International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V70I5P238

Volume 70 Issue 5, 346-356, May 2022
© 2022 Seventh Sense Research Group®

Original Article

Fuzzy Anomalous Rules-based Car Hacking Detection using Lasso Regression Approach

S. Senthil Kumar[1] and S. Mythili[2]

[1]*Time Research Scholar, Department of Computer Science, Kongunadu Arts and Science College, (Autonomous),Coimbatore, Tamil Nadu, India.*
[2]*Associate Professor & Head, Department of Information Technology, Kongunadu Arts and Science College, (Autonomous), Coimbatore, Tamil Nadu, India.*

[1*]ssksnsmca@gmail.com

Received: 04 March 2022 Revised: 07 May 2022 Accepted: 29 May 2022 Published: 31 May 2022

Abstract - *Car hacking is the exploitation of vulnerabilities within cars' software, hardware, and communication systems. Various kinds of attacks can be injected to perform car hacking, affecting the electronic control unit to exploit the vulnerability. Predicting whether car hacking is present or not is the most difficult task. In the previous research, we introduced the Lasso Regression-based Improved Anomalous Detection Algorithm (LR-IADS).The main aim of the research work is to implementa credit card dataset to predict whether the fraudulent transaction is happening in the environment. In this research work, anomalous fuzzy rules were created initially with the help of attributes chosen from the database. Based on the Gini index, information gain, and gain ratio, we choose the attributes here. The lasso regression analysis method helps to do the rule pruning on the generated anomalous rules.*

At last, unexpected suspicious detection is done according to these anomalous rules by commencing the classification process. IRVMs (Improved Relevant Vector Machines) perform it based on Association Classifiers. This research work is implemented on the car hacking database for intrusion detection gathered from the controller area network. The complete analysis of the study work is performed in a Matlab simulation platform, demonstrating that the suggested LR-IADS approach may provide accurate car hacking detection results.

Keywords - *Anomalous Rules, Car Hacking Detection, Gini Index, Information Gain, Relevance Vector Machine.*

1. Introduction

The advancements in automobile technology have made life easier for drivers [1]. Nevertheless, since V2X technology allows for interactions with vehicles as well as anything from the outside (such as vehicles and infrastructure), security risks to vehicle ECUs (electronic control units) are increasing [2]. As a result, they build a security system to reduce the vehicle's different hazards [3]. To safeguard vehicle ECUs and their related equipment against new threats, IDSs (intrusion detection systems) for in-vehicle networks are mandatory. CANs (Controller Area Networks) are in-vehicle bus system protocols that allow ECUs to communicate effectively [4] and reliable in-vehicle networks. CANs are alsolow-cost serial buses. Attackers can readily access CAN buses since they employ broadcasts without authentication, posing serious security concerns. Attackers may, for example, insert malicious packets into CAN buses by exploiting their weakness in one of many external interfaces. [5]. Many modern cars with an infotainment communication module are also vulnerable to assaults via an OTA update module. These threats might result in serious car failures and threats to driver safety.

To ensure data integrity in all ECUs of the system, CANs transmit multiple brief messages, including information on the vehicle state [6]. Nevertheless, CAN's communication is not protected from cyber-attacks since it lacks security features like authentication and encryption. Investigations have indicated the in-vehicle network's security vulnerabilities [7]. Because the protocols of CANs lack protective measures, attackers can use faked messages to manipulate vehicle systems. An attacker, for example, can change the car's gear by injecting messages with the specific CANs identities associated with gear functions. The onboard diagnostic (OBD-II) port, the infotainment system, or the wireless communication system can all be used to inject fabricated messages [8]. Many vulnerable components of car systems have been discovered in recent years, including the electric window raise, warning lights, airbag, and tyre pressure monitoring system (TPMS).

This is an open-access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Kvaser's Hybrid 2xCAN/LINs (Local Interconnect Networks) are adaptable dual-channel interfaces configured as CAN Buses or LINs independently with standard USB connectors. Two high-speed CANs or channels of LINs in two separate 9-pin D-SUB CANs connectors are high-performing tiny 'universal interfaces' that just require engineers interested in automotive communications. Using devices as dual-channel CAN interface, or programming them to connect two high-speed buses of LINs, and CANsto PCs or mobile computers. Kvaser Hybrid supports CANs FD and includes Kvaser TRX and basic development environments for getting started in programming these devices.

Many researchers have wanted to enhance the protocols of CAN's security by including digital signatures relying on pair-wise symmetric secret keys [9]. On the other hand, digital signatures have a significant communication cost, and CAN's bandwidth is restricted to 500 kbps [10]. Additional to the cost, replacing all of the equipment in existing vehicles is challenging. As a result, an investigation is necessary on IDSs which don't generate communication overhead and are integrated into current systems [11].

The ability of IDSs (intrusion detection systems) to continually monitor in-vehicle systems and identify fraudulent network events created by ECUs are the greatest approaches to identifying and responding to known and new threats [12]. Recent studies have been targeting the use of IDSs to detect assaults targeted on vehicles. When in-vehicle surroundings are altered, a lot of changes are required. Furthermore, because particular attacks are reflected while designing detection systems, the targets to be identified may be restricted. If an attacker gains access to IDSs, they can alter them to avoid detection.

We have applied the Lasso Regression-based Improved Anomaly Detection system for attack detection in this research work. This work utilizes the fuzzy anomalous rule generation approach, which is utilized in this work for car hacking detection. This work will generate the anomalous rules from the input car hacking data set from which it will be proved whether the car hacking happens or not. The analysis of the study is carried out in a MatLab simulation environment to confirm the suggested technique's applicability to over car hacking dataset.

The following is the overall structure of the research project: This part has a detailed description of the automotive mobiles with the need for car hacking detection explained. In section 2, a deep analysis ofvarious research methodologies attempted to predict car hacking. In section 3, the suggested research method analysis and appropriate instances and figures are given. In section 4, a discussion of simulation outcomes is given. The entire work's conclusion is given depending on simulation results are explained in Section 5.

2. Related Works

Seo et al. [13] attempted to identify new vehicle threats by merely knowing one class using IDS datasets for their Car-Hacking dataset's classification performances. The database was created by monitoring CAN activities from cars through OBD-II connectors. There are four different assault types in the dataset. Single class classification is an unsupervised learning method for classifying attack classes based on normal class information. It is hard to obtain higher efficiency while utilizing unsupervised learning since it does not employ negative examples for learning. The proposed approach presents a set of characteristics that can recognise new risks while classifying normal datasets effectively.

To identify threats to buses of CANs, Taylor et al. [14] proposed LSTMNNs (Long Short-Term Memories based neural networks) for detecting anomalies. The bus transmitters sent data words, and detectors learned to anticipate these words. Parts after the words or completely unexpected words were called Anomalies. The study's synthetic anomalies were meant to resemble attacks that were recorded in the past. According to the study, detectors could identify abnormalities generated with minimal false alarm rates. Moreover, granularitiesin bit-wise predictions can assist forensic investigators with information on discovered types of irregularities.

Jaoudi et al. [15] explored the applicability of spiking neural networks for in-vehicle cyber-attack detection. The authors show exemplary results by converting an auto encoder model to a spiking form. The authors presented a learning model comparison that shows the proposed SNN auto encoder outperforms a One-Class Support Vector Machine and an Isolation Forest. Furthermore, only a slight reduction in accuracy is observed compared to a traditional autoencoder.

Han et al. [16] presented a technique for detecting abnormal vehicle conditions. Using one-way ANOVA tests, these researchers analyzed commercial vehicles generated in-vehicle traffic. Consequently, their statistical-based detection approaches differentiated abnormal conditions of networked automobiles in the context of IoTs (Internet of Things).

From September 14, 2020, through November 27, 2020, Kang et al. [17] held the Car Hacking: Attack &Defence Challenge, in which several security firms and researchers took part. Based on the real-world vehicle environment, the participants devised a variety of injection attacks and high-performance detection techniques. The final round was dominated by rule-based and ensemble tree-based systems. Time intervals and data byte patterns were also useful in detecting threats.

Martinelli et al. [18] presented a fuzzy algorithm-based approach for detecting 4 types of attacks on protocols of CANs. Using the fuzzy NN method, the researchers were trying to identify threats targeting protocols of CANs with an accuracy ranging from 0.85 to 1.

Li et al. [19] suggested that IDSs built on a regression learning technique that uses correlated/redundant data to estimate specific parameters. To detect uncommon conditions that might indicate intrusions, the study compared estimated and observed values where more than 90% of vehicledata were anticipated using 3 kph error limits, according to results based on real-world vehicle data. Their suggested IDSs could identify and localize threats in real-time, which can be critical for achieving automobile security.

In contrast to the traditional methodology, Barletta et al. [20] developed a new distance-based mechanism for integrating the SOM networks using the K-means clustering, which were tested with car hacking databases that encompassed traffic information packets generated using CAN buseswhere more information with uneven data distributions was noticed.

Taylor et al. [21] described a sliding window method for measuring inter-packet time. An abnormal signal is generated by comparing the average timings to historical averages. The authors tested their method over a range of insertion frequencies and found that it had limitations. Also demonstrated is how a comparable metric of packet data content is ineffective for detecting abnormalities. Lastly, demonstrate how a one-class SVMs (|Support vector Machines)can utilize the same data to find abnormalities reliably.

Taylor et al. [22] created a model that defines threats depending on their influence on bus traffic, following a review of published information on vehicle hacking and traffic analyses of CANs from a 2012 Subaru Impreza. Attacks that introduce foreign packets, attacks that change packet timing, and merely modify data within packets are all classified as high-level effects in the model. Attacks using foreign packets are easily recognized. They created features that are suited for one-class classification techniques for timing-based anomalies. They applied recurrent neural networks and multivariate Markov model techniques to sequence anomaly detection and contrasted their results for packet stream data word anomalies.

Spicer et al. [23] presented a technique for identifyingmessages arriving from incorrect sources as suspicious by nodes depending on the noise contentsof signals. Their approach used MATLAB and Python to execute different data transformations and determine noisy characteristics in signals. They were then subjected to statistical analyses and assigned scores based on the usability of the information. Bestperforming features were passed through MLPs (multilayer Perceptrons) and SVMsfor comparison results. The methods performed well in predictions,with resulting prediction accuracies of 99.9% for MLPs and 99.8% for SVMs.

3. Car Hacking Detection

The hacking of cars is executed by modifying code in vulnerable areas of ECUs and controlling the vehicle's other ECUs. Automobile hackers have attempted to obtain control of different car models and make and systems,including centres for entertainment, gas gauges, airbags, steering, brakes, and accelerometers, where both driven and self-driving vehicle systems have been successfully hacked using proof-of-concept (POC) threats. The flow of the car hacking process is shown in the following Fig. 1.

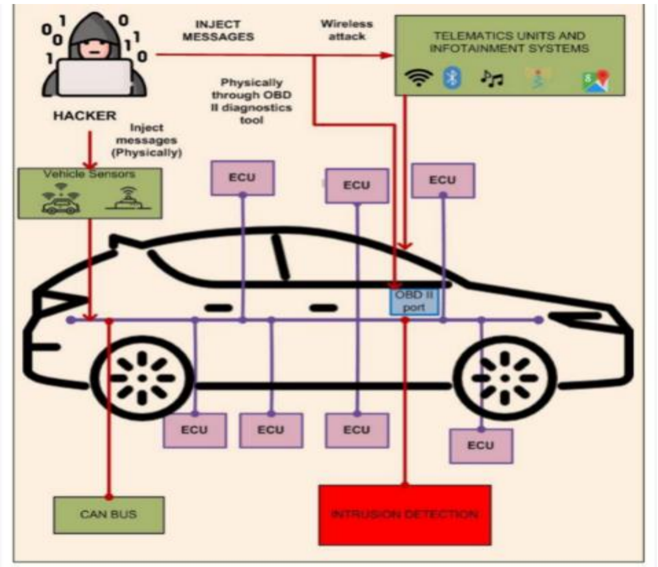
HACKER
INJECT MESSAGES
Wireless attack
TELEMATICS UNITS AND INFOTAINMENT SYSTEMS
Physically through OBD II diagnostics tool
Inject messages (Physically)
Vehicle Sensors
ECU
ECU
ECU
OBD II port
ECU
ECU
ECU
ECU
CAN BUS
INTRUSION DETECTION

Fig. 1Car hacking processing interconnections

The surface attacks on CANs buses are depicted in Fig. 1. Messages can be fed into the network directly through OBD-II connections or telematic wireless,or systems used for infotainments.

In recent days, the most threatening issue is Anomalous detections that have to be recognized as early as possible to avoid unexpected suspicious observations. So, an anonymous rule generation process should be developed to accomplish this task for early Prediction of this issue from the database. In our previous research, LR-IADS was applied on the credit card dataset to predict fraudulent transactions and proved to provide a better outcome.

4. Proposed Methodology

In this research work, LR-IADS is utilized to predict the car accurately hacking presence. The overallprocessing flow of car hacking detection is shown in the following Fig. 2.

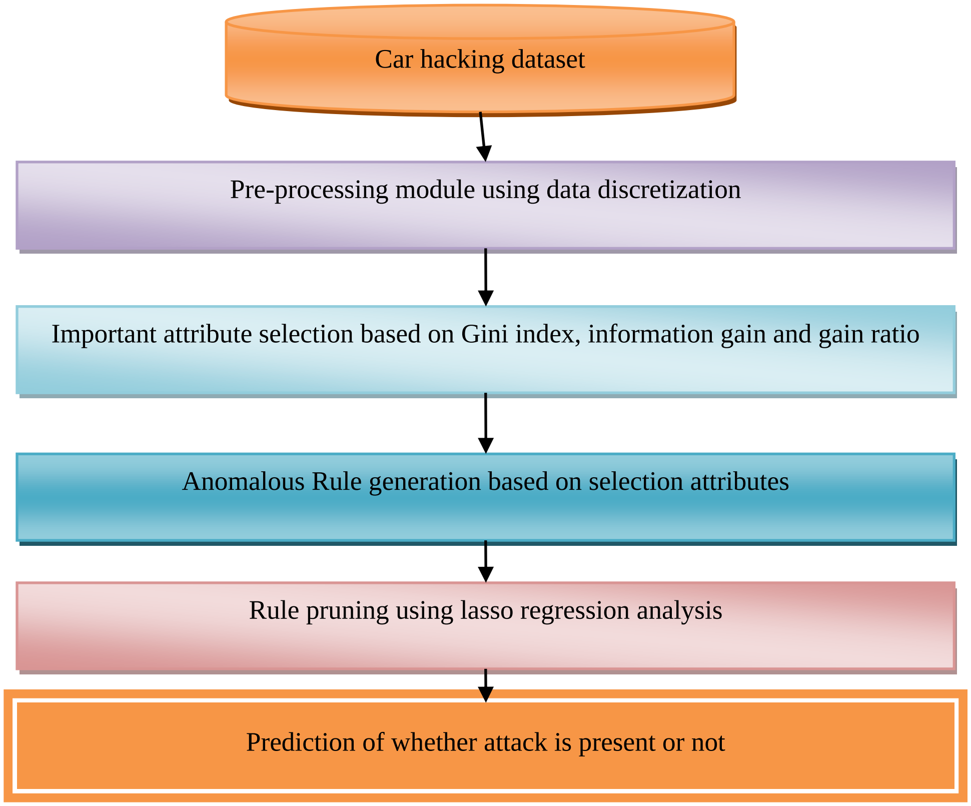
Car hacking dataset
Pre-processing module using data discretization
Important attribute selection based on Gini index, information gain and gain ratio
Anomalous Rule generation based on selection attributes
Rule pruning using lasso regression analysis
Prediction of whether attack is present or not

Fig. 2 Processing flow of proposed car hacking detection process

In Fig. 2, the overall processing flow of the proposed research methodology is given. Here, the car hacking dataset will initially be given input to the pre-processing module, where data discretization will be performed. After pre-processing, feature selection is done based on the Gini index, information gain, and gain ratio. Based on selected features, anomalous rule generation is performed. Rule pruning is performed using the lasso regression analysis method to ensure accuracy and reduce computation overhead. Finally, those reduced rule set is given as input to the improved RVM to predict whether an attack is present or not. The following sub-sections describe the proposed research methodology in detail:

4.1. Dataset Details

DoS (Denial-of-Service) assaults, fuzzy attacks, spoofing driving gears, and RPM gauges are amongst available car-hacking information in datasets created by capturing data from CANs of automobilesusing the OBD-II port while performing message injection attacks [25]. Every dataset has 300 message injection intrusions. Every intrusion lasted 3 to 5 seconds, with 30 to 40 minutes of CANs communications in every dataset.

1. DoS Attack: Every 0.3 milliseconds, messages with the CANs ID '0000' are injected. The number '0000' is the most common.

DoS attacks bringsystems or networks down, making them inaccessible to users. They work by loading target machines with continuous traffic or probes resulting in system crashes. Thus, they prevent the system's legitimate users from accessingservices or resources required by them. These attacks are executed against sites of high-profile organizations, including media, Government sites, and trading commerce sites. They predominantly attack the web servers of these sites. Though these attacks do not cause thefts or information losses, they cost organizations time and money.

2. Fuzzy Attack: Every 0.5 milliseconds, inject messages with completely random CANs ID and DATA values.

Fuzzing isa technique for automatically detecting flaws. Fuzzing applications are stressing them out and causing unexpected behaviours, resource leakages, and crashes. The procedure entails feeding computers with incorrect, unexpected, or randomized data. The processes are repeated until vulnerabilities are discovered and resulting environments are monitored. Fuzzing assaults occur when threat actors utilise fuzzing to identify zero-day exploits.

On the other hand, security experts use fuzzing approaches to test the security and stability of systems. Random messages can be injected if an attacker does not have enough information about the values of CANs IDs and the data field. Random warnings appeared on the cluster when we sent randomly generated data of various arbitrary IDs, and it rarely caused an immediate stop or flinching of the steering wheel.

3. Spoofing Attack (RPM/gear): Every 1 millisecond, messages with a certain CANs ID are linked to RPM/gear information.

Tachometers (also known as revolution counters, tachometers, rev counters, or RPM gauges) measure speeds of rotations in shafts or discs, such as motors. RPMs are typically shown on calibrated analogue dials, but digital displays are more common.On reversing vehicle traffics, attackers inject messages of CANs to control their desired functions. In the first stages, the following spoofing assaults were carried out: 1) factory mode activation warning, 2) printing higher RPM gauge, 3) engine off alert, 4) blind spot collision warning, and 5) rapid turn-on of the back camera screen. A1), 2), were added to the training set, and various CANs IDs along with3), 4), and 5) were used in the test set to improve detection complexity.

4. Gear Attack

The target vehicle was raised and held immobile for safety reasons. Apart from the drive or rear gear, participants can adjust the vehicle's settings (park or neutral gear, changing the steering wheel, opening/closing doors). Because the participants remotely accessed the attack PC due to the pandemic, staff modified the settings at the site if required. The participants were able to run the pre-submitted attack scripts to transmit attack messages to the vehicle after downloading them into the PC. Kvaser'sCANs to USB interface device was used to link the PC to the vehicle. The interface has two channels of CANs, the first of which was connected to the vehicle's C-CANs.

4.2.Dataset Attributes

Dataset attributes are listed below:

Timestamp, CANs ID, DLC, DATA[0], DATA[1], DATA[2], DATA[3], DATA[4], DATA[5], DATA[6], DATA[7], Flag.

1. Timestamp: recorded time (s)
2. CANs ID: CANs message identifier in HEX (ex. 043f)
3. DLC: number of data bytes, from 0 to 8
4. DATA [0~7] : data value (byte)
5. Flag: T or R, T indicates injected message, whereas R indicates a normal message

In the obtained dataset, numbers of normal and injected messages are depicted in Table 1.

Table 1. Dataset Overview

Type of Attack	Counts of messages	Counts of normal messages	Counts of injected messages
DoS attacks	3,665,771	3,078,250	587,521
Fuzzy attacks	3,838,860	3,347,013	491,847
Spoofing drive gear attacks	4,443,142	3,845,890	597,252
Spoofing RPM gauzes	4,621,702	3,966,805	654,897
GIDS: Attack-free (normal)	988,987	988,872	-

4.3.Dataset Preprocessing

Minimizing the item count is the main target of pre-processing concept for examining the association rule extraction algorithm for considering the items that create the local knowledge of the database. Here, the pre-processing concept uses the clustering algorithm to recognize the local knowledge and put similar knowledge together.The experimental analysis states that the proposed approach can recognize more frequent itemsets when distinguished with a random partition and the original data, creating new rules to be explored that would not be extracted in the other cases.

$$Sim\left(T_x, T_y\right) = \frac{CountT\ (T_x \cap T_y)}{CountT\ (T_x \cup T_y)} (1)$$

Consider the following vehicle communications from database

- If Timestamp = 30 sec and DLC = 7 then Flag = R
- If Timestamp = 15 sec and CANs ID = 14 and DATA [0] – 16 then Flag = T
- If Timestamp = 37 sec and Data [7] = 18 then Flag = T
- If DLC = 5 and DATA [4] = 15 then Flag = R
- If DATA [7] = 21and DATA [4] = 14 and DLC = 6 then Flag=R

Here transactions 1, 4 and 5 comes under $T_x \cap T_y$

The remaining transaction comes under $T_x \cup T_y$

$$Sim\left(T_x, T_y\right) = \frac{CountT\ (T_x \cap T_y)}{CountT\ (T_x \cup T_y)} \qquad (2)$$

$$Sim\left(T_x, T_y\right) = \frac{3}{2} = 1.05 \qquad (3)$$

4.4.Attribute Selection

Selecting the best attributes from the entire dataset to improve classification effectiveness is defined as attribute selection. It helps to reduce the items set which have been created. The chosen attributes were preceded independently upon two measures in this work. For rule generation, the attributes with maximum information gain, increased gain ratio, and minimum Gini index was chosen. The Gini index measures the inequality amongst values in a frequency distribution. It is guesstimated as

$$Gini = 1 - \sum_i p(i)^2 \qquad (4)$$

Here p(i) is known as the attributes, and i is the number of attributes.

The volume of information is known as information gain, which is achieved by identifying the value of the feature, which is the entropy of the distribution earlier than the split subtracting the entropy of the distribution following it. The least entropy is equal to the biggest information gain. So, let us consider X as the collection of entire attributes and $X_t$is called as the collection of all training examples, value(x, p (i)) with x ∈ XIt explains the value of a particular example x for attribute p(i) ∈ X, H states the entropy. The values (p (i)) function indicates the collection of all possible values of attribute X. The information gained for an attribute p(i) ∈ X is like this:

$$IG\big(Xt, p(i)\big) = H(Xt) - \sum\left(\frac{\big|\{x \in Xt \big| value\big(x, p(i)\big) = v\}\big|}{|Xt|} \cdot H(\{x \in Xt \big| value\big(x, p(i)\big) = v\})\right) \qquad (5)$$

It is assumed that the result of the number of tuples concerning the overall number of tuples in D for every output. The gain ratio is determined as:

$$Gain\ Ratio\ (A) = \frac{Gain\ (A)}{Split\ Info\ (A)} \qquad (6)$$

Where

$$Split\ Info_A(D) = -\sum_{j=1}^{v}\left(\frac{|D_j|}{|D|}\right) log_2\left(\frac{|D_j|}{|D|}\right) \qquad (7)$$

This value gave the information computed by dividing the dataset D, and it is segregated into v divisions based on v outputs which result from tests on attributes A.

The calculated Gini index, information gain, and gain ratio values are shown in Tables 2 and 3.

Table 2. Information gain, gain ratio, and Gini index values for dos and fuzzy dataset

Attribute	**DoS dataset**			**Fuzzy dataset**		
	Information gains	**Gain ratios**	**Gini indices**	**Information gains**	**Gain ratios**	**Gini indices**
Time stamp	0.0486	0.0001	0.4675	0.0225	0.0668	0.9443
CANs ID	0.0464	0.6982	NaN	0.0264	0.0305	0.9190
DLC	-0.0020	0.0965	0.4660	0.0035	0.0193	0.9166
DATA [0]	0.0448	0.3784	NaN	0.0297	0.0526	0.8784
DATA [1]	0.0481	0.2973	NaN	0.0273	0.0586	0.8592
DATA [2]	0.0348	0.3107	NaN	0.0243	0.0859	0.8168
DATA [3]	0.0421	0.3167	NaN	0.0258	0.0485	0.8665
DATA [4]	0.0170	0.2186	NaN	0.0106	0.0518	0.8138
DATA [5]	0.0294	0.2949	NaN	0.0140	0.0192	0.8659
DATA [6]	0.0446	0.2084	NaN	0.0382	0.1012	0.7413

Table 3. values of gear and rpm datasets information gains, gain ratios, and Gini indices

Attribute	**Gear Dataset**			**RPM Dataset**		
	Information gain	**Gain ratio**	**Gini index**	**Information gain**	**Gain ratio**	**Gini index**
Time stamp	0.0237	0.0005	0.6740	0.0270	0.0033	0.6672
CANs ID	0.0252	0.1015	0.6567	0.0426	0.1060	0.6316
DLC	0.0020	0.0604	0.6740	0.0036	0.0682	0.6650
DATA [0]	0.0292	0.3957	0.7991	0.0417	0.2157	0.5624
DATA [1]	0.0291	0.2558	0.4863	0.0503	0.2612	0.5007
DATA [2]	0.0257	0.3811	0.4966	0.0415	0.3178	0.4231
DATA [3]	0.0279	0.2271	0.4827	0.0475	0.2435	0.4607
DATA [4]	0.0106	0.3470	0.3577	0.0182	0.3433	0.7652
DATA [5]	0.0305	0.1571	NaN	0.0421	0.2912	0.8456
DATA [6]	0.0325	0.1322	NaN	0.0545	0.1099	NaN

The numbers of selected features based on these values are listed in Table 4.

Table 4. Selected features

	DoS database		**Fuzzy Database**		**Gear Database**		**RPM Database**	
Number of selected features	**8**		**8**		**8**		**8**	
Selected features	3	DLC	8	DATA [5]	8	DATA [4]	6	DATA [2]
	2	CANs ID	6	DATA [2]	7	DATA [3]	7	DATA [3]
	4	DATA [0]	5	DATA [1]	5	DATA [1]	4	DATA [0]
	6	DATA [2]	9	DATA [5]	6	DATA [2]	2	CANs ID
	7	DATA [3]	7	DATA [3]	2	CANs ID	3	DLC
	8	DATA [4]	3	DLC	1	Timestamp	1	Timestamp
	9	DATA [5]	2	CANs ID	3	DLC	8	DATA [4]
	10	DATA [6]	1	Timestamp	4	DATA [0]	9	DATA [5]

The proposed feature selection process will select the features specified in table 2. The major objective of the work is to attain optimally selecting features with the concern of Gini index, information gain, and gain ratio parameters, whose resultant outcome is shown in table 2.

4.5. Rule Generation

While the discretization procedure into intervals was generated for arithmetical data, the critical problem is that it is either more or below the estimated boundary values. The fuzzy set will rectify that the natural limits of attributes are considered while indicating the attributes, getting a much-attuned description to actuality. The approach employed is by anomalous fuzzy rules that are the same as those taken by fuzzy exceptions. They werebuilt on the following approaches for mining crisp anomalous rules, which were created using a formal model for fuzzy rules.:

1. Excerpt anomalous rules in transactions that fulfil the rule's antecedent.
2. The referenced rule is revised to a more restrictive rule since support for a common-sense rule that isn't a desired behaviour and discomfort to rule consistency has harmed trust in the genuine reference rule.
3. The certainty factor is utilised instead of the confidence factor. The number of acquired rules is reduced due to this certainty factor, and the gained rules are more consistent.

The study confirms that increasing $Conf(X \wedge Y \rightarrow \sim A)$ is superior as $Supp(X \rightarrow Y)$ rises. It results in support that the reference rule criterion works according to the subsequent supports $Supp(X \rightarrow Y) = supp(X\ U\ Y)$ and $supp(X\ U\ Y\ U\ A)$.

Here, 100 rules are generated for all 4 data in car hacking datasets.

4.6. Rule Pruning

For pruning or grouping rules, we use various methods, for example, cover methods, rule structure cover approaches, rule clustering, etc. One more means of selecting the association rules works according to the interestingness measures, for instance, support, confidence, correlation, etc. Here, the lasso regression analysis method helps in the rule pruning process. The Lasso regression method is utilized for the rule pruning process in this work. The processing steps of lasso regression analysis are defined below:

1. Initialize t// t free specified parameter
2. Construct matrix using attributes in the rules
3. Find scalar means of input training attributes and store them in x
4. Find the scalar mean of predictable attributes and store it in y
5. Find the variance

$$y_i - 0 - x_i T\beta = y_i - y - xT - x_i T\beta = y_i - y - x_i - xT (8)$$

6. Convert this equation into the Lagrangian form

$$min\beta \in RP1N\|y - X\beta\|22 + \lambda\|\|1 \qquad (9)$$

7. Adjust the tuning parameters λ
8. When $\lambda = 0$,then we have OLS regression
9. Constraint on attributes in the rule that shrinks coefficients towards zero
10. Identifies the most important attributes associated with the response variable

The number of rules pruned for the input datasets is shown in Table 5.

Table 5. Rule Pruning outcome

Factors	**DoS dataset**	**Fuzzy dataset**	**Gear dataset**	**RPM dataset**
Number of rules generated	100	100	100	100
Number of rules pruned	28	42	44	32
Remaining rules	72	58	56	68

4.7. Improved Relevance Vector Machine (IRVM) Based Prediction

Bayesian inference was utilized by RVMs (Relevance Vector Machines), which are machine-based learning approaches inmathematics and obtain parsimonious solutions for regression/probabilistic classification as suggested in [24]. Similar functional forms were there in RVMs to SVMs and provided probabilistic classifications. Gaussian parameters and IRVMs executed intrusion and categorizing of features. The objective of the modified RVMs wasto categorize intrusion kinds for provided car hacking datasets with better accuracy. RVM classifiers help in matching many relevant attributes.IRVM maximizes the classification accuracy through kernel function. By computing the maximum likelihood samples, it provides superior accuracy in IDSs.

The given training inputs ${\{p_i, t_i\}^n}_{i=1}$ where $pi \in R^n, t_i \in \{0, 1\}$ and n are the counts of samples. RVMs predictfor new inputs $\hat{p}$Based on functions like SVMs, they take the shapes of linear mixtures of basic functions that logistic sigmoid functions have converted.

$$q(\hat{p}, w) = \sigma(\sum_{i=1}^{n} \omega_i k(p_i, \hat{p})) = \sigma(w^T K) \qquad (10)$$

Where $k(\hat{p}) = [k(p_i, \hat{p}) \dots (p_n, \hat{p})]^T$is the kernel function vector, $w = (\omega_1 \dots \omega_n)^T$ is the weight vector, $\sigma(.)$ is the logistic sigmoid function defined by,

$$\sigma(a) = \frac{1}{1+exp\,(-a)} \qquad (11)$$

The logistic sigmoid function assures the given below symmetry property.

$$\sigma(-a) = 1 - \sigma(a) \qquad (12)$$

Consequently, the RVM scheme can be used as the posterior probability. For the input $\hat{p}$, the class c_1 posterior probability can be defined as,

$$Pr(t = 1|\hat{p}) = q(\hat{p}, w) \qquad (13)$$

Correspondingly, the class c_2 posterior probability can be defined as,

$$Pr(t = 0|\hat{p}) = 1 - p(\hat{p}, w) \qquad (14)$$

Since it uses a Bayesian probabilistic framework to train the system, the RVM model may be considered the posterior probability. The use of ARDs (Automatic Relevance Determinations) over the weight vectors$w$, with a different hyper parameter α_i for each of the weight parameters ω_iIt is a crucial aspect of RVM. Many hyper parameters are defined to high values by the inference technique, parameters that effectively drive associated weights to zero. Consequently, the kernel functions that relate to them can be pruned out, leading to a sparse model. Relevance vectors are the remaining nonzero weights of the inputs p_i.

Decisions of RVMsare pre-defined by Equation (10) and can be rewritten with only $w_{MP}$and RVMs as shown below for an input vector $\hat{p}$.

$$q(\hat{p}, w_{MP}) = \sigma\left(\sum_{p_i \in RVM}^{n} w_i k(p_i, \hat{p}) + \omega_0\right) \qquad (15)$$

Kernel function plays a significant part in the RVM decision model, as shown in Equations (10) and (15). For example, linear, polynomial, sigmoid, Gaussian Radial Basis Function (RBF), etc., are all typical kernel functions for selection. The Elliptical Radial Basis Function (ERBF) is employed for kernel function in this improved RVM.

$$(p, z) = exp(-\sum_{i=1}^{D} (p_i - z_i)^2 / (\sigma_i^2, r^2)) \qquad (16)$$

Where p and z are D-dimension feature vectors $(i.e. = (p_1, \dots\dots, p_D)^T$, $z = (z_1, \dots\dots, z_D)^T$, ris a scale factor, σ_i^2variance.

5. Results and Discussion

The suggested LR-IADS methodology is implemented in the MATLAB simulation platform for a given car hacking dataset. The research objective is to analyse the applicability of proposed LR_IADS in different real-time applications. In this work, the applicability of LR-IADS is tested on the car hacking detection process. Here performance analysis of the proposed research work is done concerningthe accuracy, precision, recall, f-measure, and Error rate. The numerical values obtained for the different car hacking datasets are shown in Table 6.

Table 6. Performancemetrics

Factor	DoS-Dataset	Fuzzy dataset	Gear dataset	RPM dataset
Accuracy	97.5	98.5	94.5	98.5
Precision	97.6263	92.1053	92.7632	98.4962
Recall	97.3822	99.1848	95.9259	98.1315
F-Measure	97.4818	95.3033	93.9757	98.3104
Error rate	2.5	1.5	5.5	1.5

The above table illustrates the numerical values obtained for the different datasets. From these numerical values, it can be learned that the LR-IADS can detect car hacking attacks efficiently. In particular, fuzzy datasets and RPM datasets can be learned and predicted with an increased accuracy level. In the following figures, graphical illustrations of the parameter values for the particular attacks are shown.

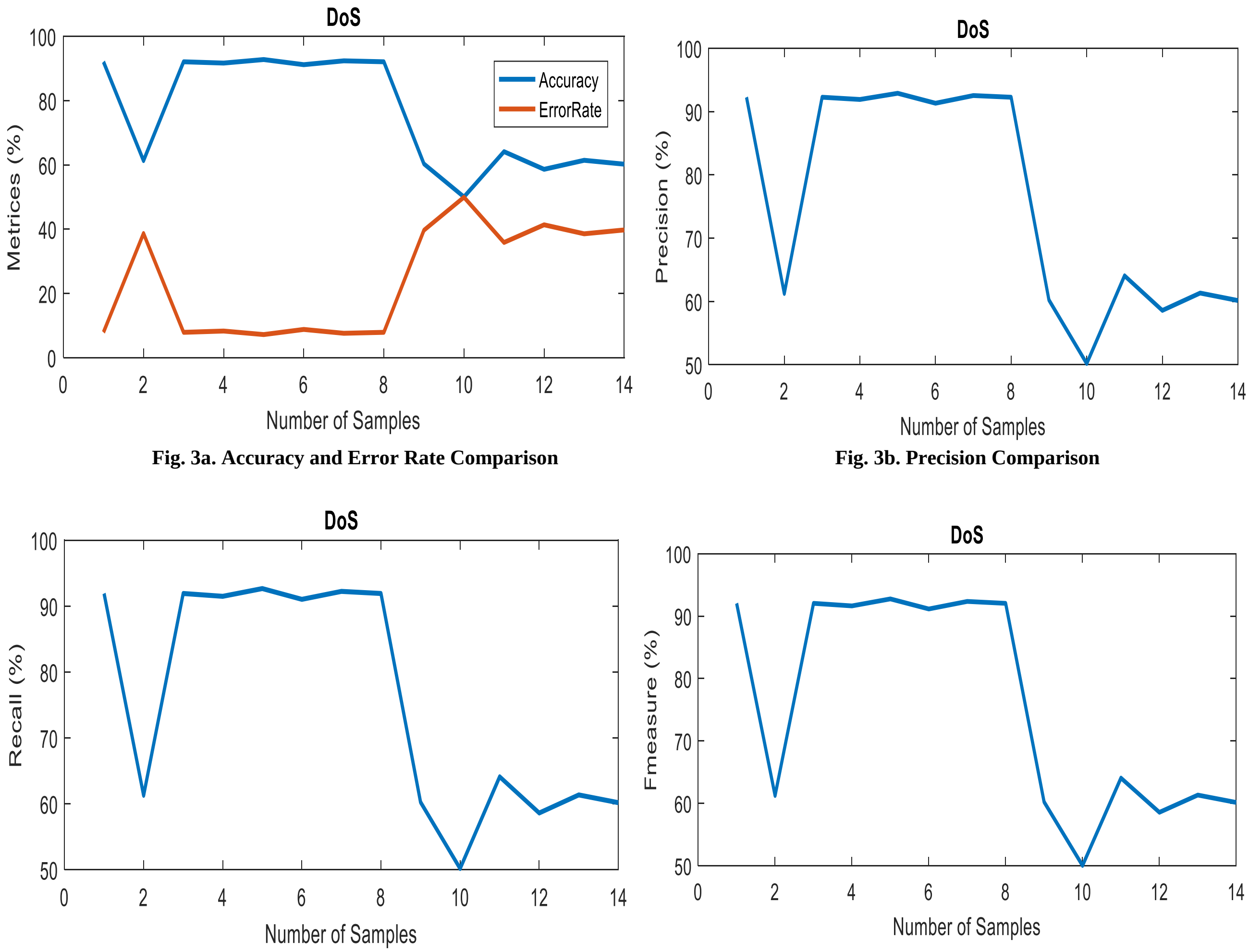

Fig. 3a. Accuracy and Error Rate Comparison

Fig. 3b. Precision Comparison

Fig. 3c. Recall Comparison

Fig. 3d.F-Measure Comparison

Fig. 3 DoS Attack Detection Process

In Fig.3, a comparison analysis of the proposed research methodology for the DoS attack dataset is given. The x-axis number of samples is shown, and y-axis metrices are depicted. From Fig. 3a, it is evident that the suggested method ensures accurate detection of DoS in the car hacking system with a lesser error rate. And also, it is observed that the proposed method providesan increased accuracy level for the increased number of data. Approximately proposed method LR-IADS attains a 97% accuracy rate of DoS attack detection in the car hacking dataset. Likewise,Fig.3b, 3c, and 3d illustrate the performance improvement of the proposed methodology against the existing technique for precision, recall, and f-measure.

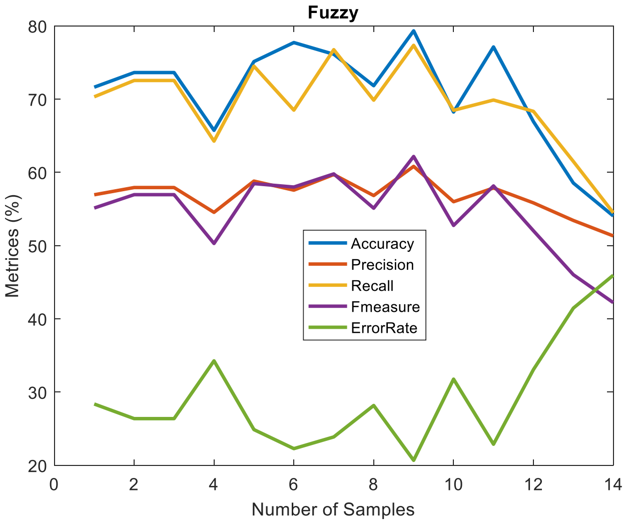

Fig. 4 Fuzzy attack detection process

In Fig. 4, a comparison analysis of the proposed research methodology for the Fuzzy attack dataset is given. The x-axis number of samples is shown, and y-axis metrices are depicted. It is evident from analyses that the suggested method ensures accurate detection of fuzzy attacks in the car hacking system with a lesser error rate. And also, it is observed that the proposed method providesan increased accuracy level for the increased number of data. Approximately proposed method LR-IADS attains a 98.5% accuracy rate of fuzzy attack detection in the car hacking dataset. Compared to other datasets, LR-IADS attainsa lesser error rate fuzzy attack dataset at 1.5.

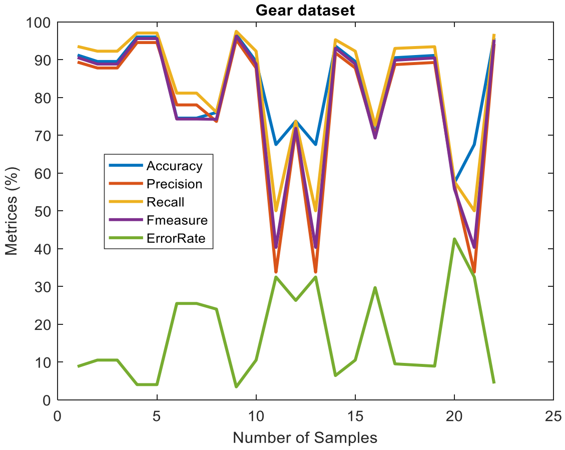

Fig. 5 Gear attack detection process

In Fig. 5, a comparison analysis of the proposed research methodology for the gear attack dataset is given. The x-axis number of samples is shown, and y-axis metrices are depicted. It is evident from analyses that the suggested method ensures accurate detection of gear attacks in the car hacking system with a lesser error rate. And also, it is observed that the proposed method providesan increased accuracy level for the increased number of data. Approximately proposed method LR-IADS attains a 94.5% accuracy rate of fuzzy attack detection in the car hacking dataset. But this is comparatively lesser than other datasets.

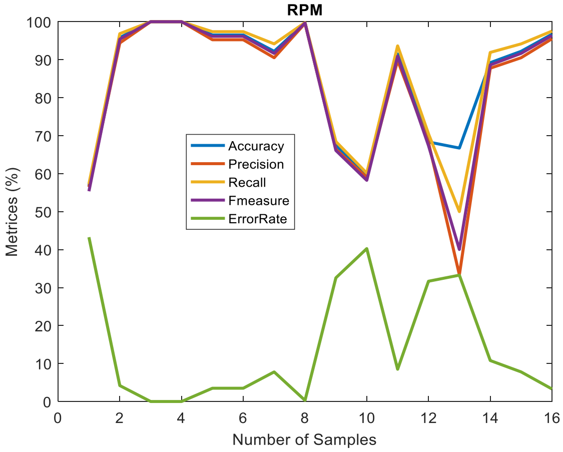

Fig. 6 RPM attack detection process

In Fig. 6, a comparison analysis of the proposed research methodology for the rpm attack dataset is given. The x-axis number of samples is shown, and y-axis metrices are depicted. It is evident from analyses that the suggested method ensures accurate detection of rpm attacks in the car hacking system with a lesser error rate. And also, it is observed that the proposed method providesan increased accuracy level for the increased number of data. Approximately proposed method LR-IADS attains a 94.5% accuracy rate of fuzzy attack detection in the car hacking dataset.

From all the analyses given above, LR-IADS seems to work better on fuzzy and RPM datasets by ensuring the highest accuracy level of 98.5% and a lesser error rate of 1.5.

6. Observation from the Simulation

Our proposed LR-IADS methodology is implemented to detect fraudulent transactions in the credit card dataset. The performance analysis has been carried out on different datasets such as soil, auto mpg, and credit card datasets. Simulation performed on the different datasetsensures accurate fraudulent transaction detection.

This research attempted to apply LR-IADS's proposed methodologyto the car hacking detection dataset to analyse its efficiency. As expected, LR-IADS ensuresan accurate detection rate with a lesser error rate. Our proposed methodology can detect various attacks like DoS, fuzzy, gear attack, and RPM attack to ensure accurate car hacking detection.

7. Conclusion

In this research work, we have utilized the LR-IADS to predict whether the car hacking happened. Here, anomalous fuzzy rules were created initially with the help of attributes chosen from the database. Based on the Gini index, information gain, and gain ratio, we choose the attributes here. The lasso regression analysis method helps to do the rule pruning on the generated anomalous rules. At last, unexpected suspicious detection is done according to these anomalous rules by commencing the classification process, and it is performed by the Improved Relevant Vector Machine based Association Classifier (IRVM). This research work is implemented on the car hacking database for the IDSs gathered from the controller area network. The overall analysis of work is carried out on the MatLab simulation platform. It is indicated that the suggested LR-IADS methodology can ensure an accurate car hacking detection outcome.

Conflicts of Interest

There are no conflicts of interest declared by the authors.

Research involving Human Participants and Animals

No animals or humans are involved in our work.

Informed consent

Not applicable as no human or animal sample was involved in this study.

References

[1] B. Fleming, Advances in Automotive Electronics [Automotive Electronics], IEEE Vehicular Technology Magazine. 10(4) (2015) 4-96.

[2] B. Leiding, and W. V. Vorobev, Enabling the V2X Economy Revolution Using a Blockchain-based Value Transaction Layer for Vehicular Ad-hoc Networks, In MCIS. (2018) 1-33.

[3] J. Contreras-Castillo, S. Zeadally, and J. A. Guerrero-Ibañez, Internet of Vehicles: Architecture, Protocols, and Security, IEEE Internet of Things Journal. 5(5) (2017) 3701-3709.

[4] B. Groza, and P. S. Murvay, Security Solutions for the Controller Area Network: Bringing Authentication to In-Vehicle Networks, IEEE Vehicular Technology Magazine. 13(1) (2018) 40-47.

[5] O. Avatefipour, and H. Malik, State-of-the-art Survey on In-Vehicle Network Communication CAN-Bus Security and Vulnerabilities, arXiv preprint arXiv:1802.01725. (2018).

[6] Y. H. Chou, T. H. Chu, S. Y. Kuo, and C. Y. Chen, An Adaptive Emergency Broadcast Strategy for Vehicular Ad Hoc Networks, IEEE Sensors Journal. 18(12) (2017) 4814-4821.

[7] P. Carsten, T. R. Andel, M. Yampolskiy, and J. T. McDonald, In-Vehicle Networks: Attacks, Vulnerabilities, and Proposed Solutions, In Proceedings of the 10th Annual Cyber and Information Security Research Conference. (2015) 1-8.

[8] L. B. Othmane, L. Dhulipala, M. Abdelkhalek, N. Multari, and M. Govindarasu, On the Performance of Detecting Injection of Fabricated Messages into the, CAN Bus, IEEE Transactions on Dependable and Secure Computing. (2020) 1-1.

[9] M. Banerjee, J. Lee, and K. K. R. Choo, A Blockchain Future for the Internet of Things Security: A Position Paper, Digital Communications and Networks. 4(3) (2018) 149-160.

[10] R. C. Staudemeyer, H. C. Pöhls, and M. Wójcik, The Road to Privacy in IoT: Beyond Encryption and Signatures, Towards Unobservable Communication, In IEEE 19th International Symposium on" A World of Wireless, Mobile and Multimedia Networks (WoWMoM). (2018) 14-20.

[11] E. Vasilomanolakis, M. Krügl, C. G. Cordero, M. Mühlhäuser, and M. Fischer, Skipmon: A Locality-Aware Collaborative Intrusion Detection System, In IEEE 34th International Performance Computing and Communications Conference (IPCCC). (2015) 1-8.

[12] L. Nishani, and M. Biba, Machine Learning for Intrusion Detection in MANET: A State-of-the-Art Survey, Journal of Intelligent Information Systems. 46(2) (2016) 391-407.

[13] J. H. Seo, Detection of Car Hacking Using One-Class Classifier, Journal of the Korea Convergence Society. 9(6) (2018) 33-38.

[14] A. Taylor, S. Leblanc, and N. Japkowicz, Anomaly Detection in Automobile Control Network Data with Long Short-Term Memory Networks, In IEEE International Conference on Data Science and Advanced Analytics (DSAA). (2016) 130-139.

[15] Y. Jaoudi, C. Yakopcic, and T. Taha, Conversion of an Unsupervised Anomaly Detection System to Spiking Neural Network for Car Hacking Identification, In 11th International Green and Sustainable Computing Workshops (IGSC). (2020) 1-4.

[16] M. L. Han, J. Lee, A. R. Kang, S. Kang, J. K. Park, and H. K. Kim,A Statistical-Based Anomaly Detection Method for Connected Cars in Internet of Things Environment, In International Conference on Internet of Vehicles. (2015) 89-97.

[17] H. Kang, B. I. Kwak, Y. H. Lee, H. Lee, H. Lee, and H. K. Kim, Car Hacking and Defense Competition on In-Vehicle Network, In Workshop on Automotive and Autonomous Vehicle Security AutoSec. (2021) 1-25.

[18] F. Martinelli, F. Mercaldo, V. Nardone, and A. Santone, Car Hacking Identification through Fuzzy Logic Algorithms, In IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). (2017) 1-7.

[19] H. Li, L. Zhao, M. Juliato, S. Ahmed, M. R. Sastry, and L. L. Yang, Poster: Intrusion Detection System for in-Vehicle Networks Using Sensor Correlation and Integration, In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. (2017) 2531-2533.

[20] V.S. Barletta, D. Caivano, A. Nannavecchia, and M. Scalera, Intrusion Detection for In-Vehicle Communication Networks: An Unsupervised Kohonen SOM Approach, Future Internet. 12(7) (2020) 119.

[21] A. Taylor, N. Japkowicz, and S. Leblanc, Frequency-Based Anomaly Detection for the Automotive CAN Bus, In World Congress on Industrial Control Systems Security (WCICSS). (2015) 45-49.

[22] A. Taylor, Anomaly-Based Detection of Malicious Activity in In-Vehicle Networks Doctoral Dissertation, Universitéd' Ottawa/University of Ottawa. (2017).

[23] M. W. Spicer, Intrusion Detection System for Electronic Communication Buses: A New Approach Doctoral Dissertation, Virginia Tech. (2018).

[24] J. Jiang, M. Li, X. Jing, and B. Lv, Research on the Performance of Relevance Vector Machine for Regression and Classification, In IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). (2015) 758-762.

[25] [Online]. Available: https://ocslab.hksecurity.net/Datasets/CAN-intrusion-dataset