RESEARCH ARTICLE

Mitigating Message Injection Attacks in Internet of Vehicles Using Deep Learning Based Intrusion Detection System

Muhammad Hammad Nawaz[1,*], Abrar Ahsan[1], Inam Ullah Khan[2], Yanan Wang[3], Mushtaq Ahmad[4] and Muhammad Shoaib Akhtar[5]

[1] Department of Statistics and Data Science, University of Mianwali, Mianwali 42200, Pakistan
[2] Department of Computer Science, Qurtuba University of Science and Information Technology, Peshawar 25000, Pakistan
[3] Department of Computer Science and Engineering, Sejong University, Seoul 05006, Republic of Korea
[4] Department of Business Informatics, Technical University of Vienna (TU Wien), Vienna, Austria
[5] Department of Electrical Engineering, University of Science and Technology, Bannu, Pakistan

Abstract

Real-time communication between autonomous vehicles, infrastructure, and their environment has facilitated the Internet of Vehicles (IoVs). Although this connectivity provides vehicular networks with significant benefits, it also introduces severe security threats, such as message injection attacks, particularly due to the heavy reliance on the Controller Area Network (CAN) protocol, which is inherently vulnerable. Electronic Control Units (ECUs) become primary targets for these attacks, leading to unsafe vehicle behaviors. To address these challenges, an Intrusion Detection System (IDS) based on deep learning architectures, Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) is proposed for detecting and classifying cyberattacks in vehicular networks. Data preprocessing techniques such as NearMiss, Random Over-Sampling (ROS), and Tomek Links are applied to handle class imbalance in the car hacking dataset. A benchmark Car-Hacking dataset, evaluation metrics including accuracy, precision, recall, and F1 score, are used to assess the performance of the models. The experimental results demonstrate that the GRU model achieves the highest accuracy of 99%, followed by LSTM with 98%, and RNN with 94%. These findings indicate that GRU outperforms the other models and, in certain configurations, detects 100% of the attacks. The proposed IDS exhibits considerable superiority over traditional deep learning approaches, presenting a promising and intelligent solution for enhancing the cybersecurity of modern IoV systems.

Keywords: internet of vehicles, car hacking, random over-sampling (Ros), nearmiss, imbalanced data, intrusion detection systems, deep learning.

Submitted: 16 May 2025
Accepted: 23 July 2025
Published: 21 September 2025

Vol. 1, **No.** 4, 2025.
10.62762/TACS.2025.560376

***Corresponding author:**
✉ Muhammad Hammad Nawaz
hammadumw@gmail.com

Citation
Nawaz, M. H., Ahsan, A., Khan, I. U., Wang, Y., Ahmad, M., & Akhtar, M. S. (2025). Mitigating Message Injection Attacks in Internet of Vehicles Using Deep Learning Based Intrusion Detection System. *ICCK Transactions on Advanced Computing and Systems*, 1(4), 208–221.

© 2025 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

1 Introduction

Most recently, deep learning has been adopted widely in many areas of healthcare, transportation, and industrial systems [1, 2]. It has been applied to common applications such as image classification and fraud prevention, or to increase cybersecurity [3, 4]. It is worth mentioning in particular the increasing importance of cybersecurity for the detection and mitigation of threats on the Internet of Things (IoT) and, even more, Industrial IoT ecosystems [5, 6]. Different deep learning architectures, including Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU), are applied to recognize cyber threats in different environments such as IoT platforms, Software Defined Networks (SDN), and cloud computing. With cyber dependency on the rise, the number of cybersecurity issues has also increased. Intrusion Detection Systems (IDS) network protection is one of the critical issues, and they need to be developed robustly to avoid cyberattacks [7]. Efficient analysis of traffic flow (or cybersecurity data) on such systems is crucial to detect anomalies and malicious activities in modern communication networks.

Due to the rapid increase of information and communication technologies, contemporary vehicles have taken the form of networked systems. In addition to autonomous vehicles, this refers to Vehicle-to-Grid (V2G) and Grid-to-Vehicle (G2V) technologies [8]. Part of these vehicles are integrated into Internet of Vehicles (IoV) frameworks, namely intra-vehicle networks (IVNs) and external communication systems. The Controller Area Network (CAN) bus is used by IVNs to pass the information among Electronic Control Units (ECUs) to enable their functions and coordinate operations. Meanwhile, smart vehicles communicate with other IoV components such as other smart vehicles, pedestrians, infrastructure, and roadside units via external communication links [9, 10]. Autonomous Electric Vehicles (AEVs) can be considered the future of mobility due to the advantages they offer over traditional vehicles [8]. Equipped with the latest technologies such as radar, cameras, GPS, AI, and cloud connectivity, AEVs operate without human control and adapt dynamically to their environments [10]. These vehicles communicate in real time via wireless technologies like 4G and 5G with other vehicles and traffic systems to improve traffic flow and provide a much safer and more comfortable driving experience. Features like auto parking, remote operation, and live traffic updates are also included [11]. However, since AEVs rely heavily on network connectivity, new cybersecurity risks are introduced, including software vulnerabilities, sensor data spoofing, remote exploitation of onboard systems, and unauthorized access to critical vehicle controls. The convergence of technologies such as 5G, artificial intelligence, and digital twins [26] is creating new paradigms for securing cyber-physical systems like IoV. These technologies offer the potential for more dynamic, adaptive, and predictive security frameworks that can respond to threats in real-time. Automotive manufacturers must impose strict cybersecurity protocols and adopt secure communication standards to ensure the safety and integrity of such systems, as illustrated in Figure 1.

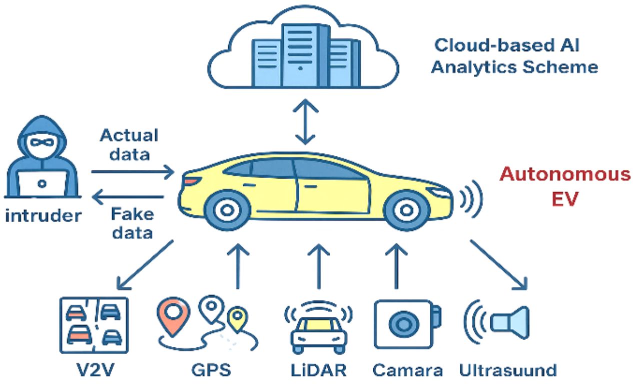

Figure 1. Intrusion system analysis for autonomous car in smart city.

The rise of deep learning in the automotive domain nowadays enables attackers to develop more artificially intelligent and targeted techniques for compromising vehicle systems. Deep learning allows for the analysis of large-scale data and the discovery of patterns and vulnerabilities that cannot be detected by human analysts. This capability can also be misused for developing advanced hacking techniques that exploit common software or hardware flaws, and for creating evasive malware that bypasses conventional security safeguards. Consequently, automakers and cybersecurity professionals must stay up to date to overcome the emerging challenge of deep learning-powered vehicle attacks [12, 13].

A deep learning-based attack detection system, specifically using architectures like RNN, LSTM, and GRU, is necessary to improve the cybersecurity posture of modern in-vehicle systems and industrial control architectures. On balanced datasets, traditional classification models may perform adequately, but real-world car hacking data is generally imbalanced and thus not feasible. This imbalance deceives models into favoring the majority class, which is inaccurate because the minority class typically represents the actual attacks. Given this, this study applies various data balancing techniques to car hacking datasets before carrying out classification using RNN, LSTM, and GRU architectures.

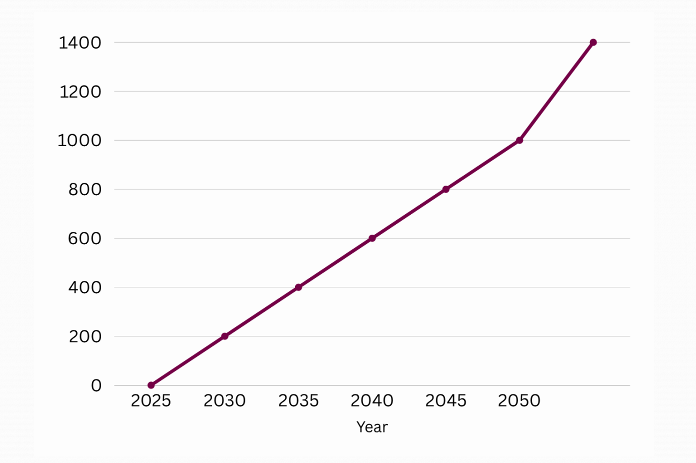

Figure 2. The forecast for the global market of self-driving cars is expected to increase by 25% yearly.

Figure 2 presents the forecast for the global market of self-driving cars with an increase of 25% yearly. The aim is to increase the model's capacity to accurately detect and predict various types of cyberattacks. The framework is evaluated by selecting some key performance indicators like accuracy, precision, and recall, and adopting them, which enables results in better performance than current schemes. Deep learning algorithms operate best if they balance the number of instances across classes, as most of them strive to minimize classification error. A model trained on such unbalanced data will achieve high accuracy by spending most of its time labeling the majority class, but many focus fails to recognize the minority class, which is usually the most relevant. Three main methods of dealing with imbalanced data are recognized.

- The Near-Miss method reduces the majority class samples that lie close to the minority class samples to balance the dataset without duplication.
- Random Over-Sampling replicates minority class samples to generate more instances easily.
- The algorithm builds clean data sets through Tomek Links by finding ambiguous cases and then removing them, which boosts model results.

The contributions of this work consist of designing a robust car attack detection framework tailored to problems arising from imbalanced data in current Internet of Vehicles (IoV) environments. To improve significantly the accuracy of intrusion detection and attack classification, the proposed solution combines deep learning algorithms with data balancing techniques. To demonstrate the effectiveness of the framework a case study is carried out over the Car-Hacking dataset under the IoV systems for smart city applications. All of the key contributions of this work are as follows:

- A deep learning architecture-based real-time threat detection methodology and a comprehensive Intrusion Detection System (IDS) for vehicular networks using LSTM, GRU, and RNN is provided. To address the problem of the vehicle cyber-attack dataset's inherent class imbalance in a framework, multiple data balancing techniques are effectively integrated with this framework.
- Preprocessing of the Car-Hacking dataset using resampling methods, namely Near-Miss, Random Over-Sampling (ROS), and Tomek Links, is done to improve the learning performance of these models. Using these techniques, the dataset can be converted from an imbalanced to a balanced state while achieving a great improvement in detection outcomes.
- The combination of ROS, GRU model, and Tomek Links is demonstrated to obtain a 100% detection rate on Internal and External vehicular communication scenarios. This proves that the proposed IDS is robust and reliable for real-world IoV Environments.

In the remainder of this paper, Section 2 outlines the related work, Section 3 describes the proposed attack prediction framework for IoV systems, Section 4 explains the high-level IDS-IoV architecture for detecting car-based attacks, Section 5 evaluates it over the dataset and evaluation criteria, and finally, Section 6 concludes the study.

2 Related Work

This section summarizes the most significant advancements in intrusion attack detection and vehicle cybersecurity proposed in previous research. The proposed solutions generally fall into two main categories: cryptographic systems and deep learning-based intrusion detection systems. Deep learning-based intrusion detection typically follows three main detection methods: signature-based, anomaly-based, and specification-based. Among these, the anomaly-based technique is the most widely adopted due to its ability to identify unusual behaviors without relying on predefined patterns. Modern vehicles are equipped with hundreds of sensors generating high-dimensional data from both internal components and external communication with infrastructure and other vehicles. Deep learning has emerged as a dominant method for analyzing this data due to its ability to model complex patterns and detect subtle anomalies, which are often missed by rule-based systems. These systems can also detect previously unseen (zero-day) attacks, making them suitable for dynamic automotive environments [14]. To reflect progress in this domain, several studies have examined the role of deep learning in improving security within intelligent transportation systems. One comprehensive review classified various deep learning approaches for intrusion detection in connected and autonomous vehicles, highlighting their potential to provide adaptive, scalable, and automated defense mechanisms.

Recent works have leveraged benchmark automotive cybersecurity datasets, such as the Car-Hacking dataset, and developed specialized models that target specific attack types like DoS, Fuzzy, Spoofing RPM, and Spoofing gear. Instead of using classical classification models, modern approaches have turned to advanced neural network architectures. For example, Convolutional Neural Networks (CNNs) have been used to identify Denial-of-Service and spoofing threats by processing raw CAN bus data directly. While some CNN-based models show promising results offline, real-time deployment and compatibility with legacy vehicle systems remain challenges.

Furthermore, Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) models have been proposed for their effectiveness in modeling sequential vehicle data and learning temporal patterns associated with cyberattacks [15, 16]. These models are particularly well-suited for time-series data generated by vehicle communication systems and have shown enhanced performance in detecting anomalies and classifying multiple attack types. Advanced deep learning frameworks also include optimized CNN architectures designed for Vehicle-to-Everything (V2X) communication in 5G networks. These models apply hyperparameter tuning techniques to maximize detection accuracy and minimize false positives in real-time environments. In summary, existing research demonstrates a clear shift toward deep learning as a foundational technology in modern Intrusion Detection Systems (IDS) for vehicular networks. The application of deep neural networks, particularly CNNs, RNNs, LSTMs, and GRUs, reflects an increasing awareness of the complex nature of cybersecurity in connected and autonomous vehicles. While most prior work focused on binary classification, our approach tackles the more challenging problem of multiclass classification to detect various attack types. Despite the advancements, challenges such as real-time integration, scalability, and compatibility with existing systems persist, emphasizing the need for continued research in developing robust, intelligent, and adaptive IDS frameworks for the Internet of Vehicles. The security of these frameworks is paramount, as they underpin the safe operation of all vehicle functionalities, including critical cooperative control [23] and autonomous navigation systems [25]. In such an explosion of smart cities and the standards of integration of the Internet of Vehicles (IoV), security becomes a major concern in vehicular systems. Due to the growing use of in-vehicle communication systems, the attack surface has been enlarged, and both in general and external vehicular networks have become targets for cyberattacks.

Wireless interfaces of servers can be used by attackers to carry out remote attacks, while the On-Board Diagnostics II (OBD-II) provides an opportunity to attack In-Vehicle Networks (IVNs) [16]. Several studies propose Intrusion Detection Systems (IDS) appropriate for the automotive environment to combat these threats. Figure 3 shows the IDS-protected vehicular network architecture. In this regard, our work takes a look at the use of sequential deep learning, i.e., recurrent neural networks (RNNs), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) to improve the detection of time-dependent patterns in vehicular traffic data. However, unlike traditional deep learning models, these architectures are very well suited for modeling temporal dependencies that are a necessity when operating on subtle anomalies determined in sequential automotive communication data.

Moreover, in models such as DeepSecDrive, lids offered a lightweight and interpretable IDS framework. The other approach is federated learning-based IDS (FL-IDS) that preserves data privacy and still maintains good detection abilities, and hybrid systems

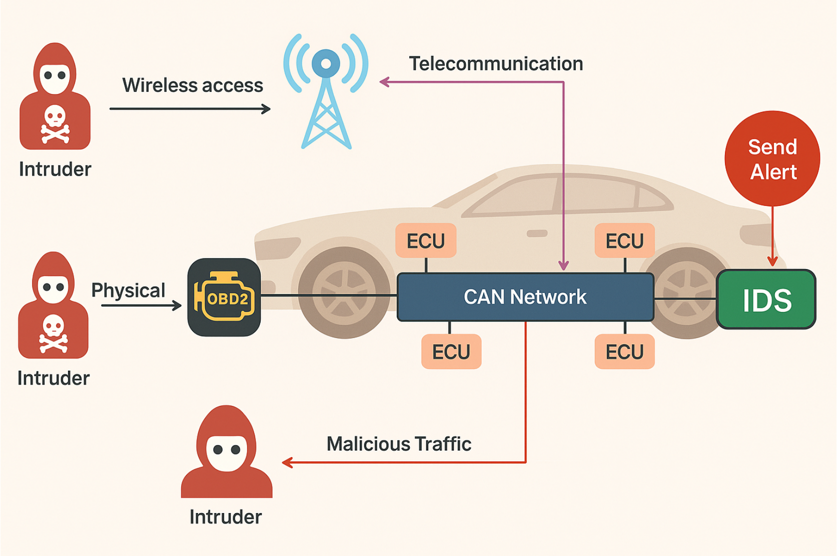

Figure 3. IDS-protected vehicular network architecture highlighting internal and external components secured through deep learning-based intrusion detection.

of LightGBM combined with neural networks to improve the performance. There has been prior work to use Convolutional Neural Network (CNN) based on architectures such as 1D-CNN, Deep CNN (DCNN), and InceptionResNet alike, to meet the objective of in-vehicle intrusion detection, and hence, yielded promising results [17]. More specifically, DeepSecDrive has further developed lightweight and interpretable IDS models. Also, federated learning-based IDS (FL IDS) approaches that keep data private, while preserving robust detection capabilities, are also considered, as well as hybrid systems composed of LightGBM with neural networks that enhance the performance. Still, little work has been mapped out for utilizing the temporal deep learning models, such as RNN, LSTM, and GRU, that might perform better than the CNN, considering the temporal nature of the attack pattern in CAN traffic. To bridge this gap, we propose to employ these recurrent models to identify an assortment of cyberattacks in the vehicular systems. We experimentally verify our design on public datasets and show that GRU, particularly, can offer both a lightweight and powerful solution to achieve high detection accuracy while maintaining efficiency at the same time. It is important to develop an Intrusion Detection System (IDS) targeted towards the Internet of Vehicles (IoV) as the cyber threats against modern vehicular systems are becoming more and more sophisticated. Sequence-based models, to name a few, such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU), have powerful capabilities in detecting temporal patterns in network traffic, but the car hacking data is highly imbalanced and therefore their performance suffers. Usually, these imbalances will result in the inability of the approach to detect the attacks, and consequently, these classes, i.e., the actual attack instances, are misclassified with high probabilities. To address this issue, the proposed approach leverages many data balancing techniques such as Near Miss [18], Random Over-Sampling (ROS), and Tomek Links, for improved model sensitivity on rare and important attack events. RNN-based models are trained to identify and classify various types of intrusions with good accuracy after balancing. As a methodology, this method attempts to decrease detection accuracy to bring in reliability and establishes a dependable IDS framework to protect the intra- & inter-vehicular communication network security.

The Car-Hacking dataset was originally collected from CAN traffic logs in simulated environments at the University of Michigan. While it offers a comprehensive view of various attack types, the synthetic nature of some attacks and the limited range of vehicle models may introduce biases. This limitation motivates future testing on diverse real-world automotive datasets.

Several cyberattacks exist in the Car-Hacking dataset that target the RPM and drive gear indicators through Denial of Service (DoS) attacks and fuzzy attacks, and spoofing attacks. The 5% portion of the Car-Hacking data contains the distribution patterns depicted in Figure 4. Technical data on the different types of attacks exists in the source material [13].

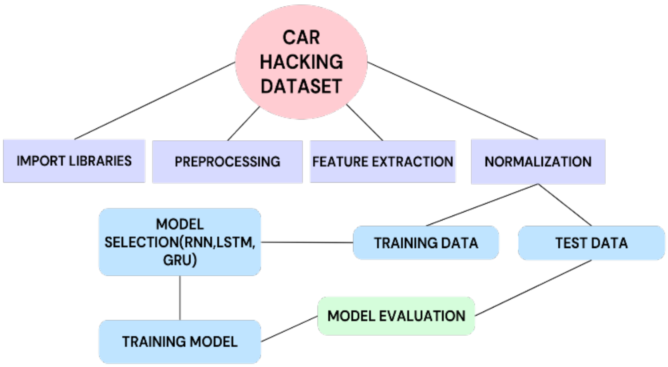

Figure 4. Class distribution for Car Hacking_5% dataset.

Several cyberattacks exist in the Car-Hacking dataset that target the RPM and drive gear indicators through Denial of Service (DoS), fuzzy, and spoofing attacks.

Each attack type in the Car-Hacking dataset targets specific vehicle functions and can disrupt normal operations:

1. Denial of Service (DoS): Overwhelms the CAN bus with high-frequency ’0000’ messages every 0.3 milliseconds, effectively jamming communication between ECUs. This may cause vehicles to stall or behave unpredictably.

2. Fuzzy Attack: Injects CAN messages with random IDs and payloads every 0.5 milliseconds, creating noise and potentially triggering unwanted behaviors in ECUs such as false alerts or improper actuation.

3. Spoofing (RPM/Gear): Sends fake data targeting RPM and gear indicators at 1 ms intervals to mislead the system into showing false readings, which can lead to unsafe driving conditions or bypassing security checks.

The dataset contains records that include six important fields:

- The DoS Attack does a high-frequency transmission of ’0000’ CAN messages every 0.3 milliseconds, thus disrupting the bus communication. At 0.5 millisecond intervals, this attack generates CAN messages with arbitrary ID and DATA values that attempt to bewilder ECU operation and cause crashes.
- The attacker sends falsified CAN messages targeting RPM and gear data through a spoofing attack at a speed of one message every millisecond to trick the system.
- The dataset contains records that include six important fields.
- The message recording time is shown precisely in seconds through the Timestamp field.
- The hexadecimal identifier of the CAN message is known as CAN ID, with an example of 043f.
- DLC serves as Data Length Code to show the DATA field length with a byte count between zero and eight.
- The message contains seven types of DATA fields consisting of individual data bytes with numbers from 0 to 7.
- Flag designates message type marked as ’T’ for a malicious event, while ’R’ stands for a genuine message.

Standard classification evaluation metrics determine the effectiveness of the proposed detection framework through accuracy, precision, recall, and F1-score. The metrics emerge from the confusion matrix that presents the performance summary of intrusion detection models according to Table 2 [26]. False positive rates (FPR) were also monitored across the models. While GRU maintained an FPR of approximately 1.2%, RNN exhibited slightly higher rates (∼3%). Minimizing false positives is crucial in vehicular systems to prevent unnecessary alerts or disruptions. Techniques such as ensemble smoothing or threshold tuning may help reduce these false detections, and will be explored in real-world validation phases.

To provide a comprehensive overview of prior works, Table 1 summarizes representative studies on vehicle system security, outlining their approaches, advantages, and limitations. This comparative analysis highlights the progress in IDS design while also emphasizing the existing gaps that motivate our proposed framework.

False positive rates (FPR) were also monitored across the models. While GRU maintained an FPR of approximately 1.2%, RNN exhibited slightly higher rates (∼3%). Minimizing false positives is crucial in vehicular systems to prevent unnecessary alerts or disruptions. Techniques such as ensemble smoothing or threshold tuning may help reduce these false detections, and will be explored in real-world

validation phases. Recent advancements have shown the potential of Federated Learning (FL) in vehicle security, enabling collaborative intrusion detection across multiple nodes without sharing sensitive data. For example, Yang et al. [9] proposed FL-based hybrid IDS for 5G-enabled IoV that maintains high accuracy while preserving privacy. Similarly, Nabil et al. [19] combined LightGBM and deep networks under FL to enhance detection performance on CAN datasets. However, these methods often face communication delays and model synchronization issues, which our future work aims to address through optimized edge-cloud orchestration.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

$$\text{F1\_Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

3 Proposed Attack Detection Framework

The research develops a strong and smart detection framework for vehicle system cyber-attack identification through the utilization of RNN and its advanced versions, LSTM and GRU deep learning models. The design of the proposed intrusion detection system appears in Figure 5, which illustrates that the system operates in two phases: Data Preparation and Pre-processing.

The choice of NearMiss, ROS, and Tomek Links over other techniques like SMOTE was intentional. SMOTE creates synthetic examples, which can lead to overfitting in time-series CAN data. Instead, NearMiss selects informative examples near class boundaries, ROS retains actual attack data by oversampling minority classes, and Tomek Links remove overlapping samples, improving class separability without generating artificial noise. These techniques work better with sequential traffic data common in vehicular systems.

The proposed deep learning-based intrusion detection framework combining data balancing (NearMiss, ROS, Tomek Links) and sequence models (RNN, LSTM, GRU) for detecting vehicle cyberattacks.

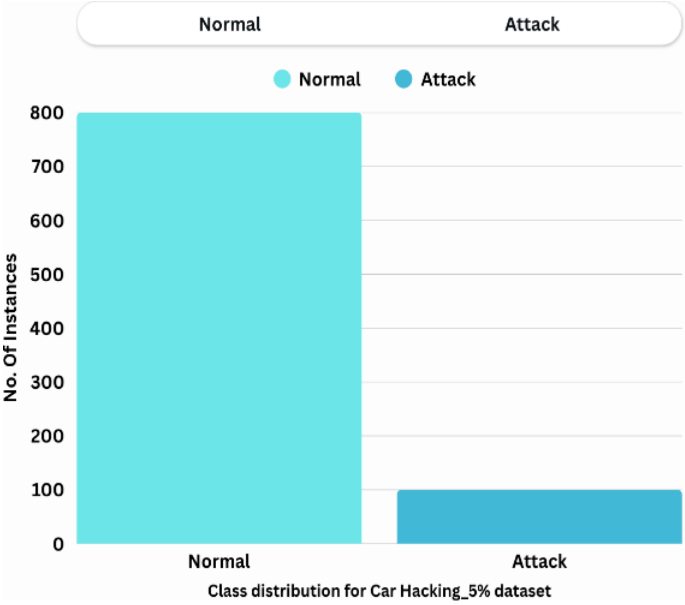

Figure 5. The system operates in two phases, including data preparation and pre-processing.

3.1 The system operates in two phases, including Data Preparation and Pre-processing

The main objective during this phase consists of addressing the built-in class imbalance that exists in the vehicular intrusion dataset. The detection process uses NearMiss and Random Over-Sampling (ROS) and Tomek Links techniques to properly resample data. The NearMiss algorithm finds important majority class examples by selecting instances that are nearest to minority class samples, thus creating a balanced situation without duplicate entries. Rapid Over-Sampling duplicates relevant instances from the minority group randomly to achieve equivalent class distributions in the dataset. By using Tomek Links, the dataset improves its class separability through the removal of ambiguous dual-class overlapping samples [19]. The preprocessing techniques make deep learning models process a balanced dataset that delivers better training outcomes by eliminating bias.

3.2 The second phase includes model training together with evaluation activities

Before partitioning the dataset, a 70:30 training and testing split is implemented after preprocessing. The RNN, together with LSTM and GRU deep learning models, conducts their training process on the data, which was balanced during the initial stage. The developers adjust each model specifically to detect both temporal dependencies along sequential patterns that exist in CAN bus traffic data. During the testing phase, each model receives evaluation through accuracy, precision, recall, and F1-score metrics to

Table 1. Advantages and disadvantages of related work on vehicle system security.

Study	Approach	Advantages	Disadvantages
[10]	IDS for both internal and external vehicular networks using CNN-based IDS	High detection rate (above 99.25%) on Car-Hacking and CICIDS2017 datasets comprehensive protection for internal and external networks	Potential complexity in implementing CNN-based IDS may require significant computational resources
[11]	Analysis of cyber-attack trends in the automotive industry	Highlights the increasing frequency and impact of cyberattacks emphasizes the need for robust cybersecurity measures	Does not propose a specific IDS solution mainly descriptive, lacking actionable recommendations
[8]	PLNet approach using deep transfer learning for in-vehicle network intrusion detection	High F1 Score (above 97%) on Car-Hacking dataset effective use of transfer learning for improved performance	Focused solely on in-vehicle networks may not address external threats comprehensively
[12]	1D-CNN-based IDS for intra-vehicle intrusion detection	Effective for time-series data analytics high accuracy in identifying IVN attacks	Limited to intra-vehicle network protection may not scale well to larger datasets or external threats
[13]	Deep CNN (DCNN) using reduced InceptionResnet for IVN attack detection	High accuracy with Car-Hacking dataset advanced model architecture for improved detection	Potentially high computational requirements focused on IVN, not addressing external threats
[14]	DeepSecDrive: deep learning architecture for early IVN attack detection	Lightweight, effective, and interpretable outperforms state-of-the-art detection approaches	Limited to in-vehicle networks may require specialized knowledge for implementation

Table 2. Confusion matrix (CM) for attack detection.

	Predicted Normal	Predicted Attack
Actual Normal	True Positive (TP)	False Negative (FN)
Actual Attack	False Positive (FP)	True Negative (TN)

determine their effectiveness in cyber-attack detection.

3.3 The third phase includes the detection and classification of vehicle cyberattacks

The Phase 3 utilizes trained deep learning models consisting of RNN, LSTM, and GRU, which evaluate real-world vehicular data within the testing dataset. The classification system determines every input sequence as either Attack (malicious) or Normal (benign) [20]. Through their effective handling of temporal sequence data, the detection system obtains a superior ability to pinpoint delicate anomalies. Standard classification metrics evaluate the model's performance by testing its ability to separate normal from intrusive vehicular network behaviors through accuracy, precision, recall, and F1-score assessment. The final proposed IDS framework features its detailed pseudocode as Algorithm 1, which organizes the process into three sequential steps starting from data preparation and continuing to model training and testing, then concluding with a performance evaluation.

Algorithm 1: Car Hacking Detection Framework using Deep Learning

```
Input  : Raw car hacking dataset $D$ with feature
         set $X$ and target labels $Y$
Output : Classification results and evaluation
         metrics
Step 1: Load required libraries (pandas,
  scikit-learn, tensorflow);
Step 2: Load dataset $D$ and split into features $X$
  and labels $Y$;
Step 3: Apply balancing techniques (NearMiss,
  Random OverSampling (ROS), Tomek Links) to
  address class imbalance;
Step 4: Partition dataset into training and testing
  sets;
Step 5: Encode categorical labels and reshape
  input data for time-series format if needed;
Step 6: Build the selected deep learning model
  (e.g., LSTM, GRU, 1D-CNN);
Step 7: Train the model with appropriate
  optimizers and callbacks (EarlyStopping,
  ReduceLROnPlateau, ModelCheckpoint);
Step 8: Evaluate model performance using
  Accuracy, Precision, Recall, F1-score, and
  Confusion Matrix;
return Classification output and evaluation
  metrics.
```

The initial processing of raw intrusion data includes the application of resampling techniques for handling the dataset's class imbalance problem. The implementation of balancing methods is essential for boosting model performance during learning operations. The data moves into a Data-Frame after processing for additional data manipulation and evaluation purposes. The data becomes usable after balancing techniques are applied for model training purposes. Evaluation metrics enable the assessment of trained models after the training process completes [22]. The complete workflow, starting from data processing involving preprocessing and training and ending with evaluation, enables the IDS to receive balanced input while conducting an extensive evaluation for vehicle intrusion threat detection.

4 Proposed High-Level IDS-IoV Framework for Predicting Vehicle Cyberattacks

To evaluate the system's applicability in real-time environments, we measured inference latency on test samples. In average, the GPU model consisted of GRU with 12 ms per prediction, whereas standard CPU environments gave 42 ms per prediction. These findings show that it is possible to implement it on edge devices allowing real-time detection. But the compute resources of real vehicle ECUs are limited, and model deployment on them would need either model pruning or quantization, both in the realm of future work.

The proposed framework for vehicle cyberattacks prediction integrates resampling techniques with deep learning algorithms to effectively simulate and detect potential intrusions in real-time IoV environments. This hybrid approach leverages the strengths of both Resampling Methods (RM) and deep Learning (DL) to enhance the robustness and accuracy of intrusion detection systems [20, 21]. Designed to assist automotive cybersecurity professionals, the framework enables proactive identification of cyber threats in realistic vehicular settings. By utilizing the Internet of Vehicles (IoV), cloud infrastructure, and intelligent learning models, it ensures continuous monitoring and defense against malicious activities across connected vehicles. The overall goal is to provide a scalable, intelligent security mechanism that adapts to evolving threats in real-time. The framework is structured into three essential stages, each contributing a specific function toward achieving an end-to-end detection and response system. These phases operate collaboratively to meet the objectives of predictive analytics and threat mitigation. The three stages, illustrated in Figure 6, are described below:

- **Stage 1**: Vehicle subsystems, including potential attack points like Electronic Control Units (ECUs), generate data that is transmitted to a cloud-based system for collection and preliminary processing.

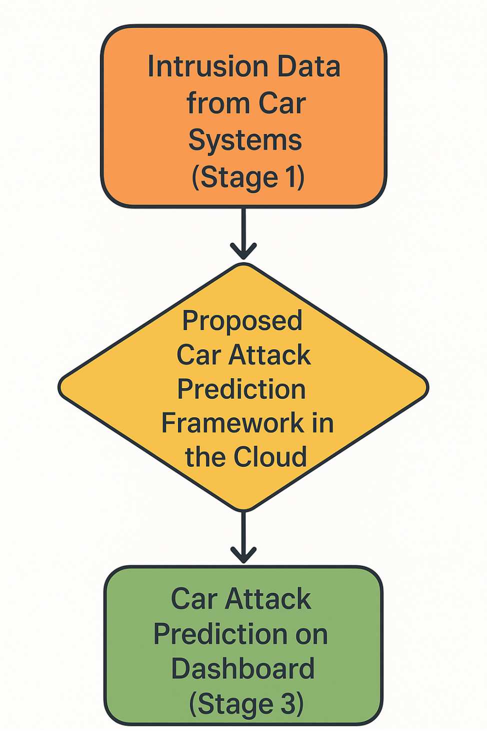
Intrusion Data from Car Systems (Stage 1)
Proposed Car Attack Prediction Framework in the Cloud
Car Attack Prediction on Dashboard (Stage 3)

Figure 6. Stages of suggested car hacking detection over the cloud system.

- **Stage 2:** Once data is collected, it is processed and stored in the cloud. At this stage, the data is structured and analyzed for indicators of suspicious or malicious activity. The system applies deep learning algorithms to forecast potential attack vectors, facilitating the timely identification of security breaches.

- **Stage 3**: A cloud-based dashboard is utilized by security analysts to monitor, visualize, and assess threats within a smart city context. The AI-driven system provides actionable insights, empowering engineers to respond to real-time vehicle threats with informed decision-making and mitigation strategies. A method based on cloud computing power and artificial intelligence functions as a system to detect and forecast vehicle attacks within smart cities. Security experts receive real-time instances from the system, which analyzes data from different sources through cloud processing, while AI technology detects patterns and potential threats. The smart city obtains improved overall security through quick and effective responses to detected potential threats.

5 Experiments and Results Analysis

The evaluation results of deep learning methods when used on car hacking data. For each type of car attack. This work has set its aim towards investigating deep learning's potential as a discriminator against malicious car attacks. The assessment of feasibility exists through the combination of robust performance as well as correct identification accuracy. The system shows significant ability to identify legitimate and deceptive information with limited incorrect classifications. We calculate precision and recall, and F1_Score for each class because the large number of normal cases should not affect our final performance evaluation.

To evaluate the automobile hacking data, the researchers partitioned it into a training segment containing 70% of the data and a testing segment consisting of 30%. Standardization procedures occurred before pre-processing was applied to all training and test datasets. Multiple DL algorithms like RNN, LSTM, and GRU received training using a randomly re-principal component analysis version of the training data. The DL models received their evaluation from the test dataset. The research used NearMiss, ROS, and TomLinks technique combinations as the resampling methods to evaluate each dataset.

While the current system performs well on known attack types in the dataset, evolving cyberattacks can exploit unknown vulnerabilities. To address this, retraining the models periodically with new data and incorporating continual learning techniques may help the system adapt to novel threats. Future research will explore adversarial training and ensemble-based uncertainty detection to enhance generalizability.

5.1 Experiments

5.1.1 *Experimental Setup*

The experiments were conducted using Google Colab Pro with access to an NVIDIA Tesla T4 GPU (16GB VRAM), 25GB RAM, and Python 3.10. TensorFlow and Keras utilized models of operation. Pandas and Scikit-learn were used to preprocess Car-Hacking dataset. Every model was trained on 50 epochs and a batch size of 64. Its learning rate consisted of 0.001 and Adam optimizer. It used early stopping to avoid overfitting. The number fields were standardized. The given dataset was divided into a 70:30 train-test split. The present setup did not involve cross-validation yet this is an area that will be looked at in future in order to provide more reliable measure of performance.

Analysis of deep learning (DL) models for the Car-Hacking dataset exists in Table 3 and Figure 7 without incorporating any resampling methods.

Table 3. DL only for the Car-Hacking dataset without resampling techniques.

Algorithm	Accuracy	Precision	Recall	F1Score
RNN	0.94	0.93	0.91	0.92
LSTM	0.98	0.97	0.96	0.96
GRU	0.99	0.98	0.97	0.98

Table 4, along with Figure 7, demonstrates a direct comparison between results gained through unaltered data without resampling for both R and T instances. Initial evaluations act as baseline measures to assess the deep learning model's raw performance before any data balancing process is applied. The implementation of the NearMiss resampling method (see Tables 5 and 7) together with Random Over-Sampling (ROS) (see Table 8) and Tomek Links resampling methods are shown in Tables 6 and 9 with corresponding Figures 8, 9, and 10. The framework combines deep learning models with resampling techniques to achieve enhanced accuracy results for all detection metrics.

Table 4. NearMiss-DL for the Car-Hacking dataset.

Algorithm	Accuracy	Precision	Recall	F1Score
RNN	0.94	0.93	0.91	0.92

Table 5. NearMiss-DL for the Car-Hacking dataset.

Algorithm	Accuracy	Precision	Recall	F1Score
LSTM	0.98	0.97	0.96	0.96

Table 6. TomekLinks-DL for the Car-Hacking dataset.

Algorithm	Accuracy	Precision	Recall	F1Score
GRU	0.99	0.98	0.97	0.98

Table 7. NearMiss-DL for the Car-Hacking dataset for Normal (R) and Attack (T).

Algorithm	Class	Precision	Recall	F1 Score
RNN	R	0.93	0.91	0.92
	T	0.93	0.91	0.92
LSTM	R	0.97	0.96	0.96
	T	0.97	0.96	0.96
GRU	R	0.98	0.97	0.98
	T	0.98	0.97	0.98

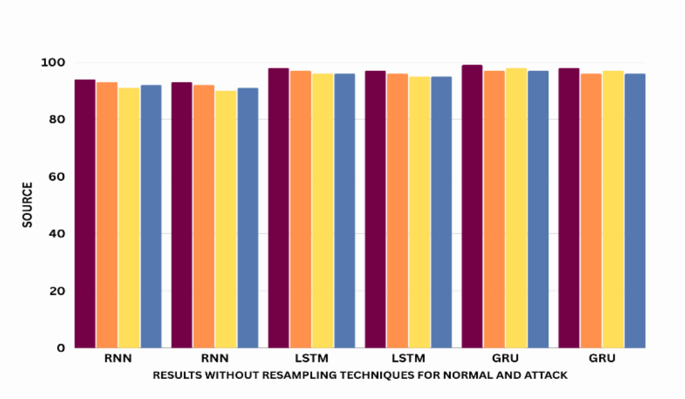

Figure 7. Results without resampling techniques for Normal (R) and Attack (T).

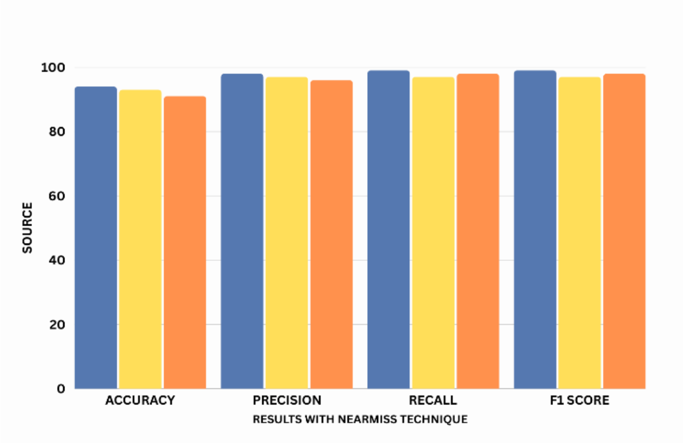

Figure 8. Results with the NearMiss technique.

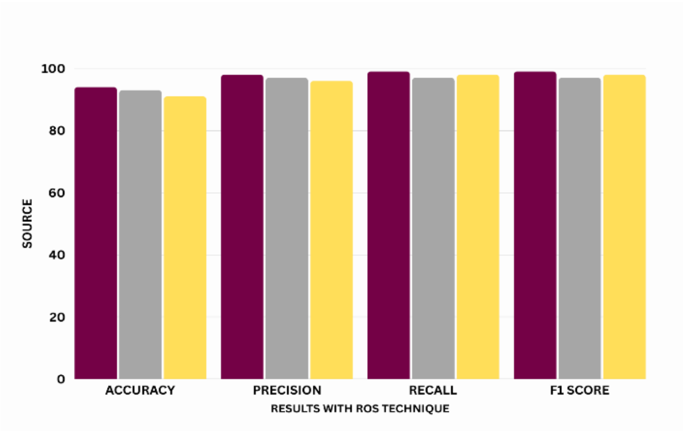

Figure 9. Results with the ROS technique.

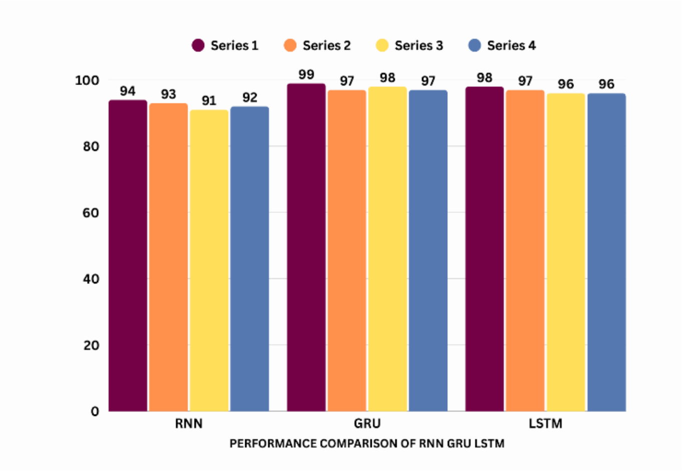

Figure 10. Performance omparisons.

Table 8. ROS-DL for the Car-Hacking dataset for Normal (R) and Attack (T).

Algorithm	Class	Precision	Recall	F1 Score
RNN	R	0.92	0.89	0.9
	T	0.94	0.93	0.93
LSTM	R	0.96	0.94	0.95
	T	0.98	0.97	0.97
GRU	R	0.97	0.95	0.96
	T	0.99	0.98	0.98

Table 9. TomekLinks-DL for the Car-Hacking dataset for Normal (R) and Attack (T).

Algorithm	Class	Precision	Recall	F1 Score
RNN	R	0.95	0.96	0.95
	T	0.91	0.86	0.88
LSTM	R	0.98	0.98	0.98
	T	0.96	0.94	0.95
GRU	R	0.99	0.99	0.99
	T	0.97	0.96	0.96

To validate the superiority of deep learning methods, we also evaluated traditional machine learning algorithms such as Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR) on the same dataset. The accuracy offered by these models was also low, as RF showed 93%, SVM 89%, and LR 85%. Whereas conventional classifiers were quicker and explicable, they could not exploit the use of sequential temporal dependencies that existed in CAN data, which were effectively used by recurrent models such as GRU. Such a comparison determines the need for deep learning in sequence-based intrusion detection that is time sensitive.

5.2 Comparative Results

As is evident in Table 10, the quantities given in the comparison demonstrate that the proposed system outperforms the current systems. The suggested method uses more features to a greater degree, significantly outperforming the previous systems in terms of accuracy, precision, recall, and F1_Score, and achieves an astounding 100% performance. The proposed system is evaluated from the results, and it outperforms the state-of-the-art methods in the accuracy, precision, recall, and F1_Score of 100%. In comparison, the system presented in [8] provides an accuracy of 98.10%, i.e., 1.9% less than ours.

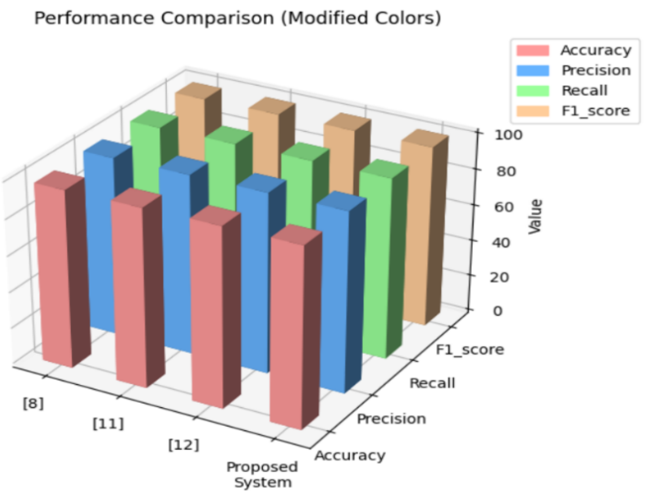

Figure 11. Comparison between the proposed system and the current work.

In the same vein, the systems presented in [11, 12] achieve 99.96% and 99.93% accuracy, respectively, and are within 0.04% and 0.07% lower than our systems. The comparative performance trends are also illustrated in Figure 11, which visually highlights the superiority of the proposed method over existing works. Finally, these results demonstrate that our proposed system is more efficient in attaining high accuracy and reliability [24].

Table 10. A comparative study of the proposed system with existing work.

Work	Accuracy%	Precision%	Recall%	F1 Score%
[8]	98.1	98.14	98.04	97.83
[11]	99.96	99.94	99.63	99.8
[12]	99.93	99.84	99.84	99.91
RNN	94	93	91	92
LSTM	98	97	96	96
GRU	99	98	97	98

6 Conclusion and Future Directions

Future work will extend the proposed IDS by integrating Federated Learning (FL) to ensure data privacy and distributed model updates without centralized data sharing. We will also explore the inclusion of Large Language Models (LLMs) and anomaly explanation frameworks to enhance interpretability. Furthermore, incorporating edge computing and lightweight model variants will improve system response time for embedded automotive environments. Addressing emerging challenges such as model poisoning in FL and adversarial evasion will be a key focus of our multi-layered security roadmap.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Khalil, R. A., Saeed, N., Masood, M., Fard, Y. M., Alouini, M. S., & Al-Naffouri, T. Y. (2021). Deep learning in the industrial internet of things: Potentials, challenges, and emerging applications. *IEEE Internet of Things Journal*, *8*(14), 11016-11040. [Crossref]

[2] Haghighat, A. K., Ravichandra-Mouli, V., Chakraborty, P., Esfandiari, Y., Arabi, S., & Sharma, A. (2020). Applications of deep learning in intelligent transportation systems. *Journal of Big Data Analytics in Transportation*, *2*(2), 115-145. [Crossref]

[3] Birchler, C., Khatiri, S., Bosshard, B., Gambi, A., & Panichella, S. (2023). Machine learning-based test selection for simulation-based testing of self-driving cars software. *Empirical Software Engineering*, *28*(3), 71. [Crossref]

[4] Zayed, S. M., Attiya, G., El-Sayed, A., Sayed, A., & Hemdan, E. E. D. (2023). An efficient fault diagnosis framework for digital twins using optimized machine learning models in smart industrial control systems. *International Journal of Computational Intelligence Systems*, *16*(1), 69. [Crossref]

[5] Raimundo, R. J., & Rosário, A. T. (2022). Cybersecurity in the internet of things in industrial management. *Applied Sciences*, *12*(3), 1598. [Crossref]

[6] Awotunde, J. B., Folorunso, S. O., Imoize, A. L., Odunuga, J. O., Lee, C. C., Li, C. T., & Do, D. T. (2023). An ensemble tree-based model for intrusion detection in industrial internet of things networks. *Applied Sciences*, *13*(4), 2479. [Crossref]

[7] Choudhary, V., Tanwar, S., & Choudhury, T. (2024). Evaluation of contemporary intrusion detection systems for internet of things environment. *Multimedia Tools and Applications*, *83*(3), 7541-7581. [Crossref]

[8] Mehedi, S. T., Anwar, A., Rahman, Z., & Ahmed, K. (2021). Deep transfer learning based intrusion detection system for electric vehicular networks. *Sensors*, *21*(14), 4736. [Crossref]

[9] Yang, L., Moubayed, A., & Shami, A. (2021). MTH-IDS: A multitiered hybrid intrusion detection system for internet of vehicles. *IEEE Internet of Things Journal, 9*(1), 616-632. [Crossref]

[10] Yang, L., & Shami, A. (2022, May). A transfer learning and optimized CNN based intrusion detection system for Internet of Vehicles. In *ICC 2022-IEEE International Conference on Communications* (pp. 2774-2779). IEEE. [Crossref]

[11] Alshathri, S., Sayed, A., & Hemdan, E. E. D. (2024). An intelligent attack detection framework for the internet of autonomous vehicles with imbalanced car hacking data. *World Electric Vehicle Journal, 15*(8), 356. [Crossref]

[12] Hossain, M. D., Inoue, H., Ochiai, H., Fall, D., & Kadobayashi, Y. (2020, December). An effective in-vehicle CAN bus intrusion detection system using CNN deep learning approach. In *GLOBECOM 2020-2020 IEEE global communications conference* (pp. 1-6). IEEE. [Crossref]

[13] Song, H. M., Woo, J., & Kim, H. K. (2020). In-vehicle network intrusion detection using deep convolutional neural network. *Vehicular Communications, 21*, 100198. [Crossref]

[14] Ding, W., Alrashdi, I., Hawash, H., & Abdel-Basset, M. (2024). DeepSecDrive: An explainable deep learning framework for real-time detection of cyberattack in in-vehicle networks. *Information Sciences, 658*, 120057. [Crossref]

[15] Khan, I. A., Moustafa, N., Pi, D., Haider, W., Li, B., & Jolfaei, A. (2021). An enhanced multi-stage deep learning framework for detecting malicious activities from autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems, 23*(12), 25469-25478. [Crossref]

[16] Ashraf, J., Bakhshi, A. D., Moustafa, N., Khurshid, H., Javed, A., & Beheshti, A. (2020). Novel deep learning-enabled LSTM autoencoder architecture for discovering anomalous events from intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems, 22*(7), 4507-4518. [Crossref]

[17] Kim, S. K., & Jang, E. S. (2022). The intelligent blockchain for the protection of smart automobile hacking. *Journal of Multimedia Information System, 9*(1), 33-42. [Crossref]

[18] Hassan, H. A., Hemdan, E. E. D., El-Shafai, W., Shokair, M., & Abd El-Samie, F. E. (2024). Detection of attacks on software defined networks using machine learning techniques and imbalanced data handling methods. *Security and Privacy, 7*(2), e350. [Crossref]

[19] Nabil, N., Najib, N., & Abdellah, J. (2024). Leveraging artificial neural networks and LightGBM for enhanced intrusion detection in automotive systems. *Arabian Journal for Science and Engineering, 49*(9), 12579-12587. [Crossref]

[20] Trovão, J. P. (2024). Advancing Automotive Technologies [Automotive Electronics]. *IEEE Vehicular Technology Magazine, 19*(1), 106-C3. [Crossref]

[21] Sajid, M., Malik, K. R., Almogren, A., Malik, T. S., Khan, A. H., Tanveer, J., & Rehman, A. U. (2024). Enhancing intrusion detection: a hybrid machine and deep learning approach. *Journal of Cloud Computing, 13*(1), 123. [Crossref]

[22] Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific reports, 12*(1), 5979. [Crossref]

[23] Liang, J., Li, Y., Yin, G., Xu, L., Lu, Y., Feng, J., ... & Cai, G. (2022). A MAS-based hierarchical architecture for the cooperation control of connected and automated vehicles. *IEEE Transactions on Vehicular Technology, 72*(2), 1559-1573. [Crossref]

[24] Zhang, X., Wang, H., Wu, B., Zhou, Q., & Hu, Y. (2023). A novel data-driven method based on sample reliability assessment and improved CNN for machinery fault diagnosis with non-ideal data. *Journal of Intelligent Manufacturing, 34*(5), 2449-2462. [Crossref]

[25] Liang, J., Tian, Q., Feng, J., Pi, D., & Yin, G. (2023). A polytopic model-based robust predictive control scheme for path tracking of autonomous vehicles. *IEEE Transactions on Intelligent Vehicles, 9*(2), 3928-3939. [Crossref]

[26] Alshathri, S., Hemdan, E. E. D., El-Shafai, W., & Sayed, A. (2023). Digital twin-based automated fault diagnosis in industrial IoT applications. *Computers, Materials & Continua, 75*(1), 183-196. [Crossref]

Muhammad Hammad Nawaz is a Data Science student in his final semester at the University of Mianwali. With three publications on Google Scholar, his research focuses on key areas of artificial intelligence. These include machine learning, deep learning, Large Language Models (LLMs), and Generative AI. His early research output highlights a strong commitment to advancing these fields. As he completes his bachelor's degree, Muhammad Hammad Nawaz is well-positioned to contribute meaningfully to the future of AI. (Email: hammadumw@gmail.com)

Abrar Ahsan is a final-year Data Science student currently enrolled in the 8th semester at the University of Mianwali. With a keen interest in cutting-edge technologies, his research areas span across Natural Language Processing (NLP), Deep Learning, Cybersecurity, and Blockchain. He is passionate about leveraging advanced computational methods to solve real-world problems and is actively involved in exploring innovative solutions within the realm of data science and emerging technologies. (Email: abrarahsan207221@gmail.com)

Inam Ullah Khan is currently pursuing a Ph.D. in Computer Science at Qurtuba University of Science and Information Technology, Peshawar, Pakistan. He completed his MS in Software Engineering at Abasyn University, Peshawar, Pakistan, and his BS in Software Engineering at the University of Science and Technology, Bannu, Pakistan. His research interests include Cybersecurity, Android Security, Machine Learning, Deep Learning, and IoT. (Email: inam1software@gmail.com)

Yanan Wang received my Bachelor's degree in Software Engineering from Harbin University, China, and am currently pursuing a Master's degree in Computer Science and Engineering at Sejong University, South Korea. My research interests include computer vision, deep learning, and machine learning. (Email: wangwynwxn@outlook.com)

Mushtaq Ahmad is currently pursuing an MS in Informatics Data Science at the Department of Business Informatics, Technical University of Vienna (TUWIEN), Austria. He completed his BS degree at Islamia College, Peshawar. His research focuses on Data Mining, NLP, Machine Learning, and Cybersecurity. (Email: mushtaq8653@gmail.com)

Muhammad Shoaib Akhtar received the B.Sc degree in electrical engineering (Telecommunication) from the University of Science and Technology Bannu, Pakistan, in 2020 and MS in electrical engineering from the same university in 2024. His area of interest includes machine learning, image processing, Data mining, Classification, Biomedical image analysis, deep neural networks, and AI health applications. (Email: engrshoaibkhanmrt@gmail)