Early Detection of Cyber-Physical Attacks in Electric Vehicles Fast Charging Stations using Machine Learning

by

Zainab Shams Warraich

A thesis submitted to the
School of Graduate and Postdoctoral Studies in partial
fulfillment of the requirements for the degree of

Master of Applied Science in Electrical and Computer Engineering

The Faculty of Engineering and Applied Science

University of Ontario Institute of Technology (Ontario Tech University)

Oshawa, Ontario, Canada

August 2021

THESIS EXAMINATION INFORMATION

Submitted by: **Zainab Shams Warraich**

Master of Applied Science in Electrical and Computer Engineering

Thesis title: EARLY DETECTION OF CYBER ATTACKS IN ELECTRIC VEHICLES CHARGING STATIONS USING MACHINE LEARNING

An oral defense of this thesis took place on September 17, 2021 in front of the following examining committee:

Examining Committee:

Chair of Examining Committee	Dr. Khalid Elgazzar
Research Supervisor	Dr. Walid Morsi Ibrahim
Examining Committee Member	Dr. Mohammad Youssef
Thesis Examiner	Dr. Xianke Lin

The above committee determined that the thesis is acceptable in form and content and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate during an oral examination. A signed copy of the Certificate of Approval is available from the School of Graduate and Postdoctoral Studies.

ABSTRACT

In smart grids, the concept of “vehicle-to-grid” allows the electric vehicles to export power to the grid to support the electric utilities in the power distribution system’s operation. The implementation of such a concept dictates the integration of a set of communication networks, which leads to numerous cyber vulnerability issues. The work in this thesis investigates the development of a novel approach that uses machine learning to early detect such denial-of-service attacks to the fast-charging stations. The study investigated the effectiveness of the proposed approach when considering different time resolutions of the advanced metering infrastructure data including hourly, half-hourly and quarter hourly. The proposed approach has been tested through MATLAB simulation environment on a microgrid equipped with renewable energy resources as well as electric vehicles in vehicle-to-grid-mode. The results have shown that the proposed approach was successful in early detecting cyberattacks at an average accuracy of nearly 98%.

Keywords: Cross validation method; cyber-physical attack; denial of service attack; decision tree; early detection.

AUTHOR'S DECLARATION

I hereby declare that this thesis consists of original work of which I have authored. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize the University of Ontario Institute of Technology (Ontario Tech University) to lend this thesis to other institutions or individuals for the purpose of scholarly research. I further authorize University of Ontario Institute of Technology (Ontario Tech University) to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my thesis will be made electronically available to the public.

Zainab Shams Warraich

STATEMENT OF CONTRIBUTIONS

The main contribution of this thesis is to introduce a new cyber-physical attack detection method that can provide optimal detection technique with less computational time.

In this research, Predictor Important is calculated to demonstrate the importance for utilizing predictor importance to boost up the detection accuracy of the classifier.

A novel approach is introduced that combines the Decision Tree classifier with k-fold cross validation method discussed in detail in Chapter#3.

This work demonstrates the effectiveness of the proposed approach through calculating different evaluation metrics. Discussed in Chapter#6.

Z. S. Warraich and W. G. Morsi, “Early detection of cyber-physical attacks on fast charging stations using machine learning” *In review.*

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the Almighty God, the Most Compassionate and the Most Merciful, for His showers of blessings throughout my research which led me to complete my thesis.

I would like to express my deep and sincere gratitude to my supervisor, Dr. Walid Morsi Ibrahim, for giving me this opportunity and providing invaluable guidance throughout my studies. His dedication, intellectual guidance, and timely advice made it possible for me to complete my thesis.

My appreciation and deep thanks to all the students within our Smart Grid and Electric Vehicles research group at Ontario Tech University for taking time out of their busy schedules to provide assistance.

Last but not least, I would like to express my appreciation to my husband, Shafi Warraich, who stood by me and always motivated me throughout this journey. A special thanks to my parents and in-laws for their love, prayers, and encouragement in every step of my life. I am extremely grateful to my mother-in-law, Razia Sultana, and father-in-law, Irfan Warraich, for their love, care, and support throughout my research studies. I would also like to express my appreciation to my uncle-in-law, Shukarullah Chaudhry, for his moral support and the keen interest shown to complete my thesis successfully.

TABLE OF CONTENTS

List of Tables

Chapter 5

List of Figures

Abbreviations

FCS	Fast Charging Stations
EVs	Electric Vehicles
ICT	Information and Communication Technology
SCADA	Supervisory Control and Data Acquisition
EVI	Electric Vehicle Infrastructure
IoT	Internet of Things
V2G	Vehicle-to-Grid
DERs	Distributed Energy Resources
RES	Renewable Energy Sources
ESS	Energy Storage Systems
EVSE	Electric Vehicle Supply Equipment
BEVs	Battery Electric Vehicles
PHEV	Plug-in-hybrid Electric Vehicle
ZEV	Zero Emission Electric Vehicles
AC	Alternating Current
DC	Direct Current
RPH	Range Per Hour

USA	United States of America
HOV	High Occupancy Vehicle
BC	British Columbia
USD	United States Dollar
UCLA	University of California Los Angeles WinSmartEV
SMERC	Smart Grid Energy Research Centre
DoS	Denial-of-Service
ML	Machine Learning
FDIA	False Data Injection Attacks
BDD	Bad Data Detection
SVM	Support Vector Machine
IDM	Intrusion Detection Model
IEC	International Electrotechnical Commission
GOOSE	Generic Object-Oriented Substation Event
PCC	Point of Common Coupling
IEDs	Intelligent Electronic Devices
SOC	State-of-Charge
IEEE	Institute of Electrical and Electronics Engineers
NEC	National Electrical Code
MILP	Mixed Integer Linear Programming

Wi-Fi	Wireless Fidelity
OLTC	On-Load Tap Changer
EWV	Enhanced Weighted Voting
PI	Predictor Importance
ROC	Rate-of-Change
DT	Decision Tree
CVM	Cross Validation Method
NARR	North America Regional Reanalysis
CWEEDS	The Canadian Weather Energy and Engineering Dataset
KPMG	Klynveld Peat Marwick Goerdeler
LOL	Loss-of-Life
DNO	Distribution Network Operator
TOU	Time-of-Use
BESS	Battery Energy Storage System
LOOCVM	Leave One Out Cross Validation Method
DAC	Distribution Automation and Control
CIA	Confidentiality, Integrity, Availability
OpenDSS	Open Distribution System Simulator
p.u.	per unit
DBM	Deep Belief Method

CART	Classification and Regression Tree

1. Introduction

1.1. Background

The smart grid is an evolution of the existing electric power gird by integrating the Information and Communication Technology (ICT) into its operation. Traditionally, the electrical infrastructure is responsible for the production, transmission, and the distribution of electric power. Recently, the information and communication infrastructure is integrated to accounts for not only providing two-way communication between the electric utilities and the customers but also for the monitoring, control, and the automation of electric power equipment. Electric Vehicles (EVs) are connected to the smart grid through charging stations and these charging stations can be private charging stations i.e., charging stations at residential areas, at workplace or public charging stations i.e., charging stations in parking lots [1]. EVs are one of the envisioned components of the smart grid [2].

The EVs can operate in two different modes: 1) charging mode and 2) vehicle-to-grid mode. In the charging mode, the vehicles batteries are charged from the grid with no power exported back to the grid. On the other hand, in the Vehicle-to-Grid (V2G) mode, the EVs are allowed to export power back to the grid at the request from the utility and hence contributing to the service provision (e.g., frequency and voltage regulation). Such services are very important for the electric utilities in the microgrids equipped with Distributed Energy Resources (DERs) such as wind farm and solar photovoltaic to avoid the frequency and voltage deviations. The DERs are considered as a viable solution to deliver clean energy to the customers. The growth in a wide variety of DERs and their utilization in the microgrids can feed to the load in different locations without their dependency on the main grid. Fulfilling the power demand of EVs from the microgrids results in zero emission and economical solution. The solution to deal with the

increasing load demand from EVs is the utilization of DERs like Renewable Energy Sources (RES) and Energy Storage Systems (ESS) [3]. Additionally, EVs are shown to emit higher emissions for fossil fuel-based countries, which necessitates the usage of RES for maximizing EVs benefits [4].

The integration of the ICT to the electricity grid opens an avenue to the smart grid to be prone to cyberattacks and hence may be considered highly vulnerable to different types of cyber attacks. Initially, hackers targeted the power grid especially Supervisory Control and Data Acquisition (SCADA) and information technology devices. The most common attacks were phishing attacks and Denial-of-Service (DoS) attacks. Currently, the introduction of Internet of Things (IoT) in Electric Vehicle Infrastructure (EVI) introduces new attacks and makes EVI highly vulnerable to cyberattacks [5].

The Electric Vehicle Supply Equipment (EVSE) are not only accountable for charging EVs but for the authorization of EVs and their connection to the grid [6]. The communication capabilities through the internet and inherent cyber physical characteristics of EVI make them highly susceptible to cyber attacks and attraction for the hackers [7].

1.2. Electric Vehicles Trend Worldwide

Battery powered EVs provide numerous benefits to the customers, provide environmental benefits, and benefits to the electric utilities. The operating cost for EVs is less as compared to the operating cost of a gasoline vehicle. Moreover, in North America [9], the gas prices are the highest, on the other hand, the average electricity price in BC was $0.124 per kWh for the residential customers in 2020 [8]. If the vehicle owners switch to EVs from gasoline, they can save up to $1800 on the fuel cost per year [9]. Furthermore, the rapid pace of the electrification in the transportation sector will help to reduce the dependency on the fossil fuels, which reduces the

levels of greenhouse gases, thus securing the environment. Last but not the least, the EVs can provide ancillary services for the electric power utilities. EVs can be utilized as DER using V2G technology and can play a vital role for peak shaving, spinning reserve and power regulation [7].

In 2020, there were more than 10 million electric cars on the road worldwide, which shows the rapid growth of EVs sales, as shown in Fig. 1.1. Nearly 20 countries worldwide have restricted the sale of the gasoline vehicles and forced all vehicle sales to be electric. In 2020, the governments globally spent USD 14 billion on EVs sales, however, EV customers have spent USD 120 billion on EVs purchase worldwide. In the USA, states like California, New York, New Jersey, and Massachusetts set their own goals to achieve net zero emission. Also, by 2035, it is anticipated to have all new light duty vehicles and light truck sales be just electric in California. The price of lithium-ion car batteries is reasonable due to which the EVs prices are reducing. The number of private and public charging stations are increasing rapidly with nearly 370 different types of EVs models are launched worldwide to meet customer's requirement [10].

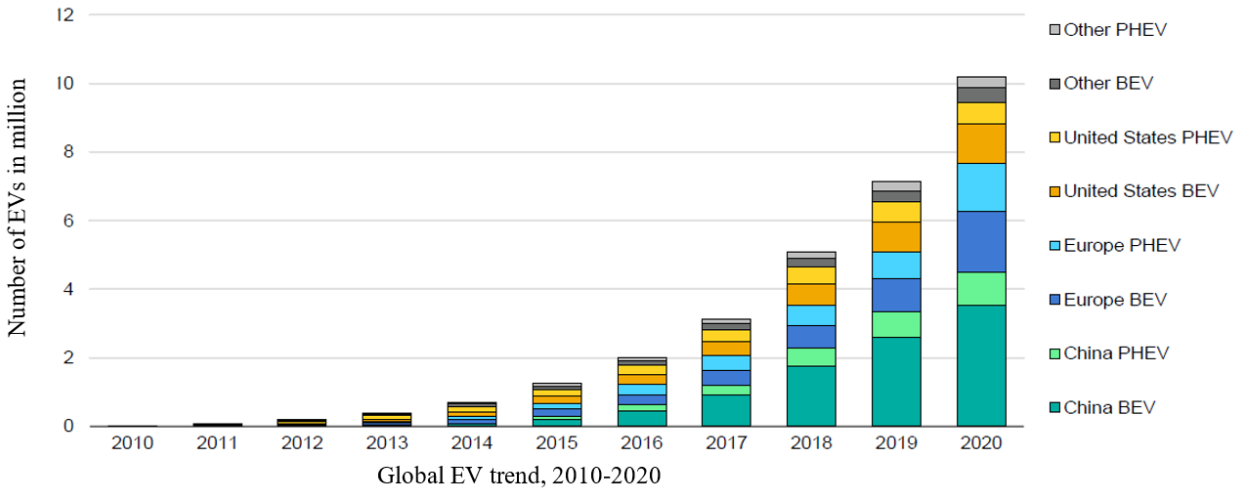

Fig. 1.1. Global EVs stock from 2010-2020 [10].

EVs trend for last decade is shown in Fig. 1.1. The trend shows that China has the highest adoption of EVs, after that is Europe, then United States and finally others. Others include countries like Australia, Brazil, Canada, Chile, India, Japan, Korea, Malaysia, Mexico, New Zealand, South Africa, and Thailand. Europe includes the EU27, Norway, Iceland, Switzerland, and the United Kingdom [10]. BEVs: Powered entirely by battery engine. Has a large rechargeable battery. BEVs are fueled with electricity from the grid. However, PHEV: Plug in hybrid electric vehicle, have rechargeable battery and petroleum fuel tank as well. EVs are fueled with electricity from the grid.

There are several EVs initiative campaigns started worldwide to increase the adoption of the EVs. The motive of the *EV30@30 Campaign*, initiated in 2017, was to have 30% of the vehicles to be electric by 2030 across 14 countries in the world. Another significant campaign launched recently in 2020, was intended to develop strategies, action plans and programmes with the agreement of government and industry stakeholders for the rapid adoption of the EVs. The objective of *GEF-7 Global Electric Mobility Programme*, launched in 2021, is to assist low and middle-income countries for their transition of the transportation sector from combustion engine to electric mobility. Moreover, *EV100* is another global initiative that includes 100 companies worldwide that are willing to shift to electric mobility and provide charging infrastructure for employees and customers by 2030. According to this initiative, it is predicted to have 4.8 million shifts to EVs, and 6500 charging locations installed worldwide by 2030. The *EV100* members were successful to deploy 169 thousand EVs, 16900 charging points installed at 2100 locations worldwide by 2020. In 2020, the number of Private Light Duty Vehicle chargers were almost 9.5 million, among which 7 million were residential charging stations and 2.5 million were at workplaces. It is forecasted to have 105 million private chargers for electric light duty vehicles by

2030 among which 80 million chargers would be at residential sites and 25 million at workplaces [10].

CleanBC Go Electric Charger Program, includes the funding of $5.4 million dollars, provides 50% cost (up to $80,000) required for the installation of per FCS in BC. Furthermore, *CleanBC Go electric fleets program*, launched in 2021, provides financial, consultation and technical support for the installation and purchase of level 2 and FCS in BC [11].

1.3. Fast Charging Station Trend Worldwide

The EVSE, which is the charging point for EVs, can provide Alternating Current (AC) or Direct Current (DC) power to the EVs. The three common charging levels include level 1, level 2 and fast charging level that are commercially available [7]. Level 1 charging mostly present in the small residential areas requires 12-24 hours to fully charge an EV, providing a Range Per Hour (RPH) charging rate of 4.5 miles/hour. Level 2 charging requires 4-6 hours to fully charge an EV, providing a RPH charging rate of 25-30 miles/hour using a 7-kW station. However, Level 3 charger, also known as DC fast charger, can provide 80% of RPH in just half an hour. The charging duration is less because DC chargers deliver high power [12]. The popularity of the EVs is increasing rapidly, as of 2017, the number of level 2 charging stations has reached 50,000 in the United States of America (USA) [13]. According to Fig. 1.2, there are more than 1.25 million EV charging stations worldwide [14].

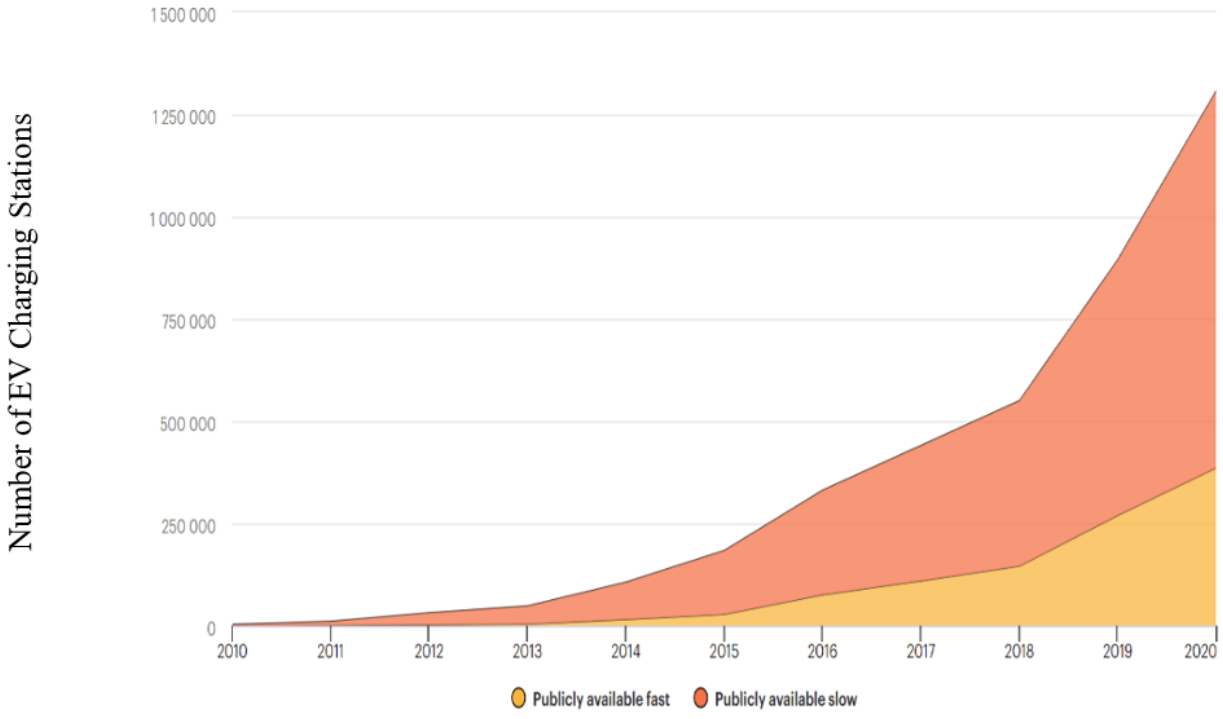

Fig. 1.2. EVs charging station trend worldwide [14].

1.4. Electric Vehicle Infrastructure trend in British Columbia

The Government of Canada has already invested over $1 billion for the deployment of electric mobility Canada [15], especially BC plays a vital role to achieve the goal of "net zero emission." In 2020, BC became the leader in the zero-emission vehicle market, the highest number of Zero-Emission Electric Vehicles (ZEV) registered in North America. BC has 2500 public charging stations, one of the largest public charging networks in the country. Also, BC has currently more than 30,000 EVs on the road and in 2019, nearly 9% of the light duty vehicles were electric. By the end of 2020, there were 54,469 light duty ZEV (Fig. 1.3) and 2500 public charging stations among which 480 were the FCS in BC (Fig.1.4) . It is expected to have 100% of the

transportation to be electric in BC by 2040, which clearly indicates that EV adoption will increase drastically in the coming years. Table 1.1 defines different types of fast charging connectors used in FCS [11].

Moreover, the BC and the Federal government is initiating some fundamental steps to achieve the goal of reducing greenhouse gas emission. Firstly, Vancouver is imposing EV charging infrastructure in new buildings, both commercial and residential. Secondly, the federal government is offering EV incentives of $2500-$5000 Canadian dollars for new EV buyers, and the BC government has allocated $50,0000 for the EV incentive program. Furthermore, BC government is also offering $5000 discount from the purchase price of battery vehicles and a discount of $1500 when an individual buys a hybrid electric vehicle. Thirdly, the early retirement vehicle program gives an opportunity to gain $6000 when a person trades in their gasoline vehicle. Last but not the least, EV users can gain access to High Occupancy Vehicle (HOV) lanes even if there is only one person in the cars and EV users can enjoy the perk of having dedicated parking just for EV users in some areas of the city of Vancouver [9]. In 2019, $1.15 million fund was allocated to create 23 EV fast chargers around BC. Moreover, the Government of BC also announced a contribution of United States Dollar (USD) $575,000 towards the fast chargers through its Clean Energy Vehicle Public Fast Charging Program [16].

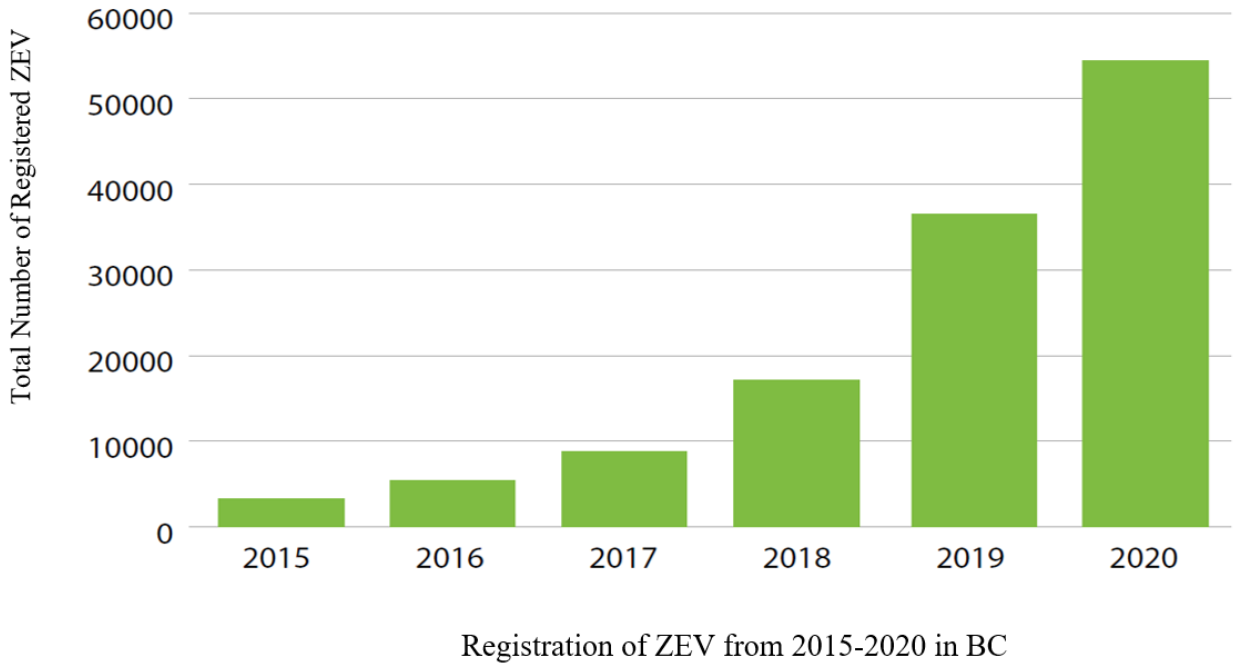

Fig. 1.3. Light duty ZEV registration in BC from 2015 to 2020 [11].

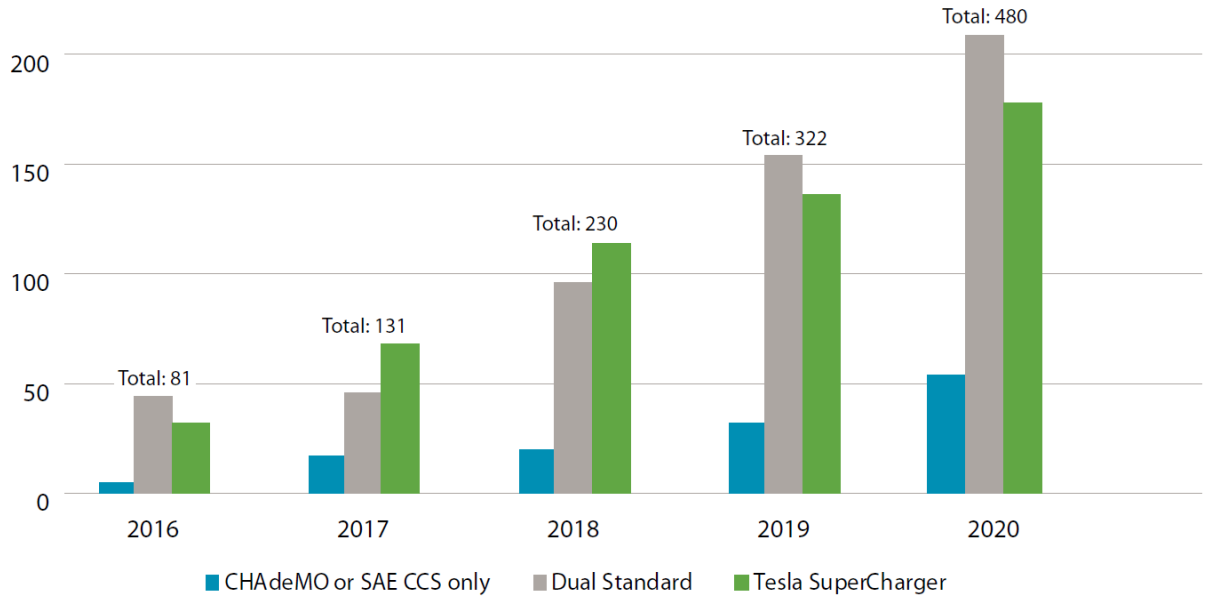

Fig. 1.4. Public fast charging growth in BC [11].

Table 1.1: Different types of fast charging connectors [11].

Charger Type	**Description**
SAE CCS	connector type for public fast charger, compatible with USA brand EVs.
CHAdeMO	connector type for public fast charger, compatible with Japanese brand EVs.
Dual Standard	Station with SAE CCS and CHAdeMO connector for public fast charger. All EVs can charge from these stations, even tesla with adapter.
Tesla Supercharger	Connector only compatible with tesla vehicles.

1.5. Vulnerabilities of Electric Vehicle Infrastructure

The cybersecurity awareness in EVs started in 2010 [12]. The research in [1] determines the vulnerabilities of the charging infrastructure along with IT security considerations in ISO/IEC and SEA standards for the secure interface between charging EV and charging spot. Different types of cyber attacks were specified along with their drastic impact, to emphasize on the need for the security measures.

In 2013, the work in [2] revealed the threat analysis in smart EV charging services. Besides, security and privacy requirements that could be utilized for the risk assessments of smart charging systems and a guideline for future techniques to secure the smart charging system was mentioned. Henceforth, the importance of smart charging was explained as uncontrolled charging could impact the load analysis especially the peak load, which will negatively impact the estimation of generation and the demand.

Correspondingly, in 2016 the work in [17] mentioned how EVI is attaining popularity among the hackers. Possible cyber threats to EVI along with hacker's motivation and concerns

were discussed. Nevertheless, in 2019, the study in [6] mentioned how the cyber-vulnerabilities in EVSE could potentially damage the availability, confidentiality, and integrity of the network of charging stations, customers and even power grids. The study in [6] emphasised on the importance of understanding the cyber and physical components interactions and their impact on each other for a secure charging station. The study divided the attacks as network-based attacks, physical attacks, and hybrid attacks and explained each type of attack. Afterwards, an overview of some software and hardware-based approaches to make EVSE more resilient to cyber attacks were determined.

In 2020, a detailed analysis of cyber vulnerabilities of EV ecosystem is determined [18]. The vulnerabilities of the power grid, EVs and FCS are discussed in a comprehensive manner. The purpose of the research was to emphasize on the importance of research and security tools necessary in the field of FCS by determining the possible detrimental impacts of cyber attacks on EV, FCS, and power grid [18].

1.6. Impact of Cyber Attacks in Fast Charging Stations

In 2019, the report in [19] determined the current state-of-the-art in the field of cybersecurity of charging infrastructure in the USA. The vulnerabilities of EVSE could impact transmission systems, distribution systems, emergency services and the manufacturing sector as well. The impact of cyber attacks could be limited or extensive.

The consequences of the cyber attacks could be limited to the failure of EV, damage to the batteries of EV, DoS for the EVs, or the compromise to the financial information of the customers. On the other hand, the effect of cyber attack could be so extreme that it can shut down or in extreme conditions could completely damage the EVSE network. The customers' confidence on electric

mobility could get affected and last but not the least, the bulk power system could be disturbed [19].

In 2019, the University of California Los Angeles (UCLA) Smart Grid Energy Research Centre (SMERC) in [20] published the recent updates on the cybersecurity research on EV infrastructure. The detailed description of the cyber vulnerabilities, attack surfaces, and possible impacts of cyber attacks were discussed. According to the research, scenarios of insider attack were explained, which could cause EV overcharging and possibly could result in explosion of EV and hence results in loss of customer's confidence. Another scenario was mentioned, where an insider attack on EV charging control centres could cause simultaneous charging of fast chargers, consequently the overloading of distribution transformer could occur and leading to power outages.

1.7. Problem Statement and Motivation

The discussions revealed the role of the FCS in increasing the popularity of the EVs among the customers. Furthermore, the role of the EVs in the service provision is important as they can provide such services to the electric utilities through the V2G operation. Such services are extremely important when considering microgrids equipped with distributed renewable energy resources to avoid frequency and voltage regulations. The integration of the ICT in FCS makes them highly susceptible to the cyber attacks. Such attacks may lead to several impacts to the electric power system operation such as increased stress to the assets as well as power outages to the customers in addition to the Denial-of-Service (DoS) provision offered by these vehicles. Therefore, there is a need for an approach to provide early detection of such cyber attacks in FCSs. The approach should be able to detect the cyber attacks before the system starts to experience the

impacts of the cyber attacks. The motivation of the research was to develop a detection technique that can detect the malicious activity in FCS, to secure the operation of FCS.

1.8. Research Objectives

The following summarizes the research objectives of this study:

- To develop an approach to early detect cyber attacks in FCSs prior to the system experience the impacts.
- To investigate the effect of different time resolutions of smart metering infrastructure data on the detection accuracy of the proposed approach.
- To assess the effectiveness of the proposed early detection approach against different types of Denial-of-Service (DoS) attacks in FCSs.
- To fine tune the Machine Learning parameter for improving the detection accuracy.

1.9. Thesis Organization

This thesis includes six chapters. Chapter 1 explains the Electric Vehicles trends in the near future and the need for Fast Charging Stations accordingly. Further, it explains the vulnerabilities of Fast Charging Stations and the importance of detecting cyber attacks followed by the problem statement and motivation. Finally, the research objectives are outlined.

Chapter 2 determined the importance of Machine Learning and their usage for the detection of cyber attacks in the distribution systems in general and after that it surveys different methods for the detection of cyber attack in Fast Charging Stations that were previously published in the literature. The advantages and the disadvantages of each method are presented and discussed. The chapter finally outlines the main research gaps of this problem and highlights the research trend that will be undertaken in this thesis.

Chapter 3 is dedicated to the explanation of the proposed Intrusion Detection Model. Firstly, this chapter explains the methodology for the data generation, along with illustrating the data preprocessing steps. Alongside, the Decision Tree and k-fold Cross Validation Method for the classifier training and attack detection is also discussed in detail. Additionally, the Intrusion Detection Process and the evaluation metrics for the classifier are illustrated in this chapter.

Chapter 4 illustrates three different types of Denial-of-Service attacks in Fast Charging Stations. Furthermore, previous research on impact analysis of uncontrolled charging on the power systems is explained, the importance for the early detection of cyber attacks and finally, is summarized.

Chapter 5 introduces the results of the intrusion detection with a detailed analysis. The algorithm is implemented while being tested rigorously for different time resolutions. The results of different time resolutions with different k-folds are presented and are discussed by comparing the classification accuracies, F-score, precision, recall and undetected rate. As well, the results of detecting the three different attacks have been presented, discussed, and compared to assess the performance of the intrusion detection systems for different attacks.

Finally, Chapter 6 presents the main conclusion and recommendation regarding the mitigation of Denial-of-Service attacks in Fast Charging Stations. Hereafter, the future work that can be based on the work presented in this thesis is presented.

2. Literature review

2.1. Introduction

In this chapter, the previous work relevant to "Cyber attacks in Power Systems" along with the previous work on "Cyber security in FCS" is reviewed. The past research for securing the power system is mentioned in the first part of this chapter however, the contributions for securing the EVSE from the cyber attacks is presented in the second part of the chapter. The literature review on the cyber attacks in power systems along with the vulnerabilities of substation are reviewed as well as the contributions being done for the stability of the power system.

Since, the main goal is to detect the vulnerabilities of FCS and to secure it from potential cyber attacks, the literature review sheds light on the contributions being done in the past for securing the FCS, thus ensuring the stability of the power system. The motive is to assess the contributions being done in the field of EVSE, making comparisons based on how accurately specific techniques were able to detect the attack and identifying the drawbacks of applying the specific techniques. Finally, this chapter identifies the research gaps that currently exist in the literature and pave the way towards the main contributions of the research work presented in this thesis.

2.2 Previous Work on Securing the Power System from the Cyber Attacks

This section presents the existing methods that are introduced in the literature for the security of power system. The previous work is classified into three different types: Cyber attack detection in Industrial Control Systems, Cyber attack detection in Substation Automation and Control Systems, and Cyber attack detection on state estimation as discussed below.

2.2.1 Cyber Attack Detection in Industrial Control Systems

Power Fingerprinting is a technique that monitors the processor power consumption in industrial control systems using external devices and execute signal processing and pattern recognition technique for the anomaly detection. Power Fingerprinting performs three major functions for the detection of the attacks. Initially, the fresh power traces from the execution of industrial control systems are captured followed by the comparison of these fresh traces with the reference stored traces and finally, the intrusion detection. In [21], the Power Fingerprinting technique is used to detect the unauthorized software modification on a basic commercial radio platform, which is PICDEM Z Evaluation Board from Microchip and consists of PIC18 processor. Moreover, the correlation index is calculated to identify tweaked samples and untweaked samples. Samples close to the stored reference traces have higher correlation value and the samples that are modified have lower correlation values. The drawback is that there is no information provided in the study regarding the detection accuracy of applying the Power Fingerprinting technique for anomaly detection.

2.2.2 Cyber Attack Detection in Substation Automation and Control Systems

In [22], several classifiers are tested to discriminate between the man-made and the natural events happening in the SCADA system. Both attacks and normal operations data must be acquired in-situ, from the system that will be monitored, and then must be appropriately tuned to minimize the false positives. The drawback of using JRipper +Adaboost for the proposed IDM is the computational complexity [23].

In [24], the cyber security test bed, which includes a real grid connected photovoltaic (PV) Supervisory Control and Data Acquisition (SCADA) system, is used to investigate the

consequences of Address Resolution Protocol (ARP) spoofing-based Man in the Middle attack. In Man in the middle attack, an attacker redirects the communication traffic between two hosts to the malicious host. The malicious host modify the communication packets and send them back to the destination host. The entire time, two hosts are unaware of any malicious activity. In the experimentation, the internet traffic and content analyzer (ITACA) software was used to analyse the network traffic in real-time. The ARP spoofing-based man in the middle attack can have devastating consequences on the power grid. One example is the false operation command to the circuit breaker, can result in load loss, decrease power supply reliability and threat to human safety. The limitations include the need for detection techniques to secure power grid from ARP spoofing-based Man in the middle attack.

In [25], the Snort software is used for the cyber security of IEC61850 based IEDs. Snort software is a rule-based intrusion detection system, and the rules are attained from the experimental data obtained by launching the simulated attacks like DoS attack, password crack attack on IEDs and by launching packet sniffing attacks. The study did not provide the numerical results, regarding the efficiency of the system.

In [26], a behaviour-based intrusion detection system for IEC 61850 based digital substation was introduced. Anomaly detection algorithm detect anomalies based on multicast messages like GOOSE and MMS. Static and dynamic features of IEC 61850 based network traffic were observed to monitor suspicious behaviour. Static feature includes response and report feature of MMS protocol-based command. However, dynamic featured includes Recency-Frequency-Monetary (RFM) of GOOSE messages. To improve the accuracy of the anomaly detection algorithm, generic traffic features like number of connections per second(cps), number of bits per second (bps), number of packets per second (pps) were also included as dynamic features.

Intrusion detection algorithm detected 24 network generated attack scenarios with the detection accuracy of 99%. Even though, the proposed algorithm performed well but it still needs to be validated on different evaluation metrics and computational time can be calculated to evaluate the computational complexity of the proposed approach. In [25], the authors proposed host-based anomaly detection and in [26] network-based intrusion detection technique was proposed. The host-based anomaly is incapable to detect anomalies at multiple hosts and network-based anomaly can have high false alarm rate. To overcome these limitations, the work in [27], proposed integrated anomaly detection system, which not only detect malicious behaviour at host and network level but is capable to detect anomalies at multiple substations simultaneously. An algorithm is proposed for temporal anomaly detection of the host-based system. Detection rules are made based on the system and security logs at substation networks and anomaly can be determined from variance between events log from different period. The host-based anomaly detection algorithm can even locate where and which type of anomaly is detected i.e., either malicious activity happened at critical settings of IED, or the status of circuit breaker is modified. The integrated anomaly detection system also includes a separate algorithm for the detection of intrusion in multicast messages in three steps. Initially, packets of SMV and GOOSE are filtered to reduce the processing burden and increase the performance of the system, then detection is performed based on pre-defined rules. Finally, the evaluation step assists in determining either the disruption is abnormal or an attack. The network-based substation vulnerability index is one of the major contributions of intrusion detection model for anomaly detection at multiple substations. However, the proposed model is evaluated on the cyber security test bed system, and it showed promising results with a negligible FPR and FNR. The model fails to detect the unknown and the

intelligent attacks and therefore the periodic update of the intrusion detection model is required for better efficiency and accuracy.

2.2.3 Cyber Attack Detection on State Estimation

In [28], the signal processing and the machine learning techniques are used for the detection of the stealthy false data injection attack on state estimation, which cannot be detected with bad data detection method. Two machine learning algorithms are used for the detection purposes. The first method uses supervised learning over labeled data and train support vector machine and the second method use semi-supervised anomaly detection algorithm for the detection. In the first method, the class labels (attack versus not attack) are given in the historical data therefore classifier can be trained to identify the attack. In second method, which is semi-supervised machine learning algorithm, the anomaly detection can be used to identify the outliers as attacks. However, authors were able to define the boundary for the safe mode and the attacked mode, but they could not find the way to distinguish whether the disturbance in the system is because of attack or because of non-attack scenarios like transmission line and generator outages. Also, there is no exchange of views on mitigating those attacks.

In [29], a survey is done to discusses major directions and recent advancements in power grid cyber attack in terms of different detection techniques, equipment protection plans, and mitigation strategies to enhance the energy delivery infrastructure resilience and operational endurance against cyber attacks. The definitions of different types of cyber attacks, potential attack surfaces, and the impacts on bulk power grids are explained. It is also described in detail how the power grid is resilience to cyber attacks and how the smart grid cyber layer should be characterized to resist cyber threats, ensuring the operational endurance and resilience. The research on some protection mechanisms in power systems against cyber adversaries is also illustrated.

2.3. Previous work on Cyber Security of Electric Vehicle Infrastructure

This section presents the existing methods that are introduced in the literature for the cyber security of electric vehicle infrastructure. The previous work is classified into three different types: Deployment of Electric Vehicles for the protection against Cyber Attacks, Impact of Cyber Attacks in Electric Vehicle Infrastructure, and Detection of Cyber Attacks in Electric Vehicle Infrastructure as discussed below.

2.3.1. Deployment of Electric Vehicles for the protection against Cyber Attacks

The work in [3] investigated the use of EV park in compliance with International Electrotechnical Commission **(**IEC) 61850 communication protocol in a microgrid as a protection scheme against cyber attacks in the microgrids. Microgrids can work in an islanding mode and support the load in case when there are issues in the power grid or in non-islanding mode when generation is less than the demand. The communication signal is sent to the logic circuit embedded in the relays during the transition from the grid connecting mode to the islanding mode to change the relay fault current settings. The relay fault settings are different for the grid connected mode and the islanded mode. In the islanded mode, the fault current level is low as compared to the grid connected mode due to the presence of DERs in the islanded mode.

A microgrid operating in three modes: grid connected mode with communication; island mode with communication and island mode without communication is considered. Three-phase to ground fault is introduced in all three modes of operation. The Generic Object-Oriented Substation Event (GOOSE) messaging protocol was used to send command signals from the Point of common coupling (PCC) to the Intelligent Electronic Device (IEDs) for shifting relay settings depending on their mode of operation. For the grid connected mode and the islanded mode with

communication, the model was stable before and after the fault. Even so, in the islanded mode with the loss of communication, the attacker compromised the communication link and the command signal from PCC was not sent to change the relay settings from the high settings in case of grid connected mode, to low settings in case of the islanding mode, which results in no fault detection. In this scenario, control of EV park identifies the attack based on the voltage and the frequency measurements from their connection points with the microgrid. By consequence, all the EVs will start to charge, which will inject high current in the system, thus required for the relay to detect the fault.

$$I_{TH} = \begin{cases} 1 & \left(I_f \geq I_G + I_M\right) + \left(I_f \geq I_{EV} + I_M\right) \\ 0 & Otherwise \end{cases} \quad (2.1)$$

$$I_{TL} = \begin{cases} 1 & \left(I_f \geq I_M\right).\left(I_C\right) \\ 0 & Otherwise \end{cases} \quad (2.2)$$

Where, I_C is the communication signal issued from the PCC to the relay to adjust its settings, I_f is the fault current, I_M is the microgrid current, I_G is the grid current and I_{EV} is the contribution current from EV park. I_{TH} is the high relay current settings in case of the grid connected, however, I_{TL} is the low relay current settings in case of the islanding mode.

There were some assumptions being made while utilising the proposed model which limits the utilization of the model. Firstly, it was assumed that EVs are connected to the EV parks for most of the day and their State-of-Charge (SOC) is enough to elevate the current to the level required for the detection. Secondly, the attack occurred during off-peak hours. Even though the proposed model has great contribution in securing the microgrid, there is still research needed for alternative solutions when the above-mentioned assumption is not met. It is necessary to investigate the impact on the distribution equipment like transformers especially in the case of all EVs requiring fast charging during peak hour with depleted batteries.

2.3.2. Impact of Cyber Attacks on Electric Vehicle Infrastructure

In [13] the research gaps for cyber-physical attacks assessment in EVI were highlighted. The purpose of the research was to highlight the areas of EVSE highly vulnerable to the cyber-physical attacks and the priority level for their security needs accordingly. Eighteen different attacks were categorised based on their class. The cyber attack and the impact could be divided into four classes:

1. Cyber-Physical:

   The attack was cyber; however, the impact can be seen on physical equipment.

2. Physical-Cyber:

   Physical attack, which results in cyber impact.

3. Cyber-Cyber:

   Cyber attack resulted in cyber impact.

4. Physical-Physical:

   Physical attack and the impacts were on physical components.

A method for the vulnerability analysis and the risk assessment of EV charging infrastructure due to cyber-physical attacks was proposed. University of California Los Angeles (UCLA) WinSmartEV charging network (Fig. 2.1) was analysed during the experimentation to identify attack vectors and attack surfaces. Several potential attack scenarios along with their impact on UCLA campus were studied. Risk assessment was dependent on two major factors; impact which determines the impact on the possibility of successful attack and the cost determining the resources required for the attackers to implement the attack. Urgency for security measures is dependent on higher value of the risk factor.

$$Risk = Impact/Cost \tag{2.3}$$

The values for the impact and the cost were decided by surveying knowledgeable researchers in the cyber attack field. Depending on the risk equation proposed in (2.3), eighteen significant attack scenarios were categorised successfully as highly vulnerable, low vulnerable and negligible impact.

The study mentioned areas that are highly vulnerable to cyber attacks but did not include the methodologies or approaches for analysing the impact of the cyber attacks on EVI. Furthermore, the research is required to detect those attacks and suggestions to improve the cyber security to have a fully secured EV charging network.

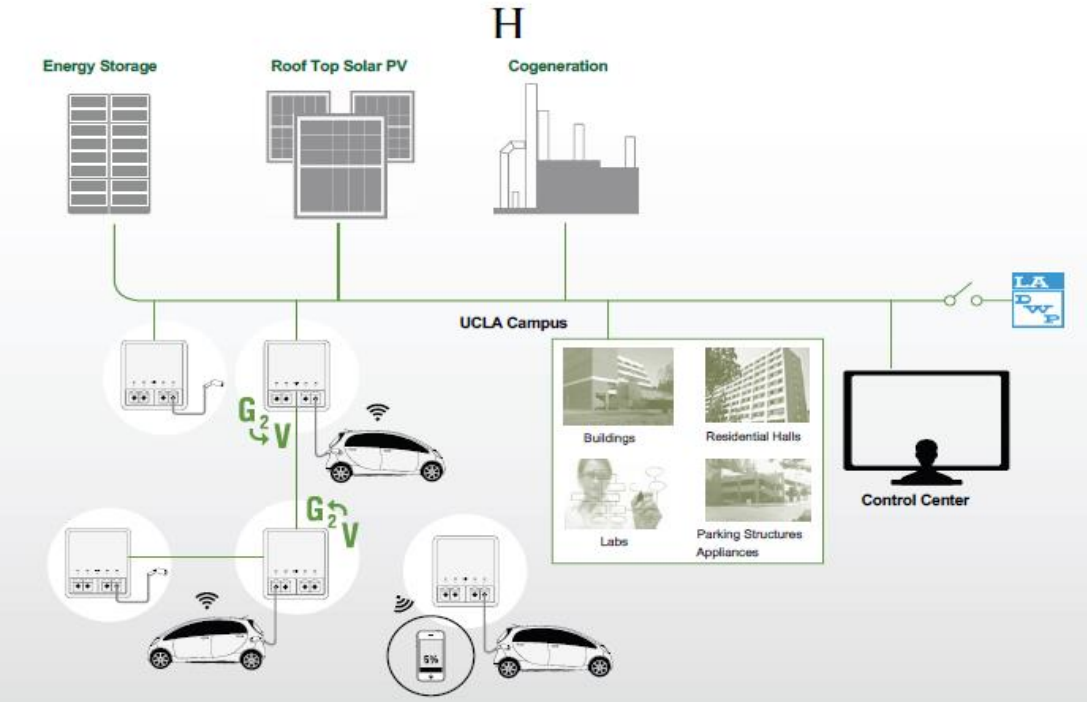

Fig. 2.1. UCLA EV WinSmartTM [13].

In [30], the vulnerabilities of the EV charging process along with the impact of the compromised EV botnets on the distribution and the transmission system is studied. An EV botnet

included compromised EVs and charging stations. Charging of EVs especially during peak hours is concerning as the total load will increase drastically which is not encouraging.

The MATPOWER software was used for the simulations to determine the impact of EV botnets on the distribution and the transmission systems. A botnet, which is a network of EV and FCS controlled by an attacker is made to charge at 7:00 in Institute of Electrical and Electronics Engineers (IEEE) 33 bus network. The power consumption increases drastically at 7:00, which causes line congestion in the distribution side of the grid, voltage dropping below the National Electrical Code (NEC) 5% voltage drop recommendation and eventually could damage the equipment at customer side leading to power outages. The study used the IEEE 39 bus system to evaluate the effect of the attack on the transmission side. The study finds out that if load altering attacks occur in a way that the protection system of the distribution network is not triggered, the impact of the attacks will be observed on the transmission side and the power outage will occur. Subsequently, the study indicated how load altering attacks on EVs and FCS can result in line congestion and voltage limit violation in the distribution systems. Simulation results in the study demonstrate that the impact of the attacks on the transmission system can result in line failure or even power outage.

The shortcomings of this study include the need for the detection and the prevention methods. There is a need for a ML algorithm for the detection of cyber attacks and detailed investigation is required to illustrate the impact of the EV botnet on the distribution transformers. The effect of multiple coordinated attacks on charging stations is required to be investigated.

2.3.3. Detection of Cyber Attacks in Electric Vehicle Infrastructure

The study in [7] investigated the impact of malware attack on EVI and its propagation into the power grid. The study described how a cyber attack can transmit from one infected EVSE to various EVSE in a city due to people travelling and the use of different EVSE for charging. The objective of the study was to introduce the optimal isolation method for the infected node to avoid further propagation of malware attacks.

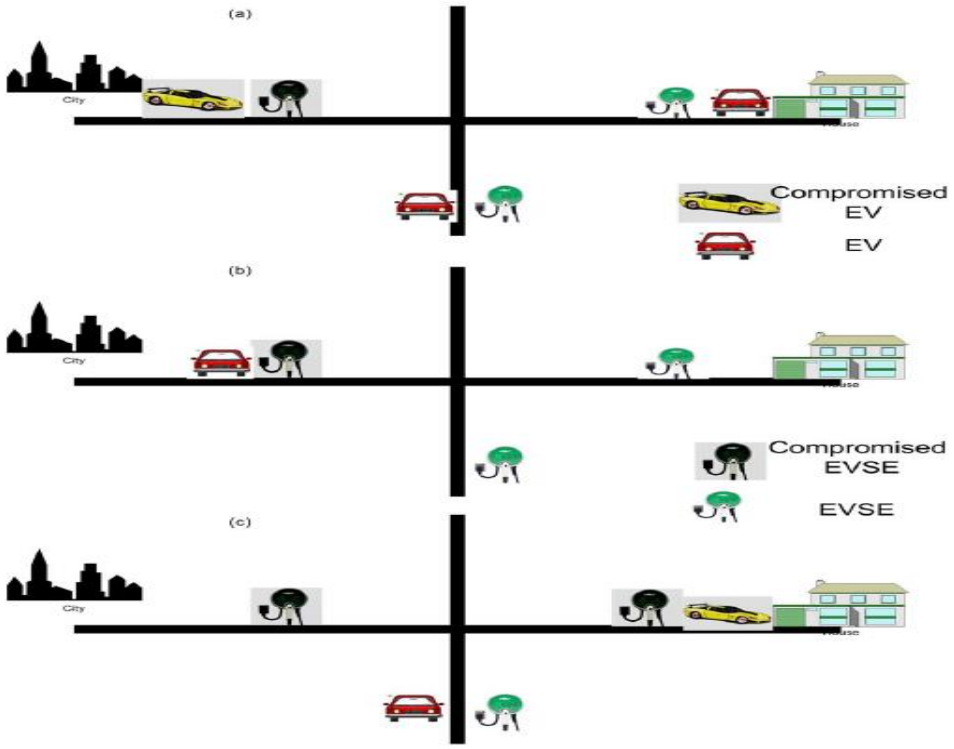

Fig. 2.2. Cyber-attack propagation in the EVI [7].

According to Fig 2.2 initially in (a) an EV from the city compromises the suburban EVSE, then in (b) a local EV charges from the compromised EVSE, consequently as shown in (c) the residential EVSE becomes infected through the local EV.

The study used a test model that includes an EVI with four EVs and three EVSEs where the cars have the possibility to charge through wireline or wirelessly. The probability that malware will spread from an infected EVSE to EV is determined by the probability (β), which is randomly chosen as 0.2 by the study. The study used a Mixed Integer Linear Programming (MILP) model to stop further propagation of worms from the infected EVs to EVSE by estimating the minimum number of EVSEs to be isolated temporarily. Experimentations were done to estimate the threat level of EVSEs. The threat levels of EVs and EVSEs are the function of probability (β). The threat level illustrates the possibility that EV and EVSE are either compromised or not, which assists in deciding whether the charging station will stay connected or disconnected temporarily from the system.

The drawback of the research in the study is that, changing the value of β (probability that the attack will propagate from EV to EVSE and vice versa) has impact on the threat level, which restricts the use of proposed model for experimentation, as accurate value for β is still needed to be investigated. Besides, research needs to be done to investigate the impact of these attacks on the smart grid cyber devices such as the distribution transformers, the phasor measurement units and the smart metering infrastructure. Furthermore, a deterministic approach is required that can detect the malware spread efficiently before the EVI become impacted. Nonetheless, the performance evaluation of the proposed approach was not assessed.

The study in [5] looked into the impact of botnet attacks on the distribution system. Power botnet includes the collection of Internet-facing devices such as air conditioner and water heaters that utilise information communication technologies such as Wireless Fidelity (Wi-Fi) and Bluetooth that are controlled by an attacker. The contribution of the study includes the impact and the detection of three different types of power botnet attacks on distribution systems using Open

Distribution System Simulator (OpenDSS). As well as they examined how the On-Load Tap Changer (OLTC) lifespans change if the transformer is under attack.

Historical sensor data from the power grid was used for the detection of attacks and no attack events. Neural network was used for the detection of power botnet attacks. Alongside, Keras deep learning library was used for the detection. The partitions of the dataset for testing-training-validation was as follows 50-45-5. Three different types of attacks were considered in which the attackers turn on and turn off the high wattage devices with different patterns. In the first attack, the devices are turned on and off as quickly as possible with a detection rate of 100%. In attack 2, the devices were turned on and off periodically and the detection was 99%. The third attack has the lowest accuracy of 80% where the attack has a 20% chance to occur in 10,000 time slots. According to the experimentation done in the study, the OLTC lifespan will be just 0.52 years if 80% of the devices are under attack. The study concluded that the power botnet attacks can have detrimental impacts on the reliability, the quality of the power supply and the hardware equipment, therefore, increasing the operational cost of the system.

Further efforts need to be made to more complicated power botnet attacks. There is still a need to study more properties of these attacks and their impact on the grid to develop supervised ML detection techniques. Neural networks are prone to overfitting, therefore, ML algorithms like Decision Tree (DT) with cross validation method needs to be investigated to increase the detection rate.

The work in [31] looks into the detection of the cyber attacks caused by FDIA. The FDIA injects the false data to the sensors, which impact the control application of the microgrids like voltage control and power control applications.

The study applied Hilbert Huang transform to extract the signal's features when a microgrid is under FDIA. Following this step, the output of the Hilbert Huang transform is input to the ensemble learner, which is used to detect the cyber attacks. An Enhanced Weighted Voting (EWV) scheme is used for implementing selective ensemble learning and Krill Herd optimization algorithm assists in selecting the optimal class specific thresholds. To train the ensemble learner, the study used a dataset, which is collected from a DC micro grid simulated in MATLAB environment.

Even though the proposed procedure was able to accurately detect FDIA with the true positive rate of 93.76%, other evaluation metrics like recall, accuracy and especially undetected rate needs to be evaluated to fully evaluate the performance of the proposed method. Furthermore, complexity of the ensemble tree classifier dictates the use of simple ML techniques like DT to reduce the computational complexity of the algorithm. The use of the Predictor Importance (PI) feature of DT could be helpful in reducing the complexity of the model and identify the features that are more informative for the detection. This will avoid the model overfitting and hence the degradation in the detection accuracy.

In [32], authors stressed on the importance of cyber security in the processors of charging devices. The core chips used in charging devices are imported chips and have hidden vulnerabilities which could cause the disclosure of customer private information, occupy the power grid, altering the customer billing information thus compromising the customer trust and the power grid operation. There is an urgent need to detect these attacks because network-based attacks are increasing with the passage of time. The charging stations which are at public sight are highly vulnerable to these types of attacks.

Charging devices are connected to the monitoring system which further connects to the power grid to provide energy to the EVs. During the charging process, the charging device monitors the battery level of EVs and state monitoring data, forwarding that information to the state level monitoring system for storing the user information for further use and for billing purposes. Network attack detection model was presented which used autoencoders for the detection of the cyber attacks by the real time monitoring of data streams. During normal operation, data stream is forwarded to the station level monitoring system however, during abnormal operation, malicious activity is detected in real time by the network attack detection model and alarm log is forward to the charging device controller, who takes further action to stop the charging process to control the attack propagation to the upper layer i.e., power grid. There are two parts of autoencoder: the encoder and the decoder. Encoder is responsible for mapping the input layer to the hidden layers for extracting new features for detection and decoder maps the hidden layer back to the original input and provides the probability of the attack.

The limitation of this paper is the lack of experimentation to evaluate the performance of the detection model. Testing needs to be done and by using different evaluation metrics the performance of the model can be calculated.

Table 2.1 shows Literature Review Table of Cyber Attacks in Fast Charging Stations.

Table 2.1: Literature Review Table of Cyber Attacks in Fast Charging Stations.

#	Targeted Area			Detection Technique		Accuracy /Measures for cyber attack detection	Early Detection
	Isolated Microgrid with V2G	Grid connected Microgrid with V2G	Non-microgrid	Machine Learning	Non-Machine Learning		
[3]	✗	✓	✗	✗	✗	✗	✗
[13]	✗	✗	✓	✗	✗	✗	✗
[30]	✗	✗	✓	✗	✗	✗	✗
[7]	✗	✗	✓	✗	✓	✗	✗
[5]	✗	✗	✓	✗	✓	✓	✗
[31]	✗	✓	✗	✗	✓	✓	✗
[32]	✗	✗	✓	✗	✓	✗	✗

✓ -Performed ✗-**not performed**

2.4. Research Gaps

In this section, the research gaps are listed based on the literature review.

- The literature review has revealed the dire need for an approach that can provide early detection of the cyber-physical attacks in FCSs. This is extremely important in the microgrids equipped with renewable DERs and when EVs in V2G mode are used in providing ancillary services to support the operation.
- The literature review has also revealed the need for a ML approach that has less computational complexity compared to the existing techniques while being effective in detecting such cyber-physical attacks in FCSs.
- There is a need to evaluate the detection accuracy when considering different time resolutions from the smart metering infrastructure data as well as the types of DoS attacks targeting the FCSs.

2.5. Summary

This chapter summarizes the previously published literature related to the cyber-physical security of the power system along with the security of FCSs and the various approaches proposed to address this problem. Initially, the chapter determined the research in the field of cyber security in the Industrial and control systems, then it discussed the research contribution for securing substation automation and control system and also the cyber attack detection on state estimation. Secondly, the chapter presented a summary of the outcome of several studies that looked at EVs park as protection techniques against cyber attacks as well as other studies that looked at the impact of cyber-attacks on the power grid. The studies that introduced detection approaches of cyber-attacks has been reviewed while highlighting the main outcomes of the studies. Furthermore, the

limitations of these studies were presented and discussed. Lastly, the research gaps were identified and the procedure to fill those gaps will be addressed in the proposed work of this thesis.

3. Intrusion Detection Model for the Detection of Cyber Attacks in Fast Charging Stations

3.1. Introduction

Intrusion is defined as a compromise to the integrity, availability, and confidentiality of a computer resource, thus jeopardising the cyber security [33]. The Intrusion Detection Model (IDM) plays a critical role in identifying those intrusions and securing the Confidentiality, Integrity, Availability (CIA) triad of cyber networks. IDM is a defense mechanism, which monitors the activities of the network and reports in response to the malicious activity in the network to the network operator [34]. There are three different types of intrusion detection methods: 1) signature-based intrusion detection method; 2) anomaly-based intrusion detection method; and 3) hybrid-based intrusion detection method.

The signature-based intrusion detection method relies on an existing datastore of previously captured attacks to detect the intrusions while the anomaly-based intrusion detection method helps to identify malicious behaviour within the network. The hybrid-based intrusion detection method is the combination of both signatures-based and anomaly-based detection method. The drawback of the signature-based intrusion detection method is that it cannot detect novel attacks due to the lack of a database [35].

In this work, anomaly-based intrusion detection method is proposed. Anomaly detectors create the profiles/patterns that represent normal behaviour, and anomaly detection method analyse those patterns and if there is any deviation from the normal pattern, anomaly is detected. Data mining concept is incorporated in an IDM to identify the intrusions in the system effectively and with less computational time. General intrusion detection process is explained in Fig 3.1. In the first step of intrusion detection, the procedure for the collection of the detection dataset is shown

along with the data preprocessing step. This chapter presents a complete intrusion detection algorithm that uses a ML classifier, DT for classification learning, and k-fold Cross Validation Method (CVM) to test the effectiveness of the proposed method. Finally, an evaluation metric for the analysis of the results will be discussed.

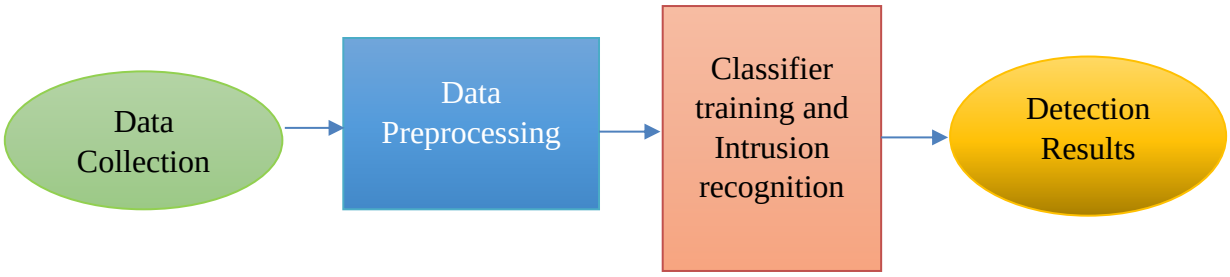

Fig. 3.1. Intrusion detection process [33].

3.2. Data Collection for the Microgrid Test System

The MATLAB SIMSCAPE model, "Power_V2G," depicted in Fig. 3.2, is used as the test model in this study and hence is used for the data collection. The microgrid is operating in standalone mode and includes a V2G system, which helps to regulate the frequency on the microgrid in the time of occurrence of an event and to support the variability in the wind farm and solar photovoltaic. In the "Power_V2G" model, the EVs in the microgrid assist to regulate the frequency of the system at the time when generation is less than demand or vice-versa [36]. Even though the EVs can help maintaining the operation of the microgrids, it is necessary to investigate the impact of the EVs charging/discharging on the microgrid operation if FCS are experiencing a cyber-physical attack. The proposed intrusion detection approach in this thesis aims to early detect the cyber attacks to the charging stations and hence provide enough time to the electric utilities before the system components physically experience the detrimental consequences of such cyber attacks.

The main components of the microgrid include a diesel generator, a wind farm, a solar photovoltaic (PV) farm, Load, V2G system, and the step-down transformer. The following represents a detailed description of each component.

Diesel Generator; The microgrid includes a diesel generator with the capacity of 16 MW. The diesel generator acts as a base power source and balances the power consumed and the power generated. The rotor speed of the diesel generator helps to determine the frequency deviation. The microgrid is in isolated mode and not connected to the main power grid. Therefore, the diesel generator is the main contributor of power generation as wind and solar power can fluctuate throughout the day since the power generation from wind and solar is not certain.

The microgrid includes two renewable resources: *solar farm* and *wind farm*. The solar and the wind are the most popular renewable energy resources currently used for the generation of clean energy. The amount of solar energy reaching the earth for an hour is more than enough to fulfil the power demand of earth for a whole year [37]. The power generation capacity for the solar farm is 8 MW. As solar irradiance fluctuates throughout the day and there is no sunlight at night, therefore, other generation units will need to fulfill the demand at the time when there is not enough solar generation.

Wind is a plentiful source of clean energy. According to the power of wind equation proposed in (3.2), wind power has a direct relation to the wind speed [38].

$$P = {}^{1}/_{2}\,\rho A v^{3} \tag{3.2}$$

Where P is wind power, ρ air density, A is rotor swept area and v is the wind speed.

The nominal wind power in the microgrid model is 4.5 MW with the nominal speed of 13.5 m/s and the maximum wind speed for wind farm is 15 m/s. If the wind speed is greater than the maximum wind speed of the wind farm, the wind farm will trip, and the power generation from

the wind farm resumes once the wind speed is in nominal wind speed range. The wind power is fixed to 1 per unit (p.u.), when the wind speed is between the nominal and the maximum wind speed.

There are two types of *loads* in the microgrids other than the EVs: a residential load and an industrial load. The residential load is 10 MW while the industrial load is 0.16 MVA. The residential load represents the community of one thousand houses and the timestep used for residential load consumption data is 60 min.

The model includes a *V2G system* consisting of 100 EVs. The rated power of fast charger is 40 kW [36].Specifications of V2G system is shown in Table 3.1. According to the microgrid model, the ratio between the EVs and household is 1:10, which means there is one EV in every tenth house. The V2G system performs two functions:

1. The EVs charging function to charge the EV batteries.

2. When connected to the system, short-term peak regulation functions at the time of the event occurrence.

Table 3.1: V2G system specifications.

Rated power of fast chargers	40 kW [36]
Rated capacity of EV battery/ EV battery size	85 kWh [39]
Charging rate	127.5 min to fully charge
EV Model	Tesla Model S [39]
EV Range	407-434.5 km [39]
Model Efficiency	90% [36]

The description of different types of vehicle's user profile is as follows:

Profile 1:

This profile includes the EV users who can charge their car at work.

Profile 2:

This profile includes the EV users who do not have the opportunity to charge their car at work.

Profile 3:

This includes the EV users who can charge their car at work but with a longer ride.

Profile 4:

This profile includes EV users who stay at home.

Profile 5:

This profile includes EV users who work on a night shift.

A three-phase (three single phase) step-down transformer with the nominal power of 20 MVA and a frequency of 60 Hz is used in the microgrid. The distribution transformer is used to step down the voltage from the primary voltage of 25 kV to the secondary voltage of 600V in the distribution lines feeding the customers. Primary and secondary side winding connection is "Wye to ground."

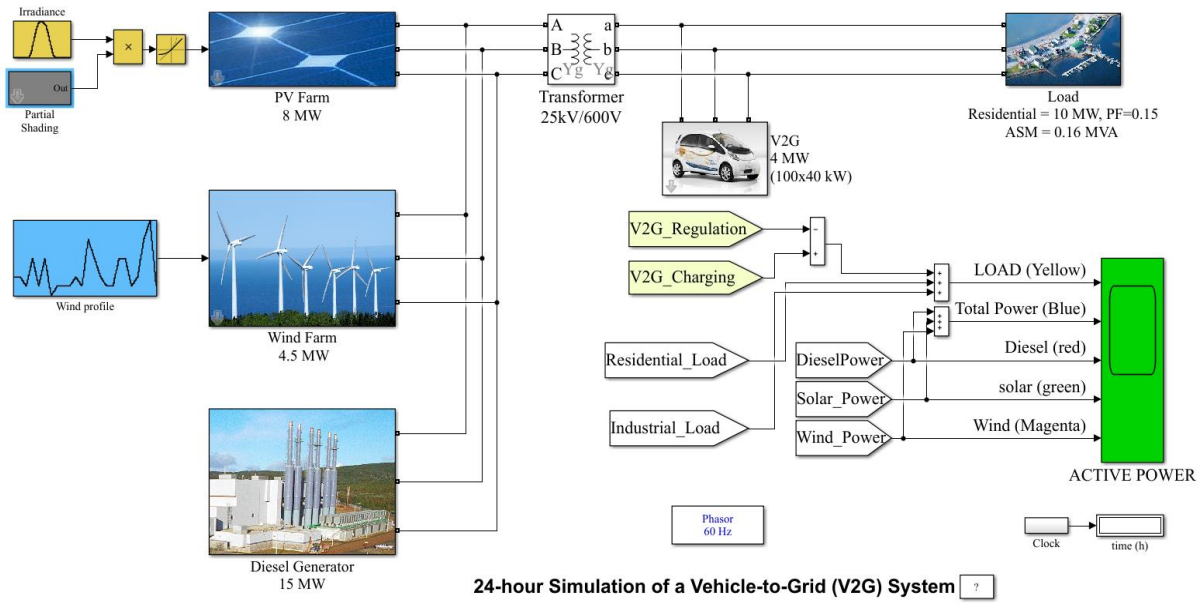

Fig. 3.2. Test system model "Power_V2G" [36].

3.3. Dataset Generation for the Cyber-Physical Attacks

To test the performance of the proposed approach for the detection of cyber attacks, a dataset is necessary for the training and the testing of the IDM. To the best knowledge of the author, there is no available dataset that can demonstrate the performance of the microgrids with EVs when FCS are under the DoS attack. With that in mind, it was necessary to create a dataset when FCS are under attack in microgrids. This allows to capture the variations in power demand of the transformer, which will be used to evaluate the performance of the proposed detection approach. The subsequent sections demonstrate how the simulations are performed in MATLAB to generate the dataset. This simulation examines a series of different scenarios corresponding to different combinations of vehicles profiles when the charging stations are under attacks and no attack. Consequently, the power demand profiles of the transformer are extracted for each scenario.

3.3.1. Procedure for the Generation of the Dataset

The following steps were used to generate the dataset:

1. The Canadian Weather Energy and Engineering Dataset (CWEEDS) [40] contains climate dataset for 564 Canadian locations. For most of the locations, the dataset period is from 1998 – 2017. The dataset provides information about temperature, dew point, wind speed and direction, solar irradiance and so on. The dataset has the hourly observation, 24/7 for the whole year. Interpolation was used to fill the missing observation gaps of 3 hours or less. For the gaps of more than 3 hours, the missing values are filled in an hourly time series file from the nearest North America Regional Reanalysis (NARR) grid point. For further readings on the climate dataset, the reader is referred to [40].

   To have a realistic analysis, a case study was chosen, where the actual wind speed and the actual solar irradiance data of BC was collected from CWEEDS and updated in the test microgrid model. The reason to choose BC dataset was the popularity of the EVs in BC. According to the survey done in Feb 2021, by Klynveld Peat Marwick Goerdeler (KPMG) which is an audit, tax and advisory firm, 70% of the Canadians would prefer to buy an EV for their next new vehicle purchase. The province wise popularity level of EVs among Canadians is the highest among BC (77%) followed by Quebec (75%), Alberta (54%), Prairie (48%) and Atlantic (55%) [41].
2. The downloaded dataset in CWEEDS website was in ".WY3" format. The dataset was converted in ".csv" file and the latest dataset, which is of year 2017 data, (the most recent data available) was captured for simulations.

3. To have diversity in the dataset, two cities of BC were chosen in this work; the city of Vancouver was chosen to represent as a large city while Abbotsford was chosen as a small city.
4. In order to study the effect of the solar and the wind profiles on the cyber-physical attacks, three days were chosen from the year 2017. A day when the solar irradiance is the highest, and the wind speed is high; a day when the solar irradiance is the lowest and the wind speed is low; a day when the wind speed and solar irradiance is close to the average wind speed and solar irradiance of the year respectively.
5. Different scenarios were implemented after updating the wind and the solar data in the microgrid for each day and for both cities. Cyber attacks were introduced with 100 EVs. Dataset was collected for all three types of attacks and no attack.
6. In order to study the effect of increasing the EVs penetration on the cyber-physical attacks, the number of EVs were increased from 100 to 125, then 150, 175 and 200. With every increment of EVs, the dataset with and without cyber attacks were generated.

3.3.2. Description of Scenarios

The Power_V2G model is used for analyzing the impact of the cyber attacks on the FCS. As mentioned earlier, there are two renewable resources: PV farm and Wind farm. The profiles of such resources were chosen from three days of the year of the two cities (Vancouver and Abbotsford) in BC. The following are the days chosen for which the wind and the solar profiles were extracted from the CWEEDS:

- **For Vancouver:**

Day 1: Wind speed is high and solar irradiance is the highest in Vancouver.

Day 2: Wind speed and solar irradiance is average in Vancouver.

Day 3: Wind speed is low and solar irradiance is the lowest in Vancouver.

- **For Abbotsford:**

Day 1: Wind speed is high and solar irradiance is the highest in Abbotsford.

Day 2: Wind speed and solar irradiance is average in Abbotsford.

Day 3: Wind speed is low and solar irradiance is the lowest in Abbotsford.

Furthermore, for each day different EVs penetration levels were tested. The penetration levels are as follows: 25%, 50%,75% and 100%.

1. **Step 1: Vancouver: For Day 1 Wind speed and Solar Irradiance Profile**

With 100 cars in Vancouver: The distribution of cars in each profile are as follows [30, 20, 20, 10, 10] [42].

- 30 represents the percentage number of vehicles in profile 1.
- 20 represents the percentage number of vehicles in profile 2.
- 20 represents the percentage number of vehicles in profile 3.
- 10 represents the percentage number of vehicles in profile 4.
- 10 represents the percentage number of vehicles in profile 5.

Following the process of importing the data for wind, solar and vehicles distribution for each profile, the simulations were performed in MATLAB SIMSCAPE considering three types of attacks and no attack. The effect of EVs penetration has been incorporated in this study with the following penetration levels:

- **Percentage EV penetration of 125%:** For this case, the number of vehicles has increased by 25% in each profile.

The same process will be done as in step 1 but with 25% increase of EVs in each vehicle profile.

- **Percentage EV penetration of 150%:** For this case, the number of vehicles has increased 50% in each profile.

The same process will be done as in step 1 but with 50% increase of EVs in each vehicle profile.

- **Percentage EV penetration of 175%:** For this case, the number of vehicles has increased 75% in each profile.

The same process will be done as in step 1 but with 75% increase of EVs in each vehicle profile.

- **Percentage EV penetration of 200%:** For this case, the number of vehicles has increased 100% in each profile.

The same process will be done as in step 1 but with 100% increase of EVs in each vehicle profile.

At the end of Step 1, there are 20 simulation results for day 1 of Vancouver.

- 5 scenarios result without attack.
- 5 scenarios result with attack 1.
- 5 scenarios result with attack 2.
- 5 scenarios result with attack 3.

2. **Step 2: Vancouver: For Day 2 Wind Speed and Solar Irradiance Profile**

The same procedure will be done as in step 1. Although the procedure is the same as in Step 1, but the wind and solar profiles are according to day 2 of Vancouver, when wind speed and solar irradiance is average.

3. **Step 3: Vancouver: For Day 3 Wind Speed and Solar Irradiance Profile**

The same procedure will be done as in step 1. Although the procedure is the same as in step 1, but the wind and the solar profiles are according to the day 3 of Vancouver, when wind speed is low, and the solar irradiance is the lowest.

4. **Step 4: Abbotsford: For Day 1 Wind Speed and Solar Irradiance Profile**

With 100 cars in Abbotsford: The distribution of cars in each profile is as follows [35, 25, 10, 20, 10] [42]

- 35 represents the percentage number of vehicles in profile 1.
- 25 represents the percentage number of vehicles in profile 2.
- 10 represents the percentage number of vehicles in profile 3.
- 20 represents the percentage number of vehicles in profile 4.
- 10 represents the percentage number of vehicles in profile 5.

Following the process of importing the data for wind, solar and vehicles distribution for each profile, the simulations were performed in MATLAB SIMSCAPE considering three types of attacks and no attack. The effect of EVs penetration has been incorporated in this study with the following penetration levels:

- **Percentage EV penetration of 125%:** For this case, the number of cars has increased by 25% in each profile.

The same process will be done as in step 4 but with 25% increase of EV in each vehicle profile.

- **Percentage EV penetration of 150%:** For this case, the number of vehicles has increased 50% in each profile.

The same process will be done as in step 4 but with 50% increase of EVs in each vehicle profile.

- **Percentage EV penetration of 175%:** For this case, the number of vehicles has increased 75% in each profile.

The same process will be done as in step 4 but with 75% increase of EVs in each vehicle profile.

- **Percentage EV penetration of 200%:** For this case, the number of vehicles has increased 100% in each profile.

The same process will be done as in step 4 but with 100% increase of EVs in each vehicle profile.

At the end of Step 4, there will be 20 simulation results for day 1 of Abbotsford.

- 5 Scenarios result without attack.
- 5 Scenarios result with attack 1.
- 5 Scenarios result with attack 2.
- 5 Scenarios result with attack3.

5. Step 5: Abbotsford: For Day 2 Wind Speed and Solar Irradiance Profile

Same procedure will be done as in Step 4. Although the procedure is the same as in Step 4, but the wind and the solar profiles are according to the day 2 of Abbotsford, when wind speed and solar irradiance is average.

6. Step 6: Abbotsford: For Day 3 Wind Speed and Solar Irradiance Profile

Same procedure will be done as in Step 4. Although the procedure is the same as in Step 4, but the wind and the solar profiles are according to day 3 of Abbotsford, when wind speed is low, and the solar irradiance is the lowest.

Table 3.2 shows the total number of scenarios in each step from Step 1 to Step 6. Moreover, Table 3.3 shows the total number of scenarios for attack 1 detection in each step from Step 1 to Step 6. Furthermore, Table 3.4 and Table 3.5 shows the total number of scenarios for attack 2 and attack 3 detection in each step from Step 1 to Step 6.

Table 3.2: Total number of the scenarios in each step.

Step number	Scenario Type	Total Number of Scenarios/ Simulations
1	Attack 1 (5 Scenarios), Attack 2 (5 Scenarios), Attack 3 (5 Scenarios), No Attack (5 Scenarios).	20
2	Attack 1 (5 Scenarios), Attack 2 (5 Scenarios), Attack 3 (5 Scenarios), No Attack (5 Scenarios).	20
3	Attack 1 (5 Scenarios), Attack 2 (5 Scenarios), Attack 3 (5 Scenarios), No Attack (5 Scenarios).	20
4	Attack 1 (5 Scenarios), Attack 2 (5 Scenarios), Attack 3 (5 Scenarios), No Attack (5 Scenarios).	20
5	Attack 1 (5 Scenarios), Attack 2 (5 Scenarios), Attack 3 (5 Scenarios), No Attack (5 Scenarios).	20
6	Attack 1 (5 Scenarios), Attack 2 (5 Scenarios), Attack 3 (5 Scenarios), No Attack (5 Scenarios).	20
Total		120

Next step was to partition the dataset into records with and without attacks. As the DT is a binary classification algorithm, therefore, there should be just two classes. Also, from ML literature, detecting a cyber attack is a binary classification problem [5]. Keeping this in mind, the above attack dataset was refined into three subsets as follows:

Table 3.3: Subset 1: Total number of the scenarios for attack 1 and no attack in each step.

Step number	Scenario Type	Total Number of Scenarios/ Simulations
1	Attack 1 (5 Scenarios), No Attack (5 Scenarios).	10
2	Attack 1 (5 Scenarios), No Attack (5 Scenarios).	10
3	Attack 1 (5 Scenarios), No Attack (5 Scenarios).	10
4	Attack 1 (5 Scenarios), No Attack (5 Scenarios).	10
5	Attack 1 (5 Scenarios), No Attack (5 Scenarios).	10
6	Attack 1 (5 Scenarios), No Attack (5 Scenarios).	10
Total		60

Table 3.4: Subset 2: Total number of the scenarios for attack 2 and no attack in each step in each step.

Step number	Scenario Type	Total Number of Scenarios/ Simulations
1	Attack 2 (5 Scenarios), No Attack (5 Scenarios).	10
2	Attack 2 (5 Scenarios), No Attack (5 Scenarios).	10
3	Attack 2 (5 Scenarios), No Attack (5 Scenarios).	10
4	Attack 2 (5 Scenarios), No Attack (5 Scenarios).	10
5	Attack 2 (5 Scenarios), No Attack (5 Scenarios).	10
6	Attack 2 (5 Scenarios), No Attack (5 Scenarios).	10
Total		60

Table 3.5: Subset 3: Total number of the scenarios for attack 3 and no attack in each step.

Step number	Scenario Type	Total Number of Scenarios/ Simulations
1	Attack 3 (5 Scenarios), No Attack (5 Scenarios).	10
2	Attack 3 (5 Scenarios), No Attack (5 Scenarios).	10
3	Attack 3 (5 Scenarios), No Attack (5 Scenarios).	10
4	Attack 3 (5 Scenarios), No Attack (5 Scenarios).	10
5	Attack 3 (5 Scenarios), No Attack (5 Scenarios).	10
6	Attack 3 (5 Scenarios), No Attack (5 Scenarios).	10
Total		60

It is worth noting that when the whole dataset was divided into three subsets, the total number of simulations came out to be 180. The reason to have three subsets from the whole dataset of attacks and no attack scenarios is because the purpose is to let the DT discriminate between the normal and the malicious scenarios. In all three subsets, the no attack scenarios are common and the same as they are normal and expected scenarios.

3.4. Data Preprocessing

Following the data collection step, the data preprocessing step is performed for the detection purpose. Initially, in the data preprocessing, the generated dataset from the microgrid will be down sampled to obtain samples for different time resolutions. The time resolutions chosen were hourly, half hourly and quarter hourly which is typical time resolution of digital meters at the substation [43]. In the subsequent step, the Rate of Change (ROC) will be calculated from 6:00 to

10:00 for each time resolution as a separate dataset for the testing and the evaluation. The duration of the attack is from 8:00 to 18:00. The complete description of the attacks and the reason to choose detection duration from 6:00 till 10:00 is provided in Chapter 4.

The ROC values are calculated from 6:00 to 10:00 because the attack is happening at 8:00 when people are going to work and need to charge their vehicles. The EV users prefer to charge their vehicles from 8:00 till almost 9:00. If they can charge their vehicles during their preferred time, and the batteries are fully charged by almost 9:00, any attack after 9:00 would not impact the distribution transformer because EVs are fully charged and there will not be load demand from EVs after 9:00.

The transformer is the only link between the generation and the distribution. Any change in the power supply pattern can be seen at the transformer. Therefore, the transformer power was captured for the detection of cyber attacks. Other features such as the frequency and the voltage at the transformer may be used for the detection of the attacks. Since the transformer is the most impacted asset in the system, the power demand profile of the transformer in Fig 4.1 is used for the feature extraction process. Specifically, the changes in the transformer demand power from 6:00 to 10:00 are used to extract the features relevant to the attacks since during this period the EVs are charged at work. The hourly change in the transformer power demand from 6:00 to 10:00 and how the corresponding ROC values are calculated is depicted in Fig.3.3 and is listed in Table.3.6. Other time resolutions such as the half-hourly and the quarter-hourly ROC of the transformer power demand are incorporated in this work. These ROC values were imported in the intrusion detection model for the detection and the corresponding results are shown in Chapter 5.

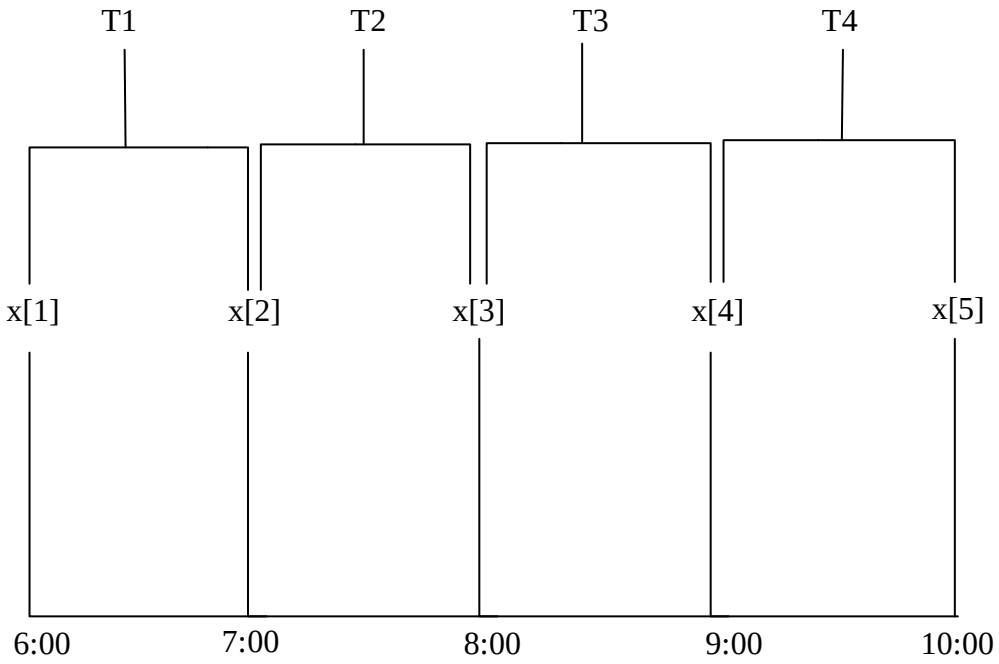

Fig. 3.3. Sample collection for the hourly time resolution.

Table 3.6: Description of sample and sequence difference at the hourly-time resolution.

Sample Point	Description of Sample Point	Sequence Difference
x[1]	PT at 6:00	
x[2]	PT at 7:00	T1=x[2]-x[1]
x[3]	PT at 8:00	T2=x[3]-x[2]
x[4]	PT at 9:00	T3=x[4]-x[3]
x[5]	PT at 10:00	T4=x[5]-x[4]

The half-hourly time resolution and how the corresponding ROC values are calculated for the half-hourly time resolution is described in Fig.3.4 and is listed in Table. 3.7. These ROC values were imported in the IDM for the detection and the corresponding results are shown in Chapter 5.

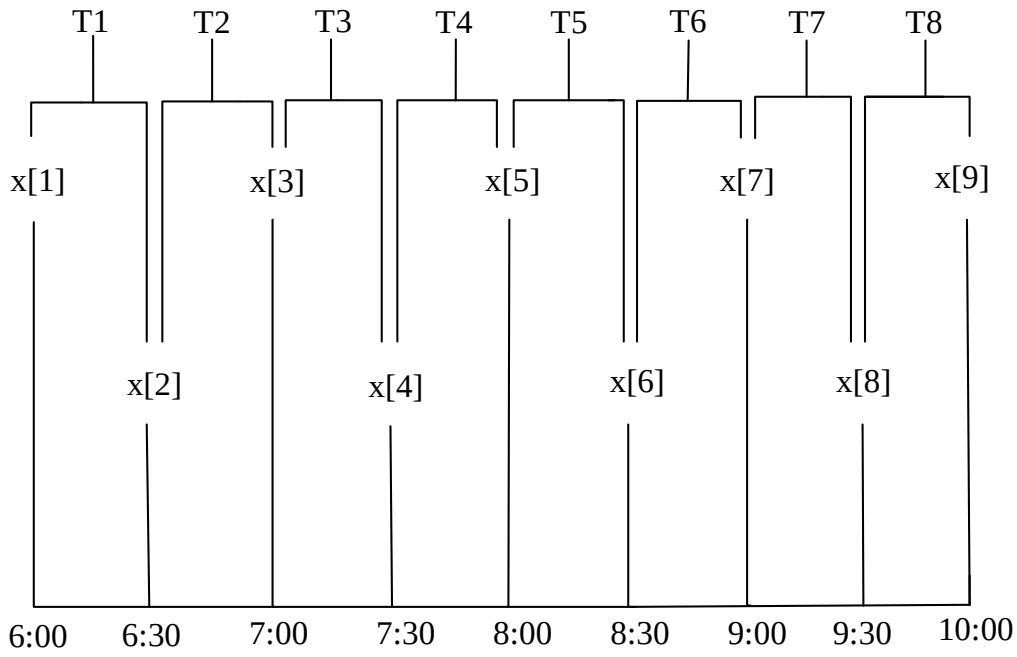

Fig. 3.4. Sample collection for the half-hourly time resolution.

Table 3.7: Description of sample and sequence difference at the half-hourly time resolution.

Sample Point	Description of Sample Point	Sequence Difference
x[1]	PT at 6:00	
x[2]	PT at 6:30	T1=x[2]-x[1]
x[3]	PT at 7:00	T2=x[3]-x[2]
x[4]	PT at 7:30	T3=x[4]-x[3]
x[5]	PT at 8:00	T4=x[5]-x[4]
x[6]	PT at 8:30	T5=x[6]-x[5]
x[7]	PT at 9:00	T6=x[7]-x[6]
x[8]	PT at 9:30	T7=x[8]-x[7]
x[9]	PT at 10:00	T8=x[9]-x[8]

Finally, the quarter-hourly time resolution and how the corresponding ROC values are calculated is described in Fig. 3.5. and is listed Table.3.8. After the sequence subtraction for three-time resolutions, the DT classifier along with the CVM was used for the detection and the model performance evaluation. These ROC values were imported in the intrusion detection model for the detection and the corresponding results are shown in Chapter 5.

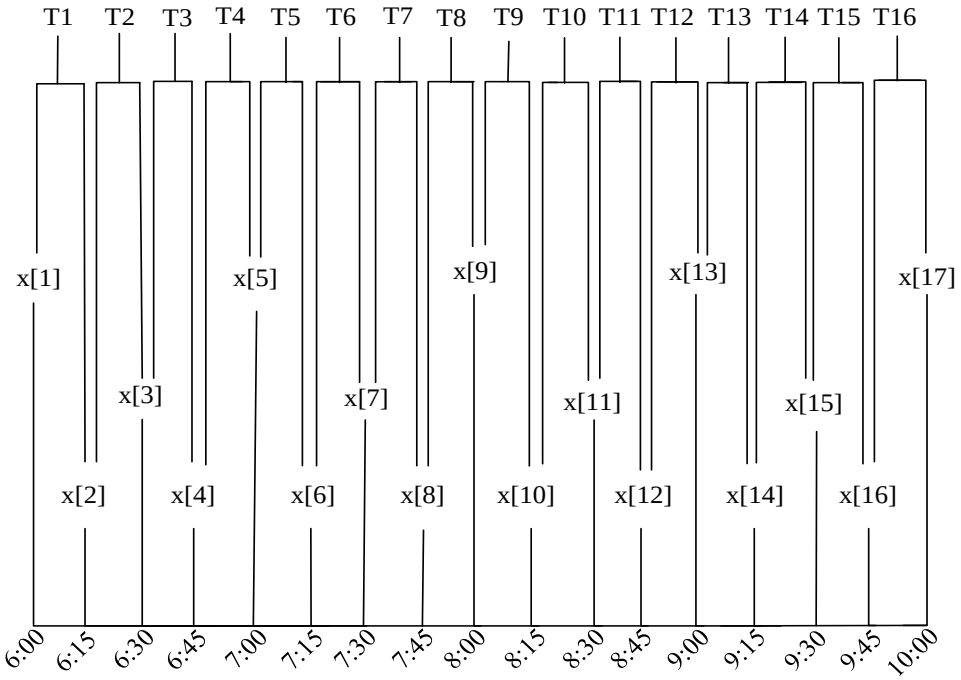

Fig. 3.5. Sample collection for the quarter-hourly time resolution**.**

Table 3.8: Description of sample and sequence difference at the quarter-hourly time resolution.

Sample Point	Description of Sample Point	Sequence Difference
x[1]	PT at 6:00	
x[2]	PT at 6:15	T1=x[2]-x[1]
x[3]	PT at 6:30	T2=x[3]-x[2]
x[4]	PT at 6:45	T3=x[4]-x[3]
x[5]	PT at 7:00	T4=x[5]-x[4]
x[6]	PT at 7:15	T5=x[6]-x[5]
x[7]	PT at 7:30	T6=x[7]-x[6]
x[8]	PT at 7:45	T7=x[8]-x[7]
x[9]	PT at 8:00	T8=x[9]-x[8]
x[10]	PT at 8:15	T9=x[10]-x[9]
x[11]	PT at 8:30	T10=x[11]-x[10]
x[12]	PT at 8:45	T11=x[12]-x[11]
x[13]	PT at 9:00	T12=x[13]-x[12]
x[14]	PT at 9:15	T13=x[14]-x[13]
x[15]	PT at 9:30	T14=x[15]-x[14]
x[16]	PT at 9:45	T15=x[16]-x[15]
x[17]	PT at 10:00	T16=x[17]-x[16]

3.5. Decision Tree-based Machine Learning Technique for the Intrusion Detection Model

In this section, the importance of using DT algorithms as opposed to other algorithms is explained in terms of comparison of various ML algorithms for the proposed IDM. The complete description of DT for the classification and its implementation into the proposed approach along with the parameter settings for the optimal accuracy is detailed in this section.

Cyber security has become essential in every field due to the dependency of the world on the internet. Cyber security includes techniques or methods, which will help to defend against cyber attacks and to secure the computer networks. The ML has been used for the detection of cyber attacks in different fields due to their capability of learning from experience without the need of programming [35]**.** Recently the popularity of the ML techniques, especially the DT, has increased for anomaly-based IDM because of their higher detection accuracy [44]. The objective of the ML classification algorithms is to accurately predict the class labels of the observations whose attribute values are known. ML, a subset of artificial intelligence, has numerous applications especially in the field of predictive data mining [45].

There are two types of ML algorithms: Supervised and Unsupervised ML algorithms. In Supervised ML, algorithms are trained using labelled dataset. A classifier builds a model based on the input dataset for classification, the most widely used classification algorithms include: SVM, DT, Artificial Neural Network and Naïve Bayes [23]. The unsupervised ML algorithms handle unlabeled dataset, use large amounts of dataset for training and are less accurate as compared to the supervised learning [33]**.** Unsupervised ML algorithms are used for the exploratory data analysis. The purpose of these algorithms includes grouping/clustering, interpretation, and visualization. The k-mean and Deep Belief Method (DBM) are the examples of unsupervised ML algorithms [46].

For intrusion detection, a ML algorithm is required which is easy to understand, has less computational and memory requirements, capable of detecting anomalies with high accuracy. The DT algorithms are well-known for extracting the classification rules from the data. The reasons behind their recognition and broad implementation includes their comparably higher or equal accuracy as compared to the accuracy of other classification models, does not need many

parameters to be adjusted in their design and last but not the least, the property of DT for easy comprehension. All these advantages have made DT a popular tool for many domains [47].

The limitations of different Supervised ML algorithms for classification purposes are listed as follows:

- Support Vector Machine (SVM) algorithm is unable to handle large or noisy dataset efficiently. The major drawback of the SVM is that it consumes an immense amount of space and time. SVM requires data trained on different time intervals to produce better results for a dynamic dataset [35].
- Naive Bayes algorithms assign 0 probability if some category in the test data set is not present in the training data set. The major limitation for Naïve Bayes classifier is that it assumes that every attribute is independent, and none of the attributes has a relationship with each other. This state of independence is technically impossible in cyberspace [35].
- Random forest algorithms have high computational cost and slow prediction generators [35].
- Artificial neural network algorithms have high cost and are time consuming [35].

A supervised ML algorithm, DT, will be used for IDM which is a simple yet broadly used classification method [23]. The DT provides a set of rules that are easy to understand and can be easily integrated with real-time technologies, which makes DT a desired classification algorithm for anomaly detection [34]. As one of the exceptional classification algorithms, DT constructs a tree, which contains crucial information to make predictions. The main advantage of DT is its visualization function, which helps to understand the operation and the assessment of DT [46]. Furthermore, DT can easily handle categorical and numerical variables in the same algorithm,

unlike many other data mining algorithms, such as SVM, and Neural Networks [48]. While classification techniques such as SVM and Bayesian classifiers might be utilized, it is desirable to use a deterministic classification technique like DT instead of other non-deterministic techniques such as SVM or Bayesian since the aim of this work is to identify attacks with the higher accuracy. The SVM utilizes optimization theory and the Bayesian relies on the probability theory. Both SVM and Naïve Bayes are non-deterministic techniques, which means that they can give multiple solutions to the same input data [23]. Moreover, the effectiveness of using the DT algorithm for intrusion detection can also be investigated in this chapter.

Each ML technique has some advantages and disadvantages. Even though ML has been used for cybersecurity, the selection of a specific ML technique for a specific field is still tricky. The DT constructs a training model based on the decision rules derived from the labeled input data to predict the class labels. The DT with a properly classified dataset and minimum number of nodes is the optimal DT. The DT is a powerful model, has numerous applications due to the simple analysis and precision, and is mostly utilized for the classification in data mining. The most used DT algorithms are Classification and Regression Tree CART (CART), Iterative Dichotomies 3 (ID3), C4.5 (Successor of ID3), CHi-squared Automatic Interaction Detector (CHAID), Unbiased and Efficient Statistical Tree (QUEST) [45]. In [45], the researchers have done a detailed literature review on the application of DT in ML and data mining tasks. According to that study, the DT showed the highest accuracy of 99.93% as compared to KNN, SVM, and Naïve Bayes algorithm. The Hunt's algorithm is the basis of the existing DT induction algorithms such as ID3, C4.5 and CART. The CART algorithm, which is used in the proposed IDM produces binary split [23].

The purpose of the DT is to create a model that predicts the targeted class label for a new test instance based on the provided input attributes [34]. A DT is a tree-based technique in which

the root node is divided into decision nodes, which are further divided to give the leaf node. Root node includes the input attributes, and the leaf node includes the class labels [45]. During the DT learning process, the decision nodes further split to form sub-trees based on the decision node attribute value, and this process is repeated until the stopping criteria are met. When growing a DT, it is necessary to choose the attribute for split with the highest information. The attribute selection is one of the essential factors that greatly impacts the efficiency of the model by selecting features for split that are important and informative for the detection process. There are several measures for choosing the best split, the most widely used includes entropy and Gini index [48]. The main issue in constructing a DT is deciding the attribute for optimal split**.** For the proposed CART DT algorithm, Gini Index is used as the split measure to choose the most appropriate splitting attribute for each node. Gini Index assists in choosing the most appropriate attribute for splitting at each node. It measures the node impurity, thus the attribute with the lowest Gini index is used for splitting. The pure node will have all the observations from the same class, thus the ideal node [23].

$$\text{Gini} = 1 - \sum_{i=1}^{k} (M_i)^2 \tag{3.3}$$

Where, M_i represents the portion of observations that belongs to a particular class k for a particular node.

Another commonly used splitting criterion is the Entropy index. It also calculates the node impurity for splitting. However, the entropy involves a log computation and Gini impurity involves a square computation. Computing the square is computationally inexpensive as compared to logarithmic function, therefore Gini impurity is preferred over entropy.

$$\text{Entropy} = -\sum_{i=1}^{k} (M_i) * \log_2 (M_i) \tag{3.4}$$

Where, M_i represents the portion of observations that belongs to a particular class for a particular node.

Another important characteristic provided by the DT is estimating the importance of predictor/feature for detection. Therefore, in this model, to obtain the highest accuracy for all three attacks at k-fold value where computational time will be less, the predictor importance function was used to determine the highest informative predictors, afterwards only the predictors with the highest importance were used for classification, which assisted to achieve the highest accuracy. The feature selection provides following advantages [49].

- Remove the irrelevant and redundant data.
- Decreases the computation time.
- Enhance the detection accuracy.
- Enhance the learning efficiency.

There are several feature selection algorithms: Filter Type Feature Selection algorithm, Wrapper Type Feature Selection algorithm, and Embedded Type Feature Selection algorithm. However, for the research problem in hand, the Embedded Type Feature Selection method is used and the Predictor Importance (PI) is calculated for determining the features with their importance in classification learning process. The Embedded approach is preferred while using DT algorithms as calculating the feature importance is the part of the model learning process. The predictor importance function determines the importance of a feature after the model is trained [50].

The PI for a classification tree is calculated by summing the difference between the risk due to the parent node and the total risk due to the children node on every predictor divided by the total number of branch nodes [51].

$$(Q\__1\text{-}Q\__2\text{-}Q_3)/N_B \quad (3.5)$$

Where, Q_1 is the node risk due to the parent node, Q_2 and Q_3 are the node risk of the two children node and N_B is the total number of branch nodes. A node risk is calculated by the node impurity multiplied by the node probability:

$$Qi = Pi * Ei, \quad (3.6)$$

Where, *Pi* is the node probability of node *i*, and *Ei* is node impurity, which is calculated by the Gini Index shown in (3.3) [51].

In summary, two steps were performed to obtain the optimal accuracy and less computational time for detection. Firstly, the quarter hourly-time resolution was used to calculate the sequence subtraction and secondly, based on the PI, the predictors with the highest importance were chosen for detection.

3.6. Cross Validation Method

Cross validation method is used to evaluate the performance of a classifier by making predictions on a new dataset, which was held out during the training. The dataset is divided into two subsets, the training subset is used for training the ML classifier while the testing subset is used for testing the performance of that classifier. The Cross validation plays a vital role in reducing the risk of overfitting and underfitting the model [52]. Several cross-validation methods that are used for performance evaluation of a classifier [23] are listed below. The reason to use the k-fold CVM in the proposed intrusion detection model is also explained below.

3.6.1. Holdout Method and its Limitations

The original labelled dataset is divided into separate training and testing dataset randomly. The model is trained on the training data and is then evaluated on the testing data set (Fig. 3.6). The partition of the training and the testing dataset is mostly 50%-50% or 70%-30% respectively. The Holdout method is usually preferred in case of a large dataset [23].

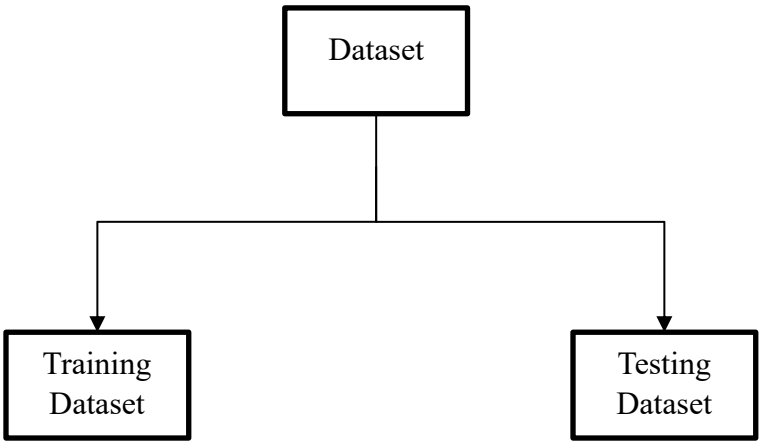

Fig. 3.6. Holdout validation method.

There are some limitations of using the holdout method. The observations in the testing set may include some important information for classification, which is absent in the training set. The holdout method cannot use the whole dataset for training therefore it may not be an appropriate method for a smaller dataset. Moreover, it is highly dependent on the composition of the training and the testing sets [23].

3.6.2. Random Subsampling and its Limitations

To overcome the limitations of the holdout method, the random subsampling is used, where the holdout method is repeated several times to improve the classifier performance. Therefore, in every iteration there is a possibility to choose every observation for testing and training [23].

The limitations of using the random subsampling are as follows: There is no control over the number of times an observation is used for testing or training. Also, some observations may be used for training often and some still may not be used for testing [23].

3.6.3. k-fold Cross Validation Method

In the k-fold CVM, the entire dataset is divided into "k" disjoint folds of approximately equal sizes. The "k-1" folds will be used for training and the rest of the fold will be used for testing. The same process will be repeated for "k" iterations. The entire dataset will be used for training and each fold will be used for testing once. Fig. 3.7 shows the 3-fold cross validation process. The error is calculated in each fold and the total error is calculated by summing up the errors for k iterations. The error is calculated at each fold as:

$$\varepsilon = \frac{1}{k}\sum_{i=1}^{k} \varepsilon_i \tag{3.7}$$

and the performance of the classifier is dependent on the average of k-fold accuracies (α) [23].

$$\alpha = 1 - \varepsilon \tag{3.8}$$

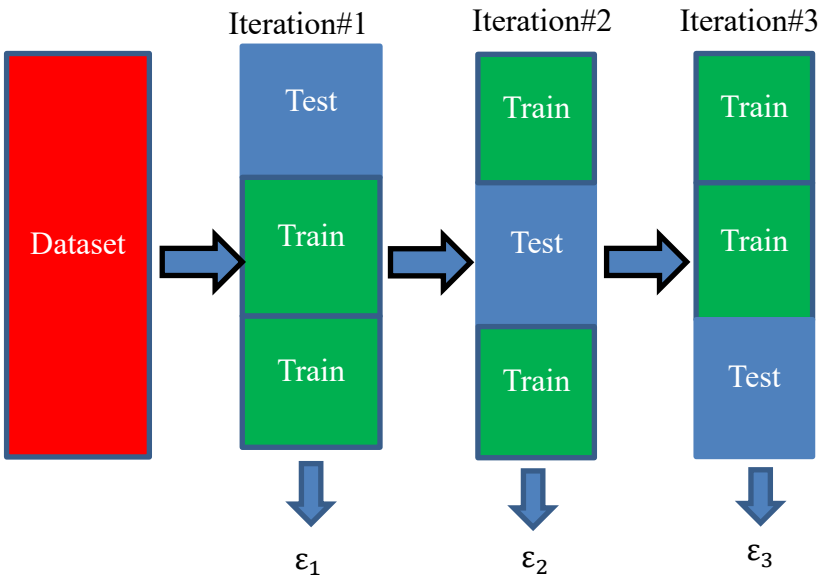

Fig.3.7. Representation of 3-fold cross validation method.

3.6.4. Leave One Out Cross Validation Method and its Limitations

Leave One Out Cross Validation Method (LOOCVM) is another type of CVM where the whole dataset is divided into k-folds and k=N where N is equal to the number of instances in the dataset. In LOOCVM, in every iteration there will be one instance used as testing and the rest as training. This method has an advantage to use maximum dataset for training and test instances are mutually exclusive.

The limitations of using the LOOCVM method are as follows: As the whole procedure will be repeated N times therefore it is computationally very expensive and has high variance [23].

3.7. Layout of the Proposed Intrusion Detection Model and the Intrusion Detection Process

The DT and the k-fold CVM are the most critical components of the proposed IDM. The power demand data of the transformer depicted in Fig 4.1 from 6:00 to 10:00 are used to calculate

the ROC of the transformer's power demand for different time resolutions. The ROC data was labeled either as 'no attack' or 'attack'. The ROC data are used as an input to the IDM for detection.

Originally, the hourly ROC was calculated using the transformer power demand data from 6:00 to 10:00 for which the DT algorithm build the detection rules while the CVM was used to evaluate the performance of the DT algorithm by computing the classification accuracy. The same procedures were applied for different time resolutions such as the half-hourly and the quarter-hourly ROC data.

Furthermore, using the quarter-hourly time resolution, based on the Embedded Feature Selection Method, the DT will prune some of the features that do not possess a great impact on the classification process and the most crucial subset of the attributes were calculated. Subsequently, the most important predictors selected by the DT were used in the IDM. Finally, the CVM is used and the evaluation metrics such as accuracy, F-score, precision, recall, and the undetected rate are presented.

Different scenarios are chosen for detection where the wind and solar profiles have different patterns as discussed in section 3.3.2. These scenarios help the decision tree to learn the detection pattern in the case; 1) when the wind speed is low and the solar irradiance is the lowest, 2) when the wind speed and solar irradiance are average, and 3) when the wind speed is high and the solar irradiance is the highest. The inclusion of these scenarios make the detection algorithm optimal for the detection because the sharp variability of the solar irradiance and the wind speed will not be detected as attack.

Figure 3.8 shows the steps for implementing the intrusion detection using different time resolutions of smart metering infrastructure.

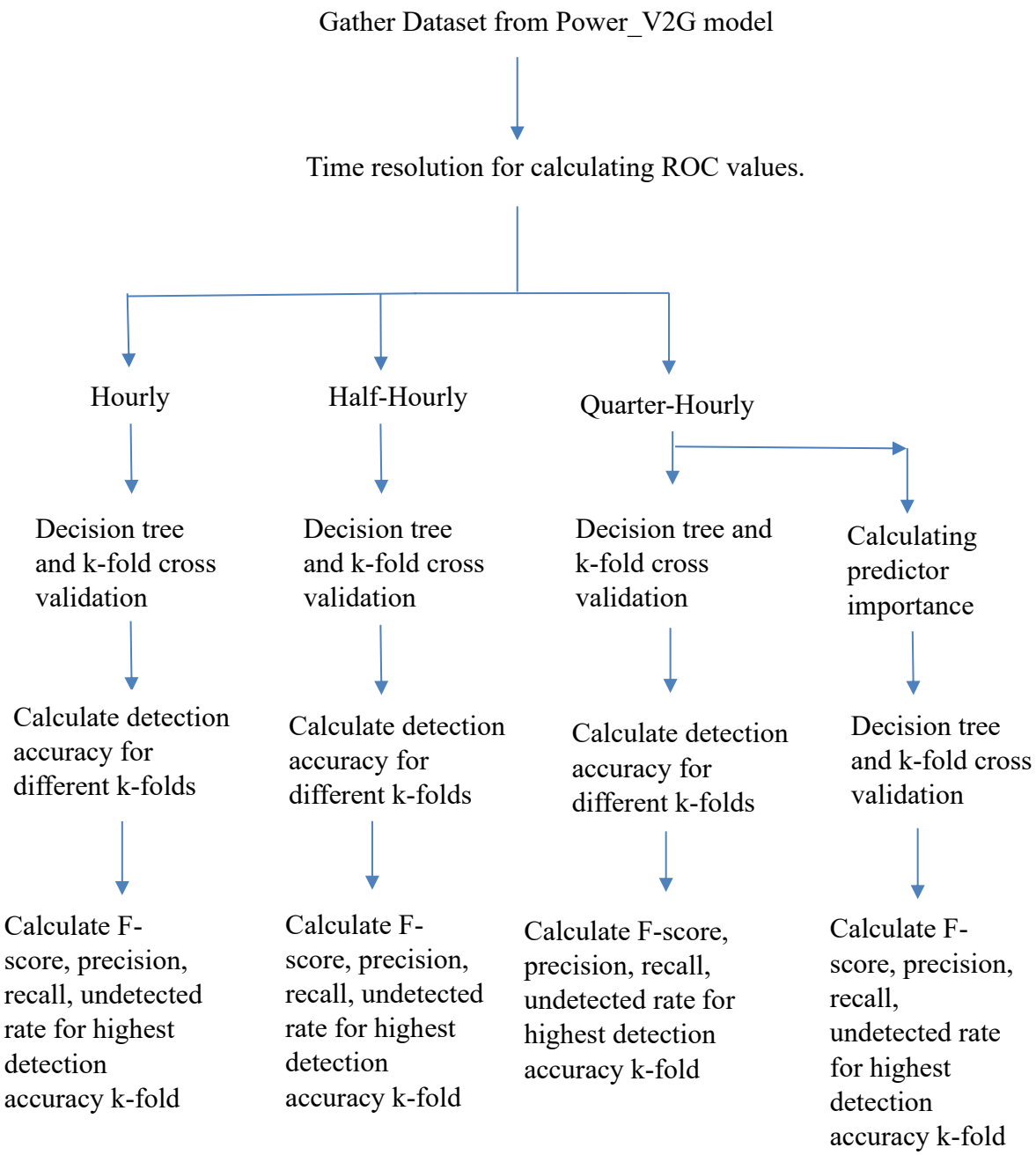

Fig.3.8. Steps for getting optimal detection results.

In a real system, the cyber attacks will be detected at the distribution automation and control (DAC) system in distribution substations. Typically, the digital recording meters will transfer the quarter-hourly load consumption pattern from the distribution transformer to the substation where the IDM is operating in the substation computers. In MATLAB environment, this dataset is typically imported from the power demand profile of the transformer shown in Fig. 4.1 for pre-processing. In the pre-processing stage, the first step is to compute the sequence difference by applying the subtraction of the readings at the quarter hourly time resolution. Furthermore, the resulted dataset is imported in the IDM where the updated dataset is fed into the DT, which detects whether it is an attack or no attack.

3.8. Evaluation Metrics

The most critical step in determining the success rate of a classifier is through the performance, which is assessed through the evaluation metrics. The measures such as the accuracy measure, precision, recall, F-score, and the undetected rate are used in this thesis to evaluate the performance of the proposed detection approach for cyber-physical attacks.

3.8.1. The Accuracy Measure

This measure gives information about the percentage of correct predictions in the model.

$$A_C = \frac{\alpha}{\beta} = (TP + TN)/(TP + FP + FN + TN) \tag{3.9}$$

Where, A_c represents the classification accuracy, α is the number of the observations correctly predicted by the classifier and β represents the total number of observations in the dataset [23]. Table 3.9 shows the description of instances.

Table 3.9: Description of instances.

Instances	Description
True Positive (TP)	Attack Scenario is classified accurately.
False Positive (FP)	No Attack Scenario is mistakenly classified as Attack Scenario.
False Negative (FN)	Attack Scenario is misclassified as No Attack Scenario.
True Negative (TN)	No Attack Scenario is distinguished accurately.

3.8.2. F-Score

The precision and the recall are used because the prediction of the attack class by the classifier is important to measure. F-score is the mean of the precision and the recall [23].

$$F - Score = 2 * TP/(2 * TP + FP + FN) \quad (3.10)$$

3.8.3. Precision

The precision metric shows the accuracy of the attack class. It measures how likely the prediction of the attack by the IDM is correct. The higher the precision is, the lower will be the false positive rate of the classifier [23]. The precision ignores the no-attack class; therefore, the recall is necessary to include.

$$Precision = TP/(TP + FP) \quad (3.11)$$

3.8.4. Recall

The recall, sensitivity, or true positive rate (TPR) computes the ratio of the attack class correctly detected. This metric shows how good the IDM is at recognizing the attacks. The

classifiers with large recall measure have very few attack scenarios misclassified as the No Attack [23].

$$Recall = TP/(TP + FN) \tag{3.12}$$

3.8.5. The Undetected Rate

The fraction of the attack scenarios misidentified as no attack by the classifier is defined as [23].

$$Undetectedrate = FN/(FN + TP) \tag{3.13}$$

3.8.6. Confusion Matrix

Confusion matrix provides the information required for the assessment of the classification model. In a confusion matrix, the number of the records correctly and incorrectly predicted are presented in the form of a Table [23] as represented in Table 3.10.

Table 3.10: Confusion matrix.

Actual Class		Predicted Class	
		Attack	No Attack
	Attack	TP	FN
	No Attack	FP	TN

3.8.7. Confidence Interval

The accuracy estimates without confidence intervals are useless [53]. Therefore, the Standard Z-test is used to estimate the confidence interval for accuracy. The mathematical formula shown below is used to determine the upper and the lower bound of accuracy at the 95% confidence interval.

$$Z_{\frac{\alpha}{2}}^2 \pm Z_{\frac{\alpha}{2}}\sqrt{Z_{\frac{\alpha}{2}}^2 + 4 \times N \times \varepsilon - 4 \times N \times \varepsilon^2}\ )/\ (2 \times \left(N + Z_{\frac{\alpha}{2}}^2\right) \quad (3.14)$$

Where, N is the total number of observations, ε determine the accuracy, which was calculated using equation 3.9, and the value of $Z_{\frac{\alpha}{2}}$ is 1.96 at 95% confidence interval [23].

3.9. Summary

This chapter describes the dataset generation procedure and the methodology used in the cyber attack detection process. The complete IDM is developed starting with the sampling rate chosen based on different time resolutions in the smart metering infrastructure, followed by the sequence subtraction, which is utilized to calculate the ROC values for each sample point. Furthermore, the ML classification is introduced where the DT classifier is described. In addition, the Feature Selection Method and its importance for the detection of the cyber attacks is determined. The method of the CVM was presented to assess the performance of the DT. Finally, the complete IDM algorithm is developed along with the description of the evaluation metrics used for the detection. The cyber-physical attacks description and their impacts on the microgrid operation are presented in Chapter 4 while the detection results following the implementation of the proposed IDM are presented in Chapter 5.

4. Description of the Cyber-Physical Attacks on Fast Charging Stations and their Impacts

4.1. Introduction

In this chapter, a detailed description of the DoS attacks on FCSs and the potential impacts due to the DoS attack are described. In general, the DoS attack means when an attacker prevents the intended user from accessing a device or service. In this work, the impacts of the DoS attacks are determined when targeting the FCSs in the microgrid system shown in Fig. 3.2 and described in Chapter 3, hence preventing the users from charging their own EVs or providing the ancillary services through V2G mode. In the microgrid model of Fig. 3.2, people are going to work and need to charge their vehicles at work. The people need to charge their vehicles from 8:00 till almost 9:00. However, the FCSs are under attack from 8:00 till 18:00, making the EV users unable to charge during that duration. The possible impacts of the DoS attacks on the power systems based on the case studies in the literature is explained to emphasize on the importance of detection.

A detailed description of the test system model is explained in Chapter 3 Section 2. According to the microgrid model, there are five different profiles for EVs which shows five different types of EV user.

Due to the introduction of the internet and communication technologies (ICT) in V2G technology, the fast charging stations become highly vulnerable to the cyber attacks. The weakest point in Power_V2G model is the V2G technology. Let's consider other components in the microgrid under attack and their consequences. The wind farm can provide maximum of 4.5 MW of power; however, the solar farm can contribute maximum of 8 MW of power. If the wind and solar farm are under attack (single or both), the diesel generator will be capable of compensating

the required generation demand in the microgrid. The maximum generation capacity of the diesel generation is 15MW, the maximum residential and industrial load demand is 10.5 MW, and EVs require maximum of 4MW of power, the total load demand is 14.5 MW. Therefore, the chances of the wind and solar farm under attack are close to zero.

Now, lets consider how V2G is the weakest point in the microgrid and the targeted area for the hacker. Firstly, the V2G technology in the microgrid provide frequency and voltage regulation whenever an event occurs. For example, if the wind farm trips, the V2G provides peak regulation at that time. Therefore, the V2G can be an ideal point for the hacker to disturb the stability of the microgrid. Secondly, V2G technology includes communication with the microgrid control system and communication between users and FCSs. The communication technology make the V2G highly vulnerable to the cyber attack as it provides the hacker an easy access to microgrid. Thirdly, if the batteries of the EVs are depleted and all EVs start to charge at the peak time, it will impact the transformer aging and the very worst case it can cause blackout.

The description of denial-of-service attacks in FCS are as follows: FCSs are compromised by hackers. EV users go to the desired location for charging EVs. They use a third-party app for authorization. Therefore, when EV users request charging, the compromised FCS does not authorize the EV thus denying the provision of service to the EV users.

4.2. Assumptions

Following assumptions are made while using the Power_V2G model.

- As described in the model, Profile 1 users are the individuals that are going to work with a normal ride. It took 2 hours for Profile 1 individuals to reach at work. The assumption for Profile 1 EV users is that they have the privilege to charge their vehicles from private FCS.

- Profile 2 users going to work with the normal ride, do not have the opportunity to charge their car at work therefore, it can be said that they prefer to charge their car at home and not at work.
- For Profile 3, the description says that EV users are going to work with a longer ride. Profile 3 EV users need almost 3 hours to reach at work. The assumption for Profile 3 EV users is that they can use public FCS for charging their EVs.
- Profile 4 users are staying at home; therefore, they are not charging at any time.
- Profile 5 users are going to work midnight; therefore, it can be said that they are not contributing toward load peak at 18:00.

The attacks and no attack scenarios are described in the Table 4.1, which lists the description of all scenarios that will be used for the explanation of the attacks.

Table 4.1: Description of cyber-physical attack scenarios

Scenario number	DESCRIPTION
Scenario 0	No Attack scenario, people can charge from Public and Private FCSs at work.
Scenario 1	Private FCS are under attack. However, Public FCS are working properly.
Scenario 2	Public FCS are under attack. However, Private FCS are working properly.
Scenario 3	Public and Private FCS are under attack

4.3. No Attack Scenario Description

Fig. 4.1 shows the no-attack scenario when FCSs are not under attack and depicts the power demand profile at the distribution transformer for 24 hours. For Scenario 0, the charging duration

of the vehicles in profile 1 is almost 25 min at 8:00, whereas the vehicles in profile 3 require almost 38 min to fully charge. The different power magnitude for Profile 1 and Profile 3 is because of the distribution of the vehicles in each profile. Profile 1 has a greater number of vehicles compared to the number of the vehicles in profile 3 (Section 3.3.2). The peak on Fig 4.1, from 8:00 to 9:00 shows the power consumption pattern due to profile 1 and profile 3 EVs. More detailed description is determined in Section 5.3. The description of different vehicle distributions was discussed in Chapter 3, Section 3.2. For every hour of driving, the EV battery is drained 10%. It requires 12.5 min to charge 10% of the EV battery. Profile 1 EVs reach home at 18:00, while Profile 3 start charging their EVs at 19:00 due to the longer ride. The vehicles with profile 4 are staying at home, therefore, they do not need to charge at any time. However, the vehicles with profile 5 are going to work at night and prefer to charge once they reach home at 4:00.

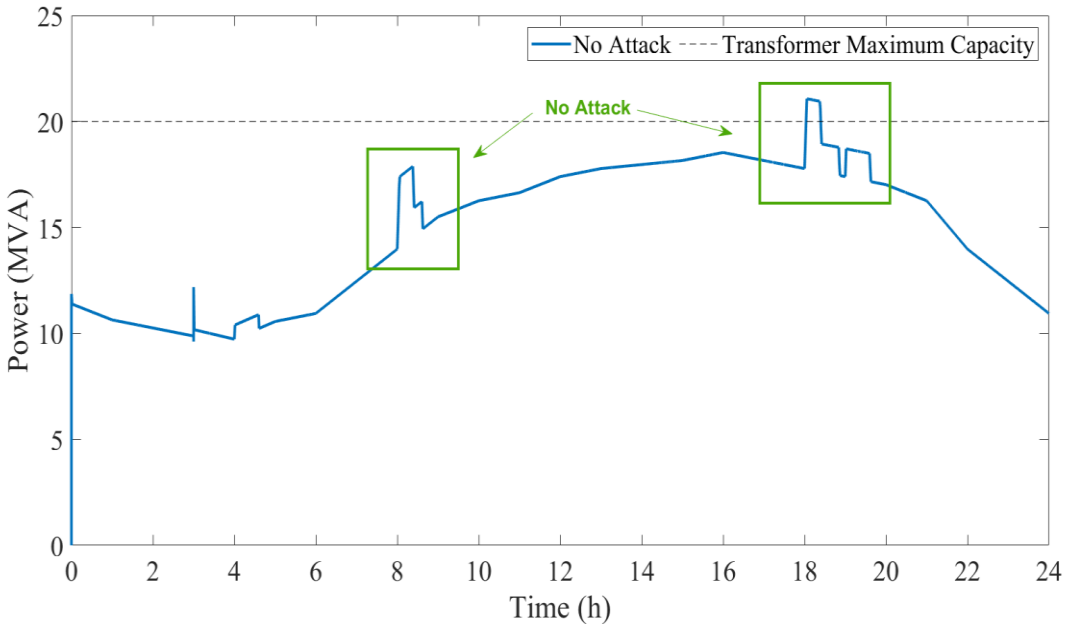

Fig. 4.1. Daily plot of the transformer's power demand with no attack.

4.4. Denial of Service Attack 1 Scenario Description

According to Attack 1, people going to work with a normal ride (Profile 1) cannot charge their vehicles at work. Private FCSs are under attack whereas the public FCSs are not impacted by the attack. Profile 1 EV users are driving to work from 6:00 to 8:00 and 20% of the battery is drained while travelling. For Scenario 0, people will be able to charge their vehicles at work and once they reach home the SOC of the cars will be 80%. For Scenario 1, people need the same amount of time to reach their work, which is 2 hours and their SOC is 80% once they reach at work, but they do not have the opportunity to charge their cars at work. Therefore, once they reach home their SOC will be 60%, and it will require 50 min to fully charge the EVs. The transformer overload duration because of scenario 1 has increased from 23 min to almost 50 min. The magnitude of the transformer power demand at 18:00 is 21 MVA. It can be seen in Fig. 4.2, during attack 1, a peak occurs at 8:00 (shown in blue), which imply that the public FCS are functioning properly and are not under the attack.

4.5. Denial of Service Attack 2 Scenario Description

According to Attack 2, people are going to work with longer rides (Profile 3) and cannot charge their vehicles at work. The public FCS are under attack while the private FCS are not impacted by the attack, and they can still charge their vehicles at work. As profile 3 vehicles need 3 hours to reach their work, profile 3 EV users are driving from 5:00 to 8:00 and 30% of the battery is drained while travelling. When they reach work, the SOC for profile 3 vehicles is 70%. For Scenario 0, people will be able to charge their vehicles at work and once they reach home, the SOC of vehicles will be 70%. As Profile 3 vehicles travel 3 hours to reach work from home and vice versa, profile 3 cars are reaching home at 19:00. The charging duration due to profile 3 vehicles in Scenario 0 will be 38 min at 19:00. For Scenario 2, people need the same duration to

reach their work that is 3 hours and their SOC is 70%, but they do not have the opportunity to charge their vehicles at work due to the attack. Therefore, once they reach home their SOC will be 40%, and it requires 76 min to fully charge the EVs. It can be seen in Fig. 4.2, during attack 2, a peak occurs at 8:00 (shown in green), which depicts that private FCS are charging. Attack 2 does not have a significant impact on the transformer overload compared to attack 1 and attack 3, however, the charging duration has increased from 38 to 75 min at 19:00.

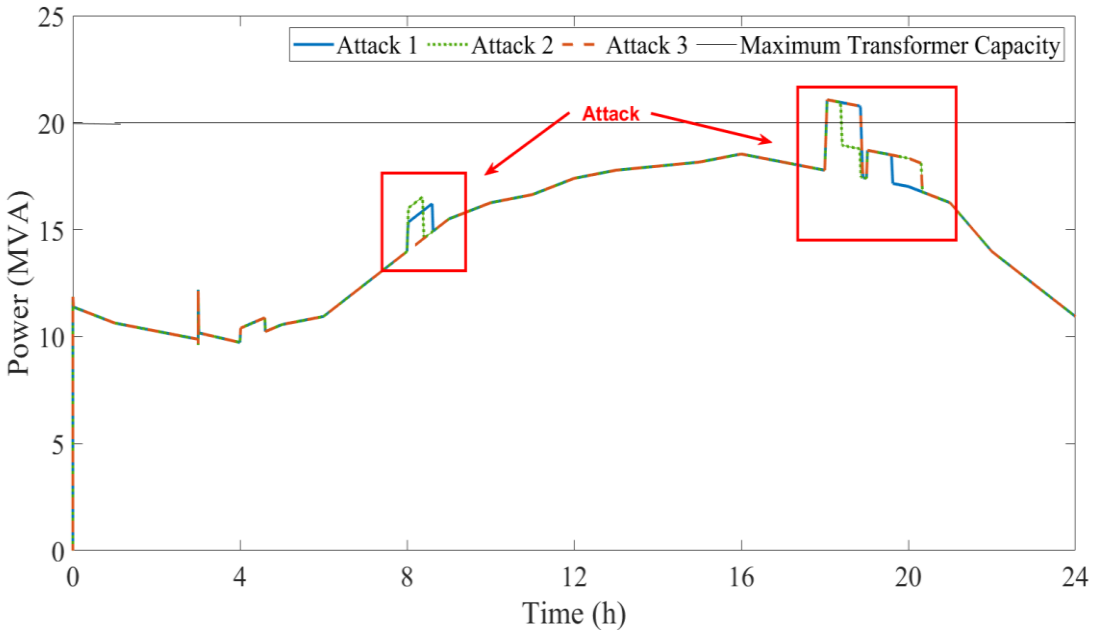

Fig. 4.2. Daily plot of the transformer's power demand with attack 1, attack 2, attack 3.

4.6. Denial of Service Attack 3 Scenario Description

According to Attack 3, people are going to work with a normal ride (Profile 1) and people are going to work with a longer ride (Profile 3) cannot charge their vehicles at work. The public and the private FCS are under attack; thus, no vehicle can charge at work. Attack 3 is the sum of attack 1 and attack 2. The consequences of the attack 3 is also the sum of consequences due to attack 1 and attack 2. In Scenario 3, as the private and the public FCS are under attack and no car

is able to charge at that time, the overload duration at 18:00 due to scenario 1 and the increase in the charging duration due to Scenario 2 is observed. There is no peak at 8:00 as shown by the red dotted line at 8:00 in Fig 4.2, which clearly confirms the attack at 8:00. It can be observed that attack 3 may be considered the most dangerous attack among all other attacks as it includes the consequences of attack 1 and attack 2.

4.7. Impact of Simultaneous Charging of Electric Vehicles on the Power Distribution System

It can be concluded from the previous section that the DoS attacks in the microgrid result in the simultaneous charging of EVs. Different case studies were presented in the literature to analyze the impact of the uncontrolled charging of EV on the power system. The presented cyber-physical attacks in the previous section show that all EVs will start to charge at the same time, which is peak hour and as a result the transformer experienced a severe overload. In this section, the impact of the simultaneous charging of EVs on different distribution networks is reviewed to understand the possible detrimental impacts of EV cyber-physical attacks on the charging station in the microgrid. The motive of this section is to understand the intensity of the impacts of such attacks and the need for its early detection. The scenario where all EVs start to charge simultaneously could result in many worst cases. There are several factors that demonstrate the impact of simultaneous charging, i.e., distribution network type, rating of transformer, load pattern and the penetration level of EVs.

The sizing of the distribution networks is typically based on an approximate electricity demand of the customers. In the case of high penetration of EVs, the transformers would have to experience unexpected load from EVs. The transformers are one of the expensive assets in the power distribution system and the average lifespan of a transformer can be 50 years [54]. For the transformer overload, the overloading is observed if the apparent power of the transformer exceeds

its rated capacity. If the load of the transformer is beyond its rated capacity, the operating temperature of the transformer increases, which eventually impacts the aging of the transformer. Overloading of the transformer does not mean the instant malfunctioning of the transformer, but it surely impacts the Loss-of-Life (LOL) of the transformer. If proper actions are not taken to avoid or resolve the overloading of the transformers, it may eventually damage the transformer and the failure of the transformer will be obvious [54].

One of the main consequences of overloading the transformers is their accelerated aging. The simultaneous charging of PEVs will significantly increase the peak demand. The measure of the impact relies on the charging behaviour of EVs and the penetration level [1]. Below are discussed various case studies, which reveal the impact of EVs simultaneous charging with different penetration levels.

In [54], a survey is done to examine the possible reasons for the transformer aging, especially the impact of EVs loading on the LOL of the transformer. Survey results depicted that the simultaneous charging of the EVs may result in voltage sag, line overload, feeder congestion and the transformer aging as well.

In [55], a case study is done in Toronto to measure the effect of EVs charger load on the distribution network. A scenario is considered where all EVs start to charge simultaneously during the peak hours in summer and winter and the impact of increase in EVs penetration to the distribution network is also stated. According to the research findings, the ambient temperature plays an important role in the overloading. The system is more sensitive to being overloaded during summer due to the high load demand and the system-current-carrying capacity reduces at high ambient temperatures. Overloading is also dependent on the size of the chargers, the chargers with the rating of 3.3 kW and less are less likely to create overloading concerns, whereas 6.6-kW

chargers could cause moderate overload for system components. A major concern is the fast chargers, even few EVs in this case can cause overload to various system components.

In [56] another case study was done in the UK to see the impact of EVs with the penetration level of 25%, 50% and 100% on the low voltage power distribution system. According to the results, even 25% of total cars on the road can be a reason for transformer overloading. Simulations were carried out with different types of load behavior. The charging levels for the simulation were 3.3 kW and 6.6 kW.

The work in [57] reviewed the impact of EVs on the power distribution system. According to [57] the impact of EVs on the distribution system depends on the penetration level of EVs, the time and the charging duration. The possible impacts with the charging of high penetration level of EVs during peak hours were as follows:

1. Components of the distribution system could become overloaded with the high penetration level of EVs, and the transformer lifespan will be affected over the long term.
2. Excessive power and energy losses.
3. Impact on the residential load curve.
4. Voltage drops and cable overload.

4.8. Importance of the Early Detection of Cyber-Physical Attacks on Fast Charging Stations

Cyber threats are increasing exponentially, making it challenging to cope with the speed of the security threats and create security solutions to prevent them [35]. Based on the previous discussion, there is a need for the protection methods that can learn from their experiences and detect the previous and new unknown attacks. In this section, some of the cyber attacks on the

power systems and their impacts are discussed. This section will determine the necessity and importance of “early” detection of the cyber attacks.

Stuxnet was the first digital weapon targeting the industrial control system of the uranium enrichment plants in Iran. On April 1, 2009, Stuxnet was first introduced in a uranium plant of Iran affecting the speed of centrifuges and the investigators suggested to replace the centrifuges. However, in January 2010 new malware was introduced, which spread faster and more dangerous than the previous one and still undetected. Moreover, in March 2010, hackers added more components to make Stuxnet more powerful. At last, in June 2010, the cyber attack was detected [58].

Recently, in 2017, a petrochemical plant in Saudi Arabia was infected by a Triton cyber attack, which is a dangerous malware. Triton affected the safety instrumented system, which is responsible for taking precautionary measures in the time of life-threatening disaster to the petrochemical plants. The safety instrumented system got disabled due to the cyber attack and the hackers used other software to damage other significant equipment of the plant. In June 2017, the plant stopped its operation and was mistakenly considered as a mechanical glitch. However, a few months later in August 2017, several systems were shut down, which caused the attention of the plant owner, the investigation was started, and Triton was detected and resolved. The attack could have caused the release of the toxic gases or in the worst case an explosion could have happened [59].

It can be concluded from above mentioned incidents that the cyber attacks are evolving and becoming highly sophisticated and thus are hard to detect. Symantec, a big cyber security company, published a report which shows that on average there are more than 200,000 web cyber

attacks occurring around the world daily in 2016 [60]. It is common these days that organizations are unaware of their system being impacted by the cyber attacks. From the above two real life events, it can be observed that organizations were unaware of the cyber threats, and it took almost a year to determine that their system was compromised due to the cyber attacks.

The existing cyber security detection techniques are not enough and require the early detection mechanism. There are various benefits provided by the early detection of the cyber attacks; firstly, it may help the security members to analyse the threat and the possible impact due to the attack before the attack compromises the system, secondly, it will give enough time to implement the mitigation techniques, secure the system from severe impacts and prevent future cyber attacks. Last but not the least, it will assist to secure the equipment thus saving a lot of money and the reputation of the company will be protected as well [60].

The early detection mechanism proposed in this research will give 9 hours to take precautionary measures to secure the microgrid from the cyber attack impacts.

4.9. Summary

This chapter presents a detailed description of three different types of DoS attacks in FCSs. The chapter provided a detailed description of the attack scenarios considered in this thesis. The impacts of the DoS on FCSs are provided while the importance of the early detection of the cyber attack is discussed. The results of implementing the proposed approach to early detect the cyber-physical attacks on FCSs are presented and discussed in Chapter 5.

5. Results and Evaluation

5.1. Introduction

This chapter assesses the performance of the proposed IDM developed in this thesis. As discussed earlier, the DT, which is an eager ML algorithm, will set up rules for the detection followed by k-fold CVM used to determine the classification accuracy. When a DT is used with a CVM, the DT acts as a proposed hypothesis and the CVM is used to compute the accuracy based on the proposed hypothesis [61]. Furthermore, different time resolutions will be tested such as hourly, half-hourly and quarter-hourly time resolution, which are available from the smart metering infrastructure. Also, different k-fold cross validation will be evaluated for each time resolution to show the effect of k-fold value on the detection accuracy of the proposed early detection approach of cyber-physical attacks on FCS. Furthermore, computational time and the percentage change of computational time from 4-fold to 5-fold and from 5-fold to 10-fold is calculated to achieve the optimal k-fold.

5.2. Detection Results

In this section, the results are presented for the IDM where the detection results with different time resolution will be presented separately to allow for a comparison. In the evaluation metric Tables shown below; the classification accuracy is calculated using (3.9), the F-score by using (3.10), the precision using (3.11), recall using (3.12) and the undetected rate using (3.13).

5.2.1. Cyber Attacks Detection Results using the Hourly Time Resolution

By using the hourly ROC in the dataset, the results are as follows. As shown in Table 5.1 (green highlighted cells show the highest accuracies), the highest detection accuracy for attack 1 is 83.33% at k-fold 3, however attack 2 showed the highest accuracy of 81.67% at k-fold 2, k-fold

6 and k-fold 11. Moreover, attack 3 showed the highest accuracy of 98.33% at various k-fold values; k-fold 4, k-fold 5 and from k-fold 8 to k-fold 20.

For the hourly time resolution, the peak between the duration 8:00-9:00 does not captured by the IDM which results in lowest accuracies for attack 1 and attack 2. However, in the case of attack 3, there is no peak for the duration between 8:00-9:00 which clearly indicates the intrusion, hence high detection accuracy. The complete description about why attack 3 has the highest accuracy and why attack 1 and attack 2 have the low accuracy among all three attacks is described in Section 5.3 and Section 5.4.

There are several reasons to represent the detection accuracy values from k-fold 1 to k-fold 20 in Table 5.1, Table 5.2 and Table 5.3 because firstly, the detection accuracy does not improve any further after k-fold 20, secondly the goal is to get lowest value of k-fold where all three attacks have the highest accuracy to achieve the less computational time approach.

Table 5.1: Detection results for attack 1, attack 2 and attack 3 in case of the hourly time resolution and different k-folds.

Attack 1 Detection Results			**Attack 2 Detection Results**			**Attack 3 Detection Results**	
k-fold value	Detection accuracy		k-fold value	Detection accuracy		k-fold value	Detection accuracy
1	0.00%		1	0.00%		1	0.00%
2	75.00%		2	81.67%		2	95.00%
3	83.33%		3	78.33%		3	95.00%
4	61.67%		4	78.33%		4	98.33%
5	76.67%		5	75.00%		5	98.33%
6	81.67%		6	81.67%		6	95.00%
7	78.33%		7	80.00%		7	96.67%
8	70.00%		8	75.00%		8	98.33%
9	80.00%		9	73.33%		9	98.33%
10	80.00%		10	80.00%		10	98.33%
11	80.00%		11	81.67%		11	98.33%
12	75.00%		12	70.00%		12	98.33%
13	75.00%		13	76.67%		13	98.33%
14	75.00%		14	78.33%		14	98.33%
15	78.33%		15	75.00%		15	98.33%
16	73.33%		16	80.00%		16	98.33%
17	76.67%		17	75.00%		17	98.33%
18	76.67%		18	76.67%		18	98.33%
19	75.00%		19	80.00%		19	98.33%
20	71.67%		20	76.67%		20	98.33%

For comparison among different time resolutions, a k-fold value is chosen where all three attacks could have the highest accuracy. In case of the hourly time resolution, k-fold 10 is chosen to evaluate the performance of IDM with different evaluation metrics. In case of the hourly time resolution, the results are as follows for k-fold 10. The detection accuracy for attack 1 and attack 2 is the same, which is 80.00%, because the correct prediction of the attacks and no attack scenarios is the same for attack 1 and attack 2. Moreover, the precision measure provides (87.50%) in case of attack 2 compared to attack 1 (82.14%) because the FP rate is small for attack 2 (3 observations were predicted wrongly as attack) as compared to attack 1 (5 observations were predicted wrongly

as attack). However, in terms of the recall measure, attack 1 showed better results (76.67%) as compared to attack 2 (70.00%). The reason why attack 1 performed better as compared to attack 2 is because the FN rate is smaller for attack 1 (7 observations were predicted wrongly as no attack) as compared to the FN rate of attack 2 (9 observations were predicted wrongly as no attack). The higher the recall values, the lower will be the undetected rate, as shown in Table 5.2. The undetected rate for attack 1 is 23.33% while the undetected rate of attack 2 is 30.00%. Moreover, the F-score measure, which is the mean of precision and recall, is high for attack 1 which is 79.31% as compared to attack 2 which is 77.78%. Attack 3 has the highest accuracy among all three attacks, which is 98.33% and the respective F-Score value is 98.31% as listed in Table 5.2. For attack 3, the recall measure is 96.67% however the precision measure is 100% (FP=0) and the undetected rate is 3.33 % as the FN=1. The confusion matrix for attack 1, attack 2 and attack 3 using the hourly time resolution are shown in Table 5.3, Table 5.4, Table 5.5.

According to Table 5.2, attack 2 is the most challenging to detect when using the hourly time resolution as it has the highest undetected rate which is 30.00%. Furthermore, it has the highest FN rate because the detection of attack 2 relies on the duration between 24.6 to 37.20, which is the duration when most of the EVs in profile 1 complete charging while those in Profile 3 are still charging.

Table 5.2: Detection results in case of the hourly time resolution for 10-fold.

Attack Type	Accuracy (%)	F-Score (%)	Precision (%)	Recall (%)	Undetected Rate (%)
Attack 1	80.00	79.31	82.14	76.67	23.33
Attack 2	80.00	77.78	87.50	70.00	30.00
Attack 3	98.33	98.31	100.00	96.67	3.33

Table 5.3: Confusion matrix for attack 1 in case of the hourly time resolution for 10-fold.

		Predicted Class	
		Attack 1	No Attack
Actual Class	Attack 1	23	7
	No Attack	5	25

Table 5.4: Confusion matrix for attack 2 in case of the hourly time resolution for 10-fold.

		Predicted Class	
		Attack 2	No Attack
Actual Class	Attack 2	21	9
	No Attack	3	27

Table 5.5: Confusion matrix for attack 3 in case of the hourly time resolution for 10-fold.

		Predicted Class	
		Attack 3	No Attack
Actual Class	Attack 3	29	1
	No Attack	0	30

It can be concluded from the results shown in Table 5.1 that the detection accuracies are low for attack 1 and attack 2. The motive is to achieve optimal detection accuracies that are above 95.00% at a same k-fold for all three attacks. Therefore, to improve detection accuracies for all three attacks the half-hourly time resolution was tested, and results are presented in Section 5.2.2.

5.2.2. Cyber Attacks Detection Results using the Half-Hourly Time Resolution

To improve the detection accuracies, the half-hourly time resolution was tested with different k-folds, and the results are presented in Table 5.6 (green highlighted cells shows the highest accuracies). It can be concluded from the Table 5.6 that the detection accuracy of attack 1 has improved from 83.33% (the highest detection accuracy obtained in the case of the hourly-time resolution) to 90.00% (the highest detection accuracy obtained in case of the half-hourly time resolution). The detection accuracy for attack 2 also improved from 81.67% (the highest value obtained in case of the hourly-time resolution) to 88.33% (the highest value obtained in case of the hourly-time resolution). Furthermore, attack 3 showed 100% detection accuracy at all k-folds other than k-fold 2 and k-fold 4. It can also be concluded that k-fold value greatly influences the detection accuracies, and therefore different k-folds should be tested to achieve the accurate and the optimal detection results.

The complete description of why attack 3 has the highest accuracy and why attack 1 and attack 2 has low accuracy among all three attacks is discussed in Section 5.3 and Section 5.4.

Table 5.6: Detection results for attack 1, attack 2 and attack 3 with the half-hourly time resolution and different k-folds.

Attack 1 Detection Results			**Attack 2 Detection Results**			**Attack 3 Detection Results**	
k-fold value	Detection accuracy		k-fold value	Detection accuracy		k-fold value	Detection accuracy
1	0.00%		1	0.00%		1	0.00%
2	90.00%		2	78.33%		2	96.67%
3	88.33%		3	80.00%		3	100.00%
4	83.33%		4	81.67%		4	96.67%
5	88.33%		5	88.33%		5	100.00%
6	90.00%		6	81.67%		6	100.00%
7	86.67%		7	85.00%		7	100.00%
8	83.33%		8	85.00%		8	100.00%
9	88.33%		9	83.33%		9	100.00%
10	88.33%		10	83.33%		10	100.00%
11	86.67%		11	85.00%		11	100.00%
12	88.33%		12	83.33%		12	100.00%
13	85.00%		13	85.00%		13	100.00%
14	86.67%		14	85.00%		14	100.00%
15	86.67%		15	86.67%		15	100.00%
16	85.00%		16	85.00%		16	100.00%
17	85.00%		17	83.33%		17	100.00%
18	86.67%		18	81.67%		18	100.00%
19	85.00%		19	86.67%		19	100.00%
20	88.33%		20	80.00%		20	100.00%

Again, for comparison among different time resolutions, a k-fold value is chosen where all three attacks could have the highest accuracy. In case of the half-hourly time resolution, k-fold 5 is chosen to evaluate the performance of the IDM with different evaluation metrics. In case of the half-hourly time resolution, the results are as follows. The accuracy measure has increased from 83.33% (highest) to now 90.00% (highest) for attack 1, because after using the half-hourly time resolution, the TP rate has increased. However, in case of attack 2, the accuracy measure has improved from 81.67% (highest) to 88.33% (highest). The reason why attack 2 accuracy has not boosted more, because even though the TP rate increased from 21 to 26, simultaneously the TN

rate has reduced from 27 to 24. On the other hand, the detection accuracy of attack 3 has reached 100% because the TP and the TN rate are 100%.

Contrarily to the hourly-time resolution, the precision measure provides (84.85%) in case of attack 1 as compared to attack 2 (87.10%). Because the FP rate has increased from 3 (hourly time resolution) to 4 (half-hourly time resolution) for attack 2 which makes the precision value drop. In case of the recall measure, attack 1 recall measure has improved from 76.67% (hourly time resolution) to 93.33% (half-hourly time resolution) and attack 2 recall measure has jumped from 70.00% (hourly time resolution) to 90.00% (half-hourly time resolution). The higher the recall values, the lower will be the undetected rate, as shown in Table 5.7. The undetected rate for attack 1 is 6.67% and the undetected rate of attack 2 is 10.00%. Attack 3 has the highest accuracy, F-score, precision, recall and the undetected rate among all three attacks as listed in Table 5.7. The confusion matrix for attack 1, attack 2 and attack 3 using the half-hourly time resolution are shown in Table 5.8, Table 5.9, Table 5.10.

In summary, the half-hourly time resolution has better results in terms of all evaluation metrics because in case of the half- hourly time resolution the IDM can capture features between 8:00-9:00 accurately as compared to when using the hourly time resolution.

Table 5.7: Detection results in case of the half-hourly time resolution for 5-fold.

Attack Type	Accuracy (%)	F-Score (%)	Precision (%)	Recall (%)	Undetected Rate (%)
Attack 1	88.33	88.89	84.85	93.33	6.67
Attack 2	88.33	88.52	87.10	90.00	10.00
Attack 3	100.00	100.00	100.00	100.00	0.00

Table 5.8: Confusion matrix for attack 1 in case of the half-hourly time resolution for 5-fold.

		Predicted Class	
		Attack 1	No Attack
Actual Class	Attack 1	28	2
	No Attack	5	25

Table 5.9: Confusion matrix for attack 2 in case of the half-hourly time resolution for 5-fold.

		Predicted Class	
		Attack 2	No Attack
Actual Class	Attack 2	27	3
	No Attack	4	26

Table 5.10: Confusion matrix for attack 3 in case of the half-hourly time resolution for 5-fold.

		Predicted Class	
		Attack 3	No Attack
Actual Class	Attack 3	30	0
	No Attack	0	30

5.2.3. Cyber Attacks Detection Results using the Quarter-Hourly Time Resolution

To further improve the detection accuracies, the quarter-hourly time resolution was used. As shown in Table 5.11 (the cells highlighted in green show the highest accuracies), the highest detection accuracy for attack 1 is 96.67% at k-fold 17, however attack 2 has the highest accuracy of 98.33% at k-fold 4, k-fold 6, k-fold 9, k-fold 11-14 and k-fold 18-20 and attack 3 detection accuracy remained at 100% from k-fold 2 to k-fold 20. The results for the quarter-hourly time resolution shows the highest accuracy for all three attacks, however, there is no common k-fold

where all three attacks have the highest accuracy. Therefore, the fine tuning of the parameters was done where the predictor importance was calculated for each attack type and by using only those predictors, the detection accuracies were calculated. The complete description is presented in Section 5.2.4.

Table 5.11: Detection results for attack 1, attack 2 and attack 3 in case of the quarter-hourly time resolution and different k-folds.

Attack 1 Detection Results			**Attack 2 Detection Results**			**Attack 3 Detection Results**	
k-fold value	Detection accuracy		k-fold value	Detection accuracy		k-fold value	Detection accuracy
1	0.00%		1	0.00%		1	0.00%
2	88.33%		2	88.33%		2	100.00%
3	95.00%		3	96.67%		3	100.00%
4	93.33%		4	98.33%		4	100.00%
5	95.00%		5	78.33%		5	100.00%
6	91.67%		6	98.33%		6	100.00%
7	93.33%		7	91.67%		7	100.00%
8	93.33%		8	95.00%		8	100.00%
9	95.00%		9	98.33%		9	100.00%
10	95.00%		10	93.33%		10	100.00%
11	95.00%		11	98.33%		11	100.00%
12	95.00%		12	98.33%		12	100.00%
13	95.00%		13	98.33%		13	100.00%
14	93.33%		14	98.33%		14	100.00%
15	95.00%		15	95.00%		15	100.00%
16	95.00%		16	96.67%		16	100.00%
17	96.67%		17	96.67%		17	100.00%
18	95.00%		18	98.33%		18	100.00%
19	95.00%		19	98.33%		19	100.00%
20	95.00%		20	98.33%		20	100.00%

5.2.4. Cyber Attacks Detection Results using the Quarter-Hourly Time Resolution with Feature Selection

Embedded Feature Selection Method was used to determine the predictor importance for each attack type. Depending on the predictor weighting, only predictors with the highest importance were used to overcome the overfitting problem, reducing the computational complexity, and improving the detection accuracy.

According to Fig. 5.1, T8 and T9 showed the highest importance for the detection of attack 1. By utilizing these predictors, the detection accuracy for attack 1 reached 96.67% for almost every k-fold value except k-fold 2, k-fold 3 and k-fold 6, see Table 5.12. Visual inspection of Fig. 5.1 also revealed that predictor T10 and T11 are the most important for predicting attack 2 class while the remaining predictors have negligible importance. According to Table 5.12, the final detection accuracy for attack 2 is 98.33% for all k-folds except for k-fold 2, k-fold 3, k-fold 5 and k-fold 10. The detection accuracy for attack 3 while using the highest importance predictor T8 remains at 100% as depicted in Table 5.12. It can be concluded from the results that by utilising the Embedded Feature Selection Method the detection accuracy has improved for most of the k-fold values especially for attack 1. For attack 1 and without Feature Selection Method the detection accuracy was the highest for only k-fold 17 (Table 5.11) however, after using the predictors with the highest importance the detection accuracy boosted for almost all the k-folds.

Table 5.12: Detection results for attack 1, attack 2 and attack 3 with the quarter-hourly time resolution, predictor importance and different k-folds.

Attack 1 Detection Results			**Attack 2 Detection Results**			**Attack 3 Detection Results**	
k-fold value	Detection accuracy		k-fold value	Detection accuracy		k-fold value	Detection accuracy
1	0.00%		1	0.00%		1	0.00%
2	91.67%		2	93.33%		2	100.00%
3	95.00%		3	96.67%		3	100.00%
4	96.67%		4	98.33%		4	100.00%
5	96.67%		5	93.33%		5	100.00%
6	93.33%		6	98.33%		6	100.00%
7	96.67%		7	98.33%		7	100.00%
8	96.67%		8	98.33%		8	100.00%
9	96.67%		9	98.33%		9	100.00%
10	96.67%		10	93.33%		10	100.00%
11	96.67%		11	98.33%		11	100.00%
12	96.67%		12	98.33%		12	100.00%
13	96.67%		13	98.33%		13	100.00%
14	96.67%		14	98.33%		14	100.00%
15	96.67%		15	98.33%		15	100.00%
16	96.67%		16	98.33%		16	100.00%
17	96.67%		17	98.33%		17	100.00%
18	96.67%		18	98.33%		18	100.00%
19	96.67%		19	98.33%		19	100.00%
20	96.67%		20	98.33%		20	100.00%

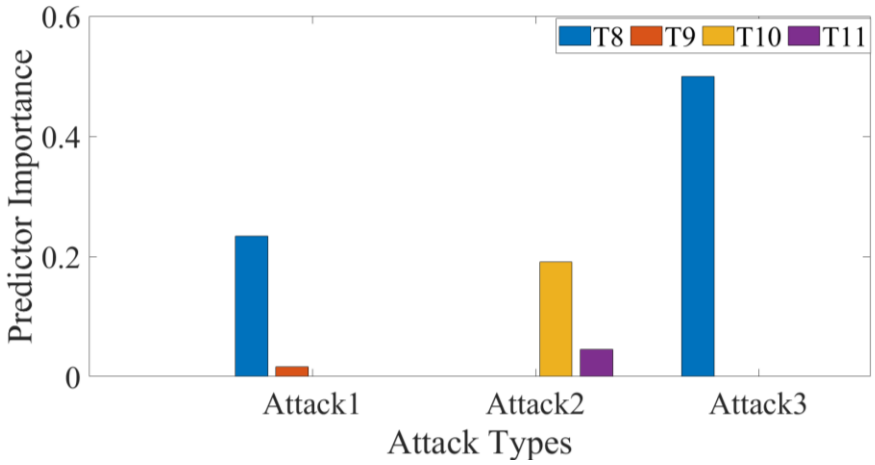

Fig.5.1. Predictor importance for all three attacks.

The settings for the final IDM are as follows:

- Decision Tree Algorithm
- Quarter-hourly time resolution with predictors having the highest weightage in predicting
- 4-fold cross validation method

By using the final IDM settings, the detection accuracy, F-score, precision, recall, and the undetected rate are shown in Table 5.13. The confusion matrix for attack 1, attack 2 and attack 3 using the quarter-hourly time resolution and the predictor importance feature are shown in Table 5.14, Table 5.15, Table 5.16.

Table 5.13: Accuracy, F-score, precision, recall and undetected rate with the final intrusion detection model

Attack Type	Accuracy	F-Score	Precision	Recall	Undetected Rate
Attack 1	96.67%	96.77%	93.75%	100.00%	0.00%
Attack 2	98.33%	98.36%	96.77%	100.00%	0.00%
Attack 3	100.00%	100.00%	100.00%	100.00%	0.00%

Table 5.14: Confusion matrix for attack 1 with the final intrusion detection model

		Predicted Class	
Actual Class		Attack	No Attack
	Attack	30	0
	No Attack	2	28

Table 5.15: Confusion matrix for attack 2 with the final intrusion detection model

Actual Class		Predicted Class	
		Attack	No Attack
	Attack	30	0
	No Attack	1	29

Table 5.16: Confusion matrix for attack 3 with the final intrusion detection model

Actual Class		Predicted Class	
		Attack	No Attack
	Attack	30	0
	No Attack	0	30

The 95% confidence interval was calculated to estimate the detection accuracies for all three attacks when the proposed intrusion detection model will be tested on new dataset. According to the able 5.17, the upper limit for the detection of attack 1 is 99.00% however the corresponding lower limit is 88.60%. Similarly, attack 2 has the capability to detect 99.70% accurately on new dataset and at least 91.10% as the lowest bound. Likewise, attack 3 has the upper bound of 100.00% and the lower bound of 93.90%.

Table 5.17: Confidence interval for three different DoS attacks with the quarter-hourly time resolution.

Attack type	95% confidence interval
Attack 1	88.60–99.00
Attack 2	91.10–99.70
Attack 3	93.90–100.00

5.3. Comparison of Detection Results for all Attacks

According to the results discussed earlier, the following points can be concluded:

1. Detection accuracy is the highest in case of the quarter-hourly time resolution.
2. In case of the quarter-hourly time resolution, the accurate pattern of the EVs charging can be captured for both the public and the private FCS. This helps to identify the behaviour of EV charging between 8:00 and 9:00
3. The detection accuracy of Attack 3 was the highest even in case of the half-hourly time resolution. Therefore, it can be concluded that the proposed detection approach is fully capable of detecting attack 3 despite such an attack was considered the most problematic in terms of its impact since both public and private FCS are under attack. When no EV can charge, the peak at 8:00 disappears, which helps to identify the cyber attack.
4. The detection accuracy of Attack 2 was found to be the second highest accuracy (98.33%) in case of the quarter-hourly time resolution. The reason is EVs using public FCSs require more time to charge the batteries. When the public FCSs are under attack and in the absence of the peak in the power demand of the transformer between the duration of 8.41 to 8.62 (see Fig. 5.4), the intrusion can be classified.
5. The detection accuracy of Attack 1 in case of the quarter-hourly time resolution was found to be 96.67%. It can be concluded that Attack 1 seems to be the most challenging to detect because the peak in the power demand of the transformer can be observed for the charging duration (8:00 to 8:62) as in no attack scenario but with lower magnitude.

Fig. 5.2. shows the accuracy plots. The Figure shows the performance of the intrusion detection methods for all three types of attack. In Fig. 5.2, the highest accuracies of all three attacks

for different time resolutions are presented for comparison. It clearly indicates the exceptional performance of the quarter-hourly time resolution with predictor importance for anomaly detection of all three attacks.

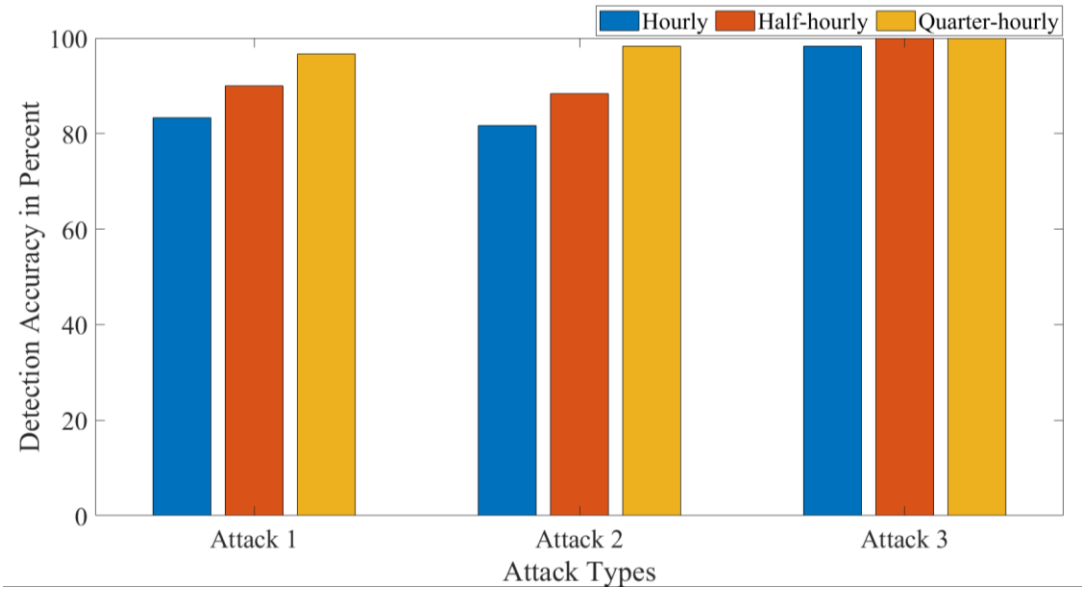

Fig. 5.2. Detection accuracy plot for different time resolutions and different types of attacks.

In Fig. 5.3, comparison of different time resolution is done in terms of the F-score (a), the precision (b), the recall (c) and the undetected rate (d). Visual inspection of Fig. 5.3 clearly indicates the outstanding performance of the intrusion detection model in case of the quarter hourly time resolution. The F-score, the precision, and recall boosted drastically when using the quarter hourly time resolution. However, the undetected rate becomes zero in case of the quarter-hourly time resolution for all three attacks.

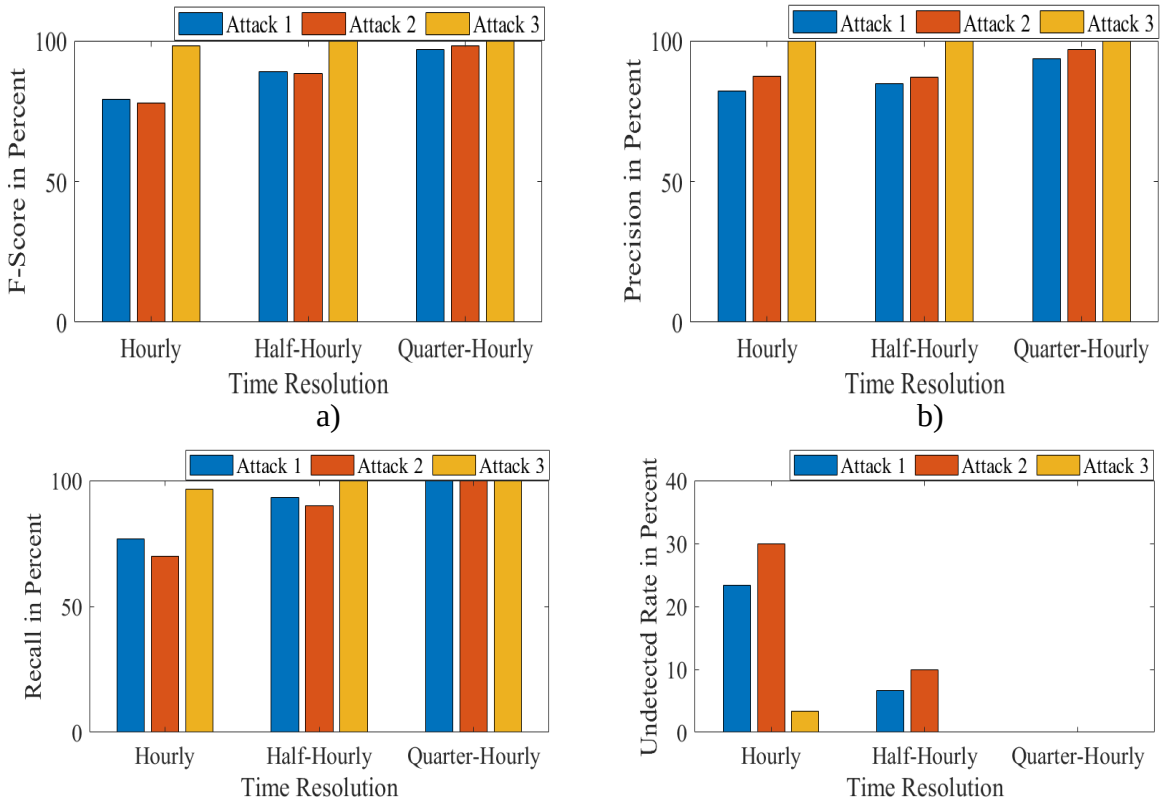

Fig. 5.3. Performance Evaluation of Intrusion Detection Model with different time resolutions.

5.4. Discussion and Analysis of the Results

As shown in Fig. 5.4, the charging duration for the vehicles in Profiles 1 and Profile 3 is from duration 8.00 to 8.625, which is almost 63% of an hour that is (63*60)/100 = 37.8 minutes. Therefore, in case of the hourly ROC data, some information is missing due to the hourly sampling of that time resolution. The accuracy increased in case of the half-hourly data since the peaks of the power demand profile between 8:00 and 9:00 are captured. Furthermore, in case of the quarter-hourly time resolution, more information is kept which helps increasing the detection accuracy.

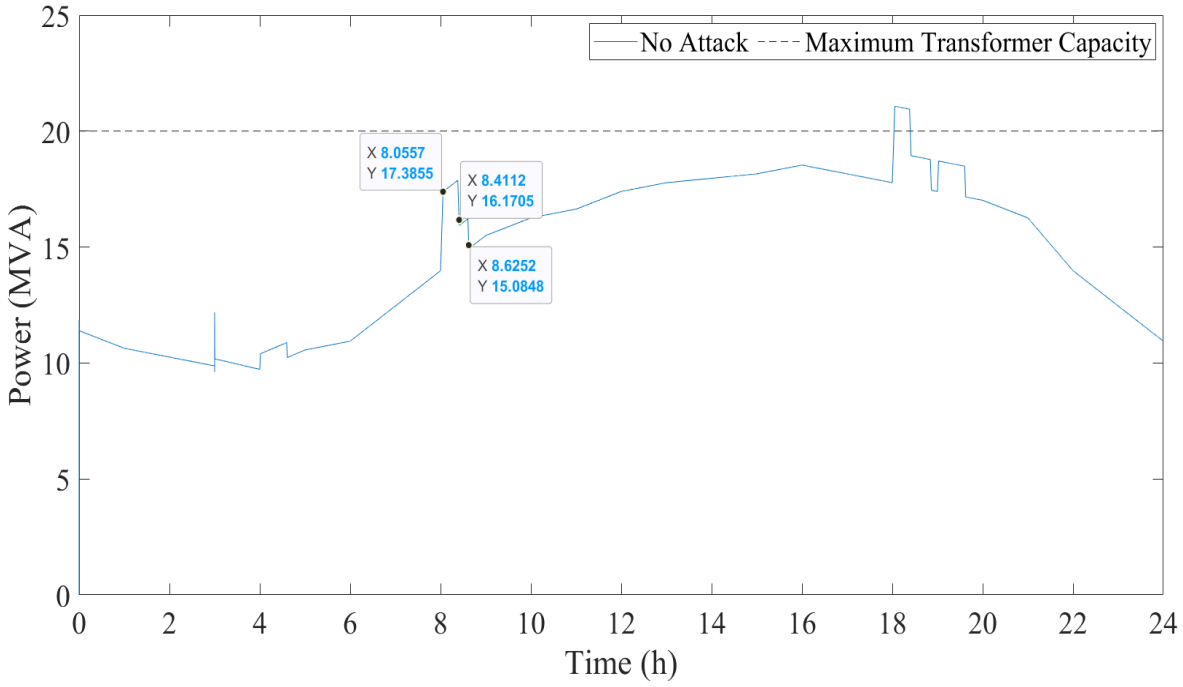

Fig. 5.4. Power of transformer pattern for a day with no attack.

For attack 1, the private FCS were under attack and the EVs in profile 1 cannot charge, but the EVs in profile 3, which are using the public FCS can still charge their vehicles at work. Consequently, the magnitude of the transformer power demand at 8:00 are lower compared to when there is no attack. As shown in Fig. 5.4, when a microgrid is not under attack, the magnitude of the transformer power demand is 17.38 MVA, which reduces to 15.52MVA due to attack 1 (see Fig. 5.5). The lower-than-expected transformer power demand will assist to discriminate between attack 1 and no attack scenarios. As the time resolution increases, the DT becomes able to learn the pattern of attack 1 and no attack scenarios with improved accuracy. The highest detection accuracy for attack 1 was 83.33% when the hourly dataset was used. The detection accuracy reached 96.67% in case of the quarter-hourly time resolution, because the power demand of the transformer peak due to the EVs in profile 1 are not there, which indicates the presence of the DoS attack.

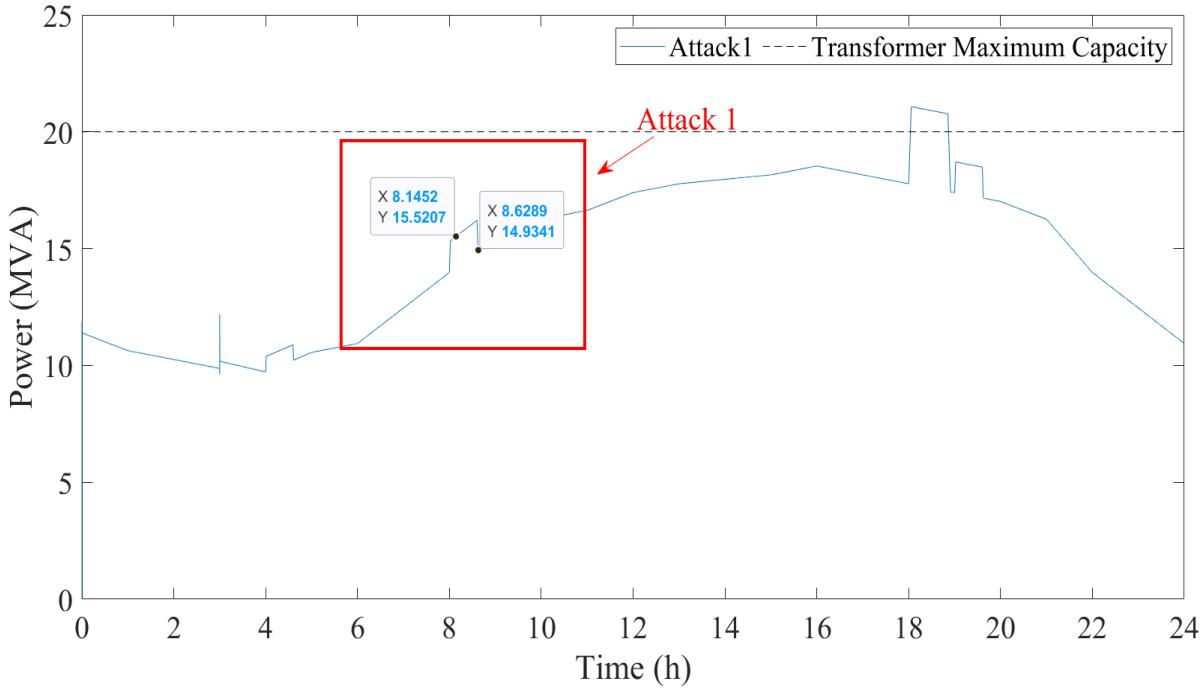

Fig. 5.5. Power of transformer pattern for a day with attack 1.

For attack 2, the public FCSs were under attack and the EVs in Profile 3 cannot charge, but the EVs in profile 1, which are the people using private FCSs can still charge their vehicles at work. Consequently, the magnitude of the transformer power demand at 8:00 are lower compared to when there is no attack. As shown in Fig. 5.4, when a microgrid is not under attack, the magnitude of the transformer power demand is 17.38MVA, however, this is reduced to 16 MVA due to attack 2 (see Fig. 5.6). The number of cars in Profile 1 is more than the number of cars in Profile 3, See Chapter 3, Section 3.2. The lower-than-expected transformer power demand will assist to discriminate between attack 2 and no attack scenarios. As the time resolution increases, the DT becomes able to learn the pattern of attack 2 and no attack scenarios with improved accuracy. The highest detection accuracy for attack 2 was 81.67% when the hourly dataset was used. The detection accuracy reached 98.33% in case of the quarter-hourly time resolution, because the power demand of the transformer peak due to the EVs in profile 3 are not there, which indicates the presence of the DoS attack.

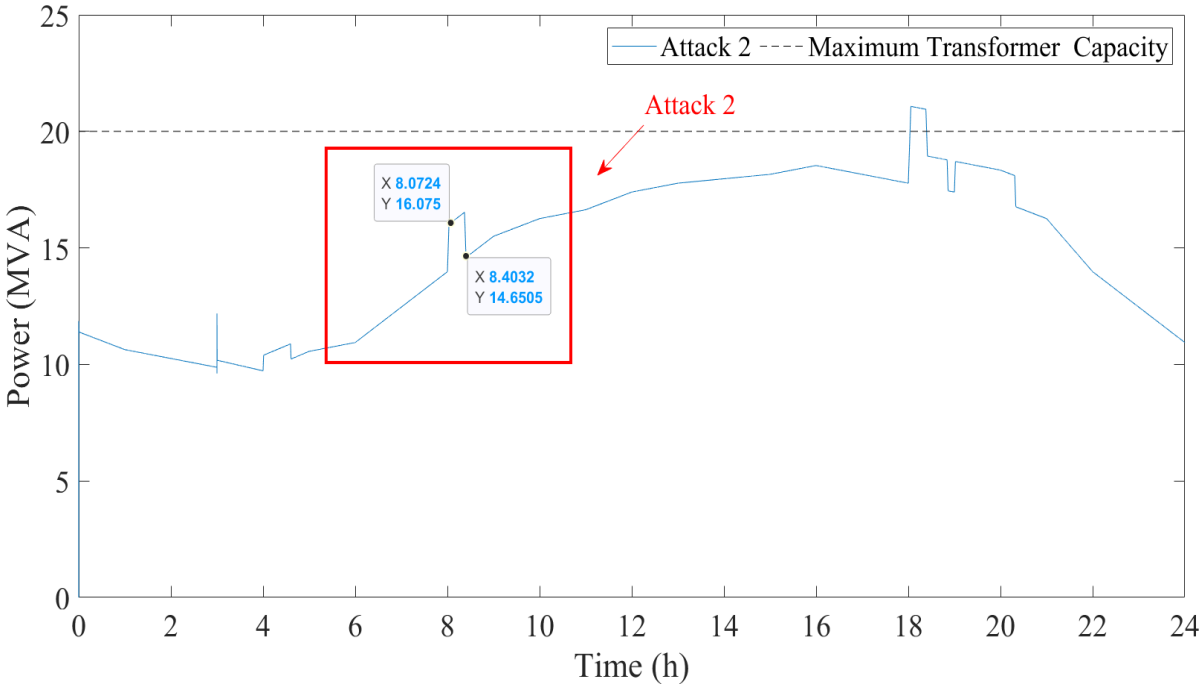

Fig. 5.6. Power of transformer pattern for a day with attack 2.

According to the results shown in Table 5.13, the detection accuracy for attack 3 is the highest among all the other attacks due to the absence of the peak in the power demand profile in attack 3 between 8:00 and 9:00, which confirms the malicious activity (see Fig. 5.7). With the hourly time resolution, the detection accuracy was 98.33%, which improved to 100% in case of the half-hourly time resolution and remained the same with increasing the time resolution.

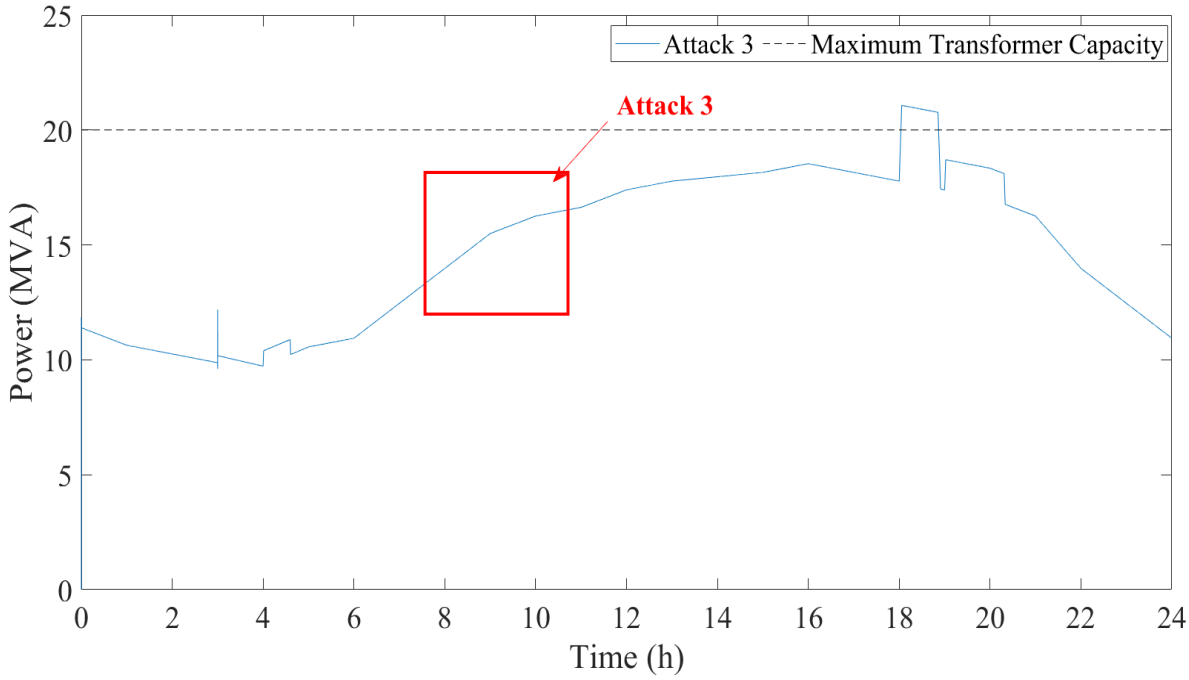

Fig. 5.7. Power of transformer pattern for a day with attack 3.

5.5. Summary

In this chapter, the IDM was implemented for the detection of cyber-physical attacks in FCS. The proposed approach was tested for three different types of "DoS Attack". The intrusion detection model was tested by using different time resolutions (hourly ROC, half-hourly ROC, quarter-hourly ROC). It was concluded from the results that the quarter-hourly time resolution outperformed the other time resolutions. Moreover, the performance of the proposed approach was assessed based on the detection accuracy, F-Score, precision, recall and the undetected rate. The results have shown that the proposed approach was capable of detecting all cases of attack 3 (100% accuracy). On the other hand, the proposed approach was found to be able to classify 98.33% of the cases in attack 2, and 96.67% of the cases in attack 1.

5.6. Computation Complexity of the Proposed Detection Approach

The computational time was calculated using "cpu" time in MATLAB to show the impact of higher k-folds on computational timing for the detection of the DoS attacks. For comparison

purposes, the 4-fold is compared with 5-fold and 10-fold for all three attacks in case of the final IDM (using the Quarter-Hourly Time Resolution with Feature Selection IDM) to show the computational effectiveness of 4-fold.

In Table 5.18, computational time was calculated in case of the final IDM using 4-fold, 5-fold and 10-fold for attack 1. The visual inspection of Table 5.18 shows that while moving from 4-fold to 5-fold the percentage change in computational time is 5%. However, the percentage change in computational time increase drastically from 4-fold to 10-fold which is 44%.

Table 5.18: Computational time in case of the final IDM for attack 1 along with percentage change.

k-fold	cpu time	percentage change
4	0.2263	0%
5	0.2367	5%
10	0.3264	44%

Similarly, In Table 5.19, computational time was calculated in case of the final IDM using 4-fold, 5-fold and 10-fold for attack 2. The Table 5.19 shows that while moving from 4-fold to 5-fold the percentage change in computational time is 12%. However, the percentage change in computational time increase significantly from 4-fold to 10-fold which is 26%.

Table 5.19: Computational time in case of the final IDM for attack 2 along with percentage change.

k-fold	cpu time	percentage change
4	0.2066	0%
5	0.232	12%
10	0.2922	26%

Moreover, In Table 5.20, computational time was again calculated in case of the final IDM (Quarter-Hourly Time Resolution) using 4-fold, 5-fold and 10-fold for attack 3. The Table 5.20 depicts that while moving from 4-fold to 5-fold the percentage change in computational time is

8%. However, the percentage change in computational time increase significantly from 4-fold to 10-fold which is 32%.

Table 5.20: Computational time in case of the final IDM for attack 3 along with percentage change.

k-fold	cpu time	percentage change
4	0.2247	0%
5	0.2423	8%
10	0.2969	32%

Furthermore, the average percentage change for all three-attacks detection while using 4-fold, 5-fold and 10-fold is calculated and shown in Table 5.21. It can be concluded from Table 5.21 that 4-fold is the best option to have optimal accuracy and optimal computational time as well. As moving from 4-fold to 10-fold the percentage change results in 36%.

Table 5.21: Computational time in case of the final IDM for attack 3 along with percentage change.

Average cpu time	percentage change
Average percentage change of all three attacks from 4-fold to 5-fold	7%
Average percentage change of all three attacks from 4-fold to 10-fold	36%

The simulation and testing presented in this work are conducted on a DELL XPS 13 9380 loaded with the Windows 10 Operating System with the following processor: Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz to 1.99 GHz. The IDM are built, trained, evaluated, and tested on the MATLAB. MATLAB is a widely used programming platform, used by millions of scientists and engineers worldwide for programming, data interpretation and model construction [62].

6. Conclusions and Recommendations

6.1. Conclusion

The work presented in this thesis aims to detect the DoS attacks in FCSs thus contributing towards the cyber security of the power system. After reviewing the state-of-the-art literature, it became evident that firstly, there is not enough research conducted to secure the power system from cyber-physical attack in FCSs. Secondly, the limited approaches presented in the past lack the capability of detecting the attacks before the cyber-physical impact is evident. Therefore, the main contribution of the proposed intrusion detection model is the capability of the model to early detect the cyber-physical attacks (i.e., ahead of time) and hence providing enough time to the DNO to take precautionary measures.

Furthermore, in this research the feasibility of the DT method for the learning and classification of cyber-physical attacks was explored. A DT classifier is typically a top-down greedy approach, which provides a rapid and effective method for classification. The DT is a rule-based classification technique, the constructed tree first performs a binary split on the given features based on the Gini index. Then, the DT algorithm divides the tree into different branches and recursively partition the training datasets into smaller subsets until all the subsets belong to a single class. Moreover, the 4-fold cross validation method was used for the performance evaluation of DT classifier. In the proposed work, 75% of the data is used to train the classifier while the remaining 25% of the dataset was used to test the classifier. Additionally, the effect of different time resolutions for the power demand profile of the transformer on the detection and classification accuracy was investigated. The results have shown that the highest detection/classification accuracy can be obtained when using the quarter-hourly time resolution (the time resolution for

final intrusion detection model). To further improve the accuracy of the final intrusion detection model, the most salient features are kept, the research proposed the use of the predictor importance and determining the predictors that hold the highest detection information, thus using only most informative predictors before building the model.

The proposed approach was tested by creating a dataset based on a "Power_V2G" microgrid test system in MATLAB whereby using different penetration levels of cars and different wind and solar profiles a data set was generated, with and without introducing cyber attacks. The final intrusion detection model also leads to better performance in terms of the detection accuracy, F-score, precision, recall and the undetected rate. Using the proposed IDM the detection accuracy for attack 1 reached to 96.67%, the F-score measure is 96.77%, and the precision measure has reached to 93.75%. For attack 2, the detection accuracy is 98.33%, the F-score measure is 98.36%, and the precision measure is 96.77%. Attack 3 has optimal results in terms of the detection accuracy, the F-score measure, and the precision measure (100%). The recall measure for all three attacks has reached to 100% however, there is no undetected event for all three attacks using the quarter-hourly time resolution with the predictor importance in case of 4-fold CVM. In terms of computational time 4-fold is better than 5-fold or 10-fold.

6.2. Recommendations

Based on the work presented in this thesis, the quarter-hourly time resolution is recommended for future cyber attacks detection, as the quarter-hourly time resolution can capture the accurate pattern of the load demand due to EV charging thus detecting cyber-physical attacks with optimal accuracies. When all smart meters are capable of capturing load profile data even after every 5 minutes, these smart meters not only capture electricity consumption pattern for residential customers but for commercial and industrial customers as well. There are more than

100,000 smart meters installed in USA [63]. Research can be done while using 5 minutes dataset for improving the detection accuracy of the proposed approach. While increasing the time resolution, the computational time may increase as well.

Predictor importance with the quarter hourly time resolution will help to remove the redundant data thus increasing the detection accuracy and decreasing the computational time. 4-fold CVM will help to evaluate the performance of classifier with less computational time.

Furthermore, two recommendations can be established in order to mitigate the cyber attacks. The first is the Time-of-Use (TOU) pricing and the second is to use Battery Energy Storage System (BESS).

The Time of use pricing is already popular and preferred by the electricity consumers. According to the time of use program, the electricity consumers are charged differently throughout the day depending on the demand. During the peak hours, the demand is high, and the cost of electricity is high as well. Contrarily, during off-peak hours, the demand is less, and customers are charged less during that duration [64].Therefore, having TOU plans for EV users will force them to charge their vehicles during the off-peak hours and it will assist in reducing the burden on distribution transformers.

The BESS is the second recommendation to cope up with the impact of cyber attacks in microgrids. Cost and availability play a vital role when deciding on an energy storage system. The popularity of lithium-ion batteries is increasing due to the cost effectiveness feature. Also, lithium-ion batteries are better than lead batteries in terms of cost, number of cycles and lifetime as well. Due to these benefits, lithium-ion batteries will be a preferable solution in terms of energy storage systems in near future.

6.3. Future Work

Some next steps that can be taken to build the work presented in the thesis are: investigating the effect of other ML techniques, incorporation of other features like voltage drop in IDM and interconnected microgrids.

There is no publicly available dataset for the detection of the DoS attack in FCS. Therefore, publicly available cyber attack dataset, which includes the dataset based on the real cyber attacks that occurred in the past could assist the future researchers for the research in the field of cyber security in FCS. This approach will assist to provide some realistic and practical results for the detection of cyber attacks by researchers.

References

[1] R. Falk, S. Fries, and S. Ag, “Electric Vehicle Charging Infrastructure - Security Considerations and Approaches,” Jun. 2012, pp. 58–64.

[2] M. A. Mustafa, N. Zhang, G. Kalogridis, and Z. Fan, “Smart electric vehicle charging: Security analysis,” in *2013 IEEE PES Innovative Smart Grid Technologies Conference, ISGT 2013*, Feb. 2013, pp. 1–6.

[3] H. F. Habib, A. O. Hariri, A. ElSayed, and O. A. Mohammed, “Deployment of electric vehicles in an adaptive protection technique for riding through cyber attack threats in microgrids,” in *2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I CPS Europe)*, 2017, pp. 1–6.

[4] A. Mohammad, R. Zamora, and T. T. Lie, “Integration of electric vehicles in the distribution network: A review of PV based electric vehicle modelling,” *Energies*, vol. 13, no. 17, p. 4541, 2020,

[5] L. Wang, L. Pepin, Y. Li, F. Miao, A. Herzberg, and P. Zhang, “Securing Power Distribution Grid Against Power Botnet Attacks,” 2019, pp. 1–5.

[6] R. Gottumukkala, R. Merchant, A. Tauzin, K. Leon, A. Roche, and P. Darby, “Cyber-physical System Security of Vehicle Charging Stations,” in *2019 IEEE Green Technologies Conference(GreenTech)*, 2019, pp. 1–5.

[7] S. Mousavian, M. Erol-Kantarci, and T. Ortmeyer, “Cyber Attack Protection for a Resilient Electric Vehicle Infrastructure,” *2015 IEEE Globecom Work. (GC Wkshps)*, pp. 1–6, 2015.

[8] Rylan Urban, “Electricity Prices in Canada 2021,” Mar. 2021. https://www.energyhub.org/electricity-prices/

[9] "Quickscan of EV Market in British Columbia and Vancouver," May 2020. [Online]. Available: https://www.rvo.nl/sites/default/files/2020/06/Quickscan-EV-Market-British-Columbia-and-Vancouver.pdf

[10] International Energy Agency, "Global EV Outlook 2021," 2021. https://iea.blob.core.windows.net/assets/ed5f4484-f556-4110-8c5c-4ede8bcba637/GlobalEVOutlook2021.pdf

[11] Government of British Columbia, "Zero-Emission Vehicle Update," 2020. [Online]. Available: https://www2.gov.bc.ca/assets/gov/farming-natural-resources-and-industry/electricity-alternative-energy/transportation/2020_zero_emission_vehicle_update.pdf.

[12] G. S. Morrison, "Threats and Mitigation of Ddos Cyberattacks Against the U.S. Power Grid Via EV Charging," Wright State University, 2018.

[13] D. N. Reeh, F. C. Tapia, Y.-W. Chung, B. Khaki, C. Chu, and R. Gadh, "Vulnerability Analysis and Risk Assessment of EV Charging System under Cyber-Physical Threats," *2019 IEEE Transp. Electrif. Conf. Expo*, pp. 1–6, 2019.

[14] iea, "Global EV Data Explorer," Apr. 2021. https://www.iea.org/articles/global-ev-data-explorer

[15] Government of Canada, "Canada Invests in New EV Fast Chargers Across British Columbia," 2021. www.canada.ca/en/natural-resources-canada/news/2021/06/canada-invests-in-new-ev-fast-chargers-across-british-columbia.html. (accessed Jun. 08, 2021).

[16] T&D World, "Canada Invests US$1.15 million on Fast Chargers for Electric Vehicles," Feb. 2019. https://www.tdworld.com/electrification/article/20972220/canada-invests-us115-million-on-fast-chargers-for-electric-vehicles (accessed Apr. 03, 2021).

[17] S. Ahmed and F. M. Dow, “Electric Vehicle and Charging Station Technology as Vulnerabilities Threaten and Hackers Crash the Smart Grid,” Oct. 2016, vol. 3, no. 10, pp. 98–103.

[18] S. Acharya, Y. Dvorkin, H. Pandžić, and R. Karri, “Cybersecurity of Smart Electric Vehicle Charging: A Power Grid Perspective,” *IEEE Access*, vol. 8, pp. 214434–214453, 2020,

[19] Sandia National Laboratories, “Grid and Charging Infrastructure,” Sep. 2021. [Online]. Available: https://www.osti.gov/servlets/purl/1706221.

[20] UCLA Smart Grid Energy Research Center, “UC-Lab Center for Electricity Distribution Cybersecurity,” Mar. 2019. [Online]. Available: https://pdcs.engr.ucr.edu/sites/g/files/rcwecm2241/files/2019-03/RGw2019.pdf

[21] J. H. Reed and C. R. A. Gonzalez, “Using Power Fingerprinting (Pfp) To Monitor the Integrity and Enhance Security of Computer Based Systems,” vol. 1, no. 19, 2013, [Online]. Available: https://patentimages.storage.googleapis.com/b8/0b/f9/f7e1b7044c3598/US20130318607A1.pdf

[22] R. C. Borges Hink, J. M. Beaver, M. A. Buckner, T. Morris, U. Adhikari, and S. Pan, “Machine learning for power system disturbance and cyber-attack discrimination,” in *2014 7th International Symposium on Resilient Control Systems (ISRCS)*, 2014, pp. 1–8.

[23] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, US ed. Addison Wesley, 2005. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20%5C&path=ASIN/0321321367

[24] Y. Yang *et al.*, “Man-in-the-middle attack test-bed investigating cyber-security

vulnerabilities in Smart Grid SCADA systems," in *International Conference on Sustainable Power Generation and Supply (SUPERGEN 2012)*, 2012, pp. 1–8.

[25] U. K. Premaratne, J. Samarabandu, T. S. Sidhu, R. Beresh, and J.-C. Tan, "An Intrusion Detection System for IEC61850 Automated Substations," *IEEE Trans. Power Deliv.*, vol. 25, no. 4, pp. 2376–2383, 2010,

[26] Y. Kwon, H. K. Kim, Y. H. Lim, and J. I. Lim, "A behavior-based intrusion detection technique for smart grid infrastructure," *2015 IEEE Eindhoven PowerTech*, 2015, pp.1-6.

[27] J. Hong, C.-C. Liu, and M. Govindarasu, "Integrated Anomaly Detection for Cyber Security of the Substations," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1643–1653, 2014,

[28] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, and Z. Han, "Detecting Stealthy False Data Injection Using Machine Learning in Smart Grid," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1644–1652, Sep. 2017,

[29] T. Nguyen, S. Wang, M. Alhazmi, M. Nazemi, A. Estebsari, and P. Dehghanian, "Electric Power Grid Resilience to Cyber Adversaries: State of the Art," *IEEE Access*, vol. 8, pp. 87592–87608, 2020,

[30] O. Gul M Khan, E. El-Saadany, A. Youssef, and M. Shaaban, "Impact of Electric Vehicles Botnets on the Power Grid." Mar. 2021.

[31] H. Cui, X. Dong, H. Deng, M. Dehghani, K. Alsubhi, and H. M. A. Aljahdali, "Cyber Attack Detection Process in Sensor of DC Micro-Grids Under Electric Vehicle Based on Hilbert–Huang Transform and Deep Learning," *IEEE Sens. J.*, vol. 21, no. 14, pp. 15885–15894, Jul. 2021,

[32] Q. Wang *et al.*, "Research on Identification Method and Device with Active Immune Attack," in *Journal of Physics Conference Series*, Sep. 2020, vol. 1646, p. 12152.

[33] G. V Nadiammai and M. Hemalatha, “Effective approach toward Intrusion Detection System using data mining techniques,” *Egypt. Informatics J.*, vol. 15, no. 1, pp. 37–50, Jan. 2013,

[34] K. Rai, M. Devi, D. Professor, and A. Guleria, “Decision Tree Based Algorithm for Intrusion Detection,” *Int. J. Adv. Netw. Appl.*, vol. 07, pp. 2828–2834, Jan. 2016.

[35] K. Shaukat *et al.*, “Performance Comparison and Current Challenges of Using Machine Learning Techniques in Cybersecurity,” *Energies*, vol. 13, no. 10, p. 2509, 2020,

[36] I. The MathWorks, “Simscape Electrical,” 2021.

[37] edfenergy, “Types of renewable energy.” https://www.edfenergy.com/for-home/energywise/renewable-energy-sources

[38] W. Tong, “Fundamentals of Wind Energy,” in *Wind power generation and wind turbine design*, W. Tong, Ed. Southampton, USA: WIT Press, 2010, pp. 1–15.

[39] Tesla Motors Inc, “Tesla Model S Electric Vehicle Catalog”, [Online]. Available: https://www.tesla.com/sites/default/files/tesla-model-s.pdf

[40] Government of Canada, “Environment Canada, Canadian Weather Energy Engineering Datasets (CWEEDS).” https://climate.weather.gc.ca/prods_servs/engineering_e.html

[41] KPMG LLP, “The next new vehicle purchase for nearly 70 per cent of Canadians will be an electric model: KPMG in Canada survey,” 2021. https://www.newswire.ca/news-releases/the-next-new-vehicle-purchase-for-nearly-70-per-cent-of-canadians-will-be-an-electric-model-kpmg-in-canada-survey-889637501.html (accessed May 03, 2021).

[42] J. Axsen, S. Goldberg, and J. Bailey, “Electrifying Vehicles: Insights from the Canadian Plug-in Electric Vehicle Study,” 2015.

[43] W. H. Kersting, “Introduction to Distribution Systems,” in *Distribution system modeling*

and analysis, THIRD., Boca Raton, USA: CRC Press, 2012, pp. 1–9.

[44] S. S., "Decision Tree: A Machine Learning for Intrusion Detection," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 6S4, p. 5, Apr. 2019,

[45] B. Jijo and A. Mohsin Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 1, pp. 20–28, 2021.

[46] J. Martínez Torres, C. Iglesias Comesaña, and P. J. García-Nieto, "Review: machine learning techniques applied to cybersecurity," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 10, pp. 2823–2836, 2019,

[47] X. Wang, X. Liu, W. Pedrycz, and L. Zhang, "Fuzzy rule based decision trees," *Pattern Recognit.*, vol. 48, no. 1, pp. 50–59, 2015,

[48] K. Kim, "A Hybrid classification algorithm by subspace partitioning through semi-supervised decision tree," *Pattern Recognit.*, vol. 60, pp. 157–163, 2016,

[49] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, Mar. 2018.

[50] Mathworks, "Introduction to Feature Selection," 2021. https://www.mathworks.com/help/stats/feature-selection.html

[51] Mathworks, "predictorImportance," 2021. https://www.mathworks.com/help/stats/compactclassificationensemble.predictorimportance.html

[52] MATLAB & Simulink., "Cross-Validation." https://www.mathworks.com/discovery/cross-validation.html..html

[53] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Proc. 14th Int. Jt. Conf. Artif. Intell.*, vol. 2, pp. 1137–1143, Aug. 1995.

[54] R. Godina, E. M. G. Rodrigues, J. C. O. Matias, and J. P. S. Catalão, “Effect of Loads and Other Key Factors on Oil-Transformer Ageing: Sustainability Benefits and Challenges,” *Energies*, vol. 8, no. 10, pp. 12147–12186, Oct. 2015,

[55] M. A. Awadallah, B. N. Singh, and B. Venkatesh, “Impact of EV Charger Load on Distribution Network Capacity: A Case Study in Toronto,” *Can. J. Electr. Comput. Eng.*, vol. 39, no. 4, pp. 268–273, 2016,

[56] K. Zafred, J. Nieto-Martin, and E. Butans, “Electric Vehicles - Effects on domestic low voltage networks,” *2016 IEEE Int. Energy Conf. ENERGYCON 2016*, pp. 4–8, Apr. 2016,

[57] M. Sepahi, A. Mirzaei, and E. Saadati, “Impacts of electric vehicles on power distribution system,” Sep. 2016.

[58] Reuters, “Factbox: Cyber warfare expert’s timeline for Iran attack,” Dec. 2011. https://www.reuters.com/article/us-cyberattack-iran-idUSTRE7B10AV20111202

[59] A. DAVIS, “Triton is the world’s most murderous malware, and it’s spreading,” Mar. 2019. https://www.technologyreview.com/2019/03/05/103328/cybersecurity-critical-infrastructure-triton-malware (accessed Jul. 03, 2021).

[60] H. H. Ambusaidi, “Cyber Threats Early Detection Countermesures and Benefits,” *IJRDO - J. Comput. Sci. Eng.*, vol. 3, no. 8, pp. 1–9, Aug. 2017.

[61] H. Blockeel and J. Struyf, “Efficient Algorithms for Decision Tree Cross-validation.,” *J. Mach. Learn. Res.*, vol. 3, pp. 621–650, 2002.

[62] Mathworks, “MATLAB.” https://www.mathworks.com/products/matlab.html

[63] “AMI in Review.pdf,” 2021.

[64] Hydro One, “Time-of-Use (TOU).” https://www.hydroone.com/rates-and-billing/rates-and-charges/electricity-pricing-and-costs