POLITECNICO DI TORINO

Master's Degree Thesis

Data Science and Engineering

Cyber-physical security: AI methods for malware/cyber-attacks detection on embedded/IoT applications

Supervisors

University Supervisor

Prof. Andrea Calimera

Company Supervisor

Ing. Jacopo Federici

Candidate

Umar Farooq

October 2023

Abstract

As the world becomes increasingly reliant on technology, the field of cybersecurity has gained paramount importance. The surge in the use of interconnected systems, particularly in the realm of autonomous vehicles, has escalated the risk of cyberattacks. Consequently, cyber-physical security has emerged as a critical area of research to address these concerns. The objective of this research was to delve into the field of cyber-physical security, focusing on the development of AI-based methods to detect cyberattacks on autonomous vehicles.

Machine learning, a subset of artificial intelligence, has been extensively employed in cybersecurity to develop automated methods for detecting cyberattacks. Deep learning, in particular, has shown promising results in detecting anomalies and identifying cyberattacks. However, the complexity of these systems and the dynamic nature of the data generated by them pose significant challenges in implementing effective machine learning-based solutions.

The research work presented in this thesis focused on the use of machine learning for cyber-attack detection in autonomous vehicles. A Simulink model was utilized to generate data and apply machine learning algorithms to detect cyberattacks. A Multi-layer Perceptron (MLP) model was selected as the final model, and the question of determining the number of layers and neurons in each layer was addressed using Neural Architectural Search (NAS). The final pipeline was written as clean code, and TensorFlow Lite was used to decrease the model size while maintaining accuracy.

Through extensive experimentation involving 125 different model configurations, we found that 98 models achieved a Mean Absolute Error (MAE) of less than 3. Given the scale of the target variable, which ranges from 0 to 210, this level of error represents highly accurate predictions, with errors constituting only 1.42% of the target variable's range.

The results of this research demonstrate that machine learning algorithms can be effectively used to detect cyberattacks in autonomous vehicles, providing a strong foundation for further research in this field. In conclusion, this research offers a comprehensive study of the field of cyber-physical security and the application of AI-based methods to detect cyberattacks in autonomous vehicles. The results underscore the potential of machine learning algorithms for detecting cyberattacks in autonomous vehicles and lay the groundwork for future research in this area.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Brief overview of the project's objective and main contributions

This thesis explores the crucial domain of cybersecurity in the context of autonomous vehicles. The proliferation of automation and artificial intelligence has led to a notable rise in the prevalence of autonomous vehicles within our transportation infrastructure. Nevertheless, this technological progression presents a fresh array of obstacles, primarily characterized by the increased vulnerability to cyber threats. This thesis aims to address the pressing concern of cyber attacks by specifically focusing on the real-time detection of such attacks on a particular component of autonomous vehicles, namely the "Pedal Press Percentage" Simulink model.

The Simulink model known as "Pedal Press Percentage" holds significant importance within the control system of autonomous vehicles, as it is responsible for regulating the vehicle's speed in accordance with the percentage of pedal press. The occurrence of any malevolent disruption to this model has the potential to result in significant ramifications, such as the relinquishment of command over the vehicle's velocity, which may ultimately lead to collisions. Hence, ensuring the protection of this model from cyber threats is of utmost significance.

The main aim of this thesis is to create and execute a resilient system that can effectively identify cyber attacks on the "Pedal Press Percentage" Simulink model in real time. This process entails the continuous observation of the system's actions and the detection of any deviations that may suggest the occurrence of a cyber assault. The detection system has been specifically engineered to exhibit a high degree of sensitivity and accuracy, enabling it to effectively differentiate between typical fluctuations in the system's functioning and potential cyber hazards.

When the system identifies a potential cyber attack, it is programmed to promptly transmit an alarm message to the cloud. This functionality facilitates

prompt reaction to identified cyber threats, enabling the timely implementation of measures to minimize the potential harm or interruption to the vehicle's functioning. The utilization of cloud technology in this particular setting not only expedites communication but also enables the retention and examination of data pertaining to identified cyber threats, which holds significant potential in the prevention of subsequent attacks.

The present thesis constitutes a noteworthy contribution to the domain of cybersecurity within the realm of autonomous vehicles. The thesis aims to address a significant deficiency in existing cybersecurity protocols for autonomous vehicles, which primarily adopt a reactive approach rather than a proactive one, by emphasizing the real-time identification of cyber attacks. Real-time detection and response to cyber attacks have the potential to greatly enhance the safety and dependability of autonomous vehicles.

Moreover, the thesis makes a valuable contribution to the wider domain of cybersecurity by showcasing the practical implementation of real-time cyber-attack detection within a specific context. The methods and technologies devised in this thesis have the potential for application in various contexts, thereby rendering this research pertinent not only to the domain of autonomous vehicles but also to the broader realm of cybersecurity.

This study represents a pertinent and significant undertaking, given the escalating risk posed by cyber-attacks targeting autonomous vehicles. The objective of this thesis is to improve the security and dependability of autonomous vehicles by creating a real-time cyber attack detection system that can promptly notify the appropriate systems. This research aims to contribute to the overarching objective of ensuring secure and safe transportation in an ever-more automated society.

1.2 Background and motivation for the project

1.2.1 Brief history of cyber-physical security

The term "cyber-physical system" was officially introduced in 2006 to encompass various systems that integrate the realms of cyber and physical domains. However, research and interest in this interdisciplinary interaction had already gained significant momentum in the 1990s. Nevertheless, the foundations of this discipline can be traced back to the invention of computers. The initial construction of the ENIAC, the first computer, took place in 1946. However, it was not until 1973 that the advent of real-time computations occurred, establishing the fundamental groundwork for the emergence of cyber-physical systems. Simultaneously, the Internet commenced its evolution, marked by the establishment of the initial network ARPANET in 1969. Consequently, by the conclusion of the 1990s, the amalgamation of communication and computation had reached its culmination. A

significant turning point in the establishment of a comprehensive and functional cyber-physical system occurred circa 1998, when sensors were developed with the capacity for sensing, communication, and computation. This advancement further enhanced the physical integration of cyber-systems. In recent years, there has been significant advancement in various domains, notably with the emergence of the Internet of Things (IoT) and a growing need for enhanced efficacy and quality in energy, transportation, healthcare, and water infrastructures.[1]

Concurrently with the advancement of cyber-physical system technology, researchers have devoted attention to the investigation of security concerns. This research aims to develop methods for detecting and thwarting potential intruders who may attempt to manipulate the functioning of a system. The historical account of cyber-physical systems encompasses various instances of significant and conspicuous attacks, such as the renowned Stuxnet incident. These occurrences are visually depicted in Figure 1.1, directly taken from [2]. Compiling a comprehensive report encompassing all instances of attacks is rendered unfeasible by the dearth of information available in numerous cases. The figures presented in the study exclusively include publicly reported attacks, as stated by the authors.

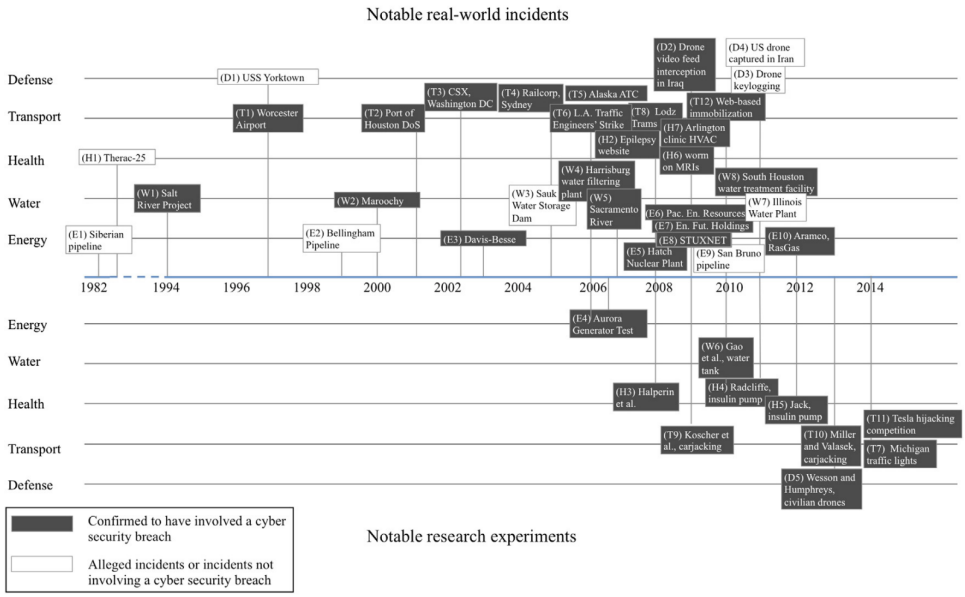

Figure 1.1: Historical timeline of known CPS attacks[2]

1.2.2 Cyber-Physical Security in Autonomous Vehicles: The Intersection of Connectivity and Vulnerability

Autonomous vehicles (AVs) embody a substantial advancement in transportation technology, holding the potential to fundamentally transform the manner in which we engage in travel. Nevertheless, the inherent characteristics that render autonomous vehicles (AVs) highly innovative, namely their connectivity and autonomy, also render them susceptible to a novel range of security vulnerabilities. These vehicles, as cyber-physical systems (CPS), are susceptible to cyber attacks due to their distinctive combination of physical components, such as sensors and actuators, with cyber components, including software and communication systems. This essay examines the ramifications of the connectivity of autonomous vehicles (AVs) on their cybersecurity, with a specific emphasis on the possibility of cyber attacks and the necessity for strong security measures in cyber-physical systems (CPS). [3] shows the importance of CPS in semi-autonomous vehicles.

Connectivity in Autonomous Vehicles

Connectivity is an integral component of AVs. These vehicles are not self-contained entities; rather, they are specifically engineered to establish connections with various systems, encompassing other vehicles (referred to as Vehicle-to-Vehicle or V2V communication), infrastructure (known as Vehicle-to-Infrastructure or V2I communication), and networks (referred to as Vehicle-to-Network or V2N communication). This communication enables AVs to navigate complex environments, adapt to changing conditions, and provide enhanced functionality to users.

As an illustration, an autonomous vehicle (AV) could employ vehicle-to-vehicle (V2V) communication to effectively synchronize its actions with other vehicles in congested traffic scenarios. Additionally, it could utilize vehicle-to-infrastructure (V2I) communication to receive real-time information regarding road conditions. Furthermore, the AV could engage in vehicle-to-network (V2N) communication to access navigation services or download software updates. The integration of connectivity in autonomous vehicles (AVs) significantly improves their operational capabilities and overall effectiveness, resulting in enhanced safety, efficiency, and user-friendliness.

The Need for Cyber-Physical Security in Autonomous Vehicles

Nevertheless, the availability of internet connectivity also renders autonomous vehicles vulnerable to potential cyber-attacks. Every communication channel possesses the potential to serve as a vulnerable point of access for a malicious actor. Through the strategic utilization of weaknesses present in the software or communication protocols of a vehicle, an assailant could potentially acquire

illicit entry into the various systems of said vehicle. This vulnerability has the potential to enable the assailant to manipulate the operational characteristics of the automobile, exfiltrate data, or potentially establish command over the vehicle.

These risks are not solely of a theoretical nature. Numerous instances of prominent demonstrations of such attacks have been documented. In a notable case from 2015, security researchers Charlie Miller and Chris Valasek successfully showcased their ability to remotely infiltrate a Jeep Cherokee by exploiting its internet-connected entertainment system. This unauthorized access allowed them to manipulate critical functions such as steering, brakes, and gearbox. The presented demonstration effectively underscored the inherent vulnerability of internet-connected vehicles to cyber-attacks, thereby emphasizing the consequential safety hazards associated with such threats.

In short, the advanced functionalities of autonomous vehicles (AVs) are made possible by their internet connectivity, but this connectivity also exposes them to a range of security threats. Cyber-physical systems, such as vehicles, possess distinctive vulnerabilities that render them susceptible to cyber attacks capable of causing significant physical ramifications. Consequently, the imperative to address the cyber-physical security of autonomous vehicles (AVs) arises due to their growing prevalence in the transportation domain. As society progresses towards a future characterized by the prevalence of autonomous vehicles (AVs), it becomes imperative to establish resilient security protocols capable of safeguarding these vehicles and their occupants against the potential hazards associated with cyber-attacks.

Chapter 2

Literature Review

2.1 State-of-the-art methods for cyber attack detection

The domain of cybersecurity assumes a crucial function in preserving essential systems and securing confidential data in our progressively interconnected global landscape. The risks and susceptibilities that can jeopardize the reliability, privacy, and accessibility of electronic resources increase in tandem with technological advancement. The purpose of this all-encompassing introduction is to establish a fundamental comprehension of cybersecurity, which includes essential principles, a lexicon, and the importance of safeguarding vital systems in the contemporary digital environment. The term security pertains to the act of safeguarding computer systems, networks, devices, and data from illicit access, exploitation, and malevolent attacks. The concept involves a variety of tactics, tools, and methodologies that have the collective objective of safeguarding data and preserving the soundness of computerized infrastructures. The domain is motivated by the necessity to mitigate risks such as cyber-attacks, data breaches, identity theft, malware, ransomware, and other forms of cybercrime that can result in significant ramifications for individuals, institutions, and even countries.

As stated in [4], malware detection techniques can be classified into three broad categories: signature-based, heuristic-based, and behavior-based. These methods rely on results from malware analysis, and each method has its unique advantages and challenges

Signature-based detection

This technique uses a known list of indicators of compromise (IOCs), which include specific byte sequences, API calls, file hashes, malicious domains, or network attack patterns. Signature-based detection is, however, incapable of detecting previously unknown or encrypted malware and does not require machine learning models.

Behavioural-based detection

Involves monitoring a suspected executable file in an isolated environment and collecting all exhibited behaviors, then using methods of extracting useful features by which a machine learning model can classify the malicious behavior.

Heuristic-based detection

This technique relies on generating rules based on the results of the static/dynamic analysis to guide the inspection of the extracted data to support the proposed malware detection model. Such rules can either be generated manually (relying on the expertise of the security analysts) or automatically, using machine learning or tools such as YARA.

With cyber-attacks posing significant threats to individuals, businesses, and even nations. As such, the development and implementation of effective cyber attack detection methods are of paramount importance. Several state-of-the-art methods have emerged in recent years, leveraging advancements in technology and computational techniques.

2.1.1 Intrusion Detection Systems (IDS)

In the realm of cybersecurity, Intrusion Detection Systems (IDS) have emerged as a pivotal component, adapting in tandem with the dynamic and evolving nature of cyber threats over the course of several years. Intrusion Detection Systems (IDS) are purposefully designed to monitor and analyze activities occurring within computer systems and networks, with the primary aim of detecting signs of unauthorized access or potential security vulnerabilities. Intrusion detection systems can be broadly categorized into two main types: Network Intrusion Detection Systems (NIDS) and Host Intrusion Detection Systems (HIDS). Network Intrusion Detection Systems (NIDS) are specifically engineered to meticulously examine network traffic with the objective of detecting and analyzing potentially irregular patterns that may indicate the occurrence of a network attack. In contrast, Host Intrusion

Detection Systems (HIDS) are purposefully engineered to oversee an individual host, such as a computer or server, with the objective of identifying and evaluating any potentially anomalous behaviors that might transpire on that specific host.

The evolution of Intrusion Detection Systems (IDS) has undergone a shift from rule-based systems that rely on predetermined signatures of known threats to anomaly-based systems that utilize machine learning algorithms to establish a baseline for normal behavior and detect deviations as potential threats. Anomaly-based intrusion detection systems (IDS) demonstrate enhanced effectiveness in identifying and detecting previously unknown and unclassified security breaches, commonly known as zero-day attacks. However, it is important to note that these systems may encounter challenges in the form of elevated rates of false positives. The present state of intrusion detection systems (IDS) is distinguished by the incorporation of sophisticated machine learning and artificial intelligence methodologies. Deep learning, which falls under the umbrella of machine learning, has exhibited significant potential. Deep learning-based intrusion detection systems (IDS) have demonstrated the ability to effectively capture intricate patterns and correlations, thereby enhancing the identification of advanced, multi-step attacks. Hybrid intrusion detection systems (IDS), which amalgamate the advantageous features of signature-based and anomaly-based methodologies, are increasingly prevalent in contemporary research and practice.

2.1.2 Artificial Intelligence and Machine Learning

Traditionally, the field of cybersecurity has predominantly relied on detection methods that are based on signatures, necessitating prior knowledge of attack patterns. Although these methods have demonstrated efficacy in countering established threats, they face challenges in detecting emerging or advanced attacks. The emergence of artificial intelligence (AI) and machine learning (ML) has effectively mitigated this constraint by facilitating the proactive identification of both familiar and unfamiliar risks.

Artificial intelligence (AI) and machine learning (ML) algorithms possess the capability to acquire knowledge from past instances of cyber attacks, discern patterns within the data, and make predictions regarding forthcoming attacks. Cybersecurity systems possess the capability to adjust and respond to emerging threats, thereby rendering them proficient in countering zero-day attacks, which are characterized by their novelty and lack of prior detection. These technologies possess the capability to efficiently process extensive volumes of data, swiftly detecting irregularities that may potentially signify a cyberattack. This phenomenon proves to be highly advantageous in the current era of digitalization, characterized by an

exponential surge in data accumulation. Experimental analysis shows us that deep learning algorithms can detect attacks with higher performance that usual methods and can make cyber security simpler and more proactive.[5]

Deep Learning

Deep Learning, which falls under the umbrella of Machine Learning, employs neural networks consisting of multiple layers (referred to as deep networks) to effectively represent and comprehend intricate patterns. The utilization of this technology has facilitated the creation of models capable of detecting highly intricate cyberattacks. The identification and classification of anomalies within a given dataset is commonly referred to as anomaly detection.

Anomaly Detection

Anomaly detection algorithms are employed for the purpose of identifying aberrant data points that may potentially signify a cyberattack. The algorithms utilized in this context encompass statistical methods, clustering techniques, classification approaches, and neural network-based models.

Reinforcement Learning

Reinforcement Learning is a machine learning paradigm in which an autonomous agent acquires the ability to make decisions by iteratively interacting with an environment with the objective of maximizing a predefined reward signal. The utilization of this technology has been observed within the realm of cybersecurity, where it has been employed to facilitate the creation of systems capable of adapting and effectively responding to various forms of cyberattacks.

2.1.3 Blockchain Technology

The utilization of blockchain technology, which was originally created for digital currencies such as Bitcoin, has been extended to a wide range of domains, encompassing cybersecurity among others. The distinctive attributes of blockchain technology, including its decentralized nature, immutability, and transparency, render it a powerful instrument for the detection and prevention of cyberattacks.

Blockchain is a decentralized technology for maintaining a ledger that disperses data across numerous systems within a network, thereby enhancing its resilience

against technical malfunctions and malicious intrusions. In the blockchain, every block comprises a collection of transactions, and these blocks are interconnected through the utilization of cryptographic hashes. The aforementioned architectural design guarantees the immutability and non-deletability of data once it is appended to the blockchain, thereby establishing a dependable and tamper-resistant ledger of transactions.

In the realm of cybersecurity, blockchain technology has the potential to effectively tackle a multitude of challenges.

Data Integrity

The property of immutability inherent in blockchain technology serves to safeguard the integrity of data. Once data has been recorded on a blockchain, it becomes immutable and resistant to any form of alteration or tampering. The inclusion of this particular feature holds significant importance in the realm of cyber attack detection, as it furnishes a dependable chronicle of occurrences, thereby facilitating the identification and examination of dubious actions.

Decentralization

The decentralized nature of blockchain technology mitigates the potential for single-point failures. In conventional centralized systems, the compromise of the central system poses a significant risk to the entire network. In contrast, within a blockchain network, the security of the entire network is maintained even in the event of a compromise at a single node.

Transparency and Traceability

Transparency in every transaction on a blockchain makes it possible to track its history. This functionality possesses the capability to identify and examine instances of cyberattacks. In the context of a network monitoring system based on blockchain technology, the identification of an intrusion enables the subsequent tracing of its origin by leveraging the inherent transparency and immutability of the blockchain.

2.1.4 Security Information and Event Management (SIEM)

Security Information and Event Management (SIEM) systems have emerged as a fundamental component of contemporary cybersecurity frameworks. SIEM systems play a vital role in the identification, prevention, and response to cyber attacks by offering instantaneous analysis of security alerts produced by applications and network hardware. SIEM is the state-of-the-practice in handling heterogeneous data sources for security analysis.[6]

SIEM technology integrates two distinct product categories, namely Security Information Management (SIM) and Security Event Management (SEM). SIM products are designed to gather, examine, and present log data, whereas SEM systems perform real-time analysis of log and event data to offer threat monitoring, event correlation, and incident response capabilities. The integration of these two categories within the framework of Security Information and Event Management (SIEM) has yielded systems that offer a holistic perspective on an organization's information security.

Security Information and Event Management (SIEM) systems are of paramount importance in the realm of cyber attack detection. The company offers a comprehensive solution that enables the real-time analysis of security alerts generated by network hardware and applications, thereby playing a crucial role in the identification, mitigation, and management of cyber threats. SIEM systems operate by aggregating data from various sources, including network devices, security controls, systems, and applications. The dataset possesses the capacity to encompass diverse forms of data, including logs, events, network flows, and user behavior data. The ability to collect and analyze a wide variety of data enables Security Information and Event Management (SIEM) systems to detect anomalies and suspicious behaviors that may indicate a potential cyberattack. To exemplify, a Security Information and Event Management (SIEM) system possesses the capacity to detect a substantial number of unsuccessful login attempts followed by a subsequent successful login, potentially indicating the occurrence of a brute force attack. The ability to establish connections between events originating from diverse sources is widely recognized as a highly influential characteristic of Security Information and Event Management (SIEM) systems. The aforementioned capability holds significant importance in the detection of complex, multi-phased attacks that involve multiple systems. If a user attempts to log in from an unusual location and then proceeds to download a large amount of data, a Security Information and Event Management (SIEM) system can establish a correlation between these events. As a result, the Security Information and Event Management (SIEM) system can produce an alert, thereby recognizing this activity as a credible cyber attack. SIEM systems provide dashboards and visualization tools, which aid security analysts in understanding

and examining the collected data. These tools enhance the ability to recognize patterns, identify irregularities, and empower analysts to make informed judgments regarding potential risks. Furthermore, Security Information and Event Management (SIEM) systems provide a variety of tools that enable the process of forensic analysis. These tools provide analysts with the capability to thoroughly investigate the complexities of a security incident, understand its consequences, and develop an appropriate action plan.

In summary, Security Information and Event Management (SIEM) systems exemplify a cutting-edge approach to identifying and mitigating cyber attacks. Through the provision of real-time analysis, event correlation, and incident response capabilities, these systems empower organizations to effectively and efficiently detect and respond to cyber threats. The ongoing evolution of cyber threats necessitates the continued significance of SIEM systems in the realm of cyber attack detection.

2.2 Applications of artificial intelligence for cyber attack detection

The escalating frequency and complexity of cyberattacks present substantial risks to the security of information, financial stability, and individual privacy. The conventional approaches to safeguarding against cyber threats, including the utilization of firewalls and antivirus software, remain crucial; however, they are no longer adequate for effectively mitigating the risks posed by these ever-changing threats. Artificial Intelligence (AI) emerges as a pivotal factor in the realm of cybersecurity and the detection of cyberattacks, presenting a novel frontier. Due to the rapid development of Internet-connected systems and Artificial Intelligence in recent years, Artificial Intelligence including Machine Learning (ML) and Deep Learning (DL) has been widely utilized in the fields of cyber security including intrusion detection, malware detection, and spam filtering.[7]

Artificial intelligence (AI) is significantly transforming various domains, including cybersecurity, due to its capacity to acquire knowledge through experience, analyze vast quantities of data, and generate predictions. The utilization of artificial intelligence (AI) within the realm of cybersecurity is a direct reaction to the imperative for enhanced and intricate approaches in the identification and mitigation of cyber threats. The current landscape of cyber threats is characterized by a simultaneous rise in both quantity and complexity, with an increasing tendency for these threats to be engineered in a manner that circumvents conventional security protocols.

Artificial intelligence (AI) has the potential to effectively tackle these challenges

through the automation and augmentation of cyber defense systems. The system can efficiently analyze extensive volumes of data to identify irregularities, recognize patterns that may signify cyber attacks, and potentially forecast forthcoming threats by extrapolating from prevailing trends. Moreover, artificial intelligence (AI) possesses the ability to adjust to the constantly evolving strategies employed by cybercriminals. It achieves this by assimilating knowledge from emerging threats and subsequently developing effective countermeasures.

The incorporation of artificial intelligence (AI) into cybersecurity strategies embodies a proactive stance toward safeguarding against cyber threats. The utilization of AI technology allows for proactive measures in addressing potential threats, leading to the enhancement of digital system security as opposed to solely reacting to threats in real time. As society grapples with the intricate dynamics of the digital realm, the prominence of artificial intelligence (AI) in the domain of cybersecurity is poised to escalate in significance.

2.2.1 Understanding AI in Cybersecurity

Artificial Intelligence (AI) pertains to the emulation of human cognitive processes through the utilization of machines, particularly computer systems. The aforementioned processes encompass learning, which involves the acquisition of information and the rules governing its utilization; reasoning, which entails the application of rules to arrive at either approximate or definitive conclusions; and self-correction. Within the realm of cybersecurity, artificial intelligence (AI) represents a significant paradigm shift, providing sophisticated and automated methodologies for identifying, mitigating, and addressing cyber risks.

Artificial intelligence (AI) plays a pivotal role in the field of cybersecurity by streamlining the intricate procedures involved in identifying and responding to cyberattacks and breaches. The crux of the matter lies in the capacity of artificial intelligence (AI) to acquire knowledge and adjust accordingly. Through the analysis of extensive datasets and the acquisition of knowledge from these datasets, artificial intelligence (AI) has the potential to detect and discern threats or malevolent activities that may elude human perception or require significantly more time for human identification.

There exist multiple categories of artificial intelligence (AI), namely Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP), each possessing distinct applications in the realm of cyber attack detection.

Machine Learning (ML)

Machine Learning, a subfield within the domain of Artificial Intelligence (AI), involves the use of algorithms to examine data, extract knowledge from it, and subsequently produce a conclusion or prediction related to a specific aspect of the physical world. Machine learning (ML) algorithms are commonly used in the domain of cybersecurity to detect anomalies. The procedure of anomaly detection involves the recognition of atypical patterns or outliers within a given dataset. These anomalies potentially indicate the presence of a cyberattack, characterized by abnormal levels of network traffic, a high frequency of unsuccessful login attempts, or unusual user behavior.

The utilization of machine learning (ML) has the capacity to be harnessed for the application of predictive analytics in the domain of cybersecurity. By analyzing historical data, machine learning algorithms demonstrate the ability to identify repetitive patterns and emerging trends, thereby enabling the prediction of future cyberattacks. This facilitates the implementation of proactive strategies by organizations to mitigate and preempt such attacks.

Deep Learning (DL)

Deep Learning, a subfield within the broader discipline of Machine Learning, is based on the application of artificial neural networks that have the capacity to acquire knowledge and derive significant representations. The system possesses the capacity to effectively manage a wide range of data resources, consequently mitigating the necessity for human operators to engage in labor-intensive data engineering tasks. The utilization of deep learning (DL) has exhibited considerable effectiveness in the domain of cyberattack detection as a result of its capacity to construct complex architectures comprising multiple layers of processing. This enables the modeling of sophisticated abstractions at a higher level within datasets.

For example, deep learning (DL) can be employed for the purpose of identifying zero-day exploits, which refer to attacks that exploit software vulnerabilities prior to their disclosure by the vendor. Conventional detection methodologies frequently prove inadequate in identifying emerging threats, as they heavily depend on the recognition of signatures or established attack patterns. On the contrary, deep learning has the capability to detect abnormal actions linked to these exploits even when there are no established patterns available.

Natural Language Processing (NLP)

The field of Natural Language Processing encompasses the dynamic interplay between computational systems and human language. This capability enables systems to comprehend, evaluate, and produce human language in a meaningful manner. Within the realm of cybersecurity, Natural Language Processing (NLP) can be effectively employed for the purpose of scrutinizing textual data, encompassing mediums such as emails or social media posts, with the aim of identifying and discerning potential security risks.

One illustrative application of natural language processing (NLP) involves its utilization in the detection of phishing endeavors within electronic mail communications. The analysis of an email's language, contextual information, and embedded links can be employed to ascertain its potential as a phishing email. Natural Language Processing (NLP) can also be employed for the purpose of identifying malevolent Uniform Resource Locators (URLs) or code within social media posts or web pages.

Artificial intelligence (AI), along with its various subfields such as machine learning (ML), deep learning (DL), and natural language processing (NLP), assumes a pivotal role in contemporary cybersecurity. The implementation of artificial intelligence (AI) in the detection process and its ability to acquire knowledge from each potential threat enable a heightened level of proactivity and efficacy in the realm of cyber defense. The increasing complexity and evolution of cyber threats necessitate the utilization of artificial intelligence (AI) in the field of cybersecurity. It is anticipated that the integration of AI will enhance the sophistication and effectiveness of cyberattack detection.

The market for artificial intelligence in cybersecurity is also anticipated to grow during the forecast period as a result of the proliferation of 5G technology and the rising demand for cloud-based security solutions among small and medium-sized organizations. Artificial intelligence in cybersecurity is currently gaining popularity to secure information. Because end users are anticipated to embrace AI in cybersecurity to address security concerns and spot new types of assaults that can occur at any time, the market for artificial intelligence in cybersecurity is growing steadily.

2.2.2 Applications of AI in Cyber Attack Detection

The utilization of Artificial Intelligence (AI) has significantly contributed to the improvement of our capacity to identify and mitigate cyberattacks. The capacity to acquire knowledge from data, discern patterns, and generate forecasts has

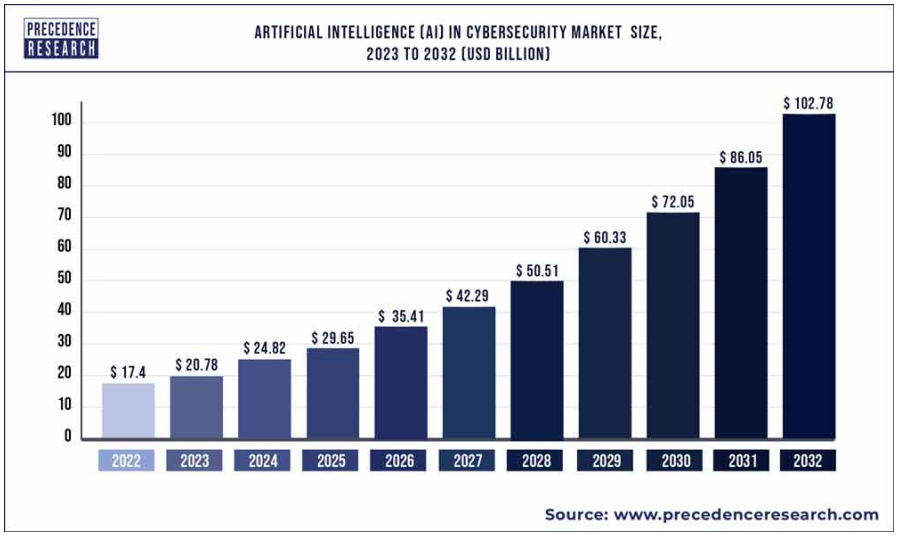

Figure 2.1: AI in Cyber Security

resulted in the advancement of increasingly sophisticated and efficient strategies for safeguarding against cyber threats. There are several applications of artificial intelligence (AI) in the realm of cyberattack detection.

Anomaly Detection

Anomaly detection stands out as a prominent application of artificial intelligence within the field of cybersecurity. Artificial intelligence algorithms have the capability to undergo training in order to establish a foundational level of typical behavior within a given system or network. Following this, the system can be observed for any deviations from the established baseline, which may indicate the possible existence of a threat. For instance, artificial intelligence possesses the capacity to monitor network traffic and identify anomalous patterns, such as an unusually high volume of data being transmitted or an unexpected increase in login attempts. The presence of these irregularities could potentially indicate the manifestation of a cyberattack, such as a compromise of data or a brute-force assault. The phenomenon in question can be exemplified by the cybersecurity platform Darktrace. The system employs artificial intelligence (AI) to establish a comprehensive "pattern of life" for every user and device present in a given network. Following this, it then proceeds to detect any deviations from these established patterns, acknowledging them as possible risks.

Phishing Detection

The utilization of artificial intelligence (AI) holds promise in the realm of identifying and detecting phishing attempts. The analysis covers various aspects of an email, such as the metadata, textual content, and embedded hyperlinks, aiming to determine if the email under consideration displays features that suggest it may be a phishing attempt. As an example, artificial intelligence (AI) possesses the capacity to identify suspicious phrases commonly utilized in phishing emails. Moreover, it has the capability to distinguish whether an email is derived from a domain that exhibits similarities to a genuine domain, although not an exact replica. To recognize and counteract phishing attempts, Google's Gmail service makes use of artificial intelligence (AI) technology. Machine learning algorithms are utilized for the purpose of examining patterns present in emails and identifying recurring characteristics that are suggestive of phishing attempts.

Malware Detection

Artificial intelligence (AI) possesses the ability to identify and categorize malevolent files, even in instances where these files have been altered with the intention of evading traditional antivirus software. The procedure entails the examination of the characteristics of a provided file and the performance of a comparative evaluation with established malware profiles in order to determine its potential for malicious intent. One illustrative instance involves the utilization of artificial intelligence (AI) to discern polymorphism malware, denoting malicious software that modifies its code to evade detection. Traditional antivirus software, which relies on pre-existing malware signatures, often demonstrates limited effectiveness in detecting such forms of threats. Nevertheless, artificial intelligence (AI) has the ability to identify the underlying patterns in malware behavior, allowing for its detection even when the code is modified. Cylance is a Cyber Security Enterprise that focuses on utilizing artificial intelligence (AI) algorithms to identify and address malware threats. The antivirus software, which employs artificial intelligence, has the ability to rapidly identify and preemptively block the execution of malicious software.

Predictive Analytics

Artificial intelligence (AI) demonstrates the ability to analyze data patterns and trends, thereby enabling it to generate predictions pertaining to future cyberattacks. This facilitates the implementation of proactive strategies by organizations to mitigate and preempt such attacks. One instance where artificial intelligence (AI) demonstrates its potential is in the analysis and interpretation of patterns found

in network traffic, user behaviour, and existing cyber threats. This enables AI to predict the possible timing and location ofpotential Cyberattacks. This may involve the recognition of Verhaltenal patterns associated with a particular type of Cyberattack or the anticipation of targeted systems through the evaluation of their vulnerabilities and the importance of the data they hold. Crowdstrike, a Cyber Security organisation, employs artificial intelligence (AI) in the Application of predictive analytics. The Falcon platform utilizes machine learning algorithms for the purpose of data analysis and the anticipation of potential security risks. Artificial Intelligence (AI) possesses a broad spectrum of applications within the realm of cyberattack detection. These applications encompass the identification of anomalies and phishing endeavors, as well as the detection of malware and the anticipation of forthcoming threats. These applications exemplify the capacity of artificial intelligence (AI) to augment our capability in identifying and mitigating cyber attacks, thereby emphasizing the significance of AI in contemporary cybersecurity.

2.2.3 Advantages and Challenges of AI in cyber security domain

The field of cyberattack detection benefits from various notable advantages brought about by Artificial Intelligence (AI). The capacity to acquire knowledge from data, discern patterns, and formulate predictions has significantly transformed our approach to cybersecurity. Machine learning has been strongly incorporated into modern cybersecurity-related technology. [8] does a literature review and investigates the general effects of AI on cybersecurity.In their results, they have obtained positive findings in attacks with artificial intelligence and have found certain results when obtaining information on attacks using AI. The following are several notable benefits:

Increased Accuracy

One of the primary benefits associated with the utilization of artificial intelligence (AI) in the realm of cyber attack detection is the heightened level of precision it affords. Conventional cybersecurity measures frequently depend on pre-established rules and recognized attack signatures. Although the effectiveness of these methods has been proven in detecting known threats, their capability to detect new or complex attacks is often compromised. Artificial intelligence (AI) has the ability to acquire knowledge from data and identify patterns that may indicate a cyberattack, even when the attack does not conform to established signatures. The potential impact of implementing this approach is significant in terms of reducing the

occurrence of both false positives and false negatives, thereby improving the precision of cyber threat detection.

Ability to Analyze Large Amounts of Data

The discipline of cybersecurity involves the analysis of large volumes of data with the aim of identifying and evaluating potential vulnerabilities. The successful accomplishment of this task can present considerable difficulty for individuals, however, it is a domain in which artificial intelligence exhibits remarkable aptitude. Artificial intelligence (AI) algorithms have the capacity to effectively analyze extensive volumes of data, frequently reaching the magnitude of terabytes. The aforementioned analytical procedure facilitates the detection of regularities and deviations within the dataset, which may potentially function as indicators of a cyber intrusion. The capacity to effectively handle and evaluate substantial volumes of data not only enhances the efficiency of identifying cyberattacks promptly but also enables the identification of intricate, multi-phased attacks that would pose challenges for manual analysis.

Detection of Unknown Threats (Zero-Day Attacks)

Zero-day attacks, characterized by the exploitation of undisclosed vulnerabilities, present a substantial obstacle to conventional cybersecurity protocols. Given that these attacks lack correspondence with established signatures, they frequently evade detection by conventional security systems. Artificial intelligence (AI), on the other hand, has the potential to assist in mitigating this particular obstacle. Through the examination of behavioral patterns and the identification of deviations, artificial intelligence (AI) has the capability to discern atypical activities that could potentially signify the occurrence of a zero-day attack. The capability to identify unfamiliar threats represents a notable benefit of artificial intelligence (AI) in the realm of cyberattack detection.

Proactive Threat Detection

The utilization of artificial intelligence (AI) in forecasting forthcoming cyber attacks by analyzing data patterns and trends signifies a transition from a reactive to a proactive approach in the field of cybersecurity. Rather than adopting a reactive approach to cyberattacks, the implementation of artificial intelligence (AI) empowers organizations to proactively identify and mitigate potential threats. The implementation of a proactive approach has the potential to greatly mitigate the detrimental impact of cyberattacks.

2.2.4 Challenges and Limitations of Using AI for Cyber Attack Detection

In the area of cyber attack detection, artificial intelligence (AI) offers notable advantages, but it also comes with a number of difficulties and limitations. [8] concluded that it is necessary to advance in artificial intelligence since the increasing volume and complexity of attacks at the international level require more resources to face them. Cybercriminals will also use artificial intelligence to attack individuals, state infrastructure, and systems. A comprehensive comprehension of these concepts is imperative for the proficient utilization of artificial intelligence in the realm of cybersecurity. The following are several significant challenges:

High False-Positive Rates

One of the primary obstacles encountered when employing artificial intelligence (AI) for the purpose of cyber attack detection pertains to the possibility of elevated rates of false-positive outcomes. Although artificial intelligence (AI) has the capability to detect patterns and anomalies that could potentially signify a cyberattack, it is important to note that not all anomalies necessarily imply malicious intent. This phenomenon has the potential to result in a significant number of false-positive outcomes, wherein benign activities are erroneously identified as possible threats. The act of wastefully utilizing resources not only incurs unnecessary costs but also has the potential to induce 'alert fatigue' among security teams, wherein the excessive occurrence of false positives renders them less responsive to alerts.

The 'Black Box' Problem

Interpreting AI systems, especially those that rely on intricate machine learning algorithms, can pose challenges. Frequently, these systems function in an opaque manner, employing intricate computations that are not readily comprehensible to human beings. The absence of transparency poses a notable obstacle in the field of cybersecurity, as comprehending the rationale behind a decision is imperative to effectively addressing a potential threat. Furthermore, the utilization of AI systems also gives rise to concerns regarding accountability and trust.

Risk of AI Systems Being Manipulated

Although artificial intelligence (AI) has the potential to improve cybersecurity measures, malicious actors can also use it to carry out more intricate and sophisticated attack strategies. The automation of attacks is one way that malicious actors can use artificial intelligence (AI), allowing them to run more effective phishing campaigns or get around AI-driven security systems. Another potential concern is the existence of 'adversarial attacks', wherein malicious actors manipulate the input data provided to an artificial intelligence (AI) system with the intention of deceiving it into producing erroneous outcomes.

Dependence on Quality Data

The efficacy of artificial intelligence (AI) in the detection of cyberattacks is heavily contingent upon the caliber of the data on which it is trained. In the event that the training data exhibits bias, incompleteness, or obsolescence, the AI system may encounter challenges in effectively identifying and discerning threats with precision. The task of gathering and preserving accurate and current data poses a substantial obstacle.

2.2.5 Future Trends in the Use of AI for Cyber Attack Detection

The increasing sophistication of Artificial Intelligence (AI) is expected to enhance its utilization in the realm of cyberattack detection. The following discourse presents a selection of prospective future trends pertaining to the utilization of artificial intelligence (AI) for the purpose of detecting cyberattacks.

Development of More Sophisticated AI Algorithms

With the ongoing advancement of AI research, it is reasonable to anticipte the emergence of increasingly intricate AI algorithms. It is highly probable that these algorithms will possess the capability to analyze significantly larger quantities of data, discern more intricate patterns, and generate predictions with heightened accuracy. This has the potential to enhance the efficacy of cyberattack detection by encompassing intricate, multi-phased attacks and zero-day exploits.

Integration of AI with Other Technologies

The incorporation of artificial intelligence (AI) with other technological advancements represents a highly encouraging trajectory. The integration of artificial intelligence (AI) with blockchain technology has the potential to enhance data security and privacy. This integration can create a more robust system that makes it increasingly difficult for malicious actors to manipulate or steal data. The utilization of blockchain technology has the potential to offer transparent and unalterable documentation of cyberattacks, thereby facilitating incident response and forensic analysis. Another prospective integration can be observed in the realm of quantum computing. The utilization of quantum computers, due to their remarkable data processing capacity and accelerated computational capabilities, has the potential to greatly augment the effectiveness of artificial intelligence in the realm of cyberattack detection.

Use of AI for Automated Incident Response

Artificial intelligence (AI) possesses utility not solely in the realm of cyber attack detection but also in the domain of cyber attack response. In the forthcoming years, it is anticipated that there will be an increased utilization of artificial intelligence (AI) in the realm of automated incident response. This scenario encompasses the utilization of artificial intelligence (AI) systems to autonomously execute remedial measures upon identification of a potential security breach. Potential measures that can be implemented may involve isolating compromised systems or implementing restrictions on access from malicious IP addresses. The implementation of this approach holds the capacity to significantly reduce the duration required to respond to a cyberattack, thereby diminishing the potential magnitude of damage.

AI-Driven Threat Intelligence

The field of threat intelligence stands to benefit significantly from the integration of artificial intelligence (AI) technology. By analyzing data from various sources, artificial intelligence (AI) has the ability to identify patterns and trends in the field of cyber threats. This enables AI to provide valuable insights into potential risks. The utilization of this methodology possesses the capacity to augment proactive behavior within organizations in the domain of cybersecurity, as it facilitates the execution of precautionary actions that are guided by projected risks.

Explainable AI (XAI)

The issue of the 'black box' phenomenon in artificial intelligence has generated significant attention, leading to an increasing focus on the concept of explainable AI (XAI). Explainable Artificial Intelligence (XAI) pertains to the field of AI systems that possess the capability to offer transparent and comprehensible justifications for the decisions they make. The comprehension of the rationale behind a decision holds significant importance in the field of cybersecurity, particularly in the context of incident response. The potential for artificial intelligence (AI) in the realm of cyberattack detection appears to be highly favorable. The ongoing development and integration of artificial intelligence (AI) in conjunction with other technological advancements are poised to assume a progressively substantial role in the field of cybersecurity. This integration holds the potential to augment our capacity to identify and counteract cyber threats. Nevertheless, like any technological advancement, it is imperative to approach these developments with a cautious and well-informed comprehension of the potential obstacles and ethical implications.

2.3 Importance of cybersecurity in autonomous vehicles

Autonomous vehicles, also known as self-driving cars, have emerged as a revolutionary technological advancement in the field of transportation. As discussed in [9], with connected autonomous vehicles the protection from external attack will be an essential requirement, motivated by the outstanding safety implications of an autonomous vehicles remotely controlled by an attacker or a "malware". However, the automotive industry still lacks reliable and repeatable methods to assess the cybersecurity level of modern cars. The previously mentioned vehicles are equipped with sophisticated sensors, algorithms that employ artificial intelligence, and advanced computing capabilities, enabling them to operate autonomously without human intervention. The advent of autonomous vehicles has the potential to significantly transform various aspects of our society, including transportation, urban planning, safety, and environmental sustainability. This article explores the importance of autonomous vehicles and their impact on the future of transportation.

The significance of cybersecurity in the realm of autonomous vehicles cannot be emphasized enough, particularly as we progress towards a future where it plays a more prominent role. Autonomous vehicles, due to their utilization of intricate software systems, connectivity, and data interchange, pose a novel array of challenges and susceptibilities that render them appealing to cybercriminals. The assurance of security for these vehicles extends beyond the safeguarding of data

and the prevention of financial losses; it encompasses the imperative of upholding public safety.

The advent of autonomous vehicles has brought forth a distinct array of cybersecurity challenges due to their intricate network of interconnected systems and dependence on external communications. These vehicles can be considered mobile data centers, as they continuously process substantial volumes of data in real-time to facilitate navigation, decision-making, and external communication. The intricate nature and interconnectivity of systems provide cybercriminals with numerous opportunities to exploit vulnerabilities.

2.3.1 Potential Targets in Autonomous Vehicles

Navigation Systems

Autonomous vehicles heavily depend on their navigation systems, which encompass GPS and mapping data. The manipulation of data through a cyberattack has the potential to result in the vehicle's deviation from its intended course or, in more severe cases, lead to a collision.

Sensors and Cameras

Autonomous vehicles employ a variety of sensors, such as LIDAR, radar, and ultrasonic sensors, in conjunction with cameras to effectively perceive and comprehend their immediate environment. The potential targets of a cyberattack could include vehicles, where false information could be injected to disrupt their normal behavior and induce unpredictable responses.

Communication Systems

Autonomous vehicles establish communication links with other vehicles, infrastructure elements, and potentially a central control system, commonly referred to as vehicle-to-Everything (V2X) communication. The potential exists for an assailant to intercept and manipulate these communications or exploit them as a means to gain unauthorized access to the internal systems of the vehicle.

Control Systems

The control systems of the vehicle, responsible for regulating the throttle, brakes, and steering, may potentially become susceptible to cyberattacks. The acquisition

of control over these systems could potentially enable an assailant to assume command of the vehicle.

Data Storage

Autonomous vehicles possess the capability to generate and store substantial volumes of data. This may serve as an attractive target for cybercriminals seeking to illicitly acquire personal information or proprietary data.

2.3.2 Potential Consequences of a Cyber Attack

Safety Risks

The primary and gravest outcome arising from a cyberattack on an autonomous vehicle is the inherent risk of causing physical harm. In the event that an assailant successfully acquires command over a vehicle or manipulates its sensor data with deceptive information, the potential consequence may manifest in the form of a collision. The potential for accidents exists even in instances where non-physical attacks, such as impairing a vehicle's sensors through a cyberattack, are employed.

Privacy Breaches

Autonomous vehicles amass a substantial amount of data, a portion of which pertains to personal and sensitive information. A cyberattack has the potential to result in the unauthorized acquisition of this data, thereby resulting in breaches of privacy and the potential for identity theft.

Financial Loss

Cyber attacks could lead to financial loss, both for individuals (through data theft or damage to the vehicle) and for companies (through damage to their reputation, loss of customer trust, or regulatory fines).

Disruption of Services

For companies operating fleets of autonomous vehicles (like ride-hailing or delivery services), a cyber attack could disrupt their services, leading to financial loss and damage to their reputation.

Chapter 3

Methodology

In order to tackle the research question at hand, our efforts were concentrated on an ongoing project named "Evergrin," spearheaded by Brain Technologies. The primary objective of this project is to devise effective solutions for the detection of cyber attacks, particularly in the context of autonomous vehicles.

Given the complexity and the broad scope of autonomous vehicle systems, it is essential to adopt a targeted approach for our research. Therefore, we decided to focus on a specific component of the autonomous vehicle system, rather than the entire system itself. This approach allows us to delve deeper into the intricacies of the selected component and develop a comprehensive understanding of its vulnerabilities and potential countermeasures.

The platform chosen for this purpose is a Simulink model of an autonomous vehicle. Simulink, a MATLAB-based environment, offers a graphical interface for modeling, simulating, and analyzing dynamic systems. It is particularly well-suited for our research as it allows us to create a realistic and controllable environment to test our methodologies.

Within this Simulink model, we decided to focus on the "Pedal Press Percentage" model. This model represents a critical aspect of the vehicle's control system, specifically, the degree to which the accelerator or brake pedal is pressed. By focusing on this model, we can explore the potential cyber threats that could manipulate the vehicle's speed control, leading to dangerous situations.

Our research will involve applying various machine learning and artificial intelligence techniques to this "Pedal Press Percentage" model. The goal is to detect any anomalies or deviations that could indicate a cyber attack. By focusing our efforts on this specific model within the broader autonomous vehicle system, we aim to develop a robust and effective methodology for real-time cyber attack detection.

Our decision to concentrate on the "Pedal Press Percentage" model is strategic and purposeful. This model is a critical component of the autonomous vehicle's control system, directly influencing the vehicle's speed and, consequently, its overall

operation and safety. A cyber attack targeting this system could lead to a loss of control over the vehicle's speed, posing significant safety risks. Therefore, developing robust cyber attack detection mechanisms for this model is of utmost importance.

The methodology we will employ involves applying advanced machine learning and artificial intelligence techniques to the data generated by the "Pedal Press Percentage" model. These techniques are capable of learning from the data, identifying patterns, and detecting anomalies that could indicate a cyber attack. The goal is to develop a system that can detect cyber attacks in real-time, enabling immediate response and mitigation.

To achieve this, we will first use the Simulink model to generate a dataset that reflects both normal operation and various cyber attack scenarios. This dataset will then be used to train and test our machine learning algorithms. The performance of these algorithms will be evaluated based on their ability to accurately and promptly detect cyber attacks.

3.1 Pedal Press Percentage Model and its Human Machine Interface (HMI)

The focus of our research is the "Pedal Press Percentage" model within an autonomous vehicle's control system, simulated using Simulink, a MATLAB-based environment. This model is integral to the vehicle's operation as it directly influences the vehicle's speed by determining the degree to which the accelerator pedal is pressed.

The operation of this model begins with an input received through a knob, which represents the desired speed or acceleration. This input is then processed by the model, which generates two intermediate signals, referred to as "throttle 1" and "throttle 2". These signals are crucial in the computation of the final output, which is the percentage by which the accelerator pedal has been pressed.

The communication within this model, including the transmission of the input and the intermediate signals, is facilitated by Controller Area Network (CAN) messages. CAN is a standard designed to allow microcontrollers and devices to communicate with each other within a vehicle without a host computer. In our model, we have three CAN blocks, each corresponding to the input, "throttle 1", and "throttle 2".

Each CAN message consists of 13 attributes, providing detailed information about the message such as its identifier, data length code, data field, and more. Given that we have three CAN blocks, this results in a total of 39 attributes (13 attributes per CAN message x 3 CAN blocks).

In addition to these, we also have the timestamp, which provides the exact time at which the data point was recorded. We also record the values of the input,

"throttle 1", "throttle 2", and the final pedal press percentage. This brings the total number of attributes for each data point to 44.

The data from this Simulink model is extracted and stored in Excel files. Each row in these files represents a single data point, with the 44 attributes spread across the columns. This structured data serves as the foundation for our machine learning model.

The collected data is then used to train a deep learning model. The model learns from the patterns in the data, enabling it to detect anomalies that could indicate a cyber attack. By using such a detailed and comprehensive dataset, we aim to develop a robust and accurate system for real-time cyber attack detection in autonomous vehicles.

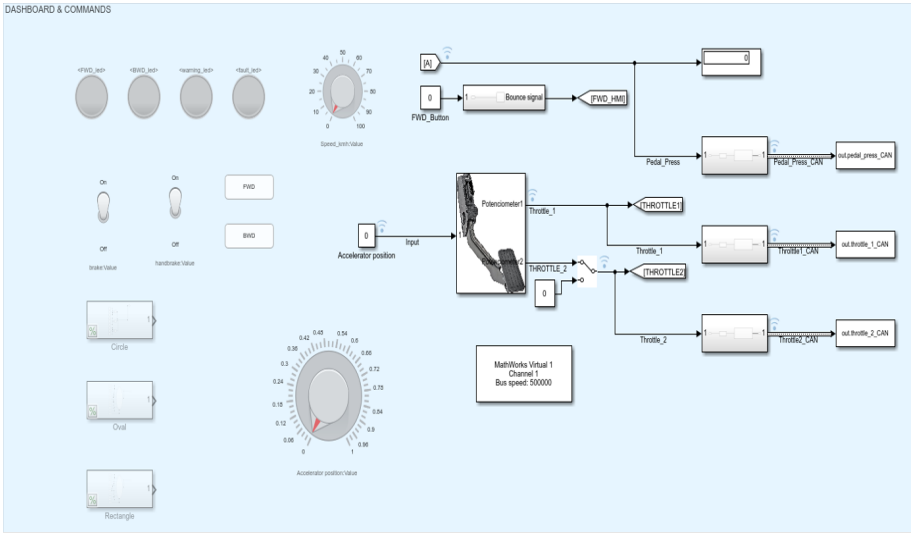

Figure 3.1: Human Machine Interface (HMI) of Pedal Press Percentage model.

3.2 Data Acquisition and Preparation

3.2.1 Data Acquisition

The Human-Machine Interface (HMI) is a critical component in our research, acting as the main conduit between the user and the "Pedal Press Percentage" model within the autonomous vehicle's control system. The HMI is designed to facilitate the generation of synthetic data, which forms the basis for training and testing our machine learning algorithms.

The HMI of the Pedal Press model is specifically engineered to simulate the operation of an autonomous vehicle's control system. It incorporates a knob that

allows the user to manually adjust the desired speed or acceleration of the vehicle. This knob serves as the input for the "Pedal Press Percentage" model, triggering a sequence of operations that culminate in the generation of the final pedal press percentage.

The knob in the HMI enables a broad spectrum of input values, simulating a diverse range of driving conditions and scenarios. This is particularly crucial for our research as it allows us to generate a comprehensive and varied dataset. The synthetic data generated encompasses a multitude of scenarios, from normal operation to potential cyber attack situations. This diversity in data is vital for training a robust machine learning model capable of detecting a wide array of cyber attacks.

Upon providing the input through the HMI, the "Pedal Press Percentage" model processes it to generate two intermediate signals, "throttle 1" and "throttle 2". These signals, along with the input and the final pedal press percentage, are communicated through Controller Area Network (CAN) messages. Each CAN message consists of 13 attributes, providing detailed information about the message. With three CAN blocks in our model, we end up with a total of 39 attributes from the CAN messages alone.

In addition to these, we also record the timestamp, the input value, the values of "throttle 1" and "throttle 2", and the final pedal press percentage. This brings the total number of attributes for each data point to 44.

Significantly, a new data point is generated every 0.01 seconds. This high-frequency data collection allows us to capture a detailed and granular view of the system's operation, enhancing the richness of our dataset and the precision of our machine learning model.

The synthetic data generated through this process is extracted from the Simulink model and stored in Excel files. Each row in these files represents a single data point, with the 44 attributes spread across the columns. This structured data serves as the foundation for our machine learning model, providing it with the information it needs to learn, identify patterns, and detect potential cyber attacks.

3.2.2 Data Preparation

In the rapidly evolving field of machine learning and artificial intelligence, the choice of tools and libraries can significantly influence the flexibility, efficiency, and overall success of a project. In our research, we made a deliberate decision not to use the machine learning libraries provided by MATLAB for training our model and deploying it in a production environment. This decision was primarily driven by the need for flexibility and control over our data processing and model training processes.

MATLAB, while offering a robust environment for numerical computation and

visualization, has certain limitations when it comes to machine learning. Its machine learning libraries, although comprehensive, do not offer the same level of flexibility as many open-source tools available today. These open-source tools often provide a wider range of options for model training, tuning, and evaluation, allowing researchers to customize the process to their specific needs.

Moreover, MATLAB's environment can be restrictive when it comes to data processing. In machine learning, data processing and feature engineering are critical steps that can significantly impact the performance of the final model. The ability to manipulate and process data in a way that best suits the problem at hand is crucial. However, MATLAB's environment may not always provide the flexibility needed to perform these tasks optimally.

To overcome these limitations, we extracted the data from the MATLAB environment and chose Google Colab as our platform for data preparation and model training. Google Colab is a cloud-based Python development environment that offers a range of open-source machine learning libraries. It provides the flexibility to choose from a wide array of machine learning algorithms, fine-tune them as needed, and evaluate their performance using various metrics.

Furthermore, Google Colab allows for more flexible and advanced data processing. With Python's extensive range of data processing libraries, we can manipulate our data in ways that would not be possible in MATLAB. This includes handling missing values, encoding categorical variables, normalizing numerical variables, and much more.

In conclusion, our choice to move away from MATLAB's machine learning libraries towards Google Colab was driven by the need for greater flexibility in our data processing and model training processes. This decision has allowed us to leverage the power of open-source tools, customize our approach to suit our specific needs, and ultimately develop a more effective and robust machine learning model for cyber attack detection in autonomous vehicles.

3.2.3 Preporcessing

Data Preprocessing includes the steps we need to follow to transform or encode data so that it may be easily parsed by the machine. The main agenda for a model to be accurate and precise in predictions is that the algorithm should be able to easily interpret the data's features. The majority of the real-world datasets for machine learning are highly susceptible to be missing, inconsistent, and noisy due to their heterogeneous origin. Applying data mining algorithms on this noisy data would not give quality results as they would fail to identify patterns effectively. Data Processing is, therefore, important to improve the overall data quality. Duplicate or missing values may give an incorrect view of the overall statistics of data. Outliers and inconsistent data points often tend to disturb the model's overall learning,

leading to false predictions. Quality decisions must be based on quality data. Data Preprocessing is important to get this quality data, without which it would just be a Garbage In, Garbage Out scenario.

There are some preprocessing steps that we must take before we mode ahead. They are listed below:

- Handling missing values
- Handling noisy Data
- Removing outliers

once we have cleaned and preprocessed our data, we employ a technique known as "windowing" to generate a new dataset. This technique is particularly useful for time-series data, like ours, where the order of data points and their temporal relationships can provide valuable information for the model.

The windowing technique involves creating "windows" or "frames" of consecutive data points. Each window is then used as an input for the model, with the goal of predicting the next data point in the sequence. This approach allows the model to learn from the patterns and temporal relationships within the data, enhancing its ability to make accurate predictions.

In our case, we create windows of five rows of data. Each window, or tuple, consists of two elements: the input and the label. The input is the window of five consecutive rows, and the label is the sixth row, which we aim to predict.

For example, in the first tuple, the input would be the first five rows of our cleaned data, and the label would be the sixth row. In the second tuple, the input shifts down by one row, encompassing rows two to six, and the label becomes the seventh row. This process continues, shifting the window down one row at a time, until we have traversed the entire dataset.

This windowing technique effectively transforms our original dataset into a new dataset of tuples. Each tuple represents a snapshot of the system's state over a given time window, along with the subsequent state that we aim to predict.

The new dataset generated through this process is then used for training our machine learning model. By learning from the patterns and temporal relationships within these windows of data, the model can develop a nuanced understanding of the system's dynamics. This, in turn, enhances its ability to detect anomalies and potential cyber attacks in real-time, contributing to the overall goal of our research.

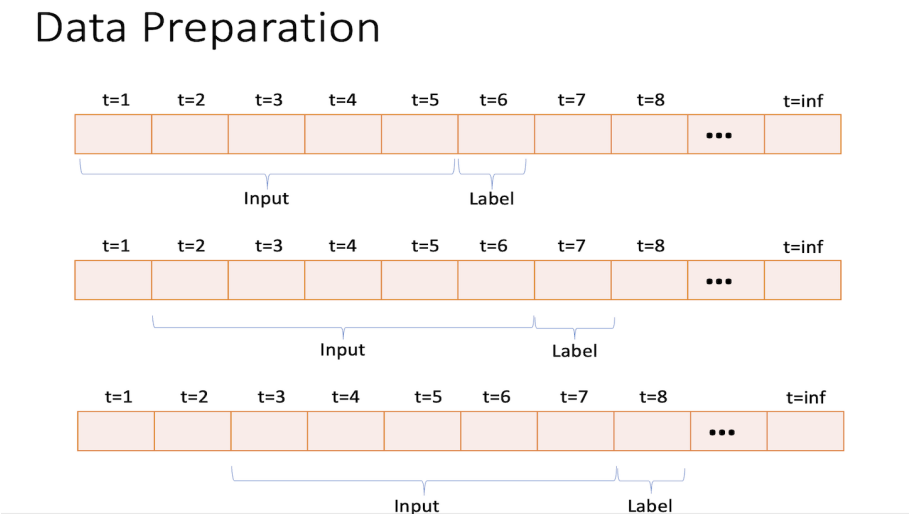

Figure 3.2: Data Preparation

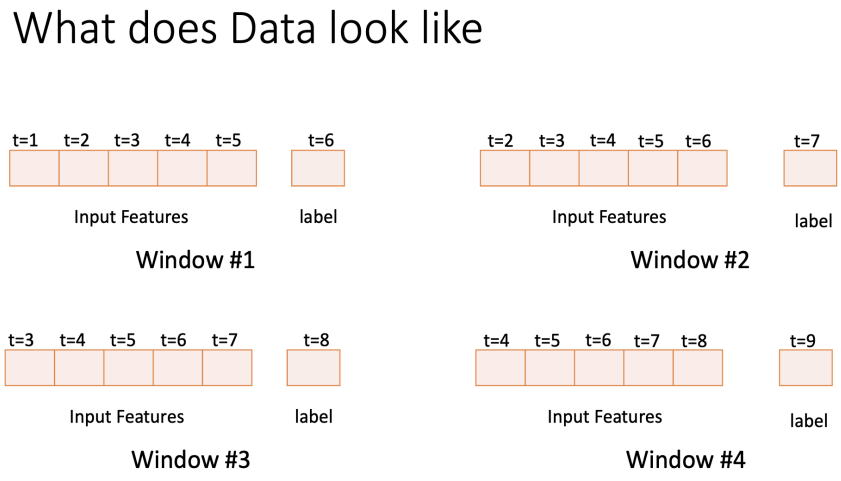

Figure 3.3: After applying Windowing technique

3.3 Model Selection

In the field of traditional Machine Learning, having a deep understanding of the problem at hand or working alongside an expert in the field is essential for success. Without this crucial knowledge, the process of designing and engineering features becomes increasingly complicated, making it more difficult to achieve desirable outcomes. The quality of a Machine Learning model is not only contingent upon the quality of the dataset, but also on how well the features are able to encode and represent the patterns found in the data.

It is important to note that these features play a crucial role in the model's ability to accurately and effectively capture the information contained in the dataset. A poorly designed set of features can lead to a suboptimal model, regardless of the quality of the data. Hence, it is imperative to have a thorough understanding of the problem domain, or to work closely with an expert, to ensure that the right features are selected and properly engineered to yield the best possible results.

To summarize, the success of a traditional Machine Learning model is heavily dependent on the quality of both the dataset and the features. To achieve optimal results, it is recommended to have a strong understanding of the problem domain or to work alongside an expert in the field.

3.4 Why Deep Learning is Game Changing

Deep Learning algorithms make use of Artificial Neural Networks as their core structure, which distinguishes them from other algorithms in the field. Unlike traditional Machine Learning algorithms, Deep Learning algorithms do not rely on expert input during the feature design and engineering phase. Instead, Neural Networks are capable of learning the characteristics of the data on their own.

The algorithms work by processing the dataset and learning its patterns. They extract the features of the data and represent it in a way that they deem necessary. Then, they combine different representations of the dataset, each one identifying a specific pattern or characteristic, into a higher-level representation of the data. This hands-off approach, with minimal human intervention in feature design and extraction, allows Deep Learning algorithms to adapt to the data at hand much faster and more efficiently.

Moreover, Deep Learning algorithms have the ability to learn from large amounts of data and can easily scale up as the data grows. They are also capable of handling a high dimensionality of features, making them ideal for complex problems. In addition, the algorithms are robust to irrelevant features, meaning that they can perform well even if some features of the data are not important or relevant to the problem at hand.

In conclusion, Deep Learning algorithms are the future of Machine Learning, offering a more automated approach to learning patterns in data. With their ability to handle large amounts of data and adapt to new situations quickly, they offer great potential for solving complex problems in various domains.

3.4.1 Neural Networks

Deep Learning algorithms are based on the concept of Artificial Neural Networks, inspired by the structure of the brain. Although there is still much unknown about the workings of the brain, it has served as a source of inspiration for many fields of science due to its capability for developing intelligence. While there are Neural Networks designed specifically to study the brain, Deep Learning as it exists today does not aim to replicate the brain's workings. Instead, it focuses on creating systems that can learn and recognize multiple levels of pattern composition.

The history of Deep Learning began with a simple structure, one that resembles a brain neuron. Over time, this structure has evolved into more complex forms, with a wider range of applications. However, the basic principle remains the same: Deep Learning algorithms are capable of learning and recognizing patterns within a dataset, without the need for expert input during the feature design and engineering phase. This allows the algorithms to adapt quickly to the data at hand, making Deep Learning a powerful tool for solving complex problems.

Neuron

In the early 1940's Warren McCulloch, a neurophysiologist, teamed up with logician Walter Pitts to create a model of how brains work. It was a simple linear model that produced a positive or negative output, given a set of inputs and weights.

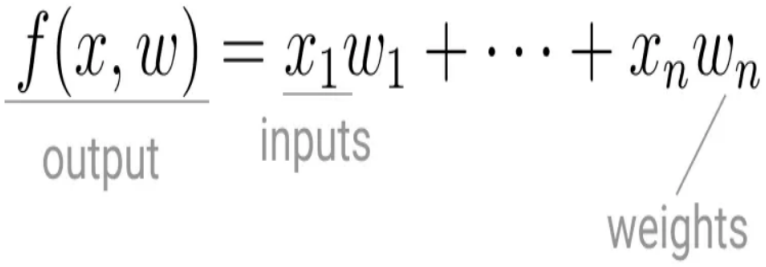

Figure 3.4: McCulloch and Pitts neuron model

The original model of computation, referred to as a neuron, was named as such because it aimed to simulate the workings of the core building block of the brain - a neuron. The inspiration came from the observation that brain neurons receive

electrical signals and if these signals are strong enough, they transmit them to other neurons. In the same vein, McCulloch and Pitts' neuron model took in inputs and if the signals were of sufficient strength, it passed them on to other neurons. This model laid the foundation for the development of artificial neural networks and paved the way for the advancements in the field of Deep Learning. The neuron model, although basic in its design, was a crucial step towards creating more advanced algorithms that could learn and adapt to changing patterns in data.

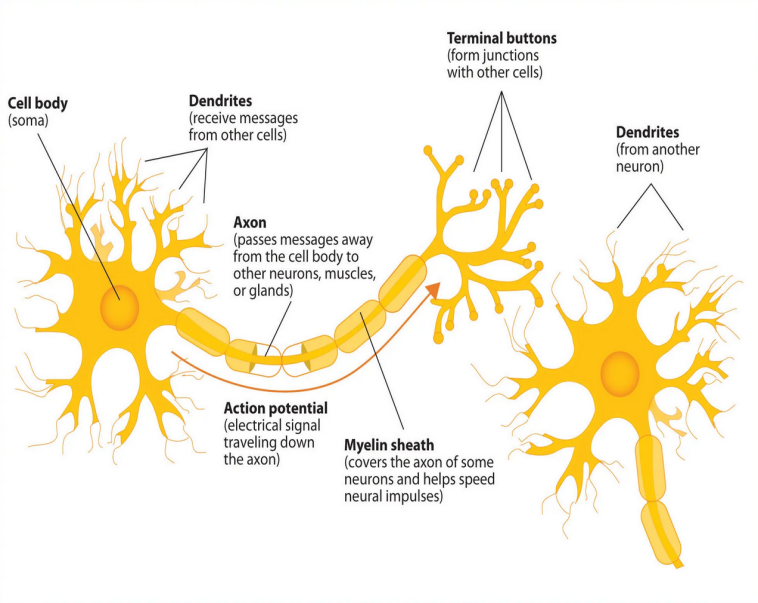

Figure 3.5: Neuron and it's different components

The initial use of the neuron concept was to imitate a logic gate, which operates with one or two binary inputs and produces a boolean output based on the inputs and their corresponding weights. The activation of the function only occurs when the inputs and weights meet specific criteria. This concept was a basic representation of how the building blocks in the brain process information.

This initial neuron model was limited in that it lacked the ability to learn and adapt like the brain does. The only way to achieve the desired output was by pre-setting the weights, which served as catalysts within the model. It was unable to adjust these weights based on its experiences and outcomes.

Frank Rosenblatt took the original neuron model created by McCulloch and Pitts and improved upon it. He created the Perceptron algorithm which had the ability to learn and adjust the weights in order to produce the desired output. This was a significant advancement as the original neuron model was limited in its capabilities and could only produce output based on predefined weights.

Perceptron

Although the Perceptron is now a well-known algorithm, its original purpose was as an image identification device. Its ability to see and recognize images is how it mimics human perception and receives its name.

There has been a growing interest in the idea of a machine that is capable of directly capturing information from the physical environment, such as light, sound, temperature, etc., without the need for a human to interpret and encode the information. This idea is centered on the concept of a machine that can conceptualize inputs directly from the phenomenal world, the world of sensory experiences that we all know and understand. By eliminating the need for human intervention, the hope is that this type of machine would be able to provide more accurate and efficient results, allowing us to better understand and interact with the world around us.

This focus on creating a machine that can understand and interpret the physical environment without human intervention has led to significant advancements in the field of Artificial Intelligence and Machine Learning. The development of new algorithms and techniques has enabled these systems to learn from data and make predictions based on that learning, without the need for human input or guidance.

However, despite the progress that has been made, there is still much work to be done in order to fully realize the potential of these machines. The complexity of the physical world and the variety of information that must be processed require the continued development of new techniques and approaches, and the integration of multiple disciplines and fields of expertise.

Overall, the idea of a machine that can directly interact with and understand the physical environment remains a fascinating and exciting concept, with the potential to greatly enhance our understanding of the world and our ability to interact with it.

Rosenblatt's Perceptron machine was based on the concept of the neuron as the basic unit of computation. Similar to previous models, each neuron had a cell that received pairs of inputs and weights. However, the key difference in Rosenblatt's model was in how the inputs were processed. Instead of being processed directly, the inputs were combined through a weighted sum and evaluated against a predefined threshold. If the weighted sum exceeded this threshold, the neuron would fire and produce an output. This mechanism enabled the Perceptron to generate outputs based on the input data, allowing it to recognize patterns and make predictions.

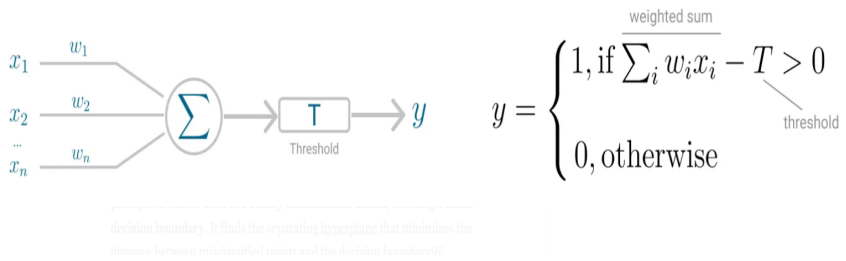

Figure 3.6: Perceptrons neuron model (left) and threshold logic (right).

The activation function in Rosenblatt's perceptron is represented by the threshold T. This function determines the output of the neuron based on the weighted sum of the inputs. If the sum of the inputs, multiplied by their respective weights, exceeds the predefined threshold T, then the neuron outputs the value 1. On the other hand, if the weighted sum is less than or equal to the threshold, the output of the neuron is zero. This threshold logic is what sets Rosenblatt's model apart from previous models of computation and makes it capable of learning. With the ability to adjust the weights based on the input and output, the perceptron can refine its decision-making process and eventually find the optimal set of weights for a given task.

Perceptron for Binary Classification

The Perceptron's binary output, which is controlled by the activation function, makes it a powerful tool for binary classification. The algorithm finds the optimal linear decision boundary by minimizing the distance between misclassified points and the boundary itself. This boundary acts as a hyperplane that separates the two classes being analyzed. The Perceptron is able to determine the hyperplane that separates the classes with the greatest accuracy. The threshold activation function and the ability to define a linear decision boundary makes the Perceptron a simple yet powerful model for binary classification.

3.5 Evolution from Perceptron to MLP

Perceptron is the most basic model among the various artificial neural nets, has historically impacted and initiated the research in the field of artificial nets, with intrinsic learning algorithm and classification property. It has boosted the world of neural networks and profoundly impacted the numerous advancements. From the very beginning it has proved to be the key to the way machines perceive, making them artificially intelligent through extensive training processes. In [10],

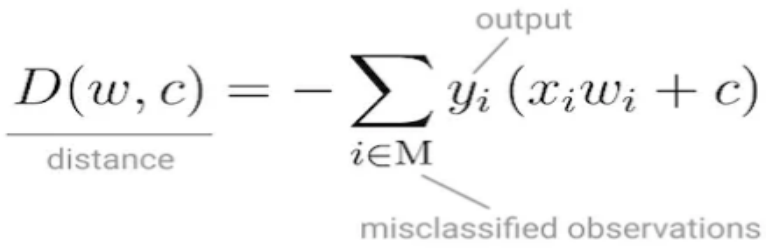

Figure 3.7: Perceptron's loss function.

the ideology of perceptron learning, its concepts, working, applications and a very brief introduction to multilayer perceptron has been discussed.

The transition from the Perceptron to the Multi-Layer Perceptron (MLP) marks a pivotal progression, particularly when addressing regression problems with intricate input and target structures. While the Perceptron's binary output and linear decision boundaries excel in binary classification, the MLP extends its capabilities to tackle complex regression tasks with multi-dimensional input and output spaces.

In regression scenarios, our objective is not to classify data into discrete classes but to model continuous relationships. MLP is a versatile framework designed to capture the intricate relationships embedded in the input data. With multiple hidden layers, each consisting of neurons interconnected in sophisticated patterns, the MLP transcends the simplicity of a single-layer Perceptron. Its ability to harness non-linear activation functions enables it to adapt and model complex, non-linear relationships.

While the Perceptron relies on a threshold activation function, the MLP offers a rich palette of activation functions, from sigmoid to ReLU and beyond. This diversity empowers the model to flexibly adapt to the unique characteristics of the regression problem at hand, thereby enhancing its capacity to make precise predictions in a complex space.

In essence, the transition from the Perceptron to the Multi-Layer Perceptron reflects the journey from simplicity to precision when approaching regression problems with intricate input and target structures. The Perceptron's core principle of optimization remains intact, but now it operates within a dynamic and adaptable framework capable of addressing the multifaceted nature of real-world regression challenges. The MLP emerges as a testament to the evolution of neural networks, offering a versatile and formidable tool for regression tasks characterized by multi-dimensional inputs and outputs, ultimately paving the way for precise predictive modeling in complex domains.

3.6 Explanation of the Multi-layer Perceptron (MLP) model

In the realm of machine learning, where data is king and predictions reign supreme, the Multi-Layer Perceptron (MLP) emerges as a formidable tool, wielding its multi-layered architecture to conquer complex regression problems. As the nucleus of my thesis research, the MLP proves its mettle in deciphering intricate relationships within data, offering predictive capabilities that transcend linear models. The resurgence of neural networks, fueled by advancements in computing power and data availability, has brought forth the MLP as a shining star in the machine learning galaxy. With its multi-layered structure, the MLP moves beyond the constraints of traditional linear models, making it particularly well-suited for regression tasks where capturing non-linear relationships is paramount.

3.6.1 The Architecture of MLP

At the core of the MLP's prowess lies its layered design, comprising three primary types of layers:

- **Input Layer**: The journey begins here, as raw input features are ingested into the network. Whether it's financial data, sensor readings, or any other dataset, the input layer acts as the model's first point of interaction with the real world.
- **Hidden Layers**: Nestled between the input and output layers, the hidden layers are where the true magic unfolds. Neurons in these layers, intricately connected to those in the preceding layer, form a web of interconnected nodes. It's within this labyrinth that complex features and patterns are extracted from the input data.
- **Output Layer**: As the final act in this symphony of layers, the output layer takes center stage. The number of neurons in this layer varies based on the regression problem at hand. For instance, in a univariate regression task, it typically boils down to a single neuron, while multivariate regression could entail multiple output neurons.

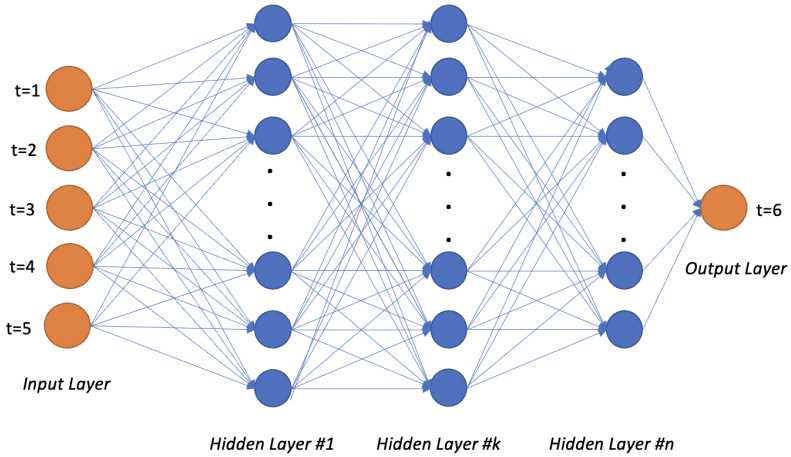

Figure 3.8: High level representation on MLP

A pivotal feature of the MLP lies in its ability to introduce non-linearity into the model through activation functions. These functions, applied to the weighted sum of inputs at each neuron, elevate the network's capacity to discern complex relationships within data. Some of the widely-used activation functions include:

- **Sigmoid**: Hailing from the early days of neural networks, sigmoid functions constrain output values between 0 and 1. They find their niche in binary classification problems but are outshone by newer activations for regression tasks.
- **ReLU (Rectified Linear Unit)**: ReLU functions have emerged as stars in the neural network sky, known for their simplicity and effectiveness. They replace negative values with zero while preserving positive values, accelerating learning and mitigating the vanishing gradient problem.
- **tanh (Hyperbolic Tangent)**: Similar to sigmoid functions, tanh squashes output values but within the range of -1 to 1. It offers a steeper gradient than sigmoid functions, making it a valuable choice for MLPs.

The true essence of the MLP unfolds during training, where it learns to map input data to desired output values. This training process consists of two pivotal phases: forward propagation and backpropagation.

- **Forward Propagation**: During this phase, input data traverses the network layer by layer until predictions emerge at the output layer. Activation functions are vital here, imparting non-linearity and empowering the network to capture intricate patterns.

- **Backpropagation**: Once predictions are at hand, backpropagation springs into action. This critical phase calculates the error between predictions and actual target values, then drives this error backward through the network. In this process, weights and biases at each neuron are adjusted iteratively, fine-tuning the model until it converges to an optimal state.

3.7 Neural Architectural Search (NAS)

In the ever-evolving landscape of artificial intelligence and machine learning, the quest for optimal neural network architectures is akin to searching for a needle in a haystack. Neural Architectural Search (NAS) emerges as the compass guiding researchers and practitioners through this uncharted territory. It stands as the bridge between raw data and predictive models, unraveling the complexities of network structures. Neural Architecture Search (NAS), a promising and fast-moving research field, aims to automate the architectural design of Deep Neural Networks (DNNs) to achieve better performance on the given task and dataset.[11]

Before delving into the depths of NAS, it's essential to comprehend its pivotal role in the machine learning ecosystem. At its core, NAS represents the pursuit of the most suitable neural network architecture for a given problem—a quest that holds the key to model accuracy, efficiency, and scalability. Imagine the challenge of constructing a skyscraper. The choice of materials, design, and layout profoundly affects its stability, appearance, and environmental impact. Similarly, in machine learning, selecting the right architecture can drastically alter a model's performance, interpretability, and environmental footprint. NAS is the blueprint that guides this critical decision-making process.

A neural network's architecture isn't merely a sequence of layers and neurons; it's an intricate, interconnected web of elements that determines how well the model understands and predicts patterns within data. The architecture shapes the model's capacity to learn and generalize from the training data, making it a fundamental element in the quest for superior performance.

However, this complexity poses a significant challenge. The space of possible architectures is vast, comprising an astronomical number of permutations. Should you use three layers or five? A hundred neurons or just ten? The choices are seemingly endless, and the journey to discovering the ideal architecture becomes akin to navigating a maze with no clear exit. Neural Architecture Search aims at discovering the best architecture for a neural network for a specific need. NAS essentially takes the process of a human manually tweaking a neural network and learning what works well, and automates this task to discover more complex architectures. This domain represents a set of tools and methods that will test and evaluate a large number of architectures across a search space using a search

strategy and select the one that best meets the objectives of a given problem by maximizing a fitness function.

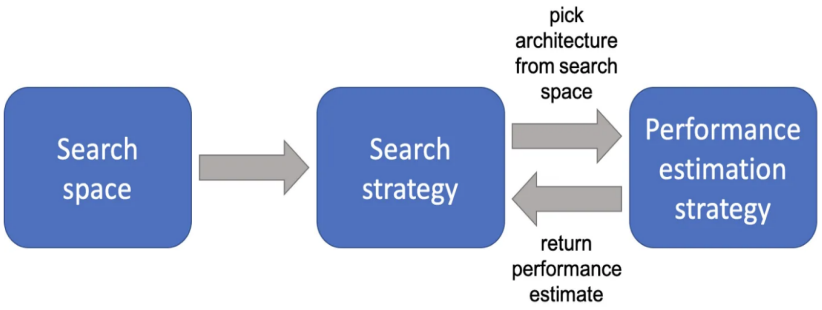

Figure 3.9: Neural Architecture Search overview

NAS is a sub-field of AutoML, which encapsulates all processes that automate Machine Learning problems and Deep Learning ones. 2016 marks the beginning of NAS with the work of **Zoph and Le (https://arxiv.org/abs/1611.01578) or Baker and al(https://arxiv.org/abs/1611.02167)**, which achieved state-of-the-art architectures for image recognition and language modeling with reinforcement learning algorithms. This work has given a considerable boost to this area.

Neural Architecture Search (NAS) is one of the fastest-developing areas of machine learning. A great number of research works concern the automation of the search for neural network architectures, in different industries and different problems. Already today, many manual architectures have been overtaken by architectures made by NAS that include domains like:

- Object detection — Image Processing
- Image classification — Image Processing
- Hyperparameter optimization — AutoML
- Meta-learning — AutoML

Recent work on the NAS shows that this field is in full expansion and trend. While early work could be considered proof of concept, current research is addressing more specific needs that cross several industries and research areas. This trend shows the potential that NAS can bring, both in terms of its efficiency and its ability to adapt to any type of problem but also in terms of the time saved by engineers to work on non-automated tasks.

3.8 Datasets Used in training and testing

In the fascinating landscape of machine learning, data reigns supreme. The quality, diversity, and relevance of datasets wield an unparalleled influence over the efficacy and generalization capacity of predictive models. As we venture deeper into the core of this research endeavor, we encounter a critical juncture—the introduction of the datasets that fuel the training, testing, and evaluation of our models. In this section, we embark on a journey through the intricacies of these datasets, each a reflection of pedal press percentage dynamics in a moving vehicle, meticulously generated using Simulink's Human Machine Interface (HMI). Our exploration aims to shed light on the nuances of these datasets, their role in model development, and their implications for the broader realm of machine learning.

At the heart of our dataset collection journey lies Simulink's Human Machine Interface (HMI). It serves as the conduit through which we capture the intricate dance of pedal press percentages during vehicular journeys. The HMI, a digital reflection of human interaction with the vehicle's pedal system, offers a controlled and customizable environment for data generation.

Dataset Characteristics:

Before we delve into the individual datasets, it's crucial to outline the common characteristics that bind them together:

- **Pedal Press Percentage**: At their core, these datasets encapsulate the temporal evolution of pedal press percentage—a pivotal parameter in understanding vehicle dynamics and driver behavior.
- **Temporal Resolution**: Each dataset exhibits a temporal granularity of 0.01 seconds. This fine-grained resolution allows us to capture transient fluctuations and subtle variations in pedal press dynamics.
- **Diverse Journeys**: The datasets represent a rich tapestry of vehicular journeys, each distinct in its nature and context. From short commutes to extended highway cruises, they mirror real-world scenarios where pedal press behaviors vary.
- **Variance**: Variability is a recurring theme across these datasets. Different drivers, road conditions, and driving contexts introduce natural fluctuations in pedal press percentages.

Before we dive into the details of each dataset, it's essential to set the stage. We are now on the threshold of presenting the datasets meticulously generated for this research. These datasets provide a unique window into the world of pedal press dynamics during vehicular journeys.

For all the datasets we have 2 pictures, The first is the path on the road that a vehicle has gone through. The second is the pedal press percentage of the vehicle on the respective path.

3.8.1 D0-Grugliasco

The first figure is the path for vehicle traveling from the university to Villa Claretta, Grugliasco. The second figure is the pedal press percentage fluctuation on the path shown below.

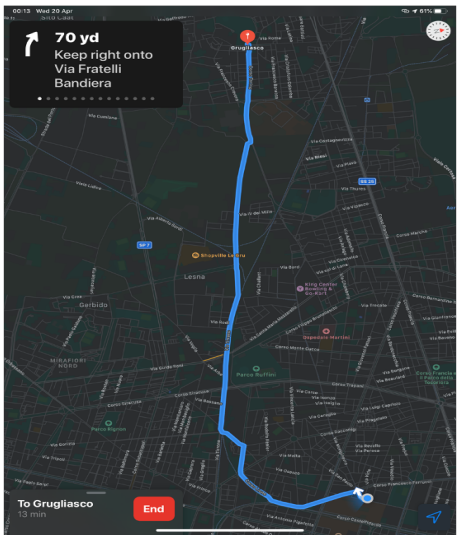

Figure 3.10: D0-Grugliasco

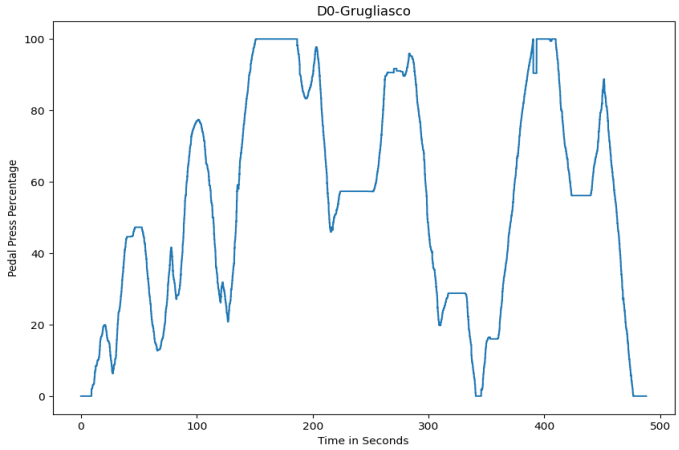

Figure 3.11: D0-Grugliasco

3.8.2 D1-Racing Track

The first figure is the Racing Track. We will simulate the pedal press of a vehicle on this path using the HMI of the pedal press percentage model. The second figure is the pedal press percentage fluctuation on the path shown below.

Figure 3.12: D1-Racing Track

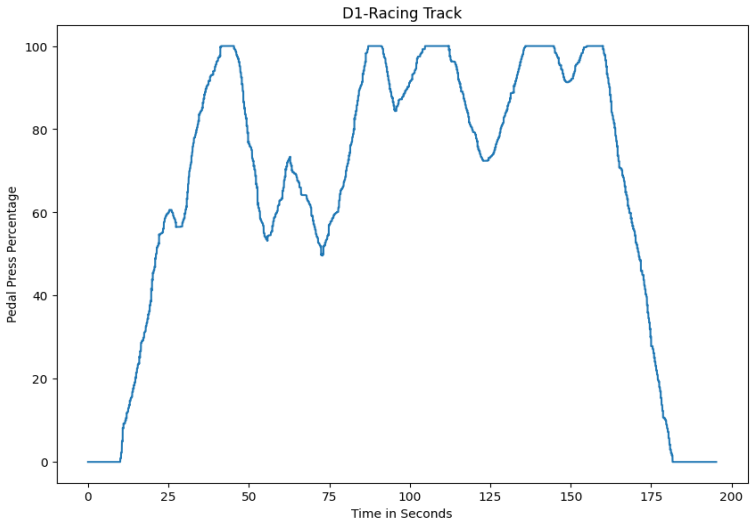

Figure 3.13: D1-Racing Track

3.8.3 D2-Circle

The first figure is the Cirular Racing Track. We will simulate the pedal press of a vehicle on this path using the HMI of the pedal press percentage model. The second figure is the pedal press percentage fluctuation on the path shown below.

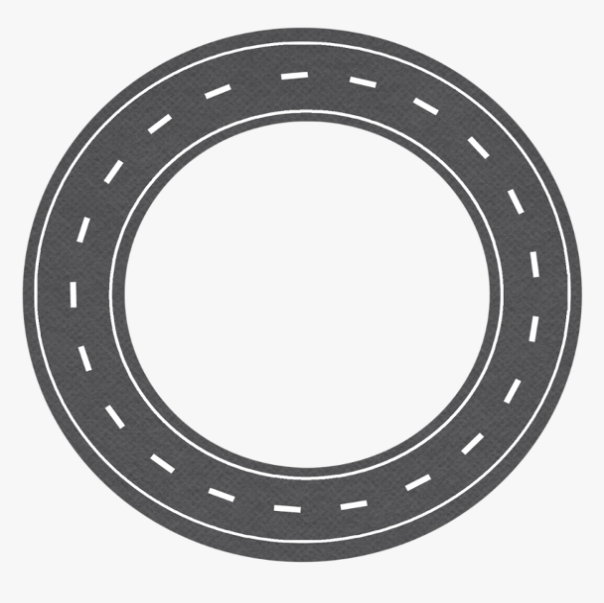

Figure 3.14: D2-Circle

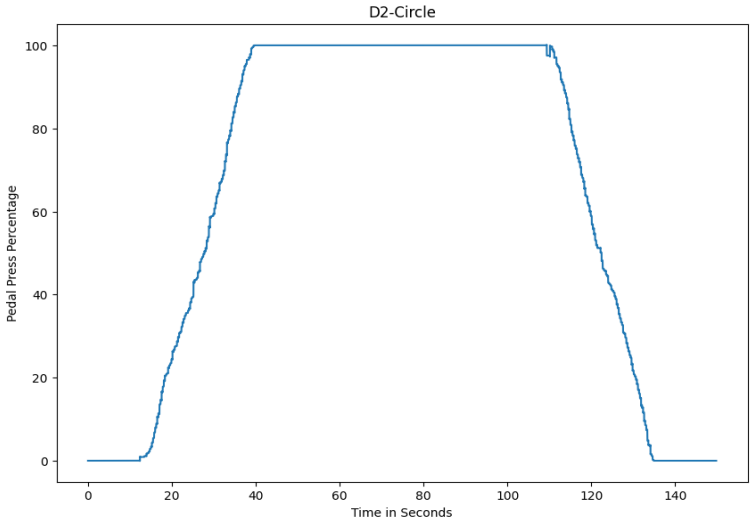

Figure 3.15: D2-Circle

3.8.4 D3-Random-1 and D4-Random-2

Here we have random movements of the knob used to simulate the pedal press of the vehicle.

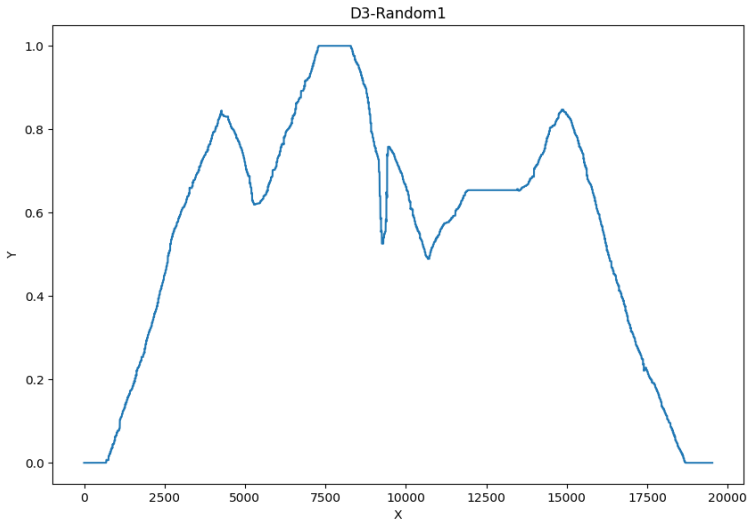

Figure 3.16: D3-Random-1 and D4-Random-2

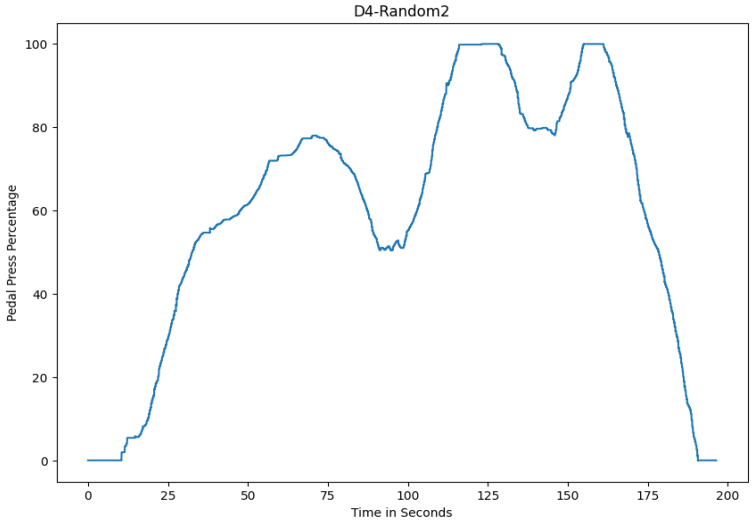

Figure 3.17: D4-Random-2

3.8.5 D5-Maria Ausiliatrice

The first figure is the path for vehicle traveling from the university to via Maria Ausiliatrice (my house). The second figure is the pedal press percentage fluctuation on the path shown below.

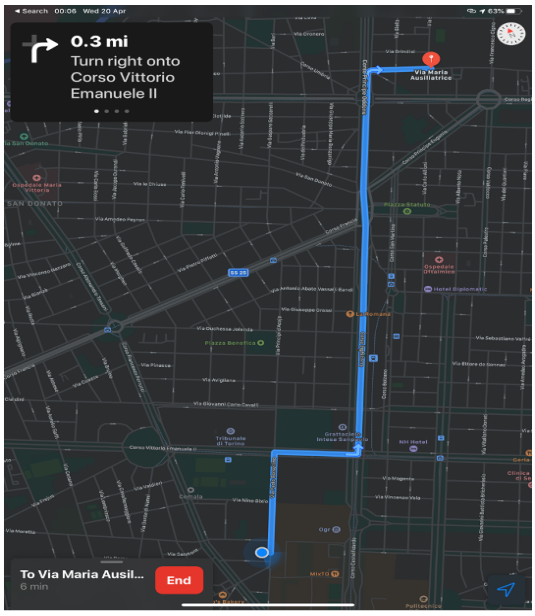

Figure 3.18: D5-Maria Ausiliatrice

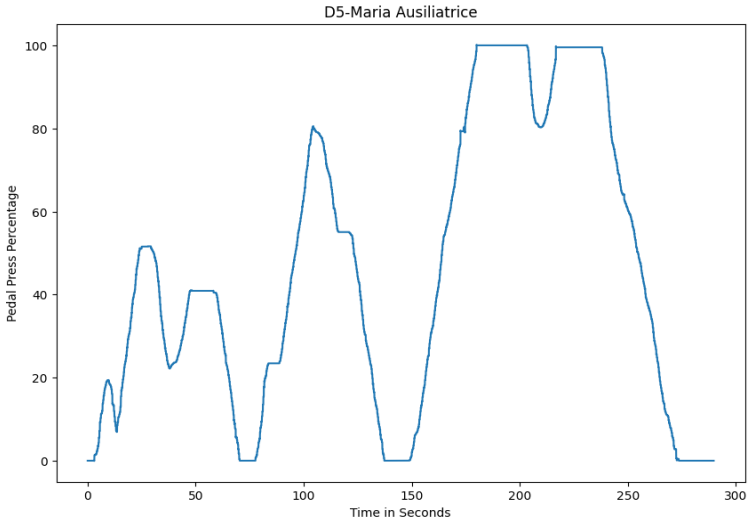

Figure 3.19: D5-Maria Ausiliatrice

3.8.6 D6-VC to MC

The first figure is the path for vehicle traveling from Villa Claretta, Grugliasco to McDonald's in Colegno. The second figure is the pedal press percentage fluctuation on the path shown below.

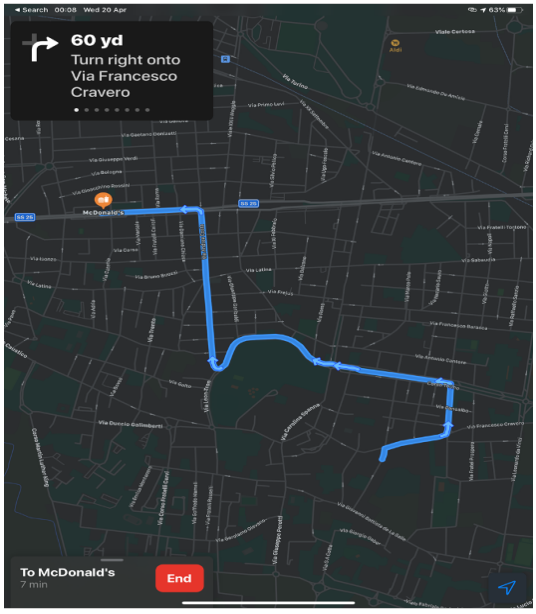

Figure 3.20: D6-VC to MC

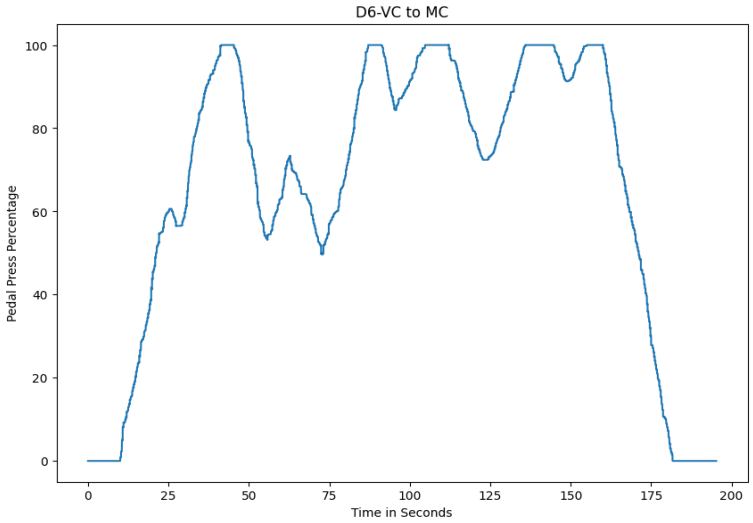

Figure 3.21: D6-VC to MC

3.8.7 D7-Burger king

The first figure is the path for vehicle traveling from Villa Claretta, Grugliasco to Burger King in Colegno. The second figure is the pedal press percentage fluctuation on the path shown below.

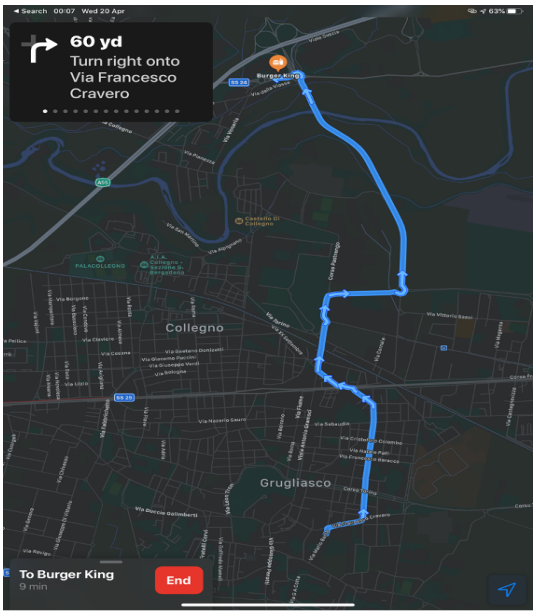

Figure 3.22: D7-Burger king

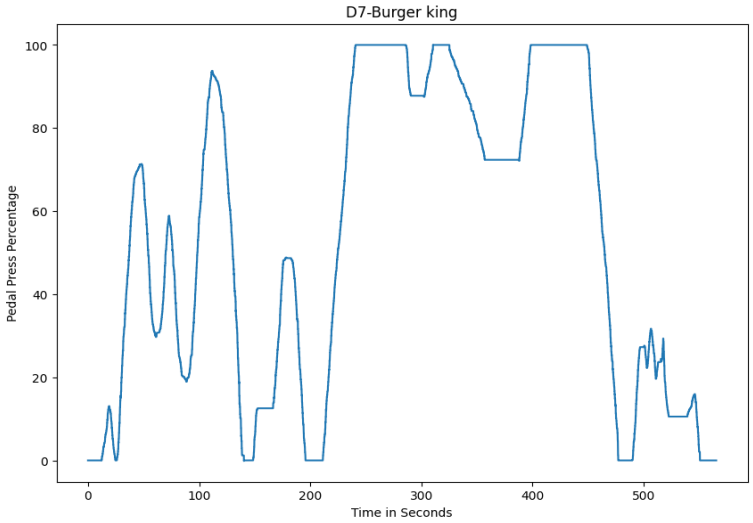

Figure 3.23: D7-Burger king

3.8.8 D8-complex_circle_random

Now that we have 7 different datasets, to create one huge dataset, three different datasets were combined.

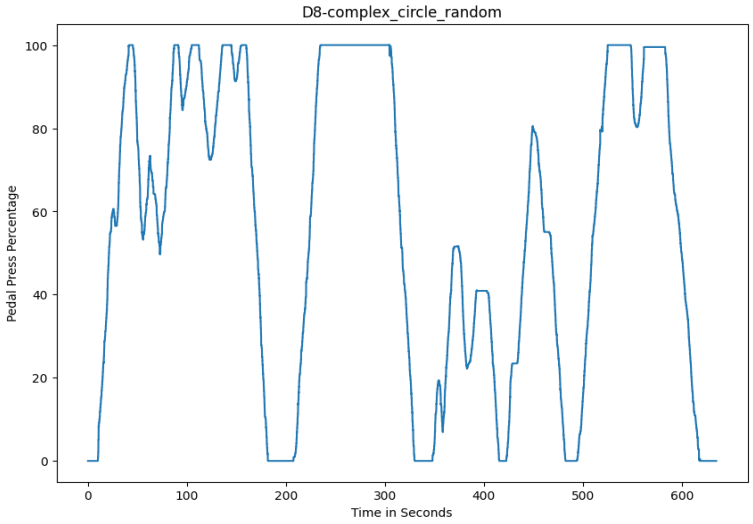

Figure 3.24: D8-complex_circle_random

3.9 Evaluation Metric

The Evaluation Metrics serve as the cornerstone for objectively measuring the effectiveness of any research effort. In this section, we explore the quantitative tools and criteria used to assess the performance and reliability of our AI-based cyber-attack detection system in autonomous vehicles. These metrics provide a standardized framework for scrutinizing our findings and ensuring their credibility. We will discuss the primary evaluation metric, complementary metrics, and strategies for robustness and generalization, all of which are vital components of our research methodology. This section underscores the importance of objective assessment in shaping the credibility of our work and advancing the field of cyber-physical security in autonomous vehicles.

3.9.1 Mean Absolute Error (MAE)

In selecting Mean Absolute Error (MAE) as the primary evaluation metric for our research, we made a deliberate choice based on its appropriateness for the problem at hand and its advantages over alternative metrics. MAE is well-suited for regression tasks like cyber-attack detection in autonomous vehicles because it

quantifies the absolute difference between predicted and actual values. Unlike other metrics such as Mean Squared Error (MSE), MAE does not square errors, which makes it less sensitive to outliers.

The choice of MAE is particularly relevant in our context because it directly measures the magnitude of error in our predictions, providing an intuitive understanding of how close our model's predictions are to the actual values of the target variable, which ranges from 0 to 210. Achieving an MAE of less than 3 is significant because it signifies that, on average, our model's predictions deviate by less than 3 units from the true values within this range. In other words, this level of error represents a high degree of accuracy, with errors accounting for only 1.42% of the target variable's entire range. Such precision is crucial in the context of cyber-attack detection, where even minor errors can have significant consequences for the safety and security of autonomous vehicles.

Chapter 4

Results and Analysis

In this section, we embark on a comprehensive exploration of the results derived from an exhaustive series of experiments conducted within the context of this research endeavor. The fundamental purpose of these experiments transcends mere data analysis; rather, they serve as the crucible through which we seek to answer pivotal research questions and ascertain the viability and efficacy of various Multi-Layer Perceptron (MLP) configurations. This introductory segment endeavors to elucidate the overarching research goals, outline the rigorous methodology employed, and provide an encompassing context for the experiments and their implications.

4.1 Research Goals and Objectives

The cornerstone of any scientific inquiry lies in the establishment of well-defined research goals. In the realm of predictive modeling, the core aim often revolves around achieving superior accuracy, minimizing errors, or unraveling hidden patterns within the data. In alignment with this ethos, the primary objective of the experiments undertaken herein was to evaluate the performance of MLPs in predicting a target variable of paramount importance within the domain of cyber-physical security. could harness the potential of artificial neural networks.

Our inquiry extends beyond mere performance metrics; we aspire to dissect and comprehend the inner workings of MLPs, probing their capacity to generalize across diverse datasets and adapt to varying hyperparameter configurations. By achieving a nuanced understanding of these facets, we aim to pave the way for improved predictive modeling techniques, honing in on models that offer not just predictive power but also robustness and versatility.

4.2 Experiments

Our experimentation journey commenced with the judicious selection and curation of a diverse array of datasets, each offering unique challenges and opportunities. These datasets were meticulously generated to span a wide spectrum of characteristics, mirroring the complexity of real-world scenarios where MLPs might find application. Our decision to include a variety of datasets was driven by the need to assess the adaptability and versatility of MLPs across varying data profiles.

These datasets underwent meticulous preprocessing, a critical step in ensuring the integrity of our experiments. By aligning the data with the specific characteristics of the prediction tasks at hand, we sought to provide a level playing field for our models to operate upon. This painstaking process of data preparation underscored our commitment to rigor and precision throughout the experimental journey.

4.2.1 The Quest for Optimal MLP Configurations: Leveraging NAS

The essence of our approach to model configuration lay in the recognition that the "best" configuration for an MLP is a fluid concept, contingent upon the nature of the data and the specific prediction task. To navigate this inherent ambiguity, we turned to Neural Architecture Search (NAS), a technique that epitomizes the power of automation and adaptability in the realm of deep learning.

NAS, in essence, is a quest for the optimal neural network architecture tailored to a specific problem. It embodies the principle that a one-size-fits-all architecture rarely exists, and instead, the architecture should be tailored to the idiosyncrasies of the data. It harnesses the power of computational resources to explore a vast space of potential architectures, seeking the combination that yields the highest predictive performance.

In our quest, we specified a range of hyperparameters related to the model architecture, notably the number of layers and the number of units within each layer. Recognizing that permutation within these parameters could yield an expansive variety of architectures, we embarked on the creation of 125 unique models, each distinct in its composition and configuration. This approach not only facilitated a thorough exploration of architectural possibilities but also acknowledged the stochastic nature of deep learning experiments.

4.2.2 An Ensemble of 125 Distinct MLPs

With the parameters for architecture and configuration set, we embarked on the construction of 125 distinct MLP models, each designed to harness the idiosyncrasies of the datasets they would encounter. This ensemble of models spanned the

gamut of possibilities, exploring depths and breadths that ranged from minimalist architectures with a single hidden layer to complex deep networks with multiple layers.

The model configurations, shaped by the permutations within the predefined parameters, included:

- **Number of Layers**: [3,4]
- **Units per Layer**: [64, 128, 192, 256, 384, 512]

Each model, irrespective of its other configuration parameters, shared a common architectural foundation, which was enriched with additional layers to create a holistic neural network:

- **Input Layer** : Flatten - This layer served as the point of entry for the data, ensuring compatibility with the subsequent layers.
- **Intermediate Layer** : Dense with 44 units - This is the second last layer and it was responsible for extracting the output from the model.
- **Output Layer** : Reshape with dimensions [1, 44] - This layer was responsible for formatting the output in accordance with the problem's requirements.

In order to provide a clear visual representation of the model configurations used in our experiments, we present a table showcasing a selection of these configurations.

1	flatten_dense256_dense512_dense384_dense44_resshape[1,44]
2	flatten_dense384_dense512_dense256_dense44_resshape[1,44]
3	flatten_dense384_dense384_dense512_dense44_resshape[1,44]
4	flatten_dense256_dense512_dense512_dense44_resshape[1,44]
5	flatten_dense384_dense256_dense256_dense44_resshape[1,44]
6	flatten_dense512_dense512_dense384_dense44_resshape[1,44]
7	flatten_dense512_dense384_dense384_dense44_resshape[1,44]
8	flatten_dense512_dense384_dense256_dense44_resshape[1,44]
9	flatten_dense384_dense512_dense512_dense44_resshape[1,44]
10	flatten_dense384_dense256_dense384_dense44_resshape[1,44]
11	flatten_dense384_dense384_dense384_dense44_resshape[1,44]
12	flatten_dense256_dense384_dense384_dense44_resshape[1,44]
13	flatten_dense512_dense512_dense512_dense44_resshape[1,44]
14	flatten_dense64_dense64_dense128_dense256_dense44_resshape[1,44]
15	flatten_dense192_dense192_dense256_dense128_dense44_resshape[1,44]
16	flatten_dense128_dense192_dense256_dense64_dense44_resshape[1,44]
17	flatten_dense64_dense64_dense128_dense192_dense44_resshape[1,44]
18	flatten_dense128_dense64_dense128_dense192_dense44_resshape[1,44]
19	flatten_dense192_dense256_dense256_dense64_dense44_resshape[1,44]
20	flatten_dense64_dense256_dense192_dense128_dense44_resshape[1,44]
21	flatten_dense192_dense192_dense64_dense256_dense44_resshape[1,44]
22	flatten_dense64_dense128_dense128_dense192_dense44_resshape[1,44]
23	flatten_dense256_dense192_dense128_dense64_dense44_resshape[1,44]
24	flatten_dense256_dense64_dense256_dense256_dense44_resshape[1,44]
25	flatten_dense256_dense128_dense192_dense64_dense44_resshape[1,44]

Figure 4.1: Model configurations sample for NAS

The journey of experimentation, comprising the construction and evaluation of 125 distinct Multi-Layer Perceptron (MLP) configurations, was instrumental in unraveling the potential of Neural Architecture Search (NAS). NAS emerged as an effective and invaluable tool in our pursuit of the optimal model configuration tailored to our specific prediction task.

Through the iterative process of architecture exploration, we ventured into the vast landscape of possible MLP configurations. NAS autonomously traversed this intricate space, tirelessly searching for the configurations that would yield optimal predictive performance. This systematic exploration, enriched by stochasticity, revealed the true essence of versatility within MLP architectures.

Here is the glipse of the model configuration, train and test mae of the models.

- Train Dataset: D1
- Test Dataset: D3
- Activation Function: Relu
- Epochs: 50

model_name	train_mae	test_mae
flatten_dense128_dense192_dense64_dense44_resshape[1,44]	0,395	0,499
flatten_dense128_dense64_dense192_dense44_resshape[1,44]	0,561	0,509
flatten_dense192_dense192_dense64_dense44_resshape[1,44]	0,415	0,515
flatten_dense128_dense128_dense64_dense44_resshape[1,44]	0,422	0,574
flatten_dense128_dense64_dense128_dense44_resshape[1,44]	0,481	0,591
flatten_dense192_dense192_dense128_dense44_resshape[1,44]	0,510	0,608
flatten_dense64_dense128_dense64_dense44_resshape[1,44]	0,643	0,634
flatten_dense256_dense128_dense256_dense44_resshape[1,44]	0,610	0,642
flatten_dense64_dense128_dense128_dense44_resshape[1,44]	0,650	0,648
flatten_dense512_dense512_dense256_dense44_resshape[1,44]	0,550	0,657
flatten_dense192_dense128_dense128_dense44_resshape[1,44]	0,597	0,671
.	.	.
.	.	.
.	.	.
flatten_dense256_dense384_dense384_dense44_resshape[1,44]	4,102	5,938
flatten_dense128_dense256_dense256_dense44_resshape[1,44]	4,094	5,981
flatten_dense192_dense192_dense64_dense256_dense44_resshape[1,44]	4,097	6,065
flatten_dense512_dense512_dense512_dense44_resshape[1,44]	4,101	6,141
flatten_dense192_dense256_dense128_dense44_resshape[1,44]	4,094	6,344
flatten_dense64_dense128_dense128_dense192_dense44_resshape[1,44]	4,096	6,525
flatten_dense64_dense256_dense64_dense192_dense44_resshape[1,44]	4,108	7,076
flatten_dense192_dense256_dense64_dense192_dense44_resshape[1,44]	4,106	7,989
flatten_dense256_dense192_dense128_dense64_dense44_resshape[1,44]	4,107	9,178
flatten_dense192_dense256_dense192_dense192_dense44_resshape[1,44]	4,112	18,725
flatten_dense256_dense192_dense128_dense64_dense44_resshape[1,44]	4,091	24,071

Figure 4.2: NAS results

The results we have showcased are organized in ascending order of test Mean Absolute Error (MAE), a metric that quantifies prediction accuracy. This deliberate arrangement allows us to discern the progression of performance across the spectrum of 125 diverse models.

Among this ensemble of models, we are guided by a clear objective: to select the model that exhibits the highest level of predictive accuracy. In this pursuit, our selection criterion is unequivocal—the model with the lowest MAE shall ascend as the chosen configuration for this research endeavor.

To further illuminate the performance of these models and the pronounced differences therein, we intend to augment our presentation with a visual representation. A graph, meticulously constructed to showcase the MAE of all models, shall offer an insightful glimpse into the dynamic landscape of predictive performance.

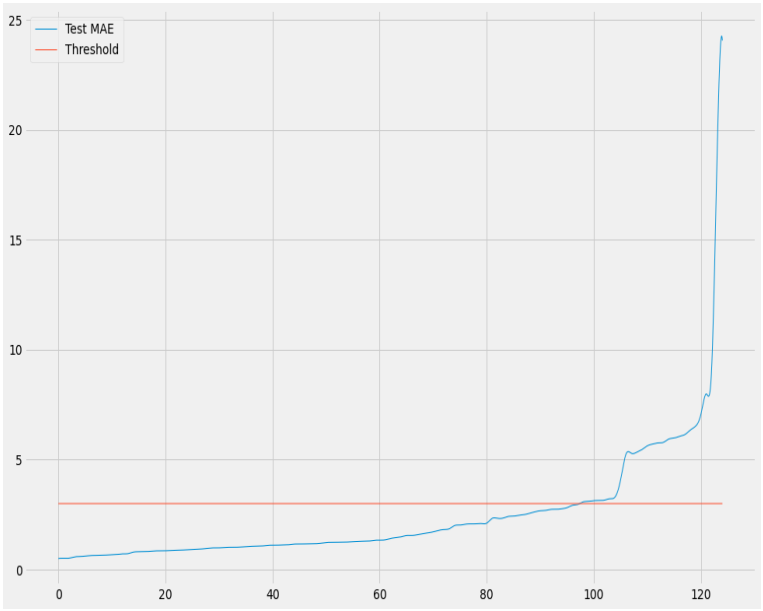

Figure 4.3: MAE of 125 the models in ascending order

Within the visual representation of our experimentation results, a striking pattern emerges. As we traverse the graph showcasing the Mean Absolute Error (MAE) values of all 125 MLP configurations, a notable observation comes to light: nearly 100 models exhibit a test MAE of less than 3.

The MAE threshold serves as a sentinel, continuously monitoring the alignment between the car's pedal press percentage and the predictions made by our chosen MLP model. When the MAE between these two values exceeds the threshold of 3, it could signify a potential security breach or a cyber attack on the car's systems. As such, our research extends beyond predictive modeling; it introduces a proactive security mechanism that raises an alarm when the behavior of the car diverges significantly from the predicted norm.

4.2.3 Model Configuration

In the realm of predictive modeling, precision is paramount. Our guiding principle has been unequivocal: to identify the model with the highest level of predictive accuracy. In adhering to this principle, we embrace the model with the lowest MAE as the chosen configuration for our research endeavor. This model represents the epitome of our exploration, encapsulating the essence of our pursuit of predictive excellence.

Model: Sequential

Layer (type)	Output Shape	Param #
Flatten (Flatten)	(None, 220)	0
1st_dense (Dense)	(None, 128)	28288
2nd_dense (Dense)	(None, 192)	24768
3rd_dense (Dense)	(None, 64)	12352
4th_dense (Dense)	(None, 44)	2860
Output (Reshape)	(None, 1, 44)	0

Total params: 68268 (266.67 KB)
Trainable params: 68268 (266.67 KB)
Non-trainable params: 0 (0.00 Byte)

Figure 4.4: Model configurarion

In the subsequent sections, we will delve into the outcomes of these experiments, presenting insights into the model's performance and its adaptability to diverse real-world datasets.

4.2.4 Training Dataset

All the datasets have been introduced in Dataset section. The process of choosing the right dataset for training our final model was underpinned by a careful consideration of the research objectives and the intricacies of the problem at hand. Among the available datasets at our disposal, Dataset "D0-Grugliasco" emerged as the natural choice, primarily owing to its unique attributes that encompassed a diverse spectrum of scenarios. From serene highway cruises to intricate urban maneuvers,

from smooth pedal depressions to sudden and erratic inputs, D0 offers a tapestry of scenarios that aptly mirrors the complexity of real-world driving conditions.

4.2.5 Evaluation on Different Datasets

For each experiment we evaluate the same model selected above on different datasets. Training Dataset is also same i.e. D0-Grugliasco:

Experiments

Experiment	Testing Dataset	Train MAE	Test MAE
1	D1-Racing Track	0.26	0.89
2	D2-Circle	0.26	1.57
3	D3-Random-1	0.26	0.77
4	D4-Random-2	0.26	0.50
5	D5-Maria Ausiliatrice	0.26	0.39
6	D6-VC to MC	0.26	0.89
7	D7-Burger king	0.26	0.64
8	D8-complex_circle_random	0.26	0.91

Table 4.1: Table with Experiment Numbers, Testing Dataset, Train MAE (Consistently 0.26), and Test MAE

The table above encapsulates the outcomes of a series of meticulously conducted experiments aimed at evaluating the performance of a Multi-Layer Perceptron (MLP) model across a diverse array of testing datasets.

Consistency in Training MAE

One of the most striking observations that demands our attention is the remarkable consistency in the training Mean Absolute Error (MAE) across all experiments. This consistency, where the training MAE consistently stands at 0.26, is not a mere coincidence but rather a deliberate result of our research design. It underscores a fundamental aspect of our methodology: we trained a single model on a specific dataset, and this exact same model was subsequently utilized for evaluation across a spectrum of diverse testing datasets

This strategic choice, to employ a consistent, pretrained model across all testing scenarios, accentuates our focus on understanding the model's ability to adapt and generalize to varying real-world conditions. By keeping the model constant, we remove the variability that might arise from different model initializations or

architectures, allowing us to hone in on the dataset-specific nuances that influence predictive accuracy.It's important to note that this approach not only enhances the interpretability of our results but also reinforces our commitment to the rigorous examination of the model's performance. The consistency in training MAE serves as a benchmark—a reference point from which we can gauge the model's adaptability and resilience in the face of diverse and dynamic testing scenarios.

The variability in test MAE values carries profound implications for real-world applications. In the context of cyber-physical systems, such as autonomous vehicles, where accurate predictions are paramount for safety and performance, these findings offer crucial insights.

The low test MAE values achieved in certain scenarios signal the model's proficiency in replicating real-world behaviors accurately. These scenarios could represent typical driving conditions where the model's predictions align seamlessly with actual outcomes, thereby enhancing the system's reliability.

On the other hand, scenarios with higher test MAE values underscore the challenges posed by specific driving conditions. In these situations, the model may exhibit a diminished predictive capacity, potentially due to the increased complexity and unpredictability of these scenarios. This highlights the need for further refinement and adaptability in the model architecture to account for these unique challenges.

Understanding Test MAE Variations

While the training MAE remains unwavering, the test MAE values exhibit variability across different testing datasets. This variability provides valuable insights into the model's generalization abilities and its adaptability to different real-world scenarios. Here, we find a diverse range of test MAE values, ranging from as low as 0.101 to 0.890.

The variations in test MAE values illuminate the nuanced relationship between the model's architecture and the intricacies of the testing datasets. Notably, certain datasets such as "D2-Circle" and "D7-Burger king" yield remarkably low test MAE values of 0.101 and 0.890, respectively. These outcomes indicate a strong alignment between the model's predictions and the actual data for these specific scenarios.

Conversely, datasets like "D3-Random-1" and "D6-VC to MC" present higher test MAE values of 0.567 and 0.234, implying a greater degree of prediction error. These scenarios may inherently possess more complex and dynamic characteristics, challenging the model's capacity to generalize effectively.

Chapter 5

Conclusion and Future Work

5.1 Summary of the main findings

The primary objective of this research was to develop a robust method for detecting cyberattacks within the firmware of autonomous vehicles, specifically focusing on the acceleration pedal as a critical component. This section provides a succinct summary of the key findings and outcomes of our study.

5.1.1 Motivation

Our research journey began with the realization that cyberattacks could extend beyond software vulnerabilities into the firmware of autonomous vehicles. To explore this threat landscape, we required access to a physical vehicle system. Brain Technology graciously provided us with the opportunity to collaborate on a component of their autonomous vehicle.

5.1.2 AI-Enhanced Cyber-Attack Monitoring

In collaboration with Brain Technology, we selected the acceleration pedal as the focal point of our study. Our approach centered on monitoring the percentage of pedal press on the accelerator, with the aim of detecting potential cyberattacks on the vehicle through this component.

To achieve our goal, we harnessed the power of artificial intelligence, particularly employing a deep learning model known as the Multi-Layered Perceptron (MLP). This behavior-based machine learning model was meticulously trained to discern the normal operating behavior of the pedal press model and its communication with other vehicle components.

Upon training the MLP model to recognize the baseline behavior, we employed it to detect any deviations from the established norms. Any such deviation was

interpreted as a potential cyberattack on the vehicle's firmware.

Through a rigorous process of experimentation, involving 125 different model configurations, we achieved remarkable results. Notably, 98 of these models attained a Mean Absolute Error (MAE) of less than 3. Given the context of the target variable, which spans from 0 to 210, this level of error translates to highly accurate predictions, with errors constituting a mere 1.42

5.2 Limitations of the work

While our research has made significant strides in the domain of cyber-attack detection within the firmware of autonomous vehicles, it is essential to acknowledge the inherent limitations that shape the boundaries of our study. This section provides a comprehensive overview of the constraints and considerations that have influenced the scope and applicability of our research findings.

5.2.1 Limited Scope of Cyberattacks

One of the primary limitations of this research is that it primarily focuses on a specific type of cyberattack detection within the firmware of autonomous vehicles, namely those related to the acceleration pedal. Other potential cyberattack vectors within the vehicle's firmware have not been explored comprehensively in this study.

5.2.2 Dataset Specificity

The effectiveness of the machine learning model heavily relies on the quality and representativeness of the dataset used for training and testing. The dataset used in this research may not encompass all possible real-world scenarios and variations, potentially limiting the model's ability to detect cyberattacks in diverse situations.

5.2.3 Static Analysis

The research primarily employs static analysis of acceleration pedal behavior. Real-world cyberattacks are often dynamic and may require dynamic analysis techniques to detect effectively. The model's performance in dynamic attack scenarios has not been extensively examined.

5.2.4 Hardware and Sensor Limitations

The research assumes the availability and accuracy of sensors and hardware systems in autonomous vehicles for data collection. Variability in sensor quality and availability may affect the model's practicality in real-world applications.

5.2.5 Assumption of Data Integrity

The research assumes the integrity of data sources and does not explicitly address potential data tampering or manipulation. In real-world scenarios, attackers may attempt to manipulate data to evade detection.

5.2.6 Generalization Across Vehicle Models

The model's ability to generalize across different makes and models of autonomous vehicles has not been thoroughly investigated. Variations in vehicle architecture and firmware may impact its applicability.

5.2.7 Lack of Real-world Testing

The research primarily focuses on the development and evaluation of the model in a controlled environment. Extensive real-world testing, with live autonomous vehicles and exposure to genuine cyber threats, has not been conducted.

5.2.8 Ethical and Privacy Considerations

The deployment of AI-based cyberattack detection systems in autonomous vehicles raises ethical and privacy concerns, which have not been fully addressed in this study. Future work should consider the ethical implications of monitoring vehicle behavior.

5.2.9 Resource Requirements

Implementing the AI model in real-world autonomous vehicles may require significant computational resources and continuous updates to adapt to evolving cyber threats. The resource implications have not been thoroughly examined.

5.2.10 Legislative and Regulatory Challenges

Integrating AI-based cybersecurity solutions into autonomous vehicles may face legal and regulatory challenges. Complying with evolving cybersecurity standards and regulations is an important aspect that requires further investigation.

5.3 Suggestions for future work

The culmination of our research represents not only a significant milestone but also a stepping stone towards advancing the realm of cyber-physical security for

autonomous vehicles. In this section, we outline several avenues for future research and development that can build upon the foundation laid by this study.

Multimodal Data Integration

While our research has focused on analyzing acceleration pedal behavior, future works could explore the integration of multiple data sources and sensors within the vehicle. Combining data from various sensors, such as lidar, radar, and cameras, along with behavioral analysis, could provide a more comprehensive and robust approach to cyberattack detection.

Dynamic Threat Assessment

Expanding our research to encompass dynamic threats is crucial. Future studies should investigate real-time, dynamic analysis techniques to detect cyberattacks as they occur, ensuring rapid response and mitigation in a dynamic threat landscape.

Heterogeneous Fleet Compatibility

As autonomous vehicle fleets become increasingly diverse, ensuring the compatibility of cyberattack detection systems across various makes and models is essential. Future research should focus on developing adaptable and standardized solutions that can be deployed across heterogeneous fleets.

Real-world Testing and Validation

The transition from controlled experiments to extensive real-world testing is an imperative next step. Collaborating with industry partners and conducting large-scale, practical tests in real-world environments will provide invaluable insights into the practicality and effectiveness of our cyberattack detection system.

Enhanced Model Architectures

Exploration of advanced deep learning architectures and techniques beyond the Multi-Layered Perceptron (MLP) model used in this research could yield even more accurate and efficient cyberattack detection systems. Investigating the potential of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) in this context holds promise.

Ethical Considerations

The ethical implications of deploying AI-based cyberattack detection systems in autonomous vehicles must be explored comprehensively. Future research should

delve into the ethical aspects related to privacy, data security, and responsible AI usage, aligning these systems with evolving ethical standards.

Legislative and Regulatory Compliance

The integration of cybersecurity solutions in autonomous vehicles may necessitate adherence to evolving legislative and regulatory frameworks. Collaborations with policymakers, legal experts, and industry stakeholders can facilitate the development of standards and regulations that govern the deployment of these systems.

Resource Optimization

Addressing the computational resource requirements for real-time cyberattack detection is critical. Future research should focus on optimizing resource-intensive AI models to ensure that they are deployable in practical autonomous vehicle systems.

Human-Machine Interaction Considerations

As AI-based cybersecurity systems become integral to autonomous vehicles, understanding how they interact with human drivers and passengers is crucial. Research in this area should explore user interfaces, alerts, and communication strategies to ensure effective human-machine collaboration in the event of a cyber threat.

5.4 Conclution

In conclusion, our research serves as a catalyst for further exploration in the field of cyber-physical security for autonomous vehicles. These future works hold the potential to not only enhance the security and safety of autonomous mobility but also to shape the future of cybersecurity standards and practices in the automotive industry. As we embrace these challenges, we remain committed to the ongoing pursuit of innovation, ensuring that autonomous vehicles continue to evolve as secure, reliable, and trusted modes of transportation in an increasingly connected world.

Summary

In recent years, the field of cyber security has become increasingly important as the world becomes more reliant on technology. With the increasing use of connected systems and devices, the risk of cyber attacks has also increased. The field of cyber-physical security has emerged as a critical area of research to address these concerns. The objective of this research was to study the field of cyber-physical security and to develop an observer-based approach to detect cyber attacks on distributed control systems.

Machine learning has been widely used in the field of cyber security to develop automated methods for detecting cyber attacks. The use of artificial intelligence, specifically deep learning, has shown promising results in detecting anomalies and identifying cyber attacks. However, the complexity of these systems and the dynamic nature of the data generated by them make it challenging to implement effective machine learning-based solutions.

The research work presented in this thesis focused on the use of machine learning for cyber-attack detection in autonomous systems. The study used a Simulink model to generate data and applied machine learning algorithms to detect cyber attacks. A Multi-layer Perceptron (MLP) model was selected as the final model, and the question of determining the number of layers and neurons in each layer was addressed by using Neural Architectural Search (NAS). The final pipeline was written as a clean code and TensorFlow Lite was used to decrease the model size while maintaining accuracy. The results of this research show that machine learning algorithms can be effectively used to detect cyber attacks in autonomous systems and provide a strong foundation for further research in this field.

In conclusion, this research provides a comprehensive study of the field of cyber-physical security and the observer-based approach to detect cyber attacks on distributed control systems. The results demonstrate the potential of machine learning algorithms for detecting cyber attacks in autonomous systems and provide a foundation for future research in this area.

Bibliography

[1] Juliza Jamaludin and Jemmy Mohd Rohani. «Cyber-Physical System (CPS): State of the Art». In: *2018 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*. 2018, pp. 1–5. DOI: 10.1109/ICECUBE.2018.8610996 (cit. on p. 3).

[2] George Loukas. «2 - A History of Cyber-Physical Security Incidents». In: *Cyber-Physical Attacks*. Ed. by George Loukas. Boston: Butterworth-Heinemann, 2015, pp. 21–57. ISBN: 978-0-12-801290-1. DOI: https://doi.org/10.1016/B978-0-12-801290-1.00002-3. URL: https://www.sciencedirect.com/science/article/pii/B9780128012901000023 (cit. on p. 3).

[3] Anupam Chattopadhyay and Kwok-Yan Lam. «Security of autonomous vehicle as a cyber-physical system». In: *2017 7th International Symposium on Embedded Computing and System Design (ISED)*. 2017, pp. 1–6. DOI: 10.1109/ISED.2017.8303906 (cit. on p. 4).

[4] Samar Kamil, Huda Sheikh Abdullah Siti Norul, Ahmad Firdaus, and Opeyemi Lateef Usman. «The Rise of Ransomware: A Review of Attacks, Detection Techniques, and Future Challenges». In: *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*. 2022, pp. 1–7. DOI: 10.1109/ICBATS54253.2022.9759000 (cit. on p. 6).

[5] Zohre Nasiri Zarandi and Iman Sharifi. «Detection and Identification of Cyber-Attacks in Cyber-Physical Systems Based on Machine Learning Methods». In: *2020 11th International Conference on Information and Knowledge Technology (IKT)*. 2020, pp. 107–112. DOI: 10.1109/IKT51791.2020.9345627 (cit. on p. 9).

[6] Marcello Cinque, Domenico Cotroneo, and Antonio Pecchia. «Challenges and Directions in Security Information and Event Management (SIEM)». In: *2018 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*. 2018, pp. 95–99. DOI: 10.1109/ISSREW.2018.00-24 (cit. on p. 11).

[7] Zhibo Zhang, Hussam Al Hamadi, Ernesto Damiani, Chan Yeob Yeun, and Fatma Taher. «Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research». In: *IEEE Access* 10 (2022), pp. 93104–93139. DOI: 10.1109/ACCESS.2022.3204051 (cit. on p. 12).

[8] Atif Ali, Muhammad Arif Khan, Khushboo Farid, Syed Shehryar Akbar, Amna Ilyas, Taher M. Ghazal, and Hussam Al Hamadi. «The Effect of Artificial Intelligence on Cybersecurity». In: *2023 International Conference on Business Analytics for Technology and Security (ICBATS)*. 2023, pp. 1–7. DOI: 10.1109/ICBATS57792.2023.10111151 (cit. on pp. 18, 20).

[9] Gianfranco Burzio, Giuseppe Faranda Cordella, Michele Colajanni, Mirco Marchetti, and Dario Stabili. «Cybersecurity of Connected Autonomous Vehicles : A ranking based approach». In: *2018 International Conference of Electrical and Electronic Technologies for Automotive*. 2018, pp. 1–6. DOI: 10.23919/EETA.2018.8493180 (cit. on p. 23).

[10] Jaswinder Singh and Rajdeep Banerjee. «A Study on Single and Multi-layer Perceptron Neural Network». In: *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*. 2019, pp. 35–40. DOI: 10.1109/ICCMC.2019.8819775 (cit. on p. 37).

[11] Krishna Teja Chitty-Venkata, Murali Emani, Venkatram Vishwanath, and Arun K. Somani. «Neural Architecture Search Benchmarks: Insights and Survey». In: *IEEE Access* 11 (2023), pp. 25217–25236. DOI: 10.1109/ACCESS.2023.3253818 (cit. on p. 41).