RESEARCH ARTICLE

Intelligent Cyber-Attack Detection for Autonomous Vehicles Using Advanced Deep Learning Models

Fida Muhammad Khan[1], **Zeeshan Ali Haider**[1,*], **Muhammad Owais Khan**[2], **Ashraf Ullah**[2], **Muhammad Shoaib Khan**[3], **Muhammad Fayaz**[4], **Mushtaq Ahmad**[5] and **Nabila**[6]

[1] Department of Computer Science, Qurtuba University of Science & Information Technology, Peshawar 25000, Pakistan
[2] Department of Computer Science, University of Science & Technology, Bannu, Pakistan
[3] School of Computer and Communication Engineering, University of Science and Technology Beijing (USTB), China
[4] Department of Computer Science and Engineering, Sejong University, Seoul 05006, Republic of Korea
[5] Department of Business Informatics, Technical University of Vienna, Vienna 1040, Austria
[6] Bangladesh University of Professionals, Dhaka, Bangladesh

Abstract

The Internet of Vehicles (IoV) has greatly influenced transportation by allowing autonomous vehicles to interact and communicate with other cars as well as with the surrounding traffic system. Even so, being interconnected comes with risks in terms of cyber attacks, for example, by injecting messages or fooling sensors through CAN systems. The study, consequently, suggests an Intrusion Detection System (IDS) that uses Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Bidirectional Encoder Representations from Transformers (BERT), and RoBERTa, to properly detect and handle these cyber threats. To solve the problem of unbalanced data, we use Random Over-Sampling (ROS), NearMiss, and Tomek Links, which makes the model work more effectively. Results from experiments show that the IDS suggested here is better than traditional machine learning models. Adopting cloud computing helps make the system more flexible and lets it be watched continually to keep AVs well protected. Experimental results show that RoBERTa achieves the highest performance with 99% accuracy and 97% F1-score across all resampling conditions, outperforming conventional machine learning baselines. An IDS in smart cities greatly benefits vehicle networks, as it hinders and stops possible cyberattacks. Future work will aim to improve IDS performance at scale and address emerging attack types.

Keywords: autonomous vehicles (AVs), cyber-attack detection, internet of vehicles (IoV), bidirectional encoder representations from transformers (BERT), cybersecurity, smart cities.

Submitted: 03 May 2025
Revised: 16 June 2025
Accepted: 09 July 2025
Published: 29 August 2025

Vol. 1, **No.** 3, 2025.
10.62762/TACS.2025.952297

***Corresponding author:**
✉ Zeeshan Ali Haider
Zeeshan.ali9049@gmail.com

Citation
Khan, F. M., Haider, Z. A., Khan, M. O., Ullah, A., Khan, M. S., Fayaz, M., Ahmad, M., & Nabila (2025). Intelligent Cyber-Attack Detection for Autonomous Vehicles Using Advanced Deep Learning Models. *ICCK Transactions on Advanced Computing and Systems*, 1(3), 180–192.

© 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

Recently, deep learning models have been widely used in most domains like healthcare, finance, transportation, and industry. Deep learning has emerged as a foundational technology for securing intelligent cyber-physical transportation systems, enabling anomaly detection and intrusion identification in vehicular communication networks through architectures such as LSTM and hybrid CNN-LSTM models [1]. Cybersecurity in IoT and Industrial IoT systems has been widely studied, with growing attention to the vulnerabilities and risk management challenges that arise from the increasing connectivity of industrial devices [2]. Various types of deep learning architectures are being explored to detect cyber threats in different environments, like software-defined networks (SDN), IoT, and cloud computing [3]. Cybersecurity has emerged as a critical area of concern, especially in newer communication systems, as interconnected systems become even more prevalent [4].

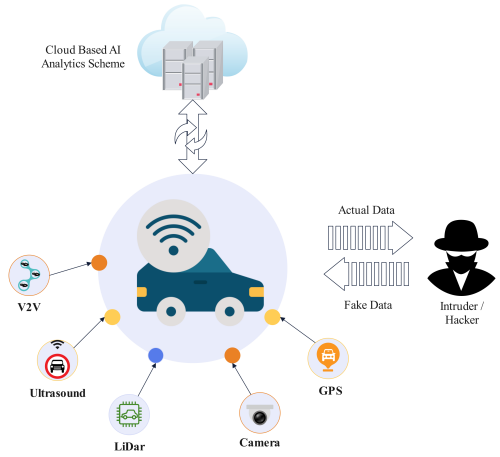


Figure 1. Cloud based AI analytics scheme for car hacking detection.

The growth of autonomous vehicles has made transportation security a critical concern. AVs form the Internet of Vehicles (IoV) through IoT connectivity, enabling real-time communication via Controller Area Networks (CAN), while simultaneously expanding the attack surface to threats such as denial-of-service, spoofing, and message injection [5, 6]. Autonomous vehicles can navigate without being driven by using radar, cameras, satellite navigation, and AI, and they communicate with outside platforms through wireless telecoms, for example, 4G and 5G [7]. However, AVs are still at risk from cyber threats such as vulnerable software, sensor tampering, and various system compromises. This is the reason cybersecurity experts are making more effort to protect AVs from such threats [8]. The use of deep learning models, including CNN, RNN, LSTM, BERT, and RoBERTa, is rising in threat detection in the automotive sector. Explainable deep learning has been applied to detect malware attacks in IoT-enabled intelligent transportation systems, offering both high detection accuracy and interpretability for security decision-making [9]. Building on this direction, cloud-based AI analytics have been proposed to enable real-time cyber-attack detection for autonomous vehicles, as illustrated in Figure 1. In such architectures, sensor data streams from sources including V2V communication, ultrasound, LiDAR, cameras, and GPS are transmitted to cloud servers for centralized analysis, where deep learning models determine whether the incoming data originates from an authorised source or constitutes an injected attack [7, 8]. Improved security for autonomous vehicles is therefore sought by integrating AI with scalable cloud computing infrastructure.

AVs interact in real time with other vehicles and infrastructure through V2V and V2I communication, exposing them to threats including message injection, sensor spoofing, and DoS attacks [10, 11]. Despite progress in IDS research, existing tools struggle with real-time detection and class-imbalanced datasets [12], motivating the deep learning-based IDS proposed in this work.

The fact that there are many more samples for the attacker than the defender leads to sluggish detection of actual attacks. Because the number of normal data points is much greater than that of attack data, traditional machine learning models have difficulties. In contrast, with models like CNN, RNN, and LSTM, programs have more accurate and consistent solutions by learning how to recognize and handle advanced threats. BERT and RoBERTa, which are Transformer-based models, make detection more accurate by capturing deep bidirectional contextual representations within sequential CAN bus traffic data. An approach to attack detection in Internet of Vehicles (IoV) security is described here, using CNN, RNN, LSTM, BERT, and RoBERTa to address unbalanced data and provide strong real-time cyber defense for both internal and external vehicle infrastructure against complicated attacks. The key contributions of this work are summarized as follows:

- We develop a smart cyberattack detection

framework utilizing deep learning models (CNN, RNN, LSTM, BERT, RoBERTa) for real-time intrusion detection in autonomous vehicle systems.

- We apply a comprehensive set of resampling strategies, including Random Over-Sampling, NearMiss, and Tomek Links, to systematically address the class imbalance inherent in car-hacking datasets and improve minority-class detection.
- To conduct performance analysis of the proposed architecture on the Car-Hacking benchmark dataset to demonstrate that deep learning models, particularly Transformer-based architectures, outperform classical machine learning approaches in identifying cyber-attacks in IoV frameworks.

This paper is structured as follows: Section 2 outlines the related work referenced in this study. Section 3 details the proposed deep learning-based high-level framework for predicting car attacks within the IoVs system. Section 4 covers the dataset, evaluation metrics, and the analysis of the results. Finally, Section 5 concludes the paper by summarizing the findings and suggesting potential areas for future work.

2 Related work

With the advent of autonomous vehicles incorporating the latest technology and communication protocols, securing vehicles is critical. Due to the rapid development of the Internet of Vehicles (IoV) in smart cities, vehicle systems are connected internally and externally. As these networks become central to enabling the use of vehicles, safeguarding them from cyberattacks is crucial for their safe and successful functioning [13]. The data-intensive and highly connected nature of these systems makes them attractive targets for cybercriminals, driving growing demand for intelligent IDS capable of detecting diverse attack types [14]. Internally, AVs rely on the Controller Area Network (CAN) bus for ECU communication, while externally they interact with roadside infrastructure and other vehicles through IoV protocols.

Advances in deep learning, especially with transformer-based architectures, have helped improve the ability to detect attacks in autonomous vehicle networks [15]. Beyond deep learning, adaptive intelligent methods such as neuro-fuzzy inference systems have also demonstrated effectiveness in detecting CAN bus cyber attacks in electric vehicles, highlighting the diversity of approaches being explored for vehicular intrusion detection [16]. Collectively, these detection approaches—spanning deep learning and adaptive intelligent methods—form the foundation of IDS solutions set up to identify distinct types of attacks in the Internet of Vehicles (IoV) system: Denial-of-Service conditions are flagged through abnormal message frequency patterns, spoofing inconsistencies are exposed by cross-referencing readings from multiple sensors such as GPS and vehicle speed sensors, and message injection is identified by analysing suspicious CAN IDs and irregular message intervals. Nevertheless, IDS solutions themselves remain vulnerable to adversarial attacks, where carefully crafted perturbations can cause misclassification of malicious traffic, posing a significant challenge to the robustness of intrusion detection in connected and autonomous vehicle networks [17]. To establish a performance reference for evaluating more advanced architectures, machine learning techniques have been applied to detect cyber attacks in Internet of Vehicles environments, providing a baseline against which deep learning models can be benchmarked [18]. Smart city environments present a unique challenge, as autonomous vehicles are exposed to a broad range of malware attacks that can compromise vehicle autonomy, increase communication latency, and threaten passenger safety, making real-time cybersecurity detection critical [19].

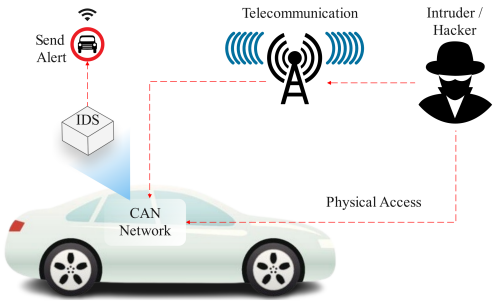


Figure 2. Intrusion detection system (IDS) for car hacking.

CAN bus helps ECUs in modern self-driving vehicles exchange information, and outside communication is necessary for interaction with road infrastructure and other vehicles on the Internet of Vehicles [20]. However, such networks are at risk of different cyberattacks like message injection, sensor spoofing, and remote exploitation [21]. As a result, using reliable Intrusion Detection System (IDS) models to prevent and respond to threats from inside and

outside the system is necessary, as deep learning-based cyber defence solutions that incorporate structured threat models, data integrity mechanisms, and model robustness against adversarial manipulation represent an important research direction for securing such systems [22]. Through the Vehicle CAN IDS architecture, displayed in Figure 2, the system is informed when an intruder is found, pointing out the possible risk from the CAN network to the telecommunication systems.

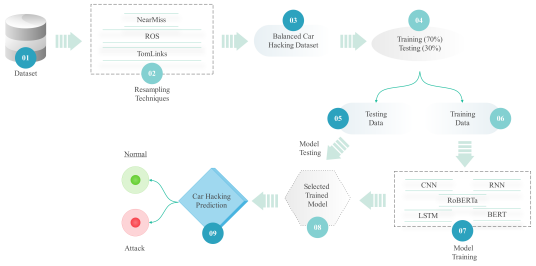


Figure 3. Workflow of the car hacking detection framework.

The insufficiency of traditional ML methods has been seen when trying to prevent new cyberattacks on autonomous vehicles, motivating the development of intelligent IDS solutions that combine multiple deep learning architectures and resampling strategies to handle evolving threats in the IoV environment [23]. Complementary approaches such as blockchain combined with artificial intelligence have also been explored to provide decentralized and tamper-resistant security guarantees for the IoV environment [24]. Earlier approaches to automated attack diagnosis in IoT systems relied on machine learning combined with frequency distribution analysis, demonstrating foundational detection capability but leaving room for improvement in capturing complex temporal and contextual patterns [25]. This motivates the adoption of advanced architectures such as CNNs, RNNs, LSTMs, BERT, and RoBERTa, which are better suited to spotting hard-to-find patterns and detecting attacks in real time. A deep learning-focused Intrusion Detection System is suggested in the paper to detect and foresee cyberattacks in autonomous vehicles. It helps solve the security challenges in real-time traffic using deep models. Comparing the models to a benchmark dataset proved that deep learning helps vehicles become more secure, efficiently discovering known and unknown attack types.

Apart from using large deep learning models, approaches like oversampling, undersampling, and hybrid resampling are added to tackle the imbalance in the dataset, stopping the IDS from strictly following the dominant class.

3 Methodology

In this paper, we introduce a completely new approach for the automatic detection of cyberattacks against autonomous vehicle systems. As shown in Figure 3.

BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa (Robustly Optimized BERT Pretraining Approach) were specifically designed to work with sequences, so they perform well at finding attacks in vehicle network traffic. BERT uses information from both the past and future and RoBERTa tweaks BERT to give it more training data and better hyperparameter settings. Both systems are effective in spotting quiet manipulations on autonomous vehicle networks. The study investigates an attack detection system built using deep learning which was studied on the Car-Hacking dataset, comprising CAN bus traffic collected from a real vehicle via the OBD-II port under both normal driving and active attack injection conditions. For this study, a representative subset of the Car-Hacking dataset comprising 818,440 records was used, of which 701,832 (85.7%) are normal CAN messages and 116,608 (14.3%) are attack messages spanning Denial-of-Service, fuzzy, and spoofing attack types. The dataset captures real-world CAN traffic with a range of attacks, with CAN ID, timestamp, and data bytes information included, like real-life vehicle networks.

3.1 Car-Hacking Dataset

The Car-Hacking dataset, originally released by the Hacking and Countermeasure Research Lab (HCRL) at Korea University [26], has been widely used to benchmark intrusion detection systems for autonomous vehicle networks. The dataset presented targets various vehicular communication systems, including DoS attacks, fuzzy attacks, and spoofing attacks (altering key vehicle data like the drive gear or the RPM gauge). Our dataset is generated such that it reflects real-time attacks and contains extensive features for the detection of attacks. The class distribution of the subset used in this study is shown in Table 1 and far fewer attack samples, which is an important factor affecting the evaluation of cyber-attack detection models.

The Car-Hacking dataset encompasses three attack scenarios: a DoS Attack, in which the attacker injects

Table 1. Car-Hacking dataset class distribution.

Class	Instances
Normal	701,832
Attack	116,608

the CAN ID message '0000' every 0.3 milliseconds, overwhelming the network with invalid data; a Fuzzy Attack, in which randomized CAN ID and data values are injected every 0.5 ms to create inconsistencies and cause network disruptions; and a Spoofing Attack (RPM/Gear), in which specific CAN ID messages related to gear and RPM data are injected every millisecond to mislead the vehicle's control units about the vehicle's status. Each record in the dataset contains five attributes: a Timestamp (the time of recording in seconds), a CAN ID (a hexadecimal message identifier, e.g., '043f'), a DLC (the number of data bytes, ranging from 0 to 8), DATA fields (the byte values DATA0 through DATA7), and a Flag indicating whether the message is normal ('R' for real) or injected ('T' for injected). Since normal messages substantially outnumber attack messages, three complementary resampling techniques are applied to address this class imbalance: Random Over-Sampling (ROS), which inflates the minority attack class by duplicating instances to stabilize the class distribution; NearMiss undersampling, which selects majority-class instances closest to minority-class boundaries, providing balance without overfitting; and Tomek Links, a data-cleaning method that identifies and removes borderline majority-minority instance pairs whose mutual proximity causes decision boundary ambiguity, thereby improving class separability in the training data.

These preprocessing techniques are necessary for training deep learning models such as CNN, RNN, LSTM, BERT, and RoBERTa. These models can identify sophisticated configurations and indicate cyber-attacks that the traditional method may not cover, and also, the deep learning models are trained and tested upon the Car-Hacking dataset to identify attacks based on real-time conditions to minimize false positives and false negatives. Model performance is evaluated with metrics like accuracy, precision, recall, and F1-score. As a result, these measures enable an overall assessment of the IDS's efficiency in accurately detecting normal and attack data with high accuracy in order to ensure that cyber threats are successfully identified and effectively neutralized within autonomous vehicle systems.

3.2 Data Preparation and Pre-processing

In this phase, we work on overcoming the class imbalance presented by the dataset, since the number of normal instances is greatly higher than the number of attack ones. When a learning model is trained on such imbalanced data, it tends to focus on the majority class while neglecting the minority attack class, leading to poor detection of actual attacks. To address this, three complementary resampling methods are applied. Random Over-Sampling (ROS) duplicates instances from the minority attack class until its size matches that of the majority class, increasing model sensitivity to attack patterns. NearMiss undersampling selects majority-class instances that are closest to minority-class instances, reducing majority-class dominance without introducing synthetic data. Tomek Links is a data-cleaning technique that identifies and removes borderline instance pairs from opposing classes whose mutual proximity creates ambiguous decision boundaries, thereby improving the separability of the training data. By applying these three resampling strategies in combination, the dataset is balanced, providing a solid foundation for training the deep learning models evaluated in this study.

3.3 Training and Testing of Deep Learning Models

After pre-processing, the dataset is partitioned into a training set and a testing set following a 70–30 split, and the pre-processed training data is used to train five state-of-the-art deep learning architectures for attack detection. Convolutional Neural Networks (CNNs) are employed to recognize spatial patterns within CAN bus records, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are used for sequential data analysis and capturing time-dependent attack patterns. Bidirectional Encoder Representations from Transformers (BERT) mines deep bidirectional contextual information from the data, making it especially valuable for large and complex traffic sequences, and the Robustly Optimized BERT Pretraining Approach (RoBERTa) extends this capability through a larger training corpus and optimized hyperparameters to harness richer contextual representations. To apply both BERT and RoBERTa to CAN bus data, CAN IDs and data bytes are serialized into token sequences, treating each CAN message field as a discrete token, enabling the models to learn bidirectional contextual patterns across sequences of network traffic in a manner analogous to sentence-level language modeling.

Different deep learning models were trained on the balanced dataset to make them perform better. To measure how well models could detect attacks, accuracy, precision, recall and F1-score were all used. The Car-Hacking dataset was collected from a real vehicle via the OBD-II port under controlled attack injection conditions, capturing authentic CAN bus traffic across multiple attack scenarios against the in-vehicle network. It includes details from different attacks such as DoS attacks, spoofing and message injection and every data point provides you with CAN ID, timestamp, data bytes and message flags. Careful balancing was done by means of vehicle-to-vehicle (V2V) simulations and by noticing which vehicle network features such as CAN bus data and sensor data were most important.

Hyperparameters like learning rate, how many layers the model has and batch size were all optimized during the training phase. Setting the learning rate to 0.001 allowed the model to converge better. CNN only had three layers, while the LSTM and RNN models included five to help capture both types of features. A batch size of 32 helped ensure the model learned fast and was not overfit, and settings were found by trying different combinations with grid search. For BERT and RoBERTa, the base-sized pre-trained models were fine-tuned for 5 epochs with a sequence length of 128 tokens, using the AdamW optimizer consistent with standard transfer learning practice for sequence classification tasks.

3.4 Car Attack Detection and Classification

In this stage, the deep learning models are validated on the testing data. The goal is to discretely detect whether each data point belongs to a normal class (negative) or an attack class (positive). Each model is evaluated based on multiple classification metrics:

- **Accuracy:** Ratio of correct predictions (of normal and attack) to total predictions.
- **Precision:** Number of true positive attack predictions over total predicted attacks.
- **Recall:** The number of true attack predictions over the total number of attacks.
- **F1-Score:** The harmonic means of the precision and recall, offering a balanced statistic for the performance of the model.

The model is assessed in terms of its speed and accuracy, making sure the streaming traffic is correctly screened into attack and non-attack traffic groups in real time. The steps for the framework workflow are to gather data, test models, make predictions, use sampling techniques (NearMiss, ROS, TomekLinks), handle data imbalances, create a test and training set, and use different models (CNN, RNN, LSTM, BERT, RoBERTa). Jamming attacks are detected using an analysis of how messages are sent over the CAN bus, spoofing is found by looking for differences in sensor readings, and applying different sensors and other attacks are identified through a detailed analysis of system data. In this case, CNNs are used to find spatial features, and LSTM models check for suspicious changes in time, mostly with DoS attacks.

Algorithm 1: Car Hacking Detection Framework Using Deep Learning

Data: Raw car hacking dataset D with features X and labels Y
Result: Trained classifiers for car hacking detection in IoVs, with evaluation metrics
`; // Prerequisites: pandas, numpy, TensorFlow/Keras/PyTorch, imbalanced-learn`
Load dataset D into a DataFrame; separate into features X and target Y;
Apply resampling to balance D: RandomOverSampler (oversampling), NearMiss (undersampling), TomekLinks (data cleaning);
Split balanced dataset into training set D_{train} (70%) and testing set D_{test} (30%);
for *each model M in {CNN, RNN, LSTM, BERT, RoBERTa}* **do**
    Train M on D_{train} with hyperparameter optimization (grid search);
    Evaluate M on D_{test}; record accuracy, precision, recall, F1-score;
end
return *Evaluation metrics for all models M*

Algorithm 1 outlines the complete pipeline for cyber-attack detection in autonomous vehicle systems. The key design decisions include: (i) the use of three complementary resampling strategies to address class imbalance rather than relying on a single technique; (ii) the selection of five architecturally distinct deep learning models to enable systematic comparative evaluation; and (iii) the use of a 70–30 train-test split with grid search hyperparameter tuning to ensure reproducibility and resistance to overfitting.

3.5 Suggested High-Level IDS-IoV Framework for Car Attack Prediction

A multilayer framework has been developed that utilizes the power of cloud computing, deep learning, and resampling algorithms to predict the potential future attacks under Internet of Vehicles (IoV) systems. This framework proficiently solves the complexities

of vehicle attack detection and prediction owing to a mixture of Resampling Techniques (RTs) and Deep Learning (DL). This top-level infrastructure is an IDS-IoV system for smart cities that will be able to provide proactive security for autonomous vehicles dealing with real-time data and time-sensitive decisions. In particular, the methodology is designed to help vehicle security engineers with the ability to anticipate a potential attack from real-world conditions by utilizing the Internet of Vehicles (IoV), Cloud Computing, and Deep Learning for ongoing monitoring and protection of vehicle communications. The frameworks also implement AI prediction systems so that security experts can analyze the visible and invisible AI systems before responding to any security threat. The architecture consists of three fundamental phases acting in concert to deliver the system goals:

Stage 1: Data Collection and Transmission to Cloud Servers

Vehicle systems transmit behavioral and network interaction data to cloud servers, which serve as the central processing hub for real-time analytics. Deep learning models (CNN, RNN, LSTM, BERT, and RoBERTa) analyze incoming data streams—including sensor readings, CAN bus traffic, and V2V/V2I communication—to detect anomalies and predict potential cyberattacks on both internal and external vehicle networks.

Stage 2: Data Processing, Analysis, and Attack Forecasting

After data is generated, it is sent to the cloud servers for processing and storage. AI models, especially deep learning algorithms, process and categorize the data within the cloud to understand the characteristics of potential cyber attackers. Cloud computing offers the necessary computational capabilities required for processing large datasets, including attack detection as well as attack predictions against vehicles. Patterns of normal and abnormal behavior are learned, and these are used to predict potential attacks. Deep learning models (e.g. LSTM, which works on sequential data, BERT, which works on both spatial and contextual analysis, and RoBERTa, which creates abstraction levels for recognizing complex patterns of understanding communication traffic) help to create highly accurate predictions of the known and unknown threats against the systems of the vehicle. Using historical data to predict future attempts at attacks enables security experts to take action before an attack even happens.

Stage 3: Cloud-Based Monitoring, Security Decision-Making, and Threat Detection

In the last stage, a cloud-based dashboard helps and monitoring application is utilized to help vehicle security investigators recognize and identify threats across wide urban areas. It gives security engineers real-time insight gathered from AI-predictive and deep learning-based analytics. Security professionals use a dedicated dashboard to access relevant data streams from vehicles, and can detect attacks and make decisions based on the severity of the attacks. For security experts, providing an overview of vehicle network behavior and potential vulnerabilities or immediate threats will improve the overall effect of the attack detection framework at this stage. Additionally, the prediction system can be on the cloud, which can help in integrating predictive analytics to enable quick decision-making to counter vehicle network attacks, thereby keeping the vehicle and passengers safe.

3.6 Cloud Computing and AI Integration

Cloud computing and AI are introduced here to strengthen smart city plans, specifically by anticipating vehicle abuses. The system continuously gathers data from many sources and uses sophisticated models to spot unusual activities and cyber threats. Using AI on the cloud helps improve the detection of cyber security threats which helps secure autonomous vehicles in any changing situation. Through cloud computing, processors can effectively handle large sets of data, supporting capability for detecting more threats. Advances in cloud computing now make it practical to analyze a lot of data collected by AVs which helps spot cyber threats in real time. Cloud systems powered by AI are able to watch network traffic in vehicles, study ways attacks occur and warn the security team whenever problems arise. Cloud platforms work well for IDS systems in smart cities because they are scalable and flexible, handling all the data from numerous vehicles to make the IoV safer.

4 Results Analysis and Assessment

In this section, we provide the report results of the deep learning-based attack detection framework applied to the Car-Hacking dataset. The performance of the proposed deep learning models (CNN, RNN, LSTM, BERT, and RoBERTa) for detecting and classifying cyber-attacks on autonomous vehicle systems in the Internet of Vehicles (IoV) is evaluated in this study. We evaluate the performance of different models based on metrics such as accuracy, precision, recall, and

F1-score.

4.1 Experimental Setup

The Car-Hacking dataset, which exhibits a substantial class imbalance favouring normal traffic over attack records, was used to train and evaluate all deep learning models. To address this imbalance prior to training, three resampling techniques were applied: Random Over-Sampling (ROS), NearMiss, and Tomek Links. The balanced dataset was then partitioned into a training set (70%) and a testing set (30%). Five deep learning architectures were trained on this balanced data: Convolutional Neural Networks (CNN), which detect spatial features and patterns within CAN bus records; Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks, which model sequential dependencies and temporal attack trends; BERT (Bidirectional Encoder Representations from Transformers), which captures deep bidirectional contextual representations from sequential CAN bus traffic by treating each CAN message field as a discrete token processed through self-attention layers; and RoBERTa (Robustly Optimized BERT Pretraining Approach), which extends BERT with a larger training corpus and optimized hyperparameters to recognize intricate, context-dependent patterns in vehicular communication traffic.

4.2 Performance Evaluation

The performance of every model was evaluated using four classification metrics: Accuracy, defined as the ratio of correctly classified instances (both normal and attack) to the total number of instances; Precision (True Positive Rate), defined as the proportion of predicted attack instances that are true positives; Recall, defined as the ratio of true positive attack predictions to the total number of actual attack instances; and F1-Score, the harmonic mean of Precision and Recall, which provides a balanced measure of model performance particularly suited to imbalanced class distributions.

Table 2 presents the baseline performance of all five deep learning models evaluated on the test dataset without any resampling, and Figure 4 visualizes the corresponding Accuracy, Precision, Recall, and F1-score for each model to facilitate direct comparison.

It has been observed that RoBERTa outperformed all other models in accuracy (99%), precision (98%), recall (95%) and F1-score (97%). Followed closely by BERT, which achieved an accuracy of 97%, precision of 96%, and recall of 93%. Other ML models with promising results, but with lower accuracy, were LSTM, CNN, and RNN, which all showed notably high recall values—an important metric in imbalanced detection scenarios such as cyberattack identification.

Table 2. Performance of models without resampling.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	95	94	91	92
RNN	94	93	92	91
LSTM	96	94	92	94
BERT	97	96	93	95
RoBERTa	99	98	95	97

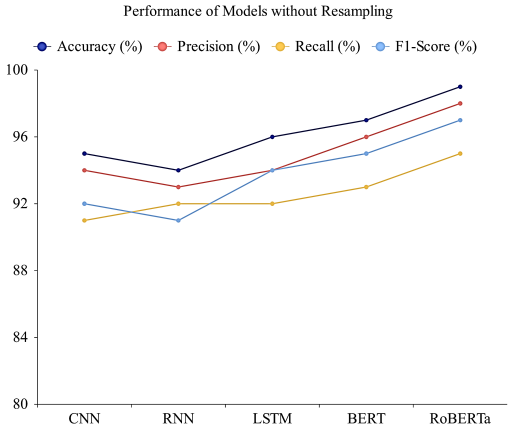


Figure 4. Performance of models without resampling.

4.3 Resampling Techniques Impact

To improve how well the models recognized attacks, we used three types of resampling methods, Random Over-Sampling (ROS), NearMiss and Tomek Links, to equalize the number of different classes in our data. With these techniques in place, the model worked much better at catching evidence of attacks. To catch less common attacks, ROS makes another instance from the same attack class. Even so, it could cause overfitting because the model will concentrate too much on repeated parts of the data. For this reason, NearMiss gets a smaller number of normal class examples and instead picks instances that are similar to the examples from the minority class. Removing close and noisy samples, Tomek Links helps raise the quality of the dataset. These techniques together allow the model to identify attacks more effectively and to not become overly specific.

4.3.1 Random Over-Sampling (ROS)

To boost the number of attack entries within the data set and to make the models used more sensitive to attack patterns, ROS was utilized. The performance for models trained with ROS is presented

in Table 3. The model performance after doing Random Over-Sampling (ROS) also has been shown in Figure 5. This illustrates the accuracy, recall, and F1-score of each model, demonstrating the consistent superiority of RoBERTa across all metrics.

This Table 3 summarizes the results of applying Random Over-Sampling (ROS) to the models. It highlights the improvement in recall and F1-score after resampling is applied.

Table 3. Comparison of model performance with different resampling techniques (random over-sampling).

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	94	92	92	93
RNN	95	93	92	93
LSTM	96	97	96	96
BERT	97	97	98	97
RoBERTa	99	98	98	97

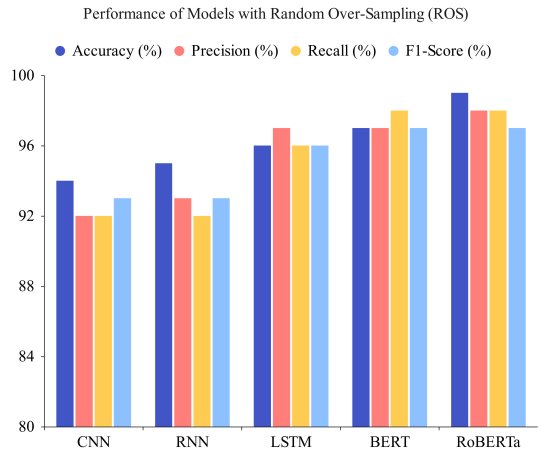


Figure 5. Performance of models with random over-sampling (ROS).

Overall, using ROS pushes RoBERTa to the top-performing position across all metrics. BERT also performs strongly under ROS, achieving recall and F1-score comparable to RoBERTa, confirming that Transformer-based architectures benefit most from over-sampling of the minority attack class. The models have been made more sensitive to detect cyber-attacks, as there is an improvement in accuracy and recall for all the models.

4.3.2 *NearMiss*

NearMiss is a method that is considered under-sampling, which removes a small number of majority examples (those samples from the normal class), in order to enhance the performance of the created model. Results for models trained with NearMiss are given in Table 4. Model performance comparison after Near Miss is illustrated in Figure 6. The graph lays down the statistics of the precision and recall and we can visually comprehend the encountered sensitivity of the attack instances of the models essentially in the case of BERT and RoBERTa.

This Table 4 presents the model performance after applying the NearMiss technique, showing the changes in precision and recall due to under-sampling.

Table 4. Comparison of model performance with NearMiss.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	94	92	88	90
RNN	93	91	85	88
LSTM	95	95	92	93
BERT	97	96	93	94
RoBERTa	98	96	95	95

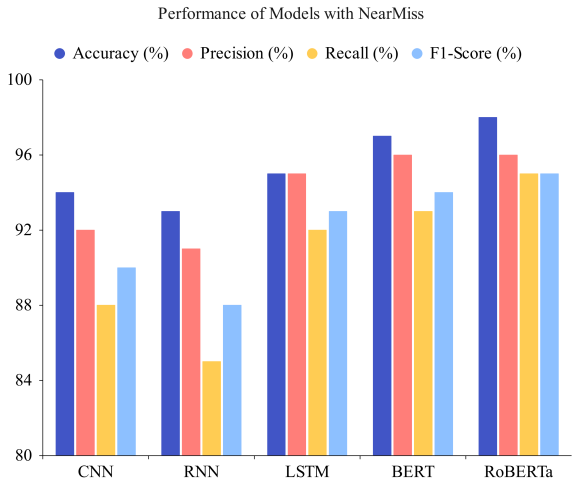


Figure 6. Performance of models with nearMiss.

While the LSTM and RoBERTa models show the most improvement in terms of recall, they are particularly better for rare attacks and outlier detection in the network. The accuracy for certain models, however, was slightly lower due to the reduced number of majority-class samples.

4.3.3 *Tomek Links*

To clean the dataset and remove overlapping instances from the majority class and from the minority class, Tomek Links was implemented. BERT and RoBERTa showed improvement in F1 scores and precision after applying this data-cleaning technique. It is noted that RoBERTa's overall accuracy marginally decreased from 99% to 98% compared to the no-resampling baseline, which is consistent with the nature of Tomek Links as a data-cleaning rather than a balancing

method; the removal of borderline instances reduces ambiguity but slightly contracts the decision boundary for the majority class. Table 5 displays the results of models trained with Tomek Links. We now use the Tomek Links data-cleaning technique, and below Figure 7 shows the performance of the models. It shows the improvements in F1-scores and precision of the cleaning process, with significant gains around RoBERTa.

This Table 5 shows the performance of models after the application of Tomek Links, highlighting improvements in F1-scores and precision.

Table 5. Comparison of model performance with Tomek Links.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	96	95	92	93
RNN	96	95	92	93
LSTM	97	96	94	94
BERT	97	97	95	95
RoBERTa	98	98	96	97

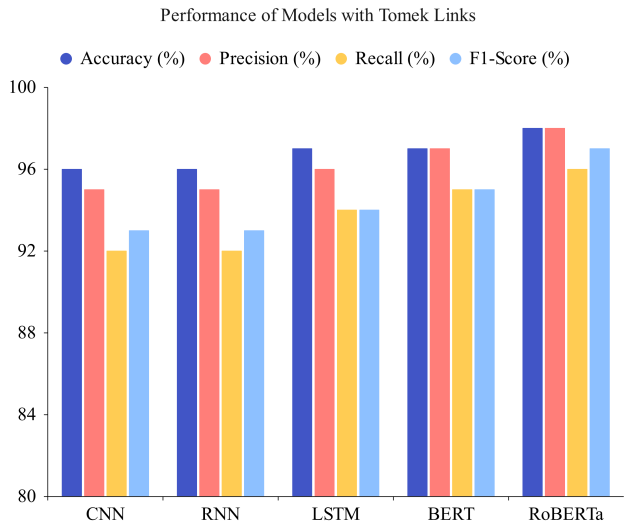


Figure 7. Performance of models with tomek links.

Both models are very suitable for the detection of such real-time attacks in the Internet of Vehicles systems, as RoBERTa was still the best-performing model and BERT also showed great progress in precision and recall.

4.4 Real-Time Attack Detection

The suggested approach with deep learning and cloud computing is able to detect any attack in real-time on autonomous vehicles. As it processes real-time data quickly, the system can spot and address cyber-attacks right away which is very important for avoiding disastrous results. Because of cloud computing, the system is better able to cope with large IoV networks. With a Denial-of-Service (DoS) attack, the IDS spots sudden changes in the number of messages and reports accordingly. When spoofing takes place and sensor readings are modified, the system checks the new values against those usual for that time of day and uses sensor fusion to see if they differ. It is also possible to see irregular movement using temporal and spatial analysis which can be linked to message injection. As a consequence, the IDS protects autonomous vehicle systems from both usual and complex attacks.

5 Discussion

5.1 Model Performance Analysis

The experimental results demonstrate that Transformer-based models, particularly RoBERTa and BERT, consistently outperform CNN, RNN, and LSTM across all resampling conditions. RoBERTa achieves the highest accuracy of 99% and F1-score of 97% without resampling, a result attributable to its robustly optimized pre-training strategy, which enables richer bidirectional contextual representations of sequential CAN bus traffic compared to recurrent architectures. BERT follows closely, benefiting from bidirectional attention over CAN ID sequences, whereas CNN, RNN, and LSTM, while effective at capturing local spatial and temporal patterns respectively, lack the long-range dependency modeling that distinguishes Transformer architectures.

The marginal performance advantage of RoBERTa over BERT can be explained by its larger training corpus, removal of the next-sentence prediction objective, and dynamic masking strategy, all of which contribute to more generalized sequence representations that transfer well to the structured token sequences derived from CAN bus data.

5.2 Comparison with Related Work

Table 6 situates the proposed framework within the existing literature on CAN bus intrusion detection using the Car-Hacking dataset.

The proposed framework achieves competitive accuracy with prior LSTM- and CNN-based approaches while introducing a key advantage: the integration of multiple complementary architectures (CNN, RNN, LSTM, BERT, RoBERTa) evaluated under three distinct resampling strategies, providing a comprehensive benchmarking perspective that

Table 6. Comparison with existing IDS methods on the Car-Hacking dataset.

Method	Model	Accuracy (%)	F1-Score (%)
Hossain et al. [11]	LSTM	99.99	99.98
Song et al. [26]	Deep CNN	99.97	99.96
Nguyen et al. [15]	Transformer (TAN)	99.10	98.80
Proposed (RoBERTa + ROS)	**RoBERTa**	**99.00**	**97.00**

individual single-model studies do not offer. Furthermore, the inclusion of Transformer-based models (BERT and RoBERTa) in the context of the Car-Hacking dataset represents a methodological contribution, as prior work has predominantly relied on recurrent or convolutional architectures.

5.3 Impact of Resampling on Detection Performance

Among the three resampling strategies evaluated, Random Over-Sampling (ROS) yields the most consistent improvement in recall and F1-score across all models, particularly for BERT and RoBERTa. NearMiss undersampling reduces majority-class dominance but at the cost of slightly lower overall accuracy for simpler models (CNN, RNN), as removing majority-class samples may discard informative boundary instances. Tomek Links, operating as a data-cleaning technique rather than a balancing technique per se, improves precision and F1-score by eliminating ambiguous borderline instances, making it most effective when combined with other resampling methods.

5.4 Limitations

Several limitations of the current study should be acknowledged. First, evaluation is conducted on a single dataset (HCRL Car-Hacking) collected from one vehicle model (Hyundai YF Sonata), which may limit the generalizability of findings to other manufacturers and CAN bus configurations. Second, the application of BERT and RoBERTa to CAN bus data involves a sequential tokenization of CAN fields that does not exploit the pre-trained linguistic knowledge of these models; performance gains therefore reflect the self-attention mechanism rather than transfer learning from natural language. Third, the framework is evaluated offline on a static dataset; real-time deployment on resource-constrained in-vehicle hardware introduces latency and computational constraints not addressed here. Future work should address these limitations through multi-dataset evaluation, lightweight model variants, and hardware-in-the-loop testing.

6 Conclusion and Future Work

Investigations in this paper presented a sophisticated Intrusion Detection System based on deep learning for connected cars within IoV framework, by applying models such as CNN, RNN, LSTM, BERT and RoBERTa for timely monitoring of cyber-attacks. To correct the unbalanced nature of the car-hacking data, we used resampling techniques (ROS, NearMiss and Tomek Links) and that improved the detection rate. Traditional machine learning models were outperformed by the framework in catching message injection, sensor spoofing and attacks designed to overload the system. Real-time data processing is possible with cloud computing which plays a key role in changing AV systems. Because of the BERT and RoBERTa models, carmakers now recognize more advanced cyber threats.

Future work will focus on expanding the dataset for broader vehicle manufacturer coverage, using multi-modal data, and testing the IDS in real-world operational environments. It will also consider reinforcement learning which helps the model respond to modern attacks and try to improve the model's performance for quick, real-time identification. Working with regulatory agencies to set up standard cybersecurity guidelines for AVs will make vehicle networks more protected. Having an IDS useful in large-scale setups is a major research challenge for the future.

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] AlEisa, H. N., Alrowais, F., Allafi, R., Almalki, N. S., Faqih, R., Marzouk, R., ... & Ibrahim, S. S.

(2023). Transforming transportation: Safe and secure vehicular communication and anomaly detection with intelligent cyber–physical system and deep learning. *IEEE Transactions on Consumer Electronics, 70*(1), 1736-1746. [Crossref]

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

[3] Setitra, M. A., & Fan, M. (2024). Detection of DDoS attacks in SDN-based VANET using optimized TabNet. *Computer Standards & Interfaces, 90*, 103845. [Crossref]

[4] Sun, X., Yu, F. R., & Zhang, P. (2021). A survey on cyber-security of connected and autonomous vehicles (CAVs). *IEEE Transactions on Intelligent Transportation Systems, 23*(7), 6240-6259. [Crossref]

[5] Korium, M. S., Saber, M., Beattie, A., Narayanan, A., Sahoo, S., & Nardelli, P. H. (2024). Intrusion detection system for cyberattacks in the Internet of Vehicles environment. *Ad hoc networks, 153*, 103330. [Crossref]

[6] Almehdhar, M., Albaseer, A., Khan, M. A., Abdallah, M., Menouar, H., Al-Kuwari, S., & Al-Fuqaha, A. (2024). Deep learning in the fast lane: A survey on advanced intrusion detection systems for intelligent vehicle networks. *IEEE Open Journal of Vehicular Technology, 5*, 869-906. [Crossref]

[7] Bergies, S., Aljohani, T. M., Su, S. F., & Elsisi, M. (2024). An IoT-based deep-learning architecture to secure automated electric vehicles against cyberattacks and data loss. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 54*(9), 5717-5732. [Crossref]

[8] Anand, M., Kumar, S. P., Selvi, M., SVN, S. K., Ram, G. D., & Kannan, A. (2023, March). Deep learning model based IDS for detecting cyber attacks in IoT based smart vehicle network. In *2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)* (pp. 281-286). IEEE. [Crossref]

[9] Wazid, M., Singh, J., Pandey, C., Sherratt, R. S., Das, A. K., Giri, D., & Park, Y. (2025). Explainable deep Learning-Enabled malware attack detection for IoT-Enabled intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, 26(5), 7231-7244. [Crossref]

[10] Sadaf, M., Iqbal, Z., Javed, A. R., Saba, I., Krichen, M., Majeed, S., & Raza, A. (2023). Connected and automated vehicles: Infrastructure, applications, security, critical challenges, and future aspects. *Technologies, 11*(5), 117. [Crossref]

[11] Hossain, M. D., Inoue, H., Ochiai, H., Fall, D., & Kadobayashi, Y. (2020). LSTM-based intrusion detection system for in-vehicle can bus communications. *IEEE Access, 8*, 185489-185502. [Crossref]

[12] Anbalagan, S., Raja, G., Gurumoorthy, S., Suresh, R. D., & Dev, K. (2023). IIDS: Intelligent intrusion detection system for sustainable development in autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems, 24*(12), 15866-15875. [Crossref]

[13] Giannaros, A., Karras, A., Theodorakopoulos, L., Karras, C., Kranias, P., Schizas, N., ... & Tsolis, D. (2023). Autonomous vehicles: Sophisticated attacks, safety issues, challenges, open topics, blockchain, and future directions. *Journal of Cybersecurity and Privacy, 3*(3), 493-543. [Crossref]

[14] Moustafa, N., Koroniotis, N., Keshk, M., Zomaya, A. Y., & Tari, Z. (2023). Explainable intrusion detection for cyber defences in the internet of things: Opportunities and solutions. *IEEE Communications Surveys & Tutorials, 25*(3), 1775-1807. [Crossref]

[15] Nguyen, T. P., Nam, H., & Kim, D. (2023). Transformer-based attention network for in-vehicle intrusion detection. *IEEE Access, 11*, 55389-55403. [Crossref]

[16] Al Isawi, O. A., Al Jaafari, K. A., & Al Sumaiti, A. S. (2025). Electric Vehicles CAN Bus Cyber Attacks Detection Using Adaptive Neuro Fuzzy Inference System. *IEEE Access, 13*, 90862-90874. [Crossref]

[17] Aloraini, F., Javed, A., & Rana, O. (2024). Adversarial attacks on intrusion detection systems in in-vehicle networks of connected and autonomous vehicles. *Sensors, 24*(12), 3848. [Crossref]

[18] Iqubal, A., & Tiwari, S. K. (2024, November). Internet of Vehicle (IoV) Cyber Attack Detection using Machine Learning Techniques. In *2024 4th International Conference on Advancement in Electronics & Communication Engineering (AECE)* (pp. 1053-1057). IEEE. [Crossref]

[19] Aurangzeb, S., Aleem, M., Khan, M. T., Anwar, H., & Siddique, M. S. (2024). Cybersecurity for autonomous vehicles against malware attacks in smart-cities. *Cluster Computing, 27*(3), 3363-3378. [Crossref]

[20] Arshad, I., Alsamhi, S. H., Qiao, Y., Lee, B., & Ye, Y. (2023). A novel framework for smart cyber defence: a deep-dive into deep learning attacks and defences. *IEEE Access, 11*, 88527-88548. [Crossref]

[21] Oseni, A., Moustafa, N., Creech, G., Sohrabi, N., Strelzoff, A., Tari, Z., & Linkov, I. (2022). An explainable deep learning framework for resilient intrusion detection in IoT-enabled transportation networks. *IEEE Transactions on Intelligent Transportation Systems, 24*(1), 1000-1014. [Crossref]

[22] Gupta, S., Maple, C., & Passerone, R. (2023). An investigation of cyber-attacks and security mechanisms for connected and autonomous vehicles. *IEEE Access, 11*, 90641-90669. [Crossref]

[23] Dasari, D. R., & Bindu, G. H. (2025). An Intelligent Intrusion Detection System in IoV Using Machine Learning and Deep Learning Models. *International Journal of Communication Systems, 38*(10), e70131. [Crossref]

[24] Bendiab, G., Hameurlaine, A., Germanos, G., Kolokotronis, N., & Shiaeles, S. (2023). Autonomous vehicles security: Challenges and solutions using

blockchain and artificial intelligence. *IEEE Transactions on Intelligent Transportation Systems, 24*(4), 3614-3637. [Crossref]

[25] Ghrib, T., Benmohammed, M., & Pandey, P. S. (2021). Automated diagnosis of attacks in internet of things using machine learning and frequency distribution techniques. *Bulletin of Electrical Engineering and Informatics, 10*(2), 950-961. [Crossref]

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



Fida Muhammad Khan received his MS in Computer Science from the University of Science and Technology, Bannu, Pakistan in 2021. Currently, he is pursuing his Ph.D. degree in Computer Science at Qurtuba University of Science and Information Technology, Peshawar, Pakistan. His research interests include Data Mining, Cybersecurity, IoT, Machine Learning, Deep Learning, and Natural Language processing (NLP). (Email: fida5073@gmail.com)



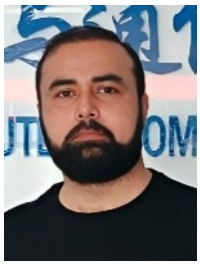
Mr. Zeeshan Ali Haider is currently pursuing a Ph.D in computer science at Qurtuba University of Science and Information Technology, Peshawar, Pakistan. He holds an MS in Computer Science from Abasyn University, Peshawar, Pakistan, where he was awarded distinction, and a BS in Computer Science from Islamia College, Peshawar, Pakistan. He has also published a book chapter in a top-tier journal. His research interests encompass a wide range of fields, including the Internet of Vehicles (IoV), Cybersecurity, Cryptography, Blockchain, Machine Learning, Deep Learning, the Internet of Things (IoT), and Data Mining. (Email: Zeeshan.ali9049@gmail.com)



Muhammad Owais Khan is an MS Computer Science candidate at the University of Science and Technology, Bannu, specializing in AI/ML, Deep Learning, and NLP. He holds a BSc in Software Engineering from the same institution and aims to advance intelligent systems for scalable problem-solving. (Email: owaiskhan.cs@outlook.com)



Ashraf Ullah received his MS degree in Computer Science from Blekinge Institute of Technology (BTH), Sweden. He completed his Ph.D. degree in Computer Science at the University of Science and Technology (UST), Bannu, Pakistan. His research interests include Information Management, Natural Language Processing (NLP), Data Science, and Artificial Intelligence (AI). (Email: ashrafbth@gmail.com)



Muhammad Shoaib Khan received his master's degree from the University of Science & Technology Bannu, Pakistan, in 2022. Currently he is pursuing his PhD degree from the School of Computer and Communication Engineering at University of Science and Technology Beijing (USTB), China, from 2023. His research interests include DevOps, software modeling and specification, IoT, network security, machine learning, big data, and anomaly detection in social networks. (Email: imkhan101@outlook.com)



Muhammad Fayaz completed his bachelor's degree from Islamia College University Peshawar and earned a master's degree in Computer Engineering, specializing in computer vision, deep learning, and machine learning, from the Department of Computer Engineering at Cyprus International University, Turkish Republic of Northern Cyprus. He is currently a research assistant at the Computer Vision and Pattern Recognition (CVPR) Laboratory at Sejong University. His research interests span computer vision, deep learning, and machine learning, with a particular focus on both medical image analysis and land cover classification. In the realm of medical image analysis, he is contributing to the development of advanced techniques for image classification and segmentation, aimed at improving diagnostic accuracy and medical decision-making. His work in land cover classification focuses on leveraging deep learning to analyze satellite and aerial imagery, enabling more effective monitoring of environmental changes, and land cover shifts, and supporting sustainable development. Through his interdisciplinary approach, Muhammad is making significant contributions to both fields, using state-of-the-art computational methods to solve complex challenges in medical imaging and environmental monitoring. (Email: muhammadfayaz@sju.ac.kr)



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)



Nabila is currently pursuing masters in Computer Science and Engineering at Bangladesh University of Professionals, Dhaka, Bangladesh. She did her Bsc in Computer Science and Engineering at Southern University Bangladesh, Chittagong, Bangladesh. Her research interests include Cryptography, Cyber Security, Machine Learning, Deep Learning and Blockchain Technology. (Email: nabila@southern.ac.bd)