A lightweight distributed ELM-based security framework for the internet of vehicles

Aziz Ullah Karimy, Putta Chandrasekhar Reddy
Department of Electronics and Communication Engineering, University College of Engineering, Science and Technology Hyderabad, Hyderabad, India

Article Info

Article history:

Received Mar 9, 2024
Revised Apr 24, 2024
Accepted May 7, 2024

Keywords:

Control area network
Cybersecurity
Distributed processing
Extreme learning machine
IoT
Internet of vehicles

ABSTRACT

The fast growth of internet of vehicles (IoV) has created a new area of connectedness, with promising safety and efficiency in transportation. However, this advancement in vehicle technology has come with significant cybersecurity risks, specifically through control area network (CAN) protocol and other communication techniques within vehicles. This experimental study suggests a machine learning (ML) based security approach based on the extreme learning machine (ELM) algorithm to address these challenges. Unlike customary neural networks, ELM is known for its fast processing, minimal training time, and high accuracy, making it preferably suitable for dynamic IoV environments. The methodology involves data preprocessing, feature selection, and employing ELM for attack classification; the algorithm's performance is evaluated using CAR-Hacking, NSL-KDD, and EdgeIIoT datasets. We also examine the significance of distributed processing to enhance the computational efficiency of the model, obtaining 89% accuracy in 3 ms run-time for external networks, and 83% accuracy with 9 ms run-time for intra-vehical networks. This newly proposed security mechanism using ELM shows very accurate results in detecting intrusions with a high recall rate and reduced computation time through distributed processing.

Corresponding Author:

Aziz Ullah Karimy
Department of Electronics and Communication Engineering, University College of Engineering
Science and Technology Hyderabad
Hyderabad, India
Email: azizullah.karimy91@gmail.com

1. INTRODUCTION

Internet of vehicles (IoV) is a highly sophisticated connected transportation technique utilizing latest technology to transfer real-time data among vehicles, infrastructure, pedestrians, and traffic management systems [1]. The communication among different interior parts is simplified via an intra_vehicle network. The network components are engine control units (ECUs), gateways, sensors, actuators, and other equipment. There are approximately 70 ECUs, enabling exchange of about 2,500 electronic signals across different parts. Each unit is responsible for a specific task and manages a particular component via a universal protocol within vehicle's network [2].

The control area network (CAN) protocol is widely employed standard for intra_vehicle networks in automotive and industrial sectors because of its flexible architecture and cost-effectiveness [3]. It significantly lowers complexity of vehicle's wiring systems, making them more intelligent and using real-time data to make decisions. This lead to improved traffic flow, fewer accidents and an efficient transportation system. However, researchers have highlighted the vulnerabilities in CAN protocol for cybercrime. Hoppe *et al.* [4] proved the exposure of CAN protocol to attacks like frame sniffing and replay; these can compromise functions like window controls, warning systems, and ABS, leading to severe consequences. Woo *et al.* [5] explained about chances of remote attacks via adversarial mobile apps, by manipulating interfaces like telematics, Bluethooh, WiFi of even OBD ports. On some other experimental study, Miller and Valasek [6] compromised infotainment system of Jeep Cherokee, led to significant recall of affected vehicles. Based on a study carried out by Keen Security Lab, pointed out the existing vulnerabilities of remote controlling features like brakes and door locks in Tesla X. They found that, CAN protocol which is core responsible for communication among various parts of the vehicle, is susceptible to denial-of-service (DoS) attack, this can happen when intruder send high-priority messages within short interval of time, making the legitimate node to stop transmitting. Similarly, on other study Nie *et al.* [7] showed about injection attack on CAN bus.

Researchers have been studying advanced ML techniques to mitigate potential threats of IoV networks by enhancing security and privacy. These techniques are well-known for its ability to demonstrate complicated data patterns, for example using multi-layered nonlinear networks. Xiao *et al.* [8] carried out studies on combination of ML techniques with edge computing to examine vehicle traffic, proposing personalized protection insights. Rosay *et al.* [9] designed deep learning (DL) based Intrusion detection systems (IDS) using multi-layer perception, evaluated its performance on automotive microprocessors with CICIDS2017 dataset. Yang *et al.* [10] suggested tree-based stacking algorithm to improve traffic analysis within IoV networks, algorithm showed optimistic results with CICIDS2017 datasets. Mehedi *et al.* [11] proposed PLeNet for intrusion detection in IoV, employing deep transfer learning, which showed promising accuracy. Li *et al.* [12] and Shone *et al.* [13] carried out an experimental study on transfer and unsupervised deep learning techniques to demonstrate effectiveness of IDS in IoV. However, they have still struggling with unbalanced datasets. Xu *et al.* [14] did an innovative experimental study using Log-Cosh variational autoencoder by generating diverse intrusion to enhance detection accuracy. However, despite all these numerous efforts to improve IoV security, it still encounters considerable challenges. Employing DL models needs vast computational resources and training time, which makes these models less feasible for real-time IoV apps, where fast reaction is necessary. On the other hand, IoV networks are quit dynamic, DL models struggle with data distribution and dynamic attack patterns. This results in model obsolescence.

In this experimental study, we propose an extreme learning machine (ELM) based security approach to detect anomalies in intra_vehicle and external vehicle networks. Since ELM is capable of processing large amounts of data with limited processing units and quickly adapt to recognize intrusion patterns in a very accurate and efficient way. Hence, it is a promising security approach for IoV networks. Furthermore, distributed processing is also employed for this study to enhance the detection rate, reduce response times, and develop scalable solutions for handling vast volume of data generated by IoV environments. The IDS based on ELM can be a significant breakthrough for IoV. This approach can use a dynamic and flexible model and continuously monitor network traffic and user behavior for anomaly detection and to avoid security breaches in IoV networks.

2. METHOD

2.1. CAN

CAN is widely used protocol in smart vehicles for communicating with each other and their integrated units for smooth operation. This protocol is exploitable to cyber threats that can be manipulated through different access points, such as spoofing and replay attacks. Intruders can exploit ODB-II connector, and gain access through malicious dongles or connected devices. Manipulation of wiring harnesses is another possible exploitation way to change security systems [5]. Infotainment is a vulnerable gateway for hackers to exploit networks through connected digital devices such as USB, Bluetooth, or Wi-Fi. They can also use software vulnerabilities of ECUs to disrupt vehicle functions or run malicious code remotely or locally [15].

2.2. ELM

After data is preprocessed, ELM is used to classify threats, ELM algorithm has high-speed processing, minimal training time, and commendable accuracy. Unlike traditional neural networks, which require iterative tuning of weights and biases, ELM randomly allocates weights and biases to hidden nodes and calculates output weights analytically. Given a train-set (x_i, t_i), where x_i represents input vector and t_i denotes target vector, ELM process can be succinctly described as follows, as noted in [16]: Initially, input weights w and biases b for the hidden nodes are assigned randomly. Subsequently, the hidden layer output matrix H is computed using an activation function g(.), expressed as H=g (w. x+b). Finally, the output weight matrix β is determined by calculating the pseudoinverse of H and then multiplying it by target matrix T, which can be formulated as $\beta = H^{\sim}. T$. This streamlined approach enables ELM to rapidly train neural networks by analytically determining the weights connecting the hidden layer to output layer, bypassing need for iterative tuning.

2.3. Proposed ELM based framework

IDS are useful method of protecting IoV networks from cybercriminal attacks; they function fast, identifying and mitigating potential intrusions [15]. The framework suggested in this paper to secure intra_vehicles and external networks of IoV by employing IDS as demonstrated in Figure 1. Hence, the suggested IDS could be integrated with intra_vehicle network for detecting anomalies [17] in CAN-Bus; at same time, it is recommended to install IDS into gateways for identifying malicious external attempts compromising vehicles [15].

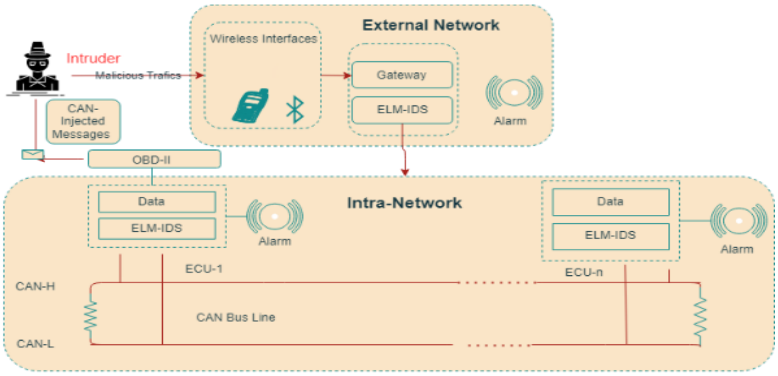


Figure 1. Proposed ELM based architecture

Since CAN-Bus runs on specified communication patterns, IDS can predict the expected behavior of vehicle’s system. With this capability, IDS monitor CAN-Bus traffic in real-time, ensuring communication sticks to the specified patterns [18]. By leveraging ELM algorithm, this framework improves IDS’s ability to distinguish between normal operations and malicious activities, delivering a delicate and quick response mechanism. The overview of proposed approach is illustrated in Figure 2, has the capability of detecting intrusions, including DoS, frame attacks, and subtle timing attacks on specific frames by observing the timestamps of frames on the bus. We examined distributed computing using socket programming to spread computational loads, speed up training and response time, and enhance efficiency of ELM algorithm with the help of server and client architecture [19]. This system is evaluated using three specific datasets, each one of them explained in details in section 3.1, Car-Hacking for intra_vehicle networks, NSL-KDD and EdgeIIoT for external networks to demonstrate a broad spectrum of testing scenarios.

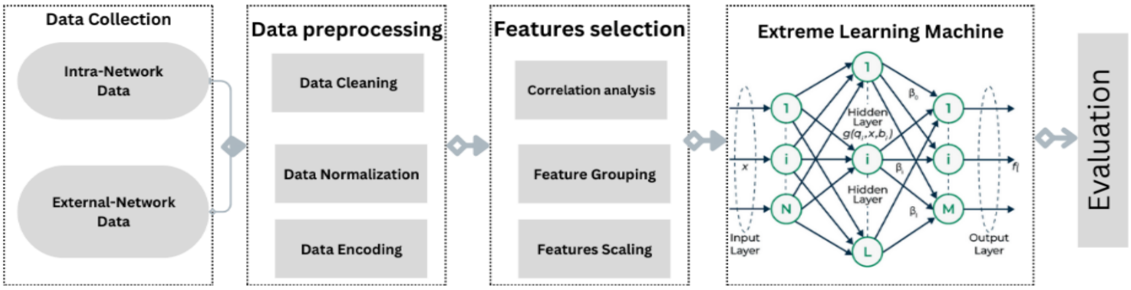


Figure 2. ELM-based IDS framework

3. RESULTS AND DISCUSSION

We implemented this experiment using Scikit-learn, Tensorflow, Scipy libraries, and socket programming for distributed processing with Python programming language. A Toshiba machine with 8 GB RAM, a 750 GB HDD, and an AMD FX-8800P 2.1 GHz processor were the hardware part for the experimental study. Methodology was evaluated using three IoV security related established datasets, the CAR-Hacking, NSL-KDD, and EdgeIIoT. To prevent risks of overfitting and skewed outcomes, and assure the model's robustness, we employed fivefold cross-validation. Accuracy, recall, precision, and F1-score metrics were used to measure the model's performances, and these metrics are calculated as [20].

$$Accuracy = \frac{TP+TN}{TP + TN+FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP + TN}{TP + FN} \tag{3}$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

The terminologies associate with the metric are:

- True positive (TP) = number of samples correctly detected as intrusion instances;
- True negative (TN) = number of samples correctly detected as normal samples;
- False positive (FP) = number of samples incorrectly detected as intrusion instances;
- False negative (FN) = number of samples incorrectly detected as normal samples.

3.1. Datasets

- Car-Hacking: dataset was generated using simulating cyber-attacks on CAN-bus, it contains two key attributes: CAN identifier (ID) and 8-byte data field (DATA [0] -DATA [7]). It consists of four simulated cyber-attacks, DoS, which overloads the network with excessive traffic; Fuzzy for malfunctioning of the system; gear spoofing, which manipulates gear display messages; and RPM spoofing, modifying engine readings [21].
- NSL-KDD: is improved version of KDD'99 dataset, and handles some drawbacks of KDD'99 dataset, such as redundancy and duplicate records. The dataset comprises normal and malicious traffic samples; there are four categories of attacks: DoS, Probe, user to root (U2R), and remote to local (R2L), with 41 incorporated features from various aspects of network connections [22].
- EdgeIIoTset: dataset contains testbed with seven layers; contains information from various IoT devices, including temperature and humidity sensors, ultrasonic sensors, and water level detectors. It covers a wide range of fourteen attacks against IoT and IoT communication protocols divided into five categories of threats, including DoS/DDoS, data collection, injection attacks, malware attacks, and man_in_the_middle attacks [23].

3.2. Data preprocessing

We outlined the specific steps for preprocessing of the data in Algorithm 1. We started methodology with a comprehensive analysis of raw data. To balance the data, SMOTE and random under sampler techniques were utilized. SMOTE is widely used technique for handling imbalanced data [24]; it creates synthetic examples rather than repeating existing samples. The mathematical procedure of SMOTE is explained in the following steps:

If x_i is a minority class sample and x_{zi} is one of its k nearest neighbors, a synthetic sample x_{new} is created as (5):

$$x_{new} = x_i + \lambda\, (x_{zi} - x_i) \tag{5}$$

where λ is a random number between 0 and 1.

We proceed with feature selection, which is essential step for identifying most correlated and significant features, to contribute in classification [25]. We applied f_classif and SelectKBest to isolate a group of highly correlated features with target variables to improve model's predictive accuracy. If F the ratio of between_group variability to within_group variability, is given by (6):

$$F = \frac{between-group\ variability}{within-group\ variability} \tag{6}$$

the exact calculation of these involves means and variances of groups and the overall dataset, which can be detailed as (7):

$$\text{Between\_group variability} = \sum_{i=1}^{n} \frac{n_i.(x_i' - x')^2}{k-1} \tag{7}$$

and the formula for within-group variability is (8):

$$\text{within\_group variability} = \sum_{i=1}^{n} \frac{\sum_{j=1}^{n_i} (x_{ij} - x'_i)^2}{N-k} \qquad (8)$$

where: n_i is the number of observations in group i, x'_i is the mean of group i, x' is the overall mean, k is the number of groups, N is the total number of observations, x_{ij} is the jth observation in the ith group.

Algorithm 1. Data preprocessing
- Validation: ensure absence of infinite or missing values in train-set and test-set.
- One-hot encoding: convert categorical features into numerical values to facilitate computation.
- Label encoding: transform all classes into numerical values, streamlining model training process.
- Column pruning: eliminate unnecessary label columns, retaining only the 'label' column indicative of attack classes.
- Data balancing: employ SMOTE for oversampling and RandomUnderSampler for undersampling to balance the dataset.
- Feature elimination: remove features with zero variation and those deemed redundant or irrelevant.
- Correlation analysis: calculate standard correlation coefficient between attributes and the target to identify predictive features.
- Feature scaling: standardize numerical features in train-set and test-set to have mean of zero and a variance of one.

3.3. Performance analysis

Performance of the proposed approach for external networks of IoV are evaluated on NSL-KDD and EdgeIIoT datasets and for intra-networks are evaluated on Car-Hacking dataset. Furthermore, in this study, we compare the effectiveness of the suggested approach against some existing models, such as SVM, DNN, and RF algorithms. For external networks the proposed approach is tested on NSL-KDD and EdgeIIoT datasets, and comparison of selected ML algorithms applied in this approach are illustrated in Table 1 and Figure 3. Considering the accuracy rate, DNN has achieved highest accuracy of (82.91%) and (96%) in both datasets, ELM with single processing unite and distributed processing has achieved the second highest accuracy of (84%) and (89%). The SVM classifier achieved the lowest accuracy of 77.52%.

On the other hand, integrating distributed processing with ELM significantly reduces training time, with NSL-KDD it reduced from (2.37697053 seconds) to (0.00972438 seconds), similarly with EdgeIIoT dataset the training and response time reduced from (6.4260058 seconds) to the lowest run-time of (0.003 seconds) among models, showing a substantial improvement in processing efficiency without compromising performance metrics [26]. Moreover, stark drop in run-time with ELM and distributed processing, enhances efficiency of IDS, mitigates the processing time issues of neural networks [27], making it more promising for real-time IoV security. Furthermore, we notice from Figures 3(a) and 3(b) ELM and DNN show progressive improvement compared to SVM and RF models, outlines advantages of applying neural networks in network security. Considering precision, recall and f1-score rate as shown in Figure 3(c), ELM and ELM+Parallel illustrated very strong performance, specifically in terms of recall and F1-score, obtaining (100%) and (97%), and (89%) and (93%) with NSL-KDD and EdgeIIoT dataset respectively, indicating their effectiveness in detecting true positive attacks.

For intra-networks, suggested approach is evaluated on Car-hacking dataset, which is based on the traffic flow CAN protocol, and results are demonstrated in Table 2 and Figure 3. Considering recall and accuracy rate, here also ELM with single and distributed processing architecture shows highest recall (95%) and superior accuracy of (83.28%) as illustrates in Figures 3(a) and 3(b) respectively, outlining its usefulness in detecting TP with minimal FN.

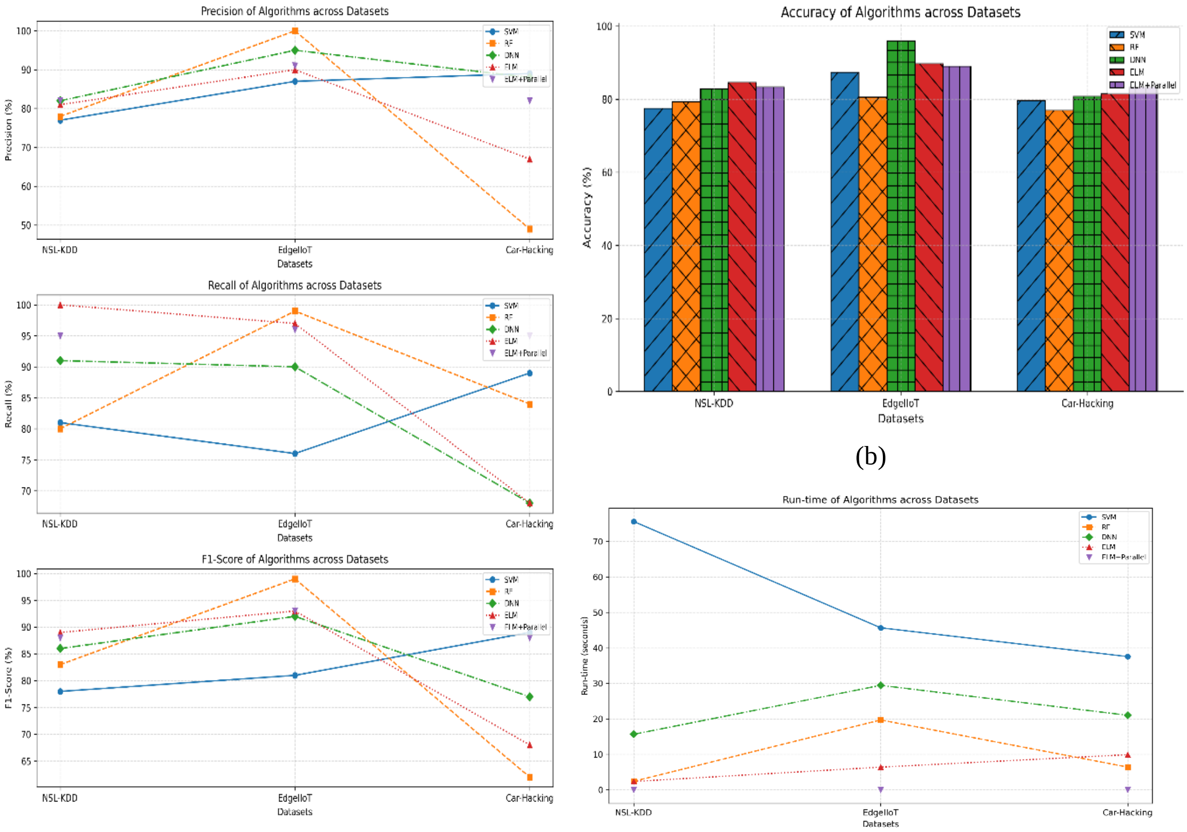
Table 1. Comparison of ML algorithms performance across datasets

Datasets	Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Run-time (seconds)
NSL-KDD	SVM	77.52	77	81	78	75.5937
	RF	79.34	78	80	83	2.4228816
	DNN	82.91	82	91	86	15.667436
	ELM	84.56	81	100	89	2.37697053
	ELM+Parallel	83.28	82	95	88	0.00972438
EdgeIIoT	SVM	87.38	87	76	81	45.6554601
	RF	80.45	100	99	99	19.70665836
	DNN	96	95	90	92	29.452039
	ELM	89.71	90	97	93	6.4260058
	ELM+Parallel	89	91	96	93	0.00313806534

Table 2. Comparison of ML algorithms performance across car-hacking dataset

Datasets	Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Run-time (seconds)
Car-Hacking	SVM	79.62	89	89	89	37.54407978
	RF	76.91	49	84	62	6.438039779
	DNN	80.72	88	68	77	21.02034759
	ELM	81.39	67	68	68	9.92550539970
	ELM+Parallel	83.28	82	95	88	0.00900697708

Further, it significantly reduces run-time from (9.92550539970 seconds) to (0.0088 seconds), through distributed processing, which emphasizes the importance of speed and efficiency for real-time apps as depicted from Figure 3(c). ELM demonstrating notable advantages over the other algorithms highlighted in survey paper by Xing *et al.* [19]. These results emphasize ability of ELM for high detection accuracy and quick response in real-time, especially in the context of IoV where the nature of network is highly dynamic [15].



(b)

(a)

(c)

Figure 3. Comparative analysis of; (a) various metrics, (b) accuracy across algorithms, and (c) run time performance of different algorithms

For real IoV networks, real-time performance of ECUs is estimated to be generally about 10 ms [3], and based on previous works that used convolutional neural network (CNN) and tree-based algorithms for intrusion diction in [8], [10], similarly, transfer learning and DNN algorithms used in [12], [13] respectively employed to train datasets, these methods require more time for computation and response. The run-time of IDS with the ELM algorithm and distributed processing method is about 0.5 ms, showing its suitability for real-time applications.

Overall, the proposed IDS, can be integrated into IoV network, it can detect various intrusions from intra_vehicle and external networks, enhancing transportation safety and efficiency. Furthermore, distributed processing techniques speed up detection tasks and can classify normal and abnormal operations within milliseconds. By safeguarding communication among vehicles, the proposed IDS strengthens the solidity of vehicles against cyber_attacks, and also helps to create safer and more efficient transportation environments. How hybrid models of DNN with ELM can enhance IDS by leveraging deep feature extraction capabilities of

DNN and rapid learning speed of ELM is worth exploring. This could result in IDS being both highly accurate and efficient, able to detect complex data patterns rapidly.

4. CONCLUSION

In this paper, we presented a novel approach for securing IoVs with help of an IDS using ELM algorithm. This suggested system can handle the critical vulnerabilities in both intra_vehicle and external vehicle networks. Particularly for intra_vehicle exposures associated with CAN protocols and other digital interfaces. With the utilization of ELM, the proposed approach demonstrated an efficient way of processing large data volumes for fast and precise anomaly detection. First, we performed data cleaning and normalization. Then, feature selection is applied based on feature correlation of CAN security data to get the best feature subset. The model is validated using benchmark datasets like NSL-KDD, EdgeIIoT for external vehicle networks, verifying its superiority in terms of detection accuracy (89.71%), recall (97%), computational resource efficiency, and a response time (3 ms) with integration of distributed processing. On the other hand, with Car-Hacking dataset for intra_vehicle networks, our approach demonstrates its ability, obtaining accuracy (83.28%), and recall (95%) with distributed processing has reduced the response time to (9 ms).

The results of this approach show the usefulness of ELM in securing IoVs networks, with significant improvement in accuracy, recall rates, and run-time efficiency with distributed processing. These findings highlight potential of ELM-based IDS in securing smart vehicles environments. Our approach demonstrated fair results in detection accuracy and response time. However, as the IoV grows, the complexity and volume of generated data will also increase, requiring more refined approaches for feature extraction and data analysis. Since deep learning is familiar for its capability to address complicated data, and extract meaningful features, can be a promising technique. In future work, we suggest examining a hybrid IDS of deep learning with extreme learning for IoV environment.

REFERENCES

[1] H. Hartenstein and K. P. Laberteaux, *VANET: Vehicular Applications and Inter-Networking Technologies*. Wiley, 2009.

[2] L. Yang, A. Moubayed, and A. Shami, “MTH-IDS: a multitiered hybrid intrusion detection system for internet of vehicles,” *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 616–632, Jan. 2022, doi: 10.1109/JIOT.2021.3084796.

[3] W. Zeng, M. A. S. Khalid, and S. Chowdhury, “In-vehicle networks outlook: Achievements and challenges,” *IEEE Communications Surveys and Tutorials*, vol. 18, no. 3, pp. 1552–1571, 2016, doi: 10.1109/COMST.2016.2521642.

[4] T. Hoppe, S. Kiltz, and J. Dittmann, “Security threats to automotive CAN networks practical examples and selected short-term countermeasures,” *Reliability Engineering and System Safety*, vol. 96, no. 1, pp. 11–25, Jan. 2011, doi: 10.1016/j.ress.2010.06.026.

[5] S. Woo, H. J. Jo, and D. H. Lee, “A practical wireless attack on the connected car and security protocol for in-vehicle CAN,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 993–1006, 2015, doi: 10.1109/TITS.2014.2351612.

[6] C. Miller and C. Valasek, “Remote exploitation of an unaltered passenger vehicle,” *Defcon 23*, vol. 2015, pp. 1–91, 2015, [Online]. Available: http://illmatics.com/Remote Car Hacking.pdf.

[7] S. Nie, L. Liu, Y. Du, and W. Zhang, “Over-the-air: how we remotely compromised the gateway, BCM, and autopilot ECUs of TESLA cars,” *BlackHat USA 2018*, vol. 1, pp. 1–19, 2018, [Online]. Available: http://www.w3.org/2000/svg%0Ahttps://www.blackhat.com/us-18/briefings/schedule/#over-the-air-how-we-remotely-compromised-the-gateway-bcm-and-autopilot-ecus-of-tesla-cars-10806.

[8] Y. Xiao, C. Xing, T. Zhang, and Z. Zhao, “An intrusion detection model based on feature reduction and convolutional neural networks,” *IEEE Access*, vol. 7, pp. 42210–42219, 2019, doi: 10.1109/ACCESS.2019.2904620.

[9] A. Rosay, F. Carlier, and P. Leroux, “Feed-forward neural network for network intrusion detection,” in *IEEE Vehicular Technology Conference*, May 2020, vol. 2020-May, pp. 1–6, doi: 10.1109/VTC2020-Spring48590.2020.9129472.

[10] L. Yang, A. Moubayed, I. Hamieh, and A. Shami, “Tree-based intelligent intrusion detection system in internet of vehicles,” in *Proceedings - IEEE Global Communications Conference, GLOBECOM*, Dec. 2019, pp. 1–6, doi: 10.1109/GLOBECOM38437.2019.9013892.

[11] S. T. Mehedi, A. Anwar, Z. Rahman, and K. Ahmed, “Deep transfer learning based intrusion detection system for electric vehicular networks,” *Sensors*, vol. 21, no. 14, p. 4736, Jul. 2021, doi: 10.3390/s21144736.

[12] X. Li, Z. Hu, M. Xu, Y. Wang, and J. Ma, “Transfer learning based intrusion detection scheme for Internet of vehicles,” *Information Sciences*, vol. 547, pp. 119–135, Feb. 2021, doi: 10.1016/j.ins.2020.05.130.

[13] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, “A deep learning approach to network intrusion detection,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 41–50, Feb. 2018, doi: 10.1109/TETCI.2017.2772792.

[14] X. Xu, J. Li, Y. Yang, and F. Shen, “Toward effective intrusion detection using log-cosh conditional variational autoencoder,” *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6187–6196, Apr. 2021, doi: 10.1109/JIOT.2020.3034621.

[15] K. Tindell, “Defending in-vehicle (CAN) buses from attacks,” in *Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium (GVSETS)*. Novi, MI: NDIA, Aug. 2022

[16] G. Bin Huang, Q. Y. Zhu, and C. K. Siew, “Extreme learning machine: a new learning scheme of feedforward neural networks,” in *IEEE International Conference on Neural Networks - Conference Proceedings*, 2004, vol. 2, pp. 985–990, doi: 10.1109/IJCNN.2004.1380068.

[17] S. Kim and R. Shrestha, *Automotive Cyber Security*. Singapore: Springer Singapore, 2020.

[18] E. Aliwa, O. Rana, C. Perera, and P. Burnap, “Cyberattacks and countermeasures for in-vehicle networks,” *ACM Computing Surveys*, vol. 54, no. 1, pp. 1–37, Jan. 2021, doi: 10.1145/3431233.

[19] L. Xing, P. Zhao, J. Gao, H. Wu, and H. Ma, “A survey of the social internet of vehicles: secure data issues, solutions, and federated learning,” *IEEE Intelligent Transportation Systems Magazine*, vol. 15, no. 2, pp. 70–84, Mar. 2023, doi: 10.1109/MITS.2022.3190036.

[20] T. M. Mitchell, Machine learning. New York: McGraw-Hill, 1997.

[21] H. Kang, B. Il Kwak, Y. H. Lee, H. Lee, H. Lee, and H. K. Kim, “Car hacking: attack and defense challenge 2020 dataset,” *IEEE DataPort*, 2021. https://dx.doi.org/10.21227/qvr7-n418.

[22] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the KDD CUP 99 data set,” in *IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009*, Jul. 2009, pp. 1–6, doi: 10.1109/CISDA.2009.5356528.

[23] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, “Edge-IIoTset: a new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning,” *IEEE Access*, vol. 10, pp. 40281–40306, 2022, doi: 10.1109/ACCESS.2022.3165809.

[24] T. Wongvorachan, S. He, and O. Bulut, “A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining,” *Information (Switzerland)*, vol. 14, no. 1, p. 54, Jan. 2023, doi: 10.3390/info14010054.

[25] S. Amraee, M. Chinipardaz, and M. Charoosaei, “Analytical study of two feature extraction methods in comparison with deep learning methods for classification of small metal objects,” *Visual Computing for Industry, Biomedicine, and Art*, vol. 5, no. 1, p. 13, Dec. 2022, doi: 10.1186/s42492-022-00111-6.

[26] R. Peng, W. Li, T. Yang, and K. Huafeng, “An internet of vehicles intrusion detection system based on a convolutional neural network,” in *Proceedings - 2019 IEEE Intl Conf on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking, ISPA/BDCloud/SustainCom/SocialCom 2019*, Dec. 2019, pp. 1595–1599, doi: 10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00234.

[27] Y. Otoum, Y. Wan, and A. Nayak, “Transfer learning-driven intrusion detection for internet of vehicles (IoV),” in *2022 International Wireless Communications and Mobile Computing, IWCMC 2022*, May 2022, pp. 342–347, doi: 10.1109/IWCMC55113.2022.9825115.

BIOGRAPHIES OF AUTHORS

Aziz Ullah Karimy is currently pursuing his Ph.D. in machine learning and IoT security at the University College of Engineering, Science and Technology, Hyderabad (JNTUH). He obtained his Bachelor’s degree in Electronics and Communication Engineering from Visvesvaraya Technological University, India, in 2016, followed by a Master’s degree in Embedded Systems from Jawaharlal Nehru Technological University, Hyderabad, in 2021. His research focuses on advancing the applications of machine learning in cyber security, specifically in securing IoT environments. He can be contacted at email: azizullah.karimy91@gmail.com.

Prof. Dr. Putta Chandrasekhar Reddy is a Senior Professor at JNTUH University College of Engineering, Science and Technology, Hyderabad. He holds a Ph.D. degree, along with Master of Technology (M.Tech) and Master of Engineering (M.E), as well as a Bachelor’s degree in Bachelor of Engineering (B.E). His primary areas of interest include wireless communication, 5G, image processing, electrical vehicles, and the IoT. He has over 30 years of experience in the academic sphere, and his career has been distinguished by mentoring over 35 Ph.D. researchers and contributing approximately 150 scholarly papers to the scientific community. He can be reached at email: drpcsreddy@jntuh.ac.in.