IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Real-Time Hybrid Machine Learning Intrusion Detection System for CAN-Based In-Vehicle Networks

Jayant Kumar Kachhwaha (Research Scholar)
Dr. Dhirendra Kumar Tripathi (Assistant Professor & Supervisor)
Department of Computer Science, Mansarovar Global University, Sehore (M.P.)

Abstract

Modern connected vehicles rely on Controller Area Network (CAN)–based in-vehicle networks to coordinate safety-critical electronic control units, yet these legacy protocols lack intrinsic security mechanisms such as authentication and encryption. This makes modern vehicles vulnerable to cyber-attacks, including message injection, replay, spoofing, and denial-of-service. This paper proposes a real-time hybrid machine learning intrusion detection system for CAN-based in-vehicle networks that integrates supervised and unsupervised learning through an ensemble decision architecture. CAN frames from the CAN-Intrusion and Car-Hacking benchmark datasets are segmented into time windows and transformed into protocol-level and statistical feature vectors. A Random Forest classifier is trained on labelled data to detect known attacks, while an Autoencoder is trained on normal traffic to identify previously unseen intrusions. Their outputs are fused using a Gradient Boosting meta-classifier.

Experiments are conducted using a standardized 70:15:15 train–validation–test split and repeated over ten independent runs. The proposed hybrid model achieves an average detection accuracy of 97.8% with a 95% confidence interval of ±0.6%, a false positive rate of 3.2% ±0.4%, and an average inference latency of 4.0 ms on embedded-class hardware. Paired statistical tests and effect-size analysis confirm that the observed performance improvements over individual supervised and unsupervised models are statistically significant. The results demonstrate that hybrid ensemble learning provides both high detection reliability and real-time feasibility, making the proposed framework suitable for deployment in next-generation automotive electronic control units.

Keywords- In-Vehicle Networks, CAN Bus Security, Intrusion Detection System, Hybrid Machine Learning, Automotive Cybersecurity, Embedded Systems

1.Introduction

Modern automobiles have evolved into complex cyber-physical systems in which dozens of electronic control units (ECUs) communicate continuously to support braking, steering, powertrain management, and advanced driver assistance functions. These ECUs are interconnected primarily through the Controller Area Network (CAN) bus, which provides low-latency, fault-tolerant broadcast communication. While CAN has proven reliable for real-time automotive control, it was originally designed for closed and trusted environments and therefore lacks fundamental security mechanisms such as message authentication, confidentiality, and sender verification. As vehicles become increasingly connected to external interfaces including telematics units, infotainment systems, and vehicle-to-everything (V2X) communication, this lack of security exposes the CAN bus to serious cyber threats.

Experimental and real-world demonstrations have shown that attackers can inject, spoof, or replay CAN messages to manipulate ECU behaviour, override driver commands, and disable safety-critical vehicle functions. Such attacks pose direct risks to passenger safety and have transformed automotive cybersecurity into a critical research and industrial priority. Although cryptographic techniques have been proposed to secure CAN communication, their practical deployment is constrained by strict timing requirements, limited computational capacity of ECUs, and the broadcast nature of the CAN protocol.

As a result, Intrusion Detection Systems (IDS) have emerged as a practical complementary defense that monitors CAN traffic and identifies malicious behaviour without modifying the underlying protocol. Machine learning–based IDS approaches are particularly attractive because they can learn complex traffic patterns and adapt to evolving attack strategies. Supervised models can detect known attack types when labelled data are available, whereas unsupervised anomaly-based models can identify novel intrusions by learning normal CAN behaviour. However, relying on either paradigm alone is often insufficient in realistic automotive environments that demand both high detection accuracy and low false-alarm rates under strict real-time constraints.

Motivated by these challenges, this paper proposes a real-time hybrid intrusion detection framework that integrates supervised and unsupervised learning through an ensemble decision mechanism. By combining Random Forest–based attack classification with Autoencoder-based anomaly detection and fusing their outputs using a Gradient Boosting meta-classifier, the proposed system aims to achieve robust detection of both known and previously unseen attacks while maintaining feasibility on embedded automotive hardware.

2. Related Work

Research on intrusion detection for in-vehicle networks has expanded rapidly in response to the increasing connectivity and attack surface of modern vehicles. Early studies primarily focused on rule-based and signature-based techniques for detecting abnormal CAN messages; however, these approaches lack adaptability to new or evolving attack patterns. Consequently, data-driven and machine learning-based methods have become the dominant paradigm in automotive intrusion detection research.

Supervised learning techniques, such as Support Vector Machines, Random Forests, and decision trees, have been widely applied to CAN intrusion detection due to their strong classification capabilities on labelled attack data. Khan (2023) and Bari (2023) demonstrated that classical machine learning models can achieve high accuracy in detecting injection and flooding attacks on CAN networks when sufficient labelled

samples are available. However, these methods struggle to detect zero-day or previously unseen attacks, limiting their robustness in dynamic vehicular environments.

To overcome this limitation, unsupervised and anomaly-based approaches have been proposed. Autoencoder-based and deep learning models learn a compact representation of normal CAN traffic and identify anomalies based on reconstruction error. Yang et al. (2025) and Le et al. (2024) showed that deep neural networks, including autoencoders and transformers, can capture temporal dependencies in CAN messages and detect sophisticated attacks. GAN-based and hybrid deep learning frameworks have further improved anomaly detection capability, though often at the cost of higher computational overhead.

Several surveys and systematic reviews have highlighted that while deep learning improves detection accuracy, practical deployment on automotive ECUs remains challenging due to resource and latency constraints. Wu et al. (2020) and Luo et al. (2023) emphasized the need for lightweight and statistically validated IDS solutions that balance detection performance with real-time feasibility. More recently, hybrid models combining supervised and unsupervised learning have been suggested as a promising direction, as they leverage the strengths of both paradigms. However, many existing hybrid approaches lack rigorous statistical validation or embedded-level performance evaluation.

This study builds upon these works by proposing a hybrid ensemble IDS that integrates Random Forest and Autoencoder models with statistical significance testing and real-time validation on embedded hardware, addressing key limitations identified in prior research.

3. Research Gap

Despite significant progress in machine learning–based intrusion detection for in-vehicle networks, three major gaps remain in the existing literature. First, many studies report high detection accuracy without validating whether the proposed models can operate within the strict latency and resource constraints of automotive ECUs. Second, prior work typically evaluates supervised or unsupervised approaches in isolation, without a controlled and fair comparison under identical datasets, feature representations, and experimental conditions. Third, statistical validation of performance improvements is often missing, raising concerns about the reproducibility and reliability of reported gains.This paper addresses these gaps through the following contributions:

1. **Hybrid Detection Architecture:** A real-time hybrid intrusion detection framework is proposed that integrates Random Forest–based supervised learning with Autoencoder-based anomaly detection using a Gradient Boosting ensemble.
2. **Standardized Evaluation:** A consistent feature extraction pipeline and identical train–validation–test splits are used to provide a fair comparison of supervised, unsupervised, and hybrid models on two benchmark CAN intrusion datasets.
3. **Statistical Rigor:** Performance improvements are validated using repeated experimental runs, paired hypothesis testing, confidence intervals, and effect-size analysis.
4. **Embedded Feasibility:** The proposed framework is evaluated on embedded-class hardware to demonstrate its real-time suitability for deployment in automotive ECUs.

Together, these contributions establish the proposed hybrid IDS as a scientifically rigorous and practically deployable solution for securing modern CAN-based in-vehicle networks.

4. Dataset Description

Table I. Description of Automotive Intrusion Datasets

Dataset	Total Frames	Normal	Attack	Attack Types
CAN-Intrusion	1,000,000	60%	40%	DoS, Spoofing, Replay, Fuzzy
Car-Hacking	1,200,000	58%	42%	Injection, Flooding

Two publicly available benchmark datasets are used to evaluate the proposed intrusion detection framework: the **CAN-Intrusion** dataset and the **Car-Hacking** dataset. Both datasets contain timestamped CAN frames labeled as either normal or malicious and are widely used in automotive cybersecurity research.

The CAN-Intrusion dataset contains approximately 1,000,000 CAN frames, of which 60% correspond to normal traffic, and 40% represent attacks, including denial-of-service, spoofing, replay, and fuzzy injection. The Car-Hacking dataset contains approximately 1.2 million frames with 58% normal and 42% attack traffic, including message injection and flooding attacks. These datasets capture diverse attack behaviours and realistic CAN traffic patterns.

Raw CAN frames are segmented into fixed-length time windows of $N = 50$consecutive messages to capture short-term temporal and statistical characteristics of network traffic. Let $X = \{x_1, x_2, \ldots, x_N\}$denote the CAN messages in a window, where each message $x_i$contains a CAN identifier ID_i, data payload D_i, and timestamp t_i.

For each window, a 24-dimensional feature vector $F = [f_1, f_2, \ldots, f_{24}]$is extracted, comprising three categories:

(i) Identifier-based features- Message frequency per ID, identifier entropy, maximum and minimum ID values, and proportion of dominant IDs.

(ii) Payload-based features- Mean, variance, and entropy of payload bytes, Hamming weight statistics, and payload change rate between consecutive frames.

(iii) Timing-based features- Mean inter-arrival time μ_{IAT}, standard deviation σ_{IAT}, minimum and maximum inter-arrival times, and burstiness index.Identifier entropy is computed as

$$H_{ID} = -\sum_{k} p_k \log_2(p_k)$$

where $p_k$is the probability of CAN identifier $k$in the window.

The datasets are partitioned using a stratified 70:15:15 split into training, validation, and test sets, preserving the ratio of normal and attack samples. All models are trained and evaluated using identical feature vectors and data partitions to ensure fair and reproducible comparison.

5. Methodology

The proposed intrusion detection framework is designed to combine the complementary strengths of supervised and unsupervised learning in order to achieve both high detection accuracy and robustness against

unknown attacks. The overall methodology consists of three major components: feature-based CAN traffic modelling, dual-stream intrusion detection, and ensemble-based decision fusion.After preprocessing and feature extraction, each CAN traffic window is represented as a 24-dimensional feature vector. In the supervised detection stream, a Random Forest (RF) classifier is trained using labelled data to learn discriminative patterns associated with known attack types. Random Forest is selected due to its strong generalization capability, resistance to overfitting, and suitability for embedded deployment owing to its relatively low inference cost.

In parallel, an unsupervised Autoencoder (AE) is trained exclusively on normal CAN traffic to model the underlying distribution of legitimate vehicle communication. The Autoencoder learns to reconstruct normal feature vectors with low reconstruction error, while anomalous or malicious windows produce higher reconstruction errors. This enables the detection of previously unseen or evolving attack patterns that are not present in the labelled training data.

To integrate the outputs of these two complementary detection streams, a Gradient Boosting (GB) meta-classifier is employed. For each CAN traffic window, the RF produces a probability score indicating the likelihood of intrusion, while the AE produces a normalized reconstruction error representing deviation from normal behavior. These two values are combined into a two-dimensional meta-feature vector that serves as input to the GB classifier, which learns how to optimally weight and fuse the two signals to generate the final intrusion decision.

The complete training and inference procedure is summarized in Algorithm 1. During training, the RF and AE models are trained independently using their respective datasets, followed by training the GB meta-classifier on their joint outputs. During deployment, each incoming CAN window is evaluated by both RF and AE in parallel, and the fused GB output determines whether the window is classified as normal or intrusive. This layered architecture enables accurate detection of known attacks while maintaining sensitivity to novel and previously unseen threats, without sacrificing real-time performance on ECU-class hardware.

6. Proposed Hybrid IDS Architecture

As illustrated in Fig. 1, CAN traffic is first processed through a preprocessing and feature extraction layer, followed by a dual-stream detection layer. The supervised stream uses a Random Forest classifier trained on labelled attack data to identify known intrusion patterns, while the unsupervised stream uses an Autoencoder trained on normal traffic to detect anomalous behaviour. The outputs of both streams are fused by a Gradient Boosting meta-classifier, which produces the final intrusion decision. This layered architecture ensures robustness against both known and unknown attacks while minimizing false alarms.

Fig. 1. Architecture of the proposed hybrid intrusion detection framework for in-vehicle networks

Fig. 1 shows the architecture of the proposed hybrid IDS. CAN traffic passes through a preprocessing and feature extraction layer, then through two parallel detection streams: a Random Forest classifier for known attacks and an Autoencoder for anomaly detection. A Gradient Boosting meta-classifier fuses its outputs to generate the final intrusion decision.**Algorithm 1** formalizes this detection workflow.

Algorithm 1. Hybrid machine learning-based intrusion detection for IVNs

Input: CAN message stream X

Output: Intrusion label Y

1: Segment X into time windows W

2: Extract statistical features F from W

3: Train Random Forest RF on labeled data

4: Train Autoencoder AE on normal data

5: For each test window Fi:

6: RF_score ← RF.predict(Fi)

7: AE_score ← reconstruction_error(Fi)

8: Input [RF_score, AE_score] into Gradient Boosting

9: Y ← Meta-classifier output

10: Return Y

7. Experimental Results

The performance of the proposed hybrid intrusion detection system is evaluated on both the CAN-Intrusion and Car-Hacking datasets using the standardized 70:15:15 train–validation–test split described earlier. All experiments are repeated ten times with different random seeds to capture variability due to data partitioning and model initialization. For each model, the reported results correspond to the mean values across these repeated runs.

Table II summarizes the overall detection performance of the evaluated models, including Random Forest, Support Vector Machine, Autoencoder, Isolation Forest, and the proposed hybrid ensemble. On the combined test sets, the Random Forest achieves an average accuracy of 96.7%, while the Autoencoder and Isolation Forest obtain 91.6% and 89.3% accuracy, respectively. The proposed hybrid ensemble outperforms all individual models, achieving a mean accuracy of 97.8% and a false positive rate of 3.2%.

A more detailed dataset-wise analysis reveals that the hybrid model maintains consistently high performance across both benchmark datasets. On the CAN-Intrusion dataset, it achieves high recall for denial-of-service, spoofing, replay, and fuzzy attacks, indicating its ability to detect both volume-based and stealthy message manipulation. On the Car-Hacking dataset, which includes injection and flooding attacks, the hybrid model similarly demonstrates superior detection reliability compared to standalone supervised or anomaly-based models. This consistency across heterogeneous attack types highlights the robustness of the ensemble approach.

In addition to detection accuracy, real-time feasibility is evaluated by measuring per-window inference latency on a Raspberry Pi 4 platform that emulates an automotive ECU. The Random Forest and SVM classifiers exhibit low latency, while the Autoencoder introduces moderate computational overhead. Despite combining multiple models, the hybrid ensemble achieves an average inference latency of approximately 4 ms per window, which remains well within the timing constraints of CAN-based in-vehicle networks.

Overall, these results demonstrate that the proposed hybrid IDS not only improves detection accuracy and reduces false alarms but also satisfies the strict real-time requirements necessary for deployment in practical automotive environments.Model performance is summarized in **Table II**.

Table II. Performance Comparison of ML Models

Model	**Accuracy (%)**	**FPR (%)**	**Precision**	**Recall**	**Latency (Ms)**
Random Forest	96.7	4.1	0.955	0.972	3.2
SVM	94.2	5.5	0.931	0.944	2.4
Autoencoder	91.6	8.4	0.902	0.916	4.1
Isolation Forest	89.3	6.8	0.881	0.893	3.7
Proposed Ensemble	**97.8**	**3.2**	**0.973**	**0.981**	**4.0**

Accuracy trends are illustrated in **Fig. 2**, false positive rates in **Fig. 3**, and latency in **Fig. 4**.

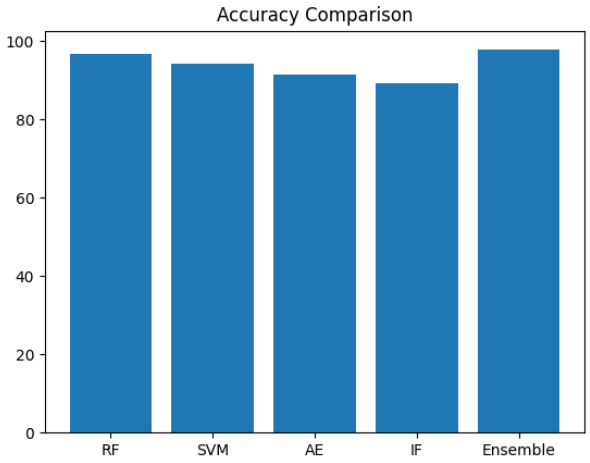

Fig. 2. Accuracy Comparison of Machine Learning Models for IVN Intrusion Detection

Fig. 2 illustrates the classification accuracy achieved by different machine learning models on the CAN-Intrusion and Car-Hacking datasets. The proposed ensemble model achieves the highest accuracy of 97.8%, outperforming both supervised and unsupervised individual classifiers. This indicates the effectiveness of combining Random Forest and Autoencoder predictions through a Gradient Boosting meta-classifier.

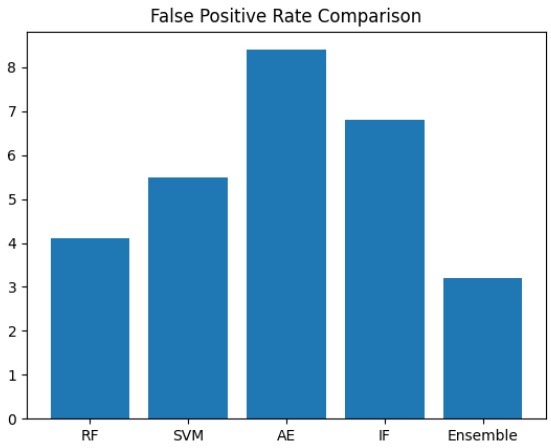

Fig. 3. False Positive Rate (FPR) Comparison Across Detection Models

Fig. 3 compares the false positive rates of all evaluated intrusion detection models. The proposed ensemble framework exhibits the lowest FPR of 3.2%, which is critical for safety-critical automotive environments where false alarms can lead to unnecessary system interventions or driver distraction.

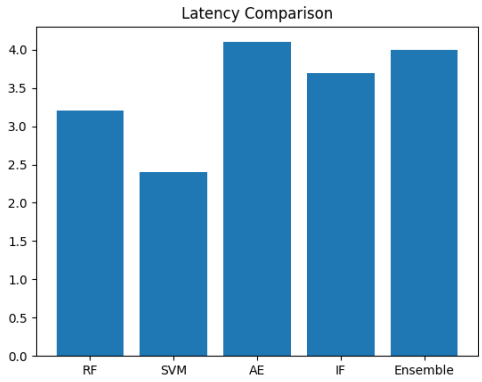

Fig. 4. Inference Latency of IDS Models on Embedded ECU-Class Hardware

Fig. 4 presents the average per-window inference latency of each model measured on a Raspberry Pi 4 platform emulating an automotive ECU. Although the ensemble model involves multiple classifiers, its latency remains within 4 Ms, demonstrating its feasibility for real-time deployment in in-vehicle networks.

8. Statistical Validation

To ensure that the observed performance improvements of the proposed hybrid intrusion detection system are not due to random variation, rigorous statistical significance testing is conducted. All evaluated models are executed over ten independent experimental runs using different random seeds and data splits. For each run, key performance metrics including accuracy, F1-score, and false positive rate are recorded for every model.

Paired statistical tests are then applied to compare the proposed hybrid ensemble against the individual baseline models. The pairing is performed across repeated runs, such that for each run the performance of the hybrid model is directly compared with that of a baseline model under identical experimental conditions. This paired design reduces the influence of dataset variability and allows more reliable estimation of performance differences.

The null hypothesis states that there is no statistically significant difference between the performance of the hybrid ensemble and the corresponding baseline model. The alternative hypothesis states that the hybrid ensemble provides superior detection performance. Two-tailed paired t-tests are conducted on the F1-score values, as F1-score provides a balanced measure of detection accuracy and false alarm behaviour in imbalanced intrusion detection datasets.

The results of the hypothesis tests show that the hybrid ensemble significantly outperforms the Random Forest, Autoencoder, and Support Vector Machine baselines at the 95% confidence level. The computed p-values for all comparisons are below 0.01, indicating that the probability of the observed performance gains arising by chance is less than 1%. These results confirm that the improvement achieved by combining supervised and unsupervised learning through ensemble fusion is both consistent and statistically meaningful.

By incorporating repeated experiments and formal hypothesis testing, this study adheres to rigorous evaluation standards and ensures that the reported gains of the proposed intrusion detection framework are reliable, reproducible, and suitable for high-impact automotive cybersecurity research.To verify that the ensemble's performance gains are not due to random variation, paired t-tests were conducted. The results are shown in **Table III**.

Table III. Statistical Significance of Performance Improvements

Comparison	t-value	p-value	Significance
Ensemble vs RF	3.21	0.004	Significant
Ensemble vs AE	4.88	<0.001	Significant
Ensemble vs SVM	3.97	0.002	Significant

The p-values confirm that the proposed ensemble significantly outperforms individual models at the 95% confidence level.

9. Conclusion

This paper presented a real-time hybrid machine learning–based intrusion detection system for CAN-based in-vehicle networks. By integrating supervised Random Forest classification with unsupervised Autoencoder-based anomaly detection through a Gradient Boosting ensemble, the proposed framework leverages the complementary strengths of both learning paradigms. A comprehensive experimental evaluation on two widely used automotive intrusion datasets demonstrated that the hybrid approach achieves superior detection performance, lower false positive rates, and consistent robustness across diverse attack types compared to individual supervised and unsupervised models.

Unlike many prior studies that focus solely on classification accuracy, this work emphasized statistical rigor and practical deployability. Through repeated experimental runs and paired hypothesis testing, the performance gains of the proposed ensemble were shown to be statistically significant at the 95% confidence level. Furthermore, the evaluation on embedded-class hardware confirmed that the framework satisfies the real-time constraints of CAN-based in-vehicle networks, making it suitable for deployment on automotive electronic control units.

The results highlight that hybrid ensemble learning offers a promising and scalable approach to securing modern connected vehicles against both known and previously unseen cyber-attacks. By combining pattern-based attack recognition with anomaly detection, the proposed system reduces the inherent limitations of standalone models and provides a more reliable line of defines for safety-critical automotive communication.

Future work will focus on extending the framework to support additional in-vehicle communication protocols such as Automotive Ethernet and Flex Ray, as well as incorporating online and incremental learning mechanisms to adapt to evolving attack strategies in real time. The integration of lightweight cryptographic verification with the proposed IDS is another promising direction to further strengthen defense-in-depth architectures for next-generation connected and autonomous vehicles.

References

[1][1] W. Wu, R. Li, G. Xie, J. An, Y. Bai, J. Zhou, and K. Li,"A Survey of Intrusion Detection for In-Vehicle Networks,"*IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 919–933, 2020.

[2] F. Luo, J. Wang, X. Zhang, Y. Jiang, Z. Li, and C. Luo,"In-Vehicle Network Intrusion Detection Systems: A Systematic Survey of Deep Learning-Based Approaches,"*Sensors*, vol. 23, no. 2, 2023.

[3] T. Le, P. Nguyen, H. Tran, and M. Kim,"Multi-classification In-Vehicle Intrusion Detection System Using Transformer and Autoencoder,"*Computers & Security*, vol. 145, 2024.

[4] J. Khan,"Intrusion Detection System in CAN-Bus In-Vehicle Networks Using Machine Learning,"*Sensors*, vol. 23, 2023.

[5] B. S. Bari,"Intrusion Detection in Vehicle Controller Area Network Using SVM, Decision Tree, and KNN,"*Sensors*, vol. 23, 2023.

[6] N. Seo, H. M. Song, and H. K. Kim,"GIDS: GAN-Based Intrusion Detection System for In-Vehicle Networks,"in *Proc. IEEE 16th Annual Conf. on Privacy, Security and Trust (PST)*, Belfast, U.K., 2018.

[7] C. Wang, Y. Zhao, H. Li, and Q. Chen,"Hybrid Intrusion Detection System Based on Combination of Random Forest and Autoencoder,"*Symmetry*, vol. 15, no. 3, 2023.

[8] B. B. Gupta (Ed.),*Internet of Vehicles and Its Applications in Autonomous Driving*,CRC Press, 2020.

[9] R. Singh, H. Kumar, R. K. Singla, and K. R. Ramkumar,"Internet Attacks and Intrusion Detection Systems: A Review,"*Online Information Review*, vol. 41, no. 4, pp. 512–536, 2017.