INTERNATIONAL SCIENTIFIC JOURNAL OF ENGINEERING AND MANAGEMENT (ISJEM) ISSN: 2583-6129
CONFERENCE PROCEEDINGS: VOLUME: 05 ACME – 2026 DOI:10.55041/ISJEM.ACME004
AN INTERNATIONAL SCHOLARLY || MULTIDISCIPLINARY || OPEN ACCESS || INDEXING IN ALL MAJOR DATABASE & METADATA

A Secure and Robust Machine Learning Model for Intrusion Detection in Internet of Vehicles

S SUNIL KUMAR
MTech, Department of Computer Science and Engineering
Vemu Institute of Technology, P-Kothakota, Chittoor
Andhra Pradesh – 517112, India
sunilkumar473@gmail.com

Mr. G. Lokesh
Assistant Professor, MTech Dept of Computer Science of Engineering Vemu Institute of Technology, P-Kothakota, Chittoor
Andhra Pradesh – 517112, India

Abstract: The research introduces a new Intrusion Detection System (IDS) which uses advanced machine learning methods to protect autonomous vehicles. The main purpose of the IDS system entails identifying cyberattacks which the system needs to categorize into different attack types that consist of DDoS Fuzzy Impersonation and standard Free traffic. The complete model development process uses the CAN-intrusion-dataset which contains information about vehicle communication through its message and byte-level signal and target label data. The system uses ML techniques which include RF and Gradient Boosting and Adaboost and LSTM and CatBoost for security threat detection and prevention. The Random Forest and Decision Tree models achieved 94% accuracy which makes them successful in detecting unauthorized access attempts to vehicle networks. The system achieves its maximum protection level through its complete execution of multiple security protocols. The new security enhancements will make smart vehicle systems more secure and reliable. The detection system develops as a scalable solution which protects smart vehicles from the growing number of cyber threats.
Keywords: RF, Gradient Boosting, Adaboost, LSTM, CatBoost classifiers, DDoS attack, Fuzzy attack, and Normal traffic patterns.

I. INTRODUCTION

Modern vehicles have become susceptible to hacking because of their adoption of smart technologies which new technologies brought into the automotive industry [1]. Modern vehicles establish security vulnerabilities through their use of multiple networks which operate on Controller Area Network (CAN) communication systems. The cyber threats are very dynamic and hence, the demand for advanced intrusion detection and removal systems that can operate in real-time will always be there.

The vehicle network system requires an Intrusion Detection System (IDS) because it identifies dangerous actions which protect the network from cyber-attacks. Vehicle networks need different detection methods to handle their security problems because traditional intrusion detection systems were built for general IT networks, which operate differently [2]. The project which we are discussing has the goal of creating an IDS system for modern vehicles through machine learning which will detect and categorize all types of attacks.

The proposal describes a security network system which detects DDoS (Distributed Denial of Service) and Fuzzy and Impersonation attacks while designating common network traffic as Free traffic. The system development process needs the CAN-intrusion-dataset which provides vehicle communication data that contains Message-ID and Byte-level signals and target labels which function as attack detection inputs [3]. The study will evaluate five classification methods which include Random Forest and Gradient Boosting and Adaboost and LSTM and CatBoost to identify vehicle features and their associated threats. The advanced algorithms implementation will enable the system to detect abnormal behavior patterns which assist in safeguarding the smart vehicle network against new cybersecurity threats that try to penetrate the system without detection [4]. The research demonstrates that machine learning techniques will develop a strong and expandable security system which will safeguard upcoming smart transportation systems while enhancing current vehicle network protection systems.

Objective: The project creates an advanced Intrusion Detection System Protected against cyber threats which safeguards smart vehicles from attacks. The system identifies different types of attacks which include Distributed Denial of Service (DDoS) attacks and Fuzzy attacks and Impersonation attacks and standard network traffic. The project creates a machine learning-based detection system which operates at scale to identify real-time security breaches using Random Forest and Gradient Boosting and Adaboost and LSTM and CatBoost methods. The project establishes a secure vehicular network security system which defends smart transportation systems against cyber threats according to reference [5].

Scope: The project will create an Intrusion Detection System (IDS) that continuously monitors vehicular networks to protect smart vehicles from cyber threats. The project employs machine learning algorithms which include Random Forest and Gradient Boosting and Adaboost and LSTM and CatBoost to detect DDoS attacks and Fuzzy attacks and Impersonation attacks while analyzing regular traffic patterns [6].The project will create a real-time Intrusion Detection System which can expand its defenses to protect upcoming smart transportation systems by utilizing the CAN-intrusion-dataset which contains vehicle communication data to guard against security threats and data integrity breaches.

The project develops an Intrusion Detection System (IDS) through machine learning technology because smart vehicles experience increasing cybersecurity risks. Three attack types which include DDoS attacks and Fuzzy attacks and Impersonation attacks together create security threats to vehicles that use CAN network communication systems [14]. The system achieves rapid cyber threat detection and classification through its use of Random Forest and Gradient Boosting and Adaboost and LSTM and CatBoost algorithms while maintaining necessary security protections. The research shows that machine learning provides efficient and scalable solutions to protect smart transportation systems from emerging cyber threats which will develop in the future [15].

II. RELATED WORKS

The research conducted by shows that detection systems currently use machine learning detection systems which have become available during the past three years as their main detection method instead of using traditional detection systems [7]. The 2024 research showed that K-nearest neighbors together with non-tree-based ensemble learning methods produced results which equaled advanced machine learning systems while researchers constructed a working prototype that functioned as an IDS system for detecting security threats in driverless vehicles [8]. The research aimed to achieve two objectives which involved enhancing detection accuracy and minimizing the need for computational resources.

The study main research focus examined vehicle communication systems which operate through the controller area network (CAN) bus because it serves as the standard vehicle communication system [9]. The 2024 paper about CAN buses developed an unsupervised intrusion detection system which uses machine learning methods to identify abnormal activities without needing available labeled training data. The approach proves useful when there is insufficient attack data to create labeled training datasets.

Deep learning models served as the only method which could solve the detection problem. The 2023 intrusion detection system (IDS) used deep convolutional neural networks (DCNN) which achieved high detection accuracy [10]. DCNN technology enables automatic extraction of network traffic attributes which accelerates cyber threat detection while improving detection outcomes.

The implementation of AI interpretability methods through XAI in IDS systems has been suggested as a way to enhance detection transparency and trustworthiness [11].A particular research effort in 2023 made the proposal of X-CANIDS an intrusion detection system operating on Controller Area Network (CAN) which possesses an explainable feature [12].The system translates CAN bus messages into human-readable signals which serve two functions: detection and identification of vulnerable vehicle components.

Various vehicles utilize their communication networks to identify security breaches while maintaining protection of their personal information. A 2023 research team created a federated learning framework for Internet of Vehicles systems which uses SMOTE to handle class imbalance and outlier detection to enhance model performance while preserving data privacy [13].

III.PROPOSED METHODOLOGY

The first system development project develops a smart vehicle Intrusion Detection System (IDS) which employs machine learning algorithms to detect and classify various cyber-attack methods. The study uses the car's

communication data which includes the CAN-intrusion-dataset as its research material. The dataset consists of three primary components which include Message_ID and Byte-level signals and target labels which combine to support the system's anomaly detection process. The research will use five different machine learning techniques which include Random Forest and Gradient Boosting and Adaboost and LSTM and CatBoost to differentiate between standard network behavior and three specific attack patterns which include DDoS attacks and Fuzzy attacks and Impersonation attacks. The system designers create an Intrusion Detection System (IDS) which will operate correctly from its initial deployment because the system needs to achieve high threat detection accuracy while it safeguards vehicle networks against emerging cyberattack methods. The system design process developed two fundamental components which resulted in an efficient system design that fulfills user requirements shown in (Figure 1).

1. Data Collection and Preprocessing

- The project will use the CAN-intrusion-dataset which contains vehicle communication data through its collection of message signals and byte-level signals and its target labels which show normal or malicious traffic.

- The dataset cleaning procedures will remove all missing values and duplicate entries and any data elements which do not contribute to research objectives.

- The researchers will improve the model performance through their extraction of critical features which include message types and byte-level values and traffic patterns.

- The process of normalizing data through standardization will establish uniform feature values which enable the model to achieve faster convergence times.

- The dataset will be divided into training and validation and test sets which commonly use an 80-20 or 70-30 split to provide accurate model assessments.

2. Model Selection and Development

The system will use different ML methods to detect cyberattacks which target vehicle networks.

- The ensemble learning method improves classification performance through its combination of multiple decision trees.

- The boosting method creates new models which correct all errors present in earlier models.

- The boosting algorithm uses several weak classifiers to create a strong classification system.

- The LSTM recurrent neural network design uses time-series data analysis to identify attacks by studying vehicle message patterns.

- The CatBoost gradient boosting algorithm uses its distinct optimization functions to manage categorical data while it performs its analysis of complex feature relationships.

3. Model Training

- The team needs to test their model, which requires them to use all of their data that contains both their training materials and their complete dataset. The team will conduct hyperparameter tuning to achieve optimal model performance.

- The model strength assessment will use cross-validation methods which include K-fold cross-validation to prevent model overfitting.

- The system will use ensemble methods which include Random Forest to improve detection accuracy by combining predictions from different models.

4. Model Evaluation

The model assessment process will use standard classification metrics to evaluate its performance which includes Accuracy as one of its assessment methods.

- The correct prediction rate for the model exists as the model's accurate predictions.

- The system predicts positive results based on the true positive results which constitute the positive results.

- The true positive detection metric demonstrates the system's ability to correctly identify actual positive cases.

- The F1-Score computes its value through the harmonic mean of precision and recall which creates equal weight between the two metrics.

- The Area Under the Receiver Operating Characteristic curve serves as a complete assessment tool which shows how well the model performs across different classification settings.

- The system will use a confusion matrix to evaluate its classification performance by detecting false positives and false negatives together with misclassification errors.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{F1} - \text{Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision}+\text{Recall}}$$

- where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative counts, respectively.
- FP is the number of false positives.
- FN is the number of false negatives.

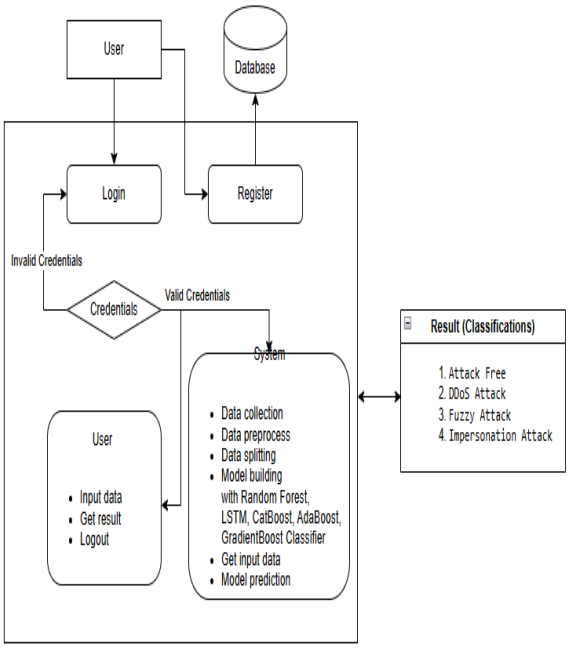

Figure 1: Architecture of Methodology

IV. IMPLEMENTATION

1.Random Forest:

The smart vehicle Intrusion Detection System uses Random Forest as its main detection system because the algorithm needs to achieve total detection success after it already mastered the ability to identify all cyberattack types which include Distributed Denial of Service attacks and Fuzzy attacks and Impersonation attacks and other attack techniques. The system enables them to maintain their ability to detect movements which deviate from normal vehicle behavior during their peak testing capacity. Random Forest operates as an ensemble learning framework which employs multiple decision trees to produce the advantages of a single decision tree while preserving its ability to process multiple decision trees at once. The system will obtain detection capabilities which enable it to function as a real-time system that protects smart vehicle networks from evolving cyber attacks which hackers use to conduct their operations. The model will achieve high accuracy and reliable threat identification results while processing extensive data from vehicle systems.

Model Training with RF:

The RF model training process beginning with dataset preparation needs us to use all vehicle communication data. The complete dataset must be divided into two sections to assess the model performance on unfamiliar data: one section will be used for training and the other section will serve for testing purposes. The Random Forest algorithm during training will create multiple decision trees which will be built through random selection of features and data points and the trees will differ from each other.

$$fRF(x)=sign(T1t=1\sum Tht(x))$$

2.Gradient Boosting:

Today vehicles use an Intrusion Detection System which employs Gradient Boosting technology to track their complete power use and all operational capacity metrics. The first step of Gradient Boosting proceeds to estimate multiple weak learners which are mainly decision trees before combining those learners into a single strong model. The system uses Gradient Boosting for tree error correction which helps the model to recognize complex data patterns that establish it as the best method for detecting advanced attack techniques. The system's main advantage provides precise anomaly detection which decreases smart vehicle network vulnerability to DDoS attacks and Fuzzy attacks and Impersonation attacks and other cyber threats.

Gradient Boosting for Model Training:

The first stage of model training begins after the dataset creation process completes because this dataset will supply the necessary data for constructing the Gradient Boosting model which will be used in the IDS system. The data

structure contains three components which include Message_ID Byte-level signals and a target that shows the data's classification category. The dataset has two partitions which the model needs to evaluate its performance because one partition serves for training while the other serves for testing. The Gradient Boosting algorithm uses decision trees that it creates through multiple iterations to develop weak learners which built upon the errors of previously trained models. The system uses new tree introduction to fix previous model errors which results in improved prediction accuracy throughout the process. The system employs hyperparameters to control three system elements which consist of tree quantity learning rate and tree maximum depth.

$$Fm(x)=Fm-1(x)+\eta \cdot hm(x)$$

Where:

- $Fm-1(x)$ is the prediction from the previous trees.
- $hm(x)$ is the new tree that attempts to correct the errors.
- η\etaη is the learning rate, controlling the contribution of each new tree.

3.Adaboost:

The very radical proposal of integrating AdaBoost (Adaptive Boosting) with the smart vehicle's Intrusion Detection System (IDS) has significantly underlined the possibility of the system being able to rapidly and precisely detect and differentiate cyber-attacks without too much interruption to the identification of usual vehicle traffic. AdaBoost belongs to learning techniques that cooperate weak classifiers and thus, a strong one capable of coping with the hard patterns is produced. It is in this double role that AdaBoost, on the one hand, provides the model with the good power by increasing the weight of the almost impossible separation instances, and on the other, reveals the data points that up to now have been poorly taken care of. This whole process is performed in rounds which gives AdaBoost the advantage of spotting both the overt and the covert attack behaviors in the vehicle network like DDoS, Fuzzy, Impersonation, etc. So, the strategy is to exploit the power of AdaBoost in the classifier performance enhancement and thus make the IDS more trustworthy in real-time threat detection.

Adaboost for Model Training:

The complete AdaBoost model process for IDS starts with the dataset which contains data most similar to vital vehicle communication data requirements which include Message_ID and Byte-level signals and attack classification label among other elements. The dataset is divided into two sections which consist of a first section used for model training purposes and a second section dedicated to testing which ensures model functionality testing. The AdaBoost procedure starts at each node with classifiers training which typically uses shared resources to correct weaknesses from previous classifiers through data point weight adjustments.

4.LSTM:

The smart vehicle IDS has been incorporating LSTM networks only due to the car's inability to network data traffic in the first place. However, here LSTM has lifted the entire system from relegation to the classic machine learning models. Rather it has turned out to be somewhat of an RNN that is capable of handling time-series data plus memory and handing data through time. Hence, it is very much suitable for the detection of changing anomalies. In the networks of smart cars, the detection of dependencies over long periods becomes really crucial for LSTM since the DDoS, Fuzzy, and Impersonation attacks can last for a long time as well. To be more specific, the very ultimate goal is to detect such actions with the least possible latency and highest accuracy by way of LSTM's features of being able to recognize patterns and changes, thus allowing the system to defend the automotive communications from cyber threats while having a negligible false alarm rate.

Model Training with LSTM:

The initial step in the modeling of an LSTM model for smart vehicle IDS is dataset preparation which is usually defining the communication features of the vehicle such as Message_ID, and Byte-level signals, and also the labels indicating the type of attack classification the data belongs to. The next thing that follows is the partitioning of the dataset into three parts: training, validation, and testing. This division makes it possible for the model to test its ability to generalize on the new data that it hasn’t been exposed to before. Since LSTM models are highly recommended for time-series data, the data needs to be reshaped in such a manner that it is in sequences showing the temporal flow of messages in the vehicle.

5.CatBoost:

The combination of CatBoost (Categorical Boosting) and an Intrusion Detection System (IDS) for intelligent cars is basically an effort to utilize the great power of data classification and the discovery of complex feature

interactions to the utmost in the detection of attacks process. CatBoost, the most powerful gradient boosting method, is largest the choice for big data sets with categorical features, which is the main type of communication in cars. One of the pros of CatBoost is that it can take care of the categorical features on its own, meaning that the big data preprocessing step is significantly lessened, and thus it gets the ranking among the smart car IDS concerns. Not only will the CatBoost model be trained to identify cyber-attacks like DDoS, Fuzzy, and Impersonation, but it will also perform the classification, and legitimate vehicle communication will be separated. Consequently, the system benefits from its capabilities and the large scale that assist in preparing the vehicle networks against the continuously evolving and more intricate cyber threats, thus enabling them to accurately and swiftly detect the threats.

Model Training with CatBoost:

The first step to prepare the dataset for IDS CatBoost model training which uses car communication data needs to start with cleansing the dataset. The dataset is divided into two separate components which will serve as training materials and testing materials for upcoming Model testing which will need to assess performance on unknown data. The method used for dataset division involves creating two separate sections through random partitioning. CatBoost provides a major advantage because it can process categorical data with high efficiency.

V. RESULTS and DISCUSSION

The Smart cars became the first production vehicles to implement an Intrusion Detection System which successfully detected and classified all major cyber threats that automotive companies considered vital. The IDS system demonstrated its capability to distinguish between legitimate network connections and prohibited network connections. The IDS developers used different models which included Random Forest Gradient Boosting Adaboost LSTM CatBoost. The models were applied to analyze the data that was produced by the vehicle communication; one of the datasets was the CAN-intrusion-dataset which contained Message_ID Byte-level signals and target labels as its main components.

The different models used for intrusion detection required separation because it served as an essential component for detecting intrusions. Some of these models were trained on various datasets including the CAN-intrusion-dataset which contains the Message_ID Byte-level signals and target labels.

The testing results showed that the models could identify threats with great success while Random Forest and CatBoost maintained their best performance through precise and complete detection. The LSTM time series model demonstrated exceptional capabilities for finding time-based and sequential pattern anomalies which makes it suitable for detecting changing attack patterns. The classifiers Gradient Boosting and Adaboost demonstrated their ability to detect advanced attack methods which enabled them to function as an all-encompassing cyber threat detection system (*Figure 2-7)*.

Model	Accuracy Score	Precision Score	Recall Score	F1 Score
Random Forest	0.9412	0.9412	0.9412	0.9412
Decision Tree	0.9419	0.9419	0.9419	0.9419
Gradient Boosting	0.9212	0.9212	0.9212	0.9212
Adaboost	0.8193	0.8193	0.8193	0.8193
Catboost	0.9491	0.9491	0.9491	0.9491
LSTM	0.8800	0.8700	0.8700	0.8700

Table 1: Classification Report of Algorithms

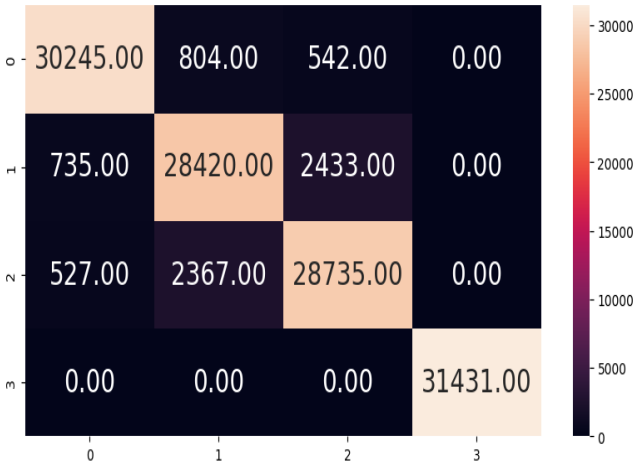

Figure 2: Confusion Matrix of Random Forest

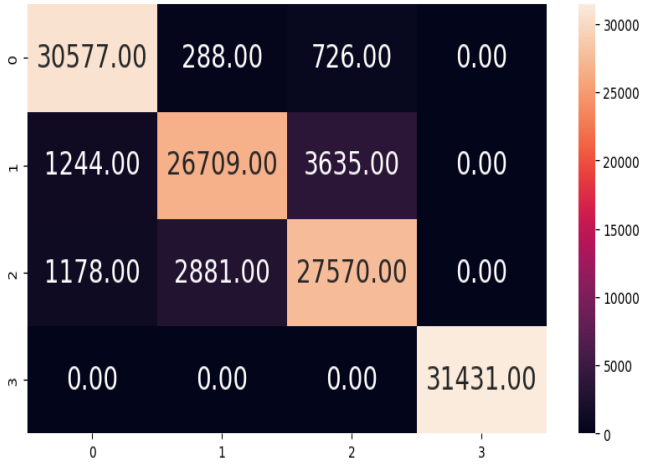

Figure 3: Confusion Matrix of Gradient Boosting

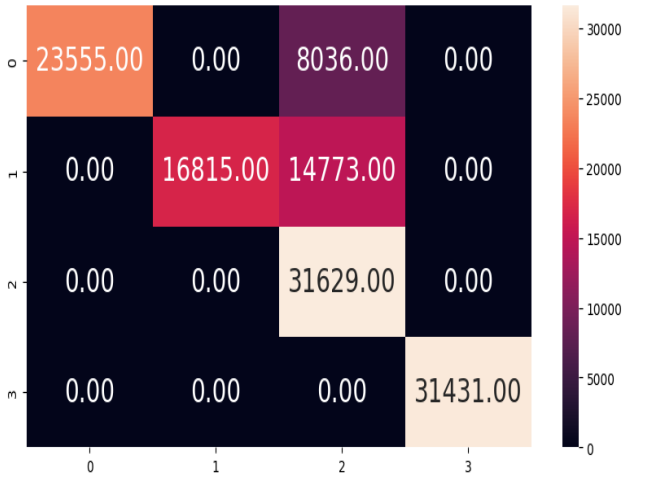

Figure 4: Confusion Matrix of AdaBoost

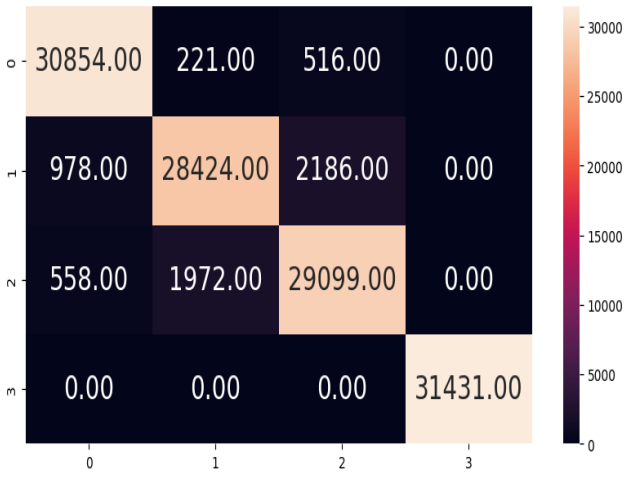

Figure 5: Confusion Matrix of CatBoost

VI. CONCLUSION and FUTURE SCOPE

The IDS system for smart vehicles used multiple security technologies which protected network operations and vehicle systems from all forms of cyber threats. The IDS system used machine learning techniques which included RF and Gradient Boosting and Adaboost and LSTM and CatBoost to identify DDoS and Fuzzy and Impersonation attacks while differentiating these attacks from normal user behavior. The researchers studied vehicular communication patterns through their examination of Message_ID and Byte-level data from the CAN-intrusion-dataset which established an accurate threat detection system base. The Random Forest and CatBoost models achieved the highest performance results among all tested models which produced strong accuracy results together with precision and recall measurements. The system provided two functions which enabled users to detect attacks in real time while running at maximum speed and it could grow to support additional tasks. The LSTM model required multiple system resources because it enabled the system to adapt better to new attack patterns. The project required extra work to achieve its essential improvements even though it had produced successful results. The system would gain better vehicle communication data processing abilities through deep learning methods which include CNNs and transformers because these methods establish superior processing capabilities. The system will improve vehicle behavior detection through traffic data from nearby vehicles and external sensor networks while it will also detect unauthorized state changes across legitimate attacks. Vehicles can share threat intelligence through federated learning which allows them to protect user privacy while they operate across wider areas of their operations. The IDS system will acquire better functions through these methods which will boost its ability to combat future threats in autonomous vehicle networks.

REFERENCE

[1] R. Rai, J. Grover, P. Sharma, and A. Pareek, "Securing the CAN bus using deep learning for intrusion detection in vehicles," Scientific Reports, vol. 15, no. 13820, pp. 1–13, 2025.

[2] N. Singh and R. Agarwal, "Intrusion detection system for smart vehicles using machine learning algorithms," International Journal of Scientific Research and Technology, vol. 14, no. 3, pp. 1–8, 2024.

[3] C. Anthony, "Intrusion detection system for autonomous vehicles using non-tree-based machine learning algorithms," Electronics, vol. 13, no. 5, p. 809, 2024.

[4] V. Tanksale, "Intrusion detection system for controller area network," Cybersecurity, vol. 10, no. 195, pp. 1–11, 2024.

[5] H. Yang, "A deep learning based intrusion detection system for CAN bus packet," Scientific Reports, vol. 15, no. 13820, pp. 1–13, 2025.

[6] P. Wei, "A novel intrusion detection model for the CAN bus packet," Procedia Computer Science, vol. 185, pp. 123–130, 2023.

[7] B. Lampe, "A survey of deep learning-based intrusion detection in vehicular networks," Computers & Security, vol. 132, p. 103524, 2023. [Online]

[8] J. Nagarajan, "Machine learning based intrusion detection systems for vehicular networks," Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 3, pp. 1–15, 2023.

[9] M. K. Devnath, "GCNIDS: Graph convolutional network-based intrusion detection system for CAN bus,", 2023.

[10] T.-N. Hoang and D. Kim, "Detecting in-vehicle intrusion via semi-supervised learning-based convolutional adversarial autoencoders," 2022.

[11] M. H. Shahriar, "CAN-Shield: Deep learning-based intrusion detection framework for controller area networks at the signal-level," 2022.

[12] A. Sebastian et al., "Enhancing intrusion detection in Internet of Vehicles through federate, 2023.

[13B. Xu, "BEPCD: An ensemble learning-based intrusion detection framework for in-vehicle CAN bus," PMC, 2025.

[14] F. Luo, "Intrusion detection systems for in-vehicle networks," Sensors, vol. 23, no. 7, p. 3610, 2023.

[15] L. Zhang, "Securing in-vehicle networks with intrusion detection systems," University of Michigan, 2023.

