Comparative Study of HVAC and HVDC Transmission Systems With Proposed Machine Learning Algorithms for Fault Location Detection

by

Bassam Albannai

A Thesis Presented in Partial Fulfillment
of the Requirement for the Degree
Master of Science

Approved March 2019 by the
Graduate Supervisory Committee:

Yang Weng, Chair
Meng Wu
Som Dahal

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

High Voltage Direct Current (HVDC) Technology has several features that make it particularly attractive for specific transmission applications. Recent years have witnessed an unprecedented growth in the number of the HVDC projects, which demonstrates a heightened interest in the HVDC technology. In parallel, the use of renewable energy sources has dramatically increased. For instance, Kuwait has recently announced a renewable project to be completed in 2035; this project aims to produce 15% of the countrys energy consumption from renewable sources. However, facilities that use renewable sources, such as solar and wind, to provide clean energy, are mostly placed in remote areas, as their installation requires a massive space of free land. Consequently, considerable challenges arise in terms of transmitting power generated from renewable sources of energy in remote areas to urban areas for further consumption.

The present thesis investigates different transmission line systems for transmitting bulk energy from renewable sources. Specifically, two systems will be focused on: the high-voltage alternating current (HVAC) system and the high-voltage direct current (HVDC) system. In order to determine the most efficient way of transmitting bulk energy from renewable sources, different aspects of the aforementioned two types of systems are analyzed. Limitations inherent in both HVAC and HVDC systems have been discussed.

At present, artificial intelligence plays an important role in power system control and monitoring. Consequently, in this thesis, the fault issue has been analyzed in transmission systems, with a specific consideration of machine learning tools that can help monitor transmission systems by detecting fault locations. These tools, called models, are used to analyze the collected data. In the present thesis, a focus on such models as linear regression (LR), K-nearest neighbors (KNN), linear support vector machine (LSVM) , and adaptive boost (AdaBoost). Finally, the accuracy of each model is evaluated and discussed. The machine learning concept introduced in the present thesis lays down the foundation for future research in this area so that to enable further research on the efficient ways to improve the performance of transmission line components and power systems.

ACKNOWLEDGEMENTS

At this moment of accomplishment, I feel deeply grateful and indebted to Prof. Yang Weng, my advisor and the chair of my thesis defense committee, for giving me the golden opportunity to work with him. His patience, enthusiasm, motivation, and immense knowledge are incommensurable. Despite his busy schedule, he was always willing to help me improve my academic and research skills, and all his assistance in guiding my thinking, progressing, and writing is deeply appreciated. Prof. Yang Weng took me first a student without research experience and then a graduate who has finished his Masters thesis, patiently and enthusiastically. I could not have had a better supervision and mentorship of my Masters thesis.

My gratitude also goes to all members of my thesis committee: Prof. Meng Wu and Dr. Som Dahal for their time in reading, reviewing, and providing valuable comments on my thesis work.

I would also like to acknowledge the financial support of the Kuwait Ministry of Higher Education (MOHE) for the full scholarship it granted me to conduct this research.

Finally, I would like thank my parents, family, friends, and classmates for their unconditional support and encouragement along this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure | Page

NOMENCLATURE

$HVDC$	High voltage direct current.
$HVAC$	High voltage alternating current.
ROW	Right-of-way.
CB	Circuit breaker.
N/A	Not Applicable.
LR	Linear regression.
KNN	K nearest neighbor.
$LSVM$	Linear support vector machine.
$AdaBoost$	Adaptive Boost.
DC	Direct current.
AC	Alternating current.
$KNEC$	Kuwait national electricity company.
Kw	Kilo watt.
KOC	Kuwait oil company.
RMS	Root mean square.
RPM	Revolution per minute.
OCR	Optical character recognition.
EDA	Exploratory data analysis.
RTD	Real time data.

1. INTRODUCTION

1.1 Kuwait Electrical System [1]

1.1.1 Electrical History of Kuwait

The discovery of oil in Kuwait, which remains a major source of national wealth, ushered the nation into the era of cultural awakening and revival in different walks of life, including social, structural, educational, and economical. Power supply has played a vital role in laying down the foundations for this awakening and in meeting the needs and requirements of such cultural march. A brief historical review provided below demonstrates the extent to which power supply has developed over the last several years. Before the construction of the first small (DC) electric plant by the Kuwait National Electricity Company (KNEC) in 1934, most people in Kuwait used kerosene lamps for lighting. After 1934, production started with two (30 kW) generators, and the power was distributed by +200 V DC line. At first, the number of consumers was rather small–in fact, by the end of the first year after the plant was constructed, there were only 60 consumers. However, by 1940, the number of electricity consumers increased to 700, which required increasing the installed capacity to 340 kW. After a period of stagnation during the Second World War, the KNEC decided to phase out the direct current system and introduced a three-phase 380/220V, 50 Hertz alternating current. Accordingly, in early 1949, a new plant comprising two (200 kW) generators was erected at Murgab (Centre of Kuwait); a year later, in 1950, a third (200kW) generator was added while the DC system was also phased out. To cope with the increasing demand for electricity, the KNEC obtained a used

[1] The information of this section were taken directly from "Kuwait Yearly Statistical report, 2017" [1]

(500 kW) generator from the Kuwait Oil Company (KOC), thereby bringing up the installed generation capacity to 1100 kW (1.1 MW). Owing the rapid progress and growth of the country, demand for electricity dramatically increased, rendering then the available plants unable to cope with this demand.

In 1951, the Government bought the shares of the KNEC and founded the Department of Electricity to adequately provide and distribute electricity supply. Since then, the Kuwait electric system has tremendously grown. At present, electricity in Kuwait is produced in three types of generation plants. In what follows, these three types of electricity generation plants are discussed in further details.

1.1.2 Kuwait Current Generating Sources

Kuwait generation plants can be broadly categorized into three types–namely, (1) gas turbine units; (2) steam turbine units; (3) and combined cycle units. Steam and combined cycle units are large capacity units, while gas units are typically small capacity units. Fig. 1.1 shows the available generation units in Kuwait as of 2017.

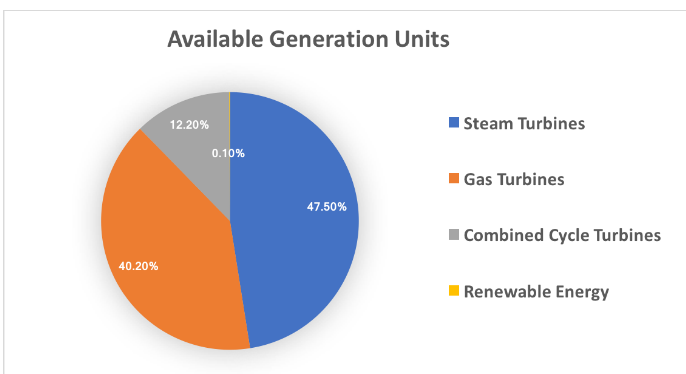


Fig. 1.1. Kuwait Available Generation Units

1.1.3 Kuwait Future Demand Estimation

At present, Kuwait is facing a rapid growth of population, which has a direct impact on the growth of the demand for electricity. Fig. 1.2 shows an estimate of the future growth of demand for electricity in the country[1].

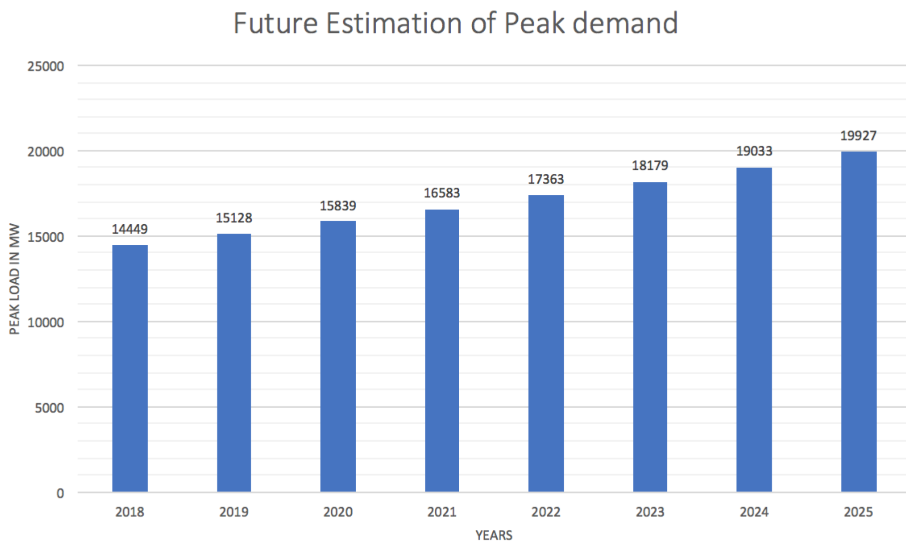


Fig. 1.2. Kuwait Future Estimation of Peak Demand

As can be seen in fig. 1.2, electricity consumption is expected to grow from 12,229 MW in 2018 to 19,927 MW in 2025. The continued industrial and urban development necessitates a considerable expansion of power production, which entails relying on natural sources of energy as part of the power production expansion.

1.1.4 Kuwait Renewable Project in 2035

As shown in fig. 1.1 in Section 1.1.2, the current Kuwait power system heavily depends on steam turbines, gas turbines, and combined cycle turbines, with each type making a different contribution to the total electricity generation (47.5%, 40.2%, and 12.2%, respectively). Of note, as can be seen in fig. 1.1, the generation of renewable energy in Kuwait remains very low, comprising mere 0.1% of the total generation.

However, by 2035, Kuwait aims to satisfy 15% of its total electricity demand by the energy obtained from renewable sources [2]. The factors that underpin this determination include the natural increase in the electricity consumption, the worlds strong interest in renewable sources of energy, and supportive climate in Kuwait (with sunny weather most of the year). Therefore, from the total expected demand of 19,927 MW in 2025 (see fig. 1.2), the power expected to be generated from renewable sources by 2035 should be around 3000MW. Moreover, Kuwait aims on investing in five islands in the country as a part of the 2035 project, which could increases the estimated demand in fig. 1.1 to more than 20,000MW.

1.1.5 Renewable Location and Transmission

An important factor in exporting the power from renewable energy producing facilities to the urban areas is the location of those facilities. Due to the fact that renewable farms require a massive land space, most renewable farms are located far away from urban areas. For instance, the facilities planned within the aforementioned renewable project in Kuwait will be constructed in the North West of Kuwait, near Saudi Arabias boarder. Consequently, in order to decrease losses in transmission operation, installation of high voltage transmission lines has to be planned. Mathematical equations for transmission losses are discussed in [3].

1.2 Transmission Lines Overview

Transmission line is conventionally defined as a conductor that transmits power from point A to point B. Overall, there are two types of current in transmission lines: alternating current (AC) and direct current (DC).

Alternating current (AC), as is clear from its name, is an electric current that which periodically reverses direction. By contrast, direct current features a constant current and voltage. A detailed description for the characteristics of AC and DC is provided in [4].

In terms of the transmission concept, transmission operation is usually performed at a high voltage level to reduce losses, and a high voltage level is conventionally defined as the one that starts with 100 kilo-volts. Therefore, typical electrical networks have transmission line at the highest voltage level through all system stages. Consequently, two terms are introduced: high- voltage alternating current (HVAC) and high-voltage direct current (HVDC). In the present thesis, more focus will be given to HVDC. However, due to such advantages as easiness in control, easiness in generation, cheaper equipment, and ability of stepping up/down voltage easily using a simple transformer, HVAC is more popular than HVDC [4].

1.2.1 What Is HVDC?

The high-voltage direct current (HVDC) system is a system used for bulk power transmission over long distances with minimum losses using overhead transmission lines or submarine cable crossings. Moreover, the technology is adopted to interconnect different power systems with varying frequencies (asynchronous interconnections). In essence, due to the limitations of HVAC such as reactive power loss, stability, current carrying capacity, operation and control, HVDC is a system of interest [5]. In the HVDC system, the transformer steps-up the generated AC voltages to the required level. The converter station takes up the electric power from one point in the three-phase AC network and rectifies it to DC, which is then transmitted through overhead lines or cables [6]. At the receiving end, an inverter converts the DC voltage back to AC, which is stepped down to the distribution voltage levels at various consumer ends. This technology is suitable for transmitting rated power range between 100-10,000MW. Fig. 1.3 shows the HVDC component configuration.

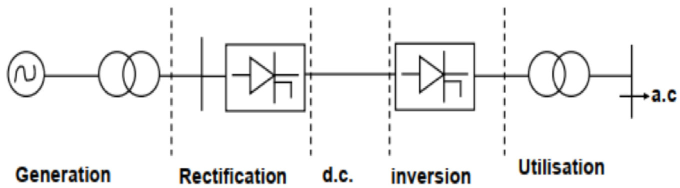


Fig. 1.3. Component Configurations of an HVDC System

1.2.2 HVDC Configurations

Depending on several factors, such as reliability, location, the arrangement of the pole and earth return, as well as the capacity to transmit bulk power, the following five HVDC system configurations can be discerned: (1) monopolar; (2) bipolar; (3) homopolar; (4) back-to-back; and (5) multi-terminal. In what follows, we discuss each of these configurations in further detail.

1.2.2.1 Monopolar Link Configuration:

A monopolar HVDC system consists of a single conductor connected to one terminal of the converter, while the other terminal is connected to the ground to form a return path. This system is conventionally used to transmit power over the sea to reduce cost [7]. The subsea cables installed using a monopolar scheme employ special electrodes for the earth return. However, this earth return path through the sea may lead to environmental concerns, such as corrosion of metallic objects. Another limitation of this system is that it is not suitable for cable crossings in freshwater and in the areas of high sensitivity of the earth. In order to overcome these challenges, the system can use a low-voltage conductor as a return path, while the DC circuit can use its own grounding. The advantages of this system configuration are as follows:

1. The system requires less conductor material, as the ground acts as the return path.

2. There is a less corona effect on the DC line due to the negative polarity of the conductor with respect to the ground.

3. The system reduces insulation costs.

Fig. 1.4 shows the Monopolar link configuration.

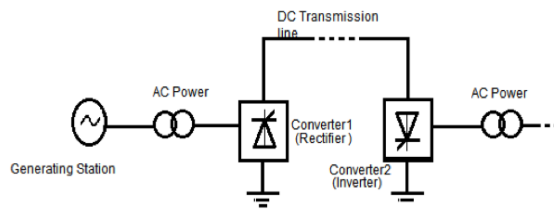


Fig. 1.4. Monopolar HVDC Configuration

1.2.2.2 Bipolar Link Configuration:

A bipolar HVDC system is a two-pole system where one conductor has a positive polarity, while the other one has a negative polarity. The advantage of this scheme over monopolar link configuration is that, whenever a fault occurs in one of the conductors, the other pole sustains the operation by acting as a monopolar link with the ground [8]. Furthermore, a bipolar link system transmits more power than a monopolar link system. In addition, there are no corrosion concerns, since the current flows in a loop and does not go through the grounded return. However, despite its advantages, the bipolar system is more expensive than a monopolar HVDC configuration due to the high cost of terminal equipment; another limitation of this system is that there are high corona losses [9]. Fig. 1.5 shows a bipolar HVDC system configuration.

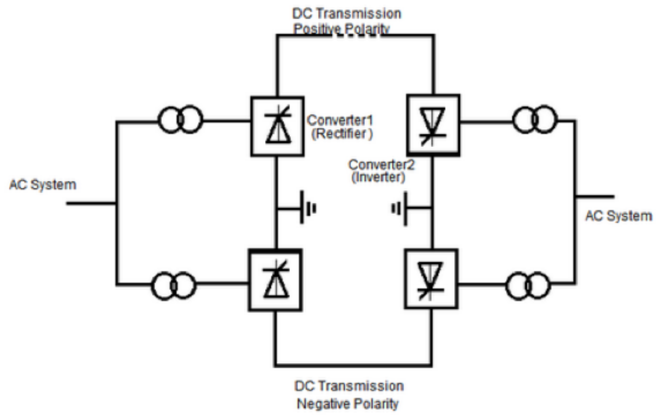


Fig. 1.5. Bipolar HVDC Configuration

1.2.2.3 Homopolar Link Configuration:

A homopolar HVDC system consists of two conductors of the same polarity, usually negative. The configuration adopts either earth or metal for its return, and its shunted poles reduce insulation cost [10]. However, this link is unpopular in the current transmission systems. Fig. 1.6 shows the configuration of a homopolar link.

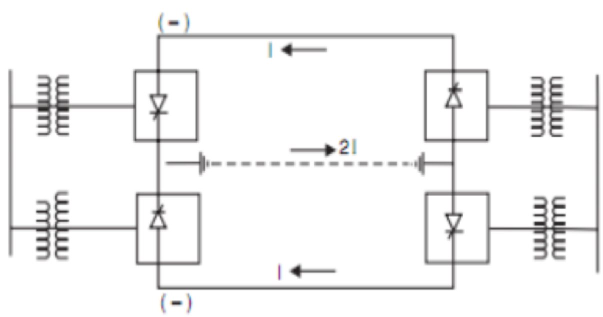


Fig. 1.6. Homopolar HVDC Configuration

1.2.2.4 Back-to-Back Configuration:

One of the primary functions of the HVDC system is asynchronous interconnection. A back-to-back HVDC configuration is a system used to connect power systems with different frequencies [11]. The back-to-back scheme is usually small and consists of two converters close to each other, as this system is mostly used for only connecting asynchronous interconnection (see figure 1.7).

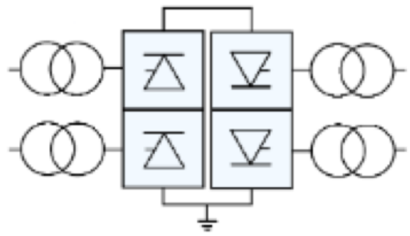


Fig. 1.7. Back-to-Back HVDC Configuration

1.2.2.5 Multi-Terminal Configuration:

Multi-Terminal HVDC configuration is a transmission system that consists of more than two converter stations (see fig. 1.8). This scheme is more complicated than monopolar and bipolar link configurations and is applied for offshore interconnections of wind farms and oil rigs [12]. Advantages of a Multi-Terminal HVDC configuration are as follows:

1. This configuration requires less conductor material, as the ground acts as a return path.
2. This configuration has low insulation cost.
3. There is less corona effect in negative polarity conductors.
4. It is possible to achieve reversal power and avoid power interruptions by transmitting power through other conductors in the event of a fault.

However, a disadvantage of Multi-Terminal HVDC configuration is that the return path can cause corrosion of metal structures, such as underground communication cables.

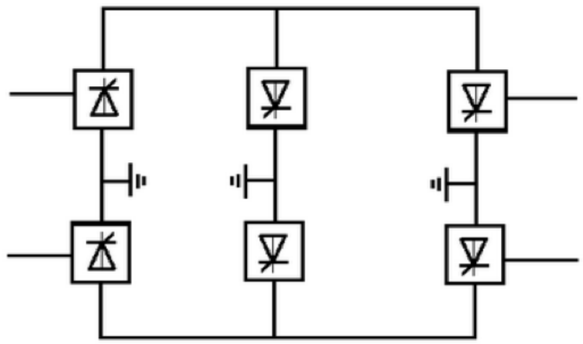


Fig. 1.8. Multi-Terminal HVDC Configuration

1.2.3 Converters Types

Modern HVDC systems use two basic converters: (1) the line-commutated current source converters (CSC) and (2) the self-commutated voltage source converters (VSCs). In Sections 1.2.3.1 and 1.2.3.2, we discuss these converters in further detail.

1.2.3.1 Current Source Converters (CSCs):

The conventional current source converter (CSC) uses thyristor valves and requires a source of potential to operate. Its building block is a three-phase, full-wave bridge called a six-pulse [5]. Figure 1.9 shows this configuration.

In the HVDC systems, the CSC converter generates harmonic currents from the surrounding AC network by absorbing reactive power, which affects electrical systems. Consequently, the filter circuit limits the AC harmonic currents and compensates the amount of reactive power absorbed by the converter. The correct converter operation depends on the AC system voltage. The control system for the DC circuit reverses this voltage in order to change the direction of power flow, while the reactor smoothens the DC current and reduces the peak current in the event of a fault [13]. The HVDC stations that use CSC converter experience a power loss of 0.5-1% per converter

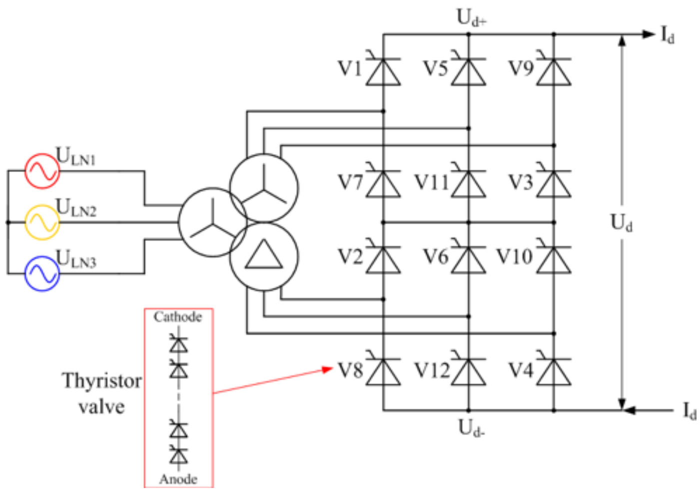


Fig. 1.9. Current Source Converters (CSCs)

station.

1.2.3.2 Voltage Source Converters (VSCs):

Voltage source converters are HVDC components that, instead of relying on line commutation for their operation, require the systems DC side to have a voltage source. Despite the polarity or the amount of current flows, the voltage source maintains the required potential across its terminals [14]. Fig. 1.10 shows the operation of a single phase two-level VSC.

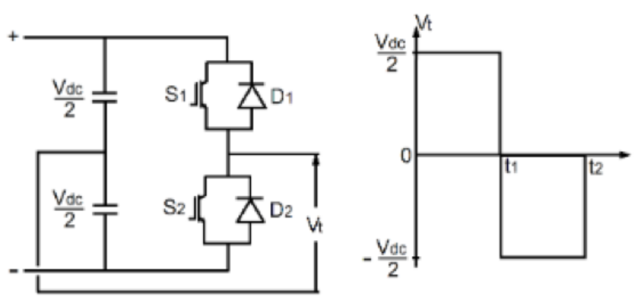


Fig. 1.10. Two-Level Basic Operation (VSC)

The system above is a half-bridge consisting of two switching cells, each with a controllable and unidirectional insulated gate bipolar transistor (IGBT). These switches are self-commutated and connected to a diode facing an anti-parallel direction, which ensures that the bridge potential has only one polarity and current can flow in both directions. Moreover, the circuit configurations allow for ON and OFF switching of the IGBTs using a Pulse Width Modulation (PWM) control scheme [13]. The split capacitors help to maintain the net voltage. Along with the two-level converter, there are other VSC configurations, such as three-level converters and modular multilevel converters. The advantages of the VSC system are as follows:

1. It has control capability for both active and reactive power.
2. Due to its flexibility from its control capability, the converter can be placed on any network.
3. Self-commutation of VSC allows for a back start, enabling the component to handle balanced three-phase voltages.
4. The converter improves voltage stability.
5. Unlike conventional line converters, VSC has no reactive power demand, but can control it to regulate the AC system.

Table 1 shows a summary of the comparison of CSC and VSC.

1.2.4 HVDC Applications

The HVDC system is effectively applied the following applications:

1. Bulk power transmission over long distances.
2. Underground and submarine cable crossings for transmission systems above 30km.

3. The asynchronous connection of the AC system with different frequencies.

4. Control and stabilization of the power system with the power flow control.

Table 1.1. Comparison Between CSC and VSC

Current Source Converter (CSC)	**Voltage Source Converter (VSC)**
The technology is already developed	The technology is still developing
Uses thyristor valves that depend on AC voltage for commutation	Uses IGBT and the system has self-commutation
Commutation failure can occur	Commutation failure can not occur
Requires reactive compensation	Does not require reactive compensation
Requires switchable AC harmonic filters	Does not require switchable AC harmonic filters
Requires converter transformers of special design	It can use conventional transformers
Requires DC voltage polarity reversal	No reversal of DC voltage polarity required because power flow can be controlled in both directions
Incurs 0.5% to 1% conversion losses of transmitted power	Incurs 1-2% conversion losses of transmitted power

1.2.5 Brief Cost Analysis for HVAC and HVDC

A key factor to consider in constructing a transmission line is the cost, including both the construction cost of the system components needed and long-term costs generated by losses. Therefore, comparing HVAC or HVDC transmission systems, many aspects should be carefully considered. Several previous studies have analyzed the HVAC and HVDC transmission line costs [14]-[15].

Based on the results of previous research [14]-[15], a study has been done on the Nelson River project in Manitoba, Canada, which started the construction in 1966 for Phase I and completed its Phase 3 in 2018 [16]. HVAC transmission line consists of three conductors (i.e. three phases), which directly impacts the right-of-way (ROW). By contrast, HVDC consists of only two conductors, which decreases the costs of the ROW as compared to HVAC. While Nelson River AC transmission line cost amounts to \$955k/mile[17], the cost of the transmission line for the DC line in the same project ranges between \$345k/mile and \$370k/mile for ±400kV to ±700kV. Therefore, the cost of the HVDC transmission line is considerably lower than the cost of HVAC transmission line in terms of the line itself. However, the main components of HVDC that consume over 50% of the whole system cost are the converters [17], including the rectifier and the inverters at each end. Many other comparison elements are being considered when comparing costs. Several previous studies have performed a detailed comparative analysis of the elements in the two transmission systems [18]-[19] (see Table 1.2 for a summary).

In terms of distance, the break-even distance for overhead transmission lines (see Figure 1.9) has been discussed in terms of line distance in [20]. As can be seen in Figure 1.9, with regard to overhead transmission line, the break-even distance is between 400 and 700 km. Therefore, in cases where the overhead transmission line is less than 400km, the AC transmission system would be the most appropriate system. On the other hand, with regard to the underwater cable, a different analysis has to be performed. In [21], a detailed comparison of an underwater cable of both HVAC and HVDC has been undertaken (see Figure 1.10 for a summary of the results). As can be seen in the figure, the underwater cable break-even distance is very short compared to the overhead transmission line break-even.

In summary, in transmission line planning, many aspects have to be considered.

Table 1.2. System Cost Elements For a Constant Power (MW) Transmitted and a Constant Transmission Length [19]

HVAC Cost Terms	**HVDC Cost Terms**
Right-of-Way	Right-of-Way
Load density per acre of ROW	Load density per acre of ROW
Transmission voltage	Transmission voltage
Conductor specifications (Size and type)	Conductor specifications (Size and type)
Substations equipment, switching stations breakers, transformers, and station civil work	Rectifier, inverter, filters, DC circuit breakers, smoothing reactors and station civil work
System reinforcement	System reinforcement
Environmental impact	Environmental impact
N/A	Conversion of voltage from AC to DC and Vice-a-Versa

The results in Figures 1.9-1.10 focus only on system losses, construction costs, and distance of the line. However, in real-life applications, many other aspects–such as the stability of the added and connected systems–should also be considered.

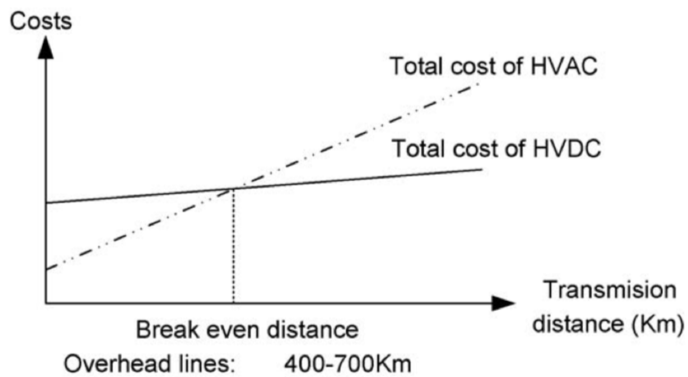


Fig. 1.11. Costs of AC and DC Overhead Lines Based on Distance [20]

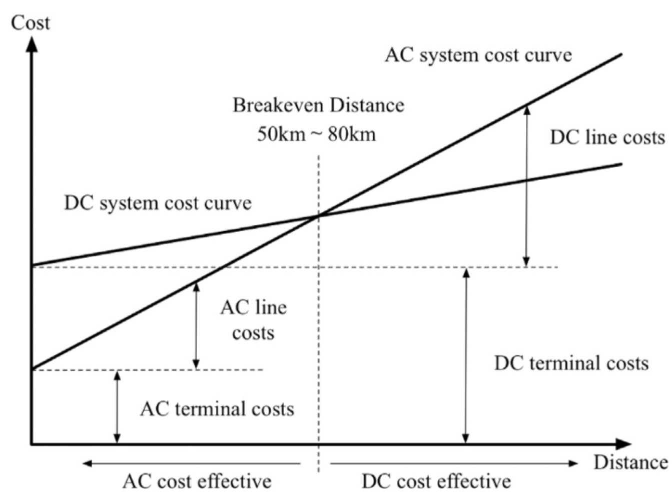


Fig. 1.12. Costs of AC and DC Underwater Cable Based on Distance [21]

1.2.6 Disadvantages of HVDC and HVAC

Both HVDC and HVAC systems have several limitations. More specifically, the disadvantages of HVDC are as follows:

1. Compared to converter stations used in the HVAC systems, converter stations used in the HVDC are expensive and complicated.

2. The design and operation of multi-terminal HVDC systems are sophisticated compared to HVAC.

3. Current and voltage harmonics are generated during conversion, which requires expensive filters.

4. The presence of high-frequency constituents in the DC transmission causes interference in the communication systems near the HVDC system.

5. The grounding of the HVDC system is complex and complicated.

In its turn, the HVAC systems have the following limitations:

1. Compared to HVDC, HVAC has a very high interference with communication lines

2. It is impossible to connect two unsynchronized HVAC (e.g., a 60Hz to a 50Hz line).

3. Compared to the HVDC systems, the HVAC systems are more likely to experience corona effects during bad weather compared to HVDC.

4. Unlike in HVDC, inductive and capacitive parameters are a limiting factor in the HVAC systems.

Therefore, both HVAC and HVDC systems have their specific advantages and disadvantages. For a better visualization and analysis on a specific electric system, and in order to obtain a conclusion for the appropriate selection, a software simulation considering the details of both systems has to be performed.

1.3 Research Purpose

The present thesis focuses on both HVAC and HVDC systems in the specific context of the Kuwait system, particularly the countrys 2035 renewable sources and islands project. We also consider the applications of artificial intelligence in the selected system. At present, Kuwait does not have any HVDC systems in its interior power system. Due to the difference in the operating frequency in Kuwait and Saudi Arabia (50Hz and 60Hz, respectively), the former is connected to the latter through a back-to-back HVDC system.

Both HVDC and HVAC systems are widely used around the world. However, in each specific case, a careful analysis is needed to determine which of the two systems would best fit the needs of a specific country or region. Consequently, despite the many and varied advantages that have made HVAC transmission more popular globally, the HVDC transmission would be preferred in numerous other cases. For instance, the elimination of challenges of synchronizing various control system operations within many power systems could become a reality. On HVDC transmission lines, there is usually a fast-acting emergency control systems, which is essential in terms of enhancing reliability and stability of power systems. However, many previous studies have demonstrated that HVDC is not necessarily the best option for transmission system.

Therefore, the present thesis aims to assess the validity and test the advantages of the HVDC system in the Kuwait context. The results are expected to benefit not only on Kuwaits system, but also those of the countrys neighbors and, more globally, demonstrate the efficiency of having an interconnected system through HVDC links in the Middle East region.

The selection of the transmission system will be based on the simulation results and analysis. Moreover, an important issue for the selected system will be analyzed and solved using artificial intelligence.

Artificial intelligence, which allows to fix a common problem at a minimal cost, can be meaningfully applied in the HVDC project. Nowadays, data are available all the time, particularly in a power system where measurement devices are installed almost everywhere in the system. However, in Kuwait, the main purpose of those measurement devices is only monitoring the power grid. However, enhancing and modernizing the power system operation in Kuwait using artificial intelligence would benefit not only Kuwait, but also its neighbor countries.

Recently, artificial intelligence applications in the Middle East region have become an object of considerable research. However, the power system is not among the priority fields for application of artificial intelligence. To illustrate, at a recent conference entitled Artificial Intelligence Week of Middle East held in Dubai, the United Arab Emirates, the main focus was on involving artificial intelligence in the government, banking and finance, and health care. In this context, the main research purpose of the present thesis is to demonstrate that artificial intelligence applications can be meaningfully used to enhance and improve Kuwaits power system. Specifically, by proposing a solution using machine learning tools, this thesis seeks to lay down the foundation for further work in the same field.

1.4 Thesis Structure

The remainder of the present thesis is structured as follows. Chapter 2 illustrates current challenges and issues of transmitting bulk energy over long distances. To this end, we compare the efficiency of the HVDC and HVAC systems in terms of necessary construction elements, delivered power, and other pertinent characteristics of those two systems. Furthermore, we also discuss the importance of stability in the power system and the factors that lead to instability of this system. The chapter concludes with the fault analysis in the power transmission system.

Chapter 3 outlines the methodology and the research design of the experiment performed in the present thesis. To this end, we start by presenting the simulation software used for studying the HVAC and HVDC systems. This is followed by a detailed presentation of the experimental set-up of the power system and the assumptions made. Finally, the system set-up is validated to ensure reliability and validity of the results.

Chapter 4 provides further detail on all steps of the experiments. Specifically, this chapter presents all types of experiments carried out throughout this thesis, discusses the results obtained from each experiment, and draws conclusions for the experimental results.

Chapter 5 focuses on the concept of artificial intelligence. We start by introducing machine learning and discuss its functionality vis-à-vis solving real-world problems. Secondly, we recapitulate on the importance of fault detection discussed in Chapter 2 and propose a machine learning tools to solve the issue at stake. For using machine learning tool, Section 5.4 describes the data collection process. Furthermore, Section 5.5 explains the tools and methods applied to the analysis of the collected data. Finally, we summarize the results, evaluate them in terms of accuracy, and discuss the findings to highlight the importance of machine learning in the power system operation and control.

2. RENEWABLE SOURCES INTEGRATION CHALLENGES AND ISSUES

2.1 Efficiency

Efficiency plays a huge role in power systems. The goal in power systems is to obtain the highest efficiency possible out of a particular system. However, many challenges could arise while trying to obtain high efficiency.

In order to decrease the losses, transmission of bulk energy either from renewable sources or from conventional generators has to be performed through the high voltage level. According to the electric loss equations discussed in [3], for both DC or AC transmission, a high level voltage for the same amount of sent power decreases the current, which, in turn, decreases the loss across the line. Therefore, in terms of efficiency, a smaller amount of energy is lost in HVDC, and DC eliminates reactive power; therefore, there is no reactive power in the DC line, so that only active power is flowing [22]. A DC line consists of two conductors for transmitting power: namely, the negative (-) conductor and the positive (+) conductor. On the other hand, the AC line for transmission consists of three lines (or three phases). Therefore, HVDC would require fewer conductors and narrower right-of-way, which results in less land uses and cheaper conductor equipment.

The main uses of HVDC include connecting offshores wind farms to onshore substations and transmitting power across the sea, where overhead lines are not applicable. This constitutes another considerable advantage of HVDC over HVAC. Specifically, the AC cables have large capacitance, which results in limiting the power transferred through the cable; therefore, in the AC case, the cable is carrying both load current and capacitive current. By contrast, a DC cable carries only load current with eliminating the capacitive current, which justifies using HVDC submarine cables for power transmission across the sea [23]. Another advantage of HVDC over HVAC is that the former eliminates the inductive voltage drop.

Given that DC voltage is constant in the whole operation, HVDC conductors carry more power compared to their HVAC counterparts. However, AC alternates periodically. Therefore, in AC, the root mean square (RMS) is considered the standard, where RMS is only about 75% of ACs peak voltage [22]. The insulation thickness and conductor spacing of the HVAC system are based on the peak voltage, rather than on the RMS value. On the other hand, since DC operates at a constant voltage, it allows the insulation and the conductor size to carry 100% of the power.

2.2 Stability and Fault Analysis

In this section, we investigate the role of system stability in the electric power system and determine the reasons of a systems instability. Stability of the system is the top priority in any secure operational electrical setup. Due to power system failures, systems can undergo major blackouts. In this context, it is essential to focus on stability of the system. In this section, we also provide an explanation about the high voltage direct current (HVDC) and how it helps the connected AC system to remain stable. Furthermore, we also introduce the fault analysis in the AC system with HVDC connected, as well as discuss the effects of the fault on the AC system and HVDC individually. Finally, this section discusses the importance of determining the fault location on a line.

2.2.1 Stability

System stability is a state of equilibrium between contradicting powers [24]. Power system instability refers to the capacity of an electric power system to maintain a given starting working condition and to recover a condition of working equilibrium after the systems exposure to an unsettling physical influence. The power system setup is an exceedingly nonlinear system that works in constantly changing conditions, with alternating loads, generator yields, topology, and key working parameters. Depending on the conditions, a disturbing influence might be little or extensive. Due to voltage fluctuations or recurrence variance which might affect the interconnected power system, an electric power system might start having stability issues [25]. There are various other factors, such as lightning, weather conditions, inappropriate wiring, vandalism, trees falling over transmission lines, aircraft collisions, excessive load, and collision of vehicles, which are harmful to the power system. These instability issues are referred to as faults in the system.

In the event of a fault, if the regular recurrence of swaying corresponds to the recurrence wavering of the generators, the engine loses synchronism, which is a fundamental condition for a power system.

Overall, for the transmission of electrical power, a high voltage direct current (HVDC) transmission system makes use of the coordinate flow with a more typical alternating current (AC) system [26]. The reasons for using high voltage direct current lines as connections in the AC transmission systems include stability, security, and affordability of these lines. It is considerably easier to control the current on the HVDC side by using terminating circuits of the thermistors installed in the two rectifiers and inverters. Exchanging activities can be performed on the AC side using AC circuit breakers (CB).

In essence, HVDC permits control transmission between unsynchronized AC transmission systems. Since power flows through an HVDC connection can be freely controlled at the stage point among the source and the load, this can settle a system against unsettling influences generated by quick changes in power [27]. In addition, HVDC permits the exchange of intensity between systems running at various frequencies, thereby enhancing the strength and economy of every electric system and thus permitting the trade of intensity between inconsistent systems.

Therefore, for transmission work and activities at longer distances, the HVDC systems might be more affordable, in that they are capable of reducing system instability and ensuring higher security. For submerged power links, HVDC keeps away from the overwhelming flows required to charge and release the link capacitance in each cycle. Therefore, in this case, the use of high voltage direct current is advisable in most electric power systems.

2.2.2 Fault analysis

A fault is basically defined as an unusual or abnormal condition in a power framework. Fault analysis, which includes determining security hardware and evaluation of the system unwavering quality, is among the key objectives in the power system setup with the AC supply and the HVDC lines [28]. At present, most is done using the high-voltage transmission system. In the event of a fault in the system, the working state of the entire system is disturbed, thereby halting the entire process. If the fault is a persistent, a serious loss of load and property harm may occur due to the blast, short circuit or fire. This can lead to dramatic economic losses.

Whenever a fault occurs in the AC system connected to the HVDC lines, the HVDC transmission lines suffer an immense loss in terms of energy and power. A transmission failure interrupts the entire power supply process [29]. Even when the system comprises a single phase, there is an estimated power loss of 30%. Despite the fact that the HVDC can bear twice its voltage before causing a failure, huge faults can still disrupt the entire system. However, when the fault is removed quickly, the power returns to its original value.

The main reason behind the fault reaching a higher level is that the AC transmission system becomes overloaded, which results in disconnection. This disconnection leads to an increased load on other lines, which has a dramatic negative impact on the entire system. The system shuts down, which causes major power failures in most regions. Main reasons for faults happening in a power include a protection failure, flashover, physical harm, or human mistake. In addition, deficiencies may be caused by either short circuits to the earth or between live conductors or might be caused by broken conductors in at least one phase.

Taken together, a short review of major factors causing faults in power systems and the consequences of these faults underscore the importance of appropriate and adequate fault analysis in power systems. In fact, fault analysis is the basic precondition to ensure security and reliability of power systems.

2.2.3 Importance of Determining Fault Location

To eliminate a fault, a crucial step is to determine the fault location in any electric system, particularly as concerns very long transmission lines. Faults might lead to fire breakouts that, in turn, can result in loss of property, death toll, and decimation of a power system. Moreover, failures can cut off power supply in various zones past the fault point in transmission and circulation arrange, prompting power outages [29]. In this context, it is essential to perform estimations of system voltages and flows amid faulty conditions, setting defensive gadgets capable of recognizing and limiting the destructive impacts of faults. Once the fault location is determined, the problem can be fixed easier, and the damages associated with the fault can be effectively reduced.

In view of the above, system stability must be the utmost priority of the electric system engineers, and all measures discussed above should be carefully considered before finalizing an electric system design and installation. Faults can be life-threatening and have a dramatically adverse impact on the economy. Therefore, it is very necessary to have a stable and secure system.

3. EXPERIMENTAL MODELLING

Research conducted in the present thesis is based on simulation. The experimental setups were simulated to better visualize the behavior of the system. Further detail on each system setup is provided in Chapter 4.

3.1 Simulation Software

In this section, we explain three software programs used in the present thesis. Sections 3.1.1-3.1.3 provide further detail on each software, clarify the reasons for choosing them, and specify our research purposes for using them in the present study.

3.1.1 ETAP

ETAP is defined as "a full spectrum analytical engineering software company specializing in the analysis, simulation, monitoring, control, optimization, and automation of electrical power systems"[30]. In industry, ETAP is one of the best simulation tools used on a daily basis. In a recent customer survey, ETAP scored 99% in overall customer satisfaction [30]. Moreover, in 2018, ETAP was awarded the product of the year by Consulting-Specifying Engineer Magazine [31]. In the present thesis, the role of ETAP is building the entire electric power system, including renewable sources, step-up transformers, HVDC link, HVAC link, step-down transformers, and loads. We use ETAP to visualize the results using its features of running power flows, both AC and DC, system efficiency, fault analysis, and system stability.

3.1.2 DigSILENT Power Factory

DigSILENT PowerFactory is defined as "a leading power system analysis software application for use in analyzing generation, transmission, distribution, and industrial systems"[32]. Its features are very similar, if not identical, to that of ETAP. However, in the present study, our purpose for using the PowerFactory simulation tool was not to build a power system or to test it. Instead, we used PowerFactory to collect Real Time Data (RTD) of the built system on ETAP. PowerFactory has the excellence function of the RTD collection and system monitoring. The real-time data were needed for our machine learning research (see Chapter 5). Without such data, testing the machine learning tool would not have been possible.

3.1.3 *Python*

Python is defined as "an interpreted, object-oriented programming language that has gained popularity because of its clear syntax and readability"[33]. An online platform called Anaconda was used to perform Python coding, as this platform contains all Python libraries needed for the present research. Python was mainly used at two stages. The first stage was the data analysis stage, while the second stage was building the required models and testing them. Detailed descriptions of the setup are provided in Chapter 5. An alternative in programming tool for machine learning could have been Matlab. However, we opted for Python, as it has many advantages over Matlab. Specifically, compared to Matlab, the coding statements in Python are more compact and readable. Furthermore, Python can be implemented on many platforms, such as Anaconda in our case, and it is free. Moreover, Python provides many choices and more graphics packages than Matlab, as well as provides controllability over the coding structures.

3.2 Experimental System Set-up

System setup was developed using ETAP; therefore, the electric power systems built in the present thesis are not a real existing systems. However, the topology of the system, ratings, frequency, and components of the built systems are typical of the Kuwaits system. For instance, in the built system, frequency was 50Hz; we also considered loads distances from the planned renewable sources location as the five islands that are going to be part of the Kuwait 2035 project. The distances selected for these five loads were estimated based on Google Maps. The reason this setup was underpinned by our consideration of the Kuwait 2035 project that involves a huge amount of renewable generation and five islands investment. Therefore, studying the power deliverability to those islands, which is a must project, and using the renewable project, our results would contribute not only to the planning stage of the project, but would also provide valuable insights in terms of the scenarios available for feeding those islands, and how are these scenarios could affecting the existing Kuwaits AC system.

Fig. 3.1 is a schematic representation of the first stage of the setup. First, a renewable farm was initially built consisting 4 photovoltaic (PV) arrays; then, inverters were used to convert DC power generated to AC power, so that a transformer can be used to step-up the voltage for the transmission purpose. Second, two transmission lines–HVAC and HVDC–were also added. However, only HVAC was actually connected in the schematic, since the switches of HVDC were disconnected. Then, a distribution substation was considered where the voltage was distributed to the five loads. The transmission line length was set at 180KM between the renewable farm and the nearest substation. The details of the transmission line parameters on ETAP matched those of real Kuwait transmission lines, so that to ensure that realistic results could be obtained from the simulation. In the setup, the 400kV transmission line voltage level and quadruple bundle conductors were also considered (see Fig. 3.1).

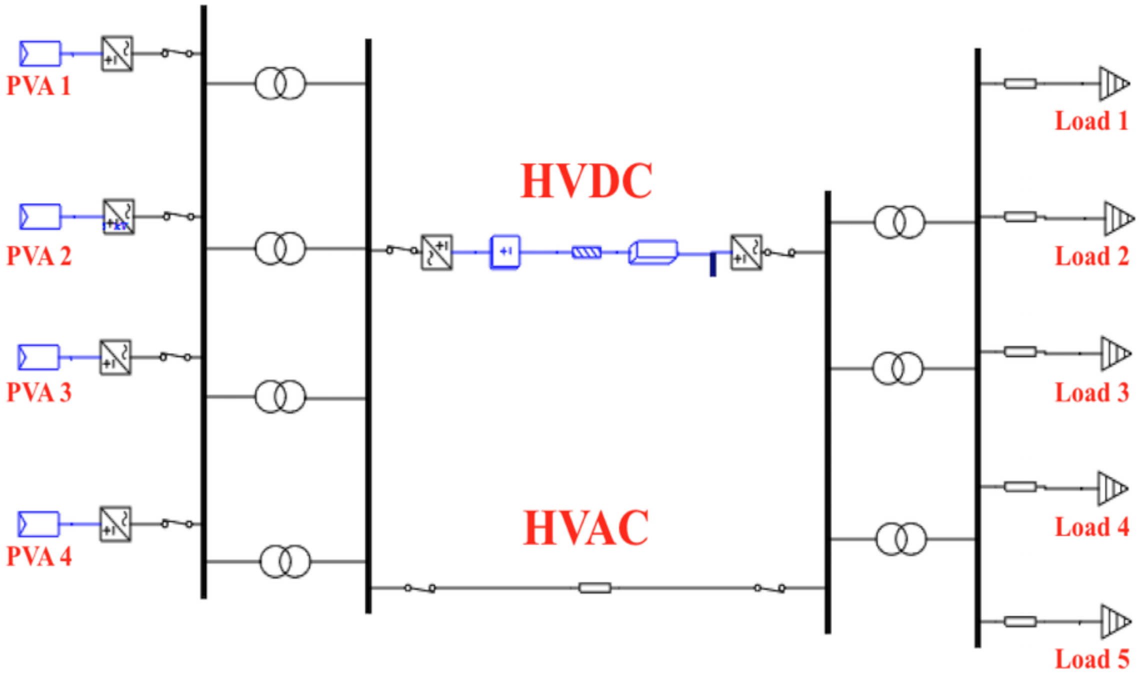


Fig. 3.1. First Stage Schematic

3.2.1 System Set-up Validation

In this section, we validate the setup shown in fig. 3.1 using the concept of voltage drop. In essence, voltage drop is the amount of electricity wasted due to the resistance of the transmission line. Previous studies have demonstrated that voltage drop increases with an increase in distance [34]. Therefore, in order to validate the system shown in fig. 3.1, the voltage drop was measured across the loads. The loads were labelled Load 1, Load 2, Load 3, Load 4, and Load 5 with the distance of 10KM, 30KM, 60KM, 75KM, and 90KM, respectively. Fig. 3.2 shows the expected results, i.e. that the load located far away from the substation would have more voltage drop percentage than the one located closer to the substation.

As can be seen in fig. 3.2, the validity of the system in terms of voltage drop is confirmed, since the tested system behaved in line with our expectation and consistently with previous research.

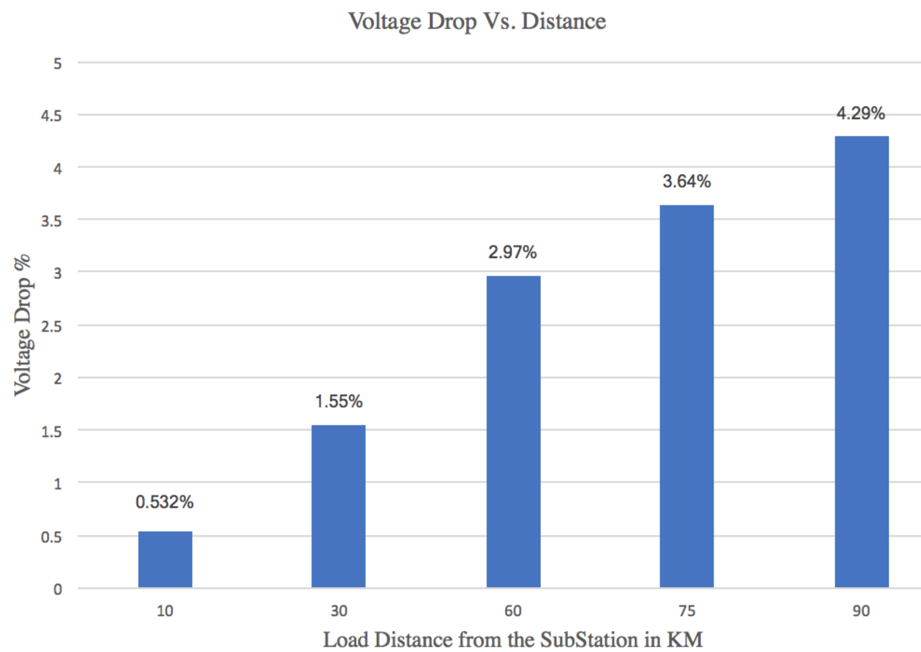


Fig. 3.2. Voltage Drop Vs. Distance

At this point, we have verified the validity of the setup shown in fig. 3.1. Therefore, further experiments will be performed on different setups (see Chapter 4 for further detail).

4. EXPERIMENTAL RESULTS AND ANALYSIS

In this chapter, we report the results of several experiments performed after a series of modifications of the setup shown in fig. 3.1 (see Chapter 3). In the first comparison experiment, a voltage drop and losses were studied for both HVAC and HVDC transmission lines shown in fig. 3.1. Each transmission line was tested separately by connecting it and disconnecting the other one. After connecting a transmission line, using the ETAP features, the losses and voltage drop across the transmission line were measured. Different transmission lengths in the range from 180Km to 700Km were considered, with a step of 50Km. The major aim of this experiment was to investigate the performance of both HVAC and HVDC in terms of distance, and how the distance affected voltage drop and losses in both systems.

4.1 Voltage Drop and Losses for HVAC Line

In this section, we report the results of the experiment when the HVAC line was connected, while the HVDC line was disconnected (see fig. 3.1). Table 4.1 reports the voltage drop values (in percent) and losses in megawatt (MW) for the HVAC system.

Table 4.1. HVAC Result Values for Losses and Voltage Drop

AC POWER FLOW		
TRANSMISSION SYSTEM SUMMARY		
LENGTH (Km)	**Losses (MW)**	**Voltage Drop (%)**
180	2.91	2.58
230	3.49	3.2
280	4.03	3.6
330	4.57	3.78
380	5.12	3.75
430	5.73	3.51
480	6.4	3.07
530	7.17	2.43
580	8.08	1.59
700	11	-1.2

4.2 Voltage Drop and Losses for HVDC Line

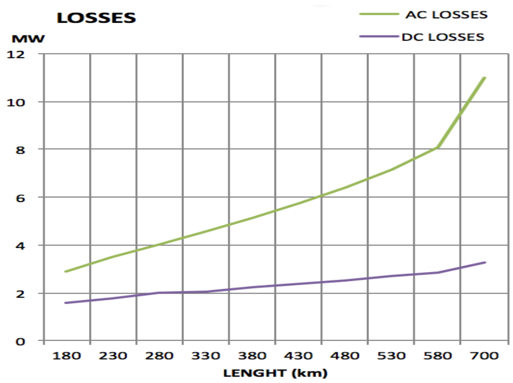
In this section, we report the results of the experiment when the HVDC line was connected, while the HVAC line was disconnected (see fig. 3.1). Table 4.2 reports the voltage drop values (in percent) and losses in megawatt (MW) for the HVDC system.

Table 4.2. HVDC Result Values for Losses and Voltage Drop

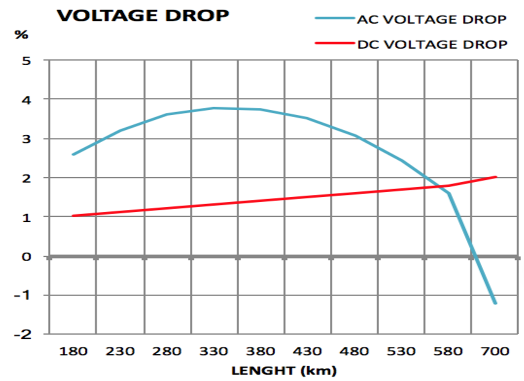
DC POWER FLOW		
TRANSMISSION SYSTEM SUMMARY		
LENGTH (Km)	**Losses (MW)**	**Voltage Drop (%)**
180	1.61	1.01
230	1.771	1.11
280	1.992	1.21
330	2.074	1.3
380	2.23	1.4
430	2.39	1.49
480	2.544	1.59
530	2.71	1.69
580	2.86	1.78
700	3.25	2.02

4.3 HVAC vs. HVDC Visualization Plots

For a better visualization of the data points obtained in Tables 4.1-4.2, two plots that compare the behavior of the two systems were drawn (see fig. 4.1).



(a) Losses in AC and DC Lines



(b) Voltage Drop Across AC and DC Lines

Fig. 4.1. Visualization Plots for Data Points in Table 4.1 and Table 4.2

System Stability

Further detail on system stability is provided in Chapter 2. In Sections 4.4, 4.5, and 4.6, we focus on the system stability by studying the behavior of several parameters in the event of a fault. In this experiment, a separate small AC system consisting of 4 AC generators and 4 step-up transformers was built. This system was connected to the same bus the renewable energy was being fed to. The idea was to apply a fault at the distribution line near one of the loads–in particular, the second load from the bottom in fig. 3.1, and the behaviors of three factors (namely, generator speed, distribution bus voltage, and load bus frequency) were studied. First, the three factors were analyzed without considering the integration of renewable sources, assuming that the AC system with its four generators was feeding the five loads by itself. Second, we analyzed the three factors upon integration of renewable energy using the HVAC transmission line along with the AC system feeding the five loads. Finally, we analyzed the three factors upon integration of renewable energy using the HVDC transmission line and considering the existence of the added AC system. The main idea in these three experiments was to visualize the results in the event of a fault in a connected AC system and to investigate how the fault would affect the system in the following three conditions: (1) without the integration of renewable energy; (2) integrating renewable energy using HVAC; and (3) integrating renewable energy using HVDC. The visualization results are presented and the conclusions are drawn in Section 4.7. Fig. 4.2 shows a modified schematic representation of the setup that was used in the experiments. In essence, this new setup is similar to the one shown in fig. 3.1, except for the fact that an AC part consisting of 4 AC generators was added.

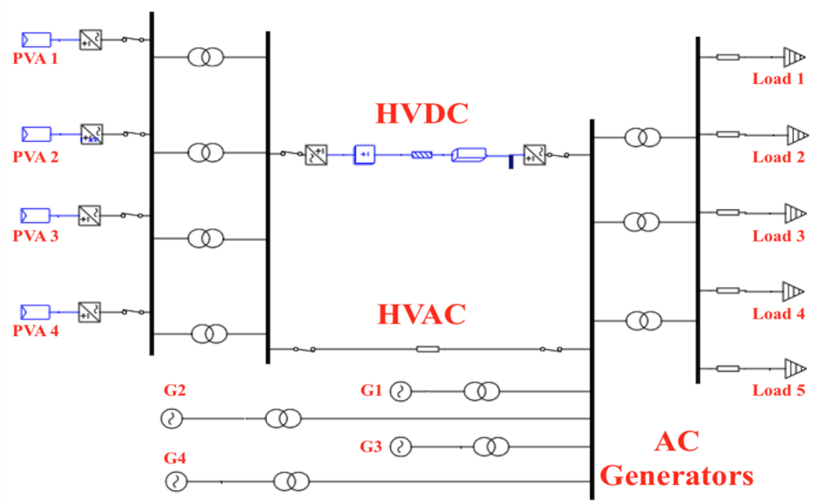


Fig. 4.2. Modified Schematic Including AC Generators Part

4.4 System Stability Without Renewable Integration

In this experiment, renewable sources and transmission lines were not included, meaning that the top left part of the schematic representation of the setup shown in fig. 4.2 was not considered. A fault was applied at the distribution line of the second load from bottom (Load 13). The fault was applied at 1.5 seconds and was cleared at 1.8 seconds.

4.4.1 Observations and Simulation Results

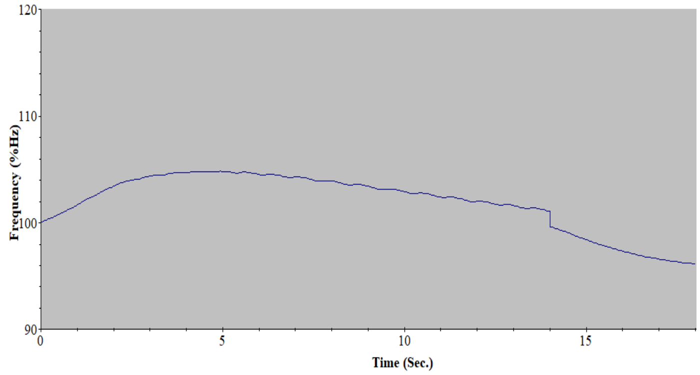
Fig. 4.3 shows the behavior of the three factors, including (a) the frequency oscillation for a period that included the fault time; (b) the substation bus voltage oscillation for a period that included the fault time; and (c) generator speed oscillation that resulted from the application of the fault (see fig. 4.3).

4.5 System Stability with Integrating Renewable Sources Through the HVAC Transmission System

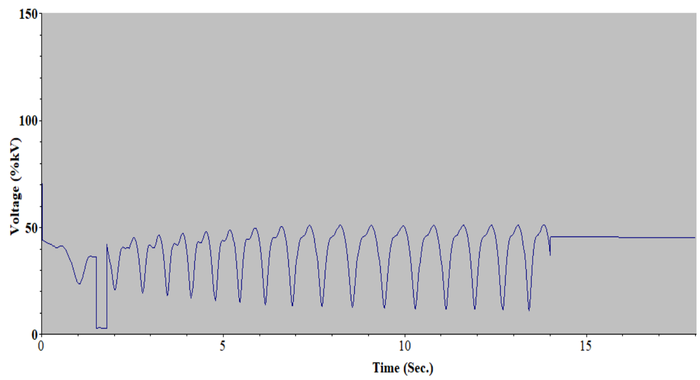
In this experiment, the renewable sources and transmission line were included, meaning that the schematic representation of the setup shown in fig. 4.2 was fully considered. However, while the HVAC line was connected, the HVDC line was disconnected. A fault was applied at the distribution line of the second load from bottom (Load 13). The fault was applied at 1.5 seconds and was cleared at 1.8 seconds.

4.5.1 Observations and Simulation Results

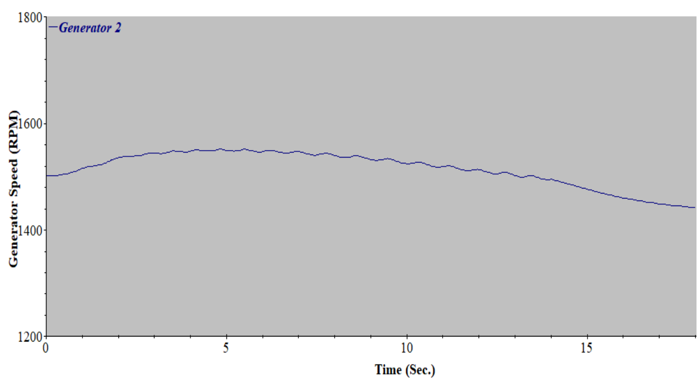
Fig. 4.4 shows the behavior of the three factors, including (a) the frequency oscillation for a period that included the fault time; (b) the substation bus voltage oscillation for a period that included the fault time, and (c) generator speed oscillation that resulted from the application of the fault (see fig. 4.4).



(a) Frequency oscillation during fault

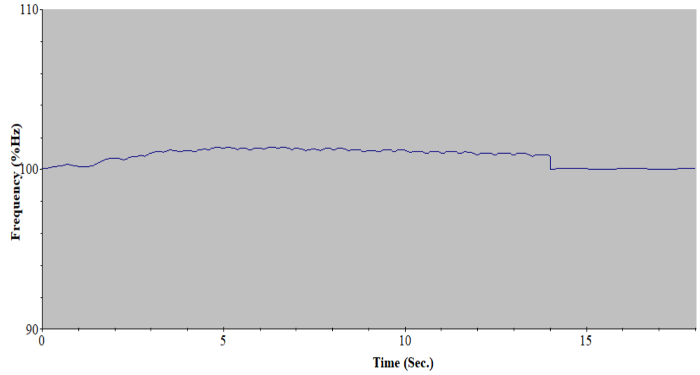


(b) Substation Bus voltage Oscillation

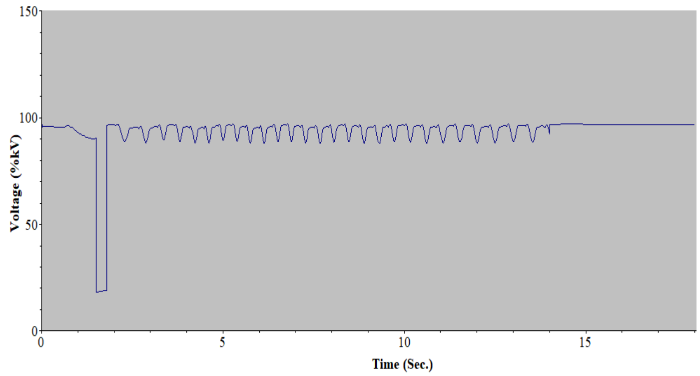


(c) Generator speed oscillation during the fault

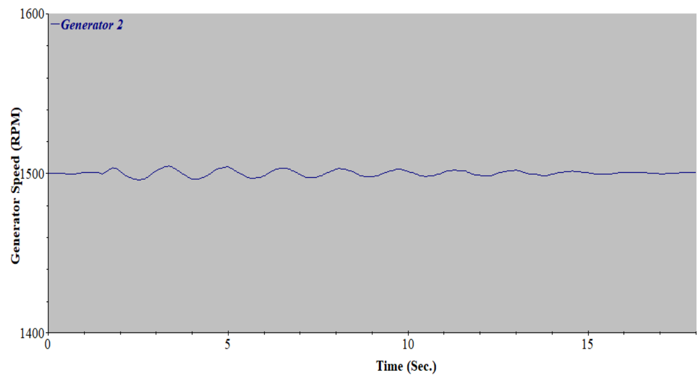
Fig. 4.3. Visualization Plots for Stability Analysis for the Case in Section 4.4 (Without Renewable Integration)



(a) Frequency Oscillation During the Fault



(b) Substation Voltage Bus Oscillation



(c) Generator Speed Oscillation During the Fault

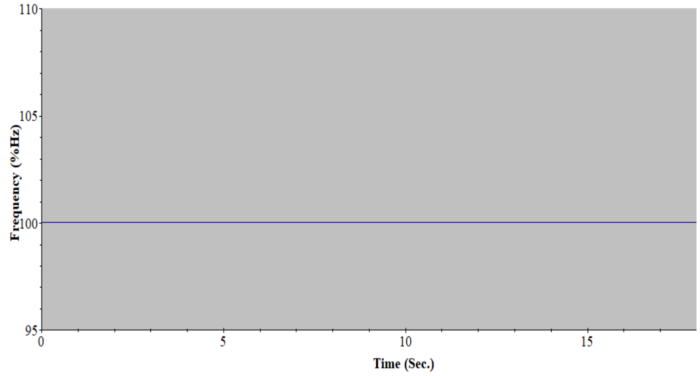
Fig. 4.4. Visualization Plots for Stability Analysis for the Case in Section 4.5 (Renewable Integration Through HVAC)

4.6 System Stability With Integrating Renewable Sources Through the HVDC Transmission System

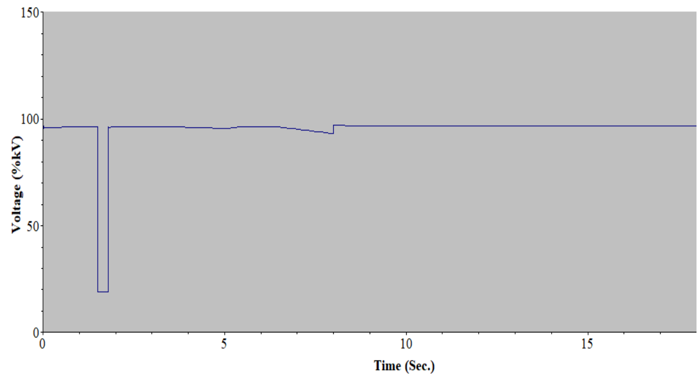
In this experiment, the renewable sources and transmission line were included, meaning that the schematic representation of the setup shown in fig. 4.2 was fully considered. However, while the HVDC line was connected, the HVAC line was disconnected. A fault was applied at the distribution line of the second load from bottom (Load 13). The fault was applied at 1.5 seconds and was cleared at 1.8 seconds.

4.6.1 Observations and Simulation Results

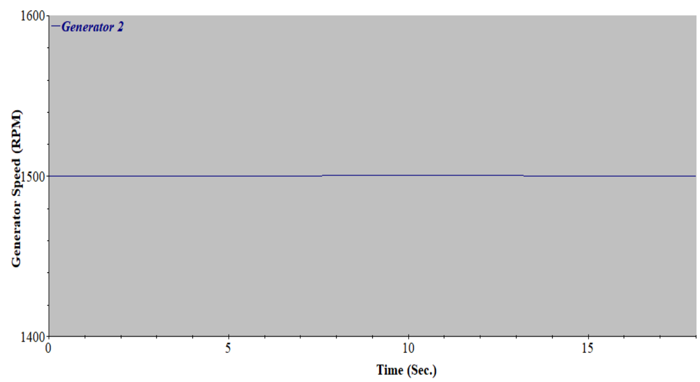
Fig. 4.5 shows the behavior of the three factors, including (a) the frequency oscillation for a period that included the fault time; (b) the substation bus voltage oscillation for a period that included the fault time; and (c) generator speed oscillation that resulted from the application of the fault (see fig. 4.5).



(a) Frequency Oscillation During the Fault



(b) Substation Bus Voltage Oscillation



(c) Generator Speed Oscillation During the Fault

Fig. 4.5. Visualization Plots for Stability Analysis for the Case in Section 4.6 (Renewable Integration Through HVDC)

4.7 Summary

To summarize, the visualizations results obtained in the three experiments outlined in Sections 4.4-4.6 are as follows.

In the first experiment described in Section 4.4 (see also fig. 4.3), frequency oscillated around the 100%, i.e. 50Hz in our case. The total simulation time was 18 seconds, and the frequency continued oscillating until the end of the simulation, which resulted in unstable frequency; even when the fault was cleared 10 seconds earlier, the oscillation was still observed. As can be seen in fig. 4.3(b), the bus voltage of the substation, which is a distribution bus, was predictably affected by the fault. However, the affection stayed for around 13 seconds after the fault clearance until it returned to its nominal value. Furthermore, as can be seen in fig. 4.3(c), one of the AC generators was chosen to test its speed (i.e. revolutions per minute, RPM), as a result of the fault. Ignoring the characteristic of the generator and inertia, the speed of the generator was tested to see how the affect would differ in the three experimental conditions. The result in fig. 4.2(c) shows an oscillation in the speed: specifically, it increased at the fault time and started to decrease after the fault was cleared. However, the speed ended up with a lower value than the initial one.

Furthermore, in the second experiment described in Section 4.5, the frequency oscillated around the 100%, i.e. 50Hz in our case. The total simulation time was 18 seconds, and the frequency continued oscillating 12.2 seconds after the fault was cleared, which predictably resulted in an unstable frequency for a very short period of time. As can be seen in fig. 4.4(b), the bus voltage of the substation was predictably affected by the fault. However, the impact was observed for around 14 seconds after the fault clearance until it returned to its nominal value. Finally, fig. 4.4(c) shows the results of testing the speed of one of the AC generators (same as for the first experiment, as discussed above, see also fig. 4.3(c)). The results show a smooth oscillation around the original value and, at the end, it returned to a very close value to the initial one.

Finally, in the third experiment described in Section 4.6, the voltage had a sharp drop at the fault period and after the fault clearance; voltage oscillated slightly around the 100% and reached about 100% at the end of the simulation. The frequency, as can be seen in fig. 4.5(a), the impact was not visible, and the frequency remained stable around 100% of the frequency value. Furthermore, fig. 4.5(c) shows the result of testing the speed of one of the AC generators (same as for the first and second experiments, see fig. 4.3(c) and fig. 4.4(c)). The result clearly shows that the speed value remained unchanged and that, at the end, the speed value was identical to the initial value before the fault. The generator speed also justifies the stability of the frequency.

In summary, in term of frequency oscillation, the third experimental condition (where renewable sources were integrated and the HVDC line was used) proved to be the most stable case, as frequency in this case were stable the whole period (see fig. 4.5(a)). Likewise, in terms of voltage stability, the third experimental condition again proved to be the most stable case, as where voltage stability persisted around 100% of the nominal voltage after the fault was cleared. This finding can be attributed to the fact that, as compared to HVAC lines, HVDC lines are characterized by a faster acting response and more transmitted power, so the voltage at the distribution bus could be fed faster than in the case when the HVAC line was used. Therefore, in the case of a quick fault (.3 second), using a HVDC line helps stabilizing the entire grid. The same holds true for the last case, where the third experimental condition again proved to be the most optimal, as the RPM was not affected at all, remaining constant the entire period (see fig 4.5(c)).

5. MACHINE LEARNING FOR FAULT DETECTION

Deriving from the importance of determining the fault location (see Section 2.2.3), in this chapter, we illustrate an artificial intelligence method to analyze and solve the problem. Recent studies have demonstrated the effectiveness of artificial intelligence in many fields, including but not limited to marketing, banking, power system, health care, and so forth. Among the well-known methods in artificial intelligence is machine learning.

5.1 What is Machine Learning?

Machine learning refers to the use of artificial intelligence that offers systems the capacity to robotically learn and advance from experience devoid of being overtly programmed. More specifically, machine learning focuses on the advancement of computer programs that can obtain data and use these data to learn in a self-reliant way [35]. The aim of machine learning is to comprehend the structure of the data and use them to construct models that can be comprehended and used by humans. While machine learning is a subdivision of computer science, it differs from conventional computational strategies. In conventional computing, a programmer sets specific algorithms of clearly programmed instructions used by computers to solve a problem. Instead, machine learning has algorithms that permit computers to learn from data inputs and to use statistical analysis to produce values within a particular range [36]. For that reason, machine learning enables computers to develop models from sample data and to make decisions based on the obtained data inputs.

In present-day world, machine learning has many and varied practical applications. For instance, machine learning is applied in the facial recognition technology used in social media sites to assist users in tagging themselves and their friends on photos. Moreover, the optical character recognition (OCR) system enables conversion of text images into movable types [36]. Furthermore, the machine learning technology is also used in navigation of self-driving automobiles to navigate in the roads. In fact, due to the changes that require higher efficacy and efficiency in manufacturing, custom execution of algorithms is normally needed for production systems. Firms are continually looking for systems that are faster, better, and require less effort to operate and have lower costs of production. Using machine learning tools helps businesses achieve higher revenue [37]. Importantly, executing algorithms help to enhance the skills necessary to find these solutions.

The machine learning technology is also used in electrical power systems–more specifically, in power transmission, generation, and maintenance. Accordingly, power firms widely use the statistical and discovery methods of machine learning for preemptive maintenance [38]. Among other applications, machine learning systems and methods are used to convert historical data from the electrical data into predictive models. Furthermore, machine learning can be used to generate transformer rankings, cable, feeder failure rankings, as well as to compute the mean time between failure estimations [39]. Machine learning also has interfaces for business management that allows for a directly incorporation of the prediction ability into decision support and corporate planning [36]. Machine learning is also beneficial in maintenance operations of power companies. Interestingly, it assists in fixing a problem proactively, instead of fixing an issue when it has already occurred. Said differently, machine learning makes it possible to prevent failures, rather than to cope with their consequences, such as cascading failures, fires and expensive emergency repairs.

A major requirement for a machine learning algorithm is data analysis. In fact, data analysis is the prerequisite for beginning a machine learning algorithm. Data analysis is a process of data collection, cleaning, aggregating, visualizing, and exploring. All these processes help in making appropriate predictions [39] and acquiring data from flat-files, spreadsheets, and databases, conducting exploratory data analysis (EDA), data reshaping, and data visualization. Furthermore, data exploration involves pursuing correlations, determining the missing content, and visualizing [36]. The building of models also includes visualization of the results, development of model diagnostics, and residual diagnostics. The machine learning algorithms can use the models to predict the future. Machine learning algorithms also require understanding of Python codes and R codes and how to operate them. To this end, Pandas library, which is useful for reshaping and aggregating the data, and Matplotlib library, which is important for data visualization, are frequently used. Similarly, Seaborn library can be used for advanced analytical processes [38]. Several basic data visualization techniques include bar charts, histograms, heat maps, and scatterplots. At this stage, the selection of the algorithm is implemented. A researcher should be specific in the selection of type and class of algorithm, as well as in the description of the system to execute. The next step of selecting a set of problems to validate and test the execution of the algorithm is available on [35]. Finally, the results of the performance of the built algorithm are evaluated based on several parameters, such as precision, F1-score, and recall (see Sections 5.4.1.1-5.4.4.1) for further detail.

5.2 Proposed Solution and Data Collection

From the definition of machine learning, it can be clearly seen that the main part of machine learning is preparing data, so that learning from data can be productive. Due to the availability of data nowadays, machine learning has become an attractive opportunity to many large companies, such as Amazon, Google, Apple, among others. In the present thesis, the fault detection problem was chosen for a machine learning tools to be performed on. The first step was to collect data of the system. Using DigSILENT PowerFactory simulation software, real-time data (RTD) feature was used to collect real-time data over the system. Fig. 5.1 shows a schematic representation of the HVDC system on the DigSILENT software.

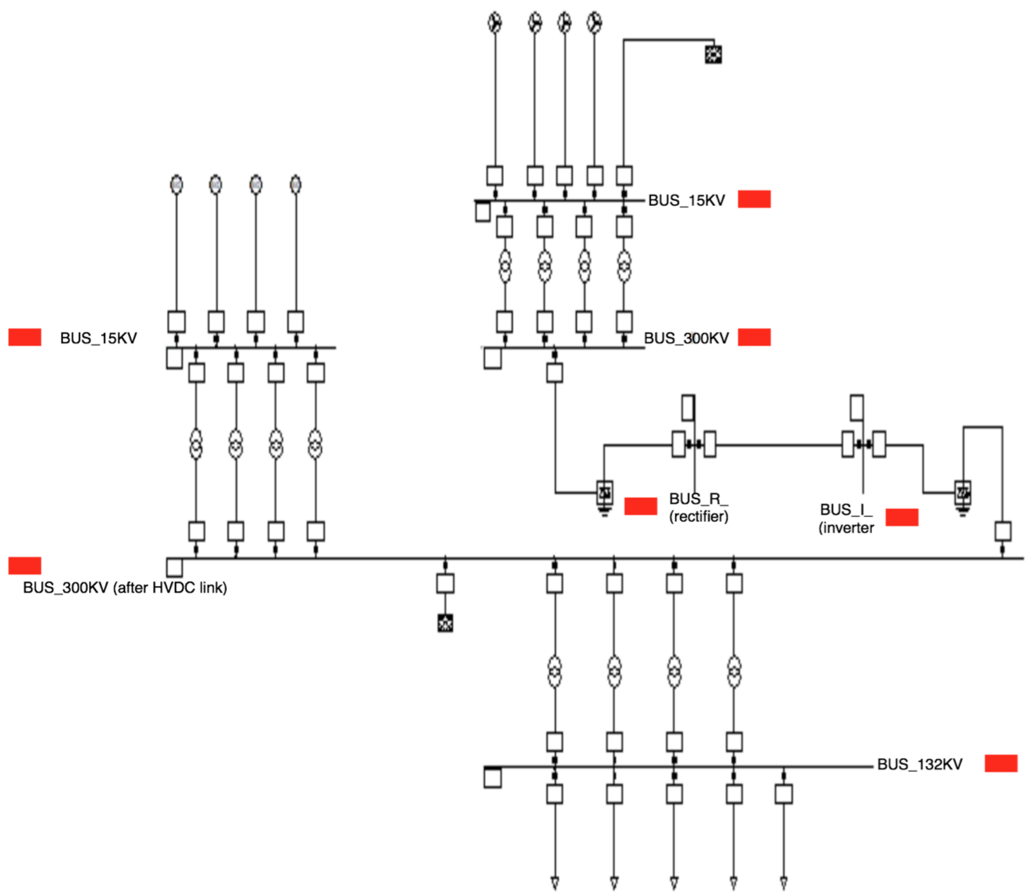


Fig. 5.1. The Schematic in DigSILENT PowerFactory for Data Collection

The problem at stake solved using machine learning, where the model was trained using existing datasets to predict the location of the fault for future cases. The proposed solution is that the model would be trained using supervised learning, which is basically a branch of machine learning that deals with pre-training the model using inputs with known output, thereby enabling the model to compute a mathematical function that can gradually learn to generalize on future unknown problems from further training on more data.

The data were collected on the HVDC line (see fig. 5.1), i.e. between the rectifier and inverter. Ten faults were applied to the HVDC line at different locations, starting from the rectifier bus and up to 99.99% of the HVDC link, which is almost at the inverter. Voltage measurements of 7 buses were taken with the step of 5 from 5% to 100% of the HVDC line length and at each fault location. The buses were the inverter bus, rectifier bus, two 300Kv buses, two 15Kv buses, and 132Kv bus (see fig. 5.1). The motivation to perform voltage measurements was to study the behavior of each bus in the event of a fault and make the model notice the observed behavior for future prediction. In machine learning terms, these seven buses were considered as features of the problem.

In all fault locations in the HVDC link, a fault was applied at 0.1 second and was cleared 0.1 second later, assuming a typical clearing time of 5 cycles at the 50Hz system frequency. Initially, 20 files of measurements were collected, where each file represented a fault location and its corresponding features. The data recording length was 6 seconds, from 0 second to 6 seconds with the step of 0.0002 to precisely measure the voltage behavior of the seven features. Consequently, the total number of data points obtained at the initial stage was around 60,000, with 3,000 data points for each fault location.

5.3 Data Analysis Stage

Data analysis is the first stage of machine learning before building a model. Feature reduction or, as it is frequently called, data cleaning is the most important part of the entire machine learning concept. The reason behind its importance is that this action studies all parameters and features of the dataset (buses voltages, in our case) Moreover, it will take the most effective features to determine the label to be considered in building the model. To find out the correlation between different features, Principal Component Analysis is generally used. Principal Component Analysis is basically a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated features into a set of values of linearly uncorrelated features called principal components [40]. However, in present thesis, different angle of data reduction is being seen.

5.3.1 Data Analysis Introduction

In the present thesis, several actions were tested before doing the elimination or data reduction. First, considering that, in reality, faults could happen anytime, different times of fault occurrences were tested. The results showed that, regardless of the time of occurrence of a fault, voltages behaviors remained stable for the same clearing time. Second, in all cases where the fault occurred at 0.3 or 0.4 or 0.2 second, the instances in the outer period of the fault remained the same. For example, if the fault happened at 0.3s and cleared at 0.4s, instances in the outer period, i.e. before 0.3s and after 0.4s, were equal to the nominal value of that particular bus.

Accordingly, based on the finding mentioned above, all instances before and after the fault occurrences were manually eliminated. Initially, the dataset was very redundant, as it had repetitive values and contained the values for instances with no faults. These values were removed, and the dataset containing only those instances when a fault was present were retained. Later on, we noticed many fluctuations even in those instances (e.g., such there were more than 1000 of such instances for each case). Further analysis showed that, in all of 20 cases, the 1000 data instances that covered the occurrence of the fault, there would come a point where the values would show very slight fluctuation. This was exactly what was needed to develop a good classifier, as, if all values were considered, too many outliers would appear and, hence, the classifier model would have performed really poorly.

Finally, after all steps of of manual reduction, further data analysis using Python was performed to find out if there was any variable which did not contribute to finding out the label/target determining the fault location.

5.3.2 Data Analysis Set-up and Results

In order to compare faulty locations and see whether they overlapped (i.e. exhibited identical behavior), graphs for all fault locations were made for all features (7 buses) shown in fig. 5.1. Overlapping features will be eliminated, as they did not really contribute much to determining the target.

Several Python libraries, such as Pandas, Seaborn, and Maplotlib, were used to draw graphs and analyze the dataset further. Our aim was to determine which features would show a weak correlation with the target class. Afterwards, if there is any weak correlation columns, those columns will be removed from the dataset.

Figures 5.2-5.8 summarize the results of testing for overlaps of the seven features.

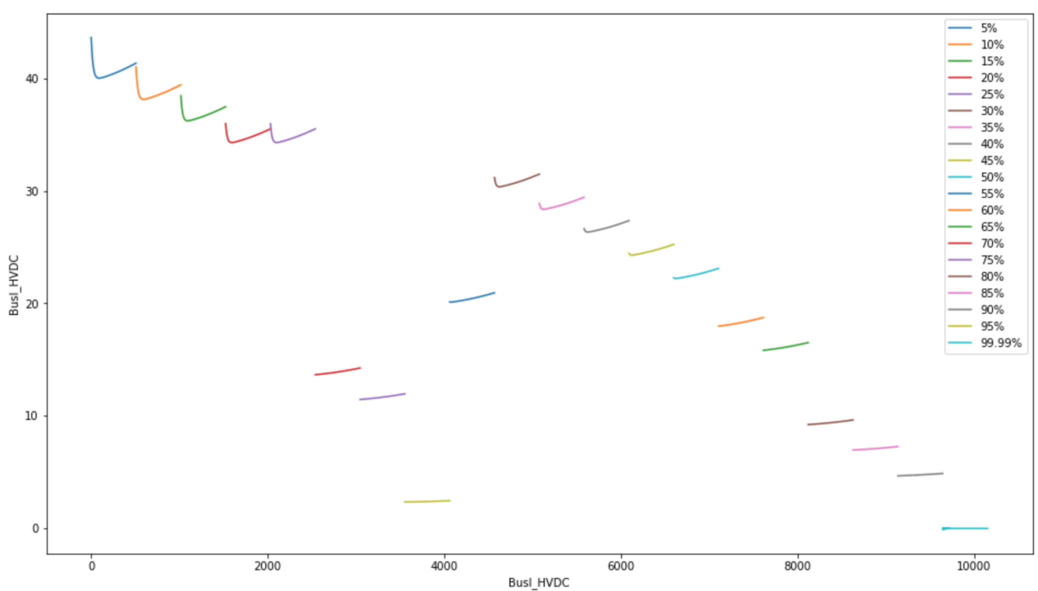


Fig. 5.2. Overlap Result for the Inverter Bus

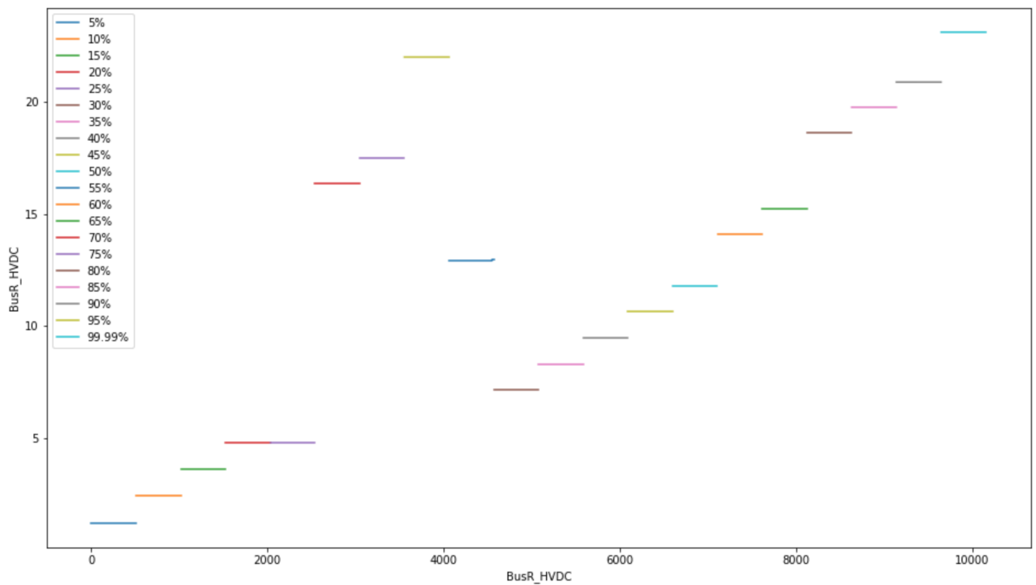


Fig. 5.3. Overlap Result for the Rectifier Bus

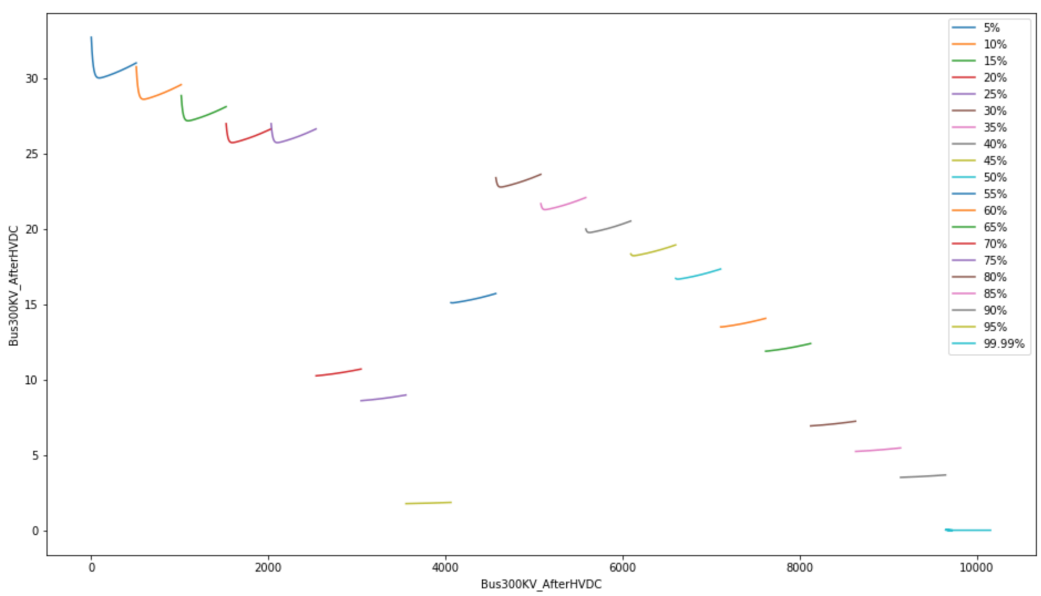


Fig. 5.4. Overlap Result for the 300Kv Bus After the HVDC Link

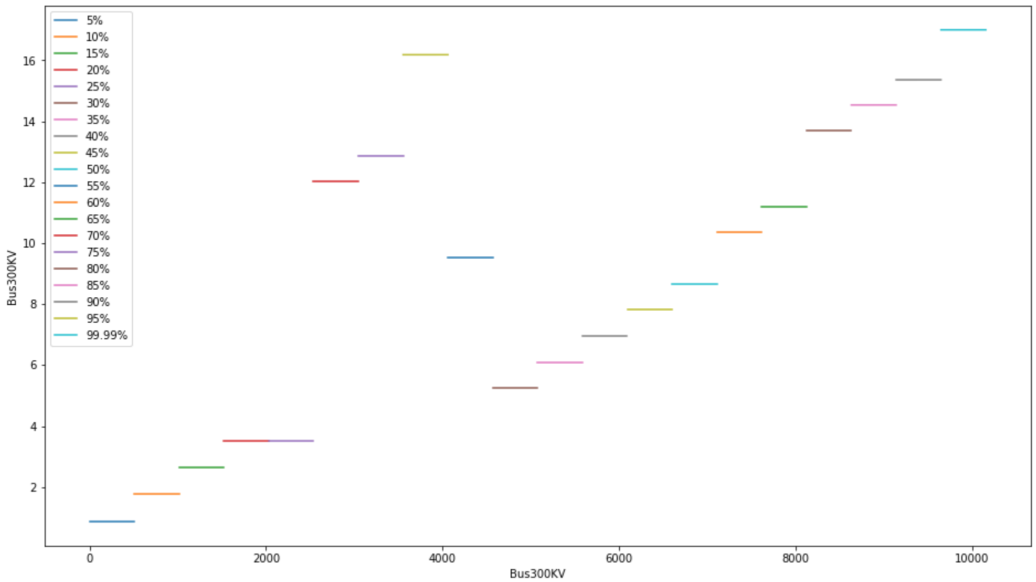


Fig. 5.5. Overlap Result for the 300Kv Bus

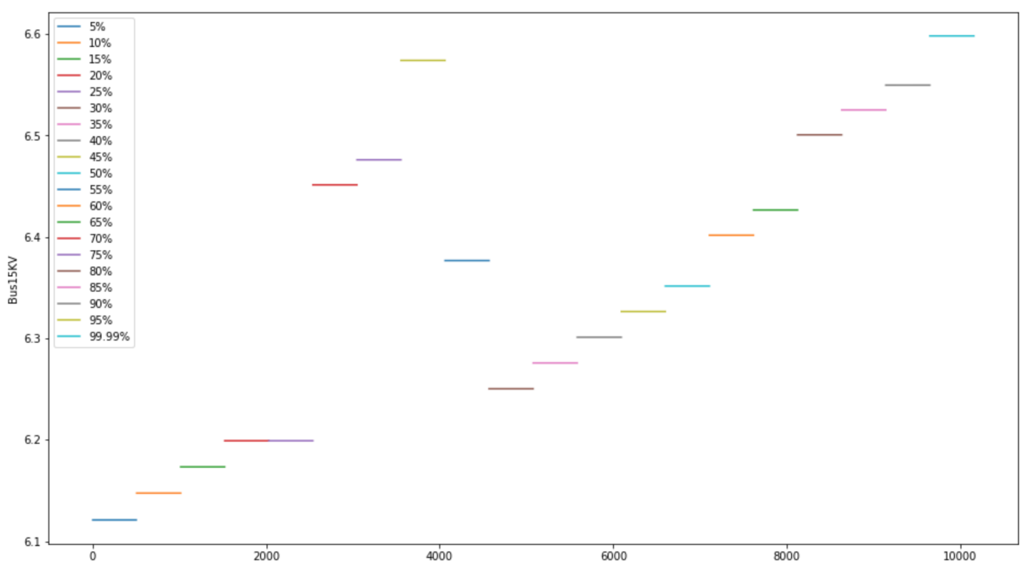


Fig. 5.6. Overlap Result for the 15Kv Bus_1

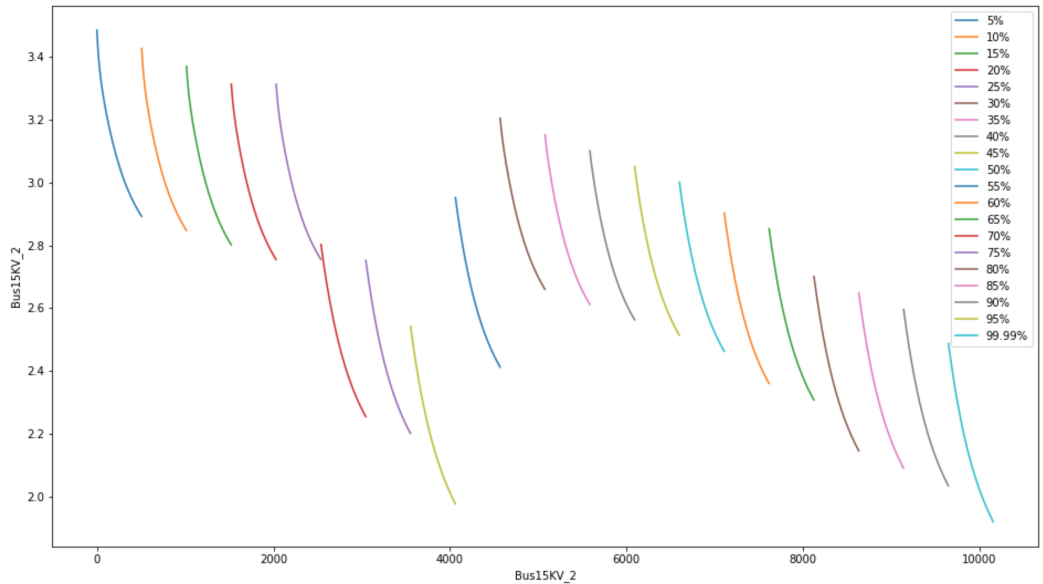


Fig. 5.7. Overlap Result for the 15Kv Bus_2

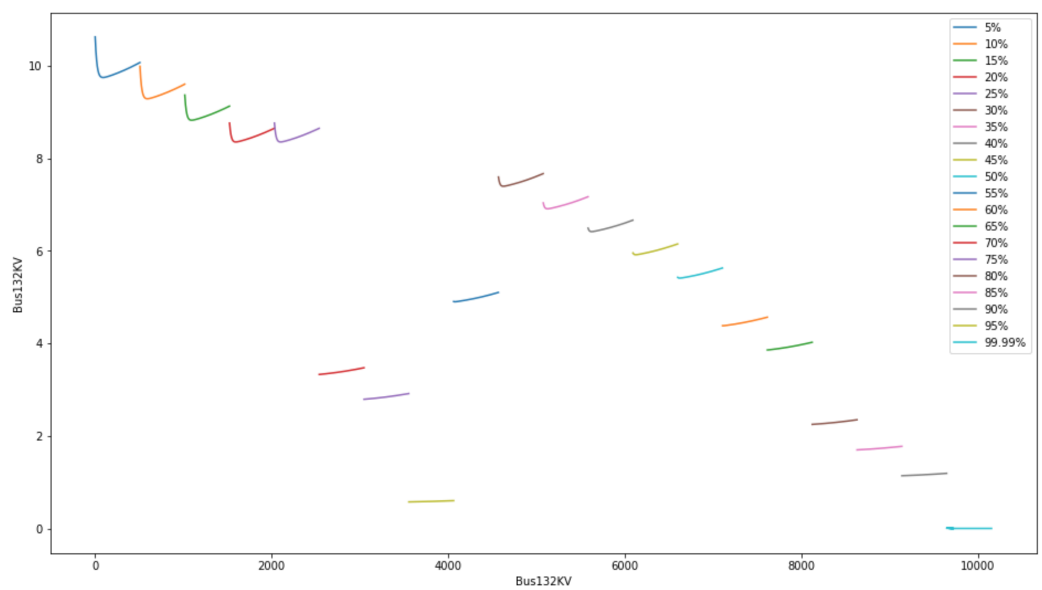


Fig. 5.8. Overlap Result for the 132Kv Bus

As can be seen in Figures 5.2-5.8, no overlap occurred in all features, which could be due to the data cleaning and instances reduction (see Section 5.3.1).

Finally, the correlation between the features was checked with the correlation matrix using the Seaborn library on Python. Using the Seaborn library, a heat map representing the contribution of each feature to finding the target label was created (see fig. 5.9).

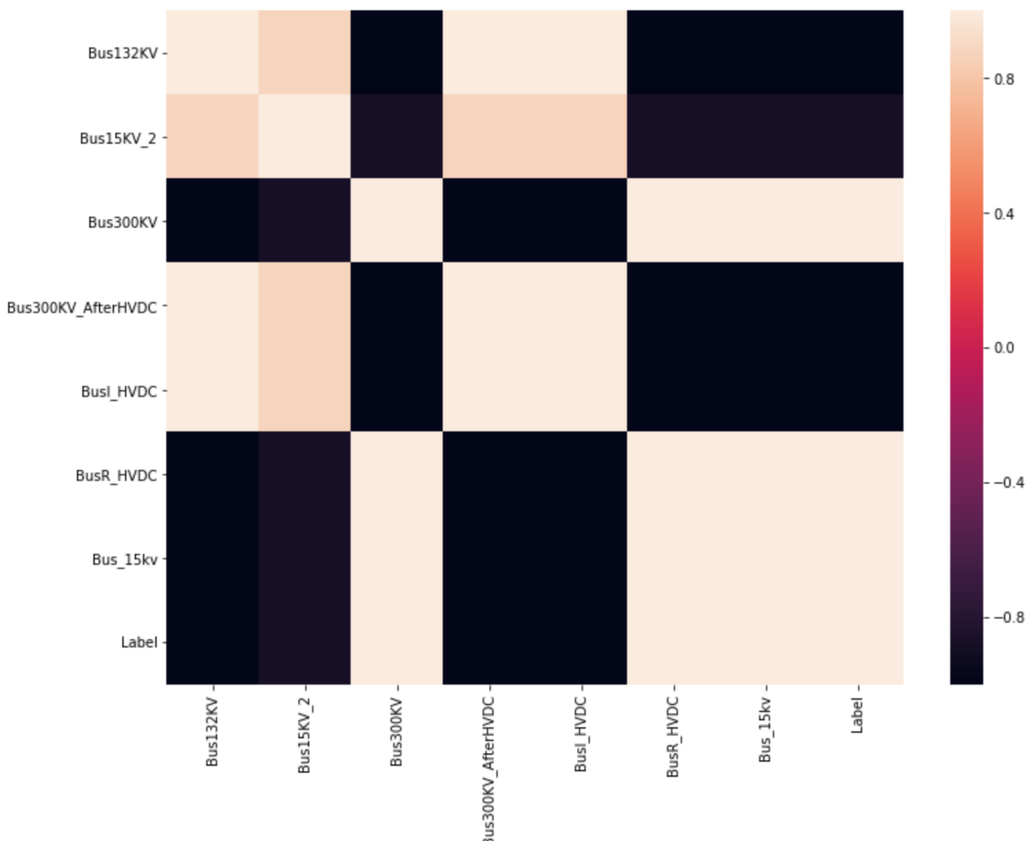


Fig. 5.9. Heatmap of the System Features

The heat map shows a clear visualization on the correlation between the features and the label. This kind of maps is particularly useful in the cases when it is necessary to eliminate several features for the training stage. In these cases, from the heat map, engineers can choose and eliminate the features that have dark spots with the label. In the present thesis, according to our heat map in fig. 5.9, the strongest features in determining the label were Bus 15Kv, BusR HVDC, and Bus 300Kv. On the other hand, the weakest features in determining the label were BusI HVDC, Bus300Kv afterHVDC, Bus15Kv 2, and Bus132Kv.

The data cleaning, processing, and analysis phase provides an insight into the ways to approach the problem at stake. At this point, all features were included in the next stage for redundant purpose in determining the target. This would conclude the data analysis stage that helps the training stage and makes the model more accurate and efficient.

5.4 Machine Learning Methods and Results

The proposed solution for the problem described in Section 5.3 is supervised learning in machine learning. Overall, there are several learning algorithms of this type, such as classification and regression. In the present thesis, the following four methods were applied to the prepared data: linear regression (LR), K-nearest neighbors (KNN), linear support vector machine (LSVM) , and adaptive boost (AdaBoost).

All methods and the mathematical concepts behind them are explained in detail in Sections 5.4.1-5.4.4. Next, each method was tested and its precision result was assessed. The test data were taken from the existing dataset. To this end, the dataset was split at the ration 70:30 ratio, with 70% used for training and 30% used for testing.

The performances of the built models were evaluated based on several factors. The results of the built models were based on the matrix called the confusion matrix, which is defined as a table that is frequently used to describe the performance of a classification model on a set of test data for which the true values are known [41]. To better understand the confusion matrix, the following four parameters have to be introduced: true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

True positive (TP) is when the model predicts a label for instances, and this label is present in the actual dataset.

True negative (TN) is when the model does not predict a label for instances, and the label is absent in the actual dataset.

False Positive (FP) is when the model predicts a label for instances, and this label is absent in the actual dataset.

False Negative(FN) is when the model does not predict a label for instances, and this label is present in the actual dataset.

Next, the following three new parameters were introduced for the real performance testing: precision, recall, and F1 score.

Mathmatically,

$$Precision = \frac{TP}{TP + FP}. \tag{5.1}$$

Precision is basically the ratio of true positive values over the total positive values. Precision shows how of all predicted instances were actually predicted correctly.

$$Recall = \frac{TP}{TP + FN}. \tag{5.2}$$

Recall, or sensitivity, is the ratio of true positive over true positive and false negative of a class. This parameter basically shows how many out of all instances in the dataset were labeled.

$$F1\_score = \frac{2 \times (Recall \times Precision)}{(Recall + Precision)}. \tag{5.3}$$

Another important term relevant for the present study is the cost function. The cost function measures how close the predicted values match the actual real values, which is important at the data training stage.

For each classifier model, Eq. (5.1)-(5.3) were used to calculate respective parameters of each model.

Finally, the targets/classes of the problem were numbered from 0 to 19, instead of location percentages, where 0 represents 5% and 19 represents 100% of the HVDC line length, with the step of 5%. Tables 5.1-5.4 summarize the results of each classifier.

5.4.1 Linear Regression

Linear regression is a relatively old and straightforward supervised machine learning algorithm that helps to a linear relationship between the input and the output of the problem, i.e. the instances and the target class.

A linear regression representation/relationship can be expressed as follows (see Eq. (5.4)):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n. \tag{5.4}$$

Where y is the target class, the x's are the input features, β_0 is the intercept, and β_1 to β_n are regression coefficients. During training, the goal is to find coefficients which minimize the cost function. To this end, gradient descent equation, an optimization algorithm to minimize the cost, was used (see Eq. (5.5)).

$$minimize \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2. \tag{5.5}$$

Initially, all coefficients are set at 0 and then gradually increase after each iteration to reduce the cost function. Therefore, linear regression helps to find the coefficient values for each feature that gives a good accuracy.

5.4.1.1 Linear Regression Results

Table 5.1 shows the performance of the linear regression classifier. Table 5.1 consists of 5 columns and 21 rows. The last row represents the average of each column individually and the total of the last column, which is the support column. Further analysis of the result is discussed in Section 5.5.

Table 5.1. Linear Regression Classification Result

Classification report				
Target	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	154
1	1.00	1.00	1.00	148
2	1.00	1.00	1.00	151
3	0.00	0.00	0.00	156
4	0.51	1.00	0.67	161
5	1.00	1.00	1.00	150
6	1.00	1.00	1.00	142
7	1.00	1.00	1.00	152
8	1.00	1.00	1.00	161
9	0.65	1.00	0.79	146
10	1.00	0.49	0.65	150
11	1.00	1.00	1.00	153
12	0.97	1.00	0.99	148
13	1.00	0.97	0.99	150
14	1.00	1.00	1.00	165
15	1.00	1.00	1.00	156
16	1.00	1.00	1.00	155
17	1.00	1.00	1.00	160
18	1.00	1.00	1.00	142
19	1.00	1.00	1.00	147
Avg/Total	0.90	0.92	0.90	3047

5.4.2 K-Nearest Neighbor

K-nearest neighbors (KNN) is one of the simplest and fastest classifications and regression algorithms; however, in our case, it was used only for classification. More specifically, KNN has three advantages that make it one of the first choices before considering any complex machine learning algorithms for a classification problem:

1. Ease of interpretation of the output
2. High speed of training and prediction
3. Strong predictive power

As suggested by its name, KNN works, by taking a vote from K-nearest neighbors of a data instance for which the model trying to find its actual class. It makes a circle that covers all K points from which a vote is needed. After taking a vote from these data instances, it can be definitely concluded that the data instance that is being considered belongs to class X. To compare with the nearest neighbors, KNN uses a relatively simple formula of distance. Some of the mostly commonly used equations are provided below (see Eq. (5.6)-(5.7)):

- Euclidian Distance

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{k} (x_i(k) - x_j(k))^2}. \tag{5.6}$$

- Manhattan Distance

$$d(x_i, x_j) = \sum_{k=1}^{k} \mid (x_i(k) - y_j(k) \mid . \tag{5.7}$$

To relate Eq. (5.6) and (5.7) with the problem case, the initial data analysis stage showed that, for each particular fault location, the instance values were in a certain range and showed a different behavior, making some sort of a cluster; therefore, the easiest way to predict a test data instance would be by finding its neighbors using one of the distance formulas shown in Eq.(5.6) and Eq. (5.7), which would allow us to find out which cluster it belongs to and thus would most likely be from that class as well.

Furthermore, the value of K has to be decided. For the value consideration, two things have to be considered:

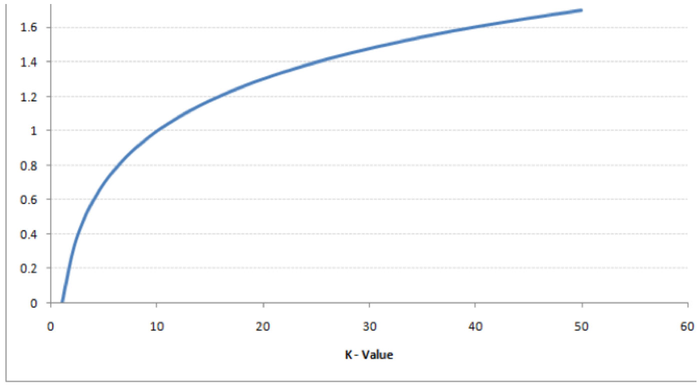
(a) Training error rate.

(b) Validation error rate.

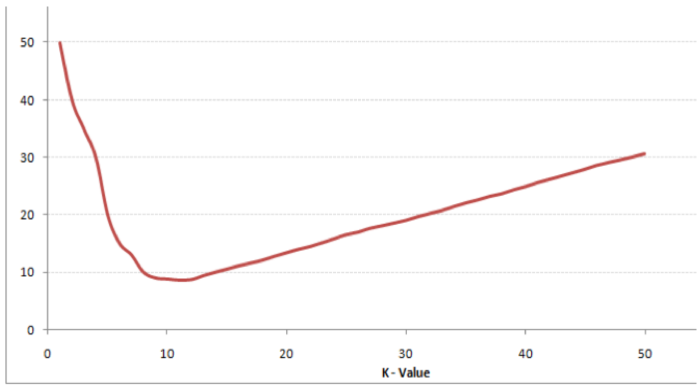
The training error rate is always zero at K = 1 (as the nearest point to a data instance is always that data instance itself), and it increases with an increase of the value of K. Fig. 5.10(a) shows a graph which illustrates this statement.

Validation error rate is the reason why data scientists do not always go with the value of K = 1. It is because it shows a different behavior; specifically, it decreases in the beginning and, on reaching a minimum point, its error rate starts to increase as well (see fig. 5.10(b) for an illustration).

Said differently, if K is chosen to be equal to 1, the model would end up overfitted. Therefore, the goal here is to find a value of K at which the validation error reaches its minima.



(a) Training Error Rate



(b) Validation Error Rate

Fig. 5.10. K-value Consideration Factors [42]

5.4.2.1 *KNN Result*

Table 5.2 shows the performance of the KNN classifier. Table 5.2 consists of 5 columns and 21 rows. The last row represents the average of each column individually and the total of the last column, which is the support column. Further analysis of the result is discussed in Section 5.5.

Table 5.2. KNN Classification Result

Classification report				
Target	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	91
1	1.00	1.00	1.00	94
2	1.00	1.00	1.00	85
3	0.21	0.23	0.22	106
4	0.28	0.26	0.27	123
5	1.00	1.00	1.00	105
6	1.00	1.00	1.00	108
7	1.00	1.00	1.00	113
8	1.00	1.00	1.00	106
9	1.00	1.00	1.00	118
10	1.00	1.00	1.00	108
11	1.00	1.00	1.00	90
12	1.00	1.00	1.00	96
13	1.00	1.00	1.00	105
14	1.00	1.00	1.00	84
15	1.00	1.00	1.00	98
16	1.00	1.00	1.00	105
17	1.00	1.00	1.00	106
18	1.00	1.00	1.00	99
19	1.00	1.00	1.00	92
Avg/Total	0.92	0.91	0.91	2032

5.4.3 Linear Support Vector Machine

Linear support vector machine (LSVM) is the fastest machine learning algorithm for multiclass classification problems, like the one addressed in the present study; LSVM is particularly useful for large datasets, as it creates a model which scales linearly with the size of the training dataset. The dataset used in the present study was not very large, as it was reduced. However, we decided to test the problem on a small scale first and test it to a higher level. Moreover, since the expected dataset was in millions, linear SVM was the perfect choice for its ability to deal with large datasets with a linear increase in computation power required, as well as the fact that the data in this problem display a clear pattern. Finally, at this point, the problem setup is using less than 10 features to predict the target class. However, in the future, predictions would need to be based on a larger number of features, and linear SVM can work with higher dimensional data with thousands of features and attributes in both sparse and dense format; therefore applying linear SVM in further research would offer many advantages in terms of scalability. To summarize, linear SVM is not only fitting the size of the current dataset, but can also be expected to perform well on a much larger dataset from the same domain.

LSVM seeks to draw a margin line between class instances. Similarly to KNNs, the current dataset was divided into clusters; therefore, what LSVM would do is draw lines to separate these classes/clusters from each other as accurately as possible, so that when user inputs a test data instance, it would fall in one of the 20 classes in our dataset, and the prediction would be very simple. A support vector is basically a frontier that best segregates the classes (see fig. 5.12 for an illustration of the concept for two classes).

Support Vector Machine works by trying to minimize the error function given in

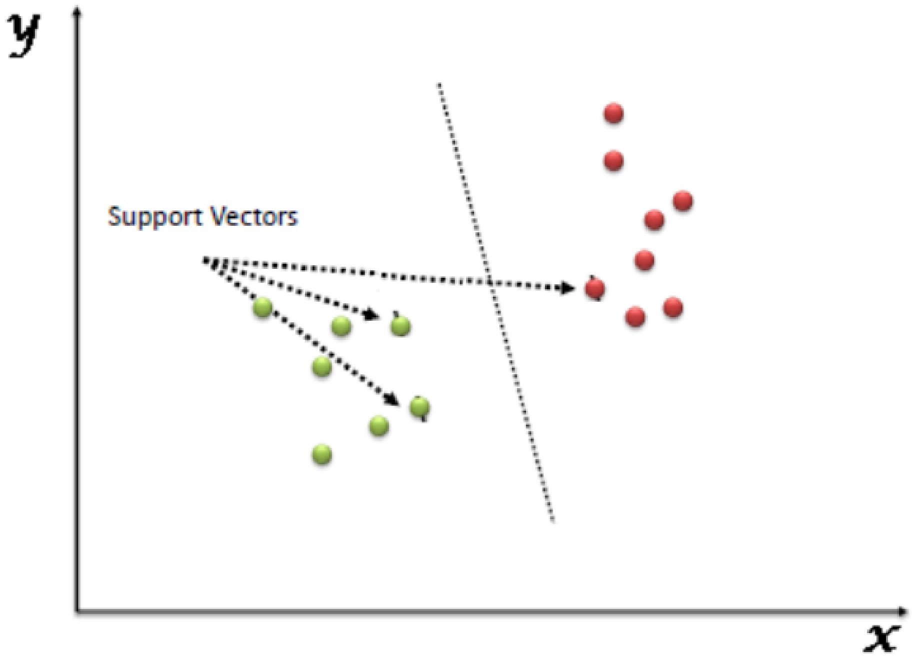


Fig. 5.11. Support Vector Machine Mechanism [43]

Eq. (5.8).

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i. \tag{5.8}$$

subject to the following constraint (see Eq. (5.9)):

$$y_i(w^T\emptyset(x_i)+b) \geq 1-\xi_i \ and \ \xi_i \geq 0,\ i=1,2,...,N. \tag{5.9}$$

"Where C is the capacity constant, w is the vector of coefficients, b is a constant, and ξ_i represents parameters for handling non-separable data (inputs). The index i labels the N training cases. Note that "$y \in \pm 1$"represents the class labels and x_i represents the independent features. The kernel $\emptyset$ is used to transform data from the input (independent) to the feature space. It should be noted that the larger the C, the more the error is penalized. Thus, C should be chosen with care to avoid over fitting"[44].

5.4.3.1 Linear Support Vector Machine Result

Table 5.3 shows the performance of the LSVM classifier. Table 5.3 consists of 5 columns and 21 rows. The last row represents the average of each column individually and the total of the last column, which is the support column. Further analysis of the result is discussed in Section 5.5.

Table 5.3. Linear SVM Classification Result

Classification report				
Target	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	91
1	1.00	1.00	1.00	94
2	1.00	1.00	1.00	85
3	0.46	1.00	0.63	106
4	0.00	0.00	0.00	123
5	1.00	1.00	1.00	105
6	1.00	1.00	1.00	108
7	1.00	1.00	1.00	113
8	1.00	1.00	1.00	106
9	1.00	1.00	1.00	118
10	1.00	1.00	1.00	108
11	1.00	1.00	1.00	90
12	1.00	1.00	1.00	96
13	1.00	1.00	1.00	105
14	1.00	1.00	1.00	84
15	1.00	1.00	1.00	98
16	1.00	1.00	1.00	105
17	1.00	1.00	1.00	106
18	1.00	1.00	1.00	99
19	1.00	1.00	1.00	92
Avg/Total	0.91	0.94	0.92	2032

5.4.4 AdaBoost

AdaBoost is short for adaptive boosting, which is a sequential ensemble method in machine learning [45]. AdaBoost combines multiple models to improve the final predictive performance. The term boosting here means combining many weak learners to create an accurate prediction; weak learners would be the classifiers that do slightly better than random guessing, i.e. have the prediction accuracy $\gg$50%. This step is performed in a sequential manner; the first classifier performs unsatisfactorily, and the second one tries to correct the errors in the first one and tries to predict harder to classify examples in the training data and so on, until accuracy gets reasonably highthen, the number of iterations is over. Each of the instance in the training dataset is weighted. The initial weight is set as follows (see Eq. (5.10)):

$$Weight(x_i) = \frac{1}{n}. \tag{5.10}$$

Where x_i is the ith training instance, and n is the number of training instances. The misclassification rate is calculated for the trained model. Typically, it is calculated as shown in Eq. (5.11).

$$Error = \frac{correct - N}{N}. \tag{5.11}$$

Where $error$ is the misclassification rate, $Correct$ is the number of training instances correctly predicted by the model, while N is the total number of training instances[46].

5.4.4.1 AdaBoost Result

Table 5.4 shows the performance of the AdaBoost classifier. Table 5.4 consists of 5 columns and 21 rows. The last row represents the average of each column individually and the total of the last column, which is the support column. Further analysis of the result is discussed in Section 5.5.

Table 5.4. AdaBoost Classification Result

Classification report				
Target	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	135
1	1.00	1.00	1.00	140
2	1.00	1.00	1.00	107
3	0.31	1.00	0.47	104
4	0.00	0.00	0.00	121
5	0.00	0.00	0.00	132
6	0.00	0.00	0.00	116
7	1.00	1.00	1.00	127
8	0.00	0.00	0.00	155
9	0.00	0.00	0.00	109
10	0.00	0.00	0.00	133
11	0.00	0.00	0.00	118
12	0.00	0.00	0.00	131
13	1.00	1.00	1.00	126
14	0.00	0.00	0.00	132
15	0.08	1.00	0.15	117
16	0.00	0.00	0.00	124
17	0.00	0.00	0.00	135
18	1.00	1.00	1.00	124
19	1.00	1.00	1.00	151
Avg/Total	0.33	0.40	0.34	2539

5.5 Summary

In summary, in this chapter, we used four classifiers to obtain results. The results showed that not all classifiers performed well on the problem, as each classifier has its own way of training on the data and subsequent testing. Therefore, based on the data structure and the behaviors of the data features, the following conclusions can be made.

Firstly, linear regression classifier performed in a good way: specifically, it yielded precision of 90%, recall of 92%, and F1 score of 90%. The reason why linear regression performed well on the problem addressed in the present thesis is that, in principle, this method seeks to find a linear relationship among all the features the problem has and the target class. Once it finds such linear relationship, it can easily do the rest of building the right algorithm of the model and then produce a model characterized by high accuracy.

Secondly, KNN has performed slightly better than the linear regression, as its precision, recall, and F1 score amounted to 92%, 91%, and 91%, respectively. The reason why KNN yielded results similar to those of linear regression is that KNN is also a regression algorithm (see Section 5.4.1). However, KNN has a different role, which yields the slightly different results by taking a vote from data instances after circling the nearest neighbors of instances.

Thirdly, LSVM also performed slightly better than the KNN classifier, when taking the average of the three columns in Table 5.2, and its precision, recall, and F1 score amounted to 91%, 94%, and 92%, respectively. The reason behind the similarity of results between KNN and LSVM is that LSVM behaves similarly to KNN in terms of separating instances, where LSVM also draws a line between clusters and separate them from each other. In the results reported in Tables 5.2 and 5.3, a considerable similarity of the results in each class can be observed. In the present thesis, it would have 20 clusters/classes and, when testing, the testing data would fall into one of those 20 classes to make the prediction.

Fourthly, AdaBoost demonstrated the worst performance, with precision of 33%, recall of 40%, and F1 score of 34%. In our case, applying AdaBoost has several limitations, as AdaBoost requires several conditions that need to be fulfilled, and, in our case, these conditions were not fully met. First, AdaBoost should be provided with a quality dataset, as it attempts to sequentially classify the misclassified instances, improving after each iteration (see Section 5.4). However, not only our dataset was insufficient for this algorithm, but also fluctuations between data for different classes were insignificant, which made it hard for AdaBoost to perform well. Second, AdaBoost should not have outliers–otherwise, the classifier would spend a significant amount of time trying to correct these cases, which makes the task almost infeasible. Yet, in the problem addressed in the present study, outlier values are of importance, as equal data instances for each class, in the same time range were selected to ensure consistency in the data. Therefore, it would be unreasonable to trade off consistency and pattern data for good accuracy in this algorithm. Therefore, the data values for each instance of a class had a wide range of up to 10, so AdaBoost predictable did not work very well on them.

6. CONCLUSION AND RECOMMENDATION

6.1 Conclusion

Performance of a transmission line can be taken into a higher level. The process of testing the two systems in Chapter 4 can be easily scaled up to testing a real system. In terms of efficiency, stability, and fault analysis, our results demonstrated that the high voltage direct current (HVDC) transmission system shows a better performance than the high voltage alternating current (HVAC) system. However, disadvantages of HVDC, such as expensive converter stations, expensive filters, and challenges of integrating HVDC with existing AC systems make considering the HVDC system to be unlikely in many cases. Therefore, a long-term vision is needed in this case.

System losses and voltage drop in the HVDC system are significantly lower than those in the HVAC system. Specifically, for the 180Km transmission line, losses in HVAC (2.91MW) were 180% higher than losses in HVDC (1.61MW). Moreover, voltage drop across the HVAC line was about 2.5 times higher than that across the HVDC line (2.58% vs. 1.01%, respectively). Moreover, in terms of system stability, HVDC proved to be much better than HVAC for integrating renewable sources. In the event of a fault, HVDC helps to stabilize the connected AC system generators, voltage buses, and frequency deviation (see Section 4.6).

Machine learning gives the HVDC system another vote over the other system. When HVDC is in service, machine learning can be meaningfully used to determine location of a fault. Armed with this information, an operator can take quick action as to prevent further damages or blackout. When a major problem occurs, making use of the power system data for prediction instead of using it for mere monitoring purposes enables taking many preventive actions.

Finally, Linear support vector machine (LSVM) showed the best performance among the three tested classifiers. Taken together, our results suggest that not every classifier can perform well on a particular problem, as was the case of AdaBoost in the present thesis. Therefore, a careful selection of a classifier after a detailed analysis of a problem can optimize the way of solving a machine learning problem. Moreover, in real case problem, predictions would need to be based on a larger number of features, and linear SVM can work with higher dimensional data with thousands of features and attributes in both sparse and dense format; therefore applying linear SVM in further research would offer many advantages in terms of scalability.

6.2 Recommendation

In the present thesis, we aimed to propose the best electrical system solution for the Kuwait 2035 project. At present, Kuwait does not have a single HVDC system in its interior power system yet, and there can be some reluctance to build such new system. However, there are several arguments in support of building the HVDC system particularly in Kuwait.

The Kuwait 2035 project aims to produce a huge amount of renewable energy and to develop five islands–and, accordingly, to invest in them by providing smart, reliable, efficient, and cost-effective power in the long term. Through an HVDC link integrating a massive amount of renewable sources in a remote area, these five islands, along with other AC systems connected, can be reliably fed. According to a recent estimate by the government, the economic turnover of the five islands project would reach \$2.2 trillion/year [47]. Therefore, in this case, building an expensive system that is cost-effective in the long term would be efficient. Therefore, supporting the islands requires undersea cables. In this respect, the HVDC transmission cables were proven to be better in transmission due to the high capacitance in AC cables that causes additional losses, as shown by our simulation results.

At present, Kuwait is part of the Gulf Countries Council (GCC) interconnected electrical network. This network includes Qatar, Bahrain, and Saudi Arabia. The overarching goal of this interconnection is to achieve electrical security and stability among GCC members by offering shared spinning reserve in case of emergencies [48]. Moreover, the project established a commercial energy market that benefits GCC countries by selling power. Consequently, the connection between Kuwait and Saudi Arabia is through the back-to-back HVDC system, as the two countries have two different operating frequencies–50 Hz and 60 Hz, respectively. Furthermore, the back-to-back HVDC system connection is between the South of Kuwait and the East of Saudi Arabia. Therefore, building an HVDC system in the West of Kuwait for integrating renewable sources will offer a valuable possibility of having another interconnection to the North of Saudi Arabia, which would increase stability in the event of emergencies either by importing or exporting power. Secondly, having more interconnection will in all probability have a strong positive impact on the commercial energy market through selling energy to the north side of Saudi Arabia instead of restricting to its Eastern side. Moreover, it would give the government an option to connect even to Iraq, which is located behind the renewable farms across the border.

In conclusion, the results of the present thesis convincingly demonstrate that installing the HVDC system can be strongly recommended for the Kuwait electric system, particularly when we consider the 2035 country's project on renewable sources and the five islands project. Considering the HVDC system will not only save money, but will also open up many opportunities for making profit, particularly through the interconnection option and through exchanging power with the asynchronous electric systems of the neighboring countries.

References

[1] Yearly Statistical Book. Ministry of Electricity and Water, Kuwait, Dec. 2017, www.mew.gov.kw/Files/AboutUs/Statistics/42/Ref713.pdf.

[2] Kuwait Institute for Scientific Research (2nd SE4All Forum Commitments) - United Nations Partnerships for SDGs Platform. United Nations, United Nations, sustainabledevelopment.un.org/partnership/?p=10396.

[3] Advantages of High Transmission Voltage. EEEGUIDE, 22 May 2018, www.eeeguide.com/advantages-of-high-transmission-voltage/.

[4] Meier, Alexandra von. Electric Power Systems: a Conceptual Introduction. Wiley, 2006.

[5] Jovcic, Dragan, and Khaled Ahmed. High-Voltage Direct-Current Transmission: Converters, Systems and DC Grids. Wiley, 2015.

[6] High Voltage DC Transmission System. Electronics Hub, 29 Oct. 2015, www.electronicshub.org/high-voltage-dc-transmission-system/.

[7] Boeck, Steven De, et al. Configurations and Earthing of HVDC Grids. 2013 IEEE Power & Energy Society General Meeting, 2013, doi:10.1109/pesmg.2013.6672808.

[8] Schon, Andre, and Mark-M. Bakran. High Power HVDC-DC Converters for the Interconnection of HVDC Lines with Different Line Topologies. 2014 International Power Electronics Conference (IPEC-Hiroshima 2014 - ECCE ASIA), 2014, doi:10.1109/ipec.2014.6870154.

[9] Chaves, Miguel, et al. HVDC Transmission Systems: Bipolar Back-to-Back Diode Clamped Multilevel Converter with Fast Optimum-Predictive Control and Capacitor Balancing Strategy. Electric Power Systems Research, vol. 81, no. 7, July 2011, pp. 14361445., doi:10.1016/j.epsr.2011.02.008.

[10] Gamit, Pallavi, et al. Comparative Analysis of HVDC and EHVAC. International Research Journal of Engineering and Technology (IRJET), vol. 02, no. 09, Dec. 2015, pp. 11651171.

[11] Konstantinou, Georgios S., et al. Analysis of Multi-Carrier PWM Methods for Back-to-Back HVDC Systems Based on Modular Multilevel Converters. IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society, Nov. 2011, pp. 43914396., doi:10.1109/iecon.2011.6120031.

[12] Liang, Jun, et al. A Multi-Terminal HVDC Transmission System for Offshore Wind Farms with Induction Generators. International Journal of Electrical Power & Energy Systems, vol. 43, no. 1, Dec. 2012, pp. 5462., doi:10.1016/j.ijepes.2012.04.063.

[13] J. Duncan, Glover, et al. Power System Analysis & Design. SI Version, Cengage Learning, 2012.

[14] Leonard A. Bateman, R.W H aywood, " Nelson River DC Transmission Project", IEEE Transactions on Power Apparatus and Systems, Vol. PAS- 88, No. 5, pp. 688-693, May 1969.

[15] R.M. Morris, A.R. Morse, J.P. Griffin, O.C. Norris-Elye, C.V. Thio, and J.S. Goodman, The Corona and Radio Interference of the Nelson River HVDC Transmission Lines, IEEE Transactions on Power Apparatus and Systems, Vol. PAS-98, No. 6, pp. 1924-1936, Nov.-Dec., 1979.

[16] Nelson River HVDC System. IEEE: Power & Energy Society, 2014.

[17] Meah, Kala, and Sadrul Ula. Comparative Evaluation of HVDC and HVAC Transmission Systems. 2007 IEEE Power Engineering Society General Meeting, July 2007, doi:10.1109/pes.2007.385993.

[18] Willis F. long, Editor "Comparison of costs and benefits for DC and AC transmission", OAK Ridge national laboratory, ORNL-6204, Feb 1987.

[19] J.P Bowles, V. Burtnyk, C.C Diemond, M.A. Lebow, E.G. Neudr of, D. Povh, E.C. Starr, C.W. Taylor, R.A. Walling, edited by C.T. WU " AC- DC economics and alternatives- 1987 panel session report", IEEE transactions on power delivery, Vol.5, No.4, pp.1241-1248, November 1990

[20] Velasco, D., et al. Power Transmission in Direct Current. Future Expectations for Colombia. Renewable and Sustainable Energy Reviews, vol. 15, no. 1, Aug. 2011, pp. 759765., doi:10.1016/j.rser.2010.08.013.

[21] YuanzeZhang, et al. Review of Modular Multilevel Converter Based Multi-Terminal HVDC Systems for Offshore Wind Power Transmission. Renewable and Sustainable Energy Reviews, vol. 61, Aug. 2016, pp. 572586.

[22] Kjrholt, sa M Halvorsdatter. HVDC Transmission Using a Bipolar Configuration Composed of an LCC and MMC. Norwegian University of Science and Technology, 2014.

[23] Karady, George G., and Keith E. Holbert. Electrical Energy Conversion and Transport: an Interactive Computer-Based Approach. IEEE Press, 2013.

[24] Biabani, Mohammed Abdul Khader Aziz, et al. Power System Stability Enhancement Using Static Synchronous Series Compensator. 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Oct. 2016, pp. 462481., doi:10.1109/scopes.2016.7955875.

[25] Mynuddin, Mohammed, et al. Stability Study of Power System. International Journal of Energy and Power Engineering , vol. 4, no. 2, Apr. 2015, pp. 4350.

[26] Kumar, Abhishek, and D. M. Akbar Hussain. HVDC (High Voltage Direct Current) Transmission System: A Review Paper. Gyancity Journal of Engineering and Technology, vol. 4, no. 2, July 2018, pp. 110., doi:10.21058/gjet.2018.42001.

[27] Xu, Peng Fei. Some Reflections on the High Voltage Direct Current Transmission. Applied Mechanics and Materials, vol. 203, 2012, pp. 355359., doi:10.4028/www.scientific.net/amm.203.355.

[28] Yu, Jing, et al. Fault Analysis of AC Circuit Based on Phase Analysis. Applied Mechanics and Materials, vol. 40-41, Nov. 2010, pp. 262265., doi:10.4028/www.scientific.net/amm.40-41.262.

[29] PATHAN, MUJIB J., and V. A KULKARNI. FAULT ANALYSIS OF HVDC TRANSMISSION SYSTEMS. International Journal of Electrical Engineering & Technology (IJEET), vol. 7, no. 3, 2016, pp. 106116.

[30] Introduction to ETAP. Load Flow Software — Load Flow Analysis — Power Flow Software — ETAP, etap.com/.

[31] Cover 1. Bt.e-Ditionsbyfry.com, Equal Opportunity Publications, Inc, 14 Sept. 2018, bt.e-ditionsbyfry.com/publication/index.php?i=526957.

[32] PowerFactory Applications. PowerFactory DIgSILENT, www.digsilent.de/en/powerfactory.html.

[33] What Is Python? - Definition from WhatIs.com. WhatIs.com, whatis.techtarget.com/definition/Python.

[34] Markgraf, Bert. Cable Length vs. Power Drop. Sciencing.com, Sciencing, 6 Dec. 2018, sciencing.com/cable-length-vs-power-drop-12184174.html.

[35] Michalski, Ryszard S., et al. Machine Learning: an Artificial Intelligence Approach. Tioga Pub. Co., 1983.

[36] Alpaydin, Ethem. Machine Learning: the New AI. The MIT Press, 2016.

[37] Lison, Pierre. An Introduction to Machine Learning. 3 Oct. 2012, heim.ifi.uio.no/plison/pdfs/talks/machinelearning.pdf.

[38] Alpaydin, Ethem. Machine Learning: the New AI. The MIT Press, 2016.

[39] Robert, Christian. Machine Learning, a Probabilistic Perspective. Chance, vol. 27, no. 2, Apr. 2014, pp. 6263., doi:10.1080/09332480.2014.914768.

[40] Austin. PCA: Principal Component Analysis -. Austin G. Walters, 14 Mar. 2015, austingwalters.com/pca-principal-component-analysis/.

[41] What Is a Confusion Matrix in Machine Learning. Machine Learning Mastery, 30 May 2018, machinelearningmastery.com/confusion-matrix-machine-learning/.

[42] Ray, Sunil, and Business Analytics. 30 Questions to Test a Data Scientist on K-Nearest Neighbors (KNN). Analytics Vidhya, 5 Sept. 2017, www.analyticsvidhya.com/blog/2017/09/30-questions-test-k-nearest-neighbors-algorithm/.

[43] Ray, Sunil, and Business Analytics and Intelligence. Understanding Support Vector Machine Algorithm from Examples (along with Code). Analytics Vidhya, 14 Sept. 2017, www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/.

[44] Sahu, Tushar Kant, and Vinay Jain. Biometric Person Identification Using Multilingual Speech Processing. International Journal for Scientific Research & Development (IJSRD), vol. 4, no. 10, 2016, pp. 687691.

[45] Donny-Hikari. A Face Detection Program in Python Using Viola-Jones Algorithm. GitHub, github.com/Donny-Hikari/Viola-Jones.

[46] Boosting and AdaBoost for Machine Learning. Machine Learning Mastery, 22 Sept. 2016, machinelearningmastery.com/boosting-and-adaboost-for-machine-learning/.

[47] Woods, Andrew. $160bn 'Five-Island Mega Project' Announced by Kuwait. Middle East, 3 Nov. 2017, middleeast.businesschief.com/marketing/1223/160bn-five-island-mega-project-announced-by-Kuwait.

[48] Gulf Cooperation Council Interconnection Authority (GCCIA), www.gccia.com.sa/P/corporate_objectives/22.