DOI: https://doi.org/10.61762/pjstvol7iss1art27052

PLATFORM - A Journal of Science & Technology

Received: 3 May 2024, Accepted: 24 June 2024, Published: 28 June 2024 Publisher: UTP Press, Creative Commons: CC BY 4.0

COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS USING THE CICIOV2024 DATASET

Nur 'Aliah Binti Amirudin*, Said Jadid Abdulkadir

Computer and Information Science Department, Universiti Teknologi PETRONAS, Malaysia

*Email: nur_22009393@utp.edu.my

ABSTRACT

Leveraging the CICIoV2024 dataset, this study evaluate the performance of different algorithms in accurately identifying and classifying intrusions in vehicular networks, specifically using LightGBM, XGBoost, CatBoost, and LCCDE algorithms. A comparative analysis conducted considered key performance metrics such as accuracy, precision, recall, and F1-score, shedding light on the strengths and limitations of each approach. Through rigorous experimentation and evaluation, the remarkable results demonstrated 100% accuracy, recall, precision, and F1-score across all models. These findings highlight the efficacy of employing advanced machine learning techniques for enhancing intrusion detection capabilities in IoV environments. The insights gained from this study can inform the development of highly accurate and reliable intrusion detection systems tailored to the unique challenges of vehicular networks.

Keywords: *CatBoost, CICIoV2024, IoV, LCCDE, LightGBM, XGBoost*

INTRODUCTION

The transportation sector, much like other industries, has undergone a profound transformation due to technological advancements. Modern vehicles now integrate an array of sensors, software, and technology, enabling them to share filtered sensor data via the Internet cloud with a network of autonomous vehicles (AUVs) [1]. This evolution has led to the emergence of the Internet of Vehicles (IoV), where interconnected vehicles communicate with each other and infrastructure elements enable advanced functionalities such as cooperative driving, real-time traffic management, and enhanced safety features [2]. However, these advancements bring new security challenges as vehicles become increasingly vulnerable to cyber threats [3].

Intrusion detection systems (IDSs) play a crucial role in ensuring the security of the IoV system by identifying malicious activities. IoVs are particularly susceptible to network threats such as spoofing and denial of service (DoS) attacks, which can have severe consequences to human safety if left undetected. For instance, compromising critical vehicle systems like braking or steering can lead to accidents, fatalities and property damage [4]. The importance of IDSs was underscored in a notable 2015 experiment, where white hat hackers infiltrated a Jeep Cherokee's system and successfully controlled the car's speed and braking, demonstrating the severity of potential risks posed by cyber-attacks on vehicles [5]. By continuously monitoring network traffic and detecting anomalous behaviour, IDSs contribute to maintaining the integrity and robustness of IoV services. However, designing effective IDS for IoV environments is challenging due to the dynamic nature of vehicular networks and diverse cyber threats [6].

Machine learning (ML)-based IDSs leverage the vast amount of data generated by interconnected vehicles to detect unauthorised infiltrations [7]. ML algorithms can be trained to identify abnormal patterns in network traffic indicative of cyber-attacks, thereby enhancing detection accuracy. However, the performance of these algorithms is influenced by various factors such as feature selection, model architecture, and dataset characteristics, necessitating careful parameter tuning for optimal results.

Despite the existing IDS solutions, significant gaps remain. Many traditional IDSs utilise older benchmark datasets, which results in a lack of adaptability required to cope with the rapidly evolving threat landscape of IoV environments [8]. Moreover, datasets that are highly imbalanced and noisy can significantly lower the performance of ML-based IDS. Consequently, there is a pressing need for more adaptable and efficient IDS solutions capable of addressing the unique challenges of IoV.

This study addresses the gaps mentioned above by utilising the novel CICIoV2024 dataset to evaluate the performance of different ML algorithms—LightGBM, XGBoost, CatBoost, and Leader Class and Confidence Decision Ensemble (LCCDE)—in accurately identifying and classifying intrusions in vehicular networks. The key contributions are as follows:

1. Comprehensive Comparative Analysis: Compared key performance metrics such as accuracy, precision, recall, and F1 score across multiple state-of-the-art ML algorithms.
2. High Performance: These experiments demonstrate remarkable results, with 100% accuracy, recall, precision, and F1-score achieved by all evaluated models.
3. Enhanced Intrusion Detection: The findings highlight the efficacy of employing advanced ML techniques for enhancing intrusion detection capabilities in IoV environments.
4. Practical Insights: The insights gained from this study can inform the development of highly accurate and reliable IDS tailored to the unique challenges of vehicular networks, addressing the limitations of prior work.

RELATED WORKS

Existing work on ML algorithms applied to IDS in the IoV especially those focusing on detecting DoS attacks and those addressing spoofing attacks was reviewed. The differences between prior work and this study proposed solution, followed by a comparative table summarising key characteristics are also highlighted.

ML Algorithms for Detecting DoS Attacks

Several studies have demonstrated the effectiveness of ML algorithms in detecting DoS attacks within vehicular networks. Gou et al. [3] an IDS based on LightGBM was proposed for detecting DoS attacks in Vehicular Ad Hoc Networks (VANETs). The study showcased LightGBM's notable accuracy and low false positive rates in IoV scenarios. The IDS utilised LightGBM's ability to handle large datasets efficiently, contributing to its high detection accuracy. Research by Dhaliwal et al. [10] presented an XGBoost-based IDS tailored for detecting DoS attacks in VANETs. This IDS achieved commendable detection rates and minimised false positives, positioning it as a viable solution for intrusion detection in IoV environments. XGBoost's parallel tree construction and regularisation techniques contributed to its superior performance.

ML Algorithms for Detecting Spoofing Attacks

Other studies have focused on detecting spoofing attacks using advanced ML algorithms. [9] developed a CatBoost-based IDS for detecting anomalous behaviour in connected vehicles. The findings indicated that CatBoost exhibited superior accuracy and robustness, particularly when handling high-dimensional and imbalanced datasets. CatBoost's advanced techniques for handling categorical features were pivotal in achieving these results. Yang et al. [11] introduced an IDS based on the Leader Class and Confidence Decision Ensemble (LCCDE) for detecting spoofing attacks in connected vehicles. The study noted LCCDE's balanced performance across different attack classes and its adaptability to dynamic IoV environments. LCCDE's ability to capture local data structures enhanced its effectiveness.

Table 1 highlights the strengths and limitations of prior works while positioning the proposed solution as a comprehensive and advanced IDS for IoV environments.

METHODOLOGY

To ensure a comprehensive understanding of the proposed solution, we first provide an overview of the approach. The overview of the proposed solution is shown in Figure 1. The methodology comprises of:

Table 1 Table of comparison between ML algorithms

Algorithm	Attack Type	Dataset	Key Characteristics	Limitations
LightGBM [3]	DoS	CICIDS2017 Car-Hacking dataset	High accuracy, low false positive rates	Limited to DoS attacks
CatBoost [9]	Spoofing	NSL-KDD	Superior accuracy handles high-dimensional data	Focuses on spoofing attacks only
XGBoost [10]	DoS	NSL-KDD	High detection rates, minimised false positives	Limited to DoS attacks
LCCDE [11]	Spoofing	CICIDS2017 Car-Hacking dataset	Balanced performance, adaptable to dynamic environments	Limited to spoofing attacks
LightGBM, XGBoost, CatBoost, LCCDE	DoS & Spoofing	CICIoV2024	Comprehensive approach, advanced preprocessing, robust evaluation	None Identified

1. Dataset collection.
2. Data preprocessing: This phase includes Z-score normalisation, k-means clustering to manage dataset size, and synthetic minority oversampling to balance class distributions.
3. Data split (80:20).
4. Model training: Evaluate multiple machine learning algorithms, including LightGBM, XGBoost, and CatBoost, to identify and classify intrusions in the IoV environment. The CICIoV2024 dataset is utilised to ensure robust performance evaluation.
5. Model testing: The best-performing model/models are utilised for LCCDE.
6. Discussion

The CICIoV2024 dataset originally employed machine learning methods, including Logistic Regression (LR), Random Forest (RF), Adaboost (AD), and Deep Neural Network (DNN). While these algorithms have demonstrated satisfactory performance in accurately identifying attacks and non-attacks, their precision, recall, and F1 scores are found to be suboptimal [6]. This may be attributed to the extensive size of the dataset. To address this limitation, various optimisation techniques and newer machine learning algorithms are likely to yield improvements in overall performance metrics such as accuracy, precision, recall, and F1-score. The purpose of this work is to develop an IDS that can enhance the effectiveness and efficiency of

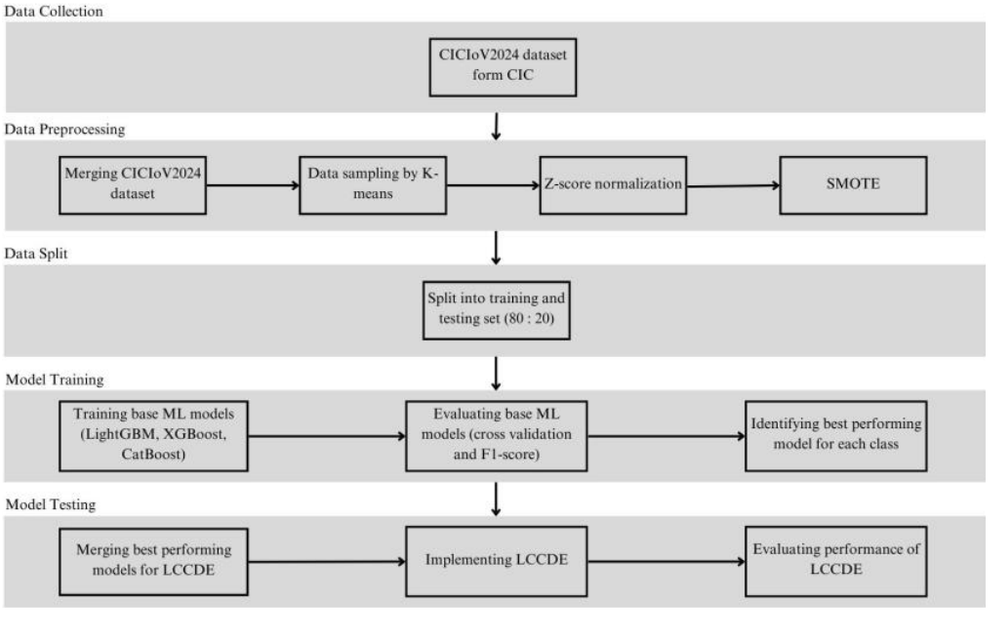

Figure 1 Proposed framework

detecting Denial of Service (DoS) and spoofing attacks in the CICIoV2024 dataset, which would ultimately benefit the automotive industry due to its enhanced reliability and security in safeguarding vehicle and passenger safety using advanced machine learning algorithms LightGBM, XGBoost, CatBoost and LCCDE.

Dataset Characteristic

The CICIoV2024 dataset is the product of extensive experimentation conducted by the Canadian Institute for Cybersecurity (CIC) on the ECUs of a 2019 Ford vehicle. This dataset provides a view of intra-vehicle communications, shedding light on the intricate interactions within the vehicle's internal network. To ensure the safety of the vehicle's occupants, researchers meticulously orchestrated five distinct attacks on the intact inner structure of the 2019 Ford car. It is noteworthy that these attacks were carefully designed to immobilise the vehicle while posing no risk to the driver or passengers. Executed through the CAN bus protocol, these attacks fall within the categories of spoofing and DoS, showcasing the vulnerabilities inherent in vehicular communication systems.

The dataset includes 12 distinct features extracted from intra-vehicular communications, meticulously documented and outlined in Table 2. Each feature serves a specific purpose in characterising the interactions within the vehicle's network and identifying potential anomalies or malicious activities. The CICIoV2024 dataset distribution is shown in Table 3.

Table 2 Features of the CICIoV2024 dataset

Feature Name	Description
ID	Arbitration
Data_0	Byte 0 of the data transmitted
Data_1	Byte 1 of the data transmitted
Data_2	Byte 2 of the data transmitted
Data_3	Byte 3 of the data transmitted
Data_4	Byte 4 of the data transmitted
Data_5	Byte 5 of the data transmitted
Data_6	Byte 6 of the data transmitted
Data_7	Byte 7 of the data transmitted
label	The identification of benign or malicious traffic
category	The category to which traffic belongs
specific class	The identification of the specific class of traffic

Table 3 Distribution of the CICIoV2024 dataset

Attack Type	Label	Count
-	Benign	1223737
DoS	DoS	74663
Spoofing	RPM	54900
	Speed	24951
	Steering Wheel	19977
	Gas	9991

Data Preprocessing

Firstly, to ensure that ML training is more efficient, the dataset undergoes Z-score normalisation, where the mean of all values is 0, and the standard deviation is 1. Datasets with largely different features that do not undergo normalisation may cause biased ML models that only emphasise large-scale features. The Z-score method implemented for each normalised feature value, x_n, is denoted by [3]:

$$x_n = \frac{x - \mu}{\sigma} \quad (1)$$

Where x is the original feature value, μ and σ are the mean and standard deviation of the feature values, respectively.

Secondly, due to the sheer size of the CICIoV2024 dataset, the k-means clustering method was applied to the majority class to reduce the size of the dataset and is denoted by [3]:

$$\sum_{i=0}^{n_k} \min_{\mu_k \in C_k} (x_i - u_j)^2 \quad (2)$$

where $(x_1, \ldots, x_n)$ is the data matrix, u_j, also called the centroid of a cluster C_k, is the mean of all the samples in C_k; and n_k is the total number of sample points in the cluster C_k. K-means has a linear time complexity of $O(nkt)$, where k, is the data size, k, is the number of clusters, and t is the number of iterations. The number of optimal clusters was identified using the Elbow Method [3].

Lastly, to obtain a balanced representation of the imbalanced dataset and ensure that the model will not be biased toward the majority class, Synthetic Minority Oversampling (SMOTE) is applied. The minority class is inflated to at least half of the number of the majority class. However, caution was taken not to oversample

the minority class too aggressively because it would lead to poor generalisation of unseen data. For each instance X in the minority class, assuming X_i is a sample randomly selected from the k nearest neighbours of X, a new synthetic instance X_n can be denoted by [3]:

$$X_n = X + rand\ (0,1) * (X_i - X),\ i = 1, 2, ..., k \qquad (3)$$

where *rand* (0,1) represents a random number in the range of (0,1). The list of notations can be found in the notations section of this paper.

Model Training

For model training, the dataset is split into a training and testing set (80:20) [11]. LightGBM is based on the gradient-boosting decision tree (GBDT) algorithm, which iteratively builds a series of decision trees to make predictions. However, what sets LightGBM apart is its novel approach to building these decision trees. Instead of using the traditional depth-wise approach, LightGBM employs a leaf-wise approach, which enables it to grow deeper trees by splitting nodes that yield the maximum reduction in loss. This approach results in more accurate models with fewer nodes, leading to faster training times and reduced memory consumption. Additionally, it implements efficient algorithms for handling categorical features and missing values, reducing the need for preprocessing and improving overall model accuracy. Overall, LightGBM is a powerful and versatile tool for a wide range of ML tasks, including classification, regression, and ranking. Its combination of speed, accuracy, and scalability makes it a popular choice among data scientists and practitioners seeking to build high-performance predictive models [12].

XGBoost, short for eXtreme Gradient Boosting, is a powerful and widely used ML algorithm known for its efficiency, effectiveness, and versatility. It belongs to the family of gradient boosting algorithms, which are ensemble learning techniques that combine the predictions of multiple individual models, typically decision trees, to produce a more accurate and robust final prediction. One of the key features of XGBoost is its innovative approach to building decision trees. Unlike traditional gradient boosting algorithms that build trees sequentially, XGBoost employs a more sophisticated optimisation technique known as gradient boosting with parallel tree construction. This technique allows XGBoost to build trees in parallel, greatly accelerating the training process [13].

CatBoost is a powerful gradient-boosting algorithm designed to handle categorical features seamlessly, making it particularly well-suited for datasets with a mix of numerical and categorical variables. CatBoost incorporates several advanced techniques to improve model performance and robustness. These include ordered boosting, which optimises the order in which the trees are built to minimise overfitting, and a novel method for handling numerical features, which allows CatBoost to automatically detect and use the most appropriate scaling for each feature [14].

Model Testing

The LCCDE model employs leader models for each class to generate the ultimate prediction within the ensemble. In cases where multiple leader models are available for different classes, LCCDE prioritises the one exhibiting the highest prediction confidence to determine the final decision. By incorporating LCCDE, the ensemble model achieves peak performance in detecting various types of attacks. LCCDE is specifically crafted to enhance the accuracy and robustness of classification models by capitalising on the distances between class centres within the feature space. During the prediction phase, LCCDE consolidates the predictions of all base classifiers to formulate a final prediction for each instance. This consolidation process can manifest in various forms, including majority voting or weighted averaging, contingent upon the specific implementation employed. One of the key advantages of LCCDE is its ability to capture the local structure of the data, allowing it to make more accurate predictions, especially in regions where the class boundaries are non-linear or complex. Additionally, by dividing the feature space into local regions, LCCDE can handle datasets with imbalanced class distributions more effectively, as it can adapt its predictions to the local characteristics of each region [10].

While previous studies have made significant contributions, this study differentiates itself in several ways, specifically on proposed IDS:

1. Leverages a combination of LightGBM, XGBoost, CatBoost, and LCCDE, aiming to capitalise on the strengths of each algorithm.
2. Utilises the CICIoV2024 dataset, which encompasses a broader range of intra-vehicle communication features and attack types compared to datasets used in previous studies.

3. The methodology includes advanced data preprocessing techniques such as Z-score normalisation, k-means clustering, and SMOTE, enhancing the performance of the proposed IDS.
4. Provide detailed comparison of our results with previous studies, focusing on accuracy, precision, recall, F1-score, and execution time to demonstrate the superiority of our approach.

RESULT AND DISCUSSION

The performance of the advanced ML models (LightGBM, XGBoost, CatBoost, LCCDE) was evaluated in terms of accuracy, precision, recall, F1-score, and execution time. Additionally, their performance with the previously used models in the original paper [6] to assess their effectiveness in handling the specific characteristics of the CICIoV2024 dataset. The results (Table 4) indicate that LightGBM, XGBoost, CatBoost and LCCDE all show superior results in terms of accuracy, precision, recall and F1-score in comparison to LR, AD, DNN and RF. LightGBM, XGBoost, CatBoost, and LCCDE all achieved an accuracy of 0.99, 1.0, 1.0, and 1.0, respectively. This indicates that all four algorithms have demonstrated exceptional performance in handling this dataset. Although the results are quite similar, each algorithm might have its unique advantages and disadvantages depending on the specific problem and dataset.

LightGBM is highly accurate and efficient in handling large-scale datasets, such as network traffic. Its fast training speed and low memory usage make it suitable for real-time applications [15]. However, its lack of interpretability and sensitivity to hyperparameters can present notable challenges. Despite these challenges, its performance on the CICIoV2024 dataset was exceptional, achieving nearly perfect accuracy and other performance metrics.

Table 4 Performance evaluation on CICIoV2024

Accuracy	Precision	Recall	F1	Execution	Time
LR	0.95	0.74	0.68	0.63	-
AD	0.87	0.14	0.17	0.15	-
DNN	0.95	0.74	0.68	0.63	-
RF	0.95	0.60	0.68	0.62	-
LightGBM	0.99	0.99	0.99	0.99	5.39s
XGBoost	1.0	1.0	1.0	1.0	4.59s
CatBoost	1.0	1.0	1.0	1.0	1 min 11s
LCCDE	1.0	1.0	1.0	1.0	1 min 43s

CatBoost offers compelling advantages in detecting DoS and spoofing attacks, including its ability to handle categorical features, mitigating overfitting, and built-in cross-validation. However, challenges such as computational resource requirements and limited interpretability should be considered [16]. Despite these challenges, CatBoost achieved perfect scores across all performance metrics in this study, demonstrating its robustness in handling complex data.

XGBoost is known for its scalability, regularisation techniques, and flexibility in effectively detecting and mitigating DoS and spoofing attacks. However, it is sensitive to hyperparameters and somewhat limited in interpretability [17]. XGBoost's performance was outstanding, achieving perfect scores in all evaluated metrics and demonstrating quick execution time, making it a practical choice for real-time applications.

LCCDE adopts a sophisticated leader model selection strategy, prioritising models with the highest prediction confidence for each class. This selective approach ensures that the ensemble prioritises the most reliable and accurate predictions for decision-making, amplifying its efficacy in detecting various attack types, including DoS and spoofing. However, it is important to address the challenges in implementation due to its complexity [18]. However, although LCCDE is complex to implement, it matched the perfect scores of other advanced models, showcasing its potential in robust intrusion detection.

Overall, the LCCDE model performed similarly to the other algorithms (LightGBM, XGBoost, CatBoost), which can be attributed to their shared objective of maximising predictive accuracy and generalisation performance. Each model may excel in certain aspects, such as handling categorical features (CatBoost), scalability (XGBoost), or interpretability (LightGBM), but they ultimately strive to achieve optimal predictive performance. If LCCDE effectively combines diverse base models and leverages their complementary strengths, it is possible to achieve results similar to those of these individual algorithms.

The execution time for each algorithm varied significantly. LightGBM and XGBoost were much faster, with execution times of 5.39 seconds and 4.59 seconds, respectively. CatBoost and LCCDE, however, had longer execution times of 1 minute 11 seconds and 1 minute 43 seconds, respectively. This difference highlights the trade-off between computational efficiency and performance. For real-time applications, LightGBM and XGBoost may be more suitable due to their faster execution times. However, for scenarios where interpretability and handling of categorical features are critical, CatBoost and LCCDE may be more advantageous despite their longer execution times.

The study conclusively demonstrates the significant potential of Machine Learning (ML) models, including LightGBM, CatBoost, XGBoost, and LCCDE, in enhancing intrusion detection systems (IDS) for the Internet of Vehicles (IoV). These models have shown remarkable performance on the CICIoV2024 dataset, achieving high metrics in accuracy, precision, recall, and F1-scores, thereby affirming the efficacy of advanced ML algorithms in strengthening the cybersecurity defenses of interconnected vehicular networks. Future research directions emphasize the need for developing adaptive IDS that can dynamically update in real-time to effectively counter new and evolving cyber threats, thereby ensuring robust security and continuous protection in the IoV ecosystem.

ACKNOWLEDGEMENT

The authors would like to acknowledge Universiti Teknologi PETRONAS for supporting the completion of this work.

REFERENCES

[1] E.-K. Lee, M. Gerla, G. Pau, U. Lee, and J.-H. Lim, "Internet of Vehicles: From intelligent grid to autonomous cars and vehicular fogs," *International Journal of Distributed Sensor Networks*, vol. 12, no. 9, p. 155014771666550, 2016.

[2] M. N. O. Sadiku, M. Tembely, and S. M. Musa, "INTERNET OF VEHICLES: AN INTRODUCTION," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 8, no. 1, p. 11, 2018.

[3] W. Gou, H. Zhang, and R. Zhang, "Multi-Classification and Tree-Based Ensemble Network for the Intrusion Detection System in the Internet of Vehicles," *Sensors*, vol. 23, no. 21, pp. 8788–8788, 2023.

[4] L. Yang, A. Moubayed, and A. Shami, "MTH-IDS: A Multi-Tiered Hybrid Intrusion Detection System for Internet of Vehicles," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 616-632, 2022.

[5] J. Golson, "Jeep hackers at it again, this time taking control of steering and braking systems," *The Verge*, 2016. [Online]. Available: https://www.theverge.com/2016/8/2/12353186/car-hack-jeep-cherokee-vulnerability-miller-valasek. [Accessed June 13, 2024].

[6] E. Carlos Pinto Neto, H. Taslimasa, S. Dadkhah, S. Iqbal, P. Xiong, T. Rahman and A. A. Ghorbani, "CICIoV2024: Advancing realistic IDS approaches against DoS and spoofing attack in IoV CAN bus," *Internet of things*, vol. 26, no. 101209, pp. 101209–101209, 2024.

[7] E. Alalwany and I. Mahgoub, "Security and Trust Management in the Internet of Vehicles (IoV): Challenges and Machine Learning Solutions," *Sensors*, vol. 24, no. 2, p. 368, 2024.

[8] I. D. Aiyanyo, H. Samuel, and H. Lim, "A Systematic Review of Defensive and Offensive Cybersecurity with Machine Learning," *Applied Sciences*, vol. 10, no. 17, p. 5811, 2020.

[9] N. Singh Bhati, "A New Intrusion Detection Scheme Using CatBoost Classifier," in *Proc. Forthcoming Networks and Sustainability in the IoT Era '20*, 2020, pp. 169—176.

[10] S. Dhaliwal, A.-A. Nahid, and R. Abbas, "Effective Intrusion Detection System Using XGBoost," *Information*, vol. 9, no. 7, p. 149, 2018.

[11] L. Yang, A. Shami, G. Stevens, and Stephen de Rusett, "LCCDE: A Decision-Based Ensemble Framework for Intrusion Detection in The Internet of Vehicles," in *Proc. GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, 2022, pp. 1-6.

[12] N. Nabil, N. Najib, and J. Abdellah, "Leveraging Artificial Neural Networks and LightGBM for Enhanced Intrusion Detection in Automotive Systems," *Arabian Journal for Science and Engineering,* 2024.

[13] C. D. Kokane, G. Mohadikar, S. Khapekar, B. Jadhao, T. Waykole, and V. V. Deotare, "Machine Learning Approach for Intelligent Transport System in IOV-Based Vehicular Network Traffic for Smart Cities," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 11s, pp. 06-16, 2023.

[14] M. S. Korium, M. Saber, A. Beattie, A. Narayanan, S. Sahoo, and P. H. J. Nardelli, "Intrusion detection system for cyberattacks in the Internet of Vehicles environment," *Ad Hoc Networks*, vol. 153, p. 103330, 2024.

[15] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Proc. 31st Conference on Neural Information Processing Systems '17*, 2017, pp. 3149-3157.

[16] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," in *Proc. 32nd Conference on Neural Information Processing Systems '18*, 2018, pp. 6639-6649.

[17] T. Chen and C. Guestrin, "XGBoost: a Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 2016, pp. 785–794.

[18] Y. Zhang, J. Zhang, and J. Wu, " Leader class and confidence decision ensemble for intrusion detection in vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no.5, pp. 4620-4633, 2019.