future internet

MDPI

Article

Multi-Class Intrusion Detection in Internet of Vehicles: Optimizing Machine Learning Models on Imbalanced Data

Ágata Palma [1], Mário Antunes [2,3], Jorge Bernardino [1,4,*] and Ana Alves [1,5]

[1] Institute of Engineering of Coimbra—ISEC, Polytechnic University of Coimbra, Rua Pedro Nunes, Quinta da Nora, 3030-199 Coimbra, Portugal; a2023113935@isec.pt (Á.P.); aalves@isec.pt (A.A.)
[2] School of Technology and Management, Polytechnic University of Leiria, 2411-901 Leiria, Portugal; mario.antunes@ipleiria.pt
[3] INESC TEC, CRACS, 4200-465 Porto, Portugal
[4] CISUC, SSE, University of Coimbra, Polo II, Pinhal de Marrocos, 3030-290 Coimbra, Portugal
[5] CISUC, LASI, University of Coimbra, Polo II, Pinhal de Marrocos, 3030-290 Coimbra, Portugal
* Correspondence: jorge@isec.pt

Abstract: The Internet of Vehicles (IoV) presents complex cybersecurity challenges, particularly against Denial-of-Service (DoS) and spoofing attacks targeting the Controller Area Network (CAN) bus. This study leverages the CICIoV2024 dataset, comprising six distinct classes of benign traffic and various types of attacks, to evaluate advanced machine learning techniques for instrusion detection systems (IDS). The models XGBoost, Random Forest, AdaBoost, Extra Trees, Logistic Regression, and Deep Neural Network were tested under realistic, imbalanced data conditions, ensuring that the evaluation reflects real-world scenarios where benign traffic dominates. Using hyperparameter optimization with Optuna, we achieved significant improvements in detection accuracy and robustness. Ensemble methods such as XGBoost and Random Forest consistently demonstrated superior performance, achieving perfect accuracy and macro-average F1-scores, even when detecting minority attack classes, in contrast to previous results for the CICIoV2024 dataset. The integration of optimized hyperparameter tuning and a broader methodological scope culminated in an IDS framework capable of addressing diverse attack scenarios with exceptional precision.

Keywords: cybersecurity; machine learning; imbalanced data; internet of vehicles (IoV); intrusion detection system (IDS); multi-class classification

Academic Editors: Olusola Tolulope Odeyomi and Temitayo Olowu

Received: 12 February 2025
Revised: 25 March 2025
Accepted: 1 April 2025
Published: 7 April 2025

Citation: Palma, Á.; Antunes, M.; Bernardino, J.; Alves, A. Multi-Class Intrusion Detection in Internet of Vehicles: Optimizing Machine Learning Models on Imbalanced Data. *Future Internet* **2025**, *17*, 162. https://doi.org/10.3390/fi17040162

1. Introduction

The Internet of Vehicles (IoV) is a cornerstone of modern intelligent transportation systems, merging vehicular communication technologies with cloud-based infrastructures to improve safety, efficiency, and real-time decision-making on the road. By enabling vehicle-to-everything (V2X) communication, IoV supports applications ranging from autonomous driving to real-time traffic management [1,2].

At the heart of vehicle communication is the Controller Area Network (CAN), a critical component of modern automotive systems. It was originally designed for use in vehicles, but also has applications in other industries such as aerospace, medical devices, and commercial machinery [3]. The CAN protocol is a lightweight and robust communication standard designed for real-time data exchange between Electronic Control Units (ECU) in vehicles [3–5]. Introduced in the 1980s, the CAN bus enables efficient message transmission without the need for a centralized control mechanism. Its ability to prioritize messages using an arbitration-based system ensures reliability even under high traffic loads [1,5,6].

The CAN protocol was originally created for conventional isolated vehicles, without incorporating any security features [3,7]. As a result, it lacks essential security features such as encryption and authentication, making it vulnerable to various cyberattacks and expanding the attack surface of vehicular systems [1,8]. Attackers can exploit these weaknesses to inject spoofed messages, disrupt communication, or cause malfunctioning in safety-critical components, demonstrating that the ability to identify intrusions in the CAN bus is a critical component of IoV security [1,4].

To address these critical security challenges, instrusion detection systems (IDS) have emerged as a critical defense mechanism in IoV environments [9–12]. These systems rely on Traditional Machine Learning (TML), Deep Learning (DL) and Deep Transfer Learning (DTL) techniques to detect anomalous patterns in network traffic that indicate potential ongoing cyberattacks [8,13]. Despite the significant transition from traditional signature-based IDS to anomaly-based systems [14], several challenges remain. One of the primary issues is the inherent imbalance in IoV datasets, where benign traffic far outweighs malicious instances, potentially leading to a biased machine learning (ML) model that fails to effectively detect minority attack classes. Another obstacle is the lack of realistic and recent IoV datasets, which poses a significant challenge to the advancement of IDS technology for IoV and hinders the ability to train and evaluate the effectiveness of these systems in realistic scenarios. Furthermore, the multi-class nature of IoV traffic requires IDS to distinguish between several types of attacks, often with overlapping characteristics [8]. Existing IDS approaches often fall short of providing robust performance in real-world IoV scenarios, that is, the ability of intrusion detection models to generalize to real-world traffic conditions, maintain high accuracy and consistency across metrics, and detect minority attack classes without bias toward dominant benign traffic.

The CICIoV2024 dataset [8] provides a benchmark for addressing these challenges. Collected from a fully equipped vehicular testbed, including the CAN bus, this dataset contains realistic vehicular traffic data, incorporating various attack scenarios such as Denial-of-Service (DoS), Gas Spoofing, Speed Spoofing, Steering Wheel Spoofing, and Revolutions Per Minute (RPM) Spoofing within an imbalanced distribution that mirrors real-world scenarios, making it an ideal dataset for evaluating IDS models under these conditions [8]. Although existing studies have applied traditional ML algorithms, such as Random Forest [5] and XGBoost [15], as well as DL architectures such as Convolutional Neural Network (CNN) [16], many have relied on artificially balanced datasets that do not capture the challenges of real-world deployments.

This study builds on previous work [8,17] by using the CICIoV2024 dataset [8] to evaluate machine learning algorithms, including Logistic Regression, Random Forest, XGBoost, ExtraTrees, and AdaBoost, without applying data balancing techniques. Optuna [18], an advanced optimization tool, is used to fine-tune model hyperparameters to ensure optimal performance. Early stopping mechanisms are also implemented to mitigate overfitting in models that are trained iteratively. By preserving the natural imbalance of the dataset, this research provides a realistic assessment of model performance in multi-class IoV environments, where benign traffic dominates.

Unlike traditional IDS-focused research, this study goes beyond intrusion detection to explore how machine learning methods, including Logistic Regression and ensemble techniques, can be adapted and optimized to address the challenges inherent in IoV traffic data. Through comprehensive analysis and optimization, we aim to contribute to a broader understanding of scalable and efficient machine learning models. These findings extend beyond security applications to address critical challenges in real-time data processing and resource-efficient model design for IoV systems. This study seeks to answer the following question:

RQ: How can machine learning models be optimized to handle imbalanced and multi-class IoV traffic data and provide robust performance in different attack scenarios?

To address these challenges and the research question, this study makes the following key contributions:

- CICIoV2024 benchmark analysis: A thorough analysis of machine learning methods such as Random Forest, AdaBoost, Logistic Regression, and Deep Neural Network (DNN) was performed on the CICIoV2024 dataset without the use of data balancing strategies. This ensures that the analysis and results reflect real-world, imbalanced traffic conditions.
- Benchmark ML models addition: Two ensemble models were added to the benchmark, Extra Trees and XGBoost. This allows for a more extensive analysis with our approach.
- Advanced Optimization: We ensure that models are not only accurate, but also computationally efficient, reducing the risk of overfitting by using Optuna to automate the search process for optimal hyperparameters and implement early stopping mechanisms.
- Benchmark improvement: With the addition of different algorithms and the use of Optuna, we were able to improve the results for all the models in the benchmark and added two more.
- Multi-Class Performance Analysis: This study highlights the model's strengths and weaknesses in detecting minority classes by thoroughly analyzing its performance in a variety of attack scenarios, including DoS, gas spoofing, speed spoofing, steering wheel spoofing, and RPM spoofing.
- Broader Implications Beyond Security: Although IDS provides a foundational framework, this research advances our knowledge of machine learning techniques for processing data in real time and creating resource-efficient models for Internet of Vehicles systems.

The remainder of this article is structured as follows. Section 2 provides a review on existing research on machine learning applications for IoV security. Section 3 describes the CICIoV2024 dataset, preprocessing techniques, and the experimental setup, including model implementations and optimization strategies. In Section 4, we present the results of the experiments and discuss them in Section 5. Finally, Section 6 summarizes the main contributions of this study and outlines possible future research directions to address the identified challenges.

2. Related Work

Instrusion detection systems are indispensable for securing IoV networks, which are increasingly vulnerable to sophisticated cyberattacks due to their highly interconnected and dynamic nature. Recent advances in machine learning, deep learning, federated learning, and optimization techniques have contributed significantly to addressing the unique challenges of IoV intrusion detection. This section reviews the existing literature with a focus on IoV security, highlighting key methodologies and open challenges, particularly in the context of CAN bus vulnerabilities and class imbalance in intrusion detection datasets.

The CICIoV2024 dataset, created and published by Neto et al. [8], is a public and realistic benchmark for evaluating IDS in IoV environments, incorporating imbalanced traffic distributions and diverse attack scenarios. However, its adoption by the research community remains limited. Neto et al. [8] presented the CICIoV2024 dataset and evaluated various machine learning models, including Random Forest, AdaBoost, Logistic Regression, and Deep Neural Networks. The authors emphasized the multi-representation of the dataset (binary, decimal, and hexadecimal) and highlighted its suitability for spoofing and DoS attack detection. However, their work focused primarily on baseline model

evaluations without integrating advanced optimization techniques, and the performance of the models dropped significantly when presented with multi-class analysis, leaving room for performance improvement.

Following the lead of Neto et al. [8], Gul and Bakir [17] used the CICIoV2024 dataset with Genetic Algorithm (GA)-based hyperparameter optimization to improve model performance. Their study showed significant accuracy improvements for Random Forest (99.64%) and Deep Neural Networks, demonstrating the potential of GA for fine-tuning IDS models. Despite its success, the computational complexity of GA may limit its scalability in real-time IoV environments. Neither study applied data balancing techniques.

Beyond the CICIoV2024 dataset, other studies have focused on specific methods for improving intrusion detection in CAN environments. Jin et al. [19] also took advantage of the GA optimization for feature selection, combined with a balanced dataset and LightGBM, resulting in a perfect accuracy and F1-score in two different CAN datasets. El-Gayar et al. [10] proposed an ensemble machine learning model that includes Random Forest(RF), XGBoost, LightGBM, and Extra Trees (ET) on a CAN dataset. The results of the ensemble model show an accuracy and F1-score of 98% and 96.9%, respectively, and for the single ML models, these values were as follows: 97.7% and 94.9% with Decision Trees; 98.2% and 93.2% with RF; 95.5 and 91.1 with XGBoost. Although the results seem promising, the dataset was also balanced, which means that these models are trained on datasets that do not accurately reflect the imbalanced nature of real-world scenarios.

Singh et al. [20] presented SELIDS, a stacking-enabled ensemble learning-based IDS for IoV. Their approach integrates RF, LightGBM, XGBoost, and CatBoost to improve intrusion detection performance, achieving near-perfect accuracy and F1-scores. Du et al. [21] presented CLUSTER, a clustering-based open-set recognition framework tailored for CAN bus traffic. Their approach uses intra-class compactness and inter-class separation to improve the recognition of unknown types of attacks. This is particularly relevant for CAN, where attacks often exhibit subtle variations that are difficult to detect.

In addition to traditional learning models, researchers have increasingly explored DL techniques for intrusion detection in IoV systems. DL models are known for their ability to easily learn complex feature representations, and have shown significant promise, particularly in detecting subtle and previously unseen attack patterns [13]. For example, Yang et al. [16] proposed a novel framework based on Convolutional Neural Networks with various hyperparameters optimizations that achieved perfect accuracy and F1-scores when applied to CAN bus data.

Conversely, traditional machine learning models remain competitive in certain scenarios due to their lower computational requirements and ease of interpretability. For example, Ullah et al. [22] developed HDL-IDS, a hybrid IDS architecture that combines convolutional and recurrent layers to capture both spatial and temporal features of CAN traffic. By analyzing time-series data from CAN messages, their system achieved an F1-score of over 99.9%. Mehedi et al. [3] used a Deep DTL-based IDS, and compared the results with TML and DL models. Although the highest F1-score achieved was in the proposed DTL model, with a value of 97.83%, the results of the TML models are higher than the DL models, specifically with Random Forest, Decision Trees, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). Also, Wang et al. proposed a lightweight IDS using MobileNetV2 [23] and MobileNetV3 [12], which achieved perfect accuracy and F1-scores but limited adaptability to novel attacks. Fu et al. [24] introduced IoV-BERT-IDS, which applies Large Language Models (LLMs) such as BERT to extract contextual semantics from CAN bus traffic. This innovative approach improved detection rates for complex attack scenarios, resulting in a 99.97% accuracy and 99.85% F1-score.

The focus on CAN-specific intrusion detection underscores the critical need for methods that prioritize optimization and effectively address the challenges posed by imbalanced datasets. Real-world CAN environments are inherently imbalanced, with benign traffic significantly outnumbering malicious instances, making it essential to evaluate models in realistic scenarios. The literature review highlights the consistent high performance of TML even under these challenging conditions. Their ability to maintain high detection rates and F1-scores across multiple types of attacks without relying on computationally intensive methods demonstrates the practicality and efficiency of TML models in resource-constrained vehicular systems. While advanced techniques such as DL and DTL models provide precision, optimization-driven approaches using TML models continue to be highly effective in ensuring scalable and reliable CAN intrusion detection in real-world applications, highlighting the importance of both methods.

In contrast to previous studies, our research focuses on evaluating machine learning models on the CICIoV2024 dataset [8] without applying artificial data balancing techniques, ensuring a more realistic evaluation of model performance under imbalanced conditions. In addition, we employ advanced hyperparameter optimization using Optuna [18] to provide fine-tuned models that balance detection accuracy across all traffic classes, including minority attacks. This comprehensive approach addresses the gaps in existing research and delivers practical and effective solutions tailored to real-world IoV environments.

3. Methodology

This study uses a systematic approach to evaluate the performance of multiple machine learning algorithms for multi-class intrusion detection in the Internet of Vehicles (IoV). The methodology integrates model training and evaluation practices to ensure an efficient analysis as shown in Figure 1. In the following, we describe the main components of the methodology.

3.1. Experimental Setup

The experimental setup, summarized in Table 1, provided the necessary resources for hyperparameter tuning and model evaluation, ensuring a successful execution of all computational tasks.

Table 1. Details of the experimental setup - ASUS laptop.

Hardware	Software
Processor: Intel® Core™ i7-8750H	Windows 11, 64-bit
CPU @ 2.20 GHz, 2.21 GHz	Jupyter Lab 4.2.5
RAM: 16.00 GB, SSD: 1TB + 256 GB	Python 3.12.7
Graphics card: Nvidia® GeForce™ GTX 1050 Ti 4GB VRAM	Optuna 4.0.0

3.2. Dataset

The experiments were conducted using the CICIoV2024 dataset [8], which stands out as the most recent and comprehensive benchmark for advancing cybersecurity research in the Internet of Vehicles (IoV) domain, specifically in CAN environments. The dataset was developed using a fully intact inner structure of a 2019 Ford vehicle, using the widely adopted standard CAN protocol for intra-vehicular communication among Electronic Control Units. While the CAN protocol is efficient for real-time communication, its lack of built-in security features exposes it to various cyber threats, making intrusion detection systems (IDS) essential for vehicular safety and reliability. This dataset captures vehicular network traffic, including both benign and malicious scenarios. It includes six different classes—benign traffic and five types of attacks: Denial-of-Service (DoS), Gas Spoofing,

Speed Spoofing, Steering Wheel Spoofing, and Revolutions Per Minute (RPM) Spoofing. Each attack scenario reflects realistic exploitation methods targeting the CAN bus, such as flooding the network to disrupt communication (DoS) or manipulating critical control data such as speed, steering angle, and gas pedal position.

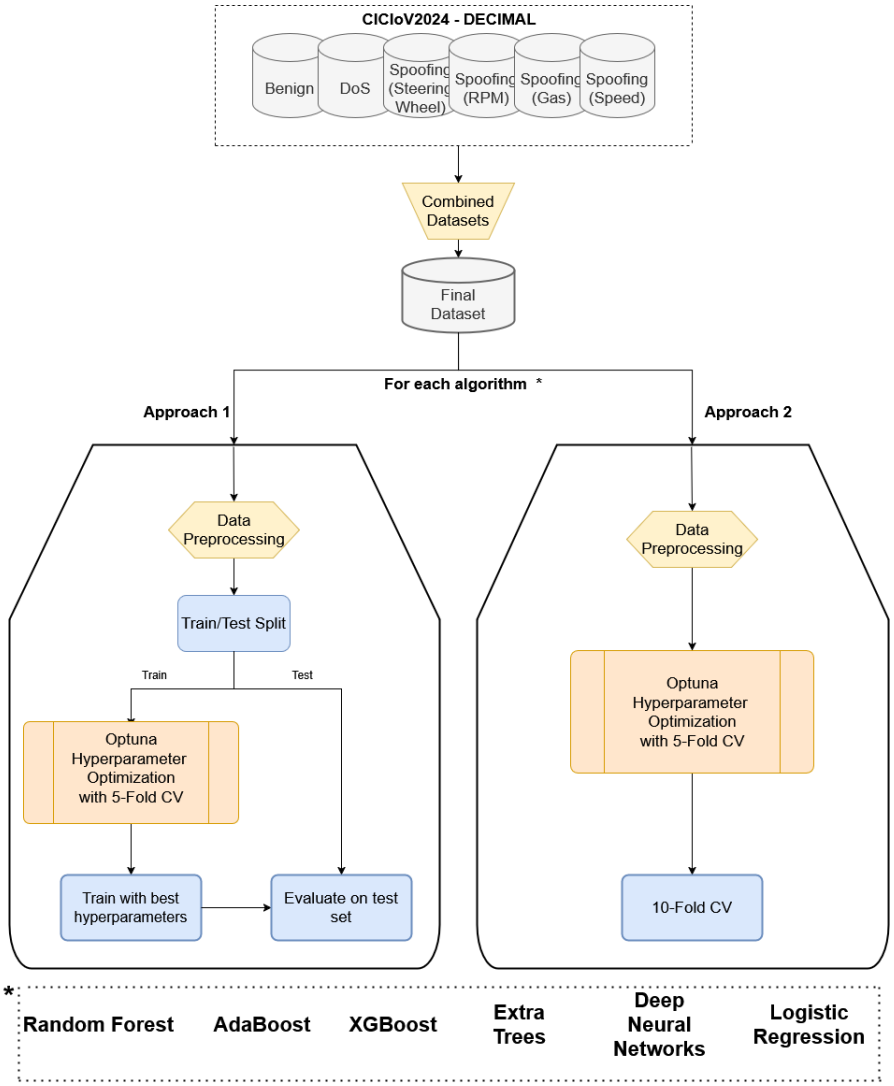

Figure 1. ML models evaluation process.

A key strength of CICIoV2024 [8] is its realism, which extends beyond the fidelity of the network traffic to include its imbalanced nature, where benign data significantly outweigh malicious instances. The data imbalance reflects real-world conditions and presents a challenge for IDS development, as minority attack classes are more difficult to detect. In addition, the features of the dataset, detailed in Table 2, include three formats—binary, decimal, and hexadecimal—providing flexibility. In this research, we used only the decimal representation of the dataset, which provides the data as numerical values suitable for machine learning analysis.

Table 2. Description of dataset features.

Feature	Description
ID	Represents the message priority and specifies the type of data being transmitted—not unique
DATA_0	Byte 0 of the transmitted data
DATA_1	Byte 1 of the transmitted data
DATA_2	Byte 2 of the transmitted data
DATA_3	Byte 3 of the transmitted data
DATA_4	Byte 4 of the transmitted data
DATA_5	Byte 5 of the transmitted data
DATA_6	Byte 6 of the transmitted data
DATA_7	Byte 7 of the transmitted data
label	Indicates whether the traffic is benign or an attack
category	Specifies the traffic classification as benign, DoS, or spoofing
specific_class	Specifies the precise class of the traffic as benign, DoS, spoofing-GAS, spoofing-RPM, spoofing-SPEED, and spoofing-STEERING_WHEEL

3.3. Data Preprocessing

In order to ensure the best performance and effectiveness of machine learning models for intrusion detection, the CICIoV2024 dataset [8] underwent several preprocessing steps to transform the data into a format suitable for training and evaluation. Label encoding and feature scaling are critical steps in the preprocessing pipeline to ensure that the dataset is optimized for machine learning. To prepare the target variable, `specific_class`, for modeling, label encoding was applied, converting categorical labels (e.g., "DoS", "Spoofing") to numerical values to ensure compatibility with machine learning algorithms. To account for differences in feature size, standard scaling was applied, transforming all numerical features to have a mean of 0 and a standard deviation of 1. This step was particularly important to optimize the performance of scale-sensitive models, such as Logistic Regression, while maintaining the integrity of tree-based methods. Together, these preprocessing steps collectively ensured that the dataset was structured, normalized, and ready for effective model training and evaluation. We did not apply resampling techniques, such as oversampling or undersampling, to maintain the dataset's natural distribution in order to preserve the natural distribution of the dataset and evaluate model performance under realistic conditions.

3.4. Optuna Optimization

Optuna [18] is an open-source framework for automating hyperparameter optimization. It is framework-agnostic, which means that it is not tied to a specific library, platform, or tool. It can be integrated with different machine learning frameworks, libraries, or workflows, allowing users to perform hyperparameter optimization regardless of the specific ML framework they are using.

In our research, we used Optuna to optimize hyperparameters over 10 trials, with each trial representing a unique set of hyperparameters. To ensure a rigorous evaluation of model performance, a five-fold cross-validation was incorporated for each trial. This approach ensures that the performance of each hyperparameter set is averaged over multiple training and validation splits, reducing overfitting and providing a reliable estimate of generalization. The optimized hyperparameters for each model are listed in Table 3 with the corresponding values found. In total, the methodology involved 50 individual training and validation runs for each model, balancing computational efficiency with rigorous evaluation.

Table 3. Optimized hyperparameters for split and 10-fold training.

Model	Hyperparameter	Split	10-Fold
RF	n_estimators	112	57
	max_depth	25	8
	min_samples_split	2	6
AdaBoost	n_estimators	169	91
	learning_rate	0.22031826639728547	0.868649991657305
XGBoost	n_estimators	346	338
	max_depth	4	6
	learning_rate	0.08257351184613305	0.12486607179583395
	subsample	0.820421224719867	0.8398820455022838
	colsample_bytree	0.6442582375605579	0.8885000829639664
LR	C	0.43873275517235	3.2376253329526534
	penalty	l1	l1
	solver	saga	saga
ET	n_estimators	94	50
	max_depth	16	34
DNN	layers	4	4
	units	128	96
	dropout_rate	0.2774088047541644	0.24552040101739658
	learning_rate	0.0009501230980605523	0.00010682901273334459
	batch_size	112	128
	epochs	45	34

3.5. Training Methods

We used two different training methods: (i) the train–test split approach and (ii) a 10-fold cross-validation to ensure a reliable evaluation of the machine learning models. In both approaches, hyperparameter optimization was performed using Optuna [18] with five-fold cross-validation, validated against the macro-averaged F1-score. This ensured that hyperparameter tuning focused on consistent performance across all classes, which is particularly important when dealing with imbalanced datasets.

3.5.1. Train–Test Split Approach

In the train–test split approach, the dataset was divided into 80% for training and 20% for testing sets, with stratified sampling to maintain the class distribution. After determining the best hyperparameters using Optuna, each model was retrained on the 80% training set and then evaluated on the 20% test set to measure generalization performance on unseen data.

3.5.2. 10-Fold Cross-Validation Approach

In the 10-fold cross-validation approach, the entire dataset was used for training and validation. After hyperparameter optimization, the models were trained using 10-fold cross-validation with the best parameters identified. Since XGBoost and DNN are prone to overfitting, early stopping was applied to both models. Performance metrics, including precision, recall, accuracy, and F1-score, were averaged across the folds to provide a rigorous evaluation of the model's capabilities.

4. Experimental Evaluation

The performance of the machine learning models was evaluated using both the train–test split and 10-fold cross-validation approaches, as described in Section 3. The results highlight the effectiveness of the proposed methods in addressing the challenges of im-

balanced multi-class intrusion detection. Key performance metrics, including precision, recall, accuracy, and F1-score, were used to evaluate model performance. Table 4 shows the execution time, in seconds, for the hyperparameter optimization and training of each model in both training approaches. The results highlight a critical trade-off between computational efficiency and model performance. While models such as Random Forest, Extra Trees, and XGBoost demonstrated perfect classification performance with relatively low computational costs, DNN required a significantly higher processing time, particularly during the hyperparameter optimization and training. In contrast, Random Forest and Extra Trees maintained both high performance and efficiency. In addition, AdaBoost and Logistic Regression also revealed some challenges in computational efficiency.

Table 4. Comparison of execution time.

Models	Train/Test Split (s) [1]		10-Fold CV (s) [1]	
	Optuna	Training	Optuna	Training
Random Forest	1830	120	1971	238
AdaBoost	5620	191	5935	496
XGBoost	1820	106	3512	1532
Logistic Regression	1745	1116	3997	3647
Extra Trees	1819	47	3271	221
Deep Neural Networks	11,010	1022	40,387	5104

[1] All times have been rounded to the nearest integer for simplicity. Time in seconds.

4.1. Train–Test Split Approach

The results of the train–test split approach are summarized in Table 5, which compares the performance of the models evaluated in this research with previous works using the same dataset and models, with the best results of the experiments highlighted in bold. The models were trained on 80% of the dataset using the best hyperparameters determined by Optuna and evaluated on the remaining 20%. Logistic Regression achieved an accuracy of 97%, with an F1-score of 97%, demonstrating strong generalization capabilities across all classes. AdaBoost achieved 98% accuracy, with challenges reflected in lower recall (81%) and precision (80%), likely due to the imbalanced nature of the dataset. RF, ET, XGBoost, and DNN demonstrated exceptional performance, achieving perfect scores (100%) across all metrics, highlighting their effectiveness in accurately detecting both majority and minority classes.

Table 5. Comparison of model performance across different studies using split approach on CICIoV2024 dataset.

Models	Neto et al. [8]				Gul and Bakir [17]				Our Study			
	Acc.	Rec.	Prec.	F1	Acc.	Rec.	Prec.	F1	Acc.	Rec.	Prec.	F1
LR	0.89	0.50	0.48	0.49	0.90	0.60	0.72	0.62	0.97	0.86	0.98	0.88
AdaBoost	0.92	0.66	0.48	0.51	0.97	0.73	0.75	0.72	0.98	0.82	0.80	0.81
RF	0.96	0.76	0.76	**0.76**	1.00	0.97	0.99	**0.97**	1.00	1.00	1.00	**1.00**
DNN	0.96	0.76	0.83	**0.78**	0.99	0.97	0.97	**0.97**	1.00	1.00	1.00	**1.00**
XGBoost	–	–	–	–	–	–	–	–	1.00	1.00	1.00	**1.00**
ET	–	–	–	–	–	–	–	–	1.00	1.00	1.00	**1.00**

4.2. 10-Fold Cross-Validation Approach

The results of the 10-fold cross-validation approach are presented in Table 6. Each model was trained using the best parameters identified by Optuna optimization, and the performance metrics were averaged across 10 folds. Random Forest, XGBoost, DNN, and Extra Trees achieved consistent and perfect performance across all metrics (100% precision,

recall, accuracy, and F1-score). Logistic Regression achieved an accuracy of 97% and and 88% F1-score, indicating its sensitivity to imbalanced datasets despite its competitive performance. AdaBoost struggled with imbalanced class distributions, achieving an F1-score of 32% but maintaining a high accuracy of 92%. The results confirm the observations obtained from the train–test split approach and underscore the reliability of ensemble methods and deep learning models in detecting different types of attacks with high precision and recall.

Table 6. Performance comparison of different models using 10-fold training.

Model	Precision	Recall	Accuracy	F1-Score
Random Forest	1.00	1.00	1.00	1.00
AdaBoost	0.30	0.33	0.92	0.32
XGBoost	1.00	1.00	1.00	1.00
Logistic Regression	0.98	0.86	0.97	0.88
Extra Trees	1.00	1.00	1.00	1.00
Deep Neural Networks	1.00	1.00	1.00	1.00

5. Discussion

Our study evaluated the performance of different machine learning models for multi-class intrusion detection in the Internet of Vehicles (IoV) using the CICIoV2024 dataset [8]. We used two different training methods: a train–test split, and a 10-fold cross-validation approach, to rigorously evaluate the performance of the models. Table 5 (Section 4) compares our results using the train–test split with those reported by Neto et al. [8] and Gul and Bakir [17]. Our optimized models, particularly Random Forest, Deep Neural Network, XGBoost, and Extra Trees, showed exceptional performance across all metrics. In contrast, the authors in [8,17] reported significantly lower performance, especially in precision and recall. Due to our optimization strategy, we were able to achieve perfect results, while Neto et al. [8] achieved the highest F1-score of 0.78 with DNN and 0.76 with RF, and Gul and Bakir [17] an F1-score of 0.9679 and 0.9745 for the same models, respectively. However, when we changed the training approach to a 10-fold cross-validation, our results, shown in Table 6, for AdaBoost were worse for every metric, compared to Neto et al. [8], Gul and Bakir [17], and our results in the train–test setting. Regarding the remaining tested models in the 10-fold cross-validation RF, DNN, ET, XGBoost, and LR-we were still able to obtain optimal values, with perfect scores for RF, ET, XGBoost, and DNN. Logistic Regression, while not achieving perfect scores, still maintained high performance, indicating its suitability for this task.

The consistent high performance of RF, XGBoost, ET, and DNN across both evaluation strategies underscores their effectiveness in handling imbalanced multi-class datasets. This can be attributed to the combination of hyperparameter optimization, which fine-tunes key parameters such as tree depth, learning rate, and the number of estimators, and its inherent ability to adapt to imbalanced datasets. In particular, tree-based ensemble methods, such as RF and ET, along with XGBoost, exploited the randomness and sequential learning inherent in their algorithms to effectively handle the skewed distribution of benign and malicious traffic. Although DNN is prone to overfitting with complex datasets, we used its deep structure to learn intricate patterns in the data, thus contributing to its perfect performance against all types of attacks.

In comparison, AdaBoost faced notable challenges, particularly in detecting minority attacks and in efficiency with a considerably high computational cost. Although it showed reasonable accuracy, its recall dropped significantly, reflecting the model's sensitivity to class imbalance in the dataset. AdaBoost iteratively adjusts the weights for misclassified instances, which in imbalanced data can amplify the influence of the dominant class, thereby reducing its ability to effectively detect the less common attack types. This underperfor-

mance is consistent with AdaBoost's limitation when dealing with imbalanced datasets [25], where additional techniques, such as adjusting class weights, are often required to improve its performance in similar real-world scenarios.

While logistic regression did not achieve perfect scores and found some challenges in computational efficiency, it maintained competitive results and outperformed AdaBoost, which is consistent with [26] when dealing with imbalanced datasets. Due to its relatively simple nature and reliance on linear decision boundaries, it was not able to capture complex interactions and patterns within the dataset as effectively as the ensemble methods and deep learning models. However, its solid performance with a high recall score demonstrates that it can still be a viable option for intrusion detection, especially when interpretability is a key consideration. The F1-score, which balances precision and recall, remained relatively high, further emphasizing the efficiency of logistic regression in handling multi-class classification tasks where the dataset is imbalanced but not overly complex.

The imbalanced nature of the CICIoV2024 dataset [8] is critical to understanding the applicability of these models in the real world. By preserving the original class distribution of the dataset, we ensured that the evaluation reflected real-world conditions where benign traffic overwhelmingly outnumbers attack traffic. Random Forest, XGBoost, Extra Trees, and DNN were able to mitigate the impact of this imbalance through their structural characteristics and optimization techniques. However, AdaBoost and Logistic Regression were not as adept, underscoring the importance of carefully selecting and tuning models when dealing with imbalanced real-world data.

The use of Optuna [18] for hyperparameter optimization contributed significantly to the performance improvements. The precise tuning of hyperparameters allowed the models to achieve their full potential by addressing specific challenges such as high-dimensional feature spaces and the imbalanced nature of the dataset. For example, Random Forest and XGBoost benefited from adjustments to tree depth and learning rates, which allowed these models to better fit the data without overfitting. DNN, on the other hand, required early stopping to avoid overfitting due to the complexity of the dataset, demonstrating the nuanced nature of optimizing deep learning models for real-world datasets.

While DNN has demonstrated impressive performance, it also presents significant computational challenges. As seen in Table 4 (Section 4), training deep neural networks requires significant computational resources, particularly when dealing with large datasets and complex architectures. The need for high memory capacity and processing power, especially during hyperparameter optimization and training iterations, makes DNN less practical for environments with limited resources or real-time processing requirements. This aspect is critical when considering the use of DNN-based IDS in IoV systems, as it may require specialized hardware or distributed computing setups to operate efficiently in real-world scenarios.

Based on our findings, we can now answer the original research question posed-"How can machine learning models be optimized to handle imbalanced and multi-class IoV traffic data and provide robust performance in different attack scenarios?". The answer is that machine learning models can be optimized for imbalanced, multi-class IoV traffic data by using ensemble-based algorithms such as Random Forest, Extra Trees, and XGBoost, coupled with systematic hyperparameter tuning using tools such as Optuna. These optimizations include adjusting tree depth, learning rates, the number of estimators, and sampling strategies within the models to ensure balanced performance across all classes.

6. Conclusions and Future Work

This research addressed the challenge of optimizing machine learning models for imbalanced, multi-class intrusion detection in the Internet of Vehicles. The results show

that ensemble models such as Random Forest, Extra Trees, and XGBoost, when paired with targeted hyperparameter tuning using Optuna [18], consistently deliver high performance. These models not only handled the imbalance in the dataset but also successfully identified minority attacks, such as Gas Spoofing and RPM Spoofing, which are critical for real-world deployment.

Deep Neural Networks also proved effective, although they required additional steps, such as early stopping to prevent overfitting. Simpler models, such as Logistic Regression and AdaBoost, were less effective in this context, as they struggled with minority class detection and were not as adaptable to the imbalanced data structure. These results highlight the importance of using realistic datasets such as CICIoV2024 [8]. By preserving the natural traffic distribution, we were able to test how well the models perform under real-world conditions, where benign traffic vastly outweighs malicious attacks. This approach avoids the risks of creating artificial scenarios that do not replicate real-world IoV environments. The study also underscores the need for IDS to evolve with the increasing complexity of connected vehicular systems to effectively address emerging threats.

Despite these promising results, there are several limitations that need to be addressed. One concern is the potential for overfitting as the results for RF, ET, XGBoost, and DNN achieved perfect scores. This could be the result of very distinct attack signatures. Nonetheless, further research is needed with different approaches to mitigate the risk for overfit, particularly with more complex models such as DNN, which may not generalize well to unseen data despite performing well on the training sets. This is particularly relevant in IoV environments, where new and unforeseen attack patterns may emerge. Overfitting could limit the model's ability to adapt when facing new threats, making it critical to implement regularization techniques or explore methods such as cross-validation across different datasets to improve model efficiency.

Additionally, while this study focused on a specific set of models, the need for diverse model exploration remains critical. By increasing the range of models considered, we could potentially improve detection performance for different types of attacks. Different models bring different strengths to the table, and exploring how combinations of models or even hybrid approaches could improve results could prove invaluable, as seen in previous studies [3,10,20,22]. Diversifying the model pool would also help to mitigate issues of model-specific limitations and provide greater flexibility in detecting complex or evolving threats in IoV systems.

Going forward, several opportunities for further research remain. Real-time detection methods that adapt to changing attack patterns could make these systems more practical for deployment. Incorporating domain-specific features of the CAN protocol could also improve accuracy and efficiency. In addition, the evaluation of these models in different IoV environments and with different datasets will be critical to confirm their broad applicability and identify areas for improvement. Ultimately, these advances underscore the importance of using realistic datasets and continuously refining IDS strategies to improve the security and reliability of vehicular networks in practical settings. This work provides a foundation for future efforts to develop more adaptable and scalable solutions for securing connected vehicles.

Author Contributions: Conceptualization, methodology, data curation: Á.P., M.A. and A.A.; visualization, investigation, software, writing—original draft preparation: Á.P.; formal analysis, validation, writing—review and editing: Á.P., M.A., A.A. and J.B.; project administration, supervision: A.A., M.A. and J.B.; funding acquisition: J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset used is available online at https://www.unb.ca/cic/datasets/iov-dataset-2024.html (accessed on 25 September 2024) and the source code of the developed ML models can be found at https://github.com/AgataPalma/MLNotebooks (accessed on 4 January 2025).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Taslimasa, H.; Dadkhah, S.; Neto, E.C.P.; Xiong, P.; Ray, S.; Ghorbani, A.A. Security issues in Internet of Vehicles (IoV): A comprehensive survey. *Internet Things* **2023**, *22*, 100809. [CrossRef]
2. Gong, W.; Yang, S.; Guang, H.; Ma, B.; Zheng, B.; Shi, Y.; Li, B.; Cao, Y. Multi-order feature interaction-aware intrusion detection scheme for ensuring cyber security of intelligent connected vehicles. *Eng. Appl. Artif. Intell.* **2024**, *135*, 108815. [CrossRef]
3. Mehedi, S.T.; Anwar, A.; Rahman, Z.; Ahmed, K. Deep transfer learning based intrusion detection system for electric vehicular networks. *Sensors* **2021**, *21*, 4736. [CrossRef]
4. Wang, S.; Zheng, B.; Liu, Z.; Fan, Z.; Liu, Y.; Dai, Y. A Lightweight Intrusion Detection System for Vehicular Networks Based on an Improved ViT Model. *IEEE Access* **2024**, *12*, 118842–118856. [CrossRef]
5. Moulahi, T.; Zidi, S.; Alabdulatif, A.; Atiquzzaman, M. Comparative Performance Evaluation of Intrusion Detection Based on Machine Learning in In-Vehicle Controller Area Network Bus. *IEEE Access* **2021**, *9*, 99595–99605. [CrossRef]
6. Nagarajan, J.; Mansourian, P.; Shahid, M.A.; Jaekel, A.; Saini, I.; Zhang, N.; Kneppers, M. Machine Learning based intrusion detection systems for connected autonomous vehicles: A survey. *Peer-to-Peer Netw. Appl.* **2023**, *16*, 2153–2185. [CrossRef]
7. Aloraini, F.; Javed, A.; Rana, O. Adversarial Attacks on Intrusion Detection Systems in In-Vehicle Networks of Connected and Autonomous Vehicles. *Sensors* **2024**, *24*, 3848. [CrossRef]
8. Neto, E.C.P.; Taslimasa, H.; Dadkhah, S.; Iqbal, S.; Xiong, P.; Rahman, T.; Ghorbani, A.A. CICIoV2024: Advancing realistic IDS approaches against DoS and spoofing attack in IoV CAN bus. *Internet Things* **2024**, *26*, 101209. [CrossRef]
9. Cheng, P.; Xu, K.; Li, S.; Han, M. TCAN-IDS: Intrusion Detection System for Internet of Vehicle Using Temporal Convolutional Attention Network. *Symmetry* **2022**, *14*, 310. [CrossRef]
10. El-Gayar, M.M.; Alrslani, F.A.; El-Sappagh, S. Smart Collaborative Intrusion Detection System for Securing Vehicular Networks Using Ensemble Machine Learning Model. *Information* **2024**, *15*, 583. [CrossRef]
11. Yang, L.; Moubayed, A.; Shami, A. MTH-IDS: A Multitiered Hybrid Intrusion Detection System for Internet of Vehicles. *IEEE Internet Things J.* **2022**, *9*, 616–632. [CrossRef]
12. Wang, S.; Wang, Y.; Zheng, B.; Cheng, J.; Su, Y.; Dai, Y. Intrusion Detection System for Vehicular Networks Based on MobileNetV3. *IEEE Access* **2024**, *12*, 106285–106302. [CrossRef]
13. Almehdhar, M.; Albaseer, A.; Khan, M.A.; Abdallah, M.; Menouar, H.; Al-Kuwari, S.; Al-Fuqaha, A. Deep Learning in the Fast Lane: A Survey on Advanced Intrusion Detection Systems for Intelligent Vehicle Networks. *IEEE Open J. Veh. Technol.* **2024**, *5*, 869–906. [CrossRef]
14. Korba, A.A.; Sebaa, S.; Mabrouki, M.; Ghamri-Doudane, Y.; Benatchba, K. A Life-long Learning Intrusion Detection System for 6G-Enabled IoV. In Proceedings of the 20th International Wireless Communications and Mobile Computing Conference, IWCMC 2024, Ayia Napa, Cyprus, 27–31 May 2024; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2024; pp. 1773–1778. [CrossRef]
15. Qin, J.; Xun, Y.; Liu, J. CVMIDS: Cloud-Vehicle Collaborative Intrusion Detection System for Internet of Vehicles. *IEEE Internet Things J.* **2024**, *11*, 321–332. [CrossRef]
16. Yang, L.; Shami, A. A Transfer Learning and Optimized CNN Based Intrusion Detection System for Internet of Vehicles. In Proceedings of the IEEE International Conference on Communications, Seoul, Republic of Korea, 16–20 May 2022; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2022; Volume 2022, pp. 2774–2779. [CrossRef]
17. Gul, M.F.; Bakir, H. Improving Attack Detection in IoV Systems using GA-based Hyperparameter Optimization. In Proceedings of the 8th International Artificial Intelligence and Data Processing Symposium, IDAP 2024, Malatya, Türkiye, 21–22 September 2024; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2024. [CrossRef]
18. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, 4–8 August 2019; Teredesai, A., Kumar, V., Li, Y., Rosales, R., Terzi, E., Karypis, G., Eds.; ACM: New York, NY, USA, 2019; pp. 2623–2631. [CrossRef]
19. Jin, F.; Chen, M.; Zhang, W.; Yuan, Y.; Wang, S. Intrusion detection on internet of vehicles via combining log-ratio oversampling, outlier detection and metric learning. *Inf. Sci.* **2021**, *579*, 814–831. [CrossRef]

20. Singh, A.P.; Chaurasia, B.K.; Tripathi, A. Stacking Enabled Ensemble Learning Based Intrusion Detection Scheme (SELIDS) for IoV. *SN Comput. Sci.* **2024**, *5*, 1000. [CrossRef]
21. Du, L.; Gu, Z.; Wang, Y.; Gao, C. Open World Intrusion Detection: An Open Set Recognition Method for CAN Bus in Intelligent Connected Vehicles. *IEEE Netw.* **2024**, *38*, 76–82. [CrossRef]
22. Ullah, S.; Khan, M.A.; Ahmad, J.; Jamal, S.S.; Huma, Z.E.; Hassan, M.T.; Pitropakis, N.; Arshad; Buchanan, W.J. HDL-IDS: A Hybrid Deep Learning Architecture for Intrusion Detection in the Internet of Vehicles. *Sensors* **2022**, *22*, 1340. [CrossRef]
23. Wang, Y.; Qin, G.; Zou, M.; Liang, Y.; Wang, G.; Wang, K.; Feng, Y.; Zhang, Z. A lightweight intrusion detection system for internet of vehicles based on transfer learning and MobileNetV2 with hyper-parameter optimization. *Multimed. Tools Appl.* **2024**, *83*, 22347–22369. [CrossRef]
24. Fu, M.; Wang, P.; Liu, M.; Zhang, Z.; Zhou, X. IoV-BERT-IDS: Hybrid Network Intrusion Detection System in IoV Using Large Language Models. *IEEE Trans. Veh. Technol.* **2024**, *74*, 1909–1921. [CrossRef]
25. Wang, W.; Sun, D. The improved AdaBoost algorithms for imbalanced data classification. *Inf. Sci.* **2021**, *563*, 358–374. [CrossRef]
26. Rahman, H.A.A.; Wah, Y.B.; He, H.; Bulgiba, A. Comparisons of ADABOOST, KNN, SVM and Logistic Regression in Classification of Imbalanced Dataset. In Proceedings of the Soft Computing in Data Science, Putrajaya, Malaysia, 2–3 September 2015; pp. 54–64.