Enhancing Cybersecurity in Internet of Vehicles: A Machine Learning Approach with Explainable AI for Real-Time Threat Detection

Tanish Patel
Dept. of CSE, School of Technology,
Pandit Deendayal Energy University
Gandhinagar, Gujarat, India
tanishpatel0106@gmail.com

Rutvij H Jhaveri
Dept. of CSE, School of Technology,
Pandit Deendayal Energy University
Gandhinagar, Gujarat, India
rutvij.jhaveri@sot.pdpu.ac.in

Dhavalkumar Thakker*
School of Computer Science,
University of Hull
Hull, United Kingdom
d.thakker@hull.ac.uk

Sandeep Verma
Dr BR Ambedkar National Institute of Technology
Jalandhar, Punjab, India
sandeepv.ec.13@nitj.ac.in

Palash Ingle
Dept. of Computer and Information Security, Sejong University
Seoul, South Korea
palash@sejong.ac.kr

ABSTRACT

The proliferation of IoV technologies has revolutionized the use of transport systems to a great level of improvement in safety and efficiency, and convenience to users. On the other hand, increased connectivity has also brought new vulnerabilities, making IoV networks susceptible to a wide range of cyber-attacks. The contribution of this paper is the in-depth study of the development and evaluation of advanced machine learning (ML) models that detect and classify network anomalies in IoV ecosystems. Several classification models have been studied in our research to achieve high accuracy for discriminating between benign and malicious traffic. This work further harnesses Explainable AI (XAI) methodologies through the LIME framework for enhanced interpretability of models' decision-making processes. Experimental results strongly advocate the strength of Random Forest and XGBoost, proving to be better on the binary and multi-class classification tasks, respectively. Due to resilience, preciseness, and scalability these models are a practical choice in real-world IoV security frameworks. Explainability integrated not only strengthens model reliability but also closes the gap between performance and interoperability in vehicular networks.

CCS CONCEPTS

• **Computing methodologies** → **Supervised learning by classification**.

KEYWORDS

Artificial Intelligence, Cybersecurity, Explainable Artificial Intelligence, Internet of Vehicles, Machine Learning

*Corresponding Author

SAC'25, March 31 –April 4, 2025, Sicily, Italy

ACM ISBN 979-8-4007-0629-5/25/03.
https://doi.org/10.1145/3672608.3707769

ACM Reference Format:
Tanish Patel, Rutvij H Jhaveri, Dhavalkumar Thakker, Sandeep Verma, and Palash Ingle. 2025. Enhancing Cybersecurity in Internet of Vehicles: A Machine Learning Approach with Explainable AI for Real-Time Threat Detection. In *The 40th ACM/SIGAPP Symposium on Applied Computing (SAC '25), March 31-April 4, 2025, Catania, Italy.* ACM, New York, NY, USA, Article 4, 8 pages. https://doi.org/10.1145/3672608.3707769

1 INTRODUCTION

The emerging paradigm of technology, the Internet of Vehicles (IoV), enables two-way communication between vehicles, infrastructure, and external networks, driving advancements in safety, efficiency, and user convenience in transportation. Increased interconnectivity stemming from IoV induces several cybersecurity problems in its systems-principally vulnerability to DoS, spoofing, and intrusion attacks. Such threats would degrade the integrity of vehicle operations and safety for the passengers, and hence, security must become an integral part of IoV development. In this dynamic environment where traditional security mechanisms face a challenge in scaling up real-time adaptability for IoV to recognize and neutralize these threats, this scenario needs IoT-based, edge-controlled, real-time security.

The innovation of machine learning (ML) models happens to be one of the most powerful methods of identifying and classifying potential cyber threats within an IoV system processing a huge volume of real-time data. At the same time, a model must be able to explain why it has arrived at a certain decision. This need further strengthens the importance of Explainable Artificial Intelligence (XAI) kind of insight into complex models' processes of decision-making and consequent trustworthiness. For instance, the study done in [22] analyzed the contribution of XAI to the improvement of the interpretability of ML models, which can be leveraged in the IoV-based application for real-time detection of threats, are thus a major step toward establishing trust in automated systems. Similarly, [12] highlights that XAI-based systems are required for real-time prediction and threat mitigation in the IoV context.

This paper proposes the development and testing of ML models against binary and multi-class classifications to solve the aforementioned cybersecurity challenges arising within IoV systems.

Understanding not only how but also why a model reaches a conclusion is important for building trust and ensuring system safety in IoV [19]. The key contributions of this work are:

(1) This study compares different ML models, such as the Support Vector Classifier, Decision Tree, and Logistic Regression against Random Forest and XGBoost, while establishing the latter to be better in IoV security contexts.
(2) The models Random Forest and XGBoost show very high accuracies of up to 99% in binary as well as multi-class classifications for benign and malicious traffic.
(3) The LIME method gives complete insight into what affects the made predictions, and as a result, the system becomes more interpretable and can be trusted.

Critical developments in IoV cybersecurity on the application of ML and XAI in real-time threat detection are discussed in the Literature review. The section on the applied methodologies covers how data collection was done, pre-processing techniques, and finally, the ML model used for network traffic classification, including applied explainability techniques like LIME. Finally, the Results section discusses in detail the models for performance evaluation of binary and multiclass classifications, in which the respective accuracy, precision, recall, and the principal F1 score are compared. In comprehension, it finishes with conclusions and future work sections that provide a summary of activities carried out to emphasize the importance of the proposed method and also directions for further research on how federated learning with blockchain technologies can work together to enhance IoV security.

2 LITERATURE REVIEW

The fact that emphasis has been relatively recent on the IoV cybersecurity domain, especially with the dramatically increasing number of connected vehicles, there is a need that demands real-time mechanisms to identify threats. Rjoub et al. [22] provided an in-depth review of XAI as relevant to how XAI develops the interpretation of ML frameworks as far as threat identification is concerned. This aspect is of peculiar importance within the context of the IoV, where decision-making must be in real-time, but with certain levels of transparency.

Páez et al. [18] reviewed reinforcement learning architectures that are still used in adaptive cybersecurity for autonomous ground vehicles. The research reveals how reinforcement learning responds to actual dynamic cyber threats in XAI integration so that at critical moments, decisions made by an AI system are understandable to operators. Furthermore, recent works such as Asaju et al. [8] continue to elaborate on them by considering AI-driven advances in the intrusion detection systems of V2X communications. Real-time threat detection and mitigation, supported by AI and ML, is a vital feature of safe data exchange between vehicles. This, no doubt, reflects how AI usage is turning around security maintenance in vehicular networks. In a related further study, P. RS et al. [23] presented an intelligent dynamic cyber-physical system targeting threat detection in 6G autonomous vehicle networks. This study presents the latest sophisticated ML approaches to protecting communications between vehicles and networks in given the rise in data traffic expected with 6G. Kumar et al. in [12] present the XAI-based cyber threat hunting system to provide real-time threat prediction. This study exploits deep learning models and provides an efficient technique for identifying and mitigating security risks, especially suited to IoV with a rapid data exchange environment. Another key contribution by Kumar et al. [11] is explaining how blockchain combined with XAI can be utilized to detect cyber threats. This work emphasizes the need to bring together blockchain-a decentralized technology-with XAI for making the threat detection mechanism transparent and safe, appropriate for distributed IoV networks. Ahmad et al. [2] further solidify the potential benefits of applying ML and blockchain technologies to enhance IoV cybersecurity. Application of these technologies in real-time network traffic helps in quicker detection of cyberattacks while ensuring that data integrity is preserved using intrusion detection systems (IDS). Ding et al. [9] proposed DeepSecDrive, a deep learning-based approach coupled with XAI to runtime detection of cyber-attacks in the in-vehicle network. Similarly, the use of inexplainable models in this framework provides reasoning for making some security decisions, therefore increasing its detection rate and gaining better trust. Real-time ML algorithms and the newly introduced XAI techniques, described by Alqahtani et. al [7], can enhance transportation security, especially in electric and flying vehicle systems. It explains how such technologies would be able to respond more effectively to cyberattacks on crucial sensors while maintaining the system's resilience and safety. The study of Ahakonye et al. [1] also talked about the intrusion detection application of XAI within IIoT, focusing on the need for transparency in allowing real-time decisions on complex industrial systems, like IoV networks. Moustafa et al. [15] made a wide-ranging overview of AI-based models for IoV intrusion detection and found out the future need to identify threats in real-time. The work described here encompassed an overview of the manifold AI techniques specifically applied on IoV data streams with the intention of attaining both greater accuracy and explainability. In another study, Almehdhar et al. [5], the authors review advanced deep learning-based intrusion detection in intelligent vehicle networks, especially federated learning. This decentralized learning approach enhances real-time detection of threats while preserving privacy. Nwakanma et al. [17] discussed the application of XAI, in an intrusion detection system for intelligent connected vehicles, proposing a system that will make ML models more interpretable to realize the detection of anomalies and respond to threats in real-time. Finally, Rani et al. [21] considered federated learning-based misbehavior detection in IoV and came up with a new distributed AI that will allow the development of real-time threat detection in interconnected vehicle networks, with a particular focus on 5G-enabled systems.

Further recent studies, as classified in Table 1 below, focus on various aspects of enhancing cybersecurity in the IoV, utilizing ML, XAI, blockchain, and other advanced techniques. These studies explore methodologies ranging from federated learning and deep learning to blockchain integration and anomaly detection in connected vehicles.

The reviewed articles present the ascendant trend of using ML, XAI, and real-time threat detection mechanisms to implement security in the IoV. The type of research, in this case, consists of three highly relevant needs for cybersecurity solutions: decentralization,

Table 1: Literature Review of Notable studies

Paper Title	Methods	Results	Related Industry	Factor of Improvement
Blockchain Intelligence for IoV: Challenges and Solutions [24]	Utilized blockchain and ML for decentralized security in IoV	Enhanced network flexibility and privacy; real-time threat detection	Automotive, Telecommunication	Improved decentralization and data privacy
Cybersecurity Threats Mitigation in IoV Communication Systems Using Reliable Clustering [10]	Clustering and deep learning-based routing strategies for VANETs	Increased network stability and reduced latency in IoV	Vehicle Networking, Smart Cities	Enhanced real-time routing and communication stability
XAI-ADS: An XAI Framework for Enhancing Anomaly Detection [16]	Integrated XAI with real-time sensor data processing for anomaly detection in autonomous systems	Improved resilience of autonomous driving systems to cyberattacks	Automotive, Autonomous Driving	Real-time decision-making and anomaly detection accuracy
Explainable AI Enabled Intrusion Detection for Secure CPS [6]	Applied XAI for real-time intrusion detection in cyber-physical systems	Improved transparency and accuracy in identifying intrusions in CPS	Cyber-Physical Systems, Industrial Automation	Higher transparency and interpretability of intrusion detection mechanisms
Real-Time Threat Intelligence Integration for Cybersecurity in Autonomous Vehicles [3]	Integrated real-time threat intelligence into deep learning frameworks for AVs	Improved detection and mitigation of threats in autonomous vehicles	Autonomous Vehicles, AI	Enhanced real-time detection and cybersecurity resilience
Security and Trust Management in IoV: Challenges and Machine Learning Solutions [4]	Trust evaluation and ML techniques for IoV security	Increased real-time communication security and trustworthiness in IoV networks	IoV, Smart Cities	Improved trust management and real-time data security

interpretability, and adaptiveness toward constantly and dynamically evolving threats. Future work should thus integrate federated learning with blockchain technologies in harnassing real-time decentralized threat detection within highly connected vehicular networks while simultaneously ensuring both data privacy and transparency. This lays out foundational steps for developing robust, scalable, and interpretable methodologies for improving the cybersecurity of IoV.

3 METHODOLOGY

The great strides in IoV-related technologies have significantly enhanced transportation safety, operational efficiency, and user convenience. Advanced sensors are linked via complicated networks, and today's vehicles pride themselves on the unparalleled level of interconnectivity. At the same time, such high levels of interconnectivity, in themselves, open the door for new vulnerabilities, thus rendering IoV systems potentially prone to a set of cyber-attacks likely to compromise the safety and integrity of vehicle operations [14].

For that matter, since huge stakes are involved, IoV security solutions must be strong and dependable. The work presented here focuses on the development and evaluation of ML models for traffic classification in IoV. More specifically, the model has to identify normal and harmful activities with a view of securing the critical infrastructures on IoV against cyber threats. Our approach was to create a model that would confirm high accuracy in the detection and interpretability of identified features but would also include understanding important features that could influence classification. [13]

In this work, we are going to adopt different ML approaches to solve two problems of classification conducted in this investigation: binary classification and multi-class classification, where various types of attacks are classified, such as DoS, and Spoofing, including Gas, RPM, Speed, SteeringWheel. In essence, taking into account the fine-grained classification of the attacks, the current work's proposed models will be very beneficial for IoV environments with real-time threat-detection demands. Several sensors will be embedded into the vehicle to record for IoV security, every aspect related to performance, such as speed, gas level, and RPM of the vehicle. Afterward, this information will be passed on to a central processing system where it will be preprocessed—such as data cleaning and scaling or encoding of features for training the model. The result from the preprocessing stage acts as input for a set of ML algorithms, like Random Forest, XGBoost, and Support Vector Classifier, that train and evaluate the model. After training the best performance model and developing others, they were selected for further analysis by LIME, as it could improve the interpretability of models' decisions. This would be useful in identifying what factor contributes to which classification and this will help in building a trustful basis in the system's predictions. Figure 1 shows the structure of the setup for IoV, reflecting at a high level the entire flow in the security system, from data collection to model deployment and retraining.

3.1 Data Collection & Preprocessing

The data implemented in this project is taken from several sensors interfaced within the IoV system, capturing vital parameters like vehicle speed, RPM, gas levels, and steering wheel positions. Sensors are currently known to provide real-time data that could be analyzed for anomalies and classified into benign or malicious behavior network traffic. With this objective, the data acquisition system was designed to provide an unbroken exchange between cars and the central processing system, thus forming one of the most vital constituents of the IoV ecosystem.

3.1.1 Data Collection from IoV Systems. This is the prototype stage of security configuration, taking traffic data from all the vehicles equipped with sensors of different types. Such vital parameters are being monitored by sensors:

- Vehicle Speed: Any fluctuations in the velocity may indicate malfunctions in the car control system.

Figure 1: IoV Security Setup Diagram

- RPMs: Revolutions per minute can tell a lot about the operating condition of a vehicle.
- Gas Level: Changes in gas levels suddenly may indicate attempts at interference or spoofing.
- Steering wheel: Changes in the steering wheel position can indicate unauthorized or unanticipated use of the vehicle.

Every sensor transmits information to a master node, which aggregates the data collected by the former and transforms it into an exponentially huge dataset describing the vehicle's current working conditions. This raw dataset undergoes processing for validation, considering coherent, complete, and ready-to-feed ML models.

3.1.2 Data Preprocessing. Raw sensor data collected is passed through several preprocessing steps that prepare it for training on the model. Data preprocessing plays a critical role for most ML models in terms of accuracy and efficiency because they transform raw data into forms appropriate for use. In this study, the following major data preprocessing activities are covered:

- Data Cleaning: Raw data collected from IoV sensors were filled with missing or incomplete records of types. Missing records were filled through imputation techniques. Since no type of important information was lost, minimum damage has been demonstrated to the data. Median values of the feature were substituted for missing values in order to preserve the integrity of the dataset.
- Feature Encoding: Since most of the features are categorical like sensor status or attack labels, the approach uses one-hot encoding as a preprocessing measure to encode the variables. This encoding procedure encodes every categorical feature as a vector of zeros and ones which can be directly input in the ML algorithms.
- Since features of the dataset included speed and RPM, of which orders are different, normalization was applied to make all features into a uniform magnitude in such a way that larger values do not dominate the training process. The normalized feature is then given by:

$$x_i^{\text{norm}} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

where, $x_{\min}$ and $x_{\max}$ represent the minimum and the maximum values for the feature respectively.

- Class Imbalance: The biggest problem that the dataset of this research project creates is a class imbalance as per the number of instances per class. Here, there are hundreds of benign traffic instances compared to malicious attack instances. We utilized SMOTE, which generates synthetic examples for the minority classes to get it nearer to equal distribution. Oversampling the underrepresented categories on attacks enables SMOTE to create a power for the ML models to learn well about all classes.
- Dimensionality Reduction: We used PCA to reduce dimensionality and hence, the reduced complexity to improve the model. PCA is very efficient in feature selection, that it picks out the most relevant features and filters redundant information, which has an optimal bearing on optimizing a dataset for efficient models.

3.2 ML Models

Once the data preprocessing phase is complete, the next step involves feeding the processed data into a set of ML models for training. We conducted experiments with different versions of ML algorithms chosen for their uniqueness in addressing classification difficulties. Through this research, we aim to develop models that not only correctly identify cyberattacks but also explain their predictions using techniques designed to improve explainability. The models employed include Random Forest, Support Vector Classifier (SVC), Decision Trees, XGBoost, and Logistic Regression as a baseline.

3.2.1 Random Forest. The random forest algorithm is an ensemble learning model that builds multiple decision trees and outputs the mode of the classes in the case of a classification task. It doesn't easily overfit and is suitable for big datasets with high dimensionality. In the random forest algorithm, each decision tree trains itself on a randomly selected subset of the data, and the final prediction combines the predictions of all trees. Mathematically, the decision trees involved can be expressed as:

$$H(x) = \sum_{i=1}^{N} \alpha_i h_i(x) \tag{2}$$

where $h_i(x)$ are the individual decision trees, and α_i are the weights assigned to each decision tree. The use of multiple decision trees helps in reducing variance and improving classification accuracy in both binary and multi-class scenarios.

3.2.2 Decision Trees. Decision Trees are simpler models compared to Random Forests but are highly interpretable. They recursively form branches from the root, with subsets of data depending on a split threshold of the feature value, until the leaves represent the final classification decision. This splitting can be done using measures such as the Gini index or Information Gain.

3.2.3 Support Vector Classifier (SVC). SVC constructs a hyperplane or a set of hyperplanes in a higher or infinite-dimensional space to separate different classes. It aims to find the hyperplane that maximizes the margin between the classes for better classification accuracy. The optimization problem solved by SVM is:

$$\min \frac{1}{2}\|w\|^2 \quad \text{subject to } y_i(w \cdot x_i + b) \geq 1, \forall i \tag{3}$$

where w is the weight vector, b is the bias term, and y_i is the class label. SVC is particularly effective in handling non-linearly separable data and helps differentiate nuanced attack types within the IoV dataset.

3.2.4 Extreme Gradient Boosting (XGBoost). XGBoost is an efficient gradient-boosting algorithm that builds models sequentially, where each new model corrects the errors made by the previous ones. The objective function of XGBoost is designed to minimize both the loss and a regularization term, as given by:

$$L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{4}$$

where l is the differentiable convex loss function, and Ω is the regularization term.

3.2.5 Logistic Regression (Baseline). Logistic regression is a simple and transparent model that has been classically used for binary classification tasks. It models the probability of input belonging to a certain class using a logistic function over a linear combination of input features. Although it is not as robust as the other models discussed in this paper, it serves as a useful baseline for comparing the performance of more advanced approaches.

3.3 Generating Explainability from LIME

To enable achieving precision without losing interpretability, LIME was applied to the best model identified above. Furthermore, LIME aids in adding more transparency to the black box by providing local explanations for every prediction made by the model. This is accomplished through input data localized near a particular instance and fitting more tractable, interpretable models, such as linear models, to approximate the functionality of the original complex model in the local setting. [20]

Algorithm 1 IoV Security Traffic Classification and Explainability

Input: Raw sensor data from IoV system
Output: Classification labels (Benign, Attack Type), Explanation for decisions

(1) **Data Collection:**
  (a) Collect real-time data from IoV sensors such as vehicle speed, RPM, gas level, and steering position.
  (b) Send the data to a central system for processing.
(2) **Data Preprocessing:**
  (a) *Data Cleaning:* Fill the missing values through imputation.
  (b) *Feature Encoding:* Translate categorical features to numerical values using one-hot encoding.
  (c) *Scaling:* Scale continuous feature variables such as speed and RPM using min-max scaling.
  (d) *Handling Class Imbalance:* Over-sample the minority classes using SMOTE within the data set.
  (e) *Dimensionality Reduction:* If relevant, reduce the dimensionality by applying PCA.
(3) **Model Training:**
  (a) Divide the dataset into training and testing sets.
  (b) Perform cross-validation over hyperparameters to tune each of the models: Random Forest, Support Vector Classifier, XGBoost, Decision Tree, and Logistic Regression.
  (c) Train models with optimal hyperparameters on the training dataset.
  (d) Evaluate models on the test set using metrics such as accuracy, precision, recall, and F1 score.
(4) **Best Model Selection:**
  (a) Select the model that maximizes the performance metric, such as overall performance (e.g., Random Forest).
(5) **Explainability with LIME:**
  (a) For each prediction, perturb the input features randomly to build a local neighborhood of instances. Fit an interpretable simple model to the noisy data (e.g., a linear model).
  (b) Identify which features contributed to the forecast.
  (c) Construct a local explanation, highlighting prominent features (e.g., RPM and speed).
(6) **Model Deployment and Feedback Loop:**
  (a) Incorporate the selected model into the IoV system to classify network traffic in real time.
  (b) Continuously assess model performance with real-time data.
  (c) Create a feedback mechanism to update the model continuously in response to emerging attacks.

LIME was selected for this study as it can provide localized interpretability for highly complex "black box" ML models without losing the "power" of such models. In the IoV, security decisions directly affect the safety and effectiveness of vehicle operations. Therefore, not only do the high classification accuracies realized need to be achieved, but also the rationale behind each prediction must be interpretable. LIME excels in this task, particularly because it approximates the behavior within the local vicinity of every instance in a complex model, providing clear and interpretive explanations. This becomes especially important in scenarios involving

diverse sophisticated cyber-attacks, such as DoS or sensor spoofing, where transparency is fundamental to building trust and allowing system decisions to be verifiable and understandable for security analysts.

Moreover, the model-agnostic nature of LIME makes it applicable to any machine-learning model, making it an all-purpose tool. This flexibility enables LIME to be used in conjunction with all the classifiers applied in this work, including Random Forest, SVC, and XGBoost. Finally, the ability of LIME to provide profound insights into the major features driving model predictions makes it the most suitable selection for improving the interpretability of our IoV security models.

LIME works by generating a neighborhood of perturbed samples around an instance of interest and then uses the following optimization (Eq.5) function to fit the simpler model:

$$\min_{g \in G} \sum_{i=1}^{N} \pi_x(x_i') \cdot L(f(x_i'), g(x_i')) + \Omega(g) \tag{5}$$

where $L(f(x_i'), g(x_i'))$ is the loss function measuring how well the simpler model g performs with respect to the original complex model f, $\pi_x(x_i')$ is the proximity measure that assigns higher values to perturbed instances closer to the original instance, and $\Omega(g)$ is the term that imposes a complexity penalty on the simpler model.

Using LIME, it becomes possible to determine which features—such as RPM, gas levels, or even the steering wheel position—have the most influence on the decision regarding whether the network traffic is benign or malicious. More importantly, LIME enables the detection of critical patterns in sensor data that represent specific types of attacks, such as DoS or spoofing gas.

Comprehensively, the entire process stated in the study can be accustomed to algorithm 1

In a nutshell, the work presented uses all the SOTA ML models and improves their interpretability through LIME in the sense that the models can be both accurate in cyber-attack detection and as explanatory in decision-making processes. Methodologies will have a sound basis: constantly adapting to new threats but retaining the trust and dependability factor in practice.

4 RESULTS AND DISCUSSION

An evaluation was conducted to assess the effectiveness of different ML models in the context of IoV security. This includes both multi-class and binary classification tasks for which the models applied were XGBoost (XGB), Random Forest Classifier (RFC), Support Vector Classifier (SVC), Decision Tree (DT), and Logistic Regression (Log). Accuracy (*Acc*), Precision (*P*), Recall (*R*), and F1 Score (*F1*) were used as evaluation metrics in most cases. The summary of results for both multi-class and binary classification tasks is shown in Table **??**.

The models proved to be highly effective in the binary classification test setup aimed at classifying "Benign" and "Attack" traffic. Below are the key observations for each model:

- **XGB:** The XGB model performed quite well, resulting in an accuracy of up to 0.98, precision of 0.99, recall of 0.9802, and an F1 score of 0.9801. These results indicate that XGBoost can distinguish between benign and malicious network data effectively, thereby minimizing false positives and negatives.
- **RFC:** The RFC model performed better compared to other models, with an accuracy of around 0.99, precision of 0.99, recall of 0.9802 and an F1 score of 0.9851. This combination of high accuracy and balanced performance metrics makes RFC highly suitable for binary classification in IoV systems.
- **SVC:** The SVC model performed well, achieving an accuracy rate of 0.95, precision of 0.93, recall of 0.8692 and an F1 score of 0.9078. However, its somewhat lower recall rate indicates that it may occasionally fail to detect certain instances of attacks.
- **DT:** The Decision Tree model achieved an accuracy of 0.96, precision of 0.90, recall of 0.82, and an F1 score of 0.88. While the model performed adequately; its relatively low recall implies that it may fail to detect certain assault scenarios, which is critical in real-world IoV applications.
- **Logistic Regression:** The Logistic Regression model performed the least effectively in this binary classification problem. Its lower recall and F1 score indicate difficulties in accurately detecting instances of attacks.

During the multi-class classification task, the models were tasked with differentiating between various attacks, including "DoS," "Spoofing-Gas," "Spoofing-RPM," "Spoofing-Speed," "Spoofing-SteeringWheel," and "Benign." The performances exhibited some high variabilities.

- Again, promising behavior is observed in XGBoost. For example, it achieved 0.98 accuracy and precision at 0.99, a recall of 0.9802 and an F1 score of 0.9801. Such results indicate that XGBoost truly navigates complexities associated with multi-class classification and keeps a very high level of both precision and recall while changing types of attacks.
- The Random Forest Classifier showed an accuracy 0.99, while the precision, recall as well as F1 score are 0.99, 0.98, and 0.98, respectively. Overall, RFC's performance is remarkable in both binary as well as multi-class tasks. This means it can be used efficiently in the vast field of IoV systems.
- The Support Vector Classifier was marginally less accurate when applying it to the multi-class scenario, which achieved an accuracy rate of 0.94, a precision level of 0.90, a recall value of 0.818,2, and an F1 score of 0.8749. The lower value of recall shows that the Support Vector Classifier faced greater difficulties in classifying all types of attacks correctly under a multi-class scenario.
- The Decision Tree model was stable, with an accuracy rate of 0.95, a precision rate of 0.91, a recall of 0.8349, and an F1 score of 0.8887. Though the Decision Tree model was able to perform well in a multi-class setting, its performance metrics were relatively weaker compared to XGB and RFC models. This implies that the Decision Tree might not be very efficient at handling the added complexity associated with multi-class classification.
- Logistic Regression performed poorly in multi-class classification achieving an accuracy of 0.80, precision of 0.79, recall value of 0.65, and F1 score of 0.72. The respective fall of both the recall and F1 scores evidenced the issues that seem to arise in making clear-cut delineations among different classes of attacks.

The results of experiments have shown that the Random Forest Classifier scores the highest accuracy and well-balanced performance for all significance metrics over all other models both in binary and multi-class tasks. XGBoost shows great performance across all classification tasks, and hence this is a good alternative to Random Forest. Both the models have been capable of achieving good discriminative ability over the different classes of the IoV network traffic; thus, they become a good choice for real-world applications where clarity and accuracy are critical.

Models of Random Forest and XGBoost showed better performance compared to the Vector Classifier, and to the Decision Tree, at least in multi-class classification, because the recall rates deceased. Meanwhile, despite its simplicity, Logistic Regression continually failed all metrics and is likely not suitable for use in IoV systems when more complex items take place in the field of categorization.

The outputs of these experiments quite clearly indicate that high-quality ML algorithms, such as Random Forest and XGBoost are to be used for security purposes in the context of the IoV. Attack detection and classification accuracy will play a significant role in such a context. Further results will be discussed on the use of LIME in boosting the explainability and allowing insightful comprehension of the processes through which these models develop their predictions.

This further enhances performance as interpretability is provided through the application of the LIME method to ML models applied to both binary and multi-class classification tasks, allowing insights into the factors at play that drive decision-making by the model, thus leading to better understanding and therefore increasing confidence in predictions made by the model.

LIME in turn, explains features that have a high influence on the model predictions in the task of binary classification, which should classify "Benign" and "Attack" traffic.

As Figure 3 suggests, it has been inferred that one of the most striking attributes that the model utilized to predict an "Attack" is DATA_3, which has a value of -0.64. Other attributes used by the model within the framework for the decision-making at a threshold defined for the proof of whether traffic is to be categorized under the attack or not include DATA_7, DATA_0, DATA_1, and DATA_2. Attributes were assigned weights regarding the influence they had on the outcome of the prediction. DATA_7 had a weight of 1.80; these data points significantly influenced the determination.

The model is very confident in the "Attack" prediction where 1.00 is assigned as its probability for that classification while "Benign" possesses a probability of 0.00. Therefore, this level of confidence is extremely high and talks about the reliance on the model in its binary classification especially when goodwill features are known to attack the traffic.

In the multi-class classification scenario, in which the model was distinguishing among several classes of attacks, in addition to benign traffic, LIME was just as insightful about which features it considered important.

Classification of a "Spoofing" attack, as described in figure 2, "RPM" spoofing, the variable DATA_0 presented high magnitudes at 2.00, such that the model is classified squarely for "RPM." Other variables were also observed with great effects, which include DATA_1, DATA_2, and DATA_5, whose contribution was in line with the determined threshold values in making decisions by the model.

Table 2: Summary of the Results of ML Models

Model	Classification Type	Accuracy	Precision	Recall	F1 Score
XGBoost	Binary	0.98	0.99	0.98	0.98
Random Forest Classifier		0.99	0.99	0.98	0.99
Support Vector Classifier		0.95	0.93	0.87	0.91
Decision Tree		0.96	0.90	0.82	0.88
Logistic Regression		0.90	0.88	0.79	0.84
XGBoost	Multi-Class	0.98	0.99	0.98	0.98
Random Forest Classifier		0.99	0.99	0.98	0.99
Support Vector Classifier		0.94	0.90	0.82	0.87
Decision Tree		0.95	0.91	0.83	0.89
Logistic Regression		0.80	0.79	0.65	0.72

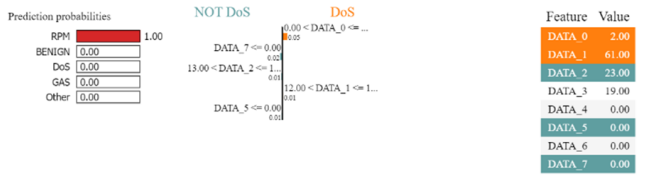

Figure 2: LIME Explanations for Multi Class Classification for a record with RPM attack

Figure 3: LIME Explanations for Binary Classification for a record with Attack Label

A multi-class context calls for a good understanding of the data, which in this case has been well complied with in distinguishing various types of attacks.

It shows an enormous amount of confidence in its "RPM" spoofing prediction in assigning a probability of 1.00, whereas all other classifications, such as "Benign," "DoS," and others, remain assigned probabilities of 0.00. Such predictions by the model are correct and reasonable both because of the highly confident nature and because of LIME's reasonableness.

Practical implementation of LIME showed that the models of ML deployed in this paper, including Random Forest Classifier and XGBoost, were not only very well-represented in terms of performance concerning accuracy and other metrics but also with interpretable results and the potential for reliability in practice. This ability to distinguish the real features from those that impact the predictions is of great value in security-based applications in the IoV, because knowing why a model has reached a certain conclusion is as important as the conclusion.

These results, which are Figures 2 and 3, indicate that an XAI, such as LIME, must be incorporated into ML applications where the security of customers is of importance. Explanations of the decision-making aspects help shed light on the working of internal mechanics of a model; therefore, models become more credible and trustworthy. As a consequence of this rise, such models become highly applicable in sensitive areas where interpretability is crucial.

5 FUTURE WORK & CONCLUSION

The study demonstrates that ML models, such as Random Forest and XGBoost, combined with XAI techniques, can significantly enhance cybersecurity in IoV systems. As the models presently obtained a precision rate of identification of valid and malicious traffic, further calibration through predictive accuracy should be carried out. Predictive models for anomaly detection would predict attack patterns based on historical data and increase the system's resilience. The successful application of XAI in LIME through this work shows the rising demand for interpretability in cybersecurity solutions since it is one of the widely demanded properties to be approached toward availability by being more transparent and trustworthy about how the decision is made by the ML model. This is particularly significant in safety-critical settings. This in turn improves the stakeholders' trust in explaining the AI decisions and hence reliability for deploying AI models in real-time IoV security systems.

Moreover, federated learning offers a transformative approach to addressing privacy and data-sharing challenges in IoV cybersecurity by enabling distributed model training across multiple vehicles without the need to centralize sensitive data. Such an implementation can ensure that raw data remains on the vehicles which significantly reduces the risk of data breaches during transmission. Additionally, the integration of federated learning into IoV cybersecurity systems can enhance scalability. In addition, by keeping the data decentralized and performing computations locally, this approach minimizes the risk of exposing sensitive information, such as location data or vehicle identifiers, to external threats.

Furthermore, federated learning could be combined with blockchain technology to create an immutable and secure ledger that records, updates and prevents tampering during model aggregation. Practical applications of this approach include the development of predictive models that can analyze patterns of IoV traffic in real-time to anticipate and mitigate cybersecurity threats before they escalate. These models could then trigger alerts or automated countermeasures to prevent potential attacks. Further studies are required to evaluate the performance and feasibility of federated learning in real-world IoV scenarios. This includes assessing the communication overhead associated with model parameter exchanges, the robustness of models trained in decentralized settings, and the potential impact on latency and computational efficiency in vehicles. Predictive models, federated learning, and blockchain technologies will undoubtedly play a pivotal role in advancing IoV cybersecurity solutions. Their integration will allow these solutions to proliferate in real-world applications in such a way that the connectivity and autonomy of IoV systems are always guaranteed to be secure, even as the technology evolves.

REFERENCES

[1] Love Allen Chijioke Ahakonye, Cosmas Ifeanyi Nwakanma, Jae Min Lee, and Dong-Seong Kim. 2024. Machine Learning Explainability for Intrusion Detection in the Industrial Internet of Things. *IEEE Internet of Things Magazine* 7, 3 (2024), 68–74.

[2] Jameel Ahmad, Muhammad Umer Zia, Ijaz Haider Naqvi, Jawwad Nasar Chattha, Faran Awais Butt, Tao Huang, and Wei Xiang. 2024. Machine learning and blockchain technologies for cybersecurity in connected vehicles. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 14, 1 (2024), e1515.

[3] M Akın. 2023. Real-Time Threat Intelligence Integration for Cybersecurity in Autonomous Vehicles—A Deep Learning Framework. *Journal of Artificial Intelligence Research and Applications* (2023).

[4] E Alalwany and I Mahgoub. 2024. Security and trust management in the internet of vehicles (IoV): Challenges and machine learning solutions. *Sensors* (2024).

[5] Mohammed Almehdhar, Abdullatif Albaseer, Muhammad Asif Khan, Mohamed Abdallah, Hamid Menouar, Saif Al-Kuwari, and Ala Al-Fuqaha. 2024. Deep learning in the fast lane: A survey on advanced intrusion detection systems for intelligent vehicle networks. *IEEE Open Journal of Vehicular Technology* (2024).

[6] L Almuqren, MS Maashi, M Alamgeer, and H Mohsen. 2023. Explainable AI enabled intrusion detection technique for secure cyber-physical systems. *Applied Sciences* (2023).

[7] Hamed Alqahtani and Gulshan Kumar. 2024. Machine learning for enhancing transportation security: A comprehensive analysis of electric and flying vehicle systems. *Engineering Applications of Artificial Intelligence* 129 (2024), 107667.

[8] Babajide J Asaju. 2024. Advancements in Intrusion Detection Systems for V2X: Leveraging AI and ML for Real-Time Cyber Threat Mitigation. *Journal of Computational Intelligence and Robotics* 4, 1 (2024), 33–50.

[9] Weiping Ding, Ibrahim Alrashdi, Hossam Hawash, and Mohamed Abdel-Basset. 2024. DeepSecDrive: An explainable deep learning framework for real-time detection of cyberattack in in-vehicle networks. *Information Sciences* 658 (2024), 120057.

[10] MV Kadam, HB Mahajan, NJ Uke, and PR Futane. 2023. Cybersecurity threats mitigation in Internet of Vehicles communication system using reliable clustering and routing. *Microprocessors and Microsystems* (2023).

[11] Prabhat Kumar, Danish Javeed, Randhir Kumar, and AKM Najmul Islam. 2024. Blockchain and explainable AI for enhanced decision making in cyber threat detection. *Software: Practice and Experience* (2024).

[12] Pankaj Kumar, Mohammad Wazid, Devesh Pratap Singh, Jaskaran Singh, Ashok Kumar Das, Youngho Park, and Joel JPC Rodrigues. 2023. Explainable artificial intelligence envisioned security mechanism for cyber threat hunting. *Security and Privacy* 6, 6 (2023), e312.

[13] Xingwang Li, Junyao Zhang, Congzheng Han, Wanming Hao, Ming Zeng, Zhengyu Zhu, and Han Wang. 2023. Reliability and security of CR-STAR-RIS-NOMA assisted IoT networks. *IEEE Internet of Things Journal* (2023).

[14] Xingwang Li, Mengle Zhao, Ming Zeng, Shahid Mumtaz, Varun G Menon, Zhiguo Ding, and Octavia A Dobre. 2021. Hardware impaired ambient backscatter NOMA systems: Reliability and security. *IEEE Transactions on Communications* 69, 4 (2021), 2723–2736.

[15] Nour Moustafa, Nickolaos Koroniotis, Marwa Keshk, Albert Y Zomaya, and Zahir Tari. 2023. Explainable intrusion detection for cyber defences in the internet of things: Opportunities and solutions. *IEEE Communications Surveys & Tutorials* 25, 3 (2023), 1775–1807.

[16] S Nazat, L Li, and M Abdallah. 2024. XAI-ADS: An explainable AI framework for enhancing anomaly detection in autonomous driving systems. *IEEE Access* (2024).

[17] Cosmas Ifeanyi Nwakanma, Love Allen Chijioke Ahakonye, Judith Nkechinyere Njoku, Jacinta Chioma Odirichukwu, Stanley Adiele Okolie, Chinebuli Uzondu, Christiana Chidimma Ndubuisi Nweke, and Dong-Seong Kim. 2023. Explainable artificial intelligence (xai) for intrusion detection and mitigation in intelligent connected vehicles: A review. *Applied Sciences* 13, 3 (2023), 1252.

[18] Andrés Páez. 2023. Explainable Reinforcement Learning Models for Adaptive Cyber Defense in Autonomous Vehicles. *Journal of Artificial Intelligence Research and Applications* 3, 2 (2023), 238–261.

[19] Ankit D Patel, Rutvij H Jhaveri, Kaushal A Shah, Ashish D Patel, Rajkumar S Rathore, Manish Paliwal, Kumar Abhishek, and Dhavalkumar Thakker. 2024. Security Trends in Internet-of-things for Ambient Assistive Living: A Review. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)* 17, 7 (2024), 18–46.

[20] Tanish Patel, Ramalingam Murugan, Gokul Yenduri, Rutvij Jhaveri, Hichem Snoussi, and Tarek Gaber. 2024. Demystifying Defects: Federated Learning and Explainable AI for Semiconductor Fault Detection. *IEEE Access* (2024).

[21] Preeti Rani, Chandani Sharma, Janjhyam Venkata Naga Ramesh, Sonia Verma, Rohit Sharma, Ahmed Alkhayyat, and Sachin Kumar. 2023. Federated learning-based misbehaviour detection for the 5G-enabled internet of vehicles. *IEEE Transactions on Consumer Electronics* (2023).

[22] Gaith Rjoub, Jamal Bentahar, Omar Abdel Wahab, Rabeb Mizouni, Alyssa Song, Robin Cohen, Hadi Otrok, and Azzam Mourad. 2023. A survey on explainable artificial intelligence for cybersecurity. *IEEE Transactions on Network and Service Management* 20, 4 (2023), 5115–5140.

[23] Ponmagal RS. 2024. An intelligent dynamic cyber physical system threat detection system for ensuring secured communication in 6G autonomous vehicle networks. *Scientific Reports* 14, 1 (2024), 20795.

[24] X Wang, H Zhu, Z Ning, and L Guo. 2023. Blockchain intelligence for internet of vehicles: Challenges and solutions. *IEEE Communications Surveys & Tutorials* (2023).