International Journal of Transport Development and Integration
Vol. 8, No. 4, December, 2024, pp. 605-612
Journal homepage: http://iieta.org/journals/ijtdi

An Efficient Intrusion Detection System in IoV Using Improved Random Forest Model

Deepthi Reddy Dasari*, Himabindu Gottumukkala

Department of Computer Science and Engineering, GITAM University, Hyderabad 502329, India

Corresponding Author Email: draavi@gitam.in

https://doi.org/10.18280/ijtdi.080412

Received: 18 April 2024
Revised: 25 October 2024
Accepted: 18 November 2024
Available online: 26 December 2024

Keywords:
Internet of Vehicles, machine learning, multi-class classification, Intrusion Detection System, cyber security

ABSTRACT

Modern cars use a hierarchical system of sensors, controlling devices, and controllers, linked via various intra-vehicle systems, to regulate and monitor the vehicle's status. Researchers have confined numerous academic papers on intrusion detection in the Internet of Things (IoT), employing data mining and machine learning (ML) techniques to secure autonomous vehicles and detect potential attacks. To identify malicious attacks on the Internet of Vehicles (IoV), however, a competent and accurate method is required. This paper presents a model for cyber-attack detection in IoV that employs tree-based ML methods, an Improved Random Forest Classifier (IRFC), and Extra Tree (ET). We build the proposed model using Improved Random Forest (IRF) and ensemble learning techniques. The proposed IRF model employs optimized feature selection and tuning strategies to enhance intrusion sensitivity and decrease false positive rates. We evaluate the proposed model's performance using the CI-CIDS 2018 dataset. Also, this work focuses mostly on the reduced feature selection and ensemble learning (EL) methods to get a high detection rate while keeping the cost of computing low. The test results show that the proposed method can find DDoS attacks and vehicle intrusions with a 0.99 accuracy rate.

1. INTRODUCTION

In the last decade, automobile manufacturers' quick adoption of different new technologies has altered the design and functionality of new vehicles (i.e., cars). Security concerns are growing in lockstep with the rapid expansion of smart car connections. An attack on the infrastructure of the Internet of Vehicles (IoV) can make vehicles less reliable and potentially lead to accidents. The World Health Organization (WHO) announced in June 2021 that vehicle accidents [1]. In a notable instance, multiple hackers successfully hijacked a car, took control of the driving and controls, and carried out deadly operations [2]. Traditional Vehicular Ad Hoc Networks (VANETs) are rapidly transforming the way we engage with IoV, devices, and infrastructures. VANETs enable communication systems between cars and equipment in Intelligent Transport Systems (ITSs), thereby transforming the vehicles mobility [3]. The transportation industry is a promising and evolving field that presents an ideal option for reducing traffic accidents and associated expenses [4]. Several of these devices, meanwhile, lack security features like gateways and proxies [5]. Since attacking or fraudulently managing automobiles on the highway constitutes a substantial risk to human life, AVs are vulnerable to network attacks with serious consequences. The following attacks are examples of potential networking risks. Current networks frequently face Denial of Service (DoS) attacks, which can severely damage the network's resources [6]. There are numerous techniques, including location spoofing, to impersonate legitimate users and deliver false GPS information to the nodes. A "port scan attack", also known as probing, is an alternative attack that could steal private information from a vehicle's systems and users [7]. Intruders, in particular, use brute-force attacks, SQL cross-site scripting (XSS), and injection attacks to get access to the web interfaces of vehicles or computers [8]. All of the aforementioned risks and threats require a powerful defense system that can repel potential attacks and facilitate inter- and intra-AV system communication. An "Intrusion Detection System" (IDS) is necessary to monitor network activity and detect abnormal traffic. Refining the reliability of IDS will minimize the number of false alarms [9]. Traditional IDSs struggle to improve recital and determine unseen attacks. ML approaches enable excellent automation for automatic detection computers. Furthermore, ML approaches offer broad potential for detecting unknown attacks. IDS is a good security method for finding suspicious data and attacks in network traffic when cars and other devices talk to each other [10]. Conventional Intrusion Detection Systems (IDS) often encounter difficulties in real-time anomaly identification owing to the substantial amount and complexity of Internet of Vehicles (IoV) data. Consequently, there is an urgent need to develop an effective Intrusion Detection System using machine learning models capable of rapidly and correctly detecting malicious activity in IoV's situations. An Improved Random Forest (RF) model, tuned for higher precision and low latency, provides a workable way to improve IDS in the IoV, leading to better classification performance and resilience.

In this current study, we employ ML-based classifier methods such as Extra Tree (ET) and Random Forest (RF). A framework IDS must have a substantial detection rate as well as a minimal computational cost. To boost accuracy, an ET, notably stacking, is utilized to save computing time. The study compares an Improved RF model for detecting intrusions with traditional IDS methods in terms of accuracy, latency, and computational efficiency. It also assesses the model's robustness in a simulated IoV environment using the CI-CIDS 2018 dataset, revealing its excellent performance.

We organize the remainder of this paper as follows: Section 2 gives an overview of related works on IoV. Section 3 presents the working of the proposed framework for IDS. Section 4 discusses the experimental and simulated results. Finally, section 5 presents the conclusion of the work.

2. RELATED WORKS

The attacker gains full access to the vehicle upon entering, enabling them to perform risky acts. In two-class and multi-class data sets, an imbalance occurs when the samples of one class include more instances than the samples of the other classes. Most traditional ML algorithms underperform in these datasets since they prefer the majority class, resulting in poor predicted performance over the minority class. This section highlights some of the most recent scientific breakthroughs in intruder detection in IoV. In the study [11], the authors proposed an FPGA-based intrusion detection approach that not only enables real-time scan capability but also finds application in a vehicle environment. They tested the suggested system on a Xilinx FPGA platform. The tests indicate that the suggested system could surpass 39 Gbps on an original FPGA platform, marking a 15% increase over current performance.

Current automobiles use the Controller Area Network (CAN) as a key mechanism to direct the interaction between the Electronic Control Units (ECU) of the in-vehicle systems, as explained by Ahmed et al. [12]. They mentioned that authorization and verification techniques are required to protect the infrastructure against cyber or malicious attacks, such as DoS and fuzzy attacks. Later, the authors describe an IDS based on deep learning architecture to secure CAN bus vehicles. They trained the VGG architecture on network attack patterns to detect malicious attacks, and the tests used the CAN-intrusion dataset. Yang et al. [13] studied intra-vehicle and exterior security breaches to recognize both recognized and unidentified vehicle threats. Tests show that the suggested system can correctly identify all known types of attacks with 99.99% accuracy on the CAN-intrusion dataset, which shows data about the inside network of the vehicle, and 99.88% accuracy on the CICIDS2017 dataset, which shows data about the outside network of the vehicle. Using CNNs and hyper-parameter optimization approaches, the authors developed EL-based IDS for IoV in the study [14]. The suggested IDS, which utilizes open benchmarking data sources such as CICIDS2017 and Carhack, demonstrated an accuracy of 99.2%. The authors also showed the reliability of the IDS for finding attacks. To build TCAN-IDS, an in-vehicle network intrusion detection model, Cheng et al. [15] proposed a temporal CNN architecture with global attention.

The proposed system original message, including an arbitration bit and a data field, into a message matrix that mirrors transmissions from a specific point in time. The model then extracts the spatial-temporal detail features. Importantly, the global attention continued to focus on significant regions based on multichannel and local pattern values, while disregarding minor changes in bytes. Lastly, a binary class element monitors abnormal traffic.

According to research, machine learning models are becoming more popular for finding intrusions in the Internet of Vehicles (IoV) because they can handle complex and high-dimensional data [16, 17]. Researchers have employed a variety of machine learning models, such as Support Vector Machines (SVM), K-Nearest Neighbours (KNN), Decision Trees, and Random Forests, to enhance the detection ability in IoV. For instance, Random Forests stand out for their ability to withstand overfitting and manage high-dimensional data, making them suitable for IoV applications [18]. For effective real-time intrusion detection, these models often need optimization for both speed and accuracy [19].

Research indicates that enhanced RF models may accommodate various intrusion types, making them adaptable for IoV security applications [20]. Nonetheless, deep learning models need substantial processing resources, rendering lightweight but optimized models, such as Random Forests with calibrated parameters, a preferred option for real-time Internet of Vehicles systems [21]. Real-time Intrusion Detection Systems in the Internet of Vehicles need models adept at managing substantial data volume and diversity with little latency. This need has prompted investigations into lightweight and efficient models, particularly those tailored for low-latency performance [22]. In the study [23], the authors suggested integrating ML models with incremental learning approaches to improve detection accuracy and adaptability in dynamic Internet of Vehicles environments.

This article discusses about three important studies that look into Intrusion Detection Systems (IDS) in the Internet of Vehicles (IoV). It focuses on ensemble learning and feature selection as ways to make IDS work better.

Wang et al. [24] observed an ensemble learning method that combines with optimized feature selection. They required to make Intrusion Detection Systems in Internet-of-Vehicles networks more accurate and efficient. Their approach combines Random Forest (RF) and Gradient Boosting Machine (GBM) classifiers with feature selection based on Principal Component Analysis (PCA). This cuts down on the number of dimensions while keeping important data. The model got better detection rates and less computational overhead by focusing on traits with high impact. This is crucial in IoV scenarios where resources are scarce. The results of this study showed that feature-reduction-based optimized ensemble models work well for real-time applications in IoV systems, providing high accuracy with low latency.

Tu and Shang [25] presented an optimized ensemble Intrusion Detection System (IDS) that uses soft voting to identify cyber threats in intelligent transportation systems' internal and external networks. The system uses three machine learning techniques: logistic regression, random forests, and decision trees to construct an integrated structure. The CICIDS2017 dataset and Car-Hacking dataset are used for external network communication and in-vehicle communication evaluations. In the study [26], the authors introduced an Intrusion Detection System for the Internet of Vehicles using ensemble learning, emphasizing the minimization of false-positive rates via efficient feature selection. Their model included AdaBoost and bagging with recursive feature elimination (RFE), which systematically identifies the most relevant features to improve the model's interpretability and performance. They tested the model with a real-world dataset of the Internet of Vehicles and found that it greatly improved detection accuracy and decreased false alarms. This shows the benefits of combining ensemble learning with RFE for IDS applications in IoV, where rapid and accurate detection is very important.

3. MATERIALS AND METHODS

Electric vehicles offer a significant advantage due to their eco-friendly features, unlike traditional motor vehicles with internal combustion engines [27]. Figure 1 summarizes the suggested IDS outline for developing a high detection model. The proposed model involves collecting sufficient network activity data, using SMOTEboost to reduce unbalanced classes, selecting features based on average significance, constructing ET and IRF models for the stack ensemble classifier, and constructing a final classifier to differentiate between normal and attack traffic, thereby reducing computational costs.

Dataset: This study uses the CICIDS-2018 dataset to assess the proposed IDS. This study utilizes the CICIDS2018 benchmark dataset to capture BENIGN and contemporary network traffic attacks. The protocols included in this dataset, including email, HTTP, and HTTPS, also incorporate the most recent network attacks. The CICIDS2018 dataset tests the proposed framework against DDoS attacks, web attacks, and infiltration.

Data pre-processing: To develop an IDS, we must first significantly influence the network traffic patterns in both benign and attack states, caused by various types of attacks. We can obtain the data from CICIDS-2018, but it must include appropriate network properties, also known as network features, for the creation of an IDS. We will preprocess the obtained network data after a few stages to make it more suitable for IDS design. Normalized data, on the other hand, is frequently more efficient for ML training. In this study, we utilized the label encoder approach to convert categorical features to numerical values. Following the conversion, we employed Eq. (1) to normalize values between 0 and 1 using the min-max normalization approach.

$$D_{norm} = \frac{d - d_{min}}{d_{max} - d_{min}} \tag{1}$$

Synthetic Minority Oversampling Technique (SMOTE):

Generally, networks maintain a stable state with insufficient attack labels, leading to the classification of network data as imbalanced. Oversampling algorithms have made errors in handling complex problems, such as generalizations or not actively addressing imbalance issues in subspace, compared to two-class imbalance situations. In this work, we apply SMOTEBoost, a data-level approach for dealing with the issue of unbalanced data.

The proposed method's major phases are SMOTE sampling and boosting. As a data-level solution, this method employs the SMOTE approach. SMOTE generates additional minority class instances for a training dataset by locating a minority class instances k-nearest neighbors (KNN) and extrapolating between that example and its neighbors to generate new instances. It also handles binary and multi-class issues. The objective is to enhance the ensemble's accuracy rate by focusing on challenging minority class cases and improving True Positives (TP).

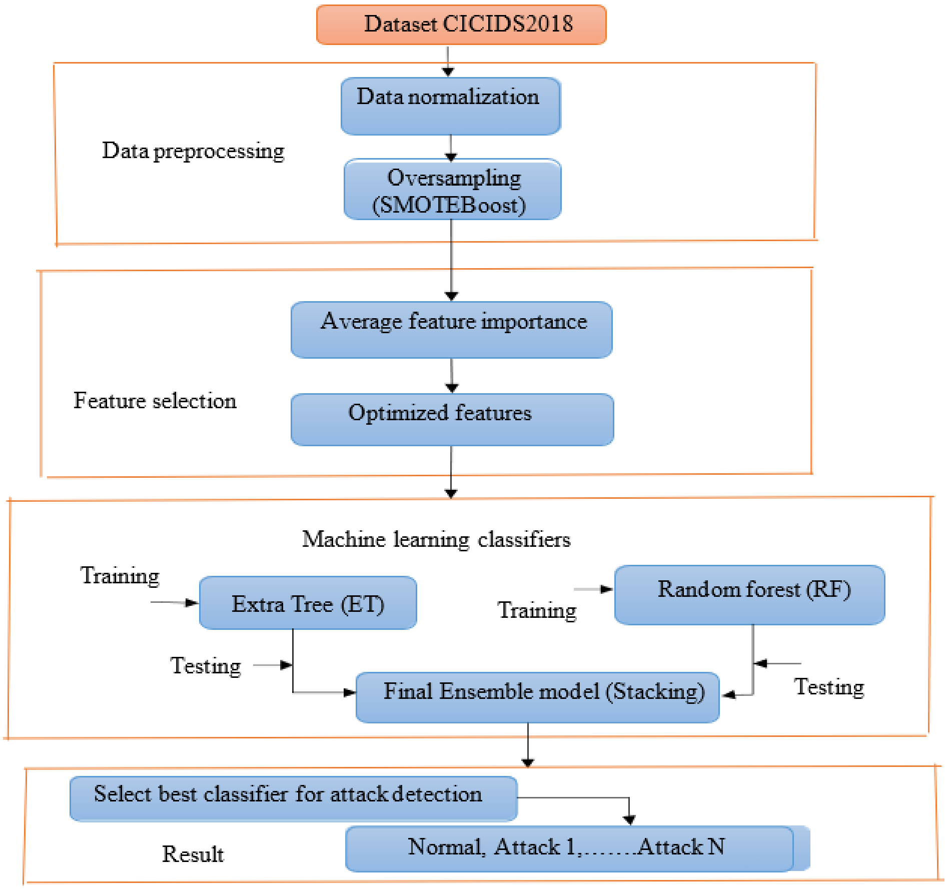
Dataset CICIDS2018
Data normalization
Data preprocessing
Oversampling (SMOTEBoost)
Average feature importance
Feature selection
Optimized features
Machine learning classifiers
Training
Extra Tree (ET)
Training
Random forest (RF)
Testing
Final Ensemble model (Stacking)
Testing
Select best classifier for attack detection
Result
Normal, Attack 1,.......Attack N

Figure 1. The proposed IDS framework

Feature selection:

It is critical to provide appropriate input to the model so that it can function successfully with the required data. Introducing redundant and consistent input may not aid prediction, but it does degrade prediction accuracy. As a result, there is a requirement to select the key features from a huge dataset and transfer them to the model. The dataset, as of CICIDS2018, has 80 features; we eliminated the class label to leave 79 features. We should reduce the 79 relevant features for an ML algorithm by retaining the significant ones and discarding the less important ones using Average feature importance in this paper.

Average feature importance:

By analyzing the average feature importance of the CICIDS 2018 dataset, we can determine which features are most effective in distinguishing between benign and malicious network traffic. Machine learning methods prioritize features based on their impact on correct classification. Feature names such as "Flow Duration," "Total Fwd Packets," or "Destination Port" could get high priority ratings if the model uses them extensively while generating decisions. In ensemble approaches, we average scores across several models, or across all trees in tree-based models, to determine average feature relevance. This averaging eliminates outliers and brings attention to characteristics that have a consistent impact. By concentrating on aspects with greater average relevance, we can simplify the dataset, remove less informative properties, and reduce computing demands without compromising accuracy.

To better understand which features (such as packet size, source/destination IP, timestamp, etc.) contribute most to differentiating between legitimate and malicious traffic, Intrusion Detection systems (IDS) in IoV need to know the average feature importance. We can use the feature priority scores to prioritize features that significantly impact intrusion detection accuracy. Some features of network packets, such as payload size and frequency, may hold significant importance in identifying outliers. We ignore features of lesser relevance to save computational burden without compromising model performance.

Classification:

Unsupervised learning is a ML approach that aids in uncovering hidden patterns and data within given datasets [28]. We compare different supervised classification techniques based on metrics like the highest detection accuracy and the least false negative predictions. Machine learning algorithms are capable of learning a huge number of malicious and benign inputs of various types and efficiently predicting them. Constructing the IDS in the proposed system to identify diverse cyber-attacks presents a multi-classification challenge, with ML methods commonly employed to address such classification problems. The selected machine learning procedures are tree-based, and they also incorporate additional trees and Random Forests.

Extra Tree:

This approach, a form of EL, gathers the outcomes of several de-correlated decision trees into a "forest" to produce a classification result. We build the Decision Trees in the Extra Trees Forest using the sample from the training process. Next, we randomly select k features from the feature set at each test node, and each tree must select the best feature to split the data using quantitative rules.

Unlike other classifiers, ETs completely randomize feature splits rather than relying on the data distribution at each node to determine appropriate thresholds. In contexts with varied sorts of intrusions, this high degree of randomization may lead to a more diversified ensemble, which in turn can improve generalizations for high-dimensional IoV data. Rapid decision-making is possible with ET trees due to their randomised structure. Fast classifications are crucial in real-time IoV systems, so this is an advantage. ET's speed allows for models with less computational overhead, which indirectly improves accuracy by preventing overfitting in complex data.

Random Forest (RF): RF utilizes the majority voting rule to identify a class from a collection of trees. Additionally, it serves as a classifier.

Improved-RFC approach: This approach employs the Random Forest technique, as well as a feature evaluator method. The multi-class trained model for classifying is chosen first in this strategy. In order to improve detection accuracy and decrease processing overhead, mathematical formulations for an Improved Random Forest (IRF) classification technique in an IDS for the IoV typically center on optimizing feature selection, classifier ensembles, and decision-making procedures. An ensemble of T decision trees, with each tree built on a subset of characteristics, makes up the Random Forest (RF) classifier. Let the feature set X = {x_1, x_2, …, x_n} represent each sample in the IoV data, and the associated class labels Y = {y_1, y_2, …, y_n} stand for things like normal or incursion.

For each decision tree *t* in the set *T*, data points we randomly select X_t and features F_t from the dataset. After receiving training from X_t and F_t, a decision tree is divided at each node according to feature criteria, such as knowledge gain. For each given input *X*, the ensemble output *H*(*X*) is the sum of the votes cast by each decision tree in the forest:

$$(X) = mode\{h_t(X)\}_{t=1}^{T} \tag{2}$$

We select features at each node using the information gain (IG) or Gini index to enhance each tree's intrusion detection capabilities. At node *N*, the information gain for feature f_i is calculated as follows:

$$IG(N, f_i) = H(N) - \sum_{j} \frac{|N_j|}{|N|} H(N_j) \tag{3}$$

In this context, *H*(*N*) denotes the entropy of the present node, Nj stands for the partitioned nodes after feature f_i splitting, and $|N|$ and $|N_j|$ denote the quantities of samples in *N* and N_j, respectively. In order to improve intrusion detection, the RF algorithm may maximize IG(*N*, f_i) and then choose features that divide the data optimally for each decision tree. The Out-of-Bag (OOB) error serves as a measure for internal validation, eliminating the need for independent test data and providing a fair assessment of the classifier's efficiency. Each data point x_i that was not used to train the $T_{OOB}(x_i)$ trees has a corresponding OOB prediction, $H_{OOB}(x_i)$:

$$H_{OOB}(x_i) = mode\big(h_i(x_i)\big)_{t \epsilon T_{OOB}(x_i)} \tag{4}$$

Optimizing the RF decision threshold to minimize FP and FN is crucial for real-time IoV situations. A cost-sensitive thresholding method can mitigate these mistakes. To find the best threshold, τ, we minimize a cost function C, where α and β are weights that represent the cost of false positives and false negatives, respectively, in an IoV environment.

$$E_{OOB} = \frac{1}{N}\sum_{i=1}^{N} 1(H_{OOB}(x_i) \neq y_i) \qquad (5)$$

4. RESULTS AND DISCUSSIONS

We implemented the proposed strategy on an I7 processor running Windows 10. We conducted all tests in an Anaconda Jupyter Notebook and Python 3.9 environment to evaluate the performance of the proposed model. Performance metrics assess the performance of a machine learning model. There are several types of performance metrics to evaluate a model; selecting the most suitable one is critical for monitoring and optimizing the model's performance. To assess a model and determine suitable classes for a system, expected performance indicators such as true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are used, despite the difficulty in observing or interpreting real data. Eqs. (6)-(9) show performance metrics for ML models evaluation the models [29].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (7)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \qquad (8)$$

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} + \text{Recall}}{\text{Precision} \times \text{Recall}} \qquad (9)$$

The following conclusions were drawn from the explanation of the machine learning model with and without the Feature Selection Technique.

When the classification data from the IoV attacks was analyzed, the IRF technique had the highest accuracy rate of 0.969, whereas the ET technique had the lowest accuracy rate of 0.96. The classification efficiency results are shown in Table 1.

Table 1. Classification ML model results in the absence of a feature selection technique

ML Model	Accuracy	Precision	Recall	F1-Score
ET	0.96	0.95	0.94	0.931
IRF (Proposed)	0.969	0.959	0.949	0.96

The IRF subset's optimum features subset was therefore added to the original set of two classifiers for further training. The dataset was separated into two parts: training (80%) and testing (20%). Table 2 displays the performance results acquired as a consequence of classification.

Table 2. Classification ML model results in the presence of a feature selection technique

ML Model	Accuracy	Precision	Recall	F1-Score
ET	0.97	0.96	0.96	0.961
IRF (Proposed)	0.99	0.99	0.99	0.98

The comparative evaluation of the proposed model, as shown in Figure 2 and one of the two classifiers shown in Figure 3 outperforms the other when using selected features to model. IRF achieved great accuracy, precision, recall, and an F1-score of 0.98, but ET performed poorly. Because regularization is a critical aspect for this sort of prediction method, the results for IRF are superior to ET. IRF, for example, is rapid to implement and delivers the highest accuracy. IRF, on the other hand, performs well even when certain missing values make the model simple to use.

The precision-recall (PR) curves are particularly beneficial for assessing the IDS's performance in the context of imbalanced classes and various attack types. They emphasize the model's ability to maintain precision as recall increases. The suggested model gets good accuracy and recall for DoS and Probe attacks, which are common and might be easier to spot because they have their own unique traffic patterns, as seen in Figure 4. For R2L, U2R, and Botnet attacks, the model demonstrates a trade-off between precision and recall, indicating that the implementation of additional features or alternative model adjustments may enhance the accuracy of detection.

From Figure 5 it is evident that the ROC curve of IRF classifier outperforms as related to ET classifier. Using a random subset of features and samples, IRF trains an ensemble of decision trees. Because of this unpredictability, IRF is able to generalize well across the many different types of intrusions seen in IoV data, which helps prevent overfitting. Reducing redundancy through improved parameters and optimised feature selection leads to enhanced detection accuracy. IRF is more effective at identifying intrusions because it has a higher AUC for capturing feature interactions in the IoV data.

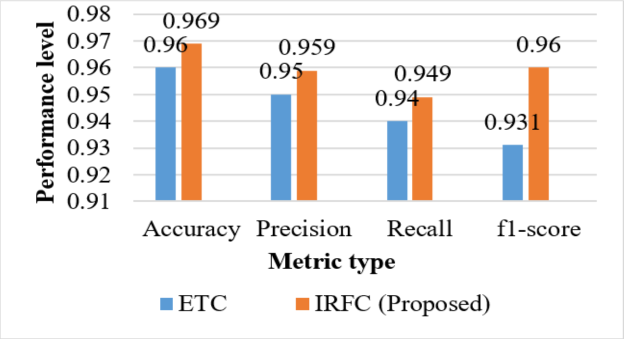

Figure 2. Comparative evaluation of proposed model

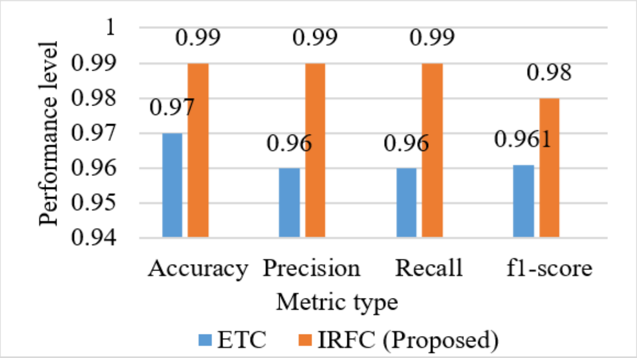

Figure 3. Comparative evaluation of proposed model (IRF) statistical metrics and ET models with feature selection

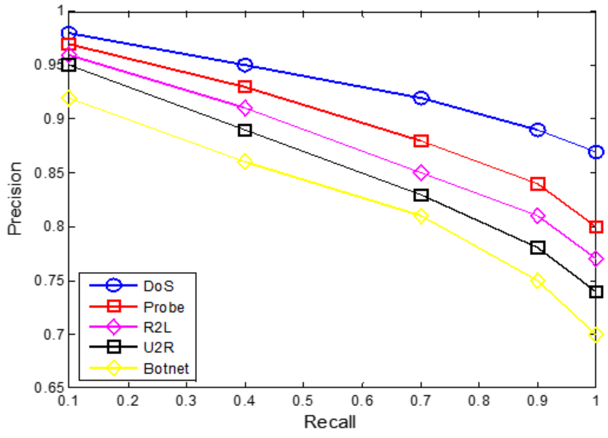

Figure 4. Precision-recall curves analysis of proposed model

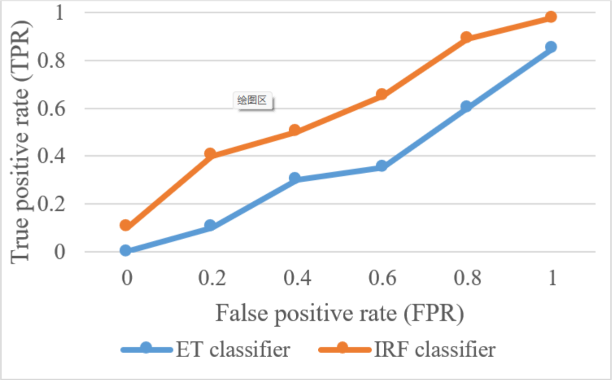

Figure 5. ROC curve of the IRF classifier and ET classifier

ROC curves also help one understand the capacity of the IRF model to differentiate between normal and each particular attack type. Higher AUC values suggest improved discriminating capacity. Table 3 shows the ROC values for every particular class of the CICIDS 2018 attack type. The Improved RF model can consistently classify different types of attacks, as shown by the high ROC-AUC of 98.2% for DoS attacks and 97.5% for Probe attacks. The lower scores for U2R of 94.8% and Botnet attacks of 91.5% suggest that further feature engineering or different detection techniques might be required to raise model performance on these more subdued attack types.

Table 3. ROC values for each specific attack type of CICIDS 2018 dataset

Attack Type	ROC-AUC Score (%)
Normal Traffic	99.1
Denial of Service (DoS)	98.2
Probe	97.5
Remote to Local (R2L)	96.0
User to Root (U2R)	94.8
Botnet	91.5

The Improved RF model reliably detects DoS and Probe attacks, as evidenced by the high ROC-AUC. Because U2R and botnet attacks got lower scores, it's possible that the model needs more feature engineering or a different way to find attacks in order to work better with these more difficult types.

Table 4 illustrates that while IRF has a slightly longer training time, it maintains an efficient prediction time that is suitable for real-time detection in IoV. The high randomization in segments in ET makes it faster to train, but it compromises a minor degree of accuracy in comparison to IRF.

Table 4. Latency and computational efficiency

Model	Training Time (s)	Prediction Time per Sample (ms)
ET	24.7	1.0
RF	28.0	1.1
Improved RF (IRF)	30.5	0.8

Table 5 presents an informative overview of the trade-offs related to tree depth for various tree models, showcasing parameters such as detection rate, memory utilisation, and inference time. In IoV-based IDS, balancing performance and resource utilization is critical. Higher model settings (more trees, deeper trees) improve detection rates, but the much higher memory and processing requirements impact real-time performance and viability for resource-limited IoV devices. The current study shows that we can adjust the IRF model to strike a balance between 99.10% detection capabilities, reasonable processing needs, and 240 MB of memory, thereby enabling efficient and effective intrusion detection in IoV instances.

Table 5. Trade-offs between performance and resource requirements of diffrerent

Model	Detection Rate (%)	Memory Usage (MB)	Inference Time (ms/Sample)
ET	95.2	400	2.5
RF	97.83	385	2.2
Improved RF (IRF)	99.10	240	1.1

We can use statistical tests like t-tests or ANOVA on the CICIDS 2018 dataset to determine any notable variations in feature distributions across different classes of network traffic, such as normal vs. attack kinds, or the types of attacks themselves. To help choose features for the Intrusion Detection Models shown in Table 6, these statistical tests may help figure out which properties are most likely to change a lot depending on the type of traffic. This could help tell the difference between normal traffic and attack traffic.

Table 6. Summary table of diffrerent features

Feature	Test	Groups Compared	p-Value
Flow Duration	t-test	Normal vs. DoS	0.03
Packet Size	ANOVA	Normal, DoS, Probe, Botnet	0.001
Source Bytes	t-test	Normal vs. Botnet	0.15
Destination Port	ANOVA	All Traffic Categories	0.05

5. CONCLUSIONS

Considering autonomous self-driving vehicles are susceptible to different network attacks, IDS are one of the most effective options for detecting network intrusions and securing vehicle networks. The presented work introduced an Intrusion Detection System (IDS) that utilizes tree-based ML techniques to detect threats. The proposed approach includes an IRF with ensemble learning algorithms, which reduces training and response times. The proposed technique's performance was evaluated on the CICIDS 2018 dataset, revealing an accuracy of 0.99, making it a viable alternative for handling CAD and multi-class classification tasks, according to the experimental results. With the detection rate of 99.10%, memory usage of 240 MB, and inference time of 1.1 ms, the proposed system is more efficient. Future work will enhance the model's adaptability to new intrusion types and improve computational efficiency for large-scale deployment in IoV ecosystems.

REFERENCES

[1] WHO. (2023). Road traffic injuries. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

[2] Golson, J. (2016). Jeep hackers at it again, this time taking control of steering and braking systems. https://www.theverge.com/2016/8/2/12353186/car-hack-jeep-cherokee-vulnerability-miller-valasek.

[3] Awang, A., Husain, K., Kamel, N., Aissa, S. (2017). Routing in vehicular ad-hoc networks: A survey on single-and cross-layer design techniques, and perspectives. IEEE Access, 5: 9497-9517. https://doi.org/10.1109/ACCESS.2017.2692240

[4] Yang, F.C., Wang, S.G., Li, J.L., Liu, Z.H., Sun, Q.B. (2014). An overview of Internet of Vehicles. China Communications, 11(10): 1-15. https://doi.org/10.1109/CC.2014.6969789

[5] Ali Alheeti, K.M., McDonald-Maier, K. (2018). Intelligent intrusion detection in external communication systems for autonomous vehicles. Systems Science & Control Engineering, 6(1): 48-56. https://doi.org/10.1080/21642583.2018.1440260

[6] Dai Nguyen, H.P., Zoltán, R. (2018). The current security challenges of vehicle communication in the future transportation system. In 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, Serbia, pp. 000161-000166. https://doi.org/10.1109/SISY.2018.8524773

[7] Sharafaldin, I., Lashkari, A.H., Ghorbani, A.A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. In Proceedings of the 4th International Conference on Information Systems Security and Privacy ICISSP, Funchal, Madeira, Portugal, pp. 108-116. https://doi.org/10.5220/0006639801080116

[8] Alshammari, A., Zohdy, M.A., Debnath, D., Corser, G. (2018). Classification approach for intrusion detection in vehicle systems. Wireless Engineering and Technology, 9(4): 79-94. https://doi.org/10.4236/wet.2018.94007

[9] Al-Dweik, A.J., Mayhew, M., Muresan, R., Ali, S.M., Shami, A. (2017). Using technology to make roads safer: Adaptive speed limits for an intelligent transportation system. IEEE Vehicular Technology Magazine, 12(1): 39-47. https://doi.org/10.1109/MVT.2016.2634462.

[10] Rahman, A., Khan, M.S.I., Razaul, M., Hasan, M., Band, S.S., Chronopoulos, A.T. (2021). Stacked intrusion detection system (IDS) using advanced machine learning (ML) approaches. Preprint.

[11] Fu, W.L., Xin, X., Guo, P., Zhou, Z. (2016). A practical intrusion detection system for Internet of vehicles. China Communications, 13(10): 263-275. https://doi.org/10.1109/CC.2016.7733050

[12] Ahmed, I., Jeon, G., Ahmad, A. (2021). Deep learning-based intrusion detection system for Internet of Vehicles. IEEE Consumer Electronics Magazine, 12(1): 117-123. https://doi.org/10.1109/MCE.2021.3139170

[13] Yang, L., Moubayed, A., Shami, A. (2021). MTH-IDS: A multitiered hybrid intrusion detection system for Internet of Vehicles. IEEE Internet of Things Journal, 9(1): 616-632. https://doi.org/10.1109/JIOT.2021.3084796

[14] Yang, L., Shami, A. (2022). A transfer learning and optimized CNN based intrusion detection system for Internet of Vehicles. In ICC 2022-IEEE International Conference on Communications, Seoul, Republic of Korea, pp. 2774-2779. https://doi.org/10.1109/ICC45855.2022.9838780

[15] Cheng, P.Z., Xu, K., Li, S.M., Han, M. (2022). TCAN-IDS: Intrusion detection system for internet of vehicle using temporal convolutional attention network. Symmetry, 14(2): 310. https://doi.org/10.3390/sym14020310

[16] Hbaieb, A., Ayed, S., Chaari, L. (2022). Federated learning based IDS approach for the IoV. In Proceedings of the 17th International Conference on Availability, Reliability and Security (ARES '22), New York, NY, USA, p. 123. https://doi.org/10.1145/3538969.3544422

[17] Alqahtani, H., Kumar, G. (2022) A deep learning-based intrusion detection system for in-vehicle networks. Computers & Electrical Engineering, 104: 108447. https://doi.org/10.1016/j.compeleceng.2022.108447

[18] Alwahedi, F., Aldhaheri, A., Ferrag, M.A., Battah, A., Tihanyi, N. (2024). Machine learning techniques for IoT security: Current research and future vision with generative AI and large language models. Internet of Things and Cyber-Physical Systems, 4: 167-185. https://doi.org/10.1016/j.iotcps.2023.12.003

[19] Rani, P., Sharma, R. (2023). Intelligent transportation system for internet of vehicles based vehicular networks for smart cities. Computers and Electrical Engineering, 105: 108543. https://doi.org/10.1016/j.compeleceng.2022.108543

[20] Lu, C.W., Cao, Y.X., Wang, Z.B. (2024). Research on intrusion detection based on an enhanced random forest algorithm. Applied Sciences, 14(2): 714. https://doi.org/10.3390/app14020714

[21] Danba, S., Bao, J.J., Han, G.R., Guleng, S., Wu, C. (2022). Toward collaborative intelligence in IoV systems: Recent advances and open issues. Sensors, 22(18): 6995. https://doi.org/10.3390/s22186995

[22] Morales, G.A., Xu, J.Y., Zhu, D.K., Slavin, R. (2022). Lightweight collaborative inferencing for real-time intrusion detection in IoT networks. In 2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta), Haikou, China, pp. 392-400. https://doi.org/10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00076

[23] Lei, Y., Wang, S.L., Zhong, M., Wang, M., Ng, T.F. (2022). A federated learning framework based on incremental weighting and diversity selection for internet of vehicles. Electronics, 11(22): 3668.

https://doi.org/10.3390/electronics11223668

[24] Wang, Y.Q., Qin, G.H., Zou, M., Liang, Y.H., Wang, G.F., Wang, K.P., Feng, Y., Zhang, Z.Z. (2024). A lightweight intrusion detection system for internet of vehicles based on transfer learning and MobileNetV2 with hyper-parameter optimization. Multimedia Tools and Applications, 83: 22347-22369. https://doi.org/10.1007/s11042-023-15771-6

[25] Tu, J., Shang, W. (2023). Enhancing intrusion detection in the internet of vehicles: An ensemble and optimized machine learning approach. In 2023 2nd International Conference on Sensing, Measurement, Communication and Internet of Things Technologies (SMC-IoT), Changsha, China, pp. 207-211. https://doi.org/10.1109/SMC-IoT62253.2023.00044

[26] Bangui, H., Ge, M.Z., Buhnova, B. (2022). A hybrid machine learning model for intrusion detection in VANET. Computing 104: 503-531. https://doi.org/10.1007/s00607-021-01001-0

[27] Rachmanto, R.A., Regannanta, F.J., Ubaidillah, Arifin, Z., Widhiyanuriyawan, D., Yohana, E., Prasetyo, S.D. (2023). Analysis development of public electric vehicle charging stations using on-grid solar power plants in Indonesia. International Journal of Transport Development and Integration, 7(3): 215-222. https://doi.org/10.18280/ijtdi.070305

[28] Gururaj, H.L., Jihadi, V., Tanuja, U., Flamini, F., Soundarya, B.C., Ravi, V. (2022). Predicting traffic accidents and their injury severities using machine learning techniques. International Journal of Transport Development and Integration, 6(4): 363-377. https://doi.org/10.2495/tdi-v6-n4-363-377

[29] Powers, D.M. (2020). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061. https://doi.org/10.48550/arXiv.2010.16061