Fault Classification and Location

Identification on Electrical Transmission

Network Based on Machine Learning Methods

A thesis submitted in partial fulfillment of the requirements for the

Bachelor’s degree of

Electrical and Electronics Engineering at

Sree Vidyanikethan Engineering College

by

# Acknowledgements

# Abstract

Power transmission network is the most important link in the country’s energy system as they carry large amounts of power at high voltages from generators to substations. Modern power system is a complex network and requires high-speed, precise, and reliable protective system. Faults in power system are unavoidable and overhead transmission line faults are generally higher compare to other major components. They not only affect the reliability of the system but also cause widespread impact on the end users. Additionally, the complexity of protecting transmission line configurations increases with as the configurations get more complex. Therefore, prediction of faults (type and location) with high accuracy increases the operational stability and reliability of the power system and helps to avoid huge power failure. Furthermore, proper operation of the protective relays requires the correct determination of the fault type as quickly as possible (e.g., reclosing relays).

With advent of smart grid, digital technology is implemented allowing deployment of sensors along the transmission lines which can collect live fault data as they contain useful information which can be used for analyzing disturbances that occur in transmission lines. In this thesis, application of machine learning algorithms for fault classification on the transmission line has been explored. They have ability to “learn” from the data without explicitly programmed and can independently adapt when exposed to new data.

# Table of Contents

Abstract .......................................................................................................................................... 3

Table of Contents ............................................................................................................................ 6

List of Figures ................................................................................................................................. 8 List of Tables .................................................................................................................................. 9

Chapter 1 : Purpose and Significance of the Research ................................................................. 10

1.1 Introduction ......................................................................................................................... 10

1.2 Basics of Power System ...................................................................................................... 11

1.3 Power Transmission Networks............................................................................................ 14

1.4 Problem Statement .............................................................................................................. 17

1.5 Contributions and Thesis Outline ........................................................................................ 19

Chapter 2 : Background and Related Work .................................................................................. 21

2.1 Introduction ......................................................................................................................... 21

2.2 Review of the Faults on the Transmission Line .................................................................. 21

2.2.1 Series Faults .................................................................................................................. 22

2.2.2 Shunt Faults .................................................................................................................. 23

2.3 Survey of the methods ......................................................................................................... 24

2.3.1 Fault Classification Techniques- Machine Learning .................................................... 25

2.3.2 Fault Location Techniques – Machine Learning .......................................................... 31

2.4 Summary .......................................................................................................................... 33

Chapter 3 : Design ........................................................................................................................ 35

3.1 Introduction ......................................................................................................................... 35

3.2 Overview ............................................................................................................................. 36

Chapter 4: Implementation and Evaluation .................................................................................. 53

4.1 Experimental Setup ............................................................................................................. 53

4.2 Methods to Evaluate the model ........................................................................................... 55

4.3 No Fault Scenario ................................................................................................................ 69

Chapter 5: Conclusion................................................................................................................... 71

5.1 Summary and Findings ....................................................................................................... 71

5.2 Challenges ........................................................................................................................... 72

5.3 Future Work ........................................................................................................................ 73

Bibliography ................................................................................................................................. 74

List of Figures

Figure 1: Building Blocks of Electric Power System [3] ............................................................. 12

Figure 2: Transmission Network .................................................................................................. 15

Figure 3: Two-terminal Transmission Line .................................................................................. 16

Figure 4: Infeed Effect at Three-terminal ..................................................................................... 18

Figure 5: Outfeed Effect at the three-terminal .............................................................................. 18

Figure 6: Phase to ground fault in time ......................................................................................... 22

Figure 7: Classification of Short Circuit faults ............................................................................. 23

Figure 8: Fault detection techniques ..............................................................................................24

Figure 9 : Artificial Neural Network ….........................................................................................28

Figure 10: Two-terminal Transmission Line Details .................................................................... 53

# List of Tables

Table 1: Line details of two terminal Circuit ................................................................................ 53

Table 2: Types of Shunt Faults ......................................................................................................55

# Chapter 1 : Purpose and Significance of the Research

## 1.1 Introduction

Modern society relies heavily upon complex and widespread electric grids for critical service capabilities such as healthcare, transportation, household heating and cooling, and industrial manufacturing to name a few. As our energy delivery systems (electric and other) age, natural disasters and man-made perturbations are expected to threaten grid integrity more often. Furthermore, urban infrastructure energy delivery networks are highly reliant on the electric grid and consequently, the vulnerability of infrastructure networks to electric grid outages is becoming a major national concern. Electric power transmission is the bulk movement of [electrical energy](https://en.wikipedia.org/wiki/Electrical_energy) from a generating site, such as a [power plant,](https://en.wikipedia.org/wiki/Power_plant) to an [electrical substation.](https://en.wikipedia.org/wiki/Electrical_substation) Essentially an electrical grid is an interconnected network for delivering [electricity](https://en.wikipedia.org/wiki/Electricity) from producers to consumers. It consists of generating stations that produce electrical power, [high voltage transmission lines](https://en.wikipedia.org/wiki/High_voltage_transmission_line) that carry [power](https://en.wikipedia.org/wiki/Power_(physics)) from distant sources to demand centers and distribution lines that connect individual [customers](https://en.wikipedia.org/wiki/Customer) or businesses. Transmission lines are a vital part of the electrical distribution system, as they provide the path to transfer power between generation and load. Transmission lines operate at voltage levels from 100kV to 1000kV and are ideally tightly interconnected for reliable operation. In recent years, advanced sensors, intelligent automation, hierarchical control, communication networks, and operations technologies (OT) have been integrated into the electric grid to enhance its performance and efficiency. These new OT devices allow for large amounts of information from numerous grid systems and transmitting needed information to operations personnel in a timely manner that could not be envisioned when previous generation and transmission systems were designed and built decades ago.

In recent years, power quality has become a main concern in power system engineering – with 85-87% of power system faults occur on distribution lines [1]. However, the faults that occur on the transmission lines (the transmission grid) though fewer have a more significant and widespread impact on the consumers. The performance of a power system is affected by faults on transmission lines, which results in interruption of power flow. As the power transmission configurations (networks) become more complex quick detection of faults and accurate estimation of fault location is critical. The rapid dispatch of repair and restoration of supply voltage is essential for minimizing local and regional economic impacts, reducing overall power outages and improving customer satisfaction.

When a fault occurs in transmission line, it initiates a transition condition. Transients produce over currents in the power system, which can damage the power system depending upon its severity of occurrence. To avoid fault recurrences and the high cost associated with finding line faults, utilities endeavor for developing more accurate fault-locating methods. Transmission protection systems are designed to identify the location of faults and isolate only the faulted section of the network. The key challenge to the transmission line protection lies in reliably detecting and isolating faults compromising the security of the system – with significant accuracy. With the advent of OT devices, new measurement devices like phasor measurement unit (PMU), Digital Fault Recorders (DFR) are often used to provide detailed information on the health the grid. These OT advances in power system has led to massive volumes of data from the continuous monitoring of transmission lines. The massive volumes of data is both a blessing and curse- large amounts of data easily can overwhelm storage facilities, but with the advent of machine learning algorithms this opens potential to implement smart and robust fault location algorithms [2].

Section 1.2 discusses the fundamental terms and concepts used in today’s electric power system. The basics and types of transmission network is presented in section 1.3. Section 1.4 describes the issue addressed in the thesis and section 1.5 lays the outline of the thesis.

## 1.2 Basics of Power System

Electric power systems are real-time energy delivery systems. Real-time meaning power is generated, transported, and supplied the moment light switch is turned on. Electric power systems are not storage systems like water systems and gas systems. Instead, generators produce the energy as the demand calls for it. Figure 1 shows the basic building blocks of an electric power system. Starting with generation, where electrical energy is produced in the power plant and then transformed in the power station to high-voltage electrical energy that is more suitable for efficient long-distance transportation. The power plants transform other sources of energy as well in the process of producing electrical energy. For example, heat, mechanical, hydraulic, chemical, solar, wind, geothermal, nuclear, and other energy sources are used in the production of electrical energy. High-voltage (HV) power lines in the transmission portion of the electric power system efficiently transport electrical energy long distances to the consumption locations. Finally, the remote substations are responsible for transforming this HV electrical energy for delivery on lower high voltage power lines called “Feeders” that are more suitable for the distribution of electrical energy. This electrical energy is again transformed to even lower voltage services for residential, commercial, and industrial consumption.

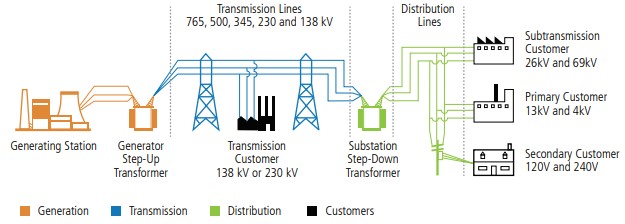


Figure 1: Building Blocks of Electric Power System [3]

The Power Generation and Distributions has four stages:

1. *Generation*: Power generation plants produce the electrical energy that is ultimately delivered to consumers through transmission lines, substations, and distribution lines. Electrical energy must be generated at the same rate at which it is consumed. A sophisticated control system is required to ensure that the [power generation](https://en.wikipedia.org/wiki/Power_generation) very closely matches the demand.
2. *Transmission*: Transmission lines are necessary to carry high-voltage electricity over long distances and connect electricity generators with electricity consumers. Transmission-level voltages are typically at or above 110,000 volts or 110 kV, with some transmission lines carrying voltages as high as 765 kV[3]. Power generators, however, produce electricity at low voltages and the generation voltage is stepped-up to transmission voltages. To make high-voltage electricity transport possible, the electricity must first be converted to higher voltages with a step-up transformer.
3. *Distribution*: Distribution systems are responsible for delivering electrical energy from the distribution substation. Most distribution systems in the United States operate at primary voltages between 12.5 kV to 34.5 kV and some operate at lower distribution voltages such as 4kV. These low-voltage distribution systems are being phased out because of their relatively excessive cost for losses (low voltage requires high currents, which means high losses). These networks carry the power to consumer units like businesses or residential entities.
4. *Load*: This stage accounts for electrical energy used by various loads on the power system. Electricity is consumed and measured several ways depending on whether the load is residential, commercial, or industrial and whether the load is resistive, inductive, and capacitive.

The electrical network's or the grid's ability to supply a clean and stable power supply is very critical on day-to-day. High power quality ideally creates a perfect power supply that is always available, has a pure noise-free, sinusoidal wave shape and is always within voltage and frequency tolerances. A well-functioning power transmission network enables:

1. *Economies of scale*: The behavior of the electricity sector is directly related to economic factors such as Gross Domestic Product (GDP). In this manner, the demand for electricity be a “thermometer” of the market. As such, growth of the economy as well as increases in purchasing power and quality of life must be accompanied by improvements in the power system, with the objective being compliance with current and future situations.
2. *Rural electrification*: Extending electrical grids into countryside will not only help cater to residential houses for lighting and household purposes but also allows for [mechanization of many farming operations](https://en.wikipedia.org/wiki/Mechanised_agriculture) especially in areas facing labor shortages.
3. *Increased transmission reliability*: Reliability refers to the extent to which customers have a continuous supply of electricity. As electricity cannot be easily stored, a reliable supply of electricity requires generators to produce electricity and the transmission and distribution networks to transport the electricity to customers in real time. Therefore, a good transmission system will ensure affordable, high-quality electric service is essential for modern life.
4. *Decreased costs*: Transmission network carries the high-voltage power from the generating sites to the distribution stations. The development and improvement of algorithms that allow the analysis and diagnosis of failures in transmission lines can have an important economic impact, for power utilities by reducing operation costs, as they enable the continuity and reliability of the electric sector.
5. *Increases potential for power pools, markets and bulk power transactions:*  A reliable transmission network will enable more advanced methods of power transfers like power pool, bulk power transfers etc. It primarily helps to balance electrical load over large network than a single utility by providing mechanism for interchange of power between two or more utilities.

Section 1.3 describes more about the transmission network and its configurations used by utilities.

## 1.3 Power Transmission Networks

The United States’ bulk electric system consists of more than 360,000 miles of transmission lines, including approximately 180,000 miles of high voltage lines, connecting to about 7,000 power plants [4]. High-voltage (up to 765 KV) transmission lines transport power long distances much more efficiently than lower voltage (12 - 34.5 KV) distribution lines for two main reasons. First, high-voltage power transmission allows for lesser resistive losses in transit which is about 6% on average in the United States [5] . This efficiency of high voltage transmission allows for the transmission of a larger proportion of the generated power to the substations and in turn to the loads, translating to operational cost savings. Second, raising the voltage to lower the current allows one to use smaller conductor sizes, or have more conductor capacity available for growth. Transmission line systems relay the power from production sites to the users. Failure of these structures can lead to power cuts and therefore disrupt the day to day life of people as well as the industries dependent on electricity.

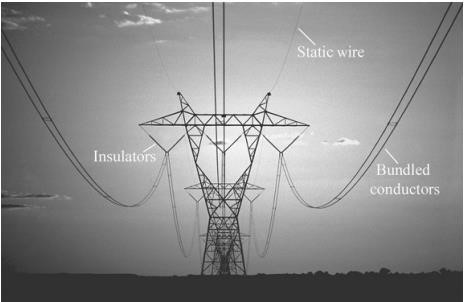


Figure 2: Transmission Network

A transmission grid is a network of power stations, transmission lines, and substations. Energy is usually transmitted within a grid with [three-phase](https://en.wikipedia.org/wiki/Three-phase_electric_power) [AC.](https://en.wikipedia.org/wiki/Alternating_current) Transmission lines are either overhead power lines or underground power cables. Overhead cables are not insulated and are vulnerable to the weather but can be less expensive to install than underground power cables. Overhead and underground transmission lines are made of aluminum alloy and reinforced with steel; underground lines are typically insulated. Figure 2 shows a three-phase 500 kV transmission line with two conductors per phase. The two conductors per phase option is called bundling. Multiple conductors are bundled together per phase to double, triple, or greater to increase the power transport capability of a power line, lower losses and improve other operating characteristics of the line such as electromagnetic fields and audible noise.

Typically, there are three types of line configurations used in the transmission network. These line configurations include (a) radial (one-terminal), (b) two-terminal, and (c) multi-terminal of which three-terminal is possibly the most prominent multi-terminal type. It should be noted that "terminals" in this context, refers to source terminals and not-tapped transformer terminals or stations. The two-terminal line configuration is the most dominant type followed by radial, and the three-terminal lines are the exceptions.

Two-terminal line systems are used for bulk power transfer and to supply loads from two power sources- Terminals A, B and are very common. Figure 4 shows the two-terminal transmission line. To obtain proper selectivity and coordination, directional distance relays [6] for phase and ground fault detection are used normally. Directional ground overcurrent relaying is sometimes applied in addition to, or in place of, directional ground distance relay functions. One or two communications-based protection groups are normally used with two-terminal line applications at the transmission voltages greater than 200KV.



Figure 3: Two-terminal Transmission Line

## 1.4 Problem Statement

Transmission lines or transmission network is a crucial part of the electric grid as it carries high voltage power from generating site to the substations where the voltage stepped-down for end-use consumption transported via distribution lines. Though the frequency of faults is much higher in distribution lines, faults on transmission lines have more widespread impact and faults in buried transmission lines take longer to locate and repair. Additionally, since the transmission lines carry high voltages, faults on these lines might lead to unsafe conditions. Therefore, safeguarding against exposed fault is the most critical task in the protection of power system. The protection schemes or mechanisms for the transmission lines become challenging as configurations of the transmission lines become increasingly complex.

Three-terminal and other multi-terminal line construction are generally a trade-off of planning economics and protection complexities. Two-terminal lines with long tap(s) supplying remote load from the main line may display many of the same protection and load ability issues as threeterminal lines. The complexity of protecting these line configurations increases from the relatively simple radial, to the more difficult two-terminal, and to the still more difficult three-terminal. Relaying three-terminal lines has been and continues to be a challenge for protection engineers

[7].

Primary and biggest challenge with protecting three terminal circuits is “Infeed”. During a fault on the transmission line, distance relay measures impedance which is equal to the positive sequence (A balanced three-phase system with the same phase sequence as the original sequence), if there are no sources of fault current on the transmission line between the line terminal where the relay is located and the fault.

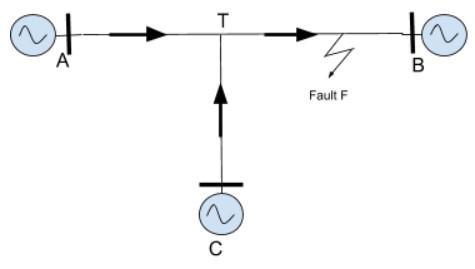


Figure 4: Infeed Effect at Three-terminal

From the Figure 6, the actual line impedance from the relay terminal (Terminal A) to the fault is not always the impedance measured by the relay. This is because the third line terminal (Terminal C) tapped (Tee point) to a line is an additional source of current for a line fault. Current will be supplied to a fault that occurs on the line section beyond the tap of Terminal C through both Terminal A and Terminal C. The voltage drops resulting from the input of fault current from each of these sources into the common section of the line will be measured by the distance relay at the Terminal A. Since the current input from Terminal C is not applied to the relay at Terminal A, the impedance measured by this relay is higher than the actual impedance from the Terminal A to the fault. The relay will under reach; that is, for a given relay setting the relay does not cover the same length of line it would if the additional current source were not present. Due to infeed, most of the impedance based and traveling wave-based methods are not successful in identifying faults and often give erroneous results [8].

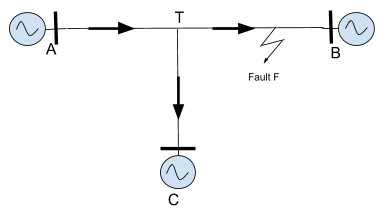


Figure 5: Outfeed Effect at the three-terminal

It is also possible to experience an “Outfeed” at the T location, in which case there will be tendency to overreach as shown in

Figure 7. This phenomenon is not too common but can cause delayed or sequential tripping at the terminals.

Thirdly, transmission lines could traverse long distances, in which case the line A-B ends up being in one region and line C-T in different region with separate set of environmental conditions which directly influences the impedance of the respective lines. This causes line non- homogeneity and since impedance on line C-T is different that of line A-B which makes fault location on line C-T trickier. Therefore, there is a need of an adaptable, resilient method for fault classification and location on transmission lines which could learn from system behavior and detect unknown faults rather than hardcoded methods (algorithms)which follow specific set of rules.

Taken all together, faults on transmission lines and the varying environmental conditions present a complex classification and detection problem. With the advent of new machine learning methods and supervised learning methods, these challenges may be more effectively addressed. Machine learning methods are based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. The ability to automatically apply complex mathematical calculations to big data – over and over, faster and faster give these algorithms potential to identify insights in the data which would be otherwise an impossible task for humans. The availability of high-resolution/high-volume data, due to the proliferation of intelligent electronic devices in smart grids, paves ground to implement more accurate and intelligent machine learning methods for fault classification and location identification on the transmission lines.

## 1.5 Contributions and Thesis Outline

Majority of the faults on the transmission lines are shunt faults and with around 5% of them being symmetric (all three phases are equally affected). This research presents supervised-learning method for classification and location of shunt faults on three-terminal and two terminal transmission lines. The main contributions of the research are:

* Two different architectures are proposed which adapts to any N-terminal in the transmission line (dimensional scaling).
* The models proposed do not require large dataset or high sampling frequency. Additionally, they can be trained quickly and generalize well to the problem.
* The first architecture is based off decision trees for its simplicity, easy visualization which have not been used earlier. In this instance, fault location method uses traveling wave-based approach for location of faults. The method is tested with performance better than expected accuracy and fault location error is less than ±1%.
* The second architecture uses single Support Vector Machine to classify ten types of shunt faults and Regression model for fault location which eliminates manual work. The architecture was tested on real data and has proven to be better than first architecture. The regression model has fault location error less than ±1% for both three and two terminals.
* Both the architectures are tested on real fault data which gives a substantial evidence of its application.

The thesis is organized into 5 chapters. Chapter 2 discusses relevant literature and presents a survey of different machine learning methods proposed over past years. The two new architectures for fault classification and location is proposed in Chapter 3. Chapter 4 details the experimental setup, implementation and results are presented. Chapter 5 summarizes the conclusions and proposes the possibilities of future work.

# Chapter 2 : Background and Related Work

## 2.1 Introduction

Given the electrical power grid is a complex power system consisting of power generating stations, high voltage transmission lines and distribution lines, fault classification and location identification is necessary to improve protection mechanisms and have reliable, high-speed protection devices. Most often, electrical faults result in mechanical or material damage to the lines or structures, which must be repaired before returning the line to service. As it is noted earlier, repair and restoration is extremely important for maintaining critical and societal services. The restoration process is hampered if the location of the fault cannot be estimated with accuracy or confidence (less than a mile). Various methods have been proposed over the years, and each method have their own merits and disadvantages.

Section 2.2 presents an overview on several types of electric faults occurring on transmission lines. In Section 2.3, to encapsulate the current state of art methods, a survey review is presented on popular machine learning algorithms used for fault classification and location on transmission lines and summaries are given in Section 2.4.

## 2.2 Review of the Faults on the Transmission Line

In an [electric power system,](https://en.wikipedia.org/wiki/Electric_power_system) a fault or fault current is any abnormal [electric current.](https://en.wikipedia.org/wiki/Electric_current) For example, a [short circuit](https://en.wikipedia.org/wiki/Short_circuit) is a fault in which current bypasses the normal load and an open-circuit fault occur if a circuit is interrupted by some failure. Transmission line carry 3-phase AC [9]. Under ideal state, all phase voltages have same maximum value but differ in phase from each other at angle 120 degrees. Shunt faults are caused by short circuit between lines. For example, a line to ground fault occurs when one conductor drops to the ground or comes in contact with the neutral conductor (ground). This causes a rapid decrease in respective phase voltage of the line involved in the fault and increase in phase current. Figure 8 shows the phase C-to-ground fault in time (µs). 𝑉𝑎, 𝑉𝑏, 𝑉𝑐 are the phase voltages of phase a, b, c and 𝐼𝑎, 𝐼𝑏, 𝐼𝑐 are the phase currents of phase a, b, c respectively. From the figure we can see that, magnitude of phase C voltage decreases and phase C current increases during the fault.

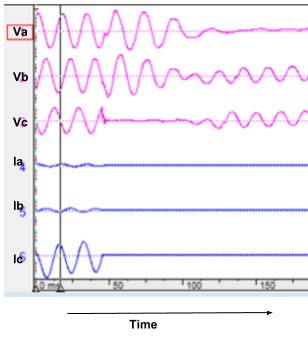


Figure 6: Phase to ground fault in time

Faults can be categorized as the shunt faults and series faults [10] described below:

#### 2.2.1 Series Faults

Series faults represent open conductor and take place when unbalanced series impedance conditions of the lines are present. These faults disturb the symmetry in one or two phases and are therefore unbalanced faults. Two examples of series fault are when the system holds one or two broken lines, or impedance inserted in one or two lines. In the real world a series faults takes place, for example, when circuit breakers control the lines and do not open all three phases, in this case, one or two phases of the line may be open while the other/s is closed [10]. Series faults are characterized by increase of voltage and frequency and fall in current in the faulted phases.

#### 2.2.2 Shunt Faults

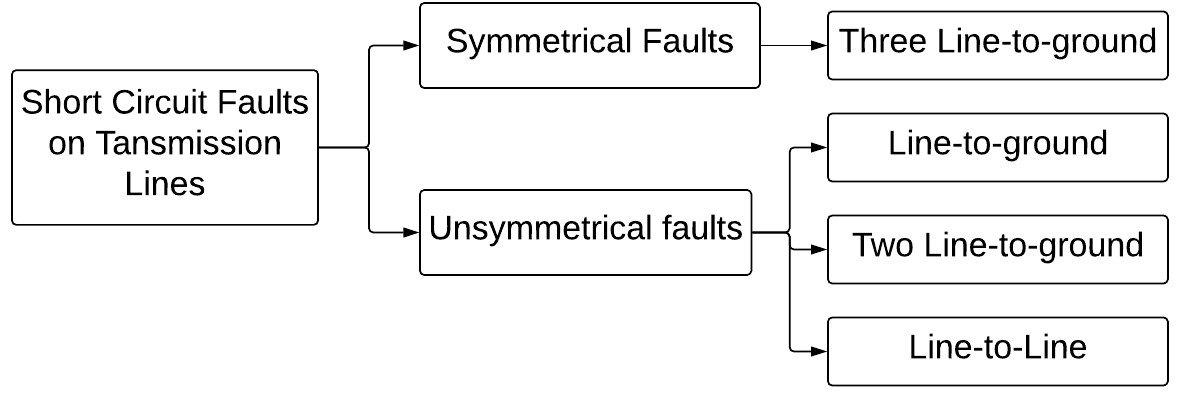


Figure 7:Classification of Short Circuit faults

There are two types of short circuit faults or shunt faults which can occur on transmission lines; balanced faults and unbalanced faults also known as symmetrical and unsymmetrical faults respectively as shown in Figure 9. In symmetrical faults, also called three phase short circuits, all the three phases are short circuited to each other and often to earth also. Such faults are balanced and symmetrical as the system remains balanced even after the occurrence of the fault. Though the symmetrical faults are rare, they generally lead to the most severe fault current flow. Most faults that occur in a power system are unsymmetrical faults involving only one or two phases. The most common type of unsymmetrical fault is a short circuit between a phase and the earth.

The shunt faults are the most common type of fault taking place in the field. They involve power conductors or conductor-to-ground or short circuits between conductors. One of the most important characteristics of shunt faults is the increment the current suffers and fall in voltage and increase frequency. Shunt faults can be classified into four categories [11].

1. *Line-to-ground fault*: This type of fault exists when one phase of any transmission lines establishes a connection with the ground either by ice, wind, falling tree or any other incident. About 70% of all transmission lines faults are classified under this category [12].

1. *Line-to-line fault*: Because of high winds, one phase could touch anther phase & line-to-line fault takes place. Approximately 15% of all transmission lines faults are considered line-to-line faults [12].
2. *Double line-to-ground*: Falling tree where two phases become in contact with the ground could lead to this type of fault. Two phases will be involved instead of one

in the line-to-ground faults scenarios. Ten percent of all transmission lines faults are under this type of faults [12].

1. *Three phase faults*: In this case, falling tower, failure of equipment or even a line breaking and touching the remaining phases can cause three phase faults. In reality, this type of fault not often exists which can be seen from its share of 5% of all transmission lines faults [12]. The first three of these faults are known as asymmetrical faults.

## 2.3 Survey of the methods

This section presents a survey on different fault classification and location identification techniques in transmission lines highlighting the implementations machine learning methods in the past. In this review, only short circuit faults are considered as they are more common.

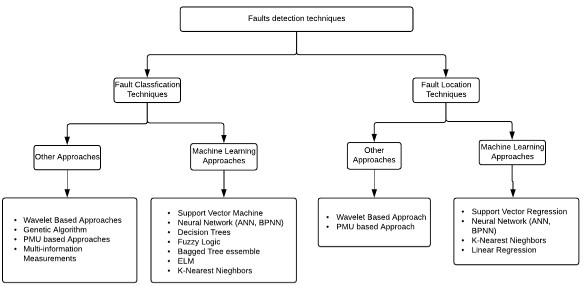


Figure 8: Fault detection techniques

The survey is mainly divided into two parts:

1. Fault classification techniques - Methods that determine the fault type
2. Fault Location Techniques - Methods that calculate the distance of the fault

Both techniques play a vital role in development of protection mechanisms for a given power system model. There have been various approaches used to develop a fast speed and reliable method to deal with faults as shown in Figure 10.

Wavelet based approaches primarily use the time difference between the traveling wave reflections which assume higher sampling rate and synchronized measurements at the terminals for fault identification making it difficult for a practical application, especially for three-terminal circuits due to infeed problem. Though this method has lower error of estimation, it has a higher computational burden [13]. Phasor measurement unit (PMU) based approaches require synchronized phasor quantities from all the terminals of the transmission lines. Genetic Algorithms are the heuristic search and optimization techniques that mimic the process of natural evolution. When applied to classify and location faults on transmission lines, they are often slow and complex to be implemented [14].

Machine learning is a subset of [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence) in the field of [computer science](https://en.wikipedia.org/wiki/Computer_science) that often uses statistical techniques to give [computers](https://en.wikipedia.org/wiki/Computer) the ability to "learn" (i.e., progressively improve performance on a specific task) with [data,](https://en.wikipedia.org/wiki/Data) without being explicitly programmed [15]. Machine learning algorithms can learn and improve themselves by studying high volumes of available data. They are very helpful in fields where traditional programming rules do-not operate or rules keep evolving. Since the faults occurring on power grid are very unlikely to be similar and power system can change depending on the future demand, use of machine learning algorithms for solving such problems might. They can benefit from learning correlation between events and can give insights helping human uncover factors causing the faults and to find the solution of complex multiobjective nonlinear systems, the above-said methods are used to get faster solution and less error.

#### 2.3.1 Fault Classification Techniques- Machine Learning

Classification of power system faults is the first stage for improving power quality and ensuring the system protection. For this purpose, a robust classifier is necessary.

Most prominent machine learning approaches for fault classification are explained below:

##### B. Neural Network

Artificial Neural Networks are computing systems vaguely inspired by the [biological neural networks](https://en.wikipedia.org/wiki/Biological_neural_network) that constitute animal brains. An ANN is based on a collection of connected units or nodes called [artificial neurons](https://en.wikipedia.org/wiki/Artificial_neuron) which loosely model the [neurons](https://en.wikipedia.org/wiki/Neuron) in a biological [brain.](https://en.wikipedia.org/wiki/Brain) Each connection, like the [synapses](https://en.wikipedia.org/wiki/Synapse) in a biological [brain,](https://en.wikipedia.org/wiki/Brain) can transmit a signal from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it. The original goal of the ANN approach was to solve problems in the same way that a [human brain](https://en.wikipedia.org/wiki/Human_brain) would. Figure 11 shows an artificial neural network which is an interconnected group of nodes, akin to the vast network of [neurons](https://en.wikipedia.org/wiki/Neuron) in a [brain.](https://en.wikipedia.org/wiki/Brain) Here, each circular node represents an [artificial neuron](https://en.wikipedia.org/wiki/Artificial_neuron) and an arrow represents a connection from the output of one artificial neuron to the input of another. However, over time, attention moved to performing specific tasks, leading to deviations from [biology.](https://en.wikipedia.org/wiki/Biology) Many researches have used neural network for fault classification and detection explained in [27].

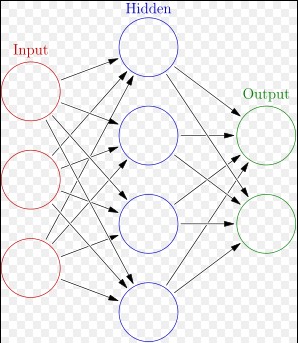


Figure 9: Artificial Neural Network [28]

Koley et. al. [29] proposed a hybrid wavelet transform and modular artificial neural network based fault detector, classifier and locator for six phase lines using single end (single terminal) data. The standard deviation of the approximate coefficients of voltage and current signals obtained using discrete wavelet transform are applied as input to the modular artificial neural network for fault classification and location.

Ahmad et. al. [30] proposed wavelet based artificial neural networks for fault classification. Discrete wavelet transforms (DWT) is used to extract high-frequency components of the aerial modal currents. A feature vector is built using the wavelets details coefficients of one level of the aerial modes and is used to train an ANN. The proposed method is tested on the simulated data with acceptable accuracies.

Rao et. al. [31] has presented a fault classification and detection method using discrete wavelet transform and artificial neural networks. Discrete wavelet Transform (DWT) is applied to the fault phase currents to obtain energy values which is used as input to train the neural network. The proposed method was tested on a simulated network values using MATLAB.

Few other researchers have used Back-Propagation Neural Network (BPNN) for identifying and classifying faults on transmission lines. Backpropagation is a method used in [artificial neural networks](https://en.wikipedia.org/wiki/Artificial_neural_network) to calculate a [gradient](https://en.wikipedia.org/wiki/Gradient) that is needed in the calculation of the [weights](https://en.wikipedia.org/wiki/Artificial_neural_network#Components_of_an_artificial_neural_network) to be used in the network. It is commonly used to train [deep neural networks,](https://en.wikipedia.org/wiki/Deep_neural_network) a term referring to neural networks with more than one hidden layer.

Saini et. al. [32] has proposed new algorithm for fault detection and classification on parallel transmission lines using Discrete Wavelet Transform (DWT) and Back-Propagation Neural Network (BPNN). Wavelet energies coefficients of alpha and beta mode currents obtained by clark’s transformation are used as input to train BPNN with two hidden layers. The proposed method is tested on different networks and fault scenarios.

##### D. Other Techniques

The other machine learning techniques used in fault classification and identification.

Jamehbozorg et. al. [36] proposed Decision Tree based method for fault classification in Double-Circuit Transmission Lines. The proposed method needs voltages and currents of only one terminal of the protected line. After detecting the exact time of fault inception and calculating the odd harmonics of the measured signals, up to the nineteenth, a decision tree algorithm is employed for recognition of the intercarrier fault type. Also, the proposed method is extended for classification of crossover faults in these transmission lines.

Mishra et. al [37] have used bagged tree ensemble technique. Bagging stands for bootstrap aggregation; whereby random samples are drawn through replacing the training datasets. Bagging is a simple method that can be employed to reduce the variance for those machine learning techniques with high variance. Post fault current is decomposed by Fast Discrete Orthonormal STransform (FDOST) and bagged tree ensemble technique is used to classify faults. The proposed method is then tested under different fault scenarios on the simulated data.

K-Nearest Neighbor (k-NN) based classification method was proposed by Majd et. al. [38]. Distances of each sample and its fifth nearest neighbor in pre-determined in the default window which determines the fault occurrence time and phase. Therefore, k-NN is applied to the instantaneous values of normalized three phase currents.

Ray et. al. [39] proposed an Extreme Learning Machine (ELM) based fault classification technique in a series compensated transmission line. Extreme learning machines are [feedforward neural networks](https://en.wikipedia.org/wiki/Feedforward_neural_network) for [classification,](https://en.wikipedia.org/wiki/Statistical_classification) regression and [feature learning](https://en.wikipedia.org/wiki/Feature_learning) with a single layer or multiple layers of hidden nodes, where the parameters of hidden nodes (not just the weights connecting inputs to hidden nodes) need not be tuned. Discrete Wavelet Transform (DWT) is used to decompose the instantaneous current signals used as input to ELM for fault classification. The proposed method is tested with simulated data.

Dasgupta et. al. [40] proposed a method for detecting and classifying transmission line faults using cross-correlation and k-Nearest Neighbor (k-NN). This method computes the cross correlation between pure and faulty current signals. Extracted features are used as input the k-NN algorithm which then computes distance of a given sample to all other samples in the set and class of the sample with least distance is predicted.

Linear Regression Index-Based Method for fault classification and detection is presented by Musa et. al. [41]. The proposed algorithm has constructed a rule as follows: when the system is running under healthy condition, the Linear Regression Coefficient Indices (LRICs) will be equal to zero; when the system is subjected to the fault condition, the LRICs of faulted phases will be greater than zero. For each possible scenario of faults, the proposed algorithm required only the three-phase current measurement of the local measurement.

#### 2.3.2 Fault Location Techniques – Machine Learning

Accurate fault location that occur on transmission lines is highly important from aspect of quick identification of weak points on the transmission line and taking respective counter measures to decrease the probability of those faults. Various approaches have been developed over the time to address these issues and in those few are hard coded. On the other hand, power of algorithms which can learn from real world pattern are explored. Figure 10 shows few fault location techniques that have been developed by the researches in the past. In this section, machine learning techniques employed in past have been summarized.

##### A. Neural Networks

Neural Networks becomes first choice when the goal is to model non-linear and complex relationships which follow real-life pattern. They generalize well which results in a model that predicts well on unseen data.

Twafik et. al. [42] proposed Artificial Neural Network (ANN) for estimating fault location on transmission lines. Prony method is used to extract the modal information from voltage or current signal. ANN are then used to estimate the fault distance based on the modal information. The model is trained and tested using the simulated data.

Fathabadi et. al. [43] proposed an hybrid framework consisting of a proposed two stage Finite Impulse Response (FIR) filter, four Support Vector Machines (SVMs), and eleven Support Vector Regressions (SVRs). The proposed two-stage FIR filter together with the SVMs are used to detect and classify short-circuit faults while the SVRs are utilized to locate short-circuit faults and predict distances.

Yadav et. al. [44] have written a comprehensive and exhaustive survey will reduce the difficulty of new researchers to evaluate different ANN based techniques with a set of references of all concerned contributions. From the survey they concluded that ANN is found to be robust, accurate, and efficient approach for transmission line fault detection, classification, localization, direction discrimination, and faulty phase selection.

##### B. Support Vector Regression (SVR)

The Support Vector Regression (SVR) uses the same principles as the SVM for classification. Because output is a real number it becomes very difficult to predict the information at hand, which has infinite possibilities. In the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM which would have already requested from the problem.

Ray et. al. [45] proposed Support Vector Machine (SVM) for fault classification and Support Vector Regression (SVR) for fault location. Fault classification consists of four SVMs each predicting the involvement of each phase and fault location consisting of SVR. Both the models have been trained by best features out of decomposed post fault current signals using Wavelet packet transform (WPT). The proposed method has been trained and tested on the simulated data and a comparison study has been carried out with methods published by other researchers.

Hosseini et. al. [46] presented hybrid method for fault classification and location. Post fault voltage samples are decomposed by discrete wavelet transform which are used with post fault current samples as input to SVM for the fault classification model. Four SVMs are used to predict the fault in each phase. Depending on the fault type in the first stage, the second stage one out of four SVRs is selected for fault location.

##### C. Other Techniques

Farshad et. al. [47] proposed a method to classify and locate single-to-ground faults using k-Nearest Neighbors (k-NN) algorithm. Various features are extracted from the voltage signals measure from a single terminal. Decomposed signals from the discrete Fourier Transform is used as input to k-NN for fault type classification. k-NN in regression mode is used for fault location. Since, current signals are not used, the proposed approach is immune against current-transformer saturation and its related errors.

Ray et. al. [48] in proposed fault location technique using extreme learning machine (ELM) in series compensated transmission line. The proposed method uses one cycle of post fault current and voltage signals which are decomposed by wavelet transform. Best features, selected by genetic algorithm are then used as input to the ELM. The method has been tested on variety of simulated data.

### 2.4 Summary

Variety of approaches have been used to increase reliability and robustness of fault classification and location methods.

Neural Network follows a black box model making the explanation for the result typically difficult to understand. Neural network-based fault classification methods show good accuracy, however the training time is quite large due to which the task becomes more complex. The artificial neural network techniques suffer from the requirement of large training data. However, these methods are hard to implement practically. If the fault can’t be identified quickly, it will produce many ill-effects such as line outages during the period of peak load leading to severe economic losses. There may be a chance for the entire grid to collapse which is called as blackout and the reliability of the system would be affected.

Few of the fault classification and location methods use fuzzy logic based architectures [13]. Fuzzy logic uses rule-based relationship for making decisions. Though they have lot less computation burden, it is tedious to develop fuzzy rules and membership functions and fuzzy outputs can be interpreted in many ways making analysis difficult. In addition, it requires lot of data and expertise to develop a fuzzy system.

Other impedance measurement-based methods for fault location depend on fundamental concept of calculating line impedances pre- and post-faults to determine the distance of the fault. However, in three terminal circuits, due to infeed, the impedance values are measured are much larger than actual line impedance which gives rise to erroneous results. Secondly, the lines A-B and C-T may be in different terrains which results in different environmental conditions. In this scenario, synchronized impedance measurements might provide erroneous results.

It appears, Current state of the art Machine Learning methods presented in above sections have tested models on simulated data which have same distribution, pattern/trend as training data. Therefore, the robustness of the methods is not completely known and the question of whether these methods are applicable to 3 terminal networks is yet to be answered.

# Chapter 3 : Design

## 3.1 Introduction

In the recent past, many researchers have proposed approaches for fault classification and location identification. However, they were not applicable to a variety of transmission network configurations, particularity 3 terminal networks and were not evaluated with real data to ensure the effectiveness of the methods in locating and classifying faults on the transmission line.

Based on the literature review and preliminary designs, the design criteria that emerged is as follows:

1. *Scalability*. Transmission network expand as the demand for the power increases every year. Therefore, fault location methods should be scalable.
2. *Confidence*. Should produce estimates that are timely, sound and reliable, otherwise the confidence in the methods would be weakened and longer lengths of transmission line would have to be examined to find the exact fault location.
3. *Network* Topology. Adaptive to different configurations, applicable to changing fault data.
4. *Relevant*. Make use of existing power line health data
5. *Extensible*. Interface with existing OT power line monitoring devices.

Based on recent advances in machine learning, it was decided to explore the utility and applicability of machine learning to fault classification and location on transmission lines. Machine learning is a form of data analysis that automates analytical model building [49]. Using algorithms that continuously assess and learn from data, machine learning algorithms enables hidden insights into complex behavior and relationships. It can handle multi-dimensional and multi-variable data in dynamic environments. A key aspect to machine learning algorithms is that they learn the behavior of the system representative and synthetic datasets to produce reliable, repeatable decisions and results. For fault classification and location on transmission lines, these attributes are desirable. In addition, machine learning methods do not require synchronized measurements at the terminals. Additionally, they can be employed in real time to monitor the grid as well.

Input data is very crucial aspect for the machine learning algorithms and the correctness of prediction is based on the quality of data that the model was trained with. Post- fault voltage or current transients from the live grid are prone to have small amount of noise (may be negligible). Therefore, to ensure the quality of the data used to train the machine learning models and to extract required features from the post-fault signal, Discrete Wavelet Transform (DWT) is used. The advantage of DWT lies in determination of the key components in the signal like energy and entropy. The training data which consists of these components is used to train the predictive models so that they can learn from the information given by these components.

In predictive modeling, the idea is to create a function which is isomorphic to original function/process which was used to generate the training data. Therefore, this predictive model can predict new data points using the “new” function. The fault type and faulty line identification on the transmission line is a classification problem – in which we want to build a classification model to classify ten fault types into the target classes. Based on the preliminary research, we focused on two machine learning based architectures. The first architecture employs decision trees and second architecture use multi-class Support Vector Machines (SVM) for fault classification and faulty line identification. Both SVM and Decision trees follow white box ( [subsystem](https://en.wikipedia.org/wiki/Subsystem) whose internals can be viewed but usually not altered) approach and are capable of handling continuous data (floating point).

Section 3.2 gives the details about the overall process for fault identification in the transmission lines.

#### 3.4.2 Fault Classification

The goal is to predict ten types of Short Circuit fault which occur on the transmission line. Figure 18 shows the process of the fault classification for two and three terminal circuits. The idea is to use divide and conquer approach on the system and collectively interpret the result.

The simulated data has phase voltage wavelet energies at each Terminal as features and four labels (A, B, C and G) each taking value of either +1 or -1. The value of +1 denotes that the phase is involved in the fault that occurred. For example, for a phase A to Ground fault, the values of A, B, C and G would be +1, -1, -1 and +1 respectively.

Simulated data set described earlier is used as input to train the decision trees and tested on the real dataset. Each of four decisions have two target classes (binary). Decision Trees corresponding to each phase are used to predict the involvement of each phase in the fault. For example, DT-Phase A is a decision tree for phase A predicts +1 if phase A is involved in the fault and -1 if it’s not. The result is interpreted as combination of prediction of 1 bit from 4 Trees. The Target space for the fault classification is ~212.

.

.

# Chapter 4: Implementation and Evaluation

## 4.1 Experimental Setup

The three terminals and two terminal networks from Dominion Energy, Virginia, USA Transmission network in North Eastern region was used to experiment the proposed architecture.

The testbed contains two-terminal and three-terminal 230KV transmission lines shown below in Figure 26 and Figure 27 with their configuration details shown tabulated in Table 1 and Table 2 respectively.



Figure 10: Two-terminal Transmission Line Details

A-B details:

Length- 24.64 mi

R=0.003

X = 0.0338

R0 = 0.0214

X0 = 0.0887

Table 1: Line details of two terminal Circuit

An evaluation framework was created used to deploy the two architectures. The both transmission models were emulated in Aspen One Liner [58] to collect the simulated data. *Aspen OneLiner* is a PC-based short circuit and relay coordination program. It is a productivity tool which is useful for simulations. In aspen one-liner, any type of fault could be simulated under different conditions and fault data for relay testing can be exported. All 10 types of shunt faults as mentioned in *Table 3* were simulated at every 1%, 2%,3%....99% of each line and instantaneous post-fault phase voltages from each terminal was collected over one cycle. Aspen OneLiner Script was written to simulate and collect the data which takes about 60 seconds.

|  |  |
| --- | --- |
| Types of Shunt faults | Fault Type |
|  | AG (phase A to Ground) |
| Line to ground faults | BG (phase B to Ground) |
| CG (phase C to Ground) |
| Double Line to ground  faults | ABG (phase A-phase B to Ground) |
| BCG (phase B-phase C to Ground) |
| ACG (phase A- phase C to Ground) |
| Three Line to ground faults | ABCG (phase A- phase B-phase C to Ground) |
| Line to Line faults | AB (phase A to phase B) |
| BC (phase B to phase C) |
| AC (phase A to phase C) |

Table 2 : Types of Shunt Faults

Python was used in the framework for data preparing which consists of computing wavelet energies for the transient post fault phase voltages collected, normalization and then assigning target classes to each sample. The simulated data was used to train and validate the models. Real fault data were collected from Digital Fault Recorders placed at the terminal of the Dominion Energy’s transmission line under study. The test set consists of real fault data which is [independent](https://en.wikipedia.org/wiki/Independence_(probability_theory)) of the training dataset (simulated) and used only to assess the performance (i.e. generalization) of a fully specified classifier. For three terminal line configuration, the test set consists of 2 samples (two faults) whereas two terminal test set consists of one sample (one fault).

Classifiers were built with scikit-learn: machine learning library in Python [60]. Training and testing time for the classifiers were within few microseconds/milliseconds. The experiment was carried on a machine with 8GB RAM, intel i5 processor and 1TB memory.

## 4.2 Methods to Evaluate the model

This section briefly describes various metrics used in evaluating the machine learning models. From the training dataset presented to the algorithm, it tries to tweak its internal parameters to better understand the data. If the model is over trained meaning the model tries to mimic the exact behavior of the training data, then it will be able to identify all the relevant information in the training data but will fail miserably when presented with the new data which is independent of the training set. Then it is said that the model is not generalizing well or that it is overfitting the training data. Below are few metrics which help in estimating the generalization capabilities of a given model:

1. *Cross validation*: One of the model validation techniques for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in a supervised setting. The goal of cross-validation is to test the model’s ability to predict new data that were not used in estimating it, to flag problems like overfitting.

To evaluate the models of the proposed model k-fold cross validation is used where k = 5. It divides the complete dataset into 5 sets, then trains model on 4 sets and evaluates performance on the remaining set. This is repeated 5 times and the 5 results can then be averaged to produce a single estimation which is used to present the result.

1. *Absolute Error (AE):* This measure is used to evaluate regression models. It is the

magnitude of the difference between the exact value and the approximation. If 𝑌𝑝𝑟𝑒𝑑𝑖𝑐𝑡𝑒𝑑 is a value of single prediction generated from a sample on all variables, and 𝑌𝑡𝑟𝑢𝑒 is the observed value of the variable being predicted, then AE of the predictor is computed as:

𝐴𝐸 = ||𝑌𝑡𝑟𝑢𝑒 – 𝑌𝑝𝑟𝑒𝑑𝑖𝑐𝑡𝑒𝑑||

## 4.3 No Fault Scenario

Under no fault scenario the power system remains at a steady state with specified current and voltage values. Therefore, the fault classification and location method could be deployed if the voltage or current values differ from the steady state values.

* human error associated with it while calculating the difference between the arrival time of the peaks. Currently, there are no software’s that can automate this process.

# Chapter 5: Conclusion

## 5.1 Summary and Findings

Transmission lines safeguard against exposed fault is the most critical task in the protection of power system. The purpose of a protective relaying is to identify the abnormal signals representing faults on a power transmission system. So, fault classification and location is necessary for reliable and high speed protective relaying.

The research work presented in this thesis provides two promising architectures for fault classification and location on transmission line using machine learning techniques. The first architecture deploys four decision tree models one for each phase for fault identification in each phase. “Divide and conquer” strategy has been used to classify faults on each phase with greatest accuracy. Each tree classifier has two possible target classes [-1, +1]. Class of ‘+1’ indicates that a phase is involved in the fault. Once the type of fault is known, travelling wave method is used to locate the fault on the transmission line. The wavelet transformation coefficients (WTCs) of the post fault transient phase voltage at a terminal is computed. The time difference between first two peaks of WTC2 is used to calculate the distance from a reference terminal. While travelling wavebased methods have few demerits, they are very reliable and can give accurate results. The method was tested on real data for both two and three terminal transmission models. The fault classification did better than baseline accuracy and the fault distance error calculated was within ±1% to ±2% of the line length.

Decision trees give better prediction with more data. Then we come to a question of how much training data is sufficient for getting a good prediction. SVM is an attractive classifier when classification process needs to be made useful and economical with smaller number of samples. SVM models have similar functional form to neural networks and radial basis functions, both popular data mining techniques. However, neither of these algorithms has the well-founded theoretical approach to regularization that forms the basis of SVM. The quality of generalization and ease of training of SVM is far beyond the capacities of these more traditional methods which makes it a better pick than decision trees.

In the second architecture, multi-class support vector machine model is used to classify 10

types of short circuit faults. When tested SVM fault classification model on real data, it outperformed decision trees with 100% accuracy on two and three terminal transmission lines. For fault location, Support Vector Regression was used for three terminal transmission line and Linear Regression for two terminal transmission lines. The error of the fault distance was same as the travelling wave-based method used in first architecture. However, regression models don’t require any manual calculations or measurements.

Since both the architecture do not use current signals, the presented approach is immune against current-transformer saturation and its related errors. Additionally, the presented method uses simulated data for training which means it can be deployed even when a new transmission line is added to the transmission network which does not have any historical records of fault. The presented architectures in this thesis are tested and validated against real data of faults. In this thesis, the test set sizes for both two and three terminal line configuration had only three samples in total. Therefore, the method needs to be tested more robustly. Overall, preliminary findings are promising but more data is needed to gain confidences in the reliability of the methods.

## 5.2 Challenges

In the initial stages, collecting the simulated data had two stages. First, the instantaneous post fault transients simulated in ASPEN were collected in .CFG as ASPEN only had one format option to save the data. After collecting the simulated data, the CFG was saved to a CSV file with help of tool called WAVEWIN (use to analyze the signals). This process creating each file manually was very tedious even to have a reasonable number of samples. To overcome this, a ASPEN script was written which could write the instantaneous values directly to the CSV file which reduced the data preparation time to 30 secs which otherwise would take several hours manually.

Secondly, since transmission line faults are rare compared to the faults on distribution lines, there was availability of limited real data to test the presented architectures. However, the preliminary results show strong potential in the architectures.

## 5.3 Future Work

In this thesis, problem of a having an adaptive, effective and resilient method for fault classification and location is addressed. There are several lines of research arising from this work which could be pursued as mentioned below:

* Since the test set considered in the thesis had just three samples in total (small size), the architectures need to be tested and evaluated with more test samples robustly in future. Additionally, architecture I had low accuracy of about 50% on the test set which may be due to low sample size. Therefore, with more test set samples, this result needs to be reevaluated. However, the preliminary results are promising to lead to further investigation.
* Four Terminal transmission lines aren’t very far from implementing. The proposed method may be possible to be extended N-terminal transmission line in the future work which would give different insights.
* The described method assumes single fault on the transmission line. However, chances of multiple faults, double circuit faults etc. though are low cannot be ignored from protection perspective.
* “Evolving Faults” are faults beginning in one phase and spreading to another phase after a few cycles. Thus, evolving fault consists of two faults: primary fault and secondary fault according to their fault inception times. Extending proposed method to classify and locate these faults would be an interesting topic for future work.
* The method proposed could be used for real-time monitoring of grid using spark-python. Once the fault has occurred, an alert could be set up from the method which would give the fault type and distance instantly.
* Complex fault sequence elements could be incorporated in data which will increase the robustness of the machine learning technique and results could be compared.

# Bibliography

s