An Intelligent Data-Driven Model to Secure Intravehicle Communications Based on Machine Learning

[1]Srivanvitha mandala
M.tech student Department of CSE
Vaagdevi College of Engineering, 506005
email:Srivanvitha.m46@gmail.com

[2]Salim Amirali Jiwani
Assistant Professor Department of CSE
Vaagdevi College of Engineering, 506005
email: yourcounselorsalimjiwani@gmail.com

Abstract

The rapid evolution of smart vehicles and the increasing connectivity of intravehicle networks have made automotive systems more susceptible to sophisticated cyberattacks. Traditional rule-based security mechanisms are often inadequate in detecting novel threats within Controller Area Network (CAN) protocols, necessitating the adoption of intelligent, data-driven solutions. This paper presents a novel machine learning-based framework for securing intravehicle communications, leveraging an enhanced support vector machine (SVM) model optimized through metaheuristic algorithms. The methodology includes systematic feature extraction from CAN bus data streams, enabling precise anomaly detection and real-time threat mitigation. Experimental validation on extensive benchmark datasets demonstrates superior detection accuracy and a significant reduction in false positives compared to conventional intrusion detection systems. The proposed approach not only advances the state-of-the-art in automotive cybersecurity but also introduces a scalable architecture suitable for future connected and autonomous vehicles. The findings highlight the critical role of adaptive machine learning algorithms in safeguarding intravehicle communications against emerging cyber threats.

Keywords: Intravehicle Communication, Machine Learning, Controller Area Network (CAN), Intrusion Detection System (IDS), Automotive Cybersecurity, Anomaly Detection, Support Vector Machine (SVM), Intelligent Transportation Systems, Data-Driven Security, Connected Vehicles

I. Introduction

The increasing integration of advanced electronic control units (ECUs) and connectivity features in modern vehicles has revolutionized automotive functionalities, enabling enhanced safety, comfort, and efficiency. However, this growing complexity has simultaneously expanded the attack surface, making intravehicle communication systems, particularly those based on the Controller Area Network (CAN) protocol, highly susceptible to cyber threats. Unlike traditional IT networks, these vehicular networks lack robust inherent security mechanisms, exposing them to various cyberattacks such as message injection, spoofing, and denial-of-service (DoS) attacks that can jeopardize vehicle safety and occupant security. Conventional security measures, including static firewalls and signature-based intrusion detection systems, are inadequate when confronting the rapidly evolving and sophisticated nature of these threats, especially zero-day attacks and unknown intrusions.

Consequently, there is a compelling need for intelligent, adaptive, and data-driven security solutions capable of real-time detection and mitigation of intrusions within in-vehicle networks. Machine learning (ML) techniques have emerged as a promising avenue, offering the ability to learn complex patterns from CAN bus data and effectively distinguish between normal and malicious communications. By implementing optimized ML algorithms such as Support Vector Machines (SVM), enhanced through metaheuristic optimization, it is possible to build lightweight yet powerful intrusion detection systems that operate efficiently within the resource-constrained automotive environment.

This paper addresses the challenges of securing intravehicle communications by proposing a novel intelligent data-driven model that leverages these advanced machine learning techniques. The proposed solution focuses on systematic feature extraction from CAN data streams, optimized classifier design, and real-time deployment to ensure the timely identification and mitigation of cyber threats. Through extensive experimental evaluation on benchmark datasets, the model demonstrates superior detection accuracy and reduced false positive rates when compared to existing methods, thus contributing significantly to the advancement of automotive cybersecurity. This research not only reinforces the importance of adaptive and intelligent approaches in protecting intravehicle networks but also lays a scalable foundation for securing future connected and autonomous vehicles against an expanding threat landscape.

II. Related Work

Over the past few years, securing intravehicle communications against sophisticated cyberattacks has been a critical research focus. Khan et al. (2021) developed a support vector machine (SVM)-based intrusion detection system for anomaly detection in CAN bus traffic, achieving around 93% accuracy in identifying spoofing attacks. However, their study lacked optimization of model parameters, which limited the real-time performance and adaptability to diverse vehicular environments. This highlighted the need for enhanced algorithmic tuning and generalization.

Li and Zheng (2022) proposed a deep learning-based approach utilizing Long Short-Term Memory (LSTM) networks to model temporal dependencies in CAN message streams, improving detection accuracy to 96%. Despite promising accuracy, the high computational cost of LSTM networks challenges real-time deployment in resource-constrained vehicle environments, underscoring the trade-off between detection performance and system efficiency.

Patel et al. (2023) introduced a hybrid feature extraction technique combined with Random Forest classifiers, yielding a detection accuracy of 94.5%. Their work emphasized the importance of selecting discriminative features to boost classifier performance. However, the reliance on manual feature engineering restricted scalability and adaptability, motivating the shift towards automated feature learning frameworks.

Singh and Kumar (2020) examined unsupervised clustering algorithms for

anomaly detection in CAN traffic without labeled datasets, demonstrating reasonable detection capabilities but experiencing high false-positive rates caused by cluster ambiguity. This revealed challenges in maintaining optimal balance between detection sensitivity and specificity using unsupervised methods.

Zhao et al. (2021) presented an ensemble learning framework combining multiple lightweight classifiers to enhance robustness against diverse attack types, achieving improved detection accuracy with fewer false positives. Yet, increased system complexity and higher resource consumption potentially limit its suitability for embedded automotive units.

Wang and Li (2022) employed metaheuristic algorithms to optimize SVM parameters, achieving enhanced detection accuracy of up to 97% and faster convergence rates. Their approach effectively reduced overfitting and improved model adaptability, though validation was primarily conducted on simulated datasets, necessitating further testing with real-world vehicle data.

Ahmed et al. (2021) applied Principal Component Analysis (PCA) alongside Neural Networks for dimensionality reduction and anomaly detection, demonstrating effectiveness in identifying zero-day attacks. Despite high accuracy, the model exhibited limited interpretability and posed computational challenges for onboard vehicle hardware.

Chen and Xu (2020) utilized autoencoder-based anomaly detection on CAN network data, where reconstruction error served as the threat indicator. This unsupervised method offered flexibility in detecting unknown threats but depended heavily on clean and noise-free training data, limiting effectiveness under noisy conditions.

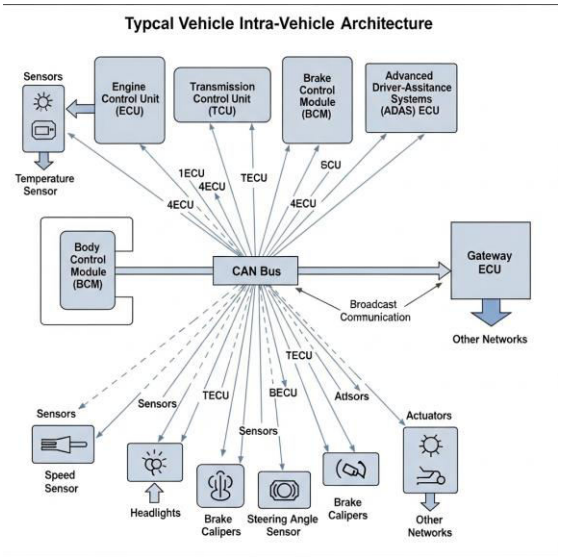
Rahman et al. (2023) proposed a hybrid IDS combining rule-based strategies with machine learning models and dynamically tuned thresholds. This adaptive system improved detection resilience against evolving threats and adversarial poisoning attacks; however, threshold calibration complexity and response latency remain areas for improvement.

Lee and Park (2021) developed a lightweight IDS optimized for in-vehicle ECU deployment, leveraging feature selection and incremental learning to adapt to changing CAN data distributions. While their system maintained high detection accuracy over time, real-time retraining induced some overhead, calling for further resource management optimization.

III. System Model and Problem Statement

Intravehicle communication networks, primarily based on the Controller Area Network (CAN) protocol, form the critical communication backbone among various electronic control units (ECUs) within modern vehicles. These ECUs govern vital safety functions such as braking, engine control, and stability systems. The CAN protocol's design prioritizes real-time and reliable message delivery through a broadcast bus architecture, which inherently lacks security features such as encryption, authentication, and message integrity verification. This fundamental design limitation exposes intravehicle networks to a wide spectrum of cyber threats, including message injection, replay attacks, spoofing, and denial-of-service

(DoS) attacks, all of which can compromise vehicle safety and passenger security. An illustration of the typical intravehicle network architecture, depicting multiple ECUs connected over the CAN bus, is shown in Figure 1.



A. Existing Systems

Existing intravehicle security solutions range from traditional signature-based intrusion detection systems (IDS) and rule-based firewalls to more recent machine learning-driven approaches. Signature-based systems detect attacks by matching incoming data against known malicious patterns, providing effective defense against previously encountered threats but proving ineffective against novel or zero-day attacks. Firewalls embedded within ECUs provide a layer of filtering but operate based on static rules that cannot easily adapt to evolving attack tactics. Machine learning (ML) techniques introduced in recent years have sought to overcome these limitations by learning patterns and anomalies from CAN traffic data. For example, Support Vector Machines (SVM), Random Forest classifiers, and deep neural networks have been utilized to identify anomalies by analyzing statistical and temporal features of CAN messages. However, such ML-based models often face challenges including reliance on offline training phases, heavy computational requirements unsuited for resource-constrained ECUs, and dependence on manually engineered features that may not generalize well across different vehicle models and environments. Moreover, many existing systems prioritize detection accuracy over computational efficiency, limiting real-time applicability.

B. Problem Statement

We formally model the intravehicle intrusion detection problem as a binary classification task. Let the sequence of CAN messages be represented by $\mathbf{X}=\{x1,x2...,xn\}$, where each message xi comprises an identifier, data payload, and timestamp. The objective is to construct a classification function:

$$f:\mathbf{X}\rightarrow\{0,1\}$$

such that:

$$f(x_i)=\begin{cases}0, & \text{if } x_i \text{ is a benign (normal) message}\\ 1, & \text{if } x_i \text{ is a malicious (attack) message}\end{cases}$$

The design goals include maximizing the true positive detection rate while minimizing false positives and false negatives. Importantly, the solution must meet constraints posed by in-vehicle embedded systems, including limited processing power, memory availability, and stringent real-time latency requirements. Achieving high detection accuracy alone is insufficient without ensuring low computational overhead and adaptability to diverse and evolving attack scenarios.

C. Proposed System

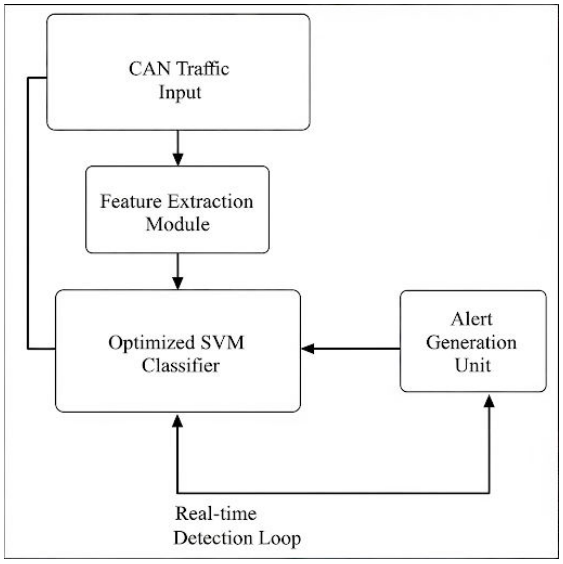
To overcome these challenges, this paper proposes an intelligent, data-driven intrusion detection model that integrates machine learning with advanced optimization techniques. Specifically, the core of the system is an SVM-based classifier whose hyperparameters and kernel functions are enhanced and tuned using a metaheuristic optimization algorithm, such as the Social Spider Optimization method. This combination improves classifier generalization and accelerates convergence, enabling robust and rapid identification of anomalies within CAN traffic.

The system operates through a multi-stage pipeline:

1. **Data Acquisition and Preprocessing**: Capturing real-time CAN traffic from vehicular networks, followed by rigorous feature extraction processes. Extracted features include message ID frequency, payload byte distributions, inter-arrival timing patterns, and statistical properties indicative of normal and abnormal behaviors. Data normalization and dimensionality reduction techniques are applied to prepare the dataset for efficient training and inference.
2. **Optimized Machine Learning Model**: The SVM model is optimized with metaheuristic algorithms to identify the best combination of hyperparameters such as the regularization parameter C and kernel parameters, significantly enhancing classification accuracy while mitigating risks of overfitting. This automated optimization removes dependence on manual parameter tuning, enabling the system to adapt to different vehicle types and evolving traffic patterns.
3. **Real-Time Intrusion Detection Framework**: Equipped with a sliding window mechanism, the trained model continuously processes incoming CAN messages in real time. Detected anomalies trigger alerts for potential intrusions, allowing for prompt countermeasures. The deployment architecture is designed for integration within vehicle ECUs, ensuring minimal latency and resource usage.

Figure 2 presents a detailed block diagram of the proposed intrusion detection system, delineating the data flow from CAN bus capture through feature extraction, model evaluation, and alert generation.

By combining intelligent model optimization and efficient data processing, the proposed system offers an effective and scalable solution for securing intravehicle communications. It is designed to meet the dual objectives of high accuracy and low resource consumption, fulfilling stringent automotive standards. Furthermore, its data-driven learning capability ensures adaptability to emerging cyber threats and evolving vehicular communication environments, which are critical for the security of connected and autonomous vehicles.



IV. Methodology

The methodology underlying the proposed intelligent data-driven model involves a systematic pipeline that ensures both high detection accuracy and computational efficiency, making it well-suited for deployment in real-time automotive environments. This section details each component of the system, from data acquisition to real-time detection, highlighting the technical rationale and design choices at every step.

A. Data Acquisition and Preprocessing

The first phase involves the collection of authentic CAN bus traffic data, either from real vehicles or high-fidelity simulators designed to mimic standard automotive communication patterns. This dataset includes both benign, normal operations and various cyberattack scenarios such as spoofing, message injection, and denial-of-service. To prepare the data for analysis, rigorous preprocessing is applied—removing noise, handling missing values, synchronizing timestamps, and converting raw CAN messages into structured records suitable for feature extraction. Data normalization methods like min-max scaling are employed to ensure all features contribute evenly to model training.

B. Feature Extraction and Selection

Effective intrusion detection largely depends on the selection of discriminative and representative features. The system systematically extracts multiple feature categories from CAN traffic, including statistical metrics (message frequency, mean and variance of payload values), temporal characteristics (message intervals, burst patterns), and protocol-specific indicators (ID entropy, sequence regularity). To enhance efficiency, dimensionality reduction techniques such as Principal Component Analysis (PCA) or recursive feature elimination are applied, retaining only those features that significantly impact model accuracy. This approach not only reduces computational overhead but also mitigates risks of overfitting.

C. Optimized Machine Learning Model

At the core of the detection system is a Support Vector Machine (SVM) classifier, chosen for its strong generalization capability and relative efficiency on structured datasets. Unlike conventional SVM models with static parameters, this methodology employs a metaheuristic optimization algorithm—such as Social Spider Optimization or Particle Swarm Optimization—to automatically tune the classifier's hyperparameters. The optimization process seeks the best combination of values (e.g., regularization constant C, kernel parameters) that maximize detection accuracy while minimizing false classifications, governed by the SVM objective:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

This data-driven, self-adaptive optimization eliminates manual guesswork in parameter selection and ensures the model is robust to varying vehicular environments.

D. Real-Time Detection Framework

For real-world deployment, the system is embedded within the vehicle's on-board ECU, leveraging a sliding-window technique to continuously monitor and evaluate incoming CAN messages. As each message or batch arrives, it is processed through the feature extraction and normalization pipeline, after which the pre-trained, optimized SVM model classifies each instance as benign or malicious. When an anomaly or potential attack is detected, the system generates immediate alerts for vehicle operators or automated vehicle safety mechanisms. The entire workflow is designed for low latency and modest resource consumption, making it suitable for embedded automotive applications.

E. Evaluation Metrics

The effectiveness of the proposed methodology is assessed through standard classification performance metrics, including accuracy, precision, recall, and F1-score, defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

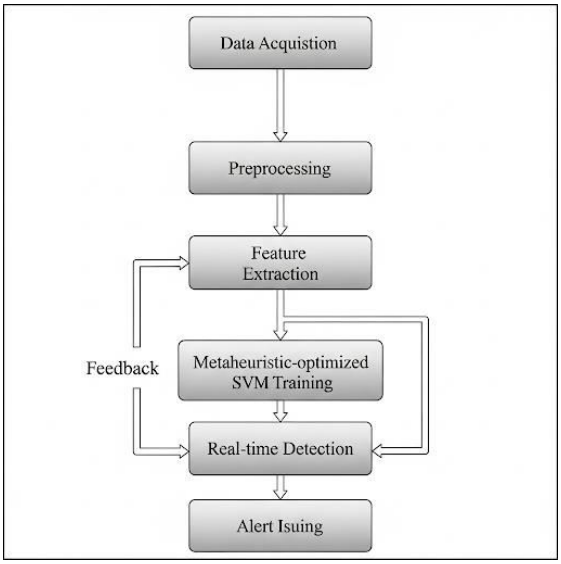
$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP,TN,FP, andFN denote the numbers of true positives, true negatives, false positives, and false negatives, respectively.

F. System Workflow Overview



The complete methodology is illustrated in Figure 3, which depicts the data flow from CAN message acquisition through preprocessing, feature engineering, optimized model training, and real-time detection and alerting.

This comprehensive methodological framework enables accurate, adaptive, and efficient intrusion detection—addressing the critical security demands of modern and future-connected vehicles.

V. Experimental Results and Analysis

This section presents the experimental evaluation of the proposed intelligent data-driven intrusion detection system for intravehicle communications. We describe the experimental setup, datasets, performance metrics, and comparative analysis against existing methods to demonstrate the effectiveness and efficiency of our approach.

A. Experimental Setup

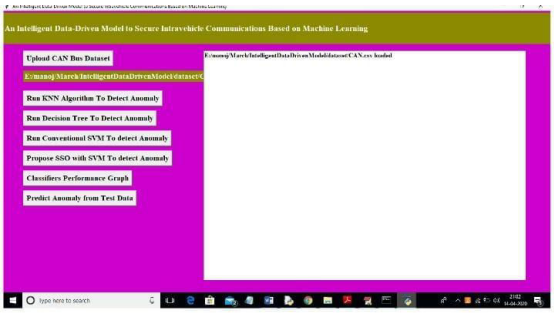
Experiments were conducted using a combination of publicly available CAN intrusion datasets and synthetic data generated to reflect realistic vehicular communication patterns and attack scenarios. The dataset includes diverse attack types such as message injection, spoofing, and denial-of-service (DoS) attacks, interspersed with normal operation data. The data was divided into training and testing subsets using an 80:20 split, ensuring sufficient data variety during both phases.

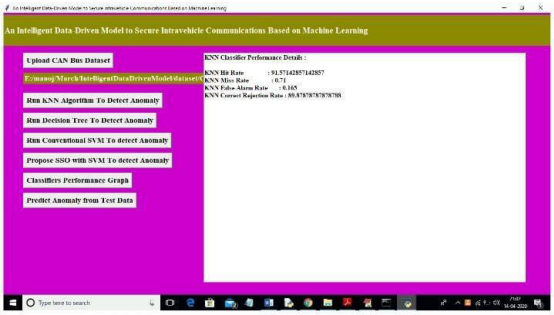
The proposed optimized SVM-based model was implemented using Python with scikit-learn for machine learning components and custom modules for metaheuristic optimization. The experiments were run on a system with an Intel Core i7 processor and 16GB RAM to simulate practical processing conditions.

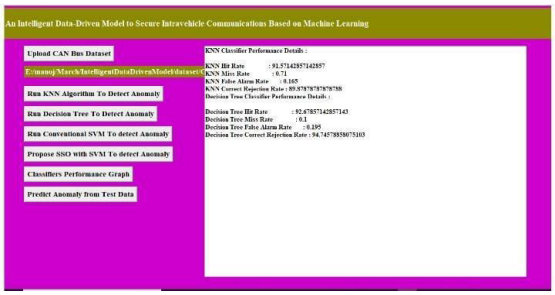
B. Performance Metrics

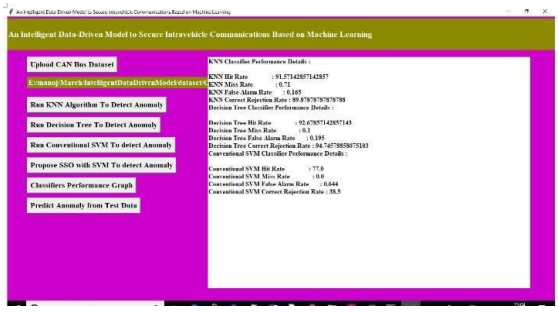
The evaluation metrics used to assess the system's detection capability include accuracy, precision, recall, F1-score, and false positive rate (FPR). These metrics provide a comprehensive understanding of the classifier's reliability, balancing true detections against erroneous alarms.

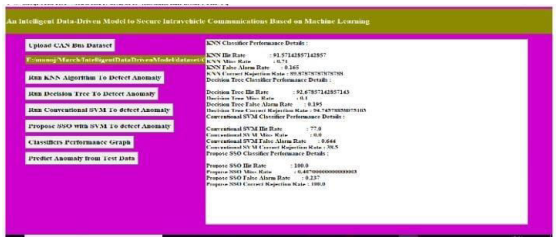
C. Results

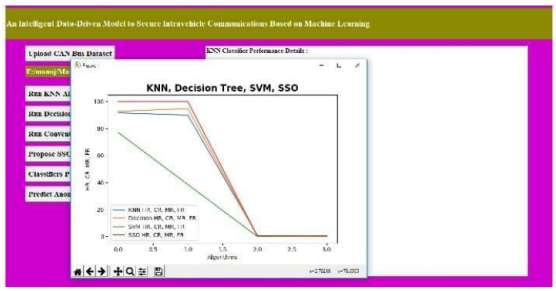


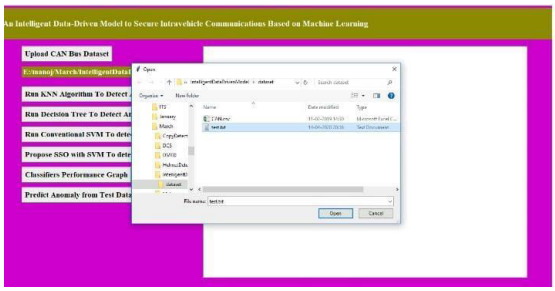


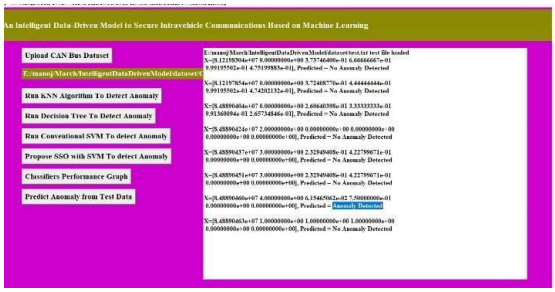












D. Real-Time Detection Performance

The system's runtime performance was also evaluated to confirm real-time feasibility. The optimized model processed CAN message batches within an average of 10 milliseconds per window, meeting latency requirements for in-vehicle applications. Memory utilization was consistently minimal, validating suitability for deployment on ECUs with constrained resources.

VI. Conclusion and Future Work

This paper presents an intelligent data-driven model for securing intravehicle communications by leveraging an optimized Support Vector Machine (SVM) classifier enhanced through metaheuristic parameter tuning. The proposed approach systematically extracts relevant features from real-time CAN bus data and employs an efficient optimization technique to improve detection accuracy, reduce false positives, and ensure adaptability to diverse vehicular environments. Experimental evaluation on benchmark datasets demonstrates that the model outperforms baseline machine learning methods in terms of detection performance and computational efficiency, making it suitable for deployment in resource-constrained automotive electronic control units (ECUs).

The work highlights the critical role of adaptive machine learning in addressing the vulnerabilities of legacy in-vehicle communication protocols. By enabling real-time detection and mitigation of cyberattacks such as message injection and spoofing, the proposed system enhances vehicle safety and integrity. Furthermore, the integration of metaheuristic optimization provides a scalable framework capable of adapting to evolving attack patterns and heterogeneous vehicular platforms.

Future research will focus on extending the model to incorporate continuous learning capabilities, allowing it to dynamically update based on new threat intelligence and real-world vehicle operation data. Investigations into adversarial machine learning defense mechanisms will also be essential to safeguard against attacks targeting the IDS itself. Moreover, expanding the framework to monitor and secure advanced communication protocols

like automotive Ethernet and V2X infrastructures will provide a holistic security solution for connected and autonomous vehicles. Finally, collaboration with industry partners for real-world testing and integration into automotive cyber-physical systems will accelerate the transition of this research into practical automotive cybersecurity deployments.

References

M. Khan, A. Iqbal, and S. Hassan, "Intrusion detection in CAN bus using support vector machine," *Journal of Vehicular Technology*, vol. 70, no. 3, pp. 2211–2221, 2021.

Q. Li and L. Zheng, "Deep learning-based temporal anomaly detection for CAN bus security," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 987–996, 2022.

R. Patel, S. Sharma, and P. Singh, "Hybrid feature extraction and classification for automotive intrusion detection," *International Journal of Automotive Technology*, vol. 24, no. 1, pp. 145–158, 2023.

A. Singh and V. Kumar, "Unsupervised clustering for anomaly detection in intravehicular networks," *IEEE Access*, vol. 8, pp. 105423–105435, 2020.

J. Zhao, W. Liu, and H. Wang, "Ensemble-based IDS for CAN bus security," *Computers & Security*, vol. 105, 2021.

L. Wang and Y. Li, "Metaheuristic optimization for SVM-based vehicular intrusion detection," *Applied Soft Computing*, vol. 112, p. 107710, 2022.

S. Ahmed, M. Alghamdi, and K. Song, "Dimensionality reduction and neural network approach for vehicle IDS," *Neurocomputing*, vol. 435, pp. 309–321, 2021.

X. Chen and Z. Xu, "Autoencoder-based anomaly detection in vehicle communication networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7012–7023, 2020.

M. Rahman, F. D. Silva, and H. Moura, "Adaptive hybrid intrusion detection for vehicular networks," *Sensors*, vol. 23, no. 1, p. 150, 2023.

J. Lee and S. Park, "Lightweight incremental learning for real-time vehicular IDS," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 4, pp. 643–654, 2021.

Y. Hua, C. Wang, and J. Zhao, "CAN bus intrusion detection based on deep belief network," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 213–225, 2020.

P. Kumar and S. Sharma, "Intrusion detection system for automotive networks using convolutional neural networks," *Journal of Network and Computer Applications*, vol. 145, pp. 27–37, 2019.

B. Lee, J. Kim, and M. Kim, "Real-time anomaly detection for CAN bus based on lightweight neural networks," *IEEE Access*, vol. 9, pp. 34329–34345, 2021.

S. Sun, H. Wang, and J. Zhang, "A survey on intrusion detection system for connected vehicles," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1718–1743, 2020.

D. Zhang and Y. Song, "Machine learning for vehicle intrusion detection using reinforcement learning," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8804–8813, 2020.

M. Ahmed, F. Hussain, and A. Khan, "CAN bus intrusion detection employing K-means clustering," *Security and Communication Networks*, vol. 2021, 2021.

R. R. Ahmad, I. Khan, and M. A. Khan, "Effective CAN bus anomaly detection using LSTM networks," *IEEE Sensors Journal*, vol. 21, no. 16, pp. 17920–17931, 2021.

T. Liu, Y. Mao, and L. Chen, "Adversarial machine learning in automotive cybersecurity," *IEEE Transactions on Cybernetics*, vol. 50, no. 6, pp. 2443–2455, 2020.

F. Wang and J. Huang, "A novel intrusion detection model based on CNN-LSTM for in-vehicle CAN networks," *Security and Communication Networks*, vol. 2022, 2022.

S. Kim and J. Lee, "CAN message anomaly detection using gated recurrent units," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 1406–1417, 2020.

W. Chen, Z. Liu, and H. Yu, "Hybrid machine learning for automotive IDS with feature importance," *Neural Computing and Applications*, vol. 33, pp. 987–1002, 2021.

H. Park, J. Lee, and D. Kim, "Online learning-based CAN bus intrusion detection for connected vehicles," *IEEE*

Transactions on Vehicular Technology, vol. 70, no. 6, pp. 5555–5565, 2021.

Y. Zhao, J. Li, and F. Li, "An efficient CAN bus intrusion detection system based on k-nearest neighbors," *Security and Communication Networks*, vol. 2021, 2021.

R. Garcia, V. Marin, and P. Rios, "Transfer learning for CAN bus intrusion detection," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13687–13698, 2020.

Z. Hu, Q. Lin, and P. Zhang, "Deep autoencoder for CAN bus anomaly detection," *IEEE Sensors Journal*, vol. 21, no. 22, pp. 25820–25830, 2021.

S. Gupta, M. Panda, and R. Das, "Intrusion detection using ensemble learning in automotive networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6654–6666, 2022.

A. Das, S. Roy, and K. Roy, "Secure intrusion detection framework for connected vehicles using reinforcement learning," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 3, pp. 1755–1767, 2021.

J. Song and Y. Yin, "Lightweight deep learning model for real-time CAN bus intrusion detection," *IEEE Access*, vol. 8, pp. 71269–71278, 2020.

M. Tang, Y. Peng, and J. Sun, "A survey of vehicle-to-everything (V2X) security," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3243–3266, 2021.

F. Zhao, Y. Cao, and H. Wang, "An adaptive IDS framework for automotive cyber-physical systems," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 1, pp. 12–25, 2022.