Deep Learning based Enhanced Intrusion Detection for Vehicular Network

Honey K Amin[1], Prof. Deepak Upadhyay[2]
[1] *PG Scholar, Master of Engineering in Cyber Security, GTU- SET*
[2] *Assistant Professor, GTU- SET*

***Abstract*—The rapid advancement of autonomous vehicle technologies has significantly improved vehicle control systems, primarily through the Controller Area Network (CAN) bus protocol. However, the inherent complexity and openness of CAN networks expose them to numerous cybersecurity issues. Despite the CAN bus's crucial role, its susceptibility to cybersecurity threats, particularly spoofing attacks, remains a significant concern. Our study presents an optimized Intrusion Detection System based on Bidirectional Long Short-Term Memory (LSTM) along with Convolutional Neural Networks (CNN). This work is designed to detect and mitigate attacks on CAN networks through the application of advanced deep learning techniques. In addition to the core CNN model, the incorporation with LSTM to enhance the system's accuracy and robustness.**

***Index Terms*—Autonomous Vehicles., Bidirectional LSTM (BiLSTM), Controller Area Network (CAN), CyberSecurity, Convolutional Neural Network (CNN), Intrusion Detection System (IDS)**

I. INTRODUCTION

Modern vehicles are becoming highly interconnected due to rapid advancements in technologies such as the Internet of Things (IoT) and the Internet of Vehicles (IoV). These developments enable seamless communication between vehicles and digital systems, significantly enhancing safety, performance, and user experience. However, this increasing connectivity also introduces substantial cybersecurity risks, particularly within a vehicle's internal communication systems.

1.1 The Role of CAN in Modern Vehicles

The evolution of Connected Vehicles (CVs) and Autonomous Vehicles (AVs) relies heavily on robust internal communication networks. These are facilitated by In-Vehicle Networks (IVNs), which manage data exchange among numerous Electronic Control Units (ECUs). At the core of IVNs lies the Controller Area Network (CAN) bus, a communication protocol developed by Bosch in the 1980s. The CAN protocol allows microcontrollers and devices to communicate without the need for a host computer, and it is widely adopted in automotive and industrial applications due to its real-time communication and reliability.

CAN frames are designed for efficiency and fault tolerance, allowing for fast data exchange between sensors, actuators, and controllers. Despite these benefits, the CAN protocol suffers from fundamental security limitations. It lacks built-in authentication and encryption mechanisms, making it vulnerable to unauthorized access, data injection, and spoofing attacks.

1.2 Security Concerns in CAN Communication

The CAN bus uses a broadcast communication model where every ECU can send and receive messages. This lack of message origin verification means that a compromised node can impersonate others and inject malicious packets. Common attacks include denial-of-service (DoS), fuzzy attacks (random data injection), and spoofing (false data replication), all of which can disrupt vehicle functionality and compromise safety.
Although traditional security measures such as encryption, firewalls, and rule-based Intrusion Detection Systems (IDS) have been implemented to counter these threats, they often fall short in detecting sophisticated, time-sensitive attacks. These conventional solutions either introduce system overhead or lack the intelligence needed to identify previously unseen attack patterns.

1.3 The Importance of Intrusion Detection Systems

Intrusion Detection Systems are critical for identifying abnormal patterns in network traffic. In the automotive context, IDS monitors the CAN bus to detect suspicious activity across ECUs. Historical research in IDS dates back to early work in network security in the 1990s, and its importance has only grown with the advent of connected automotive systems.

In CAN-based vehicle networks, IDS is tasked with detecting attacks that may not be obvious through traditional monitoring. These include subtle message spoofing or temporal manipulation of packet sequences that evade static detection rules. Therefore, IDS must evolve to handle both known and unknown threats, in real-time, and with minimal computational impact.

1.4 Deep Learning for CAN Intrusion Detection

Deep learning has emerged as a powerful approach for developing intelligent, adaptive IDS. Unlike traditional rule-based methods, deep learning models can automatically learn patterns from large datasets, enabling them to identify novel attack vectors with high accuracy.

Convolutional Neural Networks (CNNs) are effective in extracting spatial patterns from CAN data, such as byte-level anomalies. Meanwhile, Long Short-Term Memory (LSTM) networks excel at capturing temporal dependencies, making them suitable for analyzing sequences of CAN messages over time. However, traditional CNN-LSTM architectures often treat individual frames independently before modeling sequence relationships, which may reduce temporal learning effectiveness.

II. LITERATURE REVIEW

2.1 Overview of CAN Communication and Security Challenges

The Controller Area Network (CAN) bus is a foundational protocol in modern automotive systems, enabling real-time communication among Electronic Control Units (ECUs) without relying on a host computer. Despite its robustness and low latency, the CAN protocol was not originally designed with cybersecurity in mind. It lacks basic security mechanisms such as message authentication, encryption, or sender verification, leaving it vulnerable to a variety of attacks including Denial of Service (DoS), spoofing, fuzzing, and replay attacks [1][2].

As automotive systems become increasingly interconnected—especially with the growth of the Internet of Vehicles (IoV) and Autonomous Vehicles (AVs)—the risk of cyber intrusion into vehicle networks has significantly increased. Intruders can exploit these vulnerabilities to manipulate sensor data, disable safety features, or take control of vehicular functions, potentially endangering lives and causing financial loss [3][4].

2.2 Traditional Intrusion Detection Techniques

Early Intrusion Detection Systems (IDS) for CAN networks relied heavily on rule-based and statistical anomaly detection techniques. Rule-based IDS detect specific patterns or thresholds, while statistical methods learn the normal behavior of the network and flag deviations. However, these approaches suffer from limited adaptability and high false positive rates, especially in dynamic vehicular environments [5].

More recent traditional methods have employed machine learning (ML) algorithms such as Decision Trees, Random Forests, and Support Vector Machines (SVM) for intrusion detection [6]. While these ML models outperform rule-based systems in adaptability, they often depend heavily on feature engineering and may not generalize well to complex or unseen attack patterns.

2.3 Deep Learning for CAN Intrusion Detection

To address the limitations of traditional approaches, researchers have increasingly turned to deep learning (DL) techniques for IDS in CAN networks. DL models can automatically extract meaningful features from raw data, making them highly suitable for detecting complex attack patterns without manual feature selection.

2.3.1 CNN-Based IDS Models

Convolutional Neural Networks (CNNs) have been widely adopted for their ability to learn local spatial features in structured data. In the context of CAN

intrusion detection, CNNs are used to extract patterns from byte-level sequences of CAN frames [7]. However, CNNs are limited in capturing temporal dependencies between sequential messages, which is a critical aspect of identifying time-based attacks such as replay or spoofing.

2.3.2 LSTM and CNN-LSTM Models

2.4 Bidirectional LSTM-CNN Approaches

Recent advancements have introduced Bidirectional LSTM (BiLSTM) layers to enhance temporal modeling. Unlike standard LSTM, BiLSTM processes the input sequence in both forward and backward directions, allowing the model to capture past and future context simultaneously. When combined with CNN layers, the BiLSTM-CNN hybrid achieves superior performance by learning robust spatial-temporal representations [10][11].

This hybrid architecture addresses the shortcomings of previous models by:

- Preserving sequential dependencies between CAN messages.
- Enabling two-way temporal learning for more precise intrusion detection.
- Maintaining computational efficiency suitable for real-time deployment.

Studies have reported that BiLSTM-CNN models outperform traditional ML models, standalone CNNs, and CNN-LSTMs in detecting various cyberattacks in CAN networks with higher accuracy and fewer false positives.

While deep learning-based IDS models have shown promise, most existing studies either focus on spatial analysis (CNN) or unidirectional sequence learning (LSTM). There is a lack of hybrid models that efficiently integrate bidirectional temporal context with spatial features. Furthermore, few studies evaluate their performance comprehensively across different types of CAN-based attacks in real-world datasets.

This research aims to address these gaps by proposing and evaluating a BiLSTM-CNN model for CAN intrusion detection, offering improved detection performance without sacrificing computational efficiency.

III. METHODOLOGY

This section outlines the methodological framework used to develop, train, and evaluate a hybrid deep learning model for intrusion detection in Controller Area Network (CAN) systems. The proposed model leverages both Bidirectional Long Short-Term Memory (BiLSTM) and Convolutional Neural Network (CNN) layers to identify cyberattacks based on CAN traffic data.

3.1 Data Collection and Preprocessing

The dataset used in this study comprises CAN bus traffic containing five classes: Normal, Denial of Service (DoS), Fuzzy, Gear Spoofing, and RPM Spoofing attacks. Each data entry consists of an 8-byte payload representing a CAN message. The raw hexadecimal data was preprocessed and transformed into binary or numerical formats suitable for deep learning input. Label encoding was applied for multi-class classification, and normalization techniques were used to standardize the input features.

To preserve the temporal structure of the CAN data, sequences of messages were grouped and reshaped into sliding windows before feeding into the network. This ensured that temporal relationships between consecutive messages could be effectively captured.

3.2 Model Architecture

The proposed architecture integrates a Bidirectional LSTM layer followed by CNN layers, as illustrated in Figure 3.1 (architecture diagram). The BiLSTM layers process the input sequences in both forward and backward directions to capture past and future temporal dependencies. The resulting feature representations are passed through 1D CNN layers to extract spatial patterns and local dependencies among the CAN bytes.

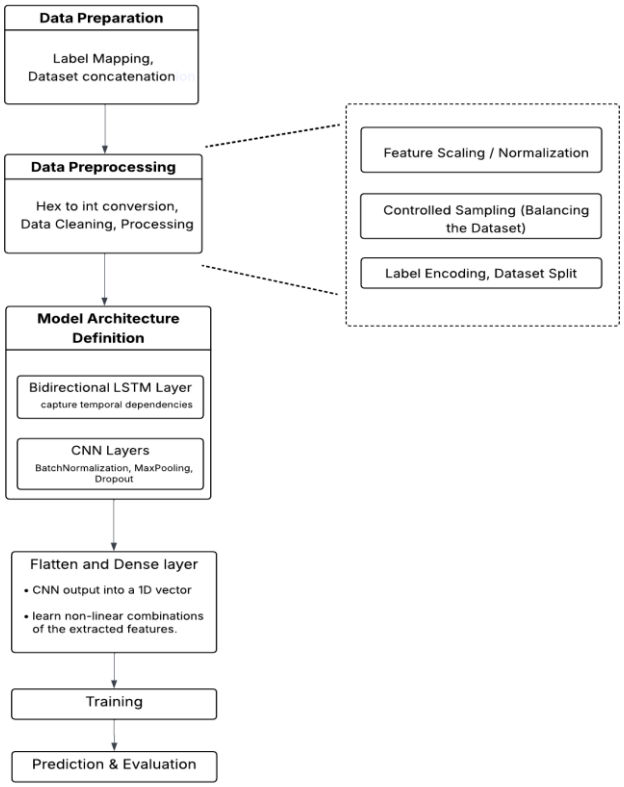

Figure 3.1: Proposed Workflow

This design reverses the traditional CNN–LSTM architecture, which often leads to partial loss of temporal information during early CNN processing. By applying BiLSTM first, the model preserves temporal dependencies across the message sequence before spatial abstraction, enhancing its ability to identify time-dependent attack patterns.

3.3 Training Procedure

The model was trained using a supervised learning approach with a categorical cross-entropy loss function and the Adam optimizer. Training was conducted over 10 epochs with early stopping to prevent overfitting. A 80:20 train-test split was applied, and performance was monitored on both training and validation sets.

Batch normalization and dropout layers were included to stabilize learning and reduce overfitting. The model's hyperparameters (such as learning rate, batch size, and number of filters/units) were selected empirically based on preliminary experiments.

Experimental Setup

This section outlines the practical implementation of the proposed Bidirectional LSTM–CNN model for detecting cyberattacks in Controller Area Network (CAN) traffic. The implementation was conducted using Python and deep learning libraries such as TensorFlow and Keras.

All experiments were performed on a system with the following configuration:

Processor: Intel Core i7

RAM: 32 GB

GPU: Google Colab

OS: Ubuntu 22.04 / Windows 11

Frameworks: Python 3.10, TensorFlow 2.x, Keras, NumPy, Pandas

The dataset contained normal and malicious CAN traffic labeled into five categories: Normal, DoS (Denial of Service), Fuzzy attack, Gear Spoofing, RPM Spoofing. Each CAN frame consists of 8 bytes. The raw hexadecimal values were converted to 64-bit binary sequences and normalized. A sliding window technique was used to segment sequences for training, with each sample consisting of a fixed-length

sequence of consecutive CAN frames.

The proposed model uses a Bidirectional LSTM layer to capture past and future dependencies, followed by 1D Convolutional and pooling layers for spatial feature extraction. Finally, dense layers are used for classification.

IV. RESULT AND DISCUSSION

In this study, we explored a deep learning model designed to identify attack behaviors within a CAN bus system. To evaluate the proposed system, we used experimental data to detect attack messages on the CAN bus. The CNN model demonstrated good accuracy, it showed signs of overfitting, indicating that it performed well on the training data but might not generalize as effectively.[9] Intrusion detection in autonomous vehicle networks is crucial for identifying malicious traffic and monitoring CAN bus systems to differentiate between normal and abnormal messages exchanged among various ECUs (Electronic Control Units).

This study presents a comparative evaluation of two deep learning models—CNN-LSTM and Bidirectional LSTM-CNN for detecting intrusion attacks in Controller Area Network (CAN) bus traffic. The primary objective was to identify attack behaviors such as DoS, fuzzing, gear spoofing, and RPM spoofing within automotive networks.

4.1 Performance of Proposed Model

The proposed BiLSTM-CNN architecture reverses the standard CNN-LSTM sequence by applying Bidirectional LSTM first. This approach captures both past and future dependencies before applying CNN for spatial pattern extraction. This preserved the temporal context, which is essential for detecting time-evolving attacks. This model achieved a higher overall accuracy of 96%, with improvements in all key performance metrics across different attack types. This model's accuracy curve showed steady learning with minimal overfitting, and its validation performance closely matched the training performance.

The CNN-LSTM model combined convolutional layers for local feature extraction with LSTM layers for capturing temporal dependencies in sequential CAN data. While it achieved a reasonably high classification accuracy of 91.06%, it showed indications of overfitting, where training accuracy was significantly better than validation accuracy. This suggests that although the model learned patterns from the training data, its generalization to unseen data was limited.

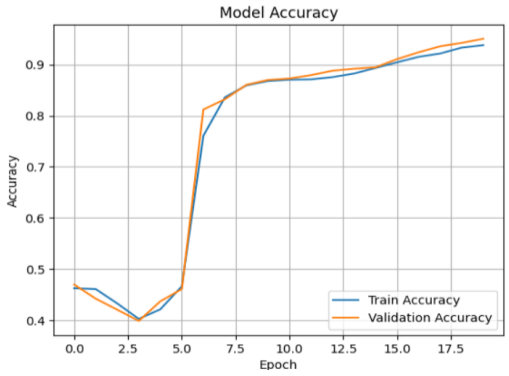

Figure 4.1 Model Accuracy

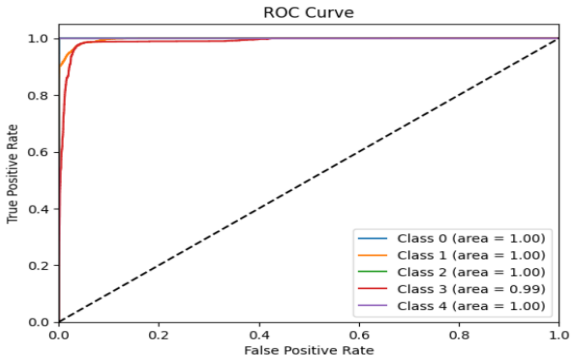

Figure 4.2 Roc Curve

The results clearly demonstrate the superiority of the BiLSTM-CNN model in terms of accuracy and generalization. Unlike CNN-LSTM, which processes each 8-byte CAN frame independently at the CNN stage (possibly losing temporal order), the BiLSTM-CNN captures temporal relationships before spatial patterns. This leads to more effective identification of complex and subtle attack patterns.

Furthermore, the hybrid approach of combining BiLSTM with CNN provides a better balance of detection accuracy and computational efficiency, making it more suitable for real-time in-vehicle intrusion detection systems.

4.2 Comparative Analysis

To evaluate the effectiveness of deep learning models in detecting intrusions in CAN bus networks, we conducted a comparative analysis between the LSTM-CNN and BiLSTM-CNN architectures using the same dataset and experimental conditions. The LSTM-CNN model utilizes a unidirectional Long Short-Term

Memory (LSTM) layer followed by convolutional layers. The LSTM captures sequential dependencies in a forward direction only, making it suitable for time-series classification tasks.

The BiLSTM-CNN architecture enhances the temporal feature extraction by employing a Bidirectional LSTM layer that processes data in both forward and backward directions. This allows the model to gain a more comprehensive understanding of the sequence, capturing relationships from both past and future data points.

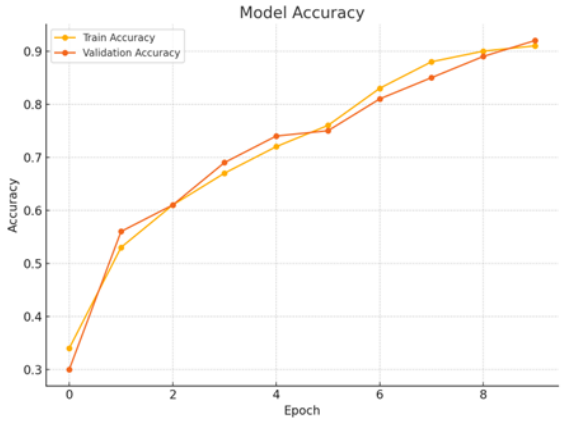

Figure 4.3 CNN-LSTM Model

Accuracy Achieved: 91.06%
Observation: While the LSTM-CNN model performs well, it is limited in its ability to fully capture context, as it only processes input in one temporal direction (past to future). This may cause loss of important future context, especially in cases where attacks exhibit patterns that depend on both preceding and succeeding messages.

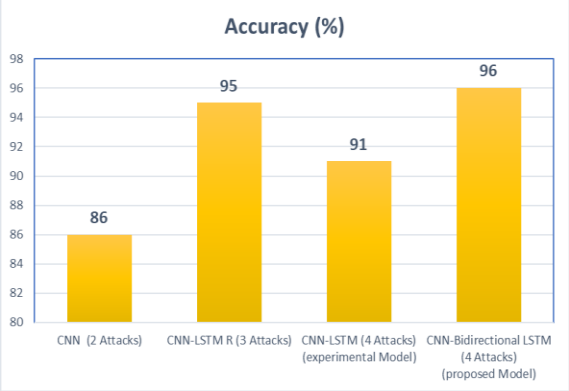

Figure 4.4: Comparative Analysis

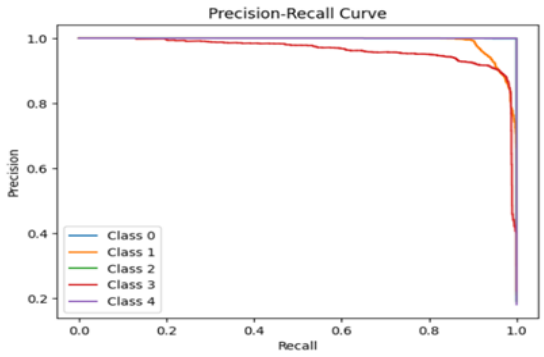

Figure 4.5: Proposed model Precision-Recall Curve

The BiLSTM-CNN consistently outperformed the LSTM-CNN across all attack types. The bidirectional structure significantly improved the model's ability to detect subtle patterns in CAN messages, leading to better classification performance and fewer false negatives. The BiLSTM-CNN model first applies Bidirectional LSTM to capture both past and future temporal dependencies across sequences. These enriched temporal representations are then passed through CNN layers for spatial feature extraction, resulting in a more comprehensive understanding of attack patterns. This reversed architecture ensures that temporal context is preserved before spatial abstraction, improving detection accuracy for complex or time-dependent attacks.

V. CONCLUSION

The proposed Bidirectional LSTM-CNN model demonstrated strong effectiveness in detecting malicious CAN bus messages, showing high accuracy in identifying and classifying various types of cyberattacks. The proposed systems demonstrated efficient detection of abnormal packets, thereby enhancing the security of the CAN bus. Additionally, these models can be extended to other security system designs within the complex infrastructures of autonomous vehicle networks, ensuring secure data processing across various vehicle communication frameworks. As a future scope, the model can be extended by incorporating more diverse and complex attack scenarios, improving its robustness and generalization as a comprehensive Intrusion Detection System (IDS).
The results demonstrate that incorporating both temporal and spatial analysis in network traffic

significantly enhances the model's ability to identify various types of intrusions such as DoS, fuzzing, spoofing, and gear attacks. This contributes to the development of more robust and intelligent Intrusion Detection Systems (IDS) for in-vehicle networks.

REFERENCES

[1] Theyazn H. H. Aldhyani, and Hasan Alkahtani. Attacks to automatous vehicles: A deep learning algorithm for cybersecurity.MDPI Sensors v 2(1), 360 2022

[2] Rajapaksha, Sampath, et al. "Ai-based intrusion detection systems for in-vehicle networks: A survey." ACM Computing Surveys 55.11 (2023): 1-40.

[3] Bari, Bifta Sama, Kumar Yelamarthi, and Sheikh Ghafoor. "Intrusion detection in vehicle controller area network (can) bus using machine learning: A comparative performance study." Sensors 23, no. 7 (2023): 3610.

[4] Baldini, Gianmarco. "Detection of cybersecurity spoofing attacks in vehicular networks with recurrence quantification analysis." Computer Communications 191 (2022): 486-499.

[5] Baldini, Gianmarco. "In-Vehicle Network Intrusion Detection System Using Convolutional Neural Network and Multi-Scale Histograms." Information 14, no. 11 (2023): 605.

[6] Adly, Salah, Ahmed Moro, Sherif Hammad, and Shady A. Maged. "Prevention of Controller Area Network (CAN) Attacks on Electric Autonomous Vehicles." Applied Sciences 13, no. 16 (2023): 9374.

[7] Taylor, A.; Leblanc, S.; Japkowicz, N. Anomaly detection in automobile control network data with long short-term memory networks. In Proceedings of the IEEE International Conference on Data Science and Advanced Analytics (DSAA 2016), Montreal, QC, Canada, 17–19 October 2016; pp. 130–139.

[8] Miller, C. Lessons learned from hacking a car. IEEE Des. Test Comput. 2019, 36, 7–9

[9] Young, C.; Zambreno, J.; Olufowobi, H.; Bloom, G. Survey of Automotive Controller Area Network Intrusion Detection Systems. IEEE Des. Test Comput. 2019, 36, 48–55.

[10] Madrigal, A.C. Inside Waymo's Secret World for Training Self-Driving Cars. In The Atlantic; Carnegie Mellon University: Pittsburgh, PA, USA, 23 August 2017.

[11] Shahroz Tariq, Sangyup Lee, Huy Kang Kim, and Simon S. Woo. 2020. CAN-ADF: Thecontroller area network attack detection framework. Computers and Security 94 (2020), 101857.

[12] Hyun Min Song, Jiyoung Woo, and Huy Kang Kim. 2020. In-vehicle network intrusion detection using deep convolutional neural network. Vehicular Communications 21 (2020), 100198.

[13] Vasudev, H.; Das, D.; Vasilakos, A.V. Secure message propagation protocols for IoVs communication components. Comput. Electr. Eng. 2020, 82, 106555.

[14] Mehedi, S.T.; Anwar, A.; Rahman, Z.; Ahmed, K. Deep Transfer Learning Based Intrusion Detection System for Electric Vehicular Networks. Sensors 2021, 21, 4736.