Advancing Intrusion Detection Systems on the Internet of Vehicles: Mitigating DoS and Spoofing Attacks in CAN Bus Networks

MSc Research Project

MSc in Data Analytics

Pushpak Attarde

Student ID: X22211721

School of Computing

National College of Ireland

Supervisor: Barry Haycock

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……. Pushpak Attarde ……

Student ID: ……x22211721……..……

Programme: ……MSc in Data Analytics…… **Year:** ……2023-2024……

Module: …MSc Research Project.……

Supervisor: ……Barry Haycock.……..……

Submission Due Date: …12th August 2024……..……

Project Title: … Advancing Intrusion Detection Systems on the Internet of Vehicles: Mitigating DoS and Spoofing Attacks in CAN Bus Networks ……..……..

Word Count: ……6980…… **Page Count**……22……..……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……Pushpak……..……

Date: ……12th August 2024……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Advancing Intrusion Detection Systems on the Internet of Vehicles: Mitigating DoS and Spoofing Attacks in CAN Bus Networks

Abstract

This report studies the “Intrusion Detection System (IDS)” methodologies for actually “identifying and mitigating” Denial of Service (DoS) and spoofing attacks in the “Controller Area Network (CAN) bus” of “Internet of Vehicles (IoV)” systems. This study specifically employs secondary data, the study utilises machine learning approaches, especially “Random Forest and Logistic Regression”, as well as deep learning models to create strong detection mechanisms. The investigation is performed in Jupyter Notebook utilising Python programming language, leveraging its comprehensive libraries and instruments for data analysis and machine learning. In this research both CICIDS 2017, and CICIoV 2024 dataset are used. The focus point is determined in the case of CICIoV 2024. This research explores the results to show that both “machine learning and deep learning approaches” greatly improve the detection accuracy and reaction time to possible threats. The Random Forest algorithm demonstrated increased precision in differentiating normal and attack traffic, while the Logistic Regression delivered beneficial understandings of the attribute significance of attack routines. The deep learning standards, on the other hand, excelled in catching complex designs and irregularities in the data. This thorough process not only supports the protection of IoV systems but also delivers a scalable framework for prospective analysis in vehicular network security.

Keywords: DoS, IoV, IDS, CAN bus, spoofing attack, Random Forest, Logistic Regression, Deep Learning, RNN, CNN, LSTM

1 Introduction

1.1 Background Of the study

IoV, or “Internet of Vehicles”, is a network of interconnected vehicles that can transfer data with each other and other entities in their environment. It's a large-scale dispersed system that combines three networks into one (Qureshi *et al*.,2020). Notably, “an inter-vehicle network, an intra-vehicle network, and vehicular mobile internet”. IoV enables vehicle-to-everything (V2X) communication utilising various wireless transmission technologies.

IoV can provide several benefits in modern transportation such as “Safety, Traffic management, Convenience, vehicle maintenance, and Environmental impact”. In terms of safety, interconnected cars can obtain real-time data about other vehicles and pedestrians on the road, which enables the drivers to adjust their driving to avoid accidents. The internet of vehicle(IoV) can also provide crash alerts. Also in this dynamic world, traffic control is a vital aspect, IoV can assist in optimizing traffic flow preventing traffic jams by monitoring traffic in real-time and communicating vehicle data” (Zhou et al., 2020). IoV can also assist with smart parking in crowded areas. In terms of comfort, IoV can provide remote certificates to a car, allowing for something such as a small door close, stolen detection of the motorcar, and a "find my vehicle" service. For vehicle supervision, IoV can allow manufacturers to employ data from automobiles to choose limitation (Manias, and Shami, 2021). In terms of Environmental impact, the efficiency of interconnected vehicles can stimulate individuals to operate “car-sharing” and public transportation better.

The “Controller Area Network (CAN)” bus is a communication protocol that permits machines to exchange data “reliably and efficiently”. It's generally utilised in vehicles, where it combines “electronic control units (ECUs)” and acts as the vehicle's nervous system. It works in multiple ways such as “Priority-driven communication, Real-time communication, simple wiring, and Error detection” (Zhang *et al*.,2020). CAN operates on a message-based protocol where appliances accept all transmissions but only transfers based on priority. The instrument with the most elevated priority transmission gets entrance to the bus. Real-time communication CAN help real-time transmission, permitting instruments to share critical management data. In terms of Simple wiring, CAN operates with only one pair of twisted wires generally referred to as CAN High and CAN Low(Bari *et al*.,2023). This facilitates wiring complicatedness and weight. In terms of Error detection, CAN combines error detection methods to guarantee dependable transmission.

The Internet of Vehicles (IoV) systems face many security threats such as “Data privacy, Software vulnerabilities, Supply chain attacks, Connection risks”. Modern vehicles generate a lot of data adding “personal details, location data, and driving habits”, which could be misused in a data violation. As more interconnected vehicles pound the roads, cyberpunks can manipulate software exposures utilising cellular networks, Wi-Fi, and hardline references to gain unauthorized hidden access to the vehicle network. Updates including malicious code can be compelled to associated cars, firmware can be compromised, and malware can put supplier processes to a halt.

Spoofing attacks interest an attacker sending notifications to the CAN network with a falsified commencement oration. This can cause the vehicle's techniques to receive the transmission as legitimate and execute orders based on the message content.

1.2 Aim and Objective

Aim

This study aims to investigate "intrusion detection system (IDS)" processes to effectively recognise and mitigate DoS and spoofing attacks in the CAN bus of IoV systems.

Objective

- ➢ To develop and execute an advanced IDS framework to detect DoS and spoofing attacks on the CAN bus in IoV conditions.
- ➢ To assess the implementation of the generated IDS framework in natural IoV techniques, concentrating on "detection accuracy, false positive rates, and response times".
- ➢ To present and validate security enhancement techniques that support the resilience of the CAN bus against growing DoS and spoofing threats in IoV systems.

1.3 Research Question

1. What is the development and execution process for an advanced IDS framework to detect DoS and spoofing attacks on the CAN bus in IoV conditions?
2. What is the implementation of the generated IDS framework in natural IoV techniques, concentrating on "detection accuracy, false positive rates, and response times"?
3. What are the effective security enhancement techniques that support the resilience of the CAN bus against growing DoS and spoofing threats in IoV systems?

1.4 Research rational

IDS offers a several benefits in real-world problems; it is essentially a vital component of an effective cybersecurity strategy. It essentially enables the protection of sensitive data from cyberattacks, addresses security incidents, and develops security responses. The CICIDS 2017, and CICIoV 2024 datasets are valuable for the determination of the research factors in this study. The primary motivation to work and focus on improving the IDS process to mitigate DoS and spoofing attacks is that IDS can help protect sensitive data from "cyberattacks, identify security incidents, and improve security responses" (Yang *et al.*, 2021). In the context of DoS and spoofing attacks in the IoV CAN bus, IDS can assist in protecting against malicious cyberattacks by monitoring "network traffic and system activities" for any unusual pattern. Along with this IDS can also assist in detecting unauthorized access, data breaches, and malware activities.

1.5 Research Gap

Multiple gaps were identified during the study's development, mostly concerning the topic of challenges and implementation. During the research, it has been found that there is a lack of detailed knowledge about intrusion detection system (IDS) approaches against the Denial of Service (DoS) and spoofing attacks on the internet of vehicles (IoV) controller area network (CAN) bus. Along with this there is also a shortage of detailed technical information on the effectiveness of IDS design, implementation, and maintenance specific to IoV CAN bus environment, including data on prevalent attacks and vulnerabilities. Also there are insufficient educational programs and training specifically tailored to IDS in IoV and practical experience. The

1.6 Summary

After completing this section, it has been concluded that this study aims to investigate developed and advanced "intrusion detection system (IDS)" techniques to effectively determine and mitigate DoS and spoofing attacks in the CAN bus of IoV systems. It has been discussed that IDS offer several advantages in the fight against various cyber threats especially in DoS and spoofing attacks

in the IoV CAN bus. This section covers the background of this topic and develops proper aims and objectives. Upcoming sections will evaluate different literature related to this topic specifically beneficial to properly achieve this research aim.

2 Related Work

2.1 Introduction

An intrusion detection system IDS is a network security technology that monitors a network activity for any suspicious activity or policy breaches. It can be executed as a software on an organization's hardware, as a physical device or as a cloud based solution. Along with these an IDS main function is to analyze the network traffic for any similar suspicious activities and packets and then compare them to the set of predefined rules and patterns or signatures that will indicate the potentially harmful activity. Once the IDS has detected the malicious activity that matches any signatures it will generate the alert and notify it to the system administrator. Once the administrator is notified of this alert, it is then reviewed and analyzed. After analyzing the threat they take the all the necessary actions to prevent any potential damage or unauthorized access.

2.2 Advanced IDS framework to detect DoS and spoofing attacks on the CAN bus in IoV condition

Nie et al., (2020), discuss the "Common Intrusion Detection Framework (CIDF)" Working Groups' efforts to deliver instruments for autonomous intrusion detection systems that will share the insights on potential threats, attacks analysis, and recommended responses. Pascale et al., (2021), paper focuses on automotive cybersecurity which includes the description of intrusion detection systems that monitor the vehicle's network traffic and components for unusual activities. If the intrusion is detected, then it corresponds the detected packet to a database to compare with the suspicious pattern or signatures and if suspicious it sends the data to a Security Operation Center (SOC). Then it is up to the manufacturing center who will evaluate the same and decide how it is going to respond. It is important to recognise Denial of Service and spoofing attacks within the CAN bus of IoV systems. Here the spoofing attack is when someone disguises a connection or identity to appear to be associated with a trusted authority. There are many methods for performing a spoofing attack such as Email spoofing, Caller ID spoofing, targets the CAN bus of IoV systems which can possibly increase the vulnerability int the IoV network. If the vulnerabilities are found by the threat actors, they can potentially take control of the vehicles. In terms of safety risks, Successful cyberattacks can pose the risk to a passenger and the pedestrians. In another research paper published by Yang et al., (2021), addresses the several challenges to the IoV that can affect the development of "intrusion detection systems (IDS)". IoV systems face safety challenges such as data integrity, key management, and permit control. Along with this security concerns also include the potential for session hijacking, replay attacks, and malware.

The implementation of intrusion detection system within the IoV systems integrates the IoT technology which will enhance the infrastructure. The framework aims to detect and respond to intrusion effectively, with a focus on real-time data collection, scalability and mobility challenges. (Philipsen et al., 2021) highlight that this intrusion detection system implementation addresses the complex architecture of vehicle networks and provides strategies for managing and mitigating security risks.

The IDS framework is executed for the detection of the intrusion which administers in the data commission technique. The IDS framework familiarizes various ingredients that aid the IDS. This framework interests a data assemblage guideline, processing of data, extraction of segments, detection policy, machine learning, and deep understanding procedure. The serviceable undertaking endows finding the key executional segments of the evaluation. (Giust et al., 2022)

examine the framework's effectiveness in practical IoV applications, that emphasizes the importance of continuous evaluation and adaption to emerging threats.

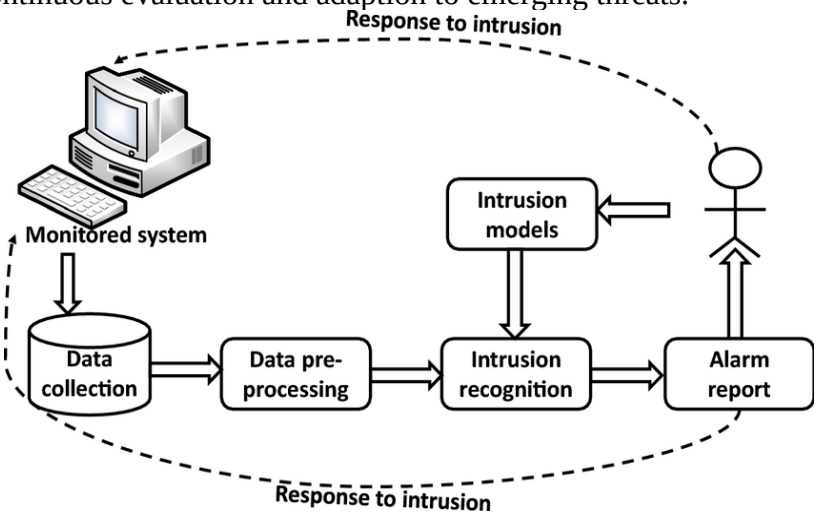


Figure 2.1: IDS framework

(Source: Sundfeldt, and Widstam, 2024)

The IDS framework application defines the involvement of the key executional elements and factors. This defines the involvement of various functional sections which introduce a monitoring system, data collection, pre-processing of the data, intrusion detection model, recognition of the intrusion, and the alarm system. This provides information about the monitoring system that is applicable to the monitoring of intrusion in the network. The data collection approach is implemented to collect the necessary data which assists in the data evaluation process. The data pre-processing approach is implemented to process the data for the execution. This provides information about the handling of the data in an informatic way. The pre-processing approach is applicable to determine the process of the finding. This defines the execution of the parametric factors. The pre-processing approach also implements the data set, sorting, filtering, null/missing data checking approach, and so on. The data-setting approach is implemented for the construction of the construction of models (Sundfeldt, and Widstam, 2024). This model provides information about the configuration of the model using machine learning, and deep learning. The intrusion detection approach is implemented to detect the intrusion factors. The recognition approach is implemented to recognize the intrusion and raise the alarm when intrusion is detected in the network. The detection approach highlights the finding of the DoS, and also the spoofing attacks. Those attacks are found in the IoT CAN bus section which needs to evaluate all the data executional parameters and factors.

2.4 Security enhancement techniques for growing DoS and spoofing threats in IoV systems

As the sophistication of the attacks is increasing there is a need for continuous advanced security enhancement techniques in IoV systems. This defines the functional execution of the data factors that are implemented for the construction of the sustainable IoV system. The implementation of the robust system highlights the introduction of the intrusion detection model that introduces the implementation of machine learning and also deep learning techniques. The implementation process defines the introduction of the investigation of the behavioral changes of the overall network section. (Taslimasa, 2023) and Tippannavar et al. (2023) discuss the integration of

machine learning and deep learning in intrusion detection system to improve the detection of behavioral anomalies and attack patterns. The enhancement helps to refine the system's response to DoS and spoofing attacks focusing on robust authentication methods. The enhancement process defines the string authentication approach to protect the internal data. The encryption process is implemented to encrypt the data and introduce secure protocols such as Transport Layer Security/ Secure Socket Layer (TLS/SSL), and Internet Protocol Security (IPSec). The protective approach is implemented to evaluate the CAN bus system. This defines the IoV execution approach which assists in the evaluation approach. The evaluation parameter supports the process of findings that provides information about the context of the provided information. (Abrar *et al.,* 2024) emphasize the importance of a comprehensive evaluation strategy that includes data analysis, model configuration and frequent auditing of security measures to adapt emerging threats efficiently.

2.5 Theoretical Underpinning

Signature-based Detection Theory

Signature-based detection is a method for identifying malicious activity by comparing network traffic to known signatures. It is a key component of security monitoring systems and is used by anti-virus software to detect threats. This is especially helpful for enhancing IDS for Dos and Spoofing episodes in IoV. Signature-based detection can recognize attacks established on characteristic patterns in network traffic, such as the number of bytes, 1s, or 0s. Apart from these, it also helps to detect attacks established on known malicious teaching sequences utilized by malware (Bhatia *et al.,* 2021). When a competition is made between network traffic and a known signature, an alert is generated.

Anomaly-based Detection Theory

Anomaly-based detection is a process of recognizing patterns in data that don't correspond to expected manners. These practices are often called "anomalies, outliers, or exceptions". Anomaly detection is employed in numerous fields, including "cybersecurity, finance, and healthcare". It has been identified that in cybersecurity, anomaly-based detection can assist in protecting methods from data breaches, financial losses, and other harmful events. It is notable that, IDS monitors system training and classifies it as either normal or anomalous. They utilize heuristics or rules, rather than practices or signatures, to detect mishandling that falls outside of standard system operation (Rajapaksha *et al.,* 2023). To completely determine attack traffic, the procedure must be prepared to recognize normal system activity. This is usually done during a training and testing phase

Specification-based Detection Theory

Specification-based detection is a technique of detecting episodes by monitoring data and lagging manners that are not permitted. It's thought to be a promising alternative to other detection methods, such as mishandling detection and abnormality detection because it incorporates their powers. Misuse detection is useful for catching known attacks, while anomaly detection can detect new ones. In specification-based detection, a program's behavioral specifications are manually specified and used as a basis for detecting attacks (Park *et al.,* 2023). For instance, a specification-based "intrusion detection system (IDS)" might celebrate "network traffic" established on specifications and traffic movements. It could even automatically create "normal and abnormal" behavioral specifications.

2.6 Conceptual Framework

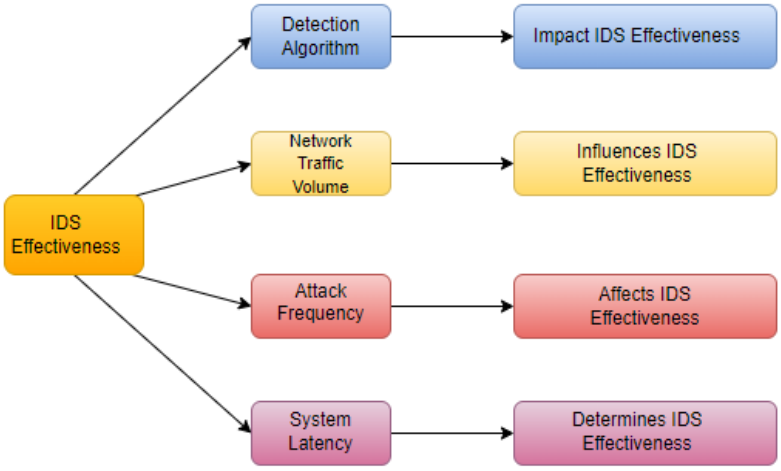


Figure 2.2: Conceptual Framework

(Source: Own-Constructed)

2.7 Literature gap

The literature gap identified in this study refers to key areas where existing research lacks the comprehensive information and details on this topic. These gaps have been in areas such as the scope of population or sample studied, research methods, data collection or analysis, and study variables or conditions. Due to the under-exploration of the impact of various electronic control units there is a significant omission identified on this topic. Although the controller area network is ideal for distributed control of complex systems through its use of a distributed arbitration architecture known as “Carrier Sense Multiple Access with Arbitration on Message Priority (CSMA/AMP)”, this important aspects have not been highlighted properly in this topic. The priority of devices which are determined by their addresses plays an important role in this process and requires the deeper discussion in future research to enhance the understanding of its impact on the overall domain(Zhang, 2023).

2.8 Summary

The overall section provides information about the details of the executional approach. The review of the papers defines the details of the overall research. The paper determines the execution of the necessary elements that are implemented for the evaluation of the major context factors. The process of finding executes the point of determination that is evaluated in the overall evaluation approach. The gap execution approach is needed to evaluate the literature portion of the overall section. This section also highlights the supportive theoretical factors which support the objectives of the research. The conceptual design section defines the overall concept of the research.

3 Research Methodology

3.1 Introduction

This chapter has been found to be focused on demonstrating the methodology which is used in this study to develop and evaluate the intrusion detection system framework in order to detect the Denial of service and spoofing attack on the internet of vehicles. It includes the research approach research philosophy data collection and others. The goal of this methodology is based on

establishing a systematic and comprehensive approach in such a manner that it can ensure reliability and validity for the findings of the research.

3.2 Research Philosophy

This research has been conducted with the help of positivism research philosophy. This research philosophy has been selected in such a manner that it has provided the use of observable and measurable facts in order to produce a reliable and objective result in the context of intrusion detection system development for the Internet of Vehicles. It also held to use the empirical data and statistical method in such a manner that it can identify and analyze the pattern of network indicators for the Denial of service and other attacks (Sharmin *et al.,* 2024). With the help of using this philosophy, the research has been found to be ensuring that the finances are based upon quantifiable evidence which can be helpful to improve the robustness and credibility of the framework which is going to be developed and the research.

3.3 Research Approach

The research approach which is chosen for this research is the detective approach which has been found to be very helpful in terms of completing the research in an appropriate. This is illustrated by the deductive research approach, which starts with the formulation of a hypothesis based upon the existing literature and theories followed by using the empirical testing of the hypothesis through the collection and analysis of the data. This procedure is furthermore identical and generous for intrusion detection technique examination as it has permitted for straining the hypothesis conveyed to the web traffic mark and incursion signature. There are several steps that are implicated in this direction, the first one is the publications review which is accomplished with the support of pinpointing existing theories and frameworks associated with the intrusion detection design heeded by materializing the hypothesis established on the anticipated pattern and differentiate of the air gridlock (Roeschlin *et al.,* 2023). The next step is based on gathering the network traffic data from the IoV environments where the data analysis is done with the help of applying statistical and machine learning techniques in order to test the hypothesis as an identifying the intrusion pattern. The framework development is also done by designing and implementing the intrusion detection system framework based on the analysis result.

3.4 Data Collection and Pre-processing

The collection of data is done from the simulated environment of IoV as well as a real-world vehicular network where the data set has included both normal and attack traffic for ensuring the comprehensive testing and training of the intrusion detection system. The preprocessing has been done with the help of several steps such as data cleaning according to which the noise is removed followed by using the irrelevant information from the data set. However, the data normalization is also done with the help of using the standardizing of the data to ensure consistent input for the machine learning models (Hafeez *et al.,* 2020). Identification and extraction of relevant features associated with the attack traffic are also performed, followed by implementing the techniques to address the missing and incomplete data entries.

3.5 Tool and Technology

The tool for this research defines the implementation of the Python tool and the coding approach. The tool defines the implementation of the Python modules/libraries. This assists in implementing the test evaluation approach. The test execution approach is implemented to evaluate the overall data. The technology defines the implementation of the data execution approach and the implementation of the data evaluation process. This defines the determination of the implementation of data execution parameters. This introduces the execution of the machine

learning technology to evaluate the data. This also involves the deep learning approach to evaluate the data.

3.6 ML/DL Model Development

The machine learning approach is implemented to construct the predictive execution approach. This provides the information about the configuration of the data model. The model construction approach introduces the ML model, and the deep learning (DL) model. The ML model introduces classification models such as Random Forest, and Logistic Regression. On the other hand, the DL model defines the deep learning model such as CNN, RNN and LSTM.

3.7 Data Analysis

The data analysis process defines the secondary data collection approach which is implemented for the evaluation. This analysis approach defines the initialization of the libraries and the data. The data processing approach is implemented to evaluate the data. This defines the checking of the data, setting of the data parameters, and so on. The data evaluation approach is implemented to configure the ML model, and also the DL model. This is implemented to evaluate the classification factors, true data evaluation values, and so on (Verma *et al.,* 2020). The accuracy evaluation approach is implemented to understand the most suitable model for determination.

3.8 Legal, Social, and Ethical Consideration

When conducting this research on improving the intrusion detection systems for the internet of vehicles, I have made sure to prioritize the ethical considerations for every step. Given how the data related to Denial of Service and spoofing attacks in CAN bus network is sensitive it is well protected. Procedures have been developed to handle the data in order to preserve the confidentiality and integrity of the information adhering to the best practices and legal requirements related to cybersecurity research.

4 Design Specification

4.1 Introduction

ML is often used to develop IDS to identify new types of attacks by analyzing network traffic. ML model can discover patterns in network data and allow the detection of possible attacks. There are several ML algorithms known for IDSs, due to high accuracy and effective performance this study uses “Random Forest and Logistic Regression”. This study also utilises “deep learning algorithms” due to it is a useful approach to intrusion detection systems (IDS) because it can learn complicated patterns in data and has increased accuracy with less activity time. DL models can learn conceptual and high-dimensional feature illustrations of IDS data by giving them through multiple hidden layers. This study analyses “CICIoV2024” data for analysis of IDS approaches against DoS and spoofing attacks in the IoV CAN bus.

4.2 Design concept and data analysis

This study shed light on using the ML approach to analyse the data as this approach helps identify patterns in data through data exploration, visualization, and mining. This study also builds models, this model can be prepared to create more authentic projections based on historical data. ML algorithms can automate repetitious tasks such as “data cleaning, preprocessing, and manual manipulation.” It allows the creation and advanced “intrusion detection system (IDS)” methods to identify and mitigate DoS and spoofing attacks in the CAN bus of IoV systems (Kocher, and Kumar, 2021). This study mainly uses Jupyter notebooks which help to analyse data by allowing users to “combine code, visualizations, and explanatory text” into a single, shareable composition. This makes it easier to communicate analytical insights and collaborate with others.

This study uses the Random Forest ML model to analyse data as they are more objective focused than most non-linear classifiers and can deliver the most elevated accuracy among category methods. They incorporate numerous decision trees to connect a development, which can lead to better accurate forecasts than individual decision trees (Ahmad *et al*., 2021). This research also uses Logistic regression for the investigation of developed and advanced IDS processes to effectively recognise and mitigate DoS and spoofing attacks in the CAN bus of IoV systems. Logistic regression is essentially valuable for exploring strategies with a binary conditional variable, such as "yes or no" or "0/1", and for indicating possible results established on earlier details.

Deep learning also helps to analyse data, this benefits computers to process the data in a way that's comparable to how the human brain processes (Thapa *et al*.,2020). Deep learning standards can identify complicated patterns in data, such as “images, text, and sounds”, to construct accurate forecasts and understandings.

Step of execution

The step of execution highlights the approach of the data analysis process. This defines the first step which is the initialization of the test process. In this case, the libraries/modules are implemented. The next section is the data reading and processing section followed by the collected data. After that, data frames are constructed for each dataset. Then the merge functionality is implemented to merge all the data and create a single data frame. After that a null checking approach is implemented to check the null value of the data. The null replacement approach is implemented to replace the null with zero. Then the visualization approach is implemented to determine some relational evaluation between various data. After that multiple ML/DL models are constructed such as RNN, Random Forest, CNN, Logistic Regression, and LSTM. The evaluation of each model parameter is highlighted in this section. Lastly the best suitable model is derived by evaluating the accuracy of various models.

Flow Chart

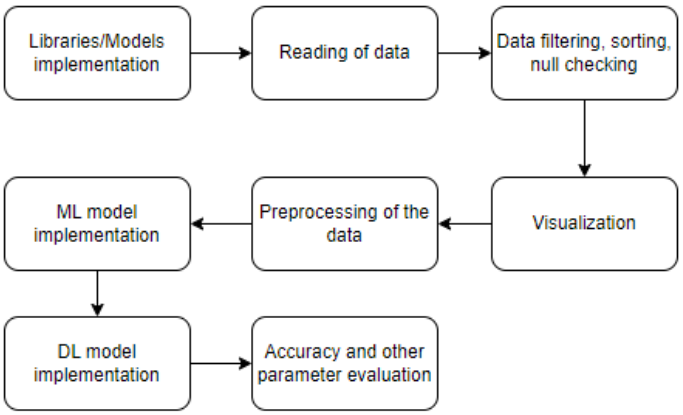


Figure 4.1: Flow chart

(Source: Self-Created)

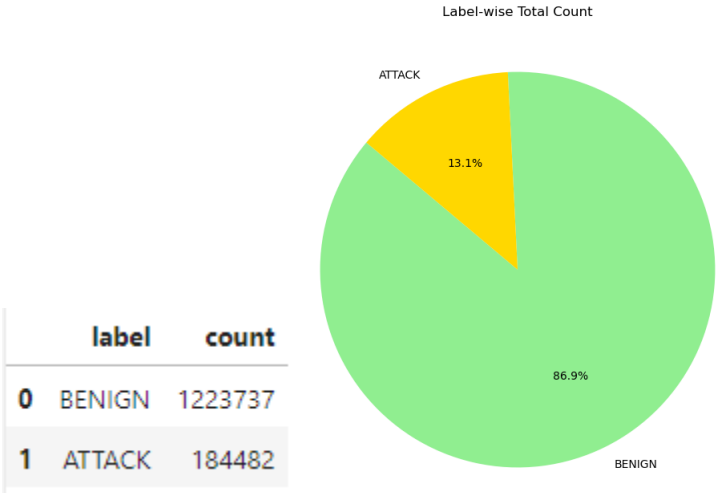
The flow chart determines the stepwise execution process which is applicable to evaluate the collected data. This defines all the necessary steps for the execution of the collected data parameters and find the most suitable attack prediction model.

5 Implementation

5.1 Experimental Evaluation and Implementation

The importing Libraries mostly contain pre-written code for common tasks. This especially allows developers to avoid reinventing the wheel and saving significant development time.

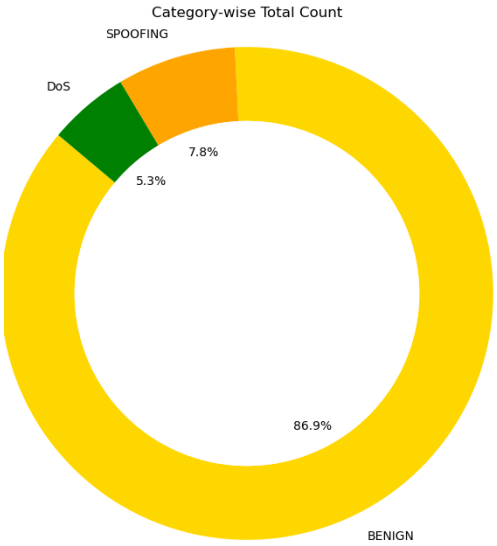
Data loading refers to the process of importing or reading data from external sources and transforming it into a design that can be utilised by the ML algorithms.



	label	count
0	BENIGN	1223737
1	ATTACK	184482

Figure 5.3: Label-wise Total Count

This figure shows a pie chart of the label-wise total count, it has been identified that a total of 13.1% of data shows attacks and 86.9% of data is Benign. In terms of number count total of 1223737 data on Benign and 184482 data for attack.



	category	count
0	BENIGN	1223737
1	SPOOFING	109819
2	DoS	74663

Figure 5.4: Category-Wise Total Count

This figure shows the category-wise total count, through this pie chart it has been identified that DoS attacks contain 5.3% and Spoofing attacks 7.8%. In terms of numeral data, it has been identified that a total of 109819 data for spoofing attacks and 74663 data for DoS attacks.

	specific_class	count
0	BENIGN	1223737
1	DoS	74663
2	RPM	54900
3	SPEED	24951
4	STEERING_WHEEL	19977

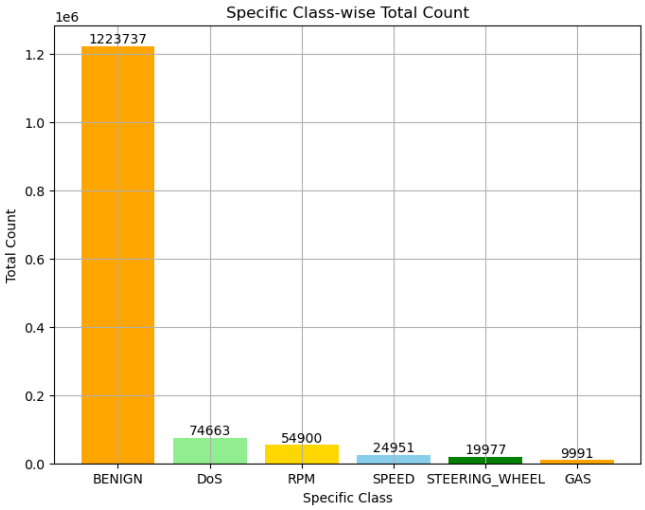


Figure 5.5: Specific Class- wise Total Count

The above figure shows bar graphs for the specific class-wise total count, the X-axis showing the specific class and the Y-axis showing the Total count.

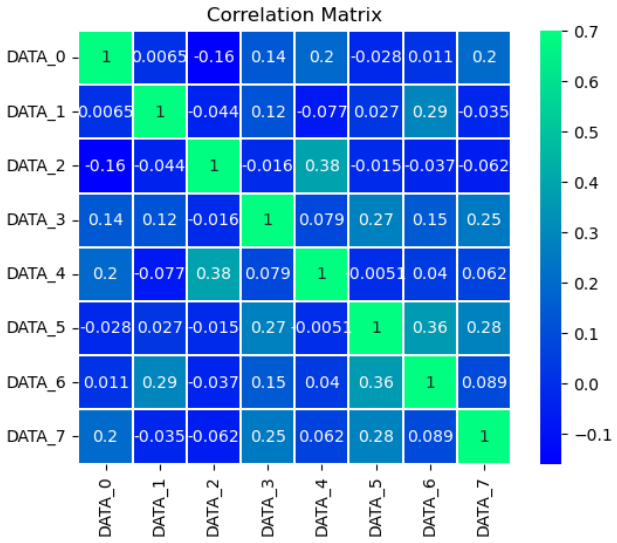


Figure 5.6: Correlation Matrix

(Source: Self-Created)

This Correlation Matrix displays the correlation coefficients between various variables in statistical data investigation. It's a reasonable tool for translating extensive data sets, identifying marks, and exploring the strength and leadership of associations between variables.

Data preprocessing is an essential phase in ML because it provides data quality and improves model performance. Data normalization, Data configuration, Data selection, Data scaling, and Integration of data from numerous authorities.

6. Evaluation

RF Report-

Accuracy value is	0.95
AUC value is	0.95
Precision value	0.74
Recall value is	0.95
F1 value is	0.83

Logistic regression is a flexible and robust method for predicting binary products or conditions, such as yes/no, success/failure, or intention that occurs/won't occur. It's often employed to comprehend connections between a conditional variable and one or more separate variables.

LR Report-

Accuracy value is	0.95
AUC value is	0.95
Precision value	0.74
Recall value is	0.95
F1 value is	0.83

6.1 Case Study 1: CICIDS 2017 Dataset

It is determined that to enhance the real-world applicability of this research's IDS analysis, the current work shifted from the CICIDS 2017 dataset to the more recent CICIov 2024 dataset. More recent products of attacks are represented in the CICIov 2024 dataset, specifically in the IoT of Vehicles (IoV). It contains not only enhanced patterns and signatures of DoS attacks and simulation that characterize current CAN-bus systems, which allows this study to train and check the IDS on data remembering the current threat techniques. Such a shift is instructed to complete sound solutions to counter everyday hazards in IoV systems' cybersecurity.

6.2 Case Study 2: RNN, CNN, AND LSTM for Decimal Data.

Thus, in the current work, the confidentiality of the work stayed ensured by employing the CICIov 2024 dataset, and, in certain, the decimal department, to detect intrusions in IoV systems. After that, this hexadecimal data was preprocessed for investigations with deep learning models including CNN, RNN and LSTM. These instances are very reasonable when it comes to analysing sequential data; for illustration, the CAN bus traffic and that is why they can recognise the patterns that are associated with the DoS and the spoofing attacks. The 100% accuracy and accuracy scored most presumably can be an illustration of over-training of the models where the learnt models can replicate the training data independently without the ability to give accurate projections of unseen data. This consequently calls for better approaches to the partitioning of the data or reasonably still approaches to data pre-processing such as regularization.

6.3 Case Study 3: RNN, CNN, AND LSTM for Hexadecimal Data.

1. RNN

The RNN model construction approach is evaluated in this section. This highlights the evaluation of the RNN by using the epochs. The parameter determination of RNN defines the evaluation of accuracy and other parameters of RNN.

RNN Report-

Accuracy value is	0.99
AUC value is	0.98
Precision value	0.99
Recall value is	0.97
F1 value is	0.98

The RNN model prediction plot is evaluated to determine the actual and the predicted labels differences. This defines the circular actual labels and crossed predicted labels for RNN.

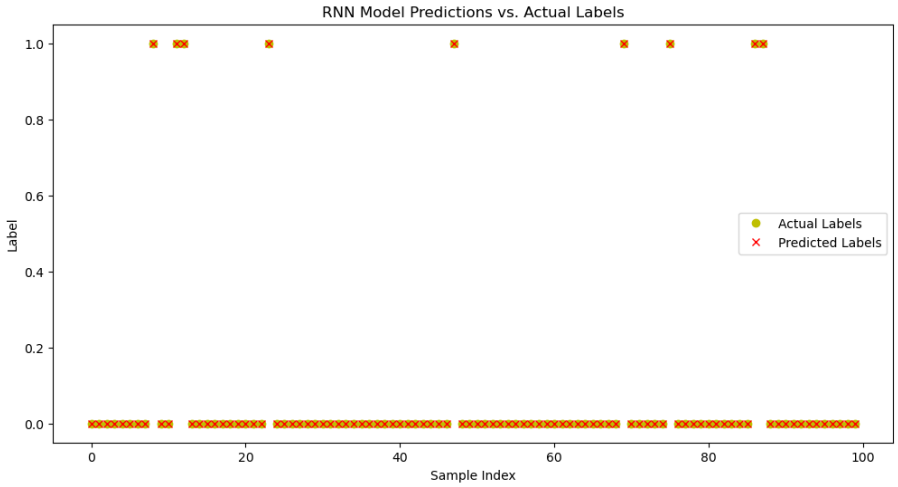


Figure 6.5: RNN Model prediction plot

(Source: Self-Created)

The model evaluated parameter for RNN is illustrated by using this plot (Bar). This highlights the parameters of the model such as Accuracy, Precision, AUC, Recall, and F1-score. In this case of RNN, all the factors have the same value which is 1.

2: LSTM

The LSTM model construction approach is evaluated in this section. This highlights the evaluation of the LSTM by using the epochs. The parameter determination of RNN defines the evaluation of accuracy and other parameters of LSTM.

LSTM Report-

Accuracy value is	0.99
AUC value is	0.98
Precision value	0.99
Recall value is	0.97
F1 value is	0.98

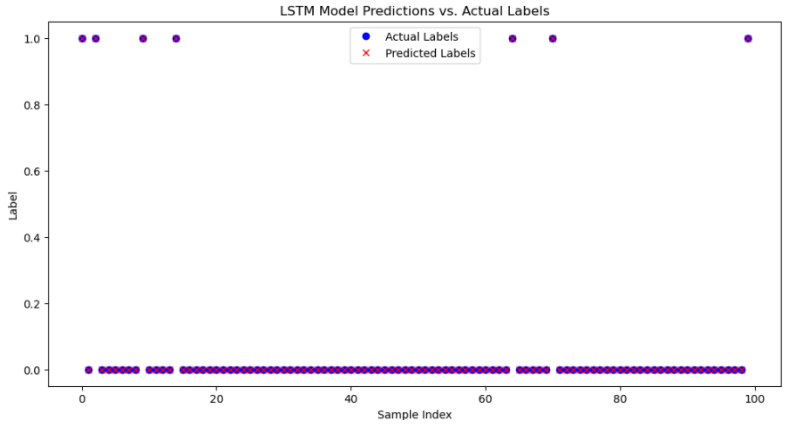


Figure 6.7: LSTM Model prediction plot

(Source: Self-Created)

The LSTM model prediction plot is evaluated to determine the actual and the predicted labels differences. This defines the circular actual labels and crossed predicted labels for LSTM.

The model evaluated parameter for LSTM is illustrated by using this plot (Bar). This highlights the parameters of the model such as Accuracy, Precision, AUC, Recall, and F1-score. In this case of LSTM, all the factors have the same value, which is 1.

3: CNN

Deep learning can seriously enhance the precision of vision and speech awards. For instance, AI representatives can utilise neural networks to recognise “objects, people, and even” sentiments in ideas.

CNN Report-

Accuracy value is	0.99
AUC value is	0.98
Precision value	1
Recall value is	0.97
F1 value is	0.98

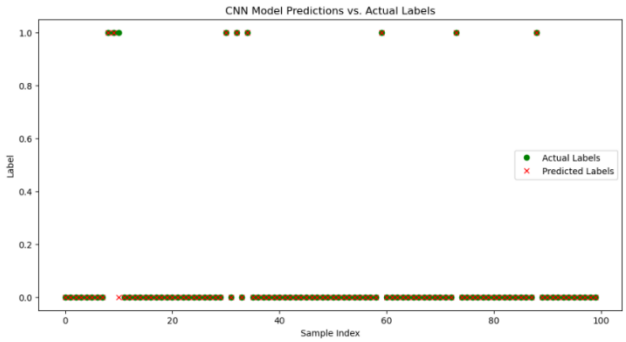


Figure 6.10: CNN Model prediction plot

(Source: Self-Created)

The CNN model prediction plot is evaluated to determine the actual and the predicted labels differences. This defines the circular actual labels and crossed predicted labels for CNN.

The model evaluated parameter for CNN is illustrated by using this plot (Bar). This highlights the parameters of the model such as Accuracy, Precision, AUC, Recall, and F1-score. In this case of CNN, all the factors have same value which is 1.

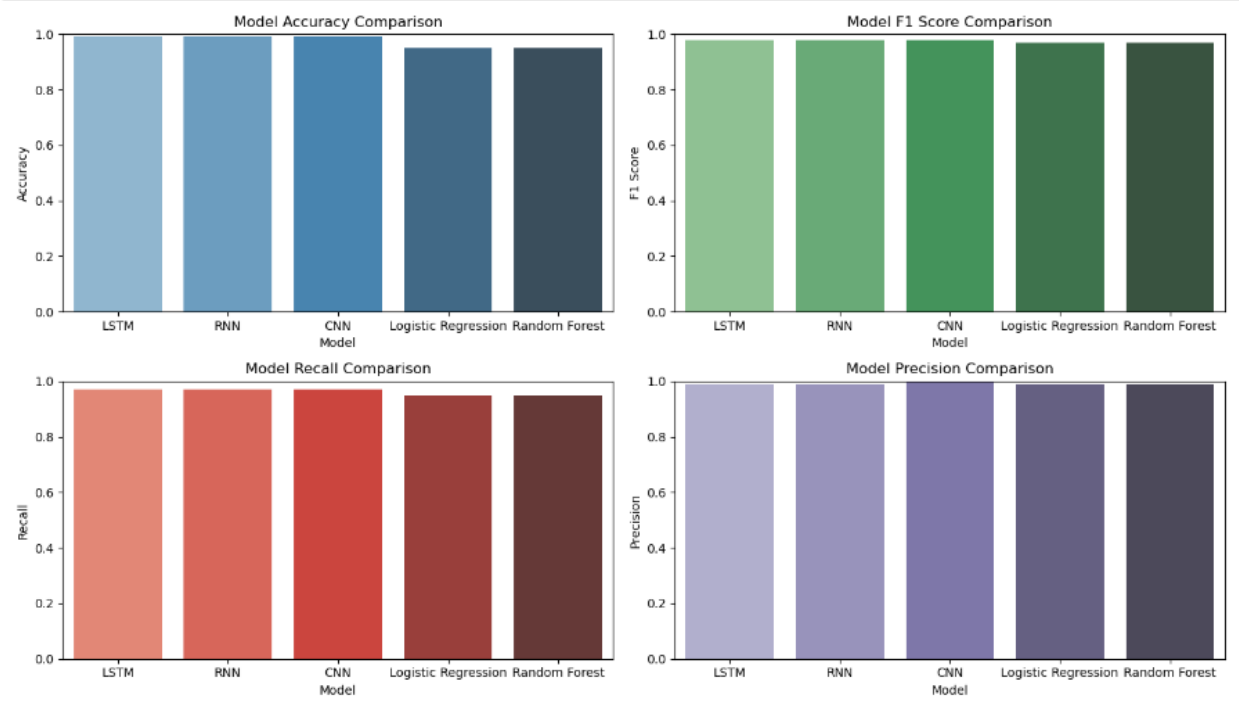


Figure 6.12: Model parameter determination

(Source: Self-Created)

The overall model parameter determination plot is demonstrated in this point of execution.

```
print(df)
```

	Model	Accuracy	F1 Score	Recall	Precision
0	LSTM	0.99	0.98	0.97	0.99
1	RNN	0.99	0.98	0.97	0.99
2	CNN	0.99	0.98	0.97	1.00
3	Logistic Regression	0.95	0.97	0.95	0.99
4	Random Forest	0.95	0.97	0.95	0.99

Figure 6.13: Model parameters value

(Source: Self-Created)

Model parameters value which are displayed in the tabular formation.

6.4 Discussion

This study vitally focused on building RF. RF standards numerous decision trees trained on additional parts of the same activity set, which can reduce conflict and boost model interpretation. It is also identified that Random forests can tolerate "binary, categorical, and numerical" components without demanding component normalization. Random forests are utilised in numerous initiatives, including banking, commodities trading, treatment, and e-commerce. The report of this random forest "Accuracy value is 0.94991, AUC value is 0.94937, Precision value is 0.74135, Recall value is 0.94864, F1 value is 0.83228".

This study also uses Logistic regression to investigate developed and advanced IDS processes to effectively recognise and mitigate DoS and spoofing attacks in the CAN bus of IoV systems. This study uses logistic regression, it is also helpful for examining numerous characteristics that influence an opposing or positive development. It can also be utilised to preprocess data, such as

sorting data with a considerable scope of importance into a more undersized, limited capacity of importance. The report of this LR is “Accuracy value is 0.94991, AUC value is 0.94937, Precision value is 0.74135, Recall value is 0.94864, F1 value is 0.83228”.

Deep learning is a subset of machine learning that uses neural networks to learn from data and identify patterns. It's considered important for DDoS attack detection because it can effectively distinguish complex patterns in network traffic data. Deep learning models can automatically extract features and classify data, making them suitable for DDoS attack detection. It is also notable that, Studies have shown that deep learning models can achieve high accuracy in DDoS attack detection, with some examples reaching 99% accuracy and 100% precision.

Therefore, it has been summarised that all models have relatively the same accuracy, precision, recall and F1 value.

7 Conclusion and Future Work

7.1 Conclusion

The research on intrusion detection system within the internet of vehicles has provided the comprehensive insights into the efficacy of different machine learning models to detect and mitigate the threats like Denial of Services and spoofing attacks on the controller area network bus. Through testing and evaluation using the datasets the study has confirmed the vital role of advanced computational models in enhancing the security framework of vehicle networks. Models random forest and logistic regression proved effective in distinguishing the normal and malicious traffic. Random forest showed high precision in identifying the attack vectors, making it more valuable for real-time security applications in IoV. Apart from random forest, techniques like CNN, RNN, and LSTM too demonstrated their capability to decode the complex and non-linear patterns in network data, indicative of cyber-attacks. The study underscores the necessity for ongoing research to refine the techniques and explore their integration into a unified security framework that protects against a broader spectrum of vulnerabilities in IoV network.

7.2 Future Work

Future research should focus on advancing the design and implementation of detection algorithms that integrate machine learning and artificial intelligence more in-depth to improve the accuracy and efficiency of detecting the Denial of Service and spoofing attack. These algorithms should be qualified of examining practices and abnormalities in real-time to fast witness and react to possible threats. Future work also highlights the implementation of the logical parameters for the execution of the data factors. It is suggested that ML and AI can greatly enhance the ability of IDS to address the difficult and growing attack practices, decreasing the false positives and improving the dependability of the design. It can be confirmed that the IDS can seamlessly combined with existing IoV procedures and additional security measures. The IDS should be designed to function in a coordinated method with other network protection tools and protocols, delivering an exhaustive protection framework.

Reference List

Bari, B.S., Yelamarthi, K. and Ghafoor, S., 2023. Intrusion detection in vehicle controller area network (can) bus using machine learning: A comparative performance study. Sensors, 23(7), p.3610.

Introduction to can (Controller Area Network) - technical articles (no date) All About Circuits. Available at: https://www.allaboutcircuits.com/technical-articles/introduction-to-can-controller-area-network/ (Accessed: 11 July 2024).

Manias, D.M. and Shami, A., 2021. Making a case for federated learning in the internet of vehicles and intelligent transportation systems. IEEE network, 35(3), pp.88-94.

Nie, L., Ning, Z., Wang, X., Hu, X., Cheng, J. and Li, Y., 2020. Data-driven intrusion detection for intelligent internet of vehicles: A deep convolutional neural network-based method. IEEE Transactions on Network Science and Engineering, 7(4), pp.2219-2230.

Pascale, F., Adinolfi, E.A., Coppola, S. and Santonicola, E., 2021. Cybersecurity in automotive: An intrusion detection system in connected vehicles. Electronics, 10(15), p.1765.

Qureshi, K.N., Din, S., Jeon, G. and Piccialli, F., 2020. Internet of vehicles: Key technologies, network model, solutions and challenges with future aspects. IEEE Transactions on Intelligent Transportation Systems, 22(3), pp.1777-1786.

Yang, L., Moubayed, A. and Shami, A., 2021. MTH-IDS: A multitiered hybrid intrusion detection system for internet vehicles. IEEE Internet of Things Journal, 9(1), pp.616-632.

Zhang, H., Meng, X., Zhang, X. and Liu, Z., 2020. CANsec: A practical in-vehicle controller area network security evaluation tool. Sensors, 20(17), p.4900.

Zhang, Y., 2024. Digital Twin for the Internet of Vehicles. In Digital Twin: Architectures, Networks, and Applications (pp. 105-120). Cham: Springer Nature Switzerland.

Zhou, H., Xu, W., Chen, J. and Wang, W., 2020. Evolutionary V2X technologies toward the Internet of vehicles: Challenges and opportunities. Proceedings of the IEEE, 108(2), pp.308-323.

Philipsen, S.G., Andersen, B. and Singh, B., 2021, November. Threats and attacks to modern vehicles. In 2021 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS) (pp. 22-27). IEEE.

Giust, A., Isoaho, J. and Adu-Kyere, A., 2022. A Study of Automotive Security-CAN Bus Intrusion detection Systems, Attack Surface, and Regulations. innovation.

Sundfeldt, F. and Widstam, B., 2024. Intrusion Detection for In-Vehicle CAN Communication Using Deep Neural Networks.

Taslimasa, H., 2023. Enhancing network intrusion detection of the Internet of Vehicles: Challenges and proposed solutions.

Tippannavar, S.S., Vanditha, M. and Nayak, P., 2023. Smart Intrusion Detection System for CAN Network Implemented using LSTM Strategy.

Abrar, M.M., Islam, R., Satam, S., Shao, S., Hariri, S. and Satam, P., 2024. GPS-IDS: An Anomaly-based GPS Spoofing Attack Detection Framework for Autonomous Vehicles. arXiv preprint arXiv:2405.08359.

Bhatia, R., Kumar, V., Serag, K., Celik, Z.B., Payer, M. and Xu, D., 2021, February. Evading Voltage-Based Intrusion Detection on Automotive CAN. In NDSS.

Rajapaksha, S., Kalutarage, H., Al-Kadri, M.O., Petrovski, A., Madzudzo, G. and Cheah, M., 2023. Ai-based intrusion detection systems for in-vehicle networks: A survey. ACM Computing Surveys, 55(11), pp.1-40.

Park, S.B., Jo, H.J. and Lee, D.H., 2023. G-idcs: Graph-based intrusion detection and classification system for can protocol. IEEE Access, 11, pp.39213-39227.

Zhang, L., 2023. Intrusion Detection Systems to Secure In-Vehicle Networks (Doctoral dissertation).
Sharmin, S., Mansor, H., Kadir, A.F.A. and Aziz, N.A., 2024. Benchmarking Frameworks and Comparative Studies of Controller Area Network (CAN) Intrusion Detection Systems: A Review. arXiv preprint arXiv:2402.06904.
Roeschlin, M., Camurati, G., Brunner, P., Mridula, S. and Srdjan, C., 2023, March. EdgeTDC: On the security of time difference of arrival measurements in CAN bus systems. In NDSS.
Hafeez, A., Rehman, K. and Malik, H., 2020. State of the art survey on comparison of physical fingerprinting-based intrusion detection techniques for in-vehicle security (No. 2020-01-0721). SAE Technical Paper.
Verma, M.E., Iannacone, M.D., Bridges, R.A., Hollifield, S.C., Moriano, P., Kay, B. and Combs, F.L., 2020. Addressing the lack of comparability & testing in can intrusion detection research: A comprehensive guide to can ids data & introduction of the road dataset. arXiv preprint arXiv:2012.14600.
Kocher, G. and Kumar, G., 2021. Machine learning and deep learning methods for intrusion detection systems: recent developments and challenges. Soft Computing, 25(15), pp.9731-9763.
Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J. and Ahmad, F., 2021. Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Transactions on Emerging Telecommunications Technologies, 32(1), p.e4150.
Thapa, N., Liu, Z., Kc, D.B., Gokaraju, B. and Roy, K., 2020. Comparison of machine learning and deep learning models for network intrusion detection systems. Future Internet, 12(10), p.167.