A comparative study on artificial intelligence-based methods for fault detection, classification, and localization in distribution lines

Miss Nanda Kumari

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering in Electrical Engineering
Department of Electrical Engineering
FACULTY OF ENGINEERING
Chulalongkorn University
Academic Year 2022

การศึกษาเปรียบเทียบวิธีเชิงปัญญาประดิษฐ์สำหรับการตรวจจับ การจำแนกประเภท และการระบุตำแหน่งความผิดพร่องในสายจำหน่าย

น.ส.นานดา คูมารี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565

Thesis Title	A comparative study on artificial intelligence-based methods for fault detection, classification, and localization in distribution lines
By	Miss Nanda Kumari
Field of Study	Electrical Engineering
Thesis Advisor	Channarong Banmonkol

Accepted by the FACULTY OF ENGINEERING, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Engineering

.... Dean of the FACULTY OF ENGINEERING

()

THESIS COMMITTEE

.... Chairman

(Att Phayomhom)

.... Thesis Advisor

(Channarong Banmonkol)

.... Examiner

(NAEBBOON HOONCHAREON)

.... Examiner

(Wijarn Wangdee)

นานดา คูมารี : การศึกษาเปรียบเทียบวิธีเชิงปัญญาประดิษฐ์สำหรับการตรวจจับ การจำแนกประเภท และการระบุตำแหน่งความผิดพร่องในสายจำหน่าย. (A comparative study on artificial intelligence-based methods for fault detection, classification, and localization in distribution lines) อ.ที่ปรึกษาหลัก : ชาญณรงค์ บาลมงคล

ในช่วงหลายปีที่ผ่านมา การเรียนรู้ของเครื่องแบบมีผู้สอน (SML) ได้แสดงให้เห็นว่ามีประสิทธิภาพในการระบุรูปแบบในชุดข้อมูลและคาดการณ์ผลลัพธ์ งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาอัลกอริทึมที่ใช้ การจำแนกด้วย SML และความสามารถของสมการถดถอย เพื่อจำแนกประเภทและระบุตำแหน่งของความผิดพร่องที่เกิดกับสายจำหน่ายไฟฟ้าอย่างมีความแม่นยำ อัลกอริทึมที่นำเสนอใช้ค่าประสิทธิผลและค่าองค์ประกอบสมมาตรลำดับศูนย์ของกระแสไฟฟ้าและแรงดันไฟฟ้าที่วัดได้จากปลายข้างหนึ่งของสายจำหน่ายเป็นข้อมูลขาเข้าแล้วส่งประเภทของความผิดพร่องเป็นข้อมูลขา การประเมินประสิทธิผลของอัลกอริทึมทำโดยการจำลองระบบไฟฟ้า IEEE 14 บัสในโปรแกรม MATLAB แล้วสร้างเหตุการณ์ผิดพร่องประเภทต่างๆ ที่ตำแหน่งและความต้านทานผิดพร่องที่หลากหลายเพื่อเก็บเป็นฐานข้อมูล อัลกอริทึมจะใช้ฐานข้อมูลและเทคนิค SML หลายประเภทได้แก่ การวิเคราะห์จำแนกเชิงเส้น (LDA) เครื่องเวกเตอร์สนับสนุน (SVM) วิธีเพื่อนบ้านใกล้ที่สุด (KNN) รวมทั้งวิธีการถดถอยแบบกำลังสองน้อยที่สุด (LMS) เพื่อเปรียบเทียบความสามารถในการจำแนกประเภทและตำแหน่งของการเกิดความผิดพร่อง นอกจากนี้ยังมีการศึกษาความอ่อนไหวต่อตัวแปรในระบบไฟฟ้า ได้แก่ ความไม่แน่นอนของหม้อแปลงเครื่องมือวัด การมีเครื่องกำเนิดไฟฟ้าหรือสายจำหน่ายหลุดออกจากระบบไฟฟ้า

สาขาวิชา วิศวกรรมไฟฟ้า ลายมือชื่อนิสิต ....

ปีการศึกษา 2565 ลายมือชื่อ อ.ที่ปรึกษาหลัก ....

6470201921 : MAJOR ELECTRICAL ENGINEERING

KEYWORD: SML, classification, regression, comparative, study

Nanda Kumari : A comparative study on artificial intelligence-based methods for fault detection, classification, and localization in distribution lines . Advisor: Channarong Banmonkol

In recent years, supervised machine learning (SML) has demonstrated its effectiveness in pattern recognition and outcome prediction within datasets. The objective of this research is to develop an algorithm that utilizes SML classification and regression equation capabilities to accurately classify and locate faults occurring in electricity distribution lines. The proposed algorithm takes the measured values of electrical current and voltage at one end of the distribution line as input data and outputs the type of fault. The algorithm evaluates its performance by simulating the IEEE 14-bus power system using MATLAB and generating various types of faults at different locations and with different fault resistances to create a comprehensive fault database. The algorithm can employ various types of SML techniques and approaches, including Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and the Least Mean Squares (LMS) regression method, to compare their abilities in classifying fault types and identifying fault locations. Additionally, the study investigates the system's vulnerability to variables such as uncertainty in transformer instrument measurements and the presence of generator or transmission line outages in the power system.

Field of Study: Electrical Engineering Student's Signature ....

Academic Year: 2022 Advisor's Signature ....

ACKNOWLEDGEMENTS

I would like to extend my heartfelt gratitude to the ASEAN/NON-ASEAN graduate scholarship program of Chulalongkorn University, Bangkok, Thailand. The scholarship came as an overwhelming opportunity to study in such a prestigious university which has been ranked No. 1 for consecutive years.

Further, I am immensely thankful to my advisor, Assistant Professor Chanarong Banmongkol, D.Eng. for his unwavering guidance and support throughout my journey at Chulalongkorn University. His invaluable insights and suggestions have played a vital role in the completion of this work. Arjan, I will remain forever grateful to you for your guidance.

I am indebted to all the faculty members who were involved in teaching the coursework. Their lectures, discussions, and academic resources have broadened my prospects and enhanced my research skills. I sincerely thank my family and friends for their constant support and understanding throughout this academic endeavor. Their belief in my ability and unconditional encouragement has always inspired me, reminding me of the importance of my patience and flexibility. I am deeply indebted to my study participants, whose willingness to share their experiences and insights made this research project possible. Their contributions provided valuable information and improved my findings. I am honored and humbled by their willingness to be a part of this study. I would like to acknowledge the tireless efforts and support of the library staff and the research support team in providing relevant materials, databases, and resource materials. Their assistance improved the accuracy and reliability of my in-depth literature review research.

Finally, I would like to thank all the individuals who have crossed my path on this learning journey and offered words of encouragement, advice, and support. Your presence made this experience even more meaningful, and I am grateful for the friends

I have made along the way.

Conclusion I would like to emphasize that it would not have been possible without the aforementioned people whose guidance, support, and belief in my abilities. I am forever grateful for their contributions and impact on my learning and personal growth.

Nanda Kumari

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

Chapter 1 Introduction

1.1 Background

Transmission and distribution lines around the globe are subjected to various faults due to the following reasons: switching surges, insulation failure, lighting, snow, conducting path failure, excessive growth in the right of way, falling of trees, creepers on the towers and poles, harsh climatic conditions, lightning surges, sudden changes in load parameters at the customer end leading to short circuits, under/over current, under voltage, unbalanced phase voltage, trespassing of animals and often surge leads to fire, loss of service and damages the equipment. The frequent faults cause wear and tear causing insulation failure, and the life span of the line and the substation equipment becomes a major concern and the same applies to Bhutan.

The distribution network in Bhutan is spread across three climatic zones. North with severe cold and snow during winter, central with heavy rainfall in summer and cold in winter. The south with extreme heat followed by thunderstorms in summer. When the fault occurs especially in such drastic climatic conditions, the information from the overcurrent relay regarding the fault doesn't suffice to figure out the location of the fault thus, the responsibility of the line restoration is entirely dependent on the operation and maintenance team. In the process of executing line restoration, the safety of the line operators comes at stake. This line fault diagnosis is of major concern to the utility.

Over the last two decades, studies on various fault detection studies and methods like phasor-based methods, traveling wave-based methods, and knowledge-based methods in terms of fault location have been extensively conducted and various researchers have made efforts to make the techniques less sophisticated for easy adoption.

The techniques developed have their drawbacks, mostly the cost being the main factor leading to difficulty in implementation in real-time and the errors and their complexities causing hindrance to the adoption in the field. Bhutan is a country that has not graduated from the list of least developed countries, thus upgrading the entire distribution network to a smart network in one go is not possible due to budget constrain followed by the unavailability proper communication network.

The frequent faults and delayed restoration cause wear and tear to the distribution components, especially the distribution transformer. With minimum to no communication link between the numerous distribution Transformers and substations, scheduling predictive and preventive maintenance is out of the question and the fault recognition and spot becomes very tedious. The responsibility of the line restoration is entirely dependent on the operation and maintenance team; thus, the restoration process becomes tiresome and time-consuming due to changes in climatic conditions like extreme rainfall with thunderstorms and freezing winter accompanied by snowfall and harsh heat of summer and autumn. In executing line restoration via the trial-and-error method, accidents occur both fatally and nonfatally.

1.2 Motivation

Bhutan Power Corporation’s (BPC) vision is to provide clients with cheap, sufficient, dependable, and high-quality electrical services, to fulfill this vision timely fault restoration is of utmost priority. To ease the fault restoration time despite the cost constraint communicable and non-communicable fault passage indicator(FPI) has been installed [1] at subtle locations however the reliability has not been assured as the FPI performance is largely affected by climatic conditions and battery durability, a good communication network. The table below shows the accident history of the last three years.

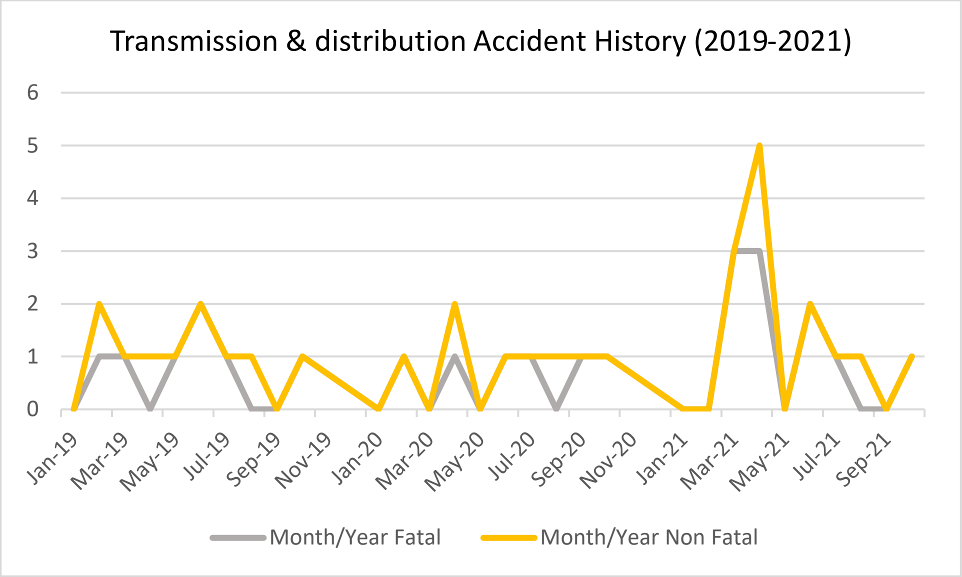


Figure 1 Distribution line Accident History(2019-2021)

From the table, it's clear that both fatal and nonfatal accidents are prominent thus the prevention of such accidents is vital for the country with such a small working population thus this research is started with the perspective of minimizing the accidents. Thus, as an alternative method to get information on the distribution during the fault condition, a simple algorithm is proposed that can directly make use of monitoring data from the SCADA (Supervisory Control and Data Acquisition) as the database and generate the required information when the fault occurs on the line. With the proper information regarding the faulted line, the operation and maintenance team can be mobilized accordingly to the faulted location to restore the line as soon as possible. In this regard develop/implement fault location techniques that can be easily applied and cost-effective to ease fault detection and location so that the burden on the operation and maintenance team can be reduced and fault restoration can be seamless during harsh climatic conditions.

1.3 Objective

As per the smart grid master plan [2] Bhutan aims to develop a fault location, isolation, and service restoration (FLISR) by 2030 thus, as a step towards the master plan the main objective of this research is as follows:

a. To develop effective fault detection, classification, and location algorithm using various supervised machine learning (SML) methods.
b. To compare different SML methods for fault detection, classification, and location using IEEE 14 bus System.
c. To carry out a parameter sensitivity study on the IEE14 bus system using SML techniques.

1.4 Scope of the research

The research will emphasize the areas that the company has planned as per the reference [3], thus with the idea of working on the distribution management master plan following is the scope of research.

a. The distribution system in Bhutan is spread across the extreme climatic zone and 90% of accidents occur during outage restoration the reliability of the service provided is affected, thus the algorithm is to be developed that can come to the rescue of the operation & maintenance team.
b. The proposed study to detect the occurrence of a fault, classify the fault types: single line to ground, double line to ground, three-phase faults, and estimate the fault location.
c. The effectiveness of various methods to be tested by modeling the real-time distribution system and IEEE 14 bus system and generating the database for study.
d. Carry out case studies with the generated database.

1.5 Anticipated contribution

a. The accidents that occur during right-of-way clearing, and line restoration during harsh climatic conditions can be greatly reduced since the operators will be aware of system health, fault type, and location beforehand.

b. With the history of event record the proposed method to effectively detect, classify and locate the fault.

c. Communication devices won't be required at both ends, the information data from the substation relay will serve the purpose thereby implication of additional cost.

d. The lineman and the operators can easily use the proposed technique without the need for expertise.

e. With the information on fault location and fault type the O&M team can be mobilized accordingly for the outage restoration.

f. The developed method will rescue the O&M team in a timely restoring the line, preventing power loss due to an outage and collective equipment life expectancy can be extended.

g. The method will be the door to enhance the reliability of distribution service and mitigation measures to reduce accidents.

Chapter 2 Theoretical background and method descriptions

To start the literature, it’s necessary to understand the system’s normal and faulted conditions. When a 50 Hz normal system functions smoothly and the system values remain within the constant limits the system is said to be normal. As the system deviates from the normal system parameters, it’s a sign of system abnormalities. When the system parameters reach the maximum permissible limit, the system is faulty, and the breaker trips to isolate the faulty section. Thus, system engineers are keenly interested in learning about the adverse effects of faults on the system and resolving the associated problems based on knowledge or research. Figure 2 shows the history of blackouts over the last 3 decades and the identified root causes are classified as natural, accidental, malicious, and cascading.

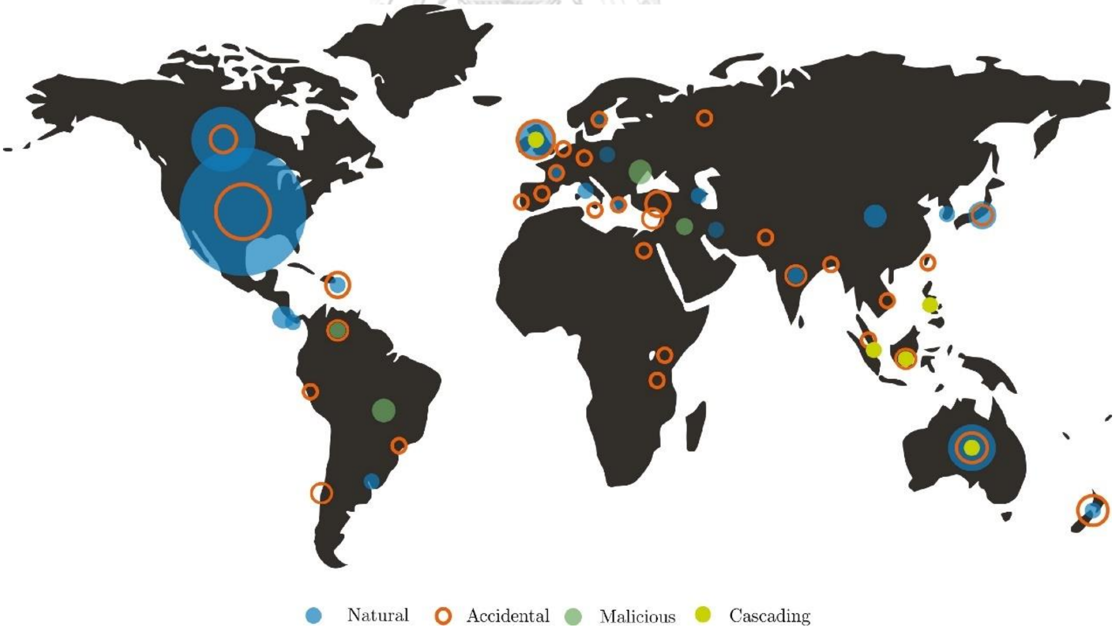


Figure 2 History of blackouts around the world with the root cause (1994-2019) [4]

2.1 Causes of faults

Frequently observed causes are conducting path failure and sudden changes in load parameters at the customer end which causes short circuits, under/over current, and under/over voltage which for instance causes fire hazards when not taken. System abnormality can be caused due to several reasons, which are listed below.

a. Insulation deterioration due to aging electrical components.
b. The swinging effect of conductors is caused by a strong wind.
c. Malfunction of joints of cables and overhead lines
d. Failure of one or more phases of a circuit breaker or conductor
e. Melting of the fuse caused by overcurrent.
f. Inadequate design, and installations.
g. Overloading and lightning surges cause insulation failure and mechanical failure.
h. Property damage by public intervention.

The figure below gives the pictorial overview of the various fault causes.

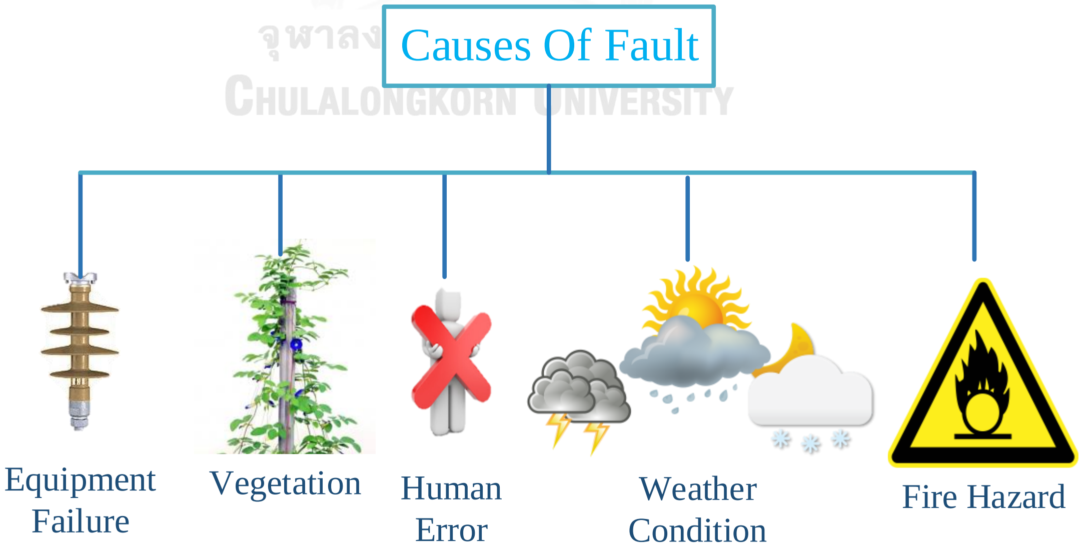


Figure 3 Causes of faults

2.2 Effects of faults

Power systems are the primary revenue generator and service provider around the globe thus the utility aims to minimize the effects of the faults and the following are frequently observed effects of the fault.

a. System reliability at stake: Loss of power in faulty are as well as interconnected areas, when not taken care of leads to a blackout.
b. Overcurrent due to fault damages costly electrical equipment.
c. Increased costs for repair and maintenance.
d. Risk to the safety of line operators: fatal and nonfatal.
e. Short-circuit ignites fires at utilities.

A descriptive illustration of the effects of fault is illustrated in given below.

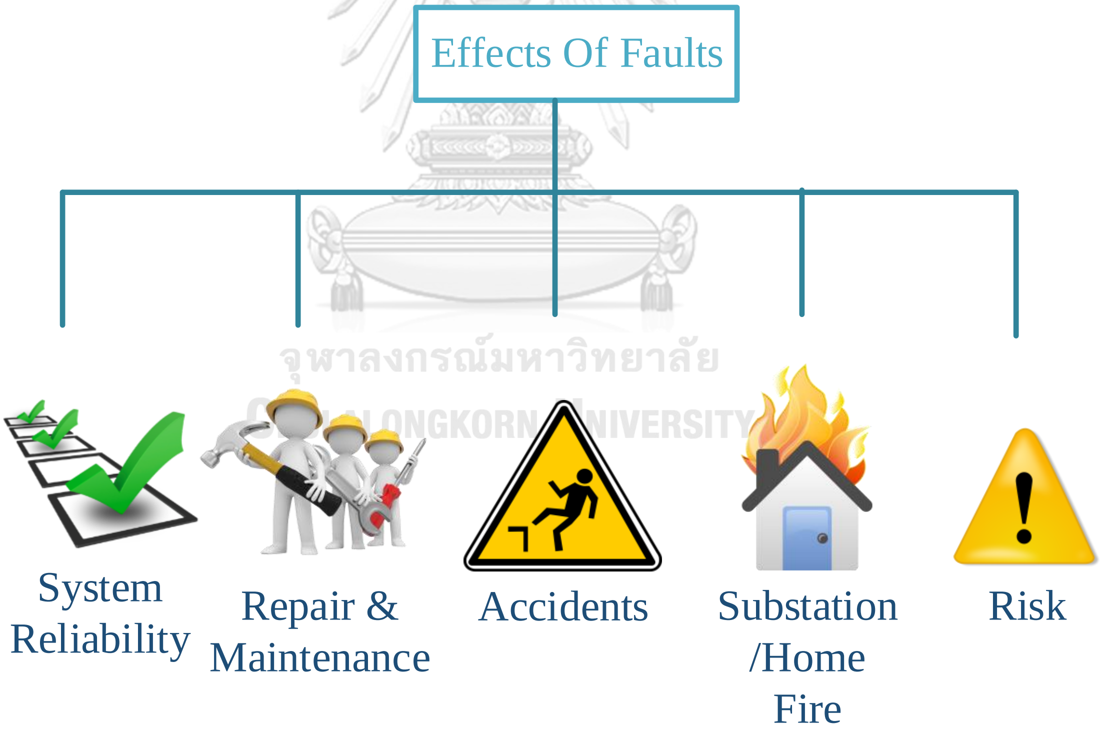


Figure 4 Effects of faults

2.3 Types of faults

As a three-phase power system behaves differently each time a fault occurs the faults can be categorized into open and short circuits which are further classified into symmetrical (balanced) and unsymmetrical(unbalanced)[5].To get a brief idea of how the system reacts and what its adverse effects let's study the fault types in detail. As per the reference [6] fault types are classified in Figure 5.

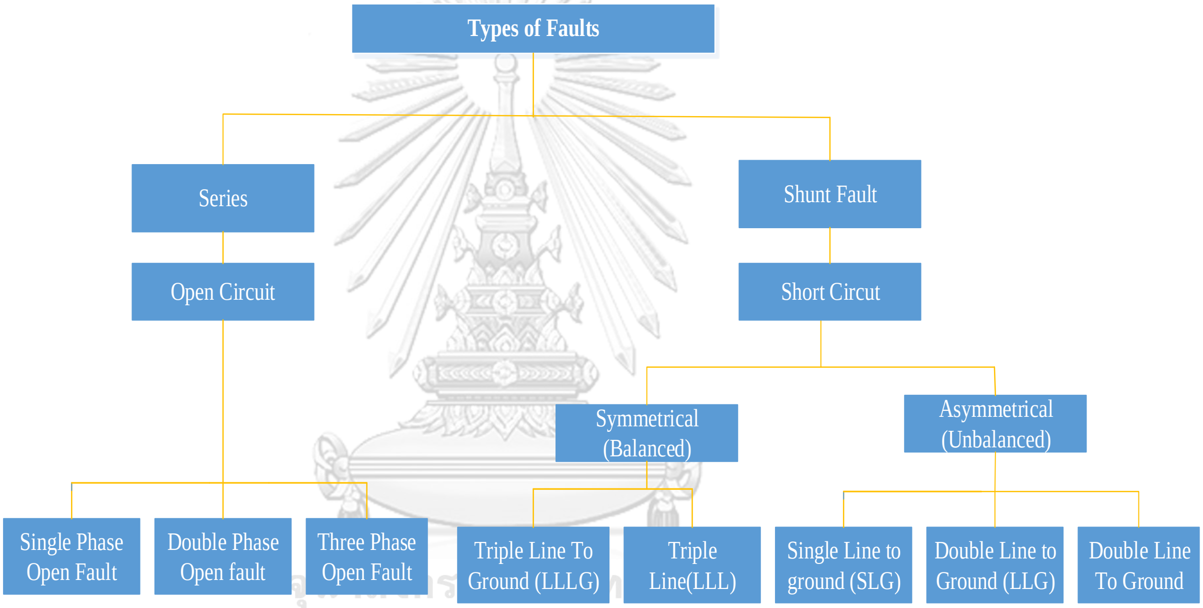


Figure 5 Fault classification

2.3.1 Series fault (open circuit fault)

Open Circuit faults, also known as series faults, occur when one more conductor fails in a three-phase system. Series fault can be single, double, and three-phases as illustrated in Figure 6, 7 & 8 where each phase of circuit is represented line with respective color code (RYB).

The faults are mainly caused by the malfunction of joints of cables and overhead lines followed by failure of one or more phases of a circuit breaker or conductor and melting of the fuse. This fault can be unsymmetrical or unbalanced except for three phases of open fault.

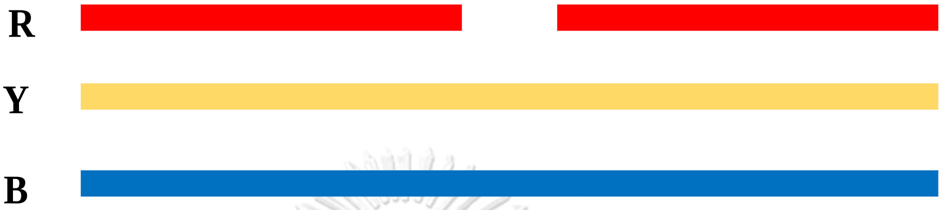


Figure 6 Single-phase open fault

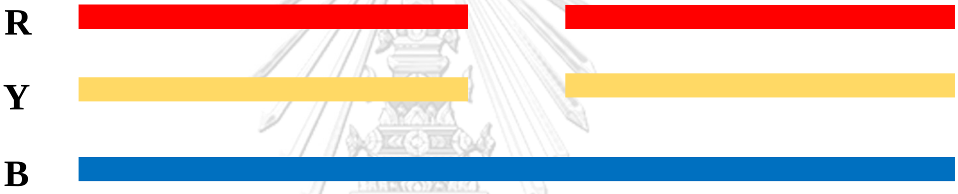


Figure 7 Double-phase open fault

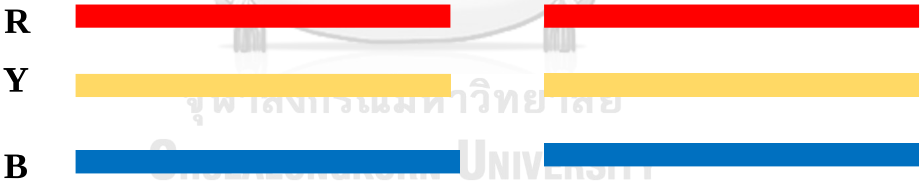


Figure 8 Three-phase open fault

The distribution network of Bhutan 66/33/11 kV carters the power from the substation to the transformer to the customer, and the load at the customer end can be balanced and unbalanced. If the transformer runs with the balanced load before the open fault, the transformer load increases and over-voltage at the transmission line is triggered to an extent that will cause a short circuit. Therefore, single and double-phase open circuit causes damage to the conductor and electrical components, system abnormalities, and insulation failure. The system can withhold the open circuit fault for a longer duration, as the open circuit fault does not generate a short circuit current, but it must be detected and rectified before it poses greater damage to the system.

2.3.2 Shunt fault

The shunt fault is further classified as balanced and unbalanced; details are given below.

2.3.2.1 Symmetrical circuit fault (balanced)

When the fault magnitudes of load currents are displaced by a 120° in phase during the fault condition such faults are known as symmetrical or balanced circuit faults which are characterized by the circuit's three-phase short-circuited. The probability of this fault rarely ranges from 2-5% of overall system faults. However, these faults hurt the power system even when the system remains in a balanced condition. The fault analysis of the system is executed using the bus impedance matrix or Thevenin's theorem by utilizing system data such as the breaking capacity of a circuit breaker, data from relays, and switch gear protective equipment.

a. Triple line fault (LLLF): When three-phase gets short-circuited.

b. Triple line to ground fault(LLLGF): When the three phases get short-circuited and encounter the ground.

2.3.2.2 Unsymmetrical circuit fault (unbalanced)

When the fault is characterized by unequal phase displacement with different fault magnitudes of load currents; then it's known as an unsymmetrical or unbalanced circuit fault. This fault is classified by both open circuit faults (single and two-phase open circuit faults) and short circuit faults excluding L-L-L and L-L-L-G faults. As per reference [5] SLG fault comprises 70% of overall system faults and the adverse effect on the system is significant.

The LLG fault usually occurs when the two conductors come in contact due to the swinging effect caused by a strong wind or any other external factors.
The double line to ground fault is severe as two lines encounter each other followed by contact with the ground. This fault accumulates to 10% of the overall system faults. An unsymmetrical circuit’s fault analysis is comparatively tedious compared to symmetrical fault analysis and methods of unsymmetrical components with system data like current and voltage magnitudes are utilized. The largest short circuit current occurs in L-G or L-L fault, and it is necessary to carry out fault analysis.

2.3.2.3 Short circuit fault

When the exceptionally low impedance of two different potential points gets connected accidentally or with the intent, it gives rise to system abnormalities known as short circuits or shunt faults. This fault is the most widely occurred fault that causes abnormally high inrush current to the electrical equipment and line igniting major damage, thus this fault needs to be rectified as early as possible. The main cause of the short circuit fault is insulation failure between the phases of the conductor, between the phase and earth conductor, or both.

The three-phase fault clear of earth and three-phase fault are also known as balanced or symmetrical, phase to phase, single line to earth, two-phase to earth and phase to phase, and single phase to earth are unsymmetrical faults. Short circuit faults are mainly caused by internal or external factors. The internal factors are a failure of electrical equipment, and lines, insulation deterioration due to aging electrical components, and inadequate design, and installations. External factors like overloading and lightning surges cause insulation failure, mechanical failure, and property damage by public intervention.

A short circuit fault is considered one of the most hazardous, as it often leads to arching and igniting fire causing an explosion of equipment like breakers and transformers. Further short circuits initiate abnormal currents in the system leading to overheating of equipment and lowering the life span of insulation. The short circuit fault also disturbs operating voltages causing the voltage to rise or drop from the permissible limits thereby affecting the quality of service provided to customers. If a Short circuit persists in the system and could not be located, it causes major power interruptions and equipment failure.

1. Single line to a ground fault (SLGF): When one phase of a conductor contacts the ground or neutral wire on a distribution line.

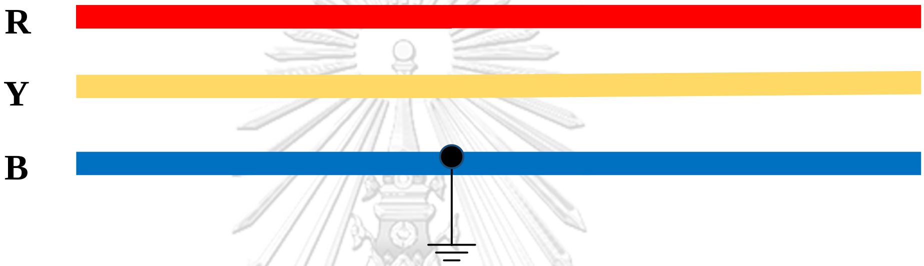


Figure 9 SLGF

2. Line-to-line fault (LLF): When strong wind causes a short circuit between two phases of the conductor.

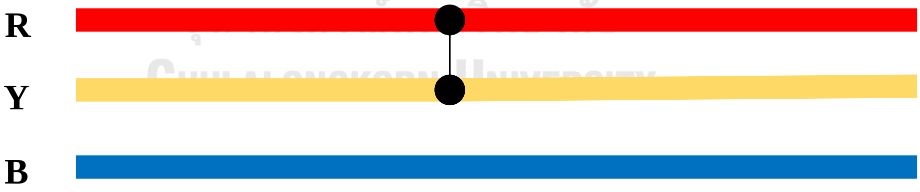


Figure 10 LLF

3. Double line to ground fault(LLGF): When the fault is associated with the falling tree which connects two-phase to the ground.

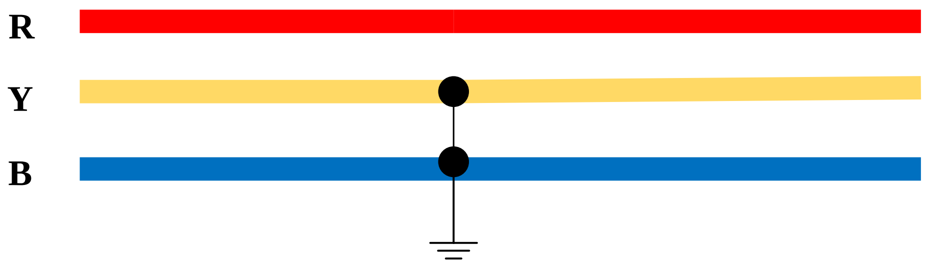


Figure 11 LLGF

4. Triple Line fault (LLLF): For example, the fault is associated with a falling tree connecting three phases of the conductor.

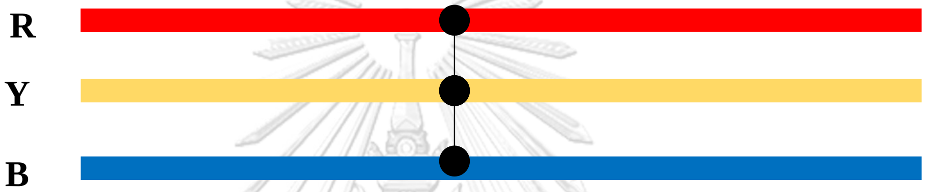


Figure 12 LLLF

5. Triple line to ground fault (LLLGF): For example, the fault is associated with a falling tree connecting three phases of the conductor and ground.

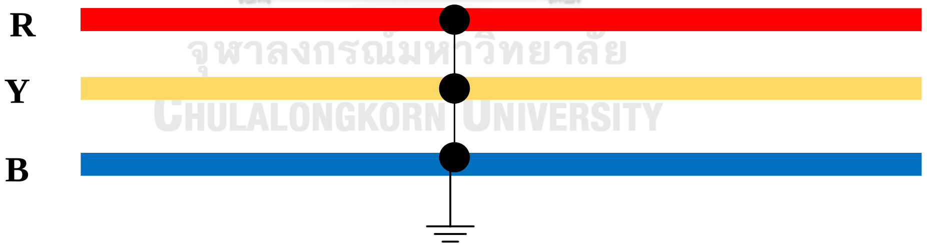


Figure 13 LLLGF

Usually, the transmission line system has relays to give system information and health, and data can be easily downloaded and analyzed in the form of waveforms or magnitudes. Whereas in the distribution system fuse and ARCBs are used and there is no coordination between the ARCB and fuse thus making the system analysis difficult and fault detection a challenge.

2.4 Measurement equipment in the substation

The power system faults are analyzed through the recorded power system parameters through various power system measurement equipment like current transformers (CT), potential/voltage (PT/VT) transformers, sequence analyzers, phasor measurement units(PMUs), and digital fault recorders. The measurement equipment is manufactured according to IEEE standards [7]. This equipment plays vital a role as the data obtained from the measurement equipment is used as a basis to evaluate the faults in the power system. The power system engineers thoroughly study the parameters and predict various stages of power system status for predictive and preventive maintenance. With the technology upgradation various power system measurements have been developed and are currently being used as per the power system requirement as discussed below.

2.4.1 Current transformer (CT)

CT is the type of transformer used to measure electrical current in a power system and it works on the principle of electromagnetic induction. It transforms high current to low current in the power system to low current that can be measured by the measuring device. To obtain reliable and accurate measurement is essential to select the correct CT with accuracy and class. The accuracy class of CT is expressed in the percentage of rated current and it’s defined as the maximum deviation of output current from actual input current under simplified conditions [7]. The classes are expressed as follows:

a. Class 0.1: This is the CT with the highest accuracy with a maximum permissible error of 0.1% of the rated current.
b. Class 0.2: This is the CT with accuracy with a maximum permissible error of 0.2% at the rated current.
c. Class 0.5: This is the CT with accuracy with a maximum permissible error of 0.5% of the rated current.

d. Class 1: This is the CT with accuracy with a maximum permissible error of 1% of rated current.
e. Class 3: This is the CT with accuracy with a maximum permissible error of 3% of rated current.

The CTs used in power systems are classified as protection, bus bar, feeder, and metering. The protection CT is used for the relaying and protection with a rated current of 1A or 0.5 with an accuracy class of 5P or 10P. Metering CT is used for revenue and billing and the rated current of CT ranges from 1A or 5A with accuracy and class of 0.2 or 0.5. Bus bar CT measures current flowing through the bus bar and the rated current CT ranges from 100A to 500 A with an accuracy of 0.5 or 1. Overall, the accuracy of CT and class are important for ensuring the reliability of the power system measurement obtained.

2.4.2 Potential/voltage transformer (PT/VT)

The potential or voltage transformer is used to measure the electrical voltage or monitor voltage fluctuations by transforming high voltage levels to the low & manageable voltage level. The primary winding is connected to the high-voltage side and the secondary winding is connected to the measuring instrument or relay. The ratio of primary voltage to secondary voltage is called the transformation ratio and it’s used to calculate the voltage of the secondary side. The PT/VT accuracy is defined as the maximum permissible between actual secondary voltage and rated secondary voltage at specified load conditions The classes are expressed as follows:

a. Class 0.1: This is the VT with the highest accuracy with a maximum permissible error of 0.1% of the rated secondary voltage.
b. Class 0.2: This is the VT with accuracy with a maximum permissible error of 0.2% of the rated secondary voltage.
c. Class 0.5: This is the VT with accuracy with a maximum permissible error of 0.5% of the rated secondary voltage.

d. Class 3: This is the VT with accuracy with a maximum permissible error of 3% of rated secondary voltage.

The VTs in the power system can be low and medium and are used for metering, protection, and monitoring. The VT is selected as per the requirement of the utility considering voltage & frequency rating, accuracy class, burden rating, and insulation level.

2.4.3 Sequence analyzer

A sequence analyzer is an instrument that analyzes the signal in a power system to determine the sequence components of the signal. The information about the sequence components can be used to detect, classify, and locate the fault.

2.4.4 Sequence components

In the study of faults, it’s necessary to know about the sequence components of the power system which often come in the form of positive negative, or zero sequences in a three-phase power system and it plays a major role in the process of understanding and creating fault detecting techniques. This sequence applies to current, impedance, and voltage; a balanced system gives rise to a positive sequence, unbalanced to negative, and grounded to zero sequence component.

In a three-phase balanced system, the current and voltage are mathematically equal, and the phasor is displaced by a 120° in clockwise rotation of ABC. This phenomenon is called a positive sequence. Positive Sequence components play an important role in power system protection as in most of the microprocessor-based relays in power systems, the positive component is utilized over-current protection which is a vital scheme of system protection.

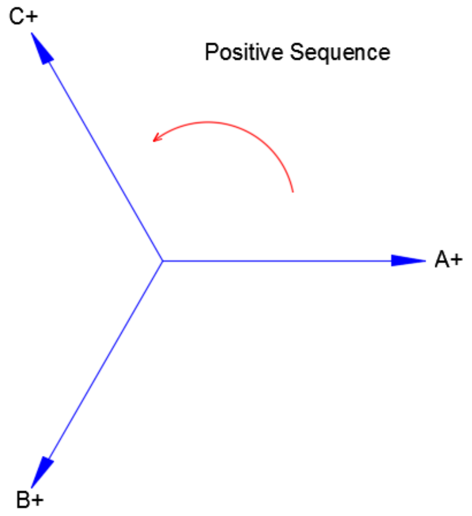


Figure 14 Positive Sequence Components

In a three-phase balanced system, the current and voltage are mathematically equal, and the phasor is displaced by a 120° in counterclockwise ABC (-) rotation direction. This phenomenon is called the negative sequence component. Negative sequence components are used by relays for directional and unbalanced protection, and it is kept as an option for the -current protection.

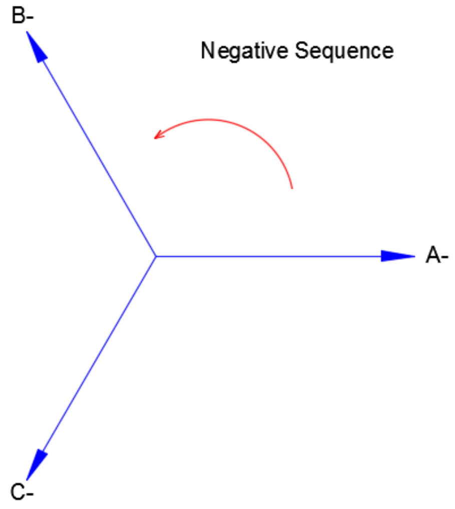


Figure 15 Negative Sequence Components

The phasor is equal in magnitude 0 degrees phase separation in the three-phase balanced system. This phenomenon is called zero Sequence.

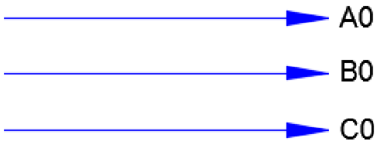


Figure 16 Zero sequence components

Zero Sequence is used for fault detection and fault calculation and as per the zero sequences formula; the neutral current (the sum of Ia, Ib, and Ic) is three times that of the zero-sequence current ($3I_0$). This concept is implemented for the identification of ground faults in grounded neutral systems by integrating three current transformers on three phase lines in parallel.

2.4.5 Phasor measurement unit (PMU)

The PMU is a device used in power systems for synchronized measurements of electrical phasors. Electrical phasors are mathematical representations of sinusoidal voltage and current used to describe AC power systems' behavior. The PMU measures phasor quantities of current and voltage given location, mostly for high-voltage transmission lines. These measurements are time-stamped and sent to the central monitoring system, where real-time monitoring takes place. The data obtained from PMUs can also be utilized in machine learning process in the electrical fields. With technology advancement data from PMU has become an important aspect for the power engineers to study the behavior of the power system and lines. The PMUs are classified according to their functions.

a . **Stand-alone PMU**: The PMU is installed independently in the power system, measures voltage and current at a single location, and sends the data to the central monitoring system for analysis.

b. **Integrated PMU**: The PMU is integrated into another device, such as a protective relay or digital fault recorder. The device is used to measure phasors and power system parameters that are used for monitoring and protection purposes.

c. **Synchronized Phasor Gateway**: This is the device that is used to monitor and distribute data collected via multiple PMUs in a system. The device helps the operators improve the efficiency of the control system.

d. **Substation PMU:** This type of PMU is installed in the substation to measure the voltage and current phasors located at the junction or interconnection point of transmission and distribution systems. This device is essential for controlling and monitoring transmission and distribution lines.

2.4.6 Digital fault recorder (DFR)

The digital fault recorder (DFR) is used to capture and retain data about faults, disturbances, and other occurrences that occur in the electricity system. DFRs are used in electric power systems to give engineers and managers useful information that will aid in locating and diagnosing power system issues. Data from voltage and current waveforms, fault position, system frequency, and other significant factors can all be recorded by DFRs. To identify the cause of the fault or disruption, assess the impact on the electricity system, and develop strategies to prevent similar issues in the future, the recorded data can be reviewed. DFRs are usually placed in the power system at key locations, such as high-voltage transmission lines or transmission substations. For research and archiving, the captured data is frequently sent to a hub. Modern power systems now place a greater emphasis on DFRs because of their potential to increase system reliability, decrease downtime, and guard against machine harm. The data obtained from the DFR has become increasingly important to understanding the power system and thus these data/signatures are used widely in machine learning models.

2.5 Fault detection, classification, and location techniques

The main concern of the generation and distribution utilities is the power system's reliability. With the rapid increase in the demand for power in all techno-driven industries and smart homes, research on improving power system reliability is on the rise. The ongoing research is described in the sections below.

2.5.1 Fault Location

In Bhutan, the distribution lines are spread across the regions with low consumption patterns and almost no proper communication. The operators rely on the information displayed in the relay during faults. In most instances, the prediction doesn't work, and its restoration work becomes tiresome. Though there are lots of the latest developments in the market, due to cost and communication constraints it's difficult to implement them. The ongoing research can be broadly explained as per [8] along with the advantages and disadvantages. The overview of the method is illustrated in Figure 17.

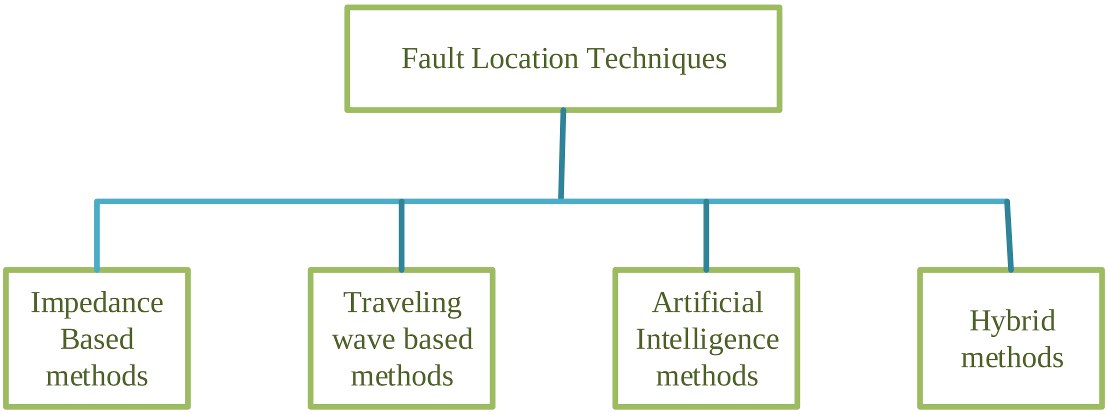

Fault Location Techniques
Impedance Based methods
Traveling wave based methods
Artificial Intelligence methods
Hybrid methods

Figure 17 Fault location techniques

2.5.2 Impedance-based method

The basic principle is illustrated in figure 4 where V_f & I_f corresponds to fault voltage and fault current. Where Vs is sending end voltage, Zs is the total line impedance, and m is the distance to the fault.

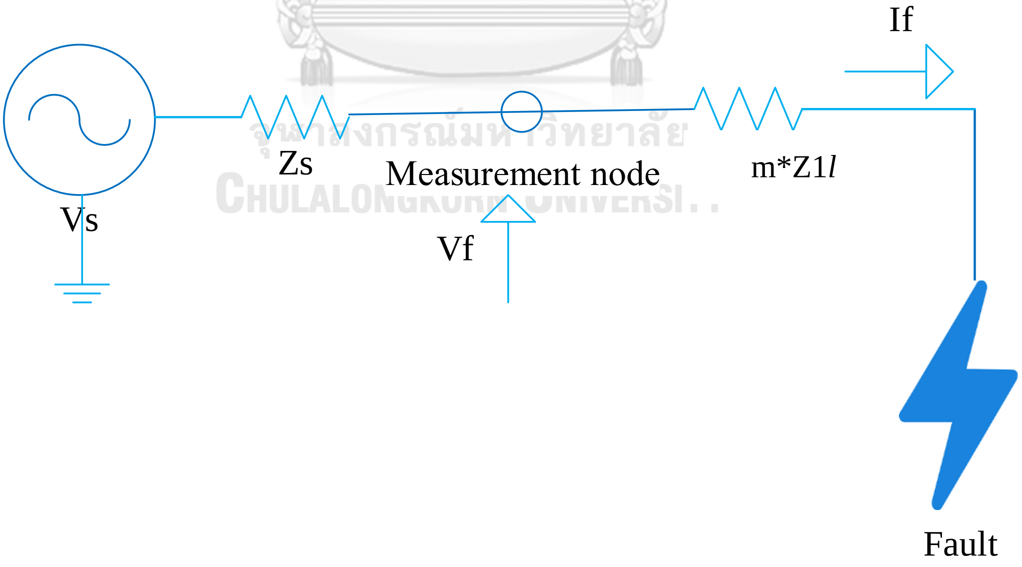


Figure 18 Impedance-based method basic diagram.

Using the ohms law, the following equation can be derived.

$$m = \frac{V_f}{I_f * Z_1 l} \qquad (1)$$

The impedance-based methods are classified as single-ended (uses measurement from one end) two ended methods (uses measurement from both ends). To meet the current requirement of the utility an algorithm using the single-end impedance method has been developed as indicated in [9] which uses three-phase current and voltage values when zero sequence impedance and positive sequence impedance are known.

The Method proposed as in reference [10] utilizes the frequency, current, and voltage recorded before and during a fault (single-phase-to-ground fault) on a radial system a system and fault location technique involves six steps.

a. Apparent Faulted Section: is implemented using fault type, current, and phasor sequence parameters.
b. Equivalent Radial System: here the apparent fault location is disregarded and loads of adjacent node is considered.
c. Load Modeling: here properties of the load are reflected by current compensation and constants and voltage are used to calculate load admittance and sequence current.
d. Voltages and Currents at the Fault and Remote end are utilized. IV. Estimating of fault location: via resistive nature of fault impedance and voltage-current properties at the fault.
e. Converting Multiple Estimates to Single Estimate: the estimates from the fault locator arrive at a single point using the software.

The setback of the method was the size of the fault locator and the software interface. The method as per reference [11] uses the current and voltage data from the fault locator and uses a separate algorithm for single phase and 3 phase computation using the impedance-based method to compute the fault location.

However, the accuracy of the algorithm depends on the accuracy of pre-fault condition determination from the substation.

The reference [12] ventures into the concept of fault location in short transmission line loads including the tapped lines indicated by lumped parameters impedance and positioning them after the fault. This method of compensating the tapped load is accurate as tapped load impedance is larger than the feeder impedance based on pre-fault and fault voltage measured from the substation. The negative sequence component is used for unbalanced faults and to reduce the source impedance error between pre-fault and fault conditions. This method was found accurate when the tapped loaded was minimum, however, as the tapped load increased the accuracy was reduced. However, impedance-based methods are widely used because of their simplicity and this method can be used with information available from a single end.

To access the feasibility of the impedance-based method as per the reference [9] meet an algorithm using the single-end method has been developed as indicated which uses three-phase current and voltage values when positive sequence impedance is known.

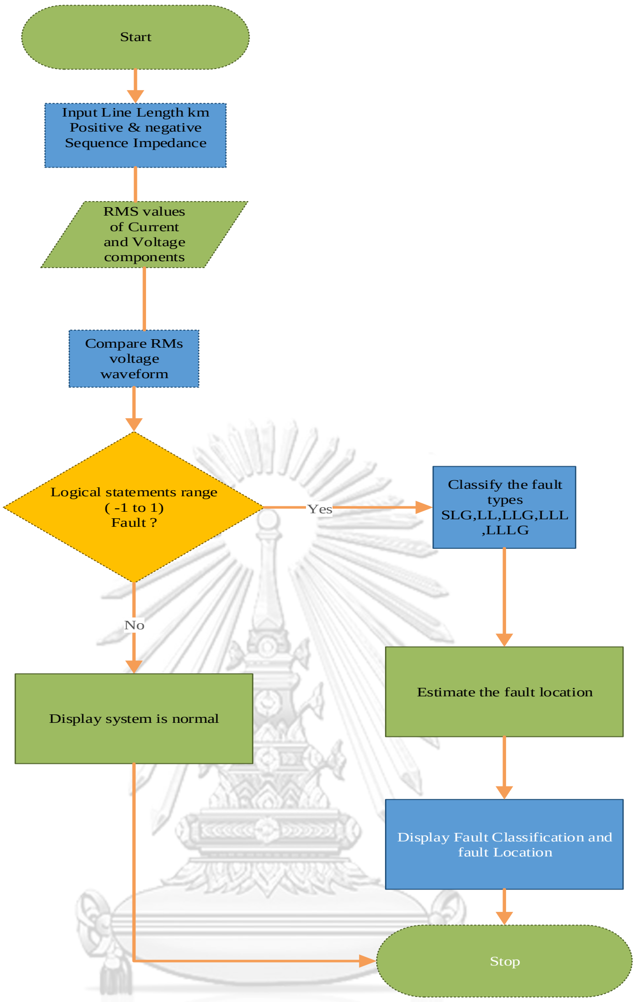


Figure 19 Logical & impedance-based flowchart.

The logical expression is formulated for the detection and classification using the fault voltage and location using the impedance-based method as indicated in [9] as equations are given in Table 1 Where Va, Vb, and Vc are three-phase voltage Ia, Ib, Ic is a three-phase current.

k is $$\frac{(Z_0\,L{-}Z_1\,L)}{3Z_1} \qquad (9)$$

ZIL positive sequence line impedance

Z0L is zero sequence impedance

m is faut location in km

Table 1 Fault location equations as per [9]

Fault type	Positive sequence impedance (mZ1L =)
AG	Va / (Ia + kIR)
BG	Vb / (Ib + kIR)
CG	Vc / (Ic + kIR)
AB or ABG	Vab / Iab
BC or BCG	Vbc / Ibc
CA or CAG	Vca / Ica
ABC	any of the aforementioned: Vab / Iab, Vbc / Ibc, Vca / Ica

As per the table the system Va, Vb, Vc , Ia , Ib & Ic can be measured at substation via the measurement equipment like current and voltage relay and the algorithm for this method uses the formulas as given in Table 1.

2.5.3 Traveling wave-based method

Traveling based method uses high-frequency waves that get initiated during the fault condition. Usually, current reflections' arrival time is used to calculate the fault location as the polarity of the waves doesn't change as it reflects in the same direction. By measuring one end equation 2 is used to calculate the fault location.

$$m = \frac{C\% * c * t}{2} \qquad (2)$$

Where C% is the relative speed of light factor.

c is the speed of light

t is the time between reflected waves

m is % distance to fault in km

Further, Figure 20 gives a simple illustration of the traveling wave which consist of 2 node A & B as measuring point. L is the total line length & f_d is the line length from one end. The t_2 & t_1 are the time taken for the waves to travel to faulted points and return.

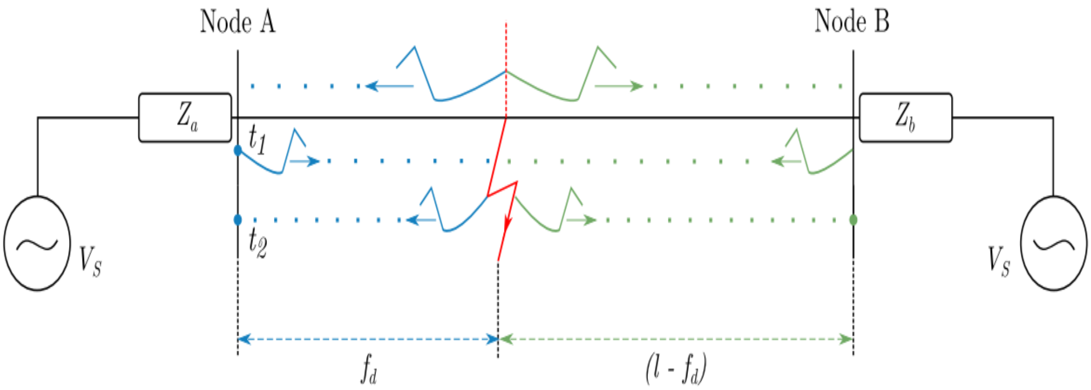


Figure 20 Traveling wave-based method basic diagram [8]

Thus, equation 2 can be further written as per [8]

$$f_d = \frac{v(t_2 - t_1)}{2} \qquad (3)$$

Where v is the velocity of propagation and the time taken to travel and return is known as the inception angle. The reference [8] explains different traveling wave methods implemented over the years. This method needs high sampling devices and thus there is a cost constrain implementation of this method in a complex system the error increases due to the requirement of synchronized observations devices like GPS (Global positioning system) and PMU(Phasor measurement unit)

2.5.4 Artificial intelligence-based method

Artificial intelligence refers to a computer system's capacity to replicate human cognitive abilities like problem-solving and learning. The AI computer system may construct reasoning that can be used to learn from the received information and make conclusions using logic, mathematics, and pattern recognition.

2.5.4.1 Machine learning

Machine learning is an application of AI. Machine learning is the practice of applying data-driven mathematical models to assist a computer in learning without being explicitly instructed. As a result, a computer system may keep picking up new skills and become better on its own as per [13]. Figure 21 shows an overview of machine learning techniques.

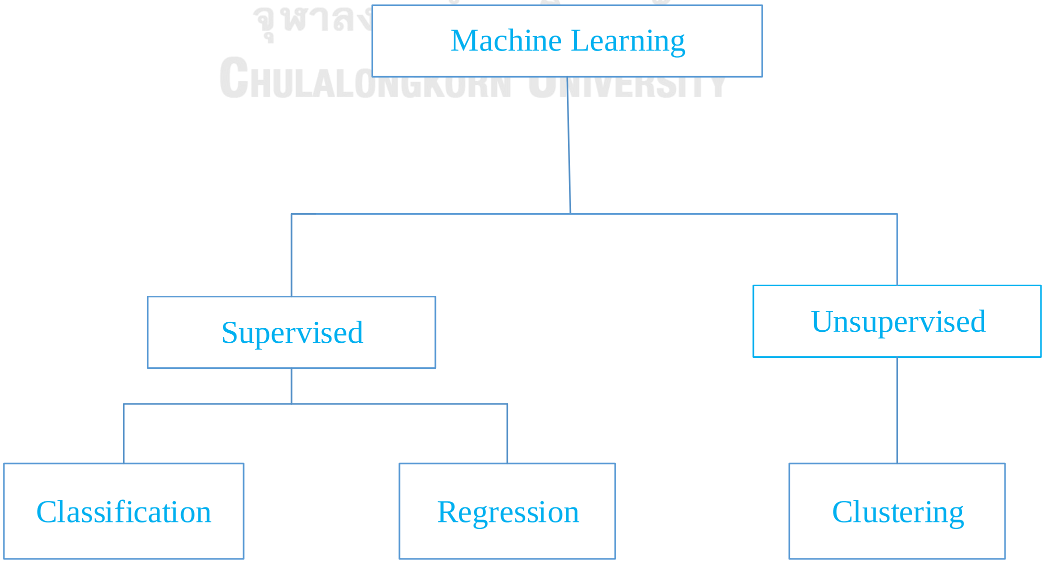


Figure 21 Classification of machine learning

2.5.4.2 Supervised machine learning (SML)

Supervised learning requires correctly labeled input and output sample data as an example to train a network or model. Supervised learning is of two types: classification & regression as in Figure 21. The classification classifies the input into a predetermined output, such as genuine or spam mail. The regression method is the most common machine language used across various fields and it predicts continuous responses, for example, the relationship between the effect of sales after advertisement or reckless driving and road accidents. The reference [14] uses supervised learning for the classification of various fault causes like faults caused by birds, and animals using the disturbance recorder files which are in the form of waveforms of various fault cases. The method was found effective, and classifications were accurate, and Table 2 shows the overview of the ML techniques.

Table 2 Machine learning techniques supervised and unsupervised.

Supervised		Unsupervised
Classification	**Regression**	
Support Vector Machine (SVM)	Linear regression (LR)	K-means, K-median
Linear discriminant analysis (LDA)	Ensemble methods (EM)	Fuzzy, C-means
Naïve Bayes (NB)	Decision trees (DT)	Hierarchical
Nearest neighbor (KNN)	Least squares (LS)	Gaussian mixture
Neural Networks (NN)	Neural networks (NN)	Neural networks

2.5.4.3 Linear discriminant analysis (LDA)

A statistical technique called discriminant analysis is used to categorize items into predetermined categories (classes) based on several predictor factors. It is applied to categorize fresh observations based on their combination of predictor values and to identify the set of variables that best distinguishes across classes. Building a discriminant function that properly distinguishes the various classes and can be used to foretell the class membership of fresh data is the aim of discriminant analysis [13].

$$D(X) = WX + W^0 \qquad (4)$$

Where "W" denotes the vector of coefficients,

"X" is the vector of predictor values for a particular observation, and

"W^0" denotes the intercept term.

By calculating the value of the function for each observation and allocating it to the class with the biggest value, the discriminant function distinguishes the classes. Using techniques like maximum likelihood estimation, the coefficients and the intercept term are inferred from the training data. Depending on the type of discriminant analysis performed, the specific shape of the discriminant function may change (e.g., linear, quadratic, Mahala Nobis).

2.5.4.4 Neural network (NN)

Inspired by the working of the human brain, a neural network consists of several nodes known as artificial neurons organized in layers and the neural network can be represented as [13]

$$y = f(z) \qquad (5)$$

where "y" represents the neuron's output.

"z" represents the inputs' weighted sum.

"f" represents the activation function.

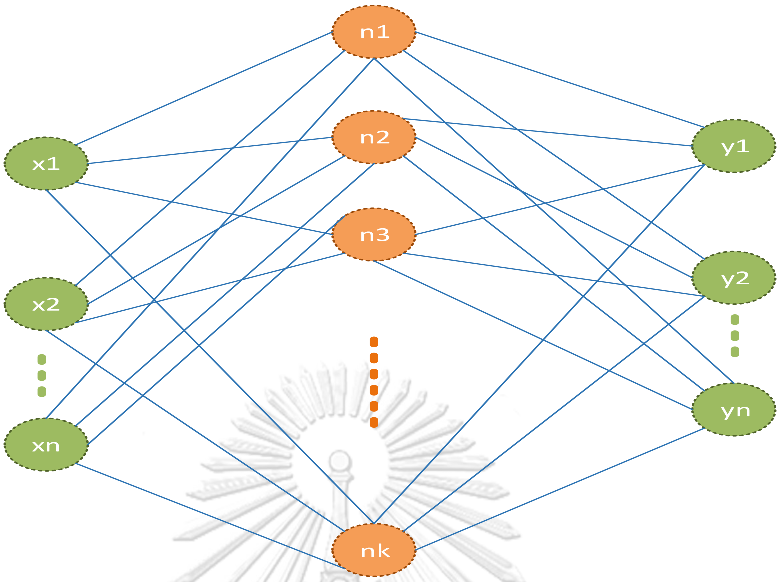


Figure 22 Neural network

With time there is a gradual shift of focus to AI-based techniques with an improved platform to perform data analysis research. AI-based methods are on the rise and all the above methods and techniques are used for data analysis across various fields. The research will focus on supervised learning and will widely explore classification and regression methods. In fault analysis, artificial neural networks are in use owing to their accuracy and their ability to understand the system behavior through existing data. It analyses the inputs and assigns them to predetermined outputs as indicated [4]. A neural network consists of an input layer, a hidden layer, and an output layer. It analyses the inputs and assigns them to predetermined outputs per reference [8].

2.5.4.5 Decision tree (DT)

A decision tree is a model that takes the form of a tree, with internal nodes representing decisions based on specific attributes, branches representing the resulting outcomes of these decisions, and leaf nodes representing either class labels or numerical values used for prediction. The decision tree algorithm recursively partitions data into subsets based on input features, selecting the feature that provides the most information gain at each internal node. Decision trees are popular and interpretable due to their ease of understanding and ability to handle both categorical and numerical data. However, they can suffer from overfitting and may not perform well with interactions between features, which can be mitigated by pruning, ensemble methods, or boosting algorithms.

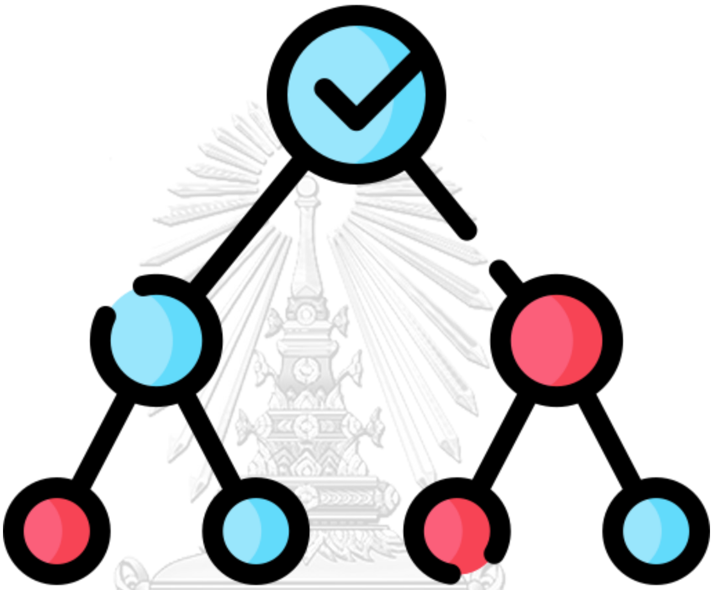


Figure 23 Decision Tree

2.5.4.6 K-nearest neighbors (KNN)

K-Nearest Neighbors (KNN) is a simple and intuitive algorithm used in supervised machine learning for classification and regression tasks. It finds the k-closest training samples (neighbors) in feature space to a new input data point and predicts the class label or value based on the majority vote or average of the k-nearest neighbors. Distance between data points is typically calculated using the Euclidean distance or other distance metrics. KNN is versatile and easy to implement but may not perform well on high-dimensional or sparse data and requires a careful selection of hyperparameters to avoid overfitting or underfitting.

2.5.4.7 Support vector machines (SVM)

Support Vector Machines (SVM) is a powerful algorithm used in supervised machine learning for classification and regression tasks. It aims to find the hyperplane that best separates data into different classes. SVM can handle both binary and multi-class classification problems and regression tasks by finding the hyperplane that best fits the data while minimizing error. SVM is advantageous due to its ability to handle high-dimensional data, flexibility in handling both linear and non-linear decision boundaries, and robustness to outliers. However, SVM can be sensitive to the choice of kernel function and hyperparameters and may be computationally expensive for large datasets. Optimal performance can be achieved through careful selection of hyperparameters and kernel functions.

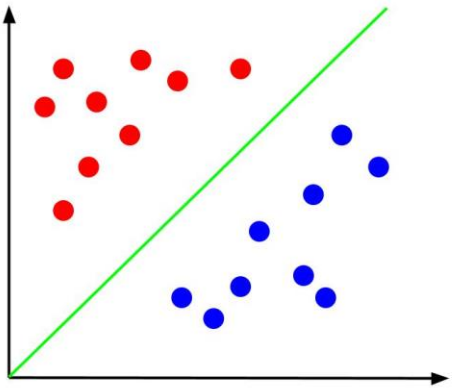


Figure 24 Support vector machine

2.5.4.8 Least square method (LSM)

The LS is a mathematical approach used to fit the line/curve to set data points. It works by minimizing the sum of the squared difference between the observed data points and the corresponding predicted values generated by the model. The equation for the LS can be expressed as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots .. + \beta_p x_p + \varepsilon \quad (6)$$

where:

y is response variable meaning the variable that will be predicted

$x_1, x_2 \ \dots x_{p,}$is the predictor variables i.e. variables used to predict the response.

β0, β1, β2, ..., βp are parameters that need to be estimated/predicted

ε is the random error term i.e difference between the predicted and observed values.

The goal of the model is to find β0, β1, β2, ..., βp that minimize the sum of of squared errors (SSE) between predicted and observed values.

$$SSE = \Sigma(yi - \hat{y}i)^2 \qquad (7)$$

Were,

yi is the observed value of the response variable for the ith data point.

$\hat{y}$i is the predicted value of the response variable for the ith data point.

To find values of β0, β1, β2, ..., βp that minimize the SSE, least squared method is used, and it takes the partial derivatives of SSE concerning each coefficient, setting them equal to zero and resulting values of β0, β1, β2, ..., βp provide best- fit for the line and curve for the given data set.

2.5.4.9 Linear regression method (LRM)

Linear regression is a statistical technique that is utilized to investigate the connection between a dependent variable, also known as an outcome or response variable, and one or more independent variables, also known as predictor or explanatory variables. The main objective of linear regression is to determine the best-fitting straight line, or hyperplane in the case of multiple independent variables, that describes the relationship between the variables. In a simple linear regression, there is only one dependent variable and one independent variable, and their relationship is illustrated by the equation of a straight line given as:

$$Y = b_0 + b_1 * X \qquad (8)$$

Here, Y represents the dependent variable, X represents the independent variable, b_0 denotes the intercept, and subscript represents the slope of the line. The goal of linear regression is to calculate the values of b0 and b1 that minimize the difference between the observed Y values and the predicted Y values based on the equation mentioned above.

2.5.5.0 Unsupervised machine learning (UML)

Whereas unsupervised learning doesn’t need the example. The system uses the data in the cluster and groups them with shared characteristics known as clustering and it’s widely used in gene sequencing, market research, and object orientation recognition like digital image processing. The machine learning techniques discussed here can be further divided into various categories and used in various research fields.

2.5.5.1 Related work

Some of the existing methods in power system are discussed here: The fault classification Distribution Management System-based fault location as per [15] The reference demonstrates how the DMS and network information can be integrated to form a distribution automation system using the existing microprocessor-based relay without much cost implication. The fault detectors are vital for the process, and they need to be correctly placed so that correct information can be retrieved despite adverse weather conditions or geographical implications. The fault location principle uses the difference between calculated current data and the one obtained from relays and fault detectors and analyzes the data via algorithm and processes through fuzzy logic to find the faults of the section with an approximate distance. The DMS user interface provides incorporation of GPS coordinates so that the real-time overview of the lines can be monitored along with a geographic view of the network and fault data and restoration options. The DMS is commercialized and is widely used around the world. However, the fault detection technique is limited to only short circuits.

The reference [7] uses an improved cuckoo search algorithm to find the fault, use current data from the field terminal unit(FTU), and perform a generic switching function. In this case, the algorithm’s accuracy depends on the accuracy of data obtained from the FTU. The methods described have their setbacks and fault locators and fault passage indicators have cost and communications constraints, followed by separate algorithms for detection, classification, and location so this paper aims to develop a user-friendly single algorithm that detects, classifies, and locates the fault.

The fault classification Distribution Management System-based fault location as per [15] reference demonstrates how the DMS and network information can be integrated to form a distribution automation system using the existing microprocessor-based relay without much cost implication. The fault detectors are vital for the process, and they need to be correctly placed so that correct information can be retrieved despite adverse weather conditions or geographical implications. The fault location principle uses the difference between calculated current data and the one obtained from relays and fault detectors and analyzes the data via algorithm and processes through fuzzy logic to find the faults of the section with an approximate distance. The DMS user interface provides incorporation of GPS coordinates so that the real-time overview of the lines can be monitored along with a geographic view of the network and fault data and restoration options. The DMS is commercialized and is widely used around the world. However, the fault detection technique is limited to only short circuits. The research will focus on supervised learning and will widely explore classification and regression methods. In fault analysis, artificial neural networks are used due to their accuracy and their ability to understand the system behavior through existing data.

2.5.5.2 Hybrid method

The combination of AI methods with the conventional method is known as the hybrid method. The main aim of this method is to combine the advantages of AI-based techniques and conventional techniques and get better accuracy in the algorithm. The hybrid method classification is illustrated in Figure 24.

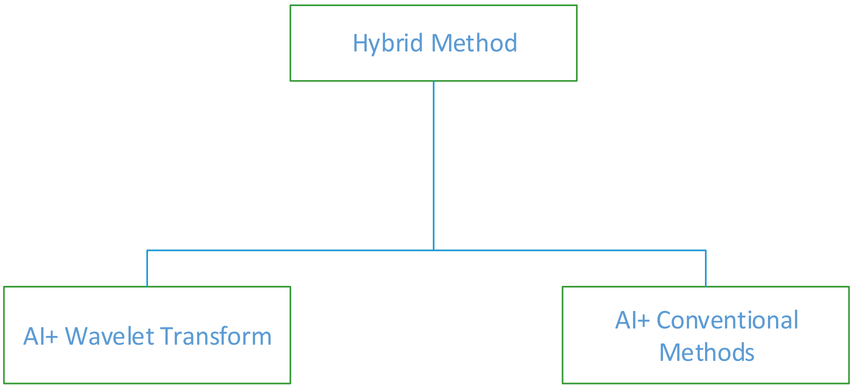


Figure 25 Hybrid method

The reference [16] uses multilevel wavelet transform, principal component analysis, support vector machines, and adaptive structural neural networks to simultaneously determine fault kind and location. In addition to introducing the methodology of the analytical approaches, a pattern-recognition approach using neural networks, and a collaborative decision-making mechanism, this study lays forth the core idea of the proposed framework. The tasks of problem detection, classification, and localization are completed in 1.28 cycles using a well-trained framework, which is far quicker than the necessary fault clearance time. As indicated in reference [8] though the algorithm's accuracy is highly anticipated, the requirement of filters, high sampling devices, complex nature, and cost constraints make it difficult to implement in the field.

2.5.5.3 Decimated wavelet decomposition (DWD)

In reference [17] wavelet theory and its application are discussed, and it is primarily used in AI and hybrid methods. The study's focus is on DWD, which involves a maximum of log2 N steps for a signal s with a duration of N. The first step involves generating two sets of coefficients, namely, approximation coefficients cA1 and detail coefficients cD1, from the signal s. This is achieved by combining s with the low-pass filter LoD and the high-pass filter HiD and then conducting dyadic decimation to obtain the approximation and detail coefficients (downsampling).

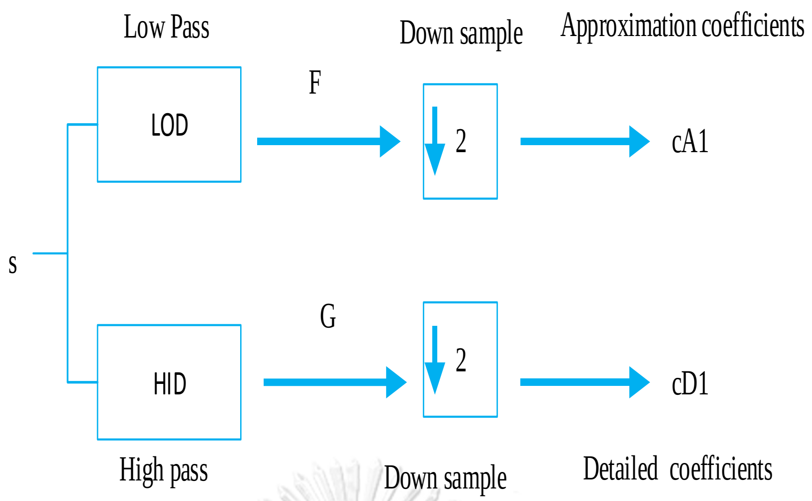


Figure 26 Wavelet decomposition

The length of each filter is 2n, where n is a positive integer. If the length of the signal s is N, then the lengths of coefficients cA1 and cD1 are Floor(N-1)/2)+n, and the lengths of signals F and G are N+2n-1. In the next step, the same approach is used to divide the approximation coefficients cA1 into two halves, generating cA2 and cD2 after replacing s with cA1.

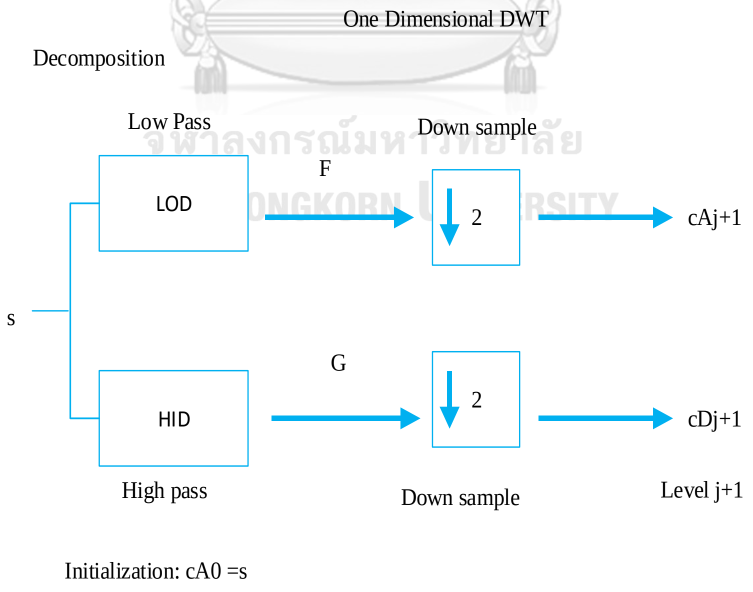


Figure 27 One-Dimensional wavelet decomposition

The signal wavelet decomposition, as determined by the level j analysis, is composed of the following elements: [cAj, cDj..., cD1]. For j = 3, this structure holds the terminal nodes of the following tree:

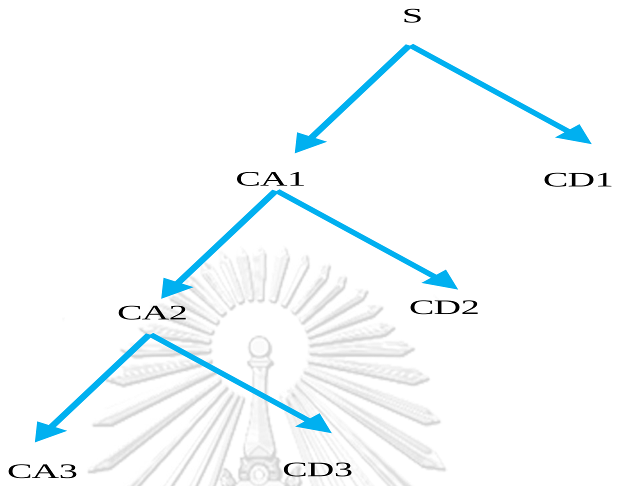


Figure 28 Structure of Decomposition tree

Wavelets are widely used to understand waveforms through the decomposition pross using the high pass and low pass filters and analyzing the small basic details of the features obtained and these features are then used to train the network or the model.

2.6 Research gap/problem statement

After detailed research, the following problem statement was identified.

a. Need for research on advanced fault localization techniques that leverage supervised machine learning approaches to improve the accuracy and efficiency of fault location estimation in distribution lines.

b. Research is needed to assess the trade-offs between accuracy, computational requirements, and implementation costs for each approach.

c. There is limited research that directly compares two or more SML approaches.

2.7 Anticipated contribution

a. The accidents that occur during right-of-way clearing, and line restoration during harsh climatic conditions can be greatly reduced since the operators will be aware of system health, fault type, and location beforehand.

b. With the use of existing devices and data the proposed method to effectively detect, classify and locate the fault.

c. Communication devices won't be required at both ends, the information data from the substation relay will serve the purpose thereby implication of additional cost.

d. The lineman and the operators can easily use the proposed technique without the need for expertise.

e. With the information on fault location and fault type the O&M team can be mobilized accordingly for the outage restoration.

f. The developed method will rescue the O&M team in a timely restoring the line, preventing power loss due to an outage and collective equipment life expectancy can be extended.

g. The method will be the door to enhance the reliability of distribution service and mitigation measures to reduce accidents.

Chapter 3 The overview of the framework

3.1 The proposed technique

AI techniques are broadly researched due to their ability to understand the system's behavior through the data sets and research exploring conventional & AI methods. The conventional method becomes outdated as technology advances and new technology is required to meet the need of communication technology. As our country has visioned in the reference [2, 3], this research work explored the possibilities to use SML mitigate current condition of the distribution line in the country using the proposed technique SML based fault detection, classification & location. With improved communication technologies followed by readily available sensors the AI based technology are growing in all the fields, and these methods have proven its ability in distribution fields as per reference [4]. Thus, this study utilizes various SML technologies to carry out the comparative study.

3.2 The research methodology

The research methodology used in this study is of utmost importance as it provides the framework for planning, executing, and analyzing the research process. It encompasses various components such as the test system, database, algorithm or trained network, and the obtained results. These components work together to ensure a systematic and reliable approach to the research.

The test system is a crucial element of the methodology, as it defines the experimental setup or environment in which the research is conducted. It includes the necessary tools & instruments that are used to collect data or perform experiments.

The database forms an integral part of the research methodology as it provides the primary source of data for analysis and evaluation. It contains the relevant information and records that are used to train the algorithm or network and to validate the obtained results. The database may consist of structured data, unstructured data, or a combination of both, depending on the nature of the research.

The algorithm or trained network represents the core component of the research methodology. It encompasses the mathematical models, statistical techniques, or machine learning approaches that are employed to analyze the data and make predictions or classifications. The algorithm or trained network is designed and optimized based on the research objectives and the characteristics of the dataset.

Finally, the results obtained through the research methodology provide valuable insights, conclusions, or predictions related to the research problem or question. These results are derived by applying the algorithm or trained network to the available data and analyzing the output. The results are typically evaluated and interpreted to draw meaningful conclusions and make informed decisions.

Figure 29 provides an overview of the proposed system, illustrating how the different components of the research methodology interact and contribute to the overall research process. It serves as a visual representation of the methodology, highlighting the flow and connection between the various stages involved in conducting the research.

.

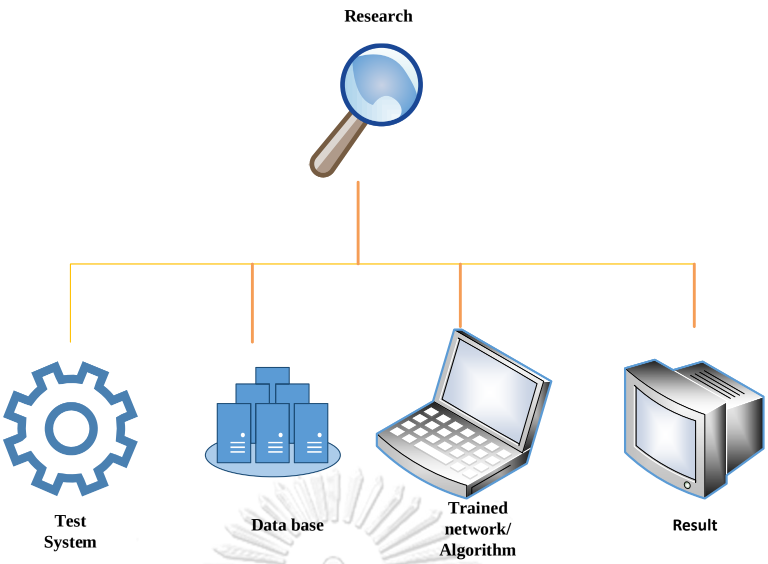


Figure 29 The methodology

3.3 The test system

IEEE 14 bus system is widely used as the benchmark to evaluate the performance of different power system analysis techniques, algorithms, and optimization methods. Researchers and engineers often use it as a test system to propose new methodologies for power system analysis. Thus this study also utilizes 14 bus system as per reference [19] is to generate the database for checking the efficiency of algorithm.

The IEEE 14 bus system consists of 14 buses with bus configuration and slack bus (reference) , PV and PQ bus. It is connected to 5 generators located at bus 2,3,5,6 & 8. and 11 loads. The system has 20 transmission lines that connect the buses. The availability of the data and its simplicity makes IEEE 14 bus a good choice for the proposed study.

The 14-bus system simulated in MATLAB to create the database. The standard IEEE 14 bus system is simulated in MATLAB/Simulink, with the transmission line parameters converted from per unit to actual values. However, the data sheet assumes zero half charging susceptance, resulting in unrealistic line lengths and capacitance. To address this, a small factor of (0.00005pu) is introduced as line charging susceptance between line 8 and line 20 to reflect real-world power system networks. This allows for more accurate representation of line length and capacitance. The modeled system uses the system data as per [20] , which further utilizes the parameters as per the reference [19]. The snapshot of the modeled system is given in Figure 31.

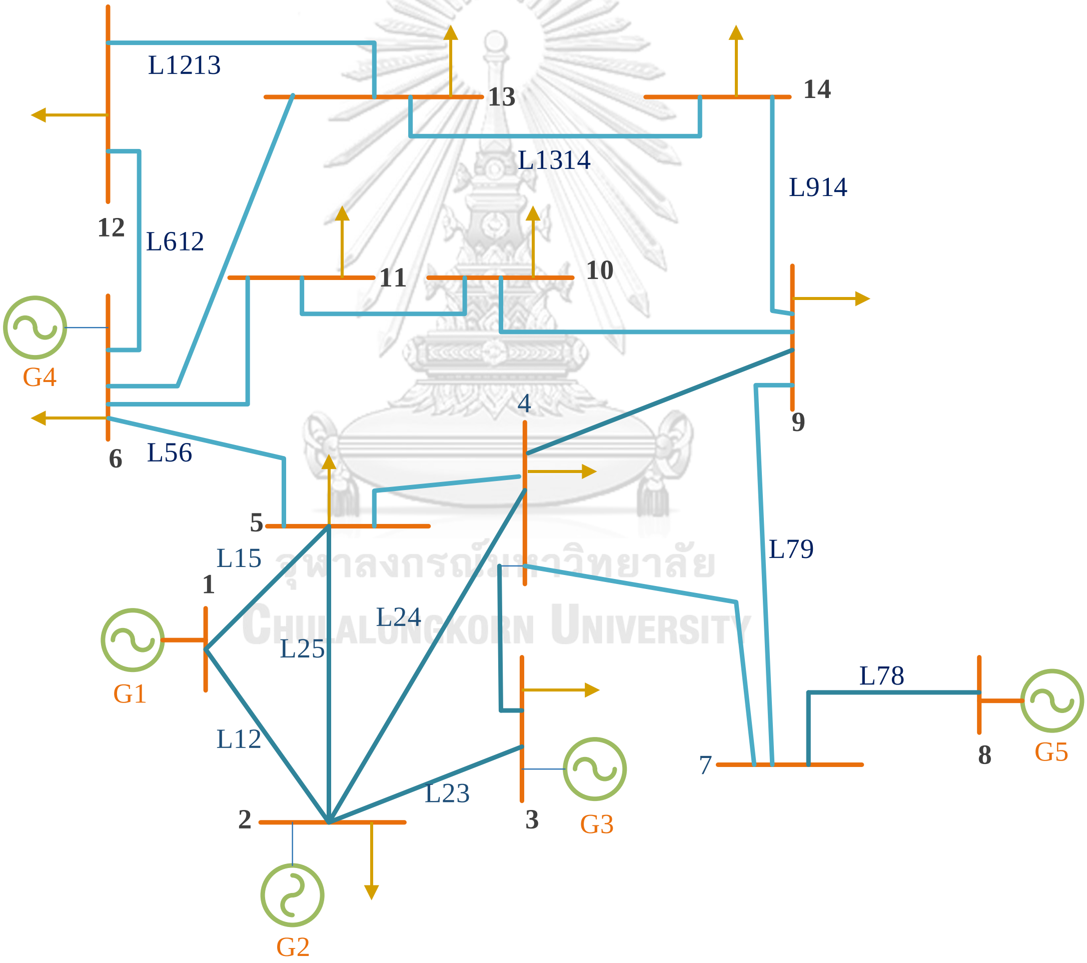


Figure 30 IEEE 14 bus system

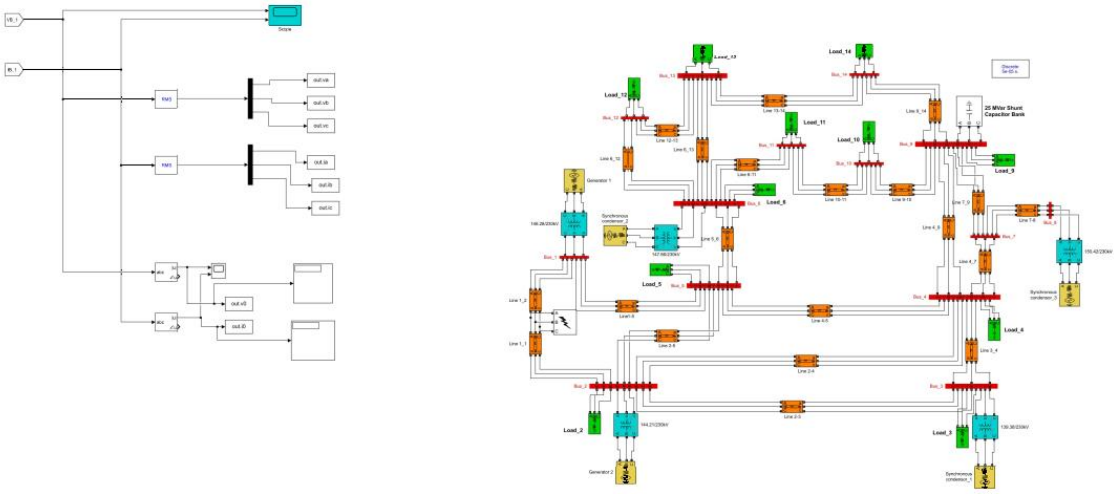


Figure 31 Snapshot of the IEEE 14 modeled in MATLAB

3.4 Fault occurrence & database generation

Various faults are applied at different fault resistance ranging from 0.01 to 200 Ω and a database is generated by applying faults at lines L12, L15, and L56 to create the database. The fault resistance ranging from 0.01 to 200 ohms was applied at line and RMS values of three-phase voltage and current (Ir, Iy, Ib, Vr, Vy, Vb) & zero sequence current, and voltage (I0, V0) collected from bus 1.

Table 3 shows different fault conditions simulated, and binary bits are assigned to classify between various fault types as given in Table 3, where ABC is the representation of three phases and D is representation of ground phase. When there is no fault and phase current, and voltage doesn’t fluctuate then the binary bit [0 0 0 0] indicating that there is no fault in all the phase and if there is fault in A phase bit assigned is [1 0 0 0].

Table 3 Sample of Fault type binary assignment

A	B	C	D	Fault Type
0	0	0	0	Normal
1	0	0	1	AG
0	1	0	1	BG
0	0	1	1	CG
1	1	0	1	ABG
1	0	1	1	ACG
0	1	1	1	BCG
1	1	1	1	ABCG
1	1	1	0	ABC
1	1	0	0	AB
1	0	1	0	AC
0	1	1	0	BC

To generate the data various shunt faults which are the most frequent in lines applied at L12, L15, L23, and L56 as indicated in Figure 30. The shunt faults are classified as a. Single line to ground (SLG): When a single-phase encounters ground and the remaining two phases remain intact. b. Double line to bottom (LLG): When two phases encounter ground and a single phase remains intact. c. The triple line to ground fault (LLLG): When all the three-phase encounter ground. d. Double line (LL): When two phases come in contact with each other. e. The triple line (LLL): When the three-phase comes in contact with each other.

All the above-mentioned faults are applied with fault resistance ranging from 0.01 to 200 ohms. and the RMS values of three phases and zero sequence current & voltage (Ir, Iy, Ib, Vr, Vy, Vb, I0, V0) are stored. Then the data is labeled as per the actual fault class & classification. A total of 5200 data was collected through multiple simulations.

Table 4 Data distribution

The training set 70% of 5200	Validation set. 15% of 5200	Testing set 15% of 5200
3640	780	780

3.5 Algorithm

The research data underwent analysis and was transformed into a predictive model using the Levenberg-Marquardt algorithm, also known as the damped least-squares (DLS) approach, as referenced in [18]. This algorithm is commonly used to address non-linear least squares problems, particularly when fitting least squares curves. The LMA incorporates techniques from both the Gauss-Newton algorithm and gradient descent, making it more robust than GNA and able to find a solution even when starting far from the minimum.

However, it may run slower than GNA for well-behaved functions and appropriate initial values. Additionally, the trust region technique to Gauss-Newton can also be applied to LMA. Donald Marquardt and independent researchers Girard, Wynne, and Morrison discovered the algorithm in 1963, as mentioned in[18]. The algorithm process flow can be understood via illustration in Figure 32 that provides a visual representation of the algorithm's process flow, and this method is deduced to resolve the drawbacks of the impedance-based methods.

The algorithm uses three phase RMS current & voltage , zero sequence current and voltage from the stored database as the input. The algorithm then distributes the database into training, testing and validation as indicated in Table 4. Next various SML methods like LDA, KNN, DT, NN and SVM are used for detection and classification. The methods like LSM, DT and LR are used for location as indicated in Figure 33. Once the network is trained an additional 15% new sets of data are simulated to check the efficiency of the trained network. Once the process is completed, the algorithm generates the results followed by the confusion matrix.

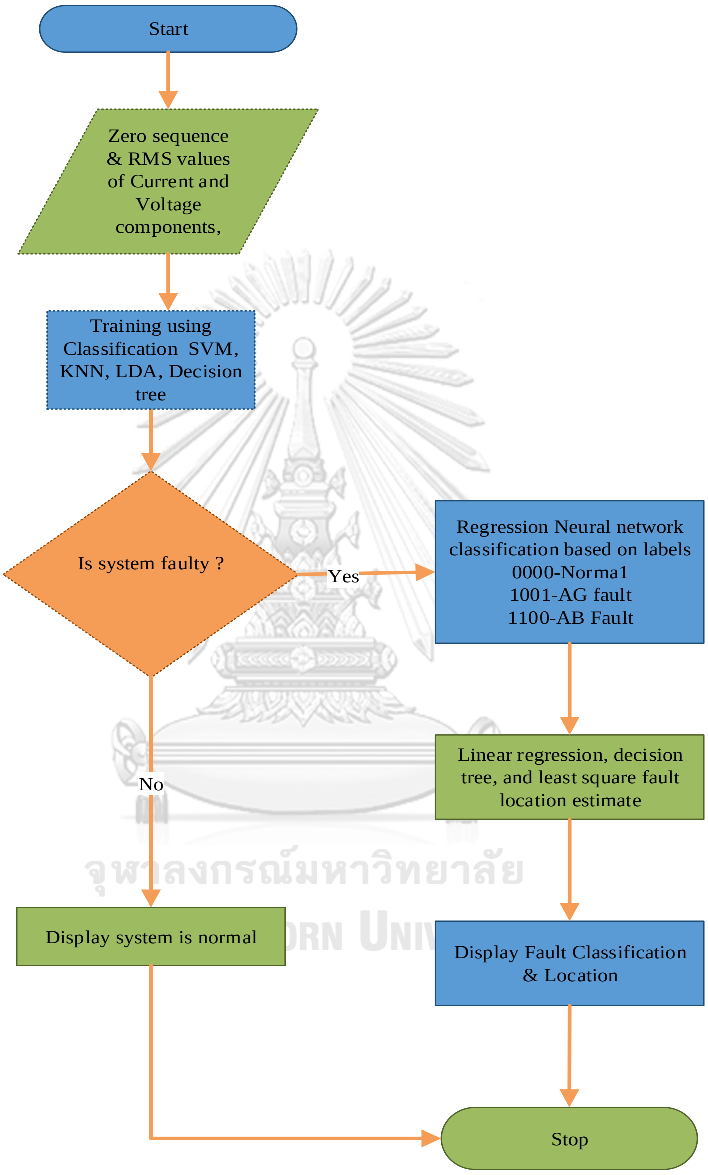

Start
Zero sequence & RMS values of Current and Voltage components,
Training using Classification SVM, KNN, LDA, Decision tree
Is system faulty ?
Yes
Regression Neural network classification based on labels
0000-Norma1
1001-AG fault
1100-AB Fault
Linear regression, decision tree, and least square fault location estimate
No
Display system is normal
Display Fault Classification & Location
Stop

Figure 32 Supervised Machine Learning based flow chart.

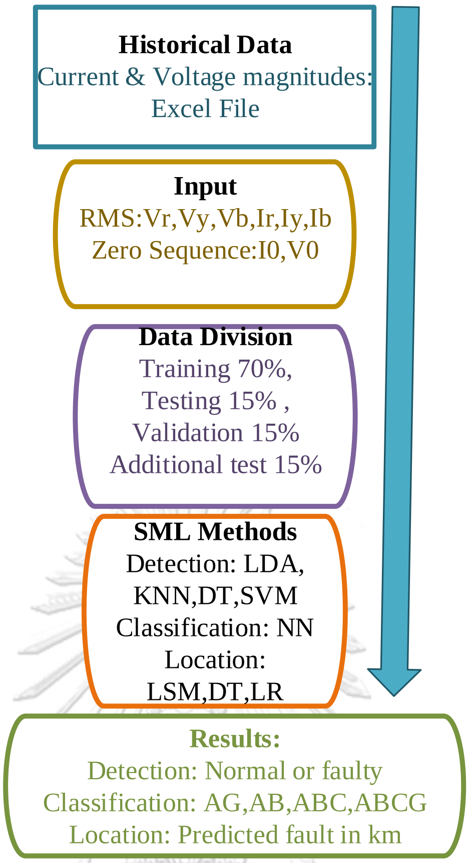


Figure 33 AI-based algorithm process flow

Chapter 5 Simulation and results

The developed algorithm's efficiency and accuracy is assessed using IEEE 14 bus system under various case studies. The simulations and results of detection, classification and location are analyzed, and various case studies are considered.

5.1 Case studies with training set without errors & outages

Developed supervised machine learning(SML) based fault detection and classification algorithm and carried out case studies considering current & voltage transformer (CT&VT) errors, generator, and line outages. The algorithm employed various SML techniques such as linear discriminant analysis (LDA), support vector machine (SVM), K-nearest neighbor (KNN), neural network (NN), and decision tree (DT) using single-ended measurement. The RMS values of three-phase voltage & current, as well as zero sequence voltage & current stored during normal & faulty conditions, are used as input and labeled class & fault classifications as outputs. The algorithm is tested using an additional test data set of different cases and accuracy compared with the SML techniques. To assess the algorithm case studies without error , with CT , VT errors , generator outage and line under maintenance is considered. Detailed case studies are given below.

5.1.1 Fault detection without measurement errors

Using the 5000 data sets the training & testing was executed as per Table 5. Three models were able to detect and identify the faulty efficiently as in Table 6 and only KNN performance was 91.7%. The detection results using LDA indicated as in confusion matrix in Figure 34 . The confusion matrix is a table used to evaluate the results of the classification model and the matrix gives the visual representation of the relationship between actual and predicted data.

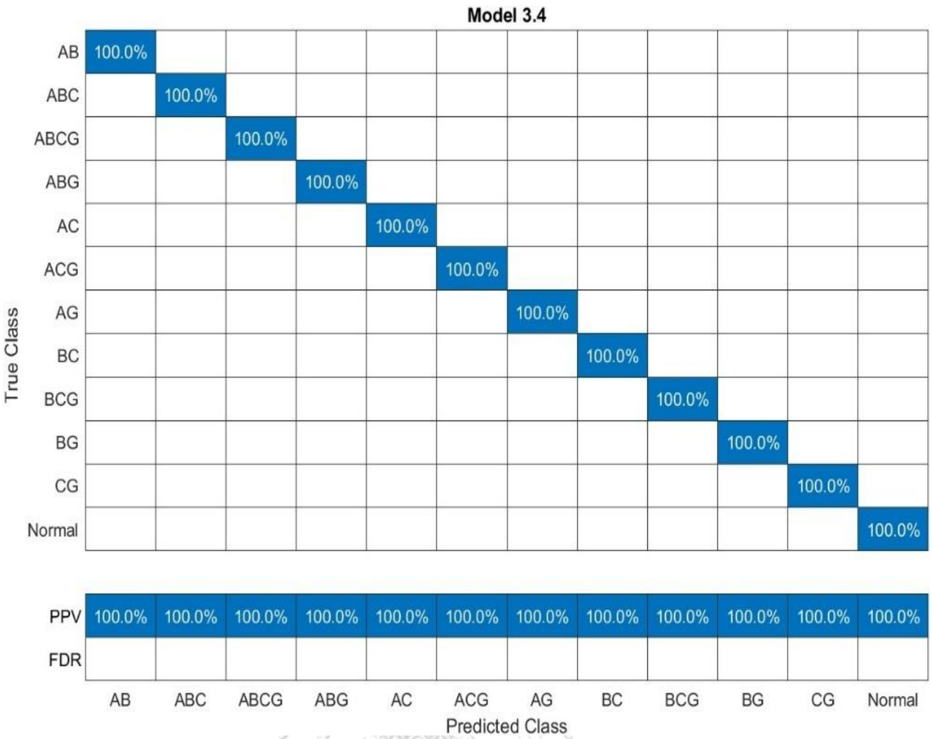


Figure 34 Detection using LDA result in positive predictive values (PPV) & false discovery rate (FDR)

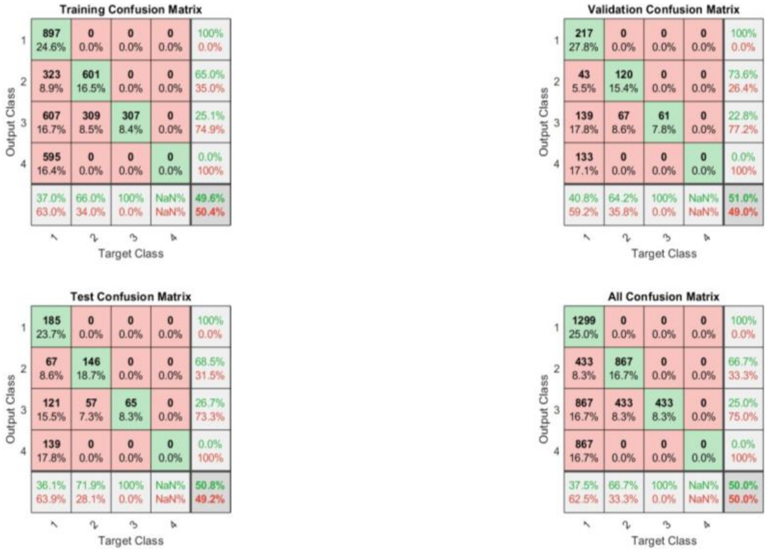
The positive predictive values (PPV) & false discovery rate (FDR) are the performance metric that provide the insights into accuracy and reliability of positive prediction made by the model.

Table 5 Detection accuracy comparison using different models

Sl. No	Machine Learning Models	Accuracy %
1	KNN	91.70%
2	SVM	100%
3	Decision Tree	100%
4	Linear Discriminant	100%

5.1.2 Fault classification without measurement errors

The regression neural network models are used to classify based on the assigned label and 100% accuracy is achieved as indicated by confusion matrix in Figure 35.



Training Confusion Matrix

	1	2	3	4	
1	897 24.6%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	323 8.9%	601 16.5%	0 0.0%	0 0.0%	65.0% 35.0%
3	607 16.7%	309 8.5%	307 8.4%	0 0.0%	25.1% 74.9%
4	595 16.4%	0 0.0%	0 0.0%	0 0.0%	0.0% 100%
	37.0% 63.0%	66.0% 34.0%	100% 0.0%	NaN% NaN%	49.6% 50.4%

Validation Confusion Matrix

	1	2	3	4	
1	217 27.8%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	43 5.5%	120 15.4%	0 0.0%	0 0.0%	73.6% 26.4%
3	139 17.8%	67 8.6%	61 7.8%	0 0.0%	22.8% 77.2%
4	133 17.1%	0 0.0%	0 0.0%	0 0.0%	0.0% 100%
	40.8% 59.2%	64.2% 35.8%	100% 0.0%	NaN% NaN%	51.0% 49.0%

Test Confusion Matrix

	1	2	3	4	
1	185 23.7%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	67 8.6%	146 18.7%	0 0.0%	0 0.0%	68.5% 31.5%
3	121 15.5%	57 7.3%	65 8.3%	0 0.0%	26.7% 73.3%
4	139 17.8%	0 0.0%	0 0.0%	0 0.0%	0.0% 100%
	36.1% 63.9%	71.9% 28.1%	100% 0.0%	NaN% NaN%	50.8% 49.2%

All Confusion Matrix

	1	2	3	4	
1	1299 25.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	433 8.3%	867 16.7%	0 0.0%	0 0.0%	66.7% 33.3%
3	867 16.7%	433 8.3%	433 8.3%	0 0.0%	25.0% 75.0%
4	867 16.7%	0 0.0%	0 0.0%	0 0.0%	0.0% 100%
	37.5% 62.5%	66.7% 33.3%	100% 0.0%	NaN% NaN%	50.0% 50.0%

Figure 35 Classification confusion matrix NN

The neural network confusion matrix provides valuable information regarding the correlation between the actual class labels and the predicted class labels in a classification task. In a confusion matrix, the rows represent the actual class labels of the data, while the columns represent the predicted class labels generated by the neural network model. Each cell in the matrix represents the number of data instances that belong to a specific class according to the actual labels and are predicted correctly or incorrectly by the model. By examining the confusion matrix, we can gain insights into how well the neural network model is performing for each class. The diagonal cells of the matrix represent the correctly predicted instances, indicating a strong correlation between the actual and predicted class labels. Conversely, the off-diagonal cells represent instances that are misclassified by the model, highlighting a discrepancy between the actual and predicted class labels.

The confusion matrix helps us understand the types of errors made by the neural network. For example, it reveals if certain classes are consistently misclassified or confused with one another. By analyzing the patterns in the matrix, we can identify areas where the model may need improvement, such as providing more training data for specific classes or adjusting the model's parameters.

The neural network confusion matrix is a useful tool for evaluating the performance of the model in classification tasks. It provides a clear representation of the relationship between the actual and predicted class labels, allowing us to assess the model's accuracy and identify areas for further optimization.

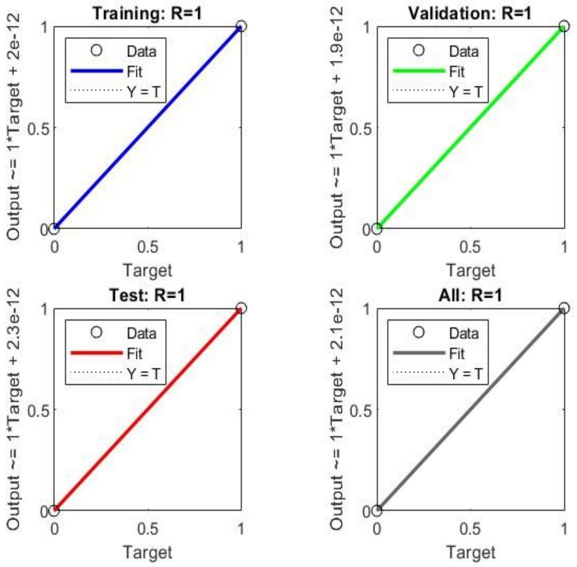


Figure 36 Regression plot using neural network.

Figure 36 depicts the relationship between actual and predicted values and the plot helps in visually understanding the correlation between the actual and predicted values. The regression R=1 indicates 100% accuracy as the model could identify all 12 different fault types accurately in all the different fault conditions. The classification pattern recognition tool was also used to classify the labeled data set, but the regression neural network outperformed the classification NN.

5.1.3 Fault location without measurement errors

The fault data collected at different fault locations are used and the algorithm predicts the output. The testing data consists of faults applied at the step of 4 km at fault resistance ranging from 0.01 to 150 ohms and the total line length is 44.47km. The plot is generated by an algorithm as indicated in Figure 37 that compares the overall accuracy using the three methods. The accuracy of least squares is 79.85%, followed by linear regression at 85.40% and the decision tree at 97.26%.

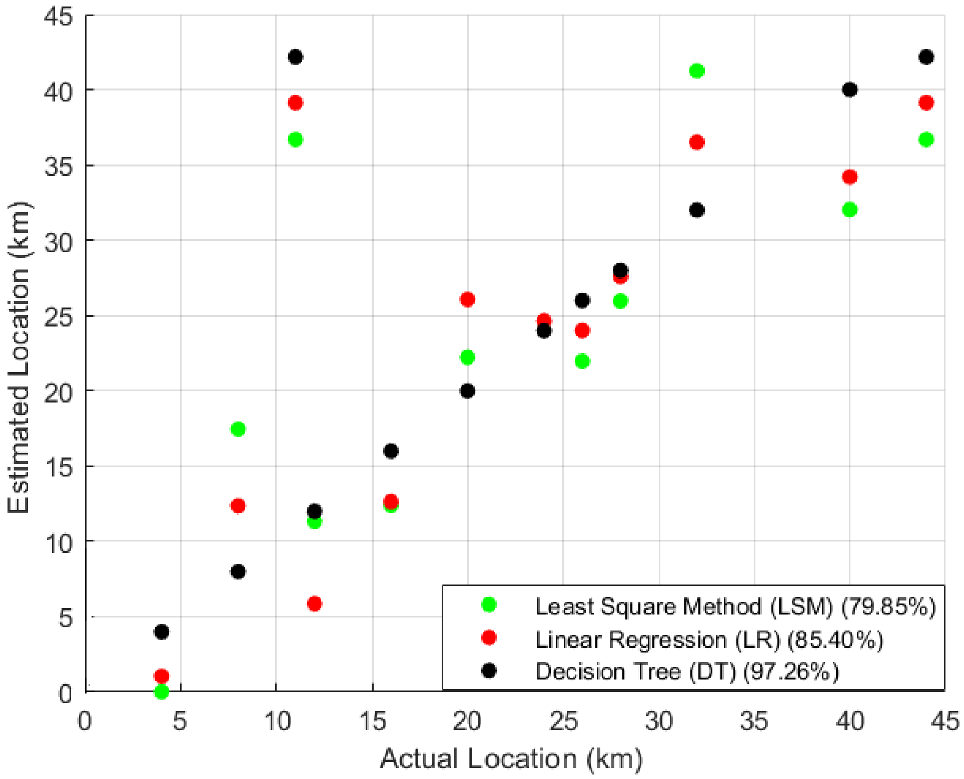


Figure 37 Overall accuracy for all fault locations

The algorithm first trains the network using the initial training, validation, and testing data. The accuracy detection was 100% where the algorithm precisely distinguishes between normal and faulty classes among 780 total testing observations. Similarly, the classification accuracy of 100% where the algorithm accurately classified all the fault types when error-free test data was used.

Initially, the network was trained using fault-free simulated datasets. These datasets were carefully designed to represent normal operating conditions without any faults or anomalies. The purpose of training the network with these fault-free datasets was to teach it the patterns and characteristics of normal data, enabling it to establish a baseline for comparison.

After the training phase, the performance of the algorithm was evaluated using a new dataset. This new dataset consisted of various cases or scenarios, each representing different fault conditions or anomalies. These cases were specifically designed to challenge the network and assess its ability to accurately detect and classify faults.

By testing the algorithm on this new dataset, aimed to evaluate its performance and measure its effectiveness in identifying and categorizing faults. The different cases within the dataset allowed for a comprehensive assessment of the algorithm's robustness and adaptability across a range of fault scenarios.

5.1.4 Detection & classification with 3% CT error

In this case, the additional test set with the current measurement with a 3% CT error is used as a test set. The test set is utilized to test with the already trained network. Table 1&2 shows detection and classification accuracy using different models SML models with a 3% error in CT measurement. The model performed poorly in the case of detection. The SVM & NN performed well with 91.66% accuracy, but DT outperformed even in this case with 100% accuracy for classification.

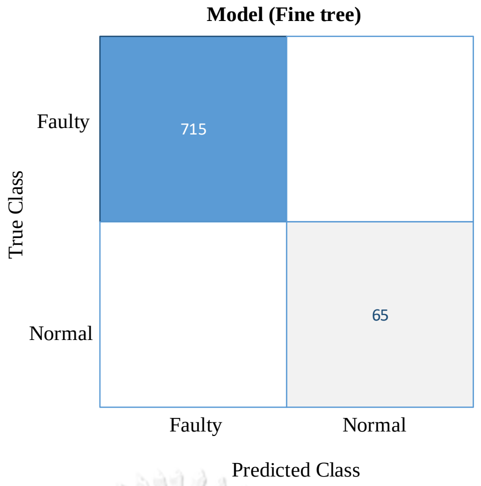


Figure 38 Detection test confusion matrix using DT

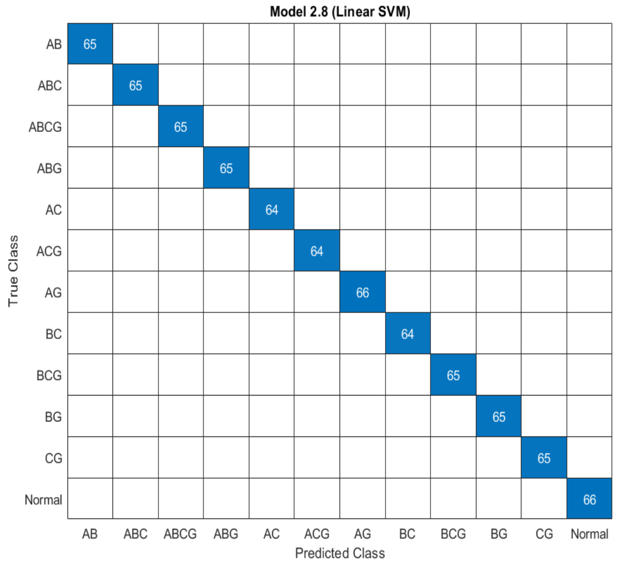


Figure 39 Classification test confusion matrix using SVM

5.1.5 Detection & classification with 3% VT error

In this case, the additional test set with VT measurement with a 3% error is used. While testing it was found that most of the SML models performed poorly but DT outperformed all with 100% accuracy in detection and LDA with 67% accuracy in classification indicated in Table 7. This shows that accuracy decreases with measurement errors in most cases.

5.1.6 Detection & classification with 3% error in CT & VT

An additional test dataset was used, which introduced a 3% error in both the Current Transformer (CT) and Voltage Transformer (VT) measurements. This dataset was designed to simulate a scenario where measurement errors are present, challenging the performance of the models. Upon evaluating the performance of various models on this additional test dataset, it was observed that all the models performed poorly. Despite the presence of measurement errors, the Linear Discriminant Analysis (LDA) model achieved a classification accuracy of 67%, as indicated in Tables 6 & 7.

The observed performance differences among the models can be attributed to their unique capabilities and the features available in the dataset. Each machine learning model has its own strengths and weaknesses, and their accuracy is influenced by the characteristics of the dataset and the specific problem being addressed.

It is important to note that the introduction of measurement errors can significantly impact the performance of the models. The 3% error in both CT and VT measurements likely affected the models' ability to accurately detect and classify faults. The complexity introduced by the measurement errors requires the models to adapt and account for these uncertainties, which can be challenging.

The variation in accuracy among the models suggests that some models may be more robust or better suited to handle measurement errors compared to others. The LDA model, despite the presence of errors, achieved a relatively higher classification accuracy of 67%. This implies that the LDA model's inherent capabilities and its ability to leverage the available features in the dataset allowed it to perform relatively better under the given conditions.

In summary, the utilization of an additional test dataset with a 3% error in CT and VT measurements revealed poor performance across all models, indicating the challenges posed by measurement errors. The varying accuracy among the models highlights the importance of selecting models that are well-suited for the problem at hand and considering their capabilities and the features present in the dataset. The accuracy achieved depends on the models' ability to handle measurement errors and leverage the available information accurately.

Table 6 Detection accuracy for different cases

Sl. No	Machine Learning Models	Error-free	3% VT error	3% CT error	3% CT& VT error
		%Acc	%Acc	%Acc	%Acc
1	KNN	100%	91.78%	8.22%	91.78%
2	LDA	100%	91.78%	8.22%	8.20%
3	NN	100%	8.30%	8.22%	8.20%
4	SVM	100%	8.30%	8.22%	8.20%
5	DT	100%	100%	8.22%	8.20%

Table 6 presents the detection accuracy of various machine learning models under different scenarios. The scenarios considered in the table include error-free conditions, a 3% CT error, a 3% CT and VT (Voltage Transformer) error, and a VT error. The table displays the accuracy values in percentage (%Acc) for each model and scenario.

1. KNN: The K-nearest neighbors (KNN) model achieved 100% accuracy in the error-free scenario. However, its accuracy dropped to 91.78% in the presence of a 3% CT error, 8.22% in the case of a 3% CT and VT error and returned to 91.78% in the VT error scenario.

2. LDA: The Linear Discriminant Analysis (LDA) model achieved 100% accuracy in the error-free scenario. However, its accuracy dropped to 91.78% in the presence of a 3%

CT error, 8.22% in the case of a 3% CT and VT error, and further decreased to 8.20% in the VT error scenario.

3. NN: The Neural Network (NN) model achieved 100% accuracy in the error-free scenario. However, its accuracy dropped to 8.30% in the presence of a 3% CT error, 8.22% in the case of a 3% CT and VT error and remained at 8.20% in the VT error scenario.

4. SVM: The Support Vector Machine (SVM) model achieved 100% accuracy in the error-free scenario. However, its accuracy dropped to 8.30% in the presence of a 3% CT error, 8.22% in the case of a 3% CT and VT error and remained at 8.20% in the VT error scenario.

5. DT: The Decision Tree (DT) model achieved 100% accuracy in the error-free scenario. However, its accuracy dropped to 100% in the presence of a 3% CT error, 8.22% in the case of a 3% CT and VT error and remained at 8.20% in the VT error scenario.

The table provides a comparative analysis of the machine learning models' detection accuracy under different error scenarios. It highlights the models' performance degradation when measurement errors are introduced. The accuracy values help assess the models' robustness and reliability in fault detection tasks. Based on the findings, it appears that the models are particularly sensitive to CT and VT errors, leading to decreased accuracy. These insights can guide the selection and improvement of machine learning models for fault detection in power systems.

Table 7 Classification accuracy for different cases

Sl. No	Machine Learning Models	Error-free	3% VT error	3% CT error	3% CT& VT error
		%Acc	%Acc	%Acc	%Acc
1	KNN	100%	8%	8.22%	8.34%
2	LDA	100%	67%	67%	67%
3	NN	100%	8%	91.66%	8.34%
4	SVM	100%	8%	91.66%	8.34%
5	DT	100%	17%	100%	8.20%

Table 8 presents the classification accuracy of various machine learning models under different scenarios. The scenarios considered in the table include error-free conditions, a 3% VT (Voltage Transformer) error, a 3% CT (Current Transformer) error, and a 3% CT and VT error. The table displays the accuracy values in percentage (%Acc) for each model and scenario.

1. KNN: The K-nearest neighbors (KNN) model achieved 100% accuracy in the error-free scenario. However, its accuracy dropped to 8% in the presence of a 3% VT error, 8.22% in the case of a 3% CT error, and slightly increased to 8.34% in the presence of a 3% CT and VT error.
2. LDA: The Linear Discriminant Analysis (LDA) model achieved 100% accuracy in the error-free scenario. However, its accuracy dropped to 67% in the presence of a 3% VT error, 67% in the case of a 3% CT error, and remained at 67% in the presence of a 3% CT and VT error.
3. NN: The Neural Network (NN) model achieved 100% accuracy in the error-free scenario. However, its accuracy dropped to 8% in the presence of a 3% VT error, increased to 91.66% in the case of a 3% CT error, and remained at 8.34% in the presence of a 3% CT and VT error.

4. SVM: The Support Vector Machine (SVM) model achieved 100% accuracy in the error-free scenario. However, its accuracy dropped to 8% in the presence of a 3% VT error, increased to 91.66% in the case of a 3% CT error, and remained at 8.34% in the presence of a 3% CT and VT error.

5. DT: The Decision Tree (DT) model achieved 100% accuracy in the error-free scenario. However, its accuracy dropped to 17% in the presence of a 3% VT error, increased to 100% in the case of a 3% CT error, and remained at 8.20% in the presence of a 3% CT and VT error.

The table provides a comparative analysis of the machine learning models' classification accuracy under different error scenarios. It reveals the models' performance degradation when measurement errors are introduced, particularly in the VT measurements. The accuracy values help evaluate the models' reliability in fault classification tasks. Based on the findings, it appears that the models perform differently in the presence of VT and CT errors, with varying degrees of accuracy. These insights can guide the selection and improvement of machine learning models for fault classification.

5.1.7 Detection & classification when the generator G2 is out of service

An additional dataset was generated by simulating a scenario where the generator G2 is out of service. This dataset was used to test the performance of an already-trained neural network. However, the models exhibited poor performance in both fault detection and classification tasks when tested with this additional dataset.

The results suggest that the absence of the G2 generator has a significant impact on the performance of the models. It implies that the fault current magnitude plays a crucial role in fault detection and classification.

The poor performance of the models in both detection and classification tasks indicates that the absence of the G2 generator and the resulting changes in fault current magnitude pose challenges for accurate fault identification and categorization. These findings highlight the importance of considering such scenarios during the training and evaluation of the neural network models.

To address this issue and improve the models' performance, it may be necessary to incorporate data that captures a wider range of fault current magnitudes, including scenarios with the generator out of service. By training the models on a more diverse dataset that encompasses various fault conditions and magnitudes, the models can learn to handle these scenarios more effectively and make accurate predictions.

In conclusion, the simulation of the additional dataset with the G2 generator out of service revealed poor performance in fault detection and classification tasks. This indicates the significance of the fault current magnitude and the challenges associated with accurately identifying and categorizing faults in such scenarios. Considering these factors and incorporating diverse training data can enhance the models' performance and improve fault detection and classification accuracy.

Table 8 Detection accuracy

Sl. No	Machine Learning Models	Error-free	G2-out of service	L25-outage
		%Acc	%Acc	%Acc
1	KNN	100%	0.00%	91.67%
2	LDA	100%	8.30%	53.21%
3	NN	100%	0.00%	91.67%
4	SVM	100%	0.00%	91.67%
5	DT	100%	8.33%	53.21%

Table 9 Classification accuracy

Sl. No	Machine Learning Models	Error-free	G2-out of service	L25-outage
		%Acc	%Acc	%Acc
1	KNN	100%	0.00%	0.00%
2	LDA	100%	8.30%	8.33%
3	NN	100%	0.00%	0.00%
4	SVM	100%	0.00%	0.00%
5	DT	100%	8.33%	8.33%

5.1.8 Detection & classification when line L2 is under shutdown /outage

Line L25 is considered to be under shutdown as labeled in Figure 30. The additional test set is simulated and utilized to test with an already-trained network. The models performed quite well for detection and poorly for classification as indicated in Table 9 & 10. This indicates that the training model data set should have sufficient data on such events, so that accuracy and reliability can be achieved.

Table 8 & 9 represent the classification and detection performance of different machine learning models considering three different cases: error-free , the G2 generator out of service, and L25 transmission line outage. The table provides accuracy values for each model in percentage (%Acc) for each scenario.

1. KNN: The K-nearest neighbors (KNN) model achieved 100% accuracy in all scenarios, meaning it made correct predictions for all instances.
2. LDA: The Linear Discriminant Analysis (LDA) model achieved 100% accuracy in the error-free scenario. However, its accuracy dropped to 8.30% when the G2 generator was out of service and 8.33% during the L25 transmission line outage.
3. NN: The Neural Network (NN) model achieved 100% accuracy in all scenarios, like the KNN model.
4. SVM: The Support Vector Machine (SVM) model achieved 100% accuracy in all scenarios, just like the KNN and NN models.
5. DT: The Decision Tree (DT) model achieved 100% accuracy in the error-free scenario. However, its accuracy dropped to 8.33% when the G2 generator was out of service and during the L25 transmission line outage.

It suggests that KNN, NN, and SVM models performed consistently well, achieving 100% accuracy in all scenarios. On the other hand, the LDA and DT models experienced reduced accuracy when specific components (G2 generator or L25 transmission line) were out of service. The table provides valuable insights into the models' behavior in different scenarios and can aid in selecting the most appropriate model for the given task.

Table 10 Comparison of supervised machine learning models

Sl. No	Algorithm	Description	Strengths	Weaknesses
1	K-nearest neighbors (KNN)	Classifies a new data point based on the majority class of its k nearest neighbors.	Simple to understand and implement. Works well with noisy data.	Can be computationally expensive for large datasets.
2	Decision trees (DT)	Creates a tree-like structure to classify data.	Easy to interpret. Can handle both categorical and continuous data.	Can be sensitive to overfitting.
3	Linear regression (LR)	Predicts a continuous value based on a linear combination of features.	Easy to understand and implement.	Can be sensitive to outliers.
4	Linear discriminant analysis (LDA)	Finds a linear combination of features that separates two or more classes.	Easy to understand and implement.	Can be sensitive to outliers.
5	Support vector machines (SVM)	Finds a hyperplane that separates two or more classes.	Very accurate for classification tasks.	Can be computationally expensive for large datasets.
6	Neural networks (NN)	A network of connected nodes that learn to predict a value based on a set of inputs.	Very accurate for both classification and regression tasks.	Can be computationally expensive to train.

5.2 Case studies with training set with errors and outages

In this section the model is trained using the database that comprises of all the fault instances both with error and without error. The capability of the model is checked and compared.

5.2.1 Detection with VT/CT error

Table 11 shows the detection accuracy of various machine learning models when errors in the training set are taken into consideration. The errors are categorized as VT (Voltage transformer) errors and CT (current transformer) errors. The table presents the accuracy percentages for each model under three scenarios: 3% VT error, 3% CT error, and 3% CT & VT error. The results obtained are as follows:

KNN: Achieves 100% accuracy in all three scenarios.

2. LDA: Maintains an accuracy of 91.8% in all three scenarios.

3. NN: Shows a significant drop in accuracy, achieving only 8.2% accuracy in the 3% VT error and 3% CT & VT error scenarios, but improves to 91.8% accuracy in the 3% CT error scenario.

4. SVM: Similarly, to NN, it exhibits low accuracy of 8.2% in the 3% VT error and 3% CT & VT error scenarios but improves to 91.8% accuracy in the 3% CT error scenario.

5. DT: Attains 100% accuracy in all three scenarios.

5.2.2 Detection with line outage & generator out of service

Similarly, Table 12 presents the detection accuracy of various machine learning models when outage data is considered in the training set. The outages are categorized as G2 (Generation 2) out-of-service and L25 outages. The following results are obtained.

1. KNN: Achieves 100% accuracy in detecting both G2 out-of-service and L25 outages.

2. LDA: Maintains an accuracy of 75% in detecting both G2 out-of-service and L25 outages.

3. NN: Shows a high accuracy of 91.67% in detecting both G2 out-of-service and L25 outages.

4. SVM: Similarly, to NN, it achieves 91.67% accuracy in detecting both G2 out-of-service and L25 outages.

5. DT: Attains 100% accuracy in detecting both G2 out-of-service and L25 outages.

These results indicate the performance of the machine learning models in accurately detecting outages when the training set comprises of outage instances along with the error instances.

Table 11 Detection accuracy with errors considered in training set

Sl. No	Machine Learning Models	3% VT error	3% CT error	3% CT& VT error
		%Acc	%Acc	%Acc
1	KNN	100%	100%	91.78%
2	LDA	91.8%	91.8%	91.78%
3	NN	8.2%	91.8%	8.22%
4	SVM	8.2%	91.8%	8.22%
5	DT	100%	100%	100%

Table 12 Detection accuracy with outage data considered in training set

Sl. No	Machine Learning Models	G2-out of service	L25-outage
		%Acc	%Acc
1	KNN	100%	100%
2	LDA	75%	75%
3	NN	91.67%	91.67%
4	SVM	91.67%	91.67%
5	DT	100%	100%

Table 13 Classification accuracy with errors considered in training set

Sl. No	Machine Learning Models	3% VT error	3% CT error	3% CT& VT error
		%Acc	%Acc	%Acc
1	KNN	100%	100%	100%
2	LDA	100%	100%	100%
3	NN	33.4%	41.7%	25%
4	SVM	33.4%	41.7%	25%
5	DT	100%	100%	100%

Table 14 Classification accuracy with outage data considered in training set

Sl. No	Machine Learning Models	G2-out of service	L25-outage
		%Acc	%Acc
1	KNN	100%	100%
2	LDA	71.03%	53.21%
3	NN	25%	33.33%
4	SVM	25%	33.33%
5	DT	100%	100%

5.2.3 Classification with VT/CT error

Table 13 provides information on the classification accuracy of different machine learning models when errors are considered in the training set. The errors are categorized as VT errors and CT errors. Here are the results for each model under three scenarios: 3% VT error, 3% CT error, and 3% CT & VT error:

1. KNN: Achieves 100% accuracy in all three scenarios, with the errors present in the training set.

2. LDA: Also maintains 100% accuracy in all three scenarios, considering errors in the training set.

3. NN: Shows a drop in accuracy, achieving 33.4% accuracy in the 3% VT error scenario, 41.7% accuracy in the 3% CT error scenario, and 25% accuracy in the 3% CT & VT error scenario.

4. SVM: Similar to NN, it exhibits reduced accuracy, achieving 33.4% accuracy in the 3% VT error scenario, 41.7% accuracy in the 3% CT error scenario, and 25% accuracy in the 3% CT & VT error scenario.

5. DT: Attains 100% accuracy in all three scenarios, even when errors are present in the training set.

These results indicate the performance of the machine learning models in classifying data accurately, taking into account the specified errors in the training set.

5.2.4 Classification with line outage & generator out of service

Table 14 provides information on the classification accuracy of different machine learning models when outage data is considered in the training set. The outages are categorized as G2 (Generation 2) out-of-service and L25 outages.

Let's examine the results:

1. KNN: Demonstrates excellent accuracy, achieving 100% in both G2 out-of-service and L25 outage detection.
2. LDA: Shows a reasonably high accuracy of 71.03% in detecting G2 out-of-service and 53.21% in detecting L25 outages.
3. NN: Exhibits lower accuracy, achieving 25% in G2 out-of-service detection and 33.33% in L25 outage detection.
4. SVM: Similarly, to NN, it also achieves 25% accuracy in G2 out-of-service detection and 33.33% accuracy in L25 outage detection.
5. DT: Performs well, achieving 100% accuracy in both G2 out-of-service and L25 outage detection.

These results indicate the performance of the machine learning models in accurately classifying outages, considering the specified outage data in the training set.

5.3 Effect of fault type in fault location

In this section, the effects of different fault types on the algorithm were tested, and the results were analyzed. The impact of fault types on the algorithm's performance was evaluated based on the accuracy percentages shown in Table 15.

According to the table, it can be observed that the fault type did not have a significant impact on the algorithm's performance. This can be attributed to the fact that the training dataset used for the algorithm consisted of instances of all 12 faults listed in Table 3. The inclusion of diverse fault types in the training dataset ensured that the algorithm was exposed to a wide range of faulted conditions.

The accuracy percentages presented in Table 15 demonstrate that the algorithm performed consistently well across different faulted conditions. Regardless of the specific fault type, the algorithm achieved high accuracy percentages, indicating its robustness and ability to handle various fault scenarios.

These findings suggest that the algorithm's training with a comprehensive dataset contributed to its effectiveness in accurately detecting and classifying faults, regardless of the specific fault type encountered. The algorithm's performance demonstrates its reliability and suitability for fault detection and analysis tasks in real-world applications.

Table 15 Effect of fault-on location

Sl. No	Machine Learning Models	SLGF	LLGF	LL	LLL	LLLG
		%Acc	%Acc	%Acc	%Acc	%Acc
1	LSM	81.3%	82%	79%	76%	74.4%
2	LR	85%	84.2%	82.4%	81%	80.2%
5	DT	98%	97%	95%	94.3%	93%

5.4 Effect of coupling in fault location

The coupling effect is the phenomenon that ours in the distribution line when two adjacent phases of the conductor intersect. The coupling occurs due to proximity of conductor or the presence of common impedance. The effect of coupling in fault location using signals propagation makes the location process complicated as per reference [21], as the fault signal may propagate through the line and affect the adjacent phase.

Since the research uses simulated event of various types of faults with RMS values of three phase current and voltage along with the sequence component current and voltage recorded, recording of such instances in the simulation is out of scope. However, there is a scope in future research using the signal database to develop the machine learning model and check the effect of coupling.

Chapter 6 Conclusion

6.1 Discussion

After conducting several case studies, it was observed that the SML (specific machine learning) model exhibited poor performance in both fault detection and classification when the generator was out of service. Additionally, during a scenario where one transmission line was shut down, the NN (Neural Network) and SVM (Support Vector Machine) models demonstrated good performance in fault detection but performed poorly in fault classification.

These findings highlight the importance of considering specific instances, such as generator outages and line shutdowns, during the training phase of the machine learning models. By including such instances in the training data, the models can be better equipped to accurately detect and classify faults in similar situations.

Moreover, other factors that should be considered during the training phase are load changes, capacitive banks, and reactors. These components play a crucial role in the overall behavior and dynamics of a power system. Incorporating information about load variations and the presence of capacitive banks and reactors in the training data will enhance the models' ability to make accurate predictions.

By training multiple SML models on comprehensive and diverse datasets that incorporate various scenarios and system conditions, it becomes possible to improve the accuracy and reliability of fault detection and classification. These models can then be used collectively to make more precise predictions, taking into account the different factors and situations that can affect the power system's behavior.

In conclusion, the findings from the case studies emphasize the need to include specific instances, such as generator outages and line shutdowns, in the training data of machine learning models for fault detection and classification. Additionally, considering load changes, capacitive banks, and reactors in the training phase enhances the models' predictive capabilities. By adopting these approaches, multiple SML models can be leveraged to make accurate predictions and improve the overall performance of fault detection and classification in power system.

6.2 Future work

Future work with the proposed SML method can be carried out as indicated below.

a. Data Collection

To begin the process, we need to collect disturbance event records in substations. These records provide valuable information about fault events that occur in the power system. By analyzing these events, we can gain insights into the underlying causes and potential fault types. The collection process involves gathering data from various substations, representing a diverse range of fault instances and types. Alongside disturbance event records, it is crucial to record phase RMS voltage and current. These measurements offer detailed information about the electrical behavior during fault events. By capturing voltage and current values at different fault instances, we can build a comprehensive dataset for training and testing the SML model. Once the data is collected, it's time to analyze the fault events. This step involves studying the recorded disturbance events and identifying their characteristics. By understanding the patterns and signatures associated with different fault types, we can effectively label and classify the data for training purposes. In addition to disturbance event records, the SML method requires sequence records for analysis. These records capture the sequential behavior of faults and are essential for identifying fault patterns.

To collect sequence records, we utilize a sequence analyzer that captures and stores the relevant data in a structured format.

b. Data label

After the data is ready, the data labeling and preposing based on the fault type via analyzing the fault signatures and assigning into labels and classes like normal and faulty , fault classes like AB, AC, BC , ABC , ABCG. To label the collected data, we analyze the fault signatures extracted from the disturbance event records. Fault signatures refer to the distinctive characteristics exhibited by different fault types. We can determine the fault type associated with each record by analyzing these signatures. Based on the fault signatures, we assign labels and classes to the data. Labels distinguish between normal and faulty instances, while classes categorize the fault types. The classes can include AB, AC, BC, ABC, ABCG, or any other fault types identified during the analysis. This labeling process ensures that the SML model can accurately differentiate between various fault scenarios. With the labeled data at hand, we proceed to prepare the dataset for training the SML model.

c. Training & Testing

Training the model involves using a dataset that contains faulty instances in various scenarios. These faulty instances are used to teach the model how to identify and handle different types of errors or problems. The dataset is carefully curated to include a wide range of scenarios, allowing the model to learn from diverse examples.

After the training phase, the model goes through a testing process. During testing, the model's capabilities are evaluated, and its algorithm is assessed to determine its accuracy. This involves inputting new data or scenarios into the model and observing its performance. If the model doesn't perform well or its accuracy is not satisfactory, improvements can be made through a process called feature engineering.

Feature engineering involves modifying or enhancing the input features that the model uses to make predictions. By refining the features or introducing new ones, the model's accuracy can be improved. This iterative process of testing, evaluating, and enhancing the model's accuracy through feature engineering helps create a more robust and effective algorithm.

6.3 Conclusion

The study initially embarked with the conventional approach using single end measurement impedance-based method and found that the method accuracy decreases as the faulted location moves far from the measurement point. Thus, the study ventured in machine learning (ML) approach and found that supervised machine learning (SML) is widely used. However, most of the studies use one or two methods thus this study proposed a comparative study using supervised machine learning (SML) concept implemented to detect, classify, and locate the fault in the power system regardless of various fault types using single end measurement.

The ability of the AI model is explored and found that if the data consist of a certain distinctive pattern SML can understand the pattern and further fitting features enabling to design predictive model without the need for system parameters like line impedance. The proposed method has reduced the vigorous calculation process of the conventional method saving time and energy. With the study, it’s clear that there is a need to consider the measurement error, line under maintenance, and all the other factors such as load changes in the training and validation phase to precisely make the prediction using all five SML models. Overall, the study gave insights into the impact of measurement equipment errors, generator out of service, and line under maintenance in the SML models.

REFERENCES

REFERENCES

[1] BPC, "Technical Assessment for Distribution System of ESD, Thimphu."

[2] B. P. Corporation, "Smart Grid Master Plan (2020-2030)."

[3] D. S. Distribution and Customer Services Department, "Distribution System Master Plan (2020-2030)."

[4] N. S. P. Stefanidou-Voziki , B. Raison , J.L. Dominguez-Garcia "A review of fault location and classification methods in distribution grids," *DOI: 10.1016/j.epsr.2022.108031,* vol. 209, 2022.

[5] S. S. Gururajapathy, H. Mokhlis, and H. A. Illias, "Fault location and detection techniques in power distribution systems with distributed generation: A review," *Renewable and Sustainable Energy Reviews,* vol. 74, pp. 949-958, 2017, doi: 10.1016/j.rser.2017.03.021.

[6] A. Abbas and M. Al-Tak, "A Review of methodologies for Fault Location Techniques in Distribution Power System," *Iraqi Journal for Electrical and Electronic Engineering,* vol. 17, no. 2, pp. 27-37, 2021, doi: 10.37917/ijeee.17.2.4.

[7] I. Team, "IEEE Standard Requirements for Instrument Transformers," *IEEE Std C,* vol. 57.

[8] P. Stefanidou-Voziki, N. Sapountzoglou, B. Raison, and J. L. Dominguez-Garcia, "A review of fault location and classification methods in distribution grids," *Electric Power Systems Research,* vol. 209, 2022, doi: 10.1016/j.epsr.2022.108031.

[9] *IEEE Guide for Determining Fault Location on AC Transmission and Distribution Lines*, I. P. a. E. Society, 2014.

[10] M. M. Saha, "Fault location method for MV cable network," presented at the 7th International Conference on Developments in Power Systems Protection (DPSP 2001), 2001.

[11] S. M. R. Das, IEEE, "A Fault Locator for Radial Subtransmission and Distribution Lines," *IEEE 0-7803-6420-1/00/$10.00* 2000.

[12] W. P. D. R. Perera, "Fault Detection in Distribution Lines Using Artificial Neural Networks," 2017, doi: 10.13140/RG.2.2.18654.54085.

[13] K. P. Murphy, *Probabilistic machine learning: an introduction*. MIT press, 2022.

[14] W. Promrat, W. Pupatanan, and W. Benjapolakul, "Fault Cause Classification on PEA 33 kV Distribution System using Supervised Machine Learning compared to Artificial Neural Network," presented at the 2021 9th International Electrical Engineering Congress (iEECON), 2021.

[15] M. P. Jarventausta P. Verho J. Partanen, "USING FUZZY SETS TO MODEL THE UNCERTAINTY IN TBE FAULT LOCATION PROCESS OF DISTRIBUTION "WORKS," *EEE Transactions on Power Delivery, Vol. 9, No. 2, April 1994,* 1994.

[16] C.-L. C. Joe-Air Jiang, "A Hybrid Framework for Fault Detection, Classification, and Location—Part I: Concept, Structure, and Methodology," *IEEE TRANSACTIONS ON POWER DELIVERY, VOL. 26, NO. 3, JULY 2011,* 2011.

[17] D. T. L. L. a. A. Yamamoto, "Wavelet Analysis: Theory and Applications," *HewlettĆPackard Journal,* 1994.

[18] R. Peeters and E. Vrije Universiteit Amsterdam. Faculteit der Economische Wetenschappen en, *Application of the Riemannian Levenberg-Marquardt algorithm to off-line system identification* (Serie research memoranda =). Amsterdam: Vrije Universiteit, Faculteit der Economische Wetenschappen en Econometrie (in English), 1993, p. 25 leaves : ill.

[19] W. H. Kersting, "Radial distribution test feeders," *IEEE Transactions on Power Systems,* vol. 6, no. 3, pp. 975-985, 1991.

[20] J. Dantuo. IEEE 14 bus System Model .

[21] do Amaral Filho NA, de Araujo LR, Penido DR, de Alcântara Vieira F. Impacts of the representation of mutual coupling between feeders in distribution systems. International Journal of Electrical Power & Energy Systems. 2019 Feb 1;105:17-27.

APPENDIX

A.I IEEE 14 Bus System data

Line number	From bus	To bus	Line impedance (*p.u.*)		Half line charging susceptance (*p.u.*)	MVA rating
			Resistance	Reactance		
1	1	2	0.01938	0.05917	0.0264	120
2	1	5	0.05403	0.22304	0.0219	65
3	2	3	0.04699	0.19797	0.0187	36
4	2	4	0.05811	0.17632	0.0246	65
5	2	5	0.05695	0.17388	0.017	50
6	3	4	0.06701	0.17103	0.0173	65
7	4	5	0.01335	0.04211	0.0064	45
8	4	7	0	0.20912	0	55
9	4	9	0	0.55618	0	32
10	5	6	0	0.25202	0	45
11	6	11	0.09498	0.1989	0	18
12	6	12	0.12291	0.25581	0	32
13	6	13	0.06615	0.13027	0	32
14	7	8	0	0.17615	0	32
15	7	9	0	0.11001	0	32
16	9	10	0.03181	0.0845	0	32
17	9	14	0.12711	0.27038	0	32
18	10	11	0.08205	0.19207	0	12
19	12	13	0.22092	0.19988	0	12
20	13	14	0.17093	0.34802	0	12

A.II Sample code for database generation

```
clc;
clear all;
close all;
warning off
%case 1
R= 0.1;
simopt=simset('solver','ode23tb','SrcWorkspace','Current','Dstworkspace','Current');
%initilize sim options
[tout,xout,yout]=sim('DistributionLine',[0 0.2],simopt);
x1=[Va Vb Vc Ia Ib Ic Vo Io]
t1=[0 0 0 0]
[tout,xout,yout]=sim('DistributionLineAG',[0 0.2],simopt);
x2=[Va Vb Vc Ia Ib Ic Vo Io]
t2=[1 0 0 1 ]
[tout,xout,yout]=sim('DistributionLineBG',[0 0.2],simopt);
x3=[Va Vb Vc Ia Ib Ic Vo Io]
t3=[0 1 0 1]
[tout,xout,yout]=sim('DistributionLineCG',[0 0.2],simopt);
x4=[Va Vb Vc Ia Ib Ic Vo Io]
t4=[0 0 1 1]
[tout,xout,yout]=sim('DistributionLineAB',[0 0.2],simopt);
x5=[Va Vb Vc Ia Ib Ic Vo Io]
t5=[1 1 0 0]
[tout,xout,yout]=sim('DistributionLineAC',[0 0.2],simopt);
x6=[Va Vb Vc Ia Ib Ic Vo Io]
t6=[1 0 1 0]
[tout,xout,yout]=sim('DistributionLineBC',[0 0.2],simopt);
x7=[Va Vb Vc Ia Ib Ic Vo Io]
t7=[0 1 1 0]
[tout,xout,yout]=sim('DistributionLineABG',[0 0.2],simopt);
x8=[Va Vb Vc Ia Ib Ic Vo Io]
t8=[1 1 0 1]
```

```
[tout,xout,yout]=sim('DistributionLineACG',[0 0.2],simopt);
x9=[Va Vb Vc Ia Ib Ic Vo Io]
t9=[1 0 1 1]
[tout,xout,yout]=sim('DistributionLineBCG',[0 0.2],simopt);
x10=[Va Vb Vc Ia Ib Ic Vo Io]
t10=[0 1 1 1]
[tout,xout,yout]=sim('DistributionLineABC',[0 0.2],simopt);
x11=[Va Vb Vc Ia Ib Ic Vo Io]
t11=[1 1 1 0]
[tout,xout,yout]=sim('DistributionLineABCG',[0 0.2],simopt);
x12=[Va Vb Vc Ia Ib Ic Vo Io]
t12=[1 1 1 1]
xdata1=[x1;x2; x3; x4; x5; x6; x7; x8; x9; x10; x11; x12;]
tdata1=[t1;t2; t3; t4; t5; t6; t7; t8; t9; t10; t11; t12;]
```

A.III Supervised learning code: Classification

```
clc;
clear;
close all;
warning off;

% Load data from Excel file into a table
data = readtable('data_11kV');

% Split data into features (X) and target (y)
X = table2array(data(:,1:8));
y = table2array(data(:,9));

% Split data into training and test sets
cv = cvpartition(y, 'HoldOut', 0.15);
X_train = X(training(cv), :);
y_train = y(training(cv), :);
```

```
X_test = X(test(cv), :);
y_test = y(test(cv), :);

% Train a Decision Tree Classifier
dt = fitctree(X_train, y_train);

% Train a Gradient Boosting Classifier
gb = fitcensemble(X_train, y_train);

% Train a Support Vector Machine Classifier
svm = fitcecoc(X_train, y_train);

% Train a Linear Discriminant Analysis Classifier
%lda = fitcdiscr(X_train, y_train);

% Train a classification neural network using fitcecoc
nn = fitcecoc(X_train, y_train);

% Train a kNN classifier using fitcknn
knn = fitcknn(X_train, y_train, 'NumNeighbors', 5, 'Standardize', true);

% Make predictions on the test set
y_pred_dt = predict(dt, X_test);
y_pred_gb = predict(gb, X_test);
y_pred_svm = predict(svm, X_test);
%y_pred_lda = predict(lda, X_test);
y_pred_nn= predict(nn, X_test);
y_pred_knn= predict(knn, X_test);

% Compute confusion matrices
cm_dt = confusionmat(y_test, y_pred_dt);
cm_gb = confusionmat(y_test, y_pred_gb);
cm_svm = confusionmat(y_test, y_pred_svm);
%cm_lda = confusionmat(y_test, y_pred_lda);
cm_nn = confusionmat(y_test, y_pred_nn);
```

```
cm_knn = confusionmat(y_test, y_pred_knn);

% Compute accuracy scores
acc_dt = sum(diag(cm_dt)) / sum(cm_dt(:))*100;
acc_gb = sum(diag(cm_gb)) / sum(cm_gb(:))*100;
acc_svm = sum(diag(cm_svm)) / sum(cm_svm(:))*100;
%acc_lda = sum(diag(cm_lda)) / sum(cm_lda(:))*100;
acc_nn = sum(diag(cm_nn)) / sum(cm_nn(:))*100;
acc_knn = sum(diag(cm_knn)) / sum(cm_knn(:))*100;

%Plot the confusion matrix
classLabels = {'Normal', 'AG', 'BG','CG','ABG', 'ACG', 'BCG','ABCG','ABC', 'AB','AC','BC'};
figure;
confusionchart(cm_knn, classLabels);
title('K nearest neghibor');
figure;
confusionchart(cm_svm, classLabels);
title('Confusion Matrix Support Vector Machine');

% Print results
fprintf('Accuracy Percentage of Decision Tree Classifier: %.2f\n', acc_dt);
fprintf('Accuracy Percentage of Gradient Boosting Classifier: %.2f\n', acc_gb);
fprintf('Accuracy Percentage of Support Vector Machine Classifier: %.2f\n', acc_svm);
%fprintf('Accuracy Percentage of Linear Discriminant Analysis Classifier: %.2f\n', acc_lda);
fprintf('Accuracy Percentage of Neural network classifer: %.2f\n', acc_nn);
fprintf('Accuracy Percentage of K-nearest neghibour classifer : %.2f\n', acc_knn);

% Extract True Positives, False Positives, True Negatives, and False Negatives
TP = cm_svm(1,1);
FP = cm_svm(2,1);
TN = cm_svm(2,2);
FN = cm_svm(1,2);

% Compute the percentage of each
percent_TP = TP / (TP + FN) * 100;
```

```
percent_FP = FP / (FP + TN) * 100;
percent_TN = TN / (FP + TN) * 100;
percent_FN = FN / (TP + FN) * 100;

% Print results
fprintf('Percentage of True Positives: %.2f\n', percent_TP);
fprintf('Percentage of False Positives: %.2f\n', percent_FP);
fprintf('Percentage of True Negatives: %.2f\n', percent_TN);
fprintf('Percentage of False Negatives: %.2f\n', percent_FN);
```

A.IV Supervised learning code: Location

```
clc;
clear;
close all;

% Load data
load location_data.mat

% Define input and output data
inputs = location_data(:,1:8);
outputs = location_data(:,9);

% Perform least squares estimation
beta = (inputs'*inputs)\inputs'*outputs;

% Estimate the location
estimated_location_leastsquare = inputs*beta;

% Calculate the residuals
residuals = outputs - estimated_location_leastsquare;

% Calculate the mean squared error
```

```
mse = mean(residuals.^2);

% calculate the accuracy
accuracy = 1 - mse / var(outputs);

% Perform linear regression
mdl_linear = fitlm(inputs, outputs);
estimated_location_linear = predict(mdl_linear, inputs);
residuals_linear = outputs - estimated_location_linear;
mse_linear = mean(residuals_linear.^2);
accuracy_linear = 1 - mse_linear / var(outputs);

% Perform decision tree regression
mdl_tree = fitrtree(inputs, outputs);
estimated_location_tree = predict(mdl_tree, inputs);
residuals_tree = outputs - estimated_location_tree;
mse_tree = mean(residuals_tree.^2);
accuracy_tree = 1 - mse_tree / var(outputs);

% Plot the results
figure;
scatter(outputs, estimated_location_leastsquare, 'blue', 'filled');
 hold on;
scatter(outputs, estimated_location_linear, 'red', 'filled');
% scatter(outputs, estimated_location_tree, 'black', 'filled');

grid on;
xlabel('Actual Location (km) ');
ylabel('Estimated Location (km) ');
legend(sprintf('Least Square (%.2f%%)', accuracy*100));

legend(sprintf('Least Square (%.2f%%)', accuracy*100), ...
      sprintf('Linear Regression (%.2f%%)', accuracy_linear*100)), ...
%       sprintf('Decision Tree Regression (%.2f%%)', accuracy_tree*100));
grid on;
```

VITA

NAME	Nanda Kumari.
DATE OF BIRTH	01 November 1989.
PLACE OF BIRTH	Haa, Bhutan.
INSTITUTIONS ATTENDED	PSNA College of Engineering and Technology, Dindigul, Affiliated to Anna University Chennai, Tamil Nadu, India.
HOME ADDRESS	Farmgoan, Lhamoizingkha, Dagana, Bhutan.
PUBLICATION	A study on fault classification and location using supervised machine learning Fault classification & Detection: Parameter sensitivity on IEEE 14 bus system using SML
AWARD RECEIVED	Best Reviewed Paper (Easy-Chair) Award Certificate of Appreciation 6th International Conference on Intelligent Computing and Optimization, April 27-28,2023| G Hua Hin Resort & Mall, Hua Hin, Thailand