The Impact of Different Feature Scaling Methods on Intrusion Detection for in-Vehicle Controller Area Network (CAN)

Siti-Farhana Lokman[1(✉)], Abu Talib Othman[1], Muhamad Husaini Abu Bakar[1], and Shahrulniza Musa[2]

[1] System Engineering and Energy Laboratory, Universiti Kuala Lumpur, Malaysian-Spanish Institute, Kulim, Malaysia
farhana.lokman@s.unikl.edu.my
[2] Universiti Kuala Lumpur, Malaysian Institute of Information Technology, Kuala Lumpur, Malaysia

Abstract. Numerous security researchers have a growing interest in the vulnerabilities of the in-vehicle Controller Area Network (CAN) bus system to cyber-attacks. The adversaries can leverage these vulnerabilities in manipulating vehicle functions and harming the drivers' safety. Some security mechanisms proposed for CAN bus in detecting anomalies have favoured over the one-class classification, where it constructs a decision boundary from normal instances. Nevertheless, the accuracy performance of the classifier is highly influenced by the data representation. Judging from this fact, this paper analyses the advantage of utilizing different feature scaling technique as in to obtain higher classification accuracy of the classifier algorithms. To serve this purpose, the CAN bus datasets in this paper are scaled using standardization, min-max, and quantile, and are evaluated using one-class classifier model used in automotive CAN bus. The results exhibit that integrating different feature scaling techniques could greatly enhance the classification accuracy of the classifiers.

Keywords: Anomaly-based detection · Neural network · Controller Area Network · Feature scaling · One-class classification

1 Introduction

Cybersecurity in vehicles will become more essential due to the rising of wireless technology embedded in the vehicle system [1]. Many security researchers have demonstrated the vulnerabilities of vehicles that focused on leveraging CAN bus network which eventually compromises the entire internal vehicle system [2]. As a result, the vehicle could be controlled by attackers to prevent it from functioning in a normal way and finally could harm the safety of the driver. One of the attack techniques called fuzzy attack showed by Koscher et al. is intended to make cyber-physical effects on the vehicle [3]. The authors aimed to take over various Electronic Control Unit (ECUs) by flooding the bus with a combination of CAN bus messages. They discovered that by performing little reversed engineering on how vehicle functions work and randomly fuzzed the CAN bus data could gain legitimate access on the entire ECUs.

M. Anbar et al. (Eds.): ACeS 2019, CCIS 1132, pp. 195–205, 2020.
https://doi.org/10.1007/978-981-15-2693-0_14

With the emerging attacks occurring in the CAN bus network, numerous researchers have proposed intrusion detection mechanisms in vehicles as a last line of defence. One of the promising security approaches is anomaly detection [4]. The inclination towards the anomaly detection approach over other approaches is due to limitations and constraints that the CAN bus system possesses. In contrast to signature-based IDS that operates in the IT desktop domain, the attack signatures are not established and documented publicly in the CAN bus field by scientific researchers of automotive manufacturers [5]. The uncertainty in predicting future and unknown attack demand makes the detection in CAN bus domain difficult. Judging from this face, thus, encourages several scholars to exploited one-class classification methods [6–10]. This method is useful when dealing with CAN bus environment where the only large corpus of normal data is available.

In the case of the CAN bus, the broadcasted CAN bus data from various ECUs are comprised of string and numeric data types and lengths, which make the packet features to have different scales. As a result, the features with larger scales dominate the small ones, thus minimizing the positive impact. Further, it has been discovered that choosing the right pre-processing techniques for data specifically feature scaling [11, 12] will make the gradient descent converge quickly, hence reduce computational cost and improve high classification accuracy performance. Based on these findings, thus encourage feature scaling steps in the model to make the classifier performance-enhanced remarkably.

To the best of our knowledge, comparing different feature scaling methods in analyzing detection rate performance has not yet been examined in the CAN bus environment. Subsequently, in this paper, we explore and analyze three different scaling methods applied on CAN bus data, and study the performance using one-class classification model. We applied standard scaling feature methods that have been widely used in CAN bus domain i.e., min-max and standardization [8, 13–15]. We compared against quantile normalization which has been effectively used to eliminate unwanted technical variation in DNA sequence domain [16, 21]. The advantage of quantile normalization could be a potential feature scaling candidate for CAN bus environment which it can remove any systematic variations in the CAN bus network [17]. We captured CAN data from Toyota Camry in order to accomplish the comparisons. Next, we study the scaled data feature results using one-class classification (OCSVM) model proposed in [8, 10, 18]. The model used to evaluate feature scaling methods in this paper is chosen based on its effectiveness in generalize, able to learn and perform rapid decisions in one-class problem in CAN bus domain.

Finally, this paper is organized into 4 sections; Sect. 1 introduces some related works on algorithms optimization specifically in the classification task. We explained the overall learning procedure in Sect. 2. As well, we described three types of feature scaling methods used to transform CAN bus datasets are presented along with the one-class classifier algorithm, the experimental setup and the CAN bus datasets used in order to validate the pre-processing techniques. Section 3 exhibits the overall detection rate results of three feature scaling techniques on four different types of attacks. Finally, we summarized the conclusions of this paper and presented some future works in Sect. 4.

2 Materials and Methods

This research undergoes six stages. The illustration depicted in Fig. 1 summarizes entire procedure. In this section, we demonstrate different approach performed in each step.

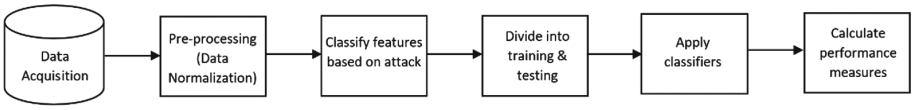

Fig. 1. Flowchart of the entire learning procedure.

2.1 Data Acquisition

In this stage, we obtained raw CAN data from Toyota Camry at low and high-speed driving. The CAN bus data acquisition process is illustrated in Fig. 2. During CAN bus data acquisition, we plugged the cable into OBD-II port which usually located under the steering wheel. The cable is also connected to the CAN bus sniffer hardware in order to communicate with the CAN bus system. We logged the CAN bus data through our laptop nearly 15 min of driving including braking, speeding, idling, and parking.

Fig. 2. Data acquisition setup through OBD-II port of Toyota Camry with CANTact device.

The structure of CAN data features (see Fig. 3) extracted from Toyota Camry are basically encompassed of CAN ID (identifier), timestamp, Cyclic Redundancy Check (CRC), Data field (with the fixed of size from 2 to 8 bytes in this paper) and finally Acknowledge (ACK) field. Nevertheless, the scope of this paper is only focused on the CAN ID and Data field. The CAN ID is used to indicate the vehicle functions e.g., steering, braking. Whereas the CAN Data field encompasses information used by the

specific vehicle function, e.g., rotating steering anticlockwise, pushing and releasing the brake.

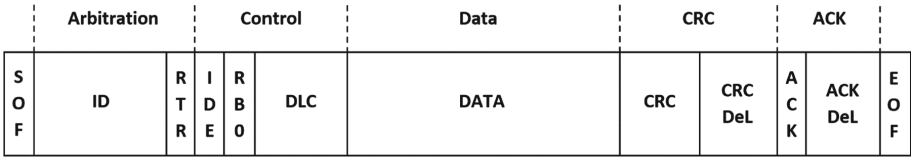

Fig. 3. CAN bus data structure [19]

To have better insight into how CAN bus traffic works, we explored and analyzed CAN data traces logged from a real vehicle. The understanding of the behaviour of CAN data is useful in order to measure how far the variance is spread out and how feature scaling can help in removing unwanted variation that exist in CAN data.

There are 103 distinct CAN IDs were found (for ease of illustration purposes we assign each ID using the numeric value from 0 to 103 shown in Fig. 4) in Toyota Camry. Nearly 2,000,000 data were transmitted (estimated data collected from 30 min of driving) at a fixed period, varying from 4 ms to 10 ms while the vehicle is moving. Figure 4 showed the different occurrences or frequencies of each ID ranging from 20 to 112,000 times throughout 30 min of driving. Meanwhile. Figure 5 exhibited the variation of Data field belong to each ID (for ease of comparison and readability, only several real IDs can be displayed in the Fig. 5). Some IDs like ID 047 and ID 300 have constant values, meaning their Data words value are always the same. However, the rest of the IDs produced multi-value of Data words. Some of the multi-value produced by ID 5AB, ID 110 and ID 609 contains an abundance and unique Data words, whereas ID 121 and ID 30C generate a smaller variation of data.

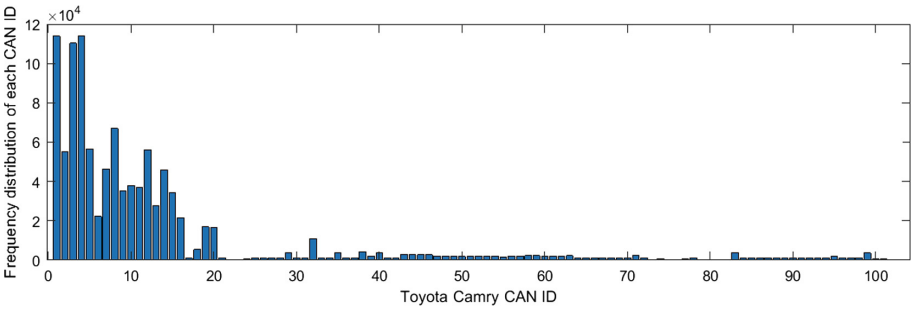

Fig. 4. The distribution of each CAN ID during 30 min of driving

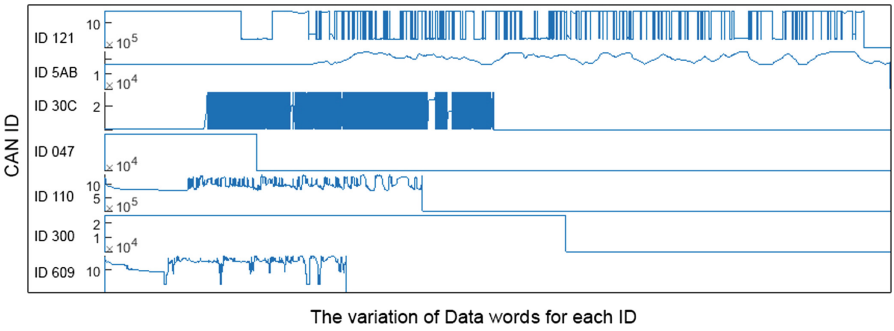

Fig. 5. The variation of Data words distributed from each CAN ID

Judging from this fact, CAN ID and Data words with small variability be swamped by the high variability, and eventually may lead to misclassification. Thus, the feature scaling method is essential to ensure the data is processed in a standard range. Besides, it also can assure the small variation introduced by some IDs due to rarely occurred in CAN bus traffic can still be distinguishable from any types of attack.

2.2 Data Pre-processing

In this stage, we scaled the CAN data using three feature scaling techniques; min-max, standardization and quantile normalization. Besides reducing dominancy impact among data features and improving training convergence, min-max and standardization methods have been commonly used in CAN bus environment as it is simple to be implemented while at the same time provide good classification results in anomaly detection problem [8, 13–15]. However, judging from the nature of CAN data behaviour motivate this study to adopt quantile normalization [20] in the CAN bus environment. Besides the increasing rate of new CAN Data words produced by a single ID that will be increasing in the future, there is also technical variation occurred caused by the signaling and clock drift in the CAN bus traffic [22]. Consequently, it induces the variation of unique CAN Data words that may slightly differ from normal. Hence, the quantile normalization proposed in this paper may be suitable in dealing with noisy environment occurred in the CAN bus in order to ensure unwanted variation can be compared with other normal features.

The necessity of proposed quantile normalization method to be used in this paper can be judged according to the preliminary data analysis such as the construction of the boxplots presented in Fig. 6. Boxplots can be effectively used in revealing similarities and difference of patterns found in the sets of observations [23]. Figure 7 compared the raw and scaled CAN data distribution throughout 30 min of driving. A large number of variations in raw CAN data distribution due to rarely occurred in the bus traffic are indicated as outliers in the boxplot.

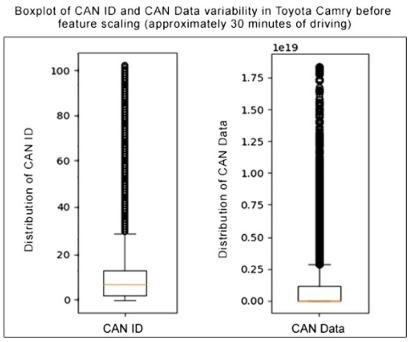

Fig. 6. Boxplots of raw CAN ID and CAN data before feature scaling methods

The figure below presented boxplots of CAN data after scaled into min-max, standardization and quantile normalization method. Based on the observation, there is a significant difference among data after feature scaling methods were applied. Min-max and standardization are very sensitive to the existence of outliers in both features. However, it may not be guaranteed whether the presence of outliers especially in CAN Data features balanced through standardization method. Nonetheless, it is worth examining the effect of this feature scaling method on a classification problem. In addition to the boxplots exhibited below, quantile normalization is robust to the presence of outliers where removing or adding outliers will still yield nearly the same transformation on the data to a defined boundary.

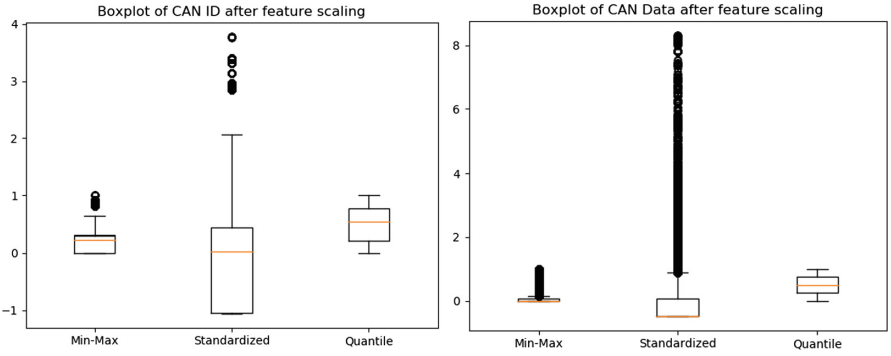

Fig. 7. Boxplots of raw CAN ID and CAN data after feature scaling methods are applied

The formulas for each scaling methods used in this paper are presented in Table 1; where x_i^j denotes as the beginning value of j^{th} feature within the i^{th} sample, whereas σ^j signifies as standard deviation, μ^j represents mean, and finally min^j max^j indicate as minimum and the maximum value of j^{th} feature calculated on all samples.

Table 1. Formulas for different scaling feature methods.

Method	Abbreviation	Formula	Function
Min-Max normalisation	Min-Max	$\frac{x_i^j - min^j}{max^j - min^j}$	Subtracts the minimum from all instances values and makes them scale from minimum to maximum. Then it divides the difference between maximum and minimum of each feature instances
Standardization	Standardized	$\frac{x_i^j - \mu^j}{\sigma^j}$	Eliminates the mean, while at the same time scales the input data into unit variance
Quantile normalisation	Quantile	$inf\{x \in \mathbb{R} : p \geq F(x)\}$	Implements a non-linear transformation where the function of a continuous random variable of each feature is mapped to the same distribution

2.3 Classification Algorithm

In this section, we discussed the one-classification algorithm used in this paper. We choose One-Class Support Vector Machine (OCSVM) [24] to study the impact of different feature scaling methods applied on the Toyota Camry data. OCSVM algorithm is considered in this paper as it has been commonly used in several works found in [8, 10, 18] and therefore offer a good baseline. Moreover, OCSVM is one of the boundary-based methods that utilize a decision function through the optimum separating margin. We employed this one class algorithm on the CAN bus data where only normal data is used during the training. As stated earlier, this type of algorithm is suitable for CAN bus problem where generally only non-anomalous CAN data is available. Thus, OCSVM can utilize its kernel function in making decision functions on complex data in the CAN bus environment. We set the algorithm to use non-linear kernel; Radial Basis Function and optimum settings of hyperparameters based on work in [25].

3 Experiment and Result

3.1 Experimental Data

In this study, we injected the CAN bus traffic with three types of attacks, the explanation of each attack is described as below:

DoS attack: this attack aims to occupy the dominant state in the CAN bus traffic by injecting higher priority of CAN data. In CAN bus domain, the priority of the data payload is determined by the CAN ID. The lower the CAN ID value indicates the higher priority of the packet. As a result, the legitimate data with lower priority ID will

drop. In this case, we continuously inserted new high priority CAN ID and CAN Data every 1 ms with the smallest constant values of '0x000' to preempt legitimate data.

Fuzzy attack: we altered both CAN Data field and CAN ID field with random values. In this case, we randomly choose CAN data in Toyota Camry and altered the ID as well as the content of the data every 3 ms.

Impersonate attack: we spoofed the CAN Data field that is related to gear and RPM with constant values of '0x000' and '0xfff' every 4 ms.

Next, we did labelling on the captured CAN bus data into normal and attack dataset. Table 2 exhibits CAN bus datasets that are used to conduct our experiment. The datasets are divided into training and test dataset.

Table 2. Formulas for different scaling feature methods.

Vehicle	CAN dataset	Attack type	CAN normal data	CAN attack data
Toyota Camry	Training	Normal	900,000	N/A
	Testing	DoS	900,000	223,430
		Impersonate (Gear)	900,000	225,144
		Impersonate (Brake)	900,000	225,046
		Fuzzy	900,000	225,156

3.2 Performance Measure

In this paper, we compared the effects of algorithm's performance on each feature scaling method in Sect. 3.3 using four performance measures; precision, recall and Receiver Operating Characteristics (ROC). The precision and recall used in this work are one of the common basic evaluation metrics in order to measure the robustness and the reliability of the anomaly-based detection model. Precision is a proportion of true positive (anomalies) divided by true positive and false positive. It shows how far the model can predict the positive class. among the samples which are indicated as a true anomaly. Whereby Recall which is known as sensitivity denotes as the proportion of relevant instances detected among the entire number of relevant instances. F-measure summarizes the trade-off between precision and recall using a various number of thresholds. The value of precision, recall and F-measure that approximately to 100% indicates the best performance of the detection model with zero false-negative and zero false-positive rates.

Meanwhile, ROC represents a diagram of a true positive rate against false positive rate based on a various number of thresholds. The ideal percentage of ROC is also 100% detection along with zero false-positive rates.

3.3 Feature Scaling Experiment

The primary focus of this work is to enhance the classification accuracy performance of the classifier algorithms using three different feature scaling techniques. Table 3 shows the overall detection rate results obtained with three feature scaling techniques (we used the abbreviation for feature scaling methods with 'MM' for min-max, 'SS' for

standardization and 'QT' as quantile normalization) using OCSVM algorithm. We also used a naming convention for impersonate attack as 'imp' for brake and gear.

The values highlighted with bold typeface specifies the best classification accuracy results. From the results shown below, some preliminary remarks can be concluded. Among the dataset, at least one feature scaling technique showed better average detection rate performance approximately 100% detection rate on DoS and impersonate attack as compared with others, which in this case, the QT method.

Table 3. The overall detection rates of OCSVM with MM, SS and QT scaling techniques applied on Toyota Camry against DoS, fuzzy and impersonate attacks.

Model	Attack	Scaling	ROC	Precision	Recall	F1_measure	Average detection rate
OCSVM	DoS	SS	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**
		MM	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**
		QT	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**
	Fuzzy	SS	29.00%	88.60%	64.30%	68.70%	62.65%
		MM	57.60%	86.70%	34.50%	34.90%	53.43%
		QT	**89.00%**	**92.80%**	**87.30%**	**88.50%**	**89.40%**
	Imp Break	SS	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**
		MM	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**
		QT	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**
	Imp Gear	SS	89.20%	68.00%	74.30%	71.00%	75.63%
		MM	44.60%	68.00%	74.30%	71.00%	64.48%
		QT	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**

It can be seen that QT normalization achieved the overall highest percentage precision, recall, F-measure and ROC for the majority of attacks after MM and followed by SS. But, its performance slightly degraded when tested on the fuzzy attack. In contrast to MM and SS method, they may perform well when injected with DoS and impersonate attack, however, they performed poorly on impersonate (gear) and fuzzy attack with an average detection rate of nearly 60% and 65%. Generally, we would say that all feature scaling methods are more affected when tested on Fuzzy attack as the randomness in the attack may have similar values like the normal data, which may, in turn, make the model confuse and lead to misclassification. However, all feature scaling techniques showed good performance on DoS and impersonate (brake) attack as these types of attack were injected with constant values that are not similar like normal data. Another important observation, the OCSVM classifier that relies heavily on distance-based calculations are affected on feature scaling, thus scaling the data with an appropriate method like QT helps in eliminating any biases that exist in raw data. Another important observation, QT normalization helps the OCSVM to converge especially when dealing with CAN bus dataset that has larger variability. Finally, the detection rate can still be improved by some tunings of the model hyperparameters. In general, we can conclude that the consequence of pre-processing step such as feature scaling relies on the characteristics of the data features and also the classifier models.

4 Conclusion

This paper examined the detection rate performance of various feature scaling techniques in CAN bus datasets from real vehicle model. When quantile normalization is used, the OCSVM, the distance-based approach is significantly improved especially when the CAN data has a large number of variabilities. Lastly, all feature scaling methods performance degraded when tested on the fuzzy attack since the attack may contain variations that similar to normal data. However, the proposed feature scaling method in this paper, quantile normalization is still robust on different types of attack. quantile normalization can be a potential candidate for pre-treatment method for CAN bus data as it showed its effectiveness in compressing the outliers in normal data and eventually make the unwanted variation comparable with normal features and still can be distinguishable from attack data.

As future work, although quantile normalization presented in this paper showed a good performance, however, we can find and analyze other potential pre-processing techniques like feature selection, feature cleaning and feature conversion to different characteristics and higher dimensional of CAN bus datasets in order to study which method improves the detection rate results. Further, regardless of the higher detection accuracy, the scaling method, as well as the algorithm, is still lacking in terms of attack patterns in automotive CAN bus domain. Thus, increasing the number of fuzzy attacks during experiment makes the performance assessment to be more precise since many possibilities of attack patterns are revealed.

References

1. Sakiz, F., Sen, S.: A survey of attacks and detection mechanisms on intelligent transportation systems: VANETs and IoV. Ad Hoc Netw. **61**, 33–50 (2017)
2. Miller, C., Valasek, C.: Remote exploitation of an unaltered passenger vehicle. Black Hat USA (2015)
3. Koscher, K., et al.: Experimental security analysis of a modern automobile. In: 2010 IEEE Symposium on Security and Privacy (SP), pp. 447–462 (2010)
4. Hoppe, T., Kiltz, S., Dittmann, J.: Security threats to automotive CAN networks – practical examples and selected short-term countermeasures. In: Harrison, M.D., Sujan, M.-A. (eds.) SAFECOMP 2008. LNCS, vol. 5219, pp. 235–248. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87698-4_21
5. Hoppe, T., Kiltz, S., Dittmann, J.: Applying intrusion detection to automotive it-early insights and remaining challenges. J. Inform. Assur. Secur. (JIAS) **4**(6), 226–235 (2009)
6. Martinelli, F., Mercaldo, F., Nardone, V., Santone, A.: Car hacking identification through fuzzy logic algorithms. In: IEEE International Conference on Fuzzy Systems, Naples (2017)
7. Tomlinson, A., Bryans, J., Shaikh, S.A.: Using a one-class compound classifier to detect in-vehicle network attacks. In GECCO 2018 Companion: Genetic and Evolutionary Computation Conference Companion. ACM, Kyoto (2018). https://doi.org/10.1145/3205651.3208223
8. Weber, M., Klug, S., Sax, E., Zimmer, B.: Embedded hybrid anomaly detection for automotive CAN communication (2018)

9. Xing, Y., Lv, C., Wang, H., Cao, D. Recognizing driver braking intention with vehicle data using unsupervised learning methods (2017)
10. Loukas, G., Vuong, T., Heartfield, R., Sakellari, G., Yoon, Y., Gan, D.: Cloud-based cyber-physical intrusion detection for vehicles using deep learning. IEEE Access **6**, 3491–3508 (2017)
11. Nawi, N.M., et al.: The effect of pre-processing techniques and optimal parameters selection on back propagation neural networks. Int. J. Adv. Sci. Eng. Inform. Technol. **7**(3), 770–777 (2017)
12. Kumar, D.A., Venugopalan, S.: The effect of normalization on intrusion detection classifiers (Naïve Bayes and J48). Int. J. Future Revolut. Comput. Sci. Commun. Eng. **3**, 60–64 (2017)
13. Kang, M.J., Kang, J.W.: Intrusion detection system using deep neural network for in-vehicle network security. PLoS One **11**(6), e0155781 (2016)
14. Wasicek, A., Weimerskirch, A.: Recognizing manipulated electronic control units (No. 2015-01-0202). SAE Technical Paper (2015)
15. Taylor, A., Leblanc, S., Japkowicz, N.: Anomaly detection in automobile control network data with long short-term memory networks. In: 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 130–139 (2016)
16. Pan, M., Zhang, J.: Quantile normalization for combining gene-expression da-tasets. Biotechnol. Biotechnol. Equip. **32**(3), 751–758 (2018)
17. Upender, B.P., Dean, A.G.: Variability of CAN network performance. In: Proceedings of the 3rd International CAN Conference ICC (1996)
18. Taylor, A., Japkowicz, N., Leblanc, S.: Frequency-based anomaly detection for the automotive CAN bus. In: 2015 World Congress on Industrial Control Systems Security (WCICSS), pp. 45–49. IEEE (2015)
19. Lokman, S.F., Othman, A.T., Bakar, M.H.A., Razuwan, R.: Stacked sparse autoencoders-based outlier discovery for in-vehicle controller area network (CAN). Int. J. Eng. Technol. **7** (4.33), 375–380 (2018). https://doi.org/10.14419/ijet.v7i4.33.26078
20. Hicks, S.C., Okrah, K., Paulson, J.N., Quackenbush, J., Irizarry, R.A., Bravo, H.C.: Smooth quantile normalization. Biostatistics **19**(2), 185–198 (2017)
21. Hansen, K.D., Irizarry, R.A., Wu, Z.: Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics **13**(2), 204–216 (2012)
22. Monot, A., Navet, N., Bavoux, B.: Impact of clock drifts on CAN frame response time distributions. In: ETFA2011, pp. 1–4. IEEE (2011)
23. Potter, K., Hagen, H., Kerren, A., Dannenmann, P.: Methods for presenting statistical information: the box plot. Vis. Large Unstr. Data Sets **4**, 97–106 (2006)
24. Moya, M.M., Hush, D.R.: Network constraints and multi-objective optimization for one-class classification. Neural Netw. **9**(3), 463–474 (1996)
25. Ghafoori, Z., Erfani, S.M., Rajasegarar, S., Bezdek, J.C., Karunasekera, S., Leckie, C.: Efficient unsupervised parameter estimation for one-class support vector machines. IEEE Trans. Neural Netw. Learn. Syst. **29**(10), 5057–5070 (2018)