Intrusion Detection System For Smart Vehicles Using Machine Learning Algorithms

Paruchuri Meghana
UG Scholar, Mohan Babu University, Tirupati, A.P, India,
meghanaparuchuri23@gmail.com

Maddana Bala Narasimhulu
UG Scholar, Mohan Babu University, Tirupati, AP, India,
balu68900@gmail.com

Kothamasu Mitesh
UG Scholar, Mohan Babu University, Tirupati, AP, India,
kothamasumitesh@gmail.com

Ravilla Harshath
UG Scholar, Mohan Babu University,, Tirupati, AP, India
Ravillaharshath1@gmail.com

Dr.Cuddapah Anitha
Associate Professor, Department of CSE, School of Computing, Mohan Babu University,, Tirupati, A.P, India,
cuddapah.anitha@mbu.asia

***Abstract*— Main feature of the research is the proposal of a novel Intrusion Detection System (IDS) for autonomous cars based on the newest machine learning techniques. The main task for the IDS will be the detection of cyberattacks and their classification which could include different types like DDoS, Fuzzy, Impersonation, and ordinary "Free" traffic. The whole process of building and testing models is based on CAN-intrusion-dataset which describes the communication among vehicles in terms of Messaged, Byte-level signals, and Target labels. Besides, some of the machine learning methods such as Random Forest, Gradient Boosting, Adaboost, LSTM, and CatBoost will be combined to perform the task of detecting and preventing threats. Thus, the identification of unauthorized persons trying to access vehicle networks will not only be very secure but also very adaptive because of the complete usage of these algorithms. Therefore, the security and reliability of the smart vehicle systems will be enhanced. In a nutshell, the development of a detection system has the intention that it would not just have a capability of scaling but at the same time also be able to guard against the ever-increasing cyber threats targeting the smart vehicles.**

Keywords— Random Forest, Gradient Boosting, Adaboost, LSTM, CatBoost classifiers.

I. INTRODUCTION

The introduction of smart technologies in the form of new ones has resulted in the modern automobiles becoming vulnerable to hacking and cyber attack which in turn made the auto industry one of the major areas of this type of crime. The high-tech cars are very much connected and reliance upon communication networks like Controller Area Network (CAN) is quite substantial and that is the reason they are open to such dangers which might cause the car to become incapacitated, the driver and passengers to be killed or even personal data to be leaked. 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.

An Intrusion Detection System (IDS) is a crucial element in the protection of a vehicle network, as it detects the malicious activities and consequently keeps the system from getting the cyber attacks. However, traditional IDSs were designed for generic IT networks whereas vehicular networks differ in some respects thus requiring the use of different techniques to solve the problem effectively. Hence, the project under discussion aims at the devising of an IDS for the state-of-the-art vehicles by means of the latest machine learning technologies for detection and categorization of attacks.

According to the proposal, the network security system is designed to identify DDoS (Distributed Denial of Service), Fuzzy, and Impersonation attacks, and also to mark the usual traffic as “Free” traffic. The progress of the system is being devised on the CAN-intrusion-dataset, which is made up of the vehicle communications data—Message-ID, Byte-level signals, and target labels—and also supplies the inputs for the attack classification. The classification methods under consideration for the study are Random Forest, Gradient Boosting, Adaboost, LSTM, and CatBoost that are going to be used for the vehicle-feature-based threat identification and grouping. These powerful algorithms being implemented will enable the system to detect anomalies very quickly thus giving the smart vehicle network a strong defense against the new cyber threats that are trying to get into the system undetected. This research has demonstrated that the utilization of machine learning techniques will not only be a highly efficient and scalable solution for the security of the smart transportation systems in the future but also a significant boost to the existing vehicular networks.

II. RELATED WORK

There has been a gradual shift in focus from traditional detection systems to machine learning based ones in the recent years. In fact, one paper in the year 2024 illustrated how non-tree-based machine learning methods such as K-nearest neighbors and ensemble learning performed just as well as their more complex counterparts and reported the creation of a strong prototype of an IDS capable of detecting the unauthorized access in the driverless cars. The goal was to improve detection accuracy while simultaneously reducing computing power requirements.

The investigation of vehicle communication and intrusion detection systems was triggered by the use of the controller area network (CAN) bus as the main research focus, which is the most widely adopted standard for vehicle communication. The 2024 paper regarding CAN buses for unsupervised intrusion detection presented an extremely novel approach of applying machine learning principles for abnormal behavior detection without the need for labeled datasets. This approach could be very useful in scenarios where there is little to no labeled attack data available.

To tackle such a difficult detection problem, deep learning models were used. The intrusion detection system (IDS) developed in 2023 using deep convolutional neural networks (DCNN) proved to be very effective. DCNN not only requires much less time for the detection of cyber threats but also automatically retrieves the significant features of the network traffic data, thus leading to a more precise detection of the cyber threats.

The use of methods of interpretability from AI (XAI) in IDS systems has been recommended as a means of increasing the detection's transparency and trust. 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. It converts the messages sent over the CAN bus into signals which are highly intelligible for humans thus, not only facilitating detection but also indicating which part of the vehicle is at risk.

The idea behind federated learning is that the different vehicles can communicate with each other to determine if there is a case of intrusion without having to expose the individual units' data. A bunch of researchers in 2023 invented a federated learning framework for the IoV (Internet of Vehicles) that not only employs the SMOTE technique to cope with class imbalance and the outlier detection technique for the model performance boost but also ensures data confidentiality

III. PROPOSED METHODOLOGY

The first system intends primarily to develop a smart vehicle's Intrusion Detection System (IDS), which will utilize machine learning algorithms for the detection and classification of different kinds of attacks. For this research, the car's communication data will be used, comprising the CAN-intrusion-dataset. This dataset contains three main features: Message_ID, Byte-level signals, and target labels, which will all work together to help the system recognize anomalies. The machine learning methods to be applied are Random Forest, Gradient Boosting, Adaboost, LSTM, and CatBoost, and they will be able to tell apart normal traffic from DDoS, Fuzzy, and Impersonation attack types. The aim is to have the IDS designed for immediate deployment, so it provides excellent threat detection while also protecting and fortifying the car networks against the ceaselessly evolving cyber threat landscape. Thus, performance and scalability were the key characteristics to be considered and the result is a very efficient and scalable system.

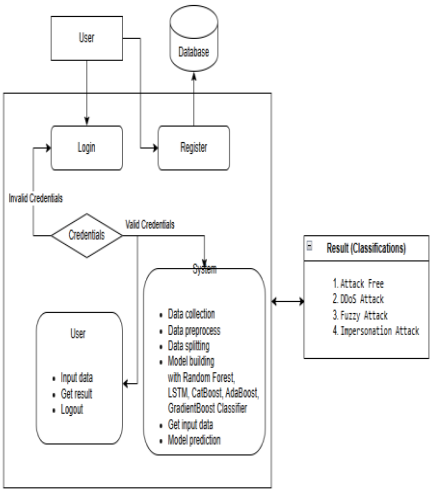

Fig 1: Architecture of Methodology

IV. EXPERIMENTAL RESULTS & ANALYSIS

Random Forest:

Currently, employing the Random Forest algorithm in the operation of smart vehicle Intrusion Detection System (IDS) is primarily focused on very precise detection, along with a mature classification covering the gamut of cyberattacks like Distributed Denial of Service (DDoS), Fuzzy, Impersonation and so forth. Besides, the retention of differences between normal vehicle traffic at the highest resolution is also one of their abilities. Random Forest, which is an ensemble learning technique composed of numerous decision trees, provides the same benefit as a single tree and at the same time avoids the problem of overfitting thereby making it the unbeatable way to boost the power of generalization. Also, the proposed system is about to possess not only the detection capability but will be operating in real-time thus providing a smart vehicle network that is safe and uninterrupted to the constantly evolving cyber attackers who are always on the lookout for new ways to launch their attacks. The model will be very accurate and reliable in identifying threats and at the same time dealing with large amounts of vehicle data.

Model Training with Random Forest:

Preparing the dataset is the first step in the Random Forest model training process which in our case is the vehicle communication data with Message_ID, Byte-level signals, and target labels designated for attack classification, and so on. To evaluate the model on new data, the entire dataset has to be divided into two parts: one for training and the other for testing. The Random Forest algorithm during the training phase will generate several decision trees that will each be formed upon the random choice of features and data points and the variety among the trees will be one of the factors.

Gradient Boosting:

The modern vehicle Intrusion Detection System (IDS) that uses Gradient Boosting has been invented for the purpose of not only detecting but also monitoring the system's power and endurance. To put it simply, Gradient Boosting is recognized as one of the most powerful ensemble learning methods that initially gives the estimate of a large number of weak learners—generally, decision trees—and then merges them into a single strong one. The mistakes of the previous tree are

corrected one at a time by Gradient Boosting, which enhances the model's capability to uncover complex patterns in the data thus making it the right choice for the sophisticated detection of the attack behaviors. One of the most important benefits is that there is an accurate anomaly detection which results in the smart vehicle networks becoming less vulnerable to various cyber threats like DDoS, Fuzzy, and Impersonation types, etc.

Gradient Boosting for Model Training:

The procedure of 'model training' has its starting point in the creation of the dataset, which is going to be used for the Gradient Boosting model that is going to be deployed for the IDS system. It contains Message_ID, Byte-level signals, and a target for classification that indicates the category to which the data belongs. For the purpose of fair assessment of the model's performance, the dataset is divided into two parts: one for training and the other for testing. The Gradient Boosting algorithm is based on a series of iterations, where the decision trees (weak learners) are fitted one by one to correct the errors made by the previously trained models. Every new tree is doing the work of eliminating the errors in the earlier models thus improving the quality of the predictions at every step. Hyperparameters such as the number of trees, learning rate, and maximum depth of trees are efficiently controlled.

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 whole process of the AdaBoost model for IDS begins with the dataset, which is the one that is most similar to the selected vehicle communication data of utmost importance, such as Message_ID, Byte-level signals, and attack classification label, etc. Then, the dataset is split into two parts - the first part is for training the model and the second part is for testing which assures the model's usability. The AdaBoost procedure is the following: at each node classifiers are trained that usually socialise and through average resources apply the weaknesses of the previous one, thus changing the weights of the data points that were classified.

LSTM:

The smart vehicle Intrusion Detection System (IDS) has been incorporating Long Short-Term Memory (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.

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 cleansing of the dataset is the first stage in the process of preparing the dataset for the IDS CatBoost model training that consists of car communication data like Message_ID, Byte-level signals, and the target labels for the

classification tasks. After that, the dataset is split into two parts - a training set and a testing set - which are subsequently used to evaluate the model's performance on unseen data. Random partitioning of the dataset into two parts is the method adopted. One of the primary benefits of CatBoost is its capacity to manage the categorical features very effectively.

.

V. Result

Smart cars were among the first mass-produced vehicles that had an Intrusion Detection System (IDS) installed to not just detect but also classify the different types of cyber-attacks such as DDoS, Fuzz, and Impersonation, which were the major concerns for the automotive sector. The IDS further had the capability to distinguish between the good and the bad traffic. The IDS developers incorporated various models like Random Forest, Gradient Boosting, Adaboost, LSTM, and 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 where the key factors were Message_ID, Byte-level signals, and the target labels.

Separation was considered a key factor in intrusion detection that necessitated the use of different models, among which were Random Forest, Gradient Boosting, Adaboost, LSTM, and CatBoost. 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.

During evaluation, the models exhibited very high detection rates and it was Random Forest and CatBoost that consistently showed the highest precision and recall. The time series model, LSTM, was brilliant in the detection of sequential and time-related anomalies; thus, it was concluded that it could be applied for dynamic attack patterns. Meanwhile, the classifiers Gradient Boosting and Adaboost proved to be very proficient in revealing complex attack scenarios which in turn made them part of the detector that was able to cope with the whole range of cyber threats.

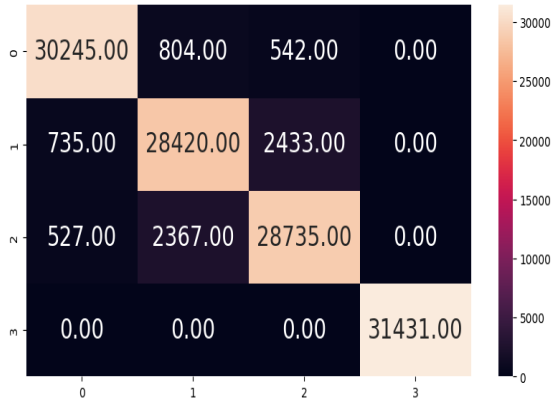

Fig 3: Confusion Matrix of Random Forest

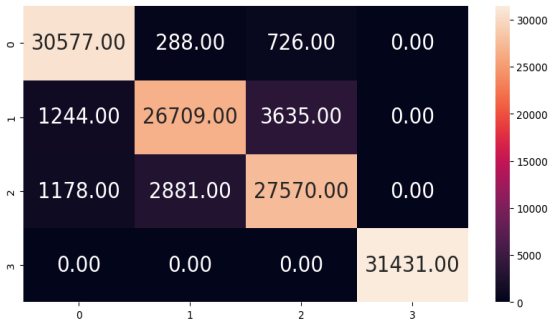

Fig 4: Confusion Matrix of Gradient Boosting

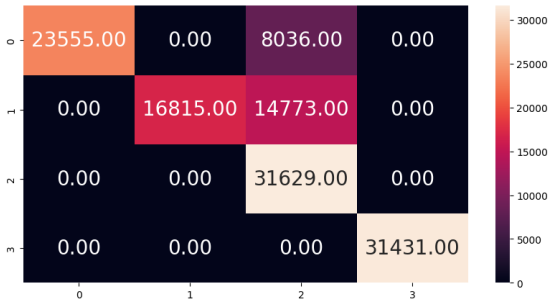

Fig 5: Confusion Matrix of Adaboost

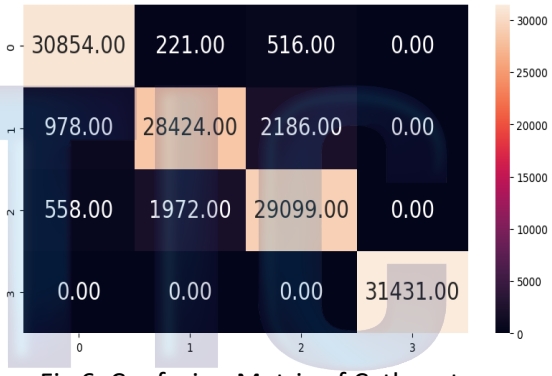

Fig 6: Confusion Matrix of Catboost

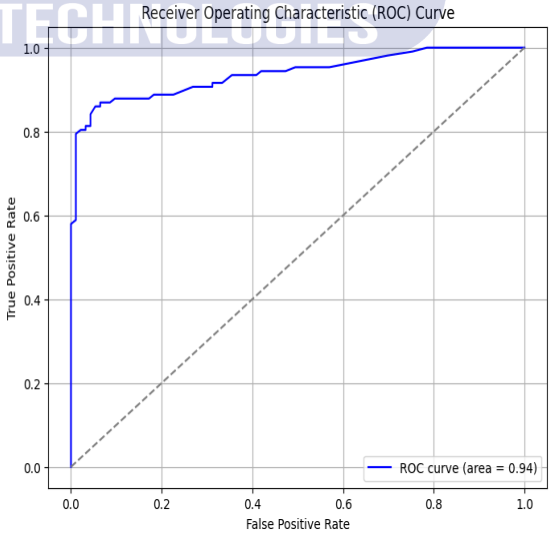

Fig 7: ROC of Random forest

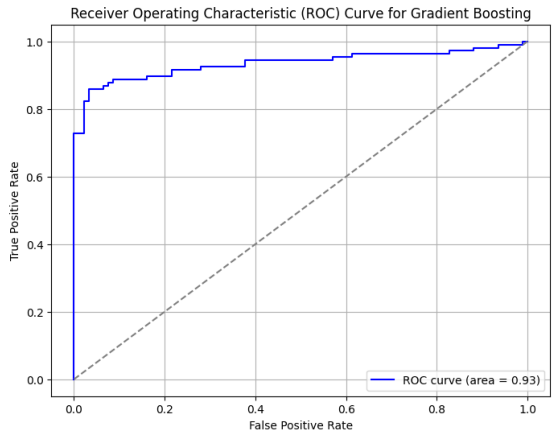

Fig 8: ROC curve of Gradient Boosting

VI. CONCLUSION

The intrusion detection system (IDS) designed for smart vehicles not only secured the network but also completely secured the vehicle system. The system had an upper hand, as it was based on machine-learning technologies, such as Random Forest, Gradient Boosting, Adaboost, Long Short-Term Memory (LSTM), and CatBoost, which eventually led to the successful detection and classification of cyber-attacks such as DDoS, Fuzzy and Impersonation. These attacks were, therefore, detected and separated from legitimate traffic. In addition, the combination of Message_ID and Byte-level signals to form the CAN-intrusion-dataset has aided in the thorough review of vehicular communication data, thereby providing a very strong foundation for threat detection with great accuracy.

Random Forest and CatBoost were indeed the best of the bunch out of all the models and the IDS achieved good accuracy, precision, and recall. For one thing, the IDS not only conducted the real-time attack detection but also promised that the system was running at the best possible level with no compromise on either speed or scalability. Moreover, the resource-intensive LSTM model was used which, however, did enhance the system's capability to defend against the evolving attack patterns in a more sophisticated manner

.

VII. FUTURE ENHANCEMENT

The intelligent vehicle anomaly detection system (IDS) has achieved to a great extent the automatic detection of all forms of cyber-attacks and is now operating with a prohibitively high identification capability. However, a key factor to be considered is that the whole system's performance will be adjusted to its maximum limit by applying the advancements rather than just applying the improvements. One such suggestion that has recently been made is the adoption of state-of-the-art deep learning approaches such as CNNs and transformers. This will allow the system to handle and analyze huge volumes of vehicle communication data. By implementing the latest and most advanced techniques, the IDS will be able to uncover even the most covert intruders and at the same time, it will lead to a substantial reduction in the occurrence of false alarms.

Moreover, the system could potentially tap into a traffic data source from car systems in the proximity or outside sensor networks and this would be considered as an additional layer of support to the system. The combination of large and complex data sources with deep analyses will certainly result in precise vehicle behavior detection which will be of great help in distinguishing between unauthorized state changes of a vehicle and possible attacks. Furthermore, with the advent of federated learning, it may turn out that the IDS and various vehicles will transfer threat intelligence amongst each other without compromising privacy or necessitating a centralized data storage, thus, greatly extending the detection area to the reach of the geographical coverage of the smart vehicle network

VIII. REFERENCES

[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," arXiv, 2023.

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

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

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

[13]. B. 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