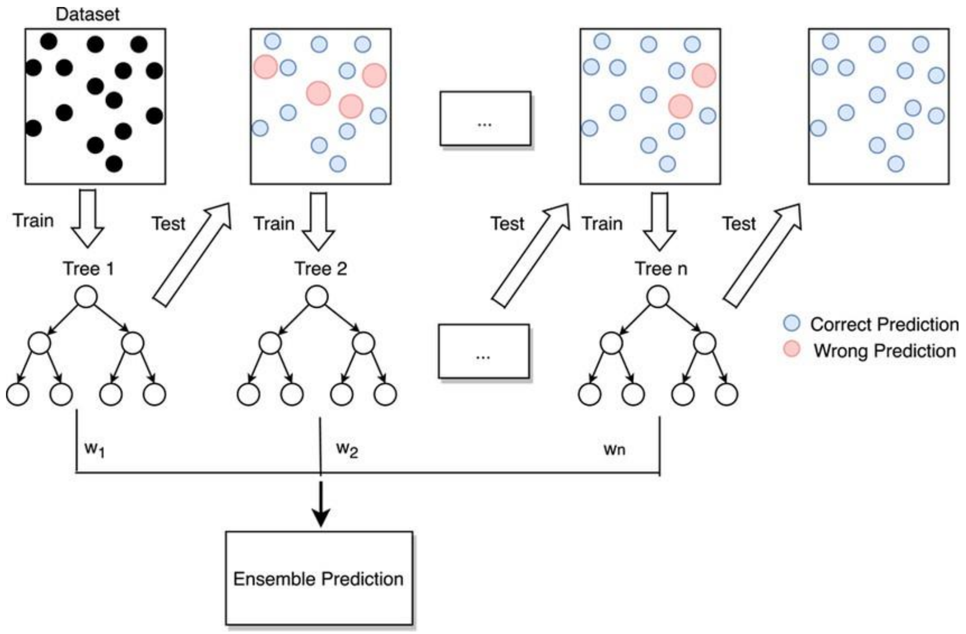


Figure 33: Neural Networks

8. Gradient Boosting:

- Gradient boosting is just another ensemble method like random forest which uses the strategy of minimizing a loss function iteratively that makes weak learners into a strong learner.
- It executes decision trees in a sequential manner, improving performance with each tree that fixes the flaws of the one before it

- Because gradient boosting can handle heterogeneous data types and is protective to overfitting, it’s kind of wise choice for fault classification problems involving a variety of feature sets



-

Figure 34: Gradient Boosting

9. MLP Classifier:

- The Multilayer Perceptron (MLP) is another example of neural network model which deals multiple layers of nodes with nonlinear activation functions that is used for classification tasks

- MLP Classifier is trained by minimizing a loss function using backpropagation, modifying the network’s weights and biases to increase classification accuracy.

- MLP Classifier is appropriate for fault classification problems involving intricate feature interactions since it can recognize complicated linkages in the data and adjust to non-linear decision bounds

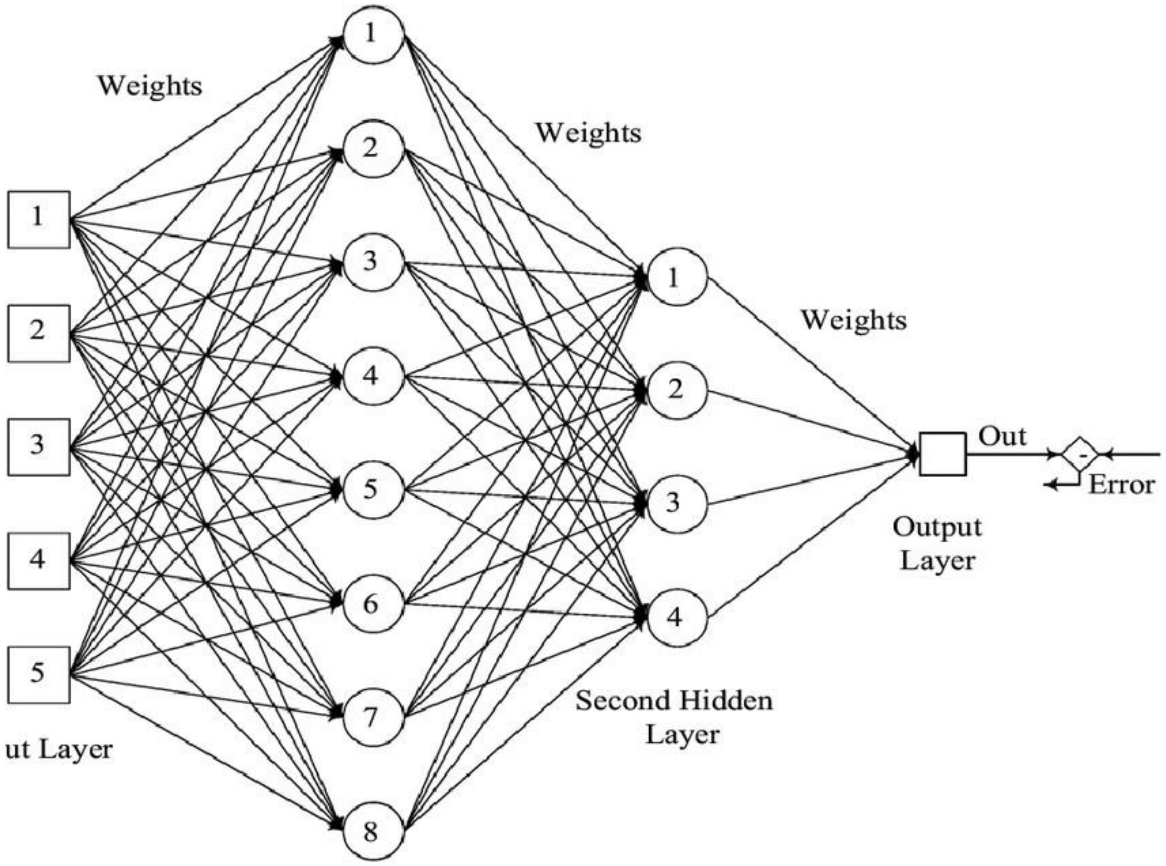


Figure 35: MLP Classifier

In conclusion, each of these machine learning methods has advantages and disadvantages, and whether or not they are appropriate for classifying faults in power systems will rely on a number of variables, including the amount of data, its complexity, and the available computing capacity

KNN Steps:

1.Import Data (from sklearn.neighbours import KneighboursClassifier)

2.knn = KneighboursClassifier(n_neighbours=3)

3. knn.fit (x_train,y_train)

4.predictx = knn.predict(x_test) 5. knn.scrore(x_test,y_test)

Logistic Regression Steps:

1.Import Data (from sklearn.linear_model import LogisticRegression)

2.LR = LogisticRegression(C=10)

3. LR.fit (x_train,y_train)

4. LR.scrore(x_test,y_test)

SVM Steps:

1.Import Data (from sklearn.svm import LinearSVC)

2.SVC = LinearSVC()

3. SVC.fit (x_train,y_train)

4. SVC.scrore(x_test,y_test)

Decision Tree Steps:

1.Import Data (from sklearn.tree import DecisionTreeClassfier)

2.tree = DecisionTreeClassifier()

3. tree.fit (x_train,y_train)

4. tree.scrore(x_test,y_test)

Random Forest Steps:

1.Import Data (from sklearn.ensemble import RandomForestClassfier)

2.forest = RandomForestClassifier()

3. forest.fit (x_train,y_train)

4. forest.scrore(x_test,y_test)

Gradient Boosting Steps:

1.Import Data (from sklearn.ensemble import GradientBoostingClassfier)

2.gbrt = GradientBoostingClassifier()

3. gbrt.fit (x_train,y_train)

4. gbrt.scrore(x_test,y_test)

MLP Classifiers Steps:

1. Import Data (from sklearn.neural_network import MLPClassfier)

2.mlp = MLPClassifier(solver=’lbgfs’).fit(x_train,y_train)

3. mlp.scrore(x_test,y_test)

4. mlp2 = MLPClassifier(solver = ‘lbfgs’, activition = ‘tanh’, hidden_layer_sizes = [10,10]).fit

5. mlp2.score(x_test,y_test)

MLP Classifiers Steps:

1. Import Data (from sklearn.neural_network import MLPClassfier)

2.mlp = MLPClassifier(solver=’lbgfs’).fit(x_train,y_train)

3.mlp.scrore(x_test,y_test)

4.mlp2 = MLPClassifier(solver = ‘lbfgs’, activition = ‘tanh’, hidden_layer_sizes = [10,10]).fit

5.mlp2.score(x_test,y_test)

Chapter 4

Chapter 4: Demonstration of Outcome Based Education (OBE)

The Outcome Based Education (OBE) is a pedagogical approach that emphasizes the attainment of specific learning outcomes by students. OBE shifts the focus from traditional teaching methodologies centered around content delivery to a student-centric approach that prioritizes measurable outcomes and competencies. In the realm of electrical engineering, where innovation and efficiency are paramount, integrating OBE principles is crucial for producing highly skilled professionals capable of addressing complex challenges in the field.

4.1 Introduction

The industry witnessed a significant transformation with the advent of Outcome Based Education (OBE).

In the context of fault detection in electrical systems using machine learning algorithms, Outcome Based Education plays a pivotal role in shaping the educational landscape. By aligning educational objectives with industry demands and real-world applications, OBE ensures that students not only grasp theoretical concepts but also develop practical skills necessary for tackling modern engineering challenges. This section provides an overview of how Outcome Based Education principles are integrated into the methodology of fault detection in electrical systems using machine learning algorithms. Through the lens of OBE, this research aims to not only advance the field of fault detection but also contribute to the holistic development of engineering students equipped with the expertise to innovate and thrive in the ever-evolving electrical engineering landscape.

4.2 Course Outcomes (COs) Addressed

The following table shows the COs addressed in EEE 4700 for Project and Thesis.

COs	CO Statement	POs	Put Tick (√) EEE 4700
CO1	Identify a contemporary real-life problem related to electrical and electronic engineering by reviewing and analyzing existing research works.	PO2	√
CO2	Determine functional requirements of the problem considering feasibility and efficiency through analysis and synthesis of information.	PO4	√
CO3	Select a suitable solution and determine its method considering professional ethics, codes and standards.	PO8	√
CO4	Adopt modern engineering resources and tools for the solution of the problem.	PO5	√
CO5	Prepare management plan and budgetary implications for the solution of the problem.	PO11	√
CO6	Analyze the impact of the proposed solution on health, safety, culture and society.	PO6	√
CO7	Analyze the impact of the proposed solution on environment and sustainability.	PO7	√
CO8	Develop a viable solution considering health, safety, cultural, societal and environmental aspects.	PO3	√
CO9	Work effectively as an individual and as a team member for the accomplishment of the solution.	PO9	√
CO10	Prepare various technical reports, design documentation, and deliver effective presentations for demonstration of the solution.	PO10	√
CO11	Recognize the need for continuing education and participation in professional societies and meetings.	PO12	√

4.3 Aspects of Program Outcomes (POs) Addressed

The following table shows the aspects addressed for certain Program Outcomes (POs) addressed in EEE 4700 for Project and Thesis.

	Statement	Different Aspects	Put Tick (√)
PO2	**Problem analysis:** Identify, formulate, research literature and analyze complex electrical and electronic engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences and engineering sciences.		√

PO4	**Investigation:** Conduct investigations of complex electrical and electronic engineering problems using research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of information to provide valid conclusions.	Design of experiments	√
		Analysis and interpretation of data	√
		Synthesis of information	√
PO6	**The engineer and society:** Apply reasoning informed by contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to professional engineering practice and solutions to complex electrical and electronic engineering problems.	Societal	√
		Health	√
		Safety	√
		Legal	√
		Cultural	√
PO7	**Environment and sustainability:** Understand and evaluate the sustainability and impact of professional engineering work in the solution of complex electrical and electronic engineering problems in societal and environmental contexts.	Societal	√
		Environmental	√
PO8	**Ethics:** Apply ethical principles embedded with religious values, professional ethics and responsibilities, and norms of electrical and electronic engineering practice.	Religious values	√
		Professional ethics and responsibilities	√
		Norms	√
PO9	**Individual work and teamwork:** Function effectively as an individual, and as a member or leader in diverse teams and in multi-disciplinary settings.	Diverse teams	√
		Multi-disciplinary settings	√
PO10	**Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.	Comprehend and write effective reports	√
		Design documentation	√
		Make effective presentations	√
		Give and receive clear instructions	√
PO11	**Project management and finance:** Demonstrate knowledge and understanding of engineering management principles and economic decision-making and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.	Engineering management principles	√
		Economic decision-making	√
		Manage projects	√
		Multidisciplinary environments	√

4.4 Knowledge Profiles (K3 – K8) Addressed

The following table shows the Knowledge Profiles (K3 – K8) addressed in EEE 4700 for Project and Thesis.

K	Knowledge Profile (Attribute)	Put Tick (√)
K3	A systematic, theory-based formulation of engineering fundamentals required in the engineering discipline	√
K4	Engineering specialist knowledge that provides theoretical frameworks and bodies of knowledge for the accepted practice areas in the engineering discipline; much is at the forefront of the discipline	√
K5	Knowledge that supports engineering design in a practice area	√
K6	Knowledge of engineering practice (technology) in the practice areas in the engineering discipline	√
K7	Comprehension of the role of engineering in society and identified issues in engineering practice in the discipline: ethics and the engineer's professional responsibility to public safety; the impacts of engineering activity; economic, social, cultural, environmental and sustainability	√
K8	Engagement with selected knowledge in the research literature of the discipline	√

4.5 Use of Complex Engineering Problems

Approach:

➔ For Problem Definition, we can thoroughly delineate the difficulties associated with electrical fault diagnosis.

➔ Data-driven modeling involves the development of machine learning models through the utilization of extensive historical fault data.

➔ Promoting interdisciplinary collaboration by facilitating cooperation between professionals in the fields of machine learning and electrical engineering.

➔ Efficient Implementation approach involves the seamless integration of models into the electrical system, which is achieved through the successful implementation of project management strategies.

➔ Continuous improvement involves the critical evaluation of real-world performance, the collection of feedback, and the iterative optimization of models.

➔ Documentation and Knowledge Transfer: Efficiently document the entire process and ease the transfer of knowledge to provide continuous assistance.

4.6 Socio-Cultural, Environmental, And Ethical Impact

The Socio-Cultural Impact in the Context of Bangladesh:

An Analysis of Workforce Dynamics:

The examination of the prospective influence of the project on the local labor force is now underway. The present study aims to evaluate the requirement of specialized skills and training in the context of introducing machine learning in fault analysis, and to examine the potential impact of this on employment dynamics within the country of Bangladesh.

The Availability and Utilization of Technology:

The analysis primarily centers on the project's measures to guarantee fair and inclusive distribution of its benefits among many social and economic groups within the context of Bangladesh. Consideration is directed towards the possible obstacles associated with accessibility and affordability.

Concept of Cultural Acceptance:

It refers to the societal recognition and embrace of diverse cultural practices, beliefs, and values. It encompasses willingness. This research is currently being conducted to examine the extent to which the project is in accordance with the cultural values and traditions of Bangladesh. The integration of cultural issues into the implementation approach is being considered in order to enhance acceptance.

Environmental Impact within the specific context of Bangladesh:

Energy Efficiency:

The focal point of the discussion pertains to the project's contribution towards enhancing energy efficiency within the specific context of Bangladesh's energy challenges. Significant importance is attributed to the potential positive implications in terms of diminishing energy usage and mitigating environmental effect.

The Utilization of Resources:

This study examines the feasibility and long-term viability of utilizing machine learning algorithms in Bangladesh, taking into account the accessibility and durability of resources required for their implementation and maintenance. The issues pertaining to limited availability of resources and the resulting environmental impacts are being considered and resolved.

Concept of Long-term Sustainability:

It refers to the ability of a system, organization, or society to maintain its operations and meet its needs. The focus of the discourse revolves around the manner in which the project aligns with Bangladesh's objectives pertaining to enduring sustainability. The evaluation takes into account its alignment with the environmental policies and objectives of the country.

The Ethical Implications within the Context of Bangladesh:

The topic of data privacy is of significant importance in contemporary society.

The significance of safeguarding data privacy within the context of Bangladesh is underscored. The discourse examines the implemented strategies aimed at safeguarding user data and guaranteeing adherence to regional data protection legislations.

The Influence of Bias on the Concept of Fairness:

This analysis takes into account the matter of bias within machine learning algorithms, with specific regard to the heterogeneous population of Bangladesh. The discourse emphasizes the measures implemented to address prejudice and uphold impartiality in the outcomes of fault analysis.

Concepts of Transparency and Accountability:

They are essential in many fields, including as public administration, business, and governance. Openness and accessibility of information are key components of transparency, as they guarantee its relevance. The discourse pertains to the level of openness exhibited by machine learning algorithms and the degree of accountability the system assumes in instances of faults or failures. This paper examines the local processes that promote accountability and openness in Bangladesh, while adhering to the country's legal and ethical requirements.

4.7 Attributes of Ranges of Complex Engineering Problem Solving (P1 – P7) Addressed

The following table shows the attributes of ranges of Complex Engineering Problem Solving (P1 – P7) addressed in EEE 4700 for Project and Thesis.

P	**Range of Complex Engineering Problem Solving**	**Put Tick (√)**
Attribute	Complex Engineering Problems have characteristic P1 and some or all of P2 to P7:	
Depth of knowledge required	**P1:** Cannot be resolved without in-depth engineering knowledge at the level of one or more of K3, K4, K5, K6 or K8 which allows a fundamentals-based, first principles analytical approach	√
Range of conflicting requirements	**P2:** Involve wide-ranging or conflicting technical, engineering and other issues	√
Depth of analysis required	**P3:** Have no obvious solution and require abstract thinking, originality in analysis to formulate suitable models	√
Familiarity of issues	**P4:** Involve infrequently encountered issues	√ √
Extent of applicable codes	**P5:** Are outside problems encompassed by standards and codes of practice for professional engineering	
Extent of stakeholder involvement and conflicting requirements	**P6:** Involve diverse groups of stakeholders with widely varying needs	√
Interdependence	**P7:** Are high level problems including many component parts or sub-problems	√

4.8 Attributes of Ranges of Complex Engineering Activities (A1 – A5) Addressed

The following table shows the attributes of ranges of Complex Engineering Activities (A1 – A5) addressed in EEE 4700 for Project and Thesis.

A	**Range of Complex Engineering Activities**	**Put Tick (√)**
Attribute	Complex activities mean (engineering) activities or projects that have some or all of the following characteristics:	
Range of resources	**A1:** Involve the use of diverse resources (and for this purpose resources include people, money, equipment, materials, information and technologies)	√
Level of interaction	**A2:** Require resolution of significant problems arising from interactions between wide-ranging or conflicting technical, engineering or other issues	√
Innovation	**A3:** Involve creative use of engineering principles and research-based knowledge in novel ways	√
Consequences for society and the environment	**A4:** Have significant consequences in a range of contexts, characterized by difficulty of prediction and mitigation	√
Familiarity	**A5:** Can extend beyond previous experiences by applying principles based approaches	√

Chapter 5

Chapter 5: Results and Discussions

5.1 Accuracy

Classification Model	Accuracy	Classification Model	Accuracy
KNN	0.746809450565434 **(74.68%)**	**Random Forest**	0.947721573760954 **(94.77%)**
Logistic Regression	0.515629158939675 **(51.56%)**	**Gradient Boosting**	0.72557060470528344 **(72.55%)**
SVM	0.46126823751863568 **(46.12%)**	**MLP Classifiers**	0.70636479037125923 **(70.63%)**
Decision Tree	0.8641381949747282 **(86.413%)**		

Table 1: Accuracy data for different classification-based machine learning algorithms

Small portion of values from Datasheet:

Input Data						Target Column
Va	Vb	Vc	Ia	Ib	Ic	Classifier
1.64228083	-0.4383	-1.20398	12.53622	-9.84595	-2.69027	Phase AG
1.694759	-0.72398	-0.97078	11.59991	-11.2543	-0.4699405	Phase BG
1.676594	-0.591	-1.08559	12.09179	-10.6296	-1.46224	Phase CG
1.6933406	-10.4041	-1.83251	12.23659	-10.4041	-1.83251	Phase AB
1.6933407	0.15196	-1.0098657	11.2903	-11.5063	0.15196	Phase BC
1.697144	-0.75286	-0.94428	11.47914	-11.3823	-0.09684	Phase AC
1.693341	1.693341	1.693341	11.71656	-11.1223	-0.59424	Phase ABG
1.681679	1.681679	1.681679	11.98988	-10.7752	-1.21472	Phase BCG

Table 2: small portion of fault values from generated datasheet

Limitations

1. This a Classification based prediction model which can’t predict the exact location of fault
2. The output accuracy varies with the data, so different data set gives different sort of result

Overview of the Work

This work explores the application of machine learning algorithms for fault analysis in electrical power systems, focusing on overhead transmission lines. Faults in transmission lines can lead to severe disruptions and pose safety hazards, making accurate and timely fault identification essential. By leveraging advanced machine learning techniques, such as LR, DT, Random Forest, SVM, k-NN, Naive Bayes, NN, Gradient Boosting, and MLP classifiers, we aim to enhance the precision and speed of fault detection and classification.

The research emphasizes the importance of understanding several types of faults, including LG, LL, DLG, and three-phase faults, and their causes, such as environmental factors, equipment failures, human errors, system overloads, aging infrastructure, and animal interference. Through a detailed analysis of fault data and feature extraction, machine learning models are trained to recognize patterns and classify faults accurately.

By implementing these models, utilities can improve system reliability and safety, minimize economic losses due to equipment damage and outages, and optimize maintenance and operations. The study highlights the potential of machine learning in supporting the modernization of power grids, integrating renewable energy sources, and ensuring compliance with regulatory requirements.

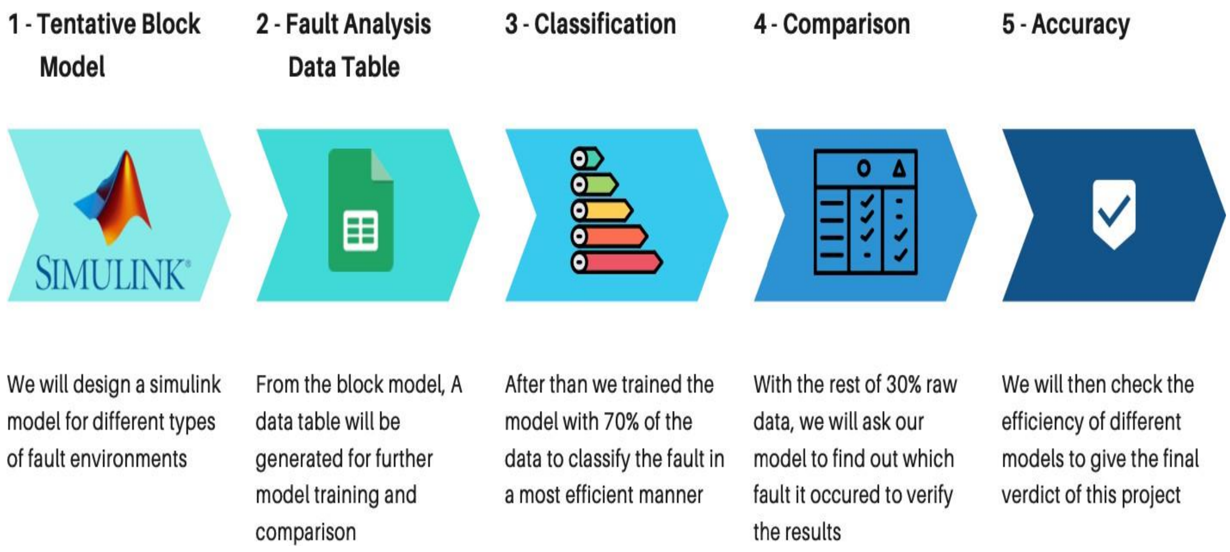


Figure 36: Process implemented

Future Work

To proceed with this fault detection journey, we would like to focus on some applications if required time is given:

1. After the fault type detection, it's better to go for more advanced task, which is to locate the exact fault based on the data with the help of deep network method or any advanced one
2. Automating the whole system, so that it gives more detailed information about the fault

References

[1] T. Goswami and U. B. Roy, "Predictive Model for Classification of Power System Faults using Machine Learning," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, pp. 1881-1885, doi: 10.1109/TENCON.2019.8929264. keywords: {Circuit faults;Power transmission lines;Machine learning;Task analysis;Power system stability;Classification algorithms;fault classification;power system;supervised machine learning}

[2] Yi Lu Murphey, M. A. Masrur, ZhiHang Chen and Baifang Zhang, "Model-based fault diagnosis in electric drives using machine learning," in IEEE/ASME Transactions on Mechatronics, vol. 11, no. 3, pp. 290-303, June 2006, doi: 10.1109/TMECH.2006.875568.
keywords: {Fault diagnosis;Machine learning;Power system modeling;Inverters;Electrical fault detection;Neural networks;Electric motors;Electronics industry;Automotive engineering;Power electronics;Electric drives;electric vehicle;fieldoriented control (FOC);fuzzy techniques;hybrid vehicle;inverter;machine learning;modelbased diagnostics;motor;neural network;power electronics},

[3] Chen, Zhicong, et al. "Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents." *Energy conversion and management* 178 (2018): 250264.

[4] Goni, Md Omaer Faruq, et al. "Fast and accurate fault detection and classification in transmission lines using extreme learning machine." *e-Prime-Advances in Electrical Engineering, Electronics and Energy* 3 (2023): 100107.

[5] Veeramsetty, V., Reddy, K.R., Santhosh, M. *et al.* Short-term electric power load forecasting using random forest and gated recurrent unit. *Electr Eng* **104**, 307–329 (2022). https://doi.org/10.1007/s00202-021-01376-5

[6] Li, Zhangling, et al. "A building electrical system fault diagnosis method based on random forest optimized by improved sparrow search algorithm." *Measurement Science and Technology* 35.5 (2024): 055110.

[7] D. Chakraborty, U. Sur and P. K. Banerjee, "Random Forest Based Fault Classification Technique for Active Power System Networks," 2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), Bangalore, India, 2019, pp. 1-4, doi: 10.1109/WIECON-ECE48653.2019.9019922. keywords: {Vegetation;Forestry;Support vector machines;Power systems;Neural networks;Power electronics;Wavelet transforms;Random Forest Tree;fault Classification;power system networks;active distribution network}

[8] Fonseca, G.A., Ferreira, D.D., Costa, F.B. *et al.* Fault Classification in Transmission Lines Using Random Forest and Notch Filter. *J Control Autom Electr Syst* **33**, 598–609 (2022). https://doi.org/10.1007/s40313-021-00844-4