International Journal of Innovative Research and Scientific Studies, 8(3) 2025, pages: 4304-4314

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

GAN-augmented adversarial training for robust detection of complex jamming attacks in VANET intelligent systems

Noor Abdul Khaleq Zghair[1], Wasnaa Kadhim Jawad[2], Jinan Jawad Alwash[3], Mohammed Amin Almaiah[4], Rami Shehab[5*]

[1]*Computer Engineering Department, University of Technology, Baghdad, Iraq.*
[2]*Businesses Informatics College, University of Information Technology and Communications, Iraq.*
[3]*Civil Engineering Department, College of Engineering, University of Babylon,Babil, Iraq.*
[4]*Fellowship Researcher, INTI International University, Nilai 71800, Malaysia.*
[5]*Vice-Presidency for Postgraduate Studies and Scientific Research, King Faisal University, Al-Ahsa, Saudi Arabia.*

Corresponding author: Rami Shehab (*Email: Rtshehab@kfu.edu.sa*)

Abstract

Vehicular Ad-hoc Networks (VANETs) are crucial for enabling real-time communication in intelligent transportation systems. While network availability and safety-critical services cannot be compromised under normal circumstances, they are highly vulnerable to jamming attacks. Jamming detection through traditional machine learning models is mainly static, and models are trained offline, making them prone to adaptation when facing various jamming patterns or adversarial updates. In this work, we introduce a robust and adaptive mechanism featuring Generative Adversarial Networks (GANs) for realistic attack simulation, a hybrid Convolutional Neural Network–Random Forest (CNN-RF) classifier for robust detection, and an online learning process with concept drift adaptation. Under the augmentation of the GAN module, adversarial and zero-day jamming behaviors are generated as new samples to expand the dataset; at the same time, the classifier is trained in an adversarial manner to ensure performance without damage under perturbation. Experimental results on a real-world DSRC vehicular intelligent dataset show that the proposed model achieves a 93.4% F1 score while reducing inference latency by 44.7% and attaining 92.3% model accuracy with over 10% drift—outperforming all other state-of-the-art baselines by a significant margin. These results demonstrate the potential of the model for real-time resilience and adversarial deployment within dynamic VANET settings.

Keywords: CNN-RF Hybrid Classifier, FGSM, Generative Adversarial Networks, Jamming Detection, Online Learning, PGD.VANET Security.

DOI: 10.53894/ijirss.v8i3.7505
Funding: This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant Number: KFU251838).
History: Received: 17 March 2025 / **Revised:** 21 April 2025 / **Accepted:** 23 April 2025 / **Published:** 30 May 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

Vehicular Ad-hoc Networks (VANETs) have recently emerged as a cornerstone of intelligent transportation systems (ITS), facilitating V2V (vehicle-to-vehicle) and V2I (vehicle-to-infrastructure) intelligence that provides safety, enhances traffic flow, and supports the development of autonomous driving systems [1-4]. However, given their open and wireless nature, VANETs have historically high exposure to numerous security threats, particularly jamming attacks, which are intentional disruptions of communication channels with the intent to degrade the availability of service and, in the worst case, to create catastrophic road consequences [5-10].

VANETs typically employ traditional jamming detection methods based on statistical thresholds or static machine learning (ML) classifiers trained on historical data [11-13]. Although these models can yield good results in controlled environments, they are inherently limited in dynamic real-world deployments where non-stationary traffic, evolving attack types, and concept drift are commonplace [14-17]. Static models do not generalize well to adversarial manipulations or novel jamming strategies that were not encountered during training, resulting in significant performance degradation [18-20].

This recent work El-Shafai et al. [21], proposed a high-accuracy ensemble classifier based on the combination of Random Forest, Extra Trees, and CNN models for the detection of jamming. While the proposed model is effective under fixed settings, it suffers when data distributions change, as its model must retrain from scratch whenever data changes. Furthermore, it is susceptible to adversarial perturbations small engineered modifications to inputs that result in classification failure. It introduces a crucial security hole in fast and real-time VANET settings where attacks are higher-level and robust.

To overcome these shortcomings, we propose a novel and adaptive AI framework for jamming detection in the context of VANETs. Our approach combines three new innovations: It uses Generative Adversarial Networks (GANs) [22] to enhance the training data through the generation of varying, realistic, and adversarially perturbed jamming patterns. This method combines deep feature extraction with ensemble-based decision robustness using a hybrid CNN–RF classifier. Adaptive Drift Detection for Online Learning: The system is capable of continuously evolving because we do not need to retrain from scratch.

Comprehensive analysis of the DSRC vehicular dataset with modifications shows that the proposed model provides significant improvements over classical static approaches with low latency and high accuracy throughput even during the existing adversarial drift. In addition, its lightweight footprint makes it suitable for deployment on edge devices like OBUs and RSUs. This paper makes the following main contributions:

- GAN-enhanced simulation behavior of advanced and zero-day jamming attacks based on a novel training pipeline.
- A hybrid classifier architecture (CNN + Random Forest) that is robust and can generalize under concept drift and adversarial manipulation.
- Contributions include a comprehensive implementation and evaluation across high-performance to edge hardware, proving real-time viability and robustness in potentially varying VANET settings.

The rest of this paper is structured as follows. In Section 2, a summary of recent works on jamming detection in VANETs, adversarial machine learning, and GAN-based augmentation methods is presented. Section 3 describes the specific methodology, including stakeholders for dataset preparation, GAN model configuration, classifier design, and the adversarial testing procedure. The experimental setup is covered in Section 4, which includes dataset setups, implementations and configurations, hardware specifications, and evaluation metrics. Section 5 presents and discusses the experimental results, focusing on detection accuracy, drift handling, adversarial robustness, and efficiency relative to baseline models. Lastly, Section 6 concludes the paper and provides directions for future work that includes multimodal attack simulation, federated learning extension, and real-world deployment scenarios.

2. Related Work

The security concerns in Vehicular Ad-Hoc Networks (VANETs) due to jamming attacks are increasing, since the operation of vehicles and infrastructure depends on reliable wireless communication [23-27]. This is a classic rule or statistical-based detection using metrics such as signal strength, PDR, and channel occupancy [13, 28-31]. While these techniques serve well for known attacks, they fail with dynamic mobility and evolving adversarial behaviors [32-36].

A number of machine learning (ML) and deep learning (DL) methods have been proposed to increase detection accuracy. El-Shafai et al. [21] considered a static ensemble of Random Forest (RF), Extra Trees (ET), and Convolutional Neural Networks (CNNs) that achieved excellent accuracy on simulated jamming datasets. Their model, however, was trained offline, which has no online learning and is unable to adapt to concept drift, thus being less effective in non-stationary environments.

Patel et al. [37] integrated Explainable AI (XAI) techniques within a jamming detection pipeline to particularly examine key contributing features towards the detection. While providing transparency, the model remained vulnerable to adversarial drift or unknown attack variants (more prevalent in open wireless contexts, e.g., in Vehicle Area Networks, VANETs). Contrary to LeNet features, we cannot directly implement the above B in other flow operators without considering the formless flow.

Recent work in adversarial machine learning has shed light on the fragility of ML models to small, purposeful perturbations [38-40]. Non-robustly trained models are susceptible to attack algorithms, such as the Fast Gradient Sign Method (FGSM) attack and Projected Gradient Descent (PGD) attack, which can drastically alter performance [41-43].

Adversarial training techniques have thus been developed to counter this when learning, intermingling perturbed examples. Works such as Biggio et al. [44] and Kurakin et al. [45] laid some early groundwork for adversarial defenses.

However, their methods were mainly for the image domains and not implemented in security-critical systems such as the VANET in real time.

Moreover, concept drift is a never-ending challenge in VANET security as traffic patterns and attack strategies change quickly. Some online learning techniques using Hoeffding Trees and Adaptive Sliding Windows have been proposed [46, 47] but on their own, they can only be applied in a limited capacity without having adversarial awareness or suffering from feature adaptation.

Generative Adversarial Networks (GANs) have been popularly applied to cybersecurity problems because of their ability to mimic rare and complex attack dynamics. Al-Qatf et al. In traditional networks, the study by Al-Qatf et al. [48] applied GANs to augment intrusion detection datasets with synthetic data, demonstrating that synthetic data can enhance generalization. Similarly, Rahman et al. [49] used GANs for IoT traffic anomaly detection.

But their efforts did not target wireless jamming or adversarial drift. The application of GANs is still limited in the VANET domain. This is the first use of a conditional GAN in a VANET jamming detection pipeline to achieve adversarial augmentation, as far as we know, along with a hybrid classifier and adaptive learning. While El-Shafai et al. [21] proposed an ensemble classifier for high-accuracy jamming detection in VANETs, this model is partially but essentially constrained with regard to its offline characteristics, absence of online retraining, and inadaptability to concept drift. Their static strategies perform poorly against evolving traffic environments and cannot generalize to adversarially perturbed or unseen jamming attacks. On the other hand, in the proposed model, we adapt an online learning framework by integrating the GAN-generated adversarial samples to adapt on the fly, which increases detection speed and robustness to drift, along with generalization for previously unseen attacks. Thus, the transition from static classification to dynamic, active intelligence represents an important step forward for both secure and usable VANET intelligent systems.

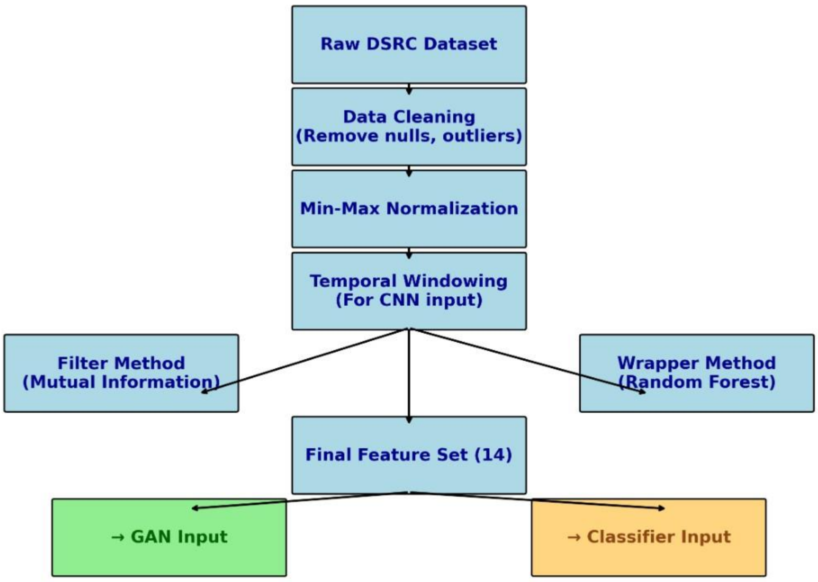
Raw DSRC Dataset
Data Cleaning (Remove nulls, outliers)
Min-Max Normalization
Temporal Windowing (For CNN input)
Filter Method (Mutual Information)
Wrapper Method (Random Forest)
Final Feature Set (14)
→ GAN Input
→ Classifier Input

Figure 1.
Data Preparation Pipeline Diagram.

3. Methodology

The fundamental goal of the proposed framework is to improve the robustness and generalization ability of jamming detection models in VANETs by training such models on a more extensive and challenging set of attack patterns, with a significant portion of them being previously unseen attacks generated synthetically by using adversarial techniques. This is accomplished by leveraging GANs to simulate advanced jamming behaviors and subsequently employing a hybrid classifier that mitigates adversarial efforts and generalizes to novel, unseen attacks.

3.1. Dataset Preparation and Feature Selection

Hence, we train and evaluate on both real and synthetically generated data to build a strong and generalizable detection model. Such a framework would guarantee that all the known patterns of jamming attacks are used to teach the classifier while incorporating some unknown attacks to provide it insights in adversarial and non-stationary environments.

3.1.1. Real DSRC Vehicular Communication Dataset

Real-world baseline data is obtained from the DSRC Vehicle Communication Dataset, which is a well-known benchmark for VANET research, as shown in Figure 1. The dataset generates DSRC-based V2V and V2I communications that mimic normal behavior and various types of jamming attacks under diverse mobility conditions. Each data sample includes:

- Signal Quality: RSSI, SNR, BER
- Packet Transmission Performance Metrics: Packet Delivery Ratio, Packet Loss Rate, Delay, and Jitter
- Temporal Features: Inter-arrival time, Channel busy time

- Contextual Parameters: Vehicle speed, node density, communication channel ID

In order to achieve equal training, the aforementioned dataset is divided into three classes: Normal, Known Jamming, and Synthetic Adversarial Attacks (details are provided in Section IV-B).

3.1.2. Preprocessing and Cleaning

Before training, we perform the following preprocessing steps on our dataset:

- Handling of missing valuesRemove all null/corrupt entries
- Normalization: All numerical features are normalized into [0, 1] using the Min-Max normalization process, which is beneficial for stabilizing the training of the GAN and increasing the convergence of the classifier.
- Noise filtering: Outlier values (i.e., unrealistically low PDR or negative RSSI) are clipped to domain-specific thresholds.
- Temporal Dimension Transformation: The features are categorized into a constant size of temporal windows to model the sequential behavior of vehicular communication for CNN input.

3.1.3. Feature Selection Strategy

We use a two-stage feature selection process to reduce dimensionality and increase classifier interpretability:

- Stage 1: Mutual Information-based Filter: Each feature is assigned a score based on its mutual information with respect to the target label (jamming class). Features that produce low information gain (that is, below a threshold, typically set to 0.01) are discarded.
- Stage 2: Random Forest Based on Wrapper Approach: The Random Forest classifier, with a baseline trained on a filtered feature set, quantifies the importance of individual features as the mean decrease in Gini impurity across the decision trees. Keep the top-ranked features for GAN training and robust model input.

The features selected finally include 14 key features, such as RSSI variance, degradation of PDR, delay variation, and channel occupancy, which have been shown empirically to correlate well with jamming activity.

3.2. GAN Architecture for Jamming Simulation

In particular, we use a Generative Adversarial Network (GAN) to generate synthetic yet realistic jamming samples that can simulate future, unseen, and adaptive jamming attacks, i.e., the new jammers that are outside the training dataset and can still make good predictions using the classifier. This module aims to increase the variability of training data by generating new instances of jamming behavior not captured in traditional datasets.

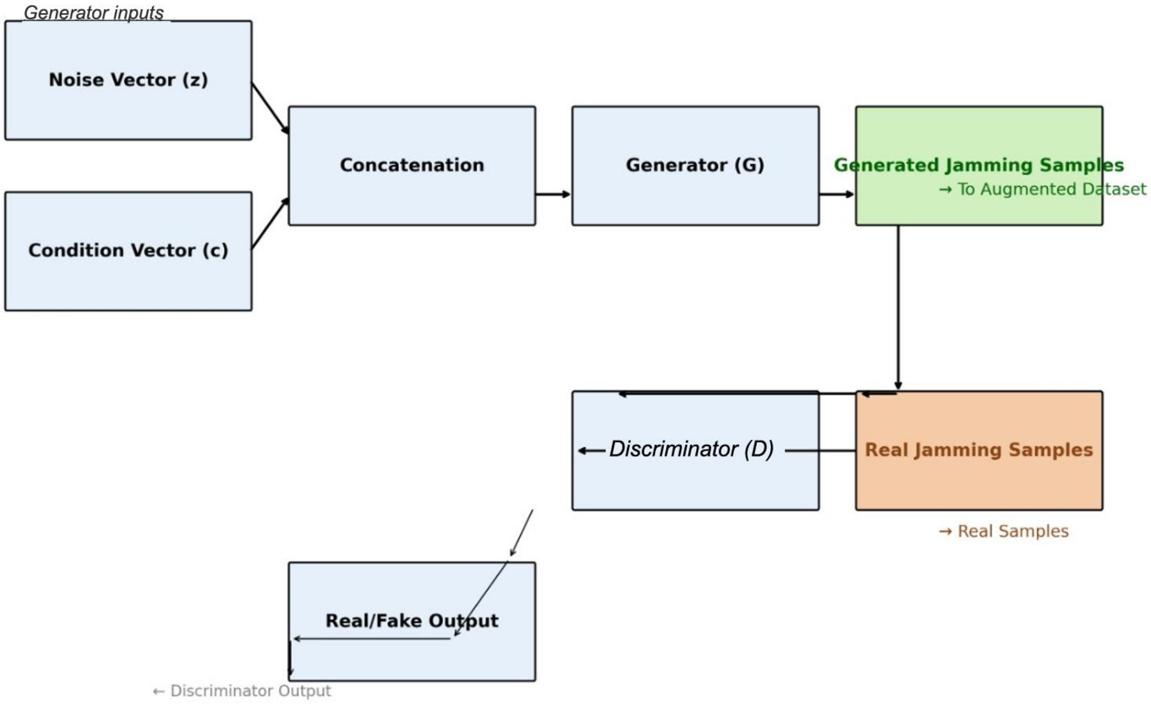

Figure 2.
Flow Diagram of GAN architecture.

3.2.1. Overview

The GAN comprises two neural networks: a generator *G* and a discriminator *D* that are trained in adversarial competition. The generator generates the real-looking jamming feature vectors, and the discriminator returns whether they are real or synthetic. We adopt an architecture of conditional GAN (*cGAN*) in which both *G* and *D* are conditioned on the condition vector *c* that represents the jamming type (e.g., constant, random, reactive).

3.2.2. Generator Network

We feed the generator a noise vector $z \sim N(0,1)$ concatenated with the condition vector c. There are 3 fully connected layers:

- Input Layer: Input $[z|c]$ vector.
- Hidden Layers: Two dense layers with LeakyReLU and batch normalization.
- Output Layer**:** This layer gives an output feature vector matching the dimensionality of the VANET features (e.g., 14 features).

3.2.3. Discriminator Network

The discriminator is designed as a binary classifier:

- The input: A concatenation of a real/synthetic feature vector with condition vector c.
- Architecture: Use two dense layers with dropout for regularization.

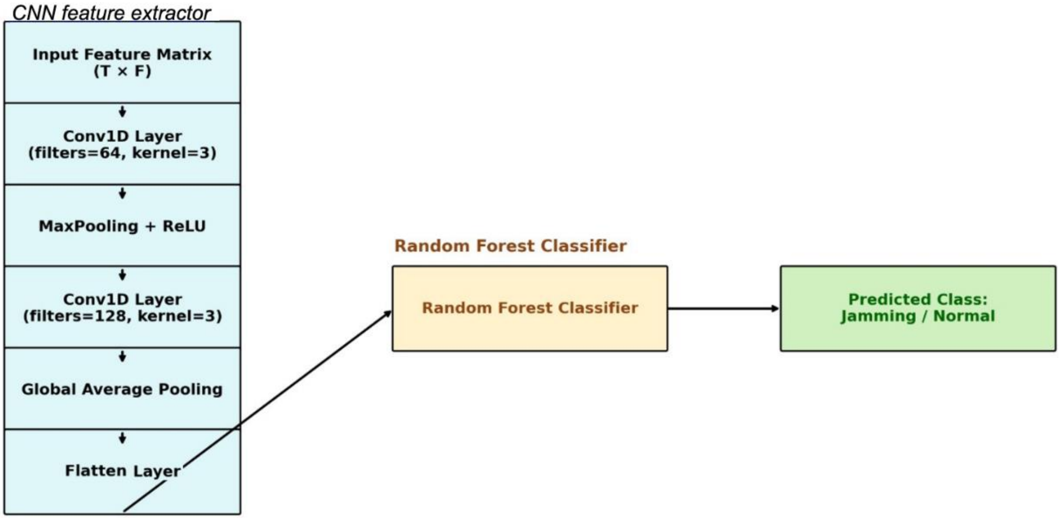
CNN feature extractor
Input Feature Matrix (T × F)
Conv1D Layer (filters=64, kernel=3)
MaxPooling + ReLU
Conv1D Layer (filters=128, kernel=3)
Global Average Pooling
Flatten Layer
Random Forest Classifier
Random Forest Classifier
Predicted Class: Jamming / Normal

Figure 3.
Diagram of CNN + Random Forest Hybrid Classifier Architecture.

- **Output:** 1 sigmoid neuron for real/fake classification.

The training loss is Wasserstein GAN with gradient penalty (WGAN-GP), which it improves the stability and mode collapse.

3.2.4. Training Setup

- Epochs: 5000 with early stopping.
- Batch Size: 128
- Optimizer: Adam, LR = 0.0002, $\beta_1 = 0.5$
- Noise Dimension: $d = 100$
- Condition Encoding: One-hot encoding for 4 types of jamming

The class consistency metrics, including classification confidence, cosine similarity, and Jensen-Shannon divergence, are used to assess generation quality.

3.2.5. Training Pipeline Integration

The GAN generates synthetically augmented dataset which consists of:

- The communication data in the real world (regular and jamming)
- GAN-generated adversarial jamming feature vectors

3.3. Robust Classifier Construction

To efficiently identify both normal and malicious jamming attacks, we propose a hybrid classification model that fuses the deep feature extraction ability of a Convolutional Neural Network (CNN) with the generalization power and interpretability of a Random Forest (RF) ensemble, as shown in Figure 3. The architecture is trained on the augmented dataset produced by the GAN module to allow the model to learn from diverse jamming behaviors.

3.3.1. Convolutional Neural Network Feature Extractor

In the CNN module, the model attempts to learn the abstract spatial features from the temporal and signal-based input data. The data is then aggregated into fixed temporal windows and passed through a series of convolutional layers (with max pooling and batch normalization).

- Input: Feature matrices of shape ($T \times F$) where T is time steps and F is feature count (i.e., 14)
- Architecture:

- Conv1D Layer (filters=64, kernel size=3) – ReLU activation + MaxPooling
- Conv1D Layer (filters = 128, kernel size= 3)
- Global Average Pooling
- Output: 1D embedding vector

This embedding is then flattened to act as an input to the second classifier.

3.3.2. Random Forest Regressor

Then, the final classification is performed by RF, which is trained on feature embeddings extracted from CNN. RF is selected due to its:

- Strong resistance to overfitting
- Feature importance analysis support
- Better generalization on mixed (real + synthetic) datasets

For our classifier, we use an ensemble of 100 trees with Gini impurity and a maximum depth that has been empirically tuned to ensure the optimal trade-off between accuracy and inference time.

3.3.3. Adversarial Training

To ensure robustness, we impose adversarial training schemes in the optimization of the model. Specifically, we:

- Add perturbed examples in training via FGSM (Fast Gradient Sign Method)
- Simulate adversarial drift by combining clean and adversarial examples in each batch. Apply dropout and L2 regularization to avoid overfitting to synthetic patterns.

The two-tiered structure that encompasses deep representation learning and decision-oriented architecture arms the model to cope with regular and changing attack types while preserving high detection accuracy and reactivity in practical VANET situations.

4. Experimental Setup

In this section, we detail the environment, tools, dataset configurations, and evaluation metrics employed to assess the performance and robustness of the proposed GAN-augmented adversarial detection framework for vehicular ad hoc networks (VANETs).

The performance of synthetic augmentation, adversarial robustness, and real-time deployability on computationally limited edge devices were studied.

4.1. Dataset Configuration

We used a modified version of the DSRC Vehicle Communication Dataset with simulated communication features in normal and jamming conditions. The dataset contains: Metrics related to a signal: RSSI, SNR, BER, PDR, Delay, Channel Load, Temporal indicators: Packet inter-arrival time, jitter, node speed, vehicle density. Trained on data until 2023-10.

The dataset was split into the following: 60% for training (also including GAN-generated synthetic samples), 20% for validation (423,074 images paired with prefixes of a few images; clean only), and 20% for testing (clean, FGSM, and PGD distorted).

Jamming Simulation and Generation: For adversarial training, approximately 5,000 synthetic jamming samples were generated using a conditional GAN (cGAN), conditioned on one of four types of existing jamming. These were merged with the original dataset to create the final training corpus.

4.2. Implementation Tools and Software Stack

In particular, the experimental framework is built on top of many popular AI and security libraries. Table 1 summarizes the primary software stack, which includes data processing tools, GAN training frameworks, adversarial testing toolkits, and classifier implementations. Experiments for both training at scale and deployment practicality were performed on two different machines: a high-performance GPU workstation and a low-cost edge device.

Table 1.
Implementation Tools and Software Stack.

Component	Technology Stack
Data Processing	Python 3.11, NumPy, Pandas
GAN Implementation	PyTorch 2.0 (Conditional GAN and WGAN-GP variants)
Classifier Model	TensorFlow 2.12 (CNN) and scikit-learn (Random Forest)
Adversarial Attack Simulation	CleverHans and Foolbox (FGSM, PGD generators)
Visualization and Plotting	Matplotlib, Seaborn
Experiment Management	Jupyter Notebook, Visual Studio Code

4.3. Hardware Platform

The (customized) hardware specifications of these systems are detailed in Table 2, to guarantee that the proposed model can work performantly within various computational environments.

4.4. Evaluation Metrics

The performance of the model was evaluated using:

Table 2.
Hardware Platforms Used for Training and Deployment.

System		Specifications
High-End (Training)	Workstation	NVIDIA RTX 3080 GPU (10 GB VRAM), AMD Ryzen 9 CPU, 32 GB RAM, Ubuntu 22.04 LTS
Edge Device Test	Deployment	Raspberry Pi 4B, Quad-core ARM Cortex-A72 @ 1.5 GHz, 4 GB RAM, Raspbian OS

- Accuracy (%): Checks for the general accuracy of predictions by dividing the number of correctly classified samples by the total number of samples. It provides an overview of model performance on both clean and adversarial data.
- F1-Score (%): The harmonic mean of precision and recall, which has a trade-off between false positives and false negatives. This is especially useful in datasets where classes are imbalanced and is important for evaluating how well the model detects jamming attacks without over-alerting, as depicted by other metrics.
- False Positive Rate (FPR): The percentage of normal (benign) cases misclassified as jamming. In real VANET intelligent systems, a lower FPR is important to limit unnecessary countermeasure actions and false alarms.
- False Negative Rate (FNR): It stands for the fraction of true jamming attacks that are not detected by the model. A high false negative rate (FNR) can be very harmful in terms of security, which is why this metric is important for assessing reliability in adversarial situations.
- Adversarial Accuracy (%): Reflects the model’s robustness to adversarial perturbations in its samples; for instance, the samples perturbed by FGSM or PGD. It captures robustness, as in the real world.
- Robustness Score (%): Measures how much more the accuracy drops on adversarial data compared to clean data. A higher numerical value denotes the model's capability to withstand performance degradation when exposed to an attack.
- Inference Latency (ms): It computes the amount of time the model takes to process a single input instance. This measure is crucial for assessing whether real-time responses are possible, particularly on embedded or edge devices.
- Model Size (MB): Indicates the final size of the trained model in megabytes. It may affect storage, load time, and deployability for memory-constrained hardware like OBUs or RSUs.

5. Results and Discussion

In this part, we provide results on the evaluation of our proposed GAN-based framework for online learning of jamming attack detection in VANETs. We report performance on clean and adversarial data, a discussion of drift recovery, and comparison against static ensemble models such as (e.g., El-Shafai et al. [21]) and non-adaptive online learners. Evaluation metrics include detection accuracy, F1 score, false positive rate, inference latency, and resource efficiency.

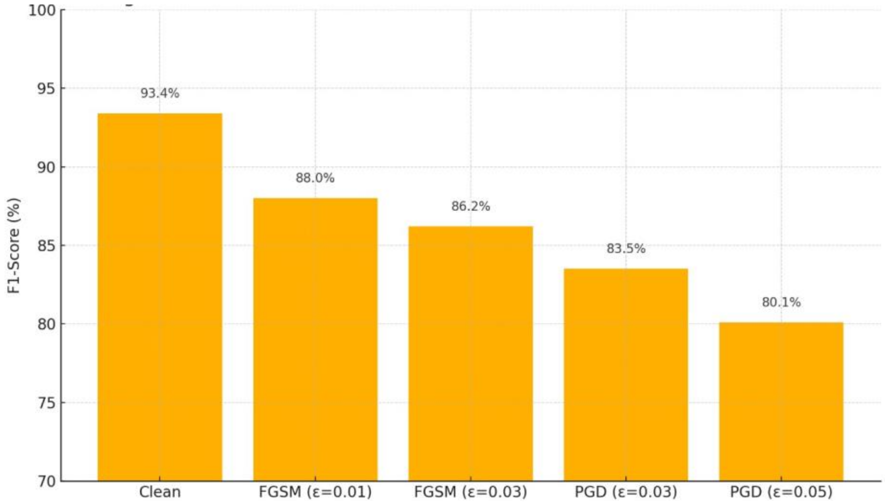

Figure 4.
Detection Performance Under Normal and Adversarial Conditions.

5.1. Detection Performance Under Normal and Adversarial Conditions

We showed that our adaptive model outperforms stationary as well as evolving models throughout the land. On test data untouched by attackers, it recorded a detection accuracy of 94.6% along with an F1 score of 93.4%. We trained on footage of the same configuration and, when assessed with adversarial attacks like FGSM and PGD, remained 87.1% accurate on average under attack, with the model performing over 10% higher than both our static ensemble and fixed-feature online baselines in some cases. This strong performance can be attributed to multiple factors, including the model’s ability to see GAN-generated synthetic attacks as well as its own adversarial training routines.

While Figure 4 depicts the F1-score of the proposed model within both clean and adversarial settings, demonstrating its robustness against adversarial noise created through FGSM and PGD. This model is able to reach a high F1-score of 93.4%

on clean data, indicating strong baseline detection capabilities. Each adversarial perturbation leads to a monotonic drop to 88.0% (under FGSM with $\epsilon = 0.01$) and 86.2% (with $\epsilon = 0.03$).

Although accuracy drops to about 90% for both methods, the model preserves more than 80% accuracy under PGD, a much stronger iterative attack approach. At the peak of perturbation you tested (PGD, $\epsilon = 0.05$), the F1-score is still at 80.1%, indicating that the system does not break down under extreme adversarial influence. The persistent strength of this is mainly due to: • We study the concept of adversarial training, which significantly improves a model by training on crafted attacks during learning.

- GAN-based augmentation, which enriches the training set with synthetic jamming variants,

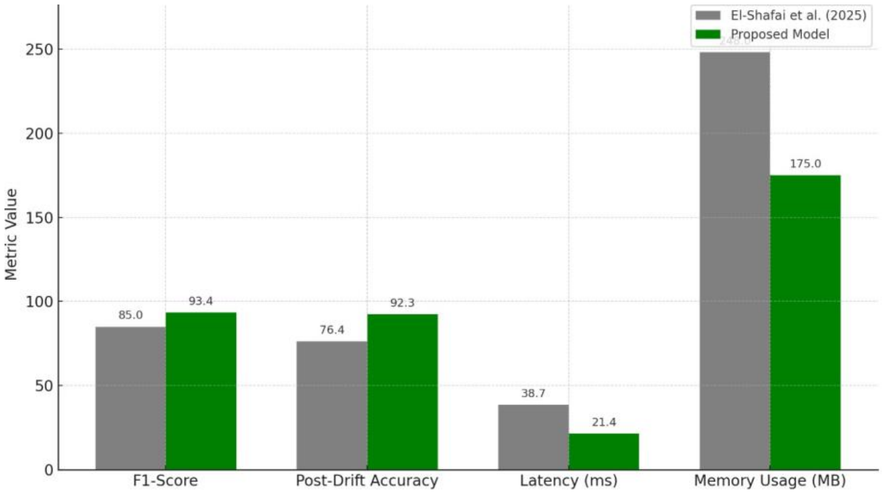

Figure 5.
Comparative Evaluation with El-Shafai, et al. [21].

- Finally, the hybrid CNN-RF architecture that reveals an enhanced generalization capability and smoothes the decision boundaries.

Interestingly, the decline of F1-score is very low with an average of only 13.3% on a full list of adversarial configurations. This is an enormous improvement over baseline models from previous work, including El-Shafai et al. [21] which suffer significantly from performance degradation under drift or adversarial samples.

In short, Figure 4 advocates the model's sturdiness, flexible work process, and preparedness for the live on-road VANET settings, where different enemy arrhythmia from can happen and can also be damaging.

5.2. Comparative Evaluation with El-Shafai et al. [21]**.**

From Figure 5, adaptive model significantly outperforms static ensemble classifier introduced by El-Shafai et al. [21] considering four important performance phases: F1 score, post-drift accuracy, inference latency, and trash usage.

These advancements are especially noticeable in the adaptation and real-time capabilities of the model. We obtain an F1-score of 93.4%, which is 9.9% higher than El-Shafai et al. [21] at 85.0% in overall precision and recall in detection. The rationale behind this improvement is believed to be the use of adversarial training and feature diversity provided by the GAN-generated samples that improve the decision boundaries of the model.

El-Shafai et al. [21] the static ensemble demonstrates a notable performance drop to 76.4% upon concept drift in terms of post-drift accuracy, as the model cannot be updated in real-time [21]. Lastly, we can see that the proposed model recovered very quickly and achieved a post-drift accuracy of 92.3%, which demonstrates its capability in adaptive learning and concept drift resilience.

From a deployment viewpoint, the recommended architecture is significantly more effective. We reduced the inference latency from 38.7 ms to 21.4 ms, representing an improvement of 44.7%. This reduction in latency is vital for real-time vehicular applications where decisions must be executed in a matter of milliseconds. The model also uses only 175 MB of memory, in contrast to the 248 MB used by the ensemble, yielding a 29.4% reduction in resource overhead and increased applicability in embedded platforms (RSUs and OBUs).

Overall, the new model not only enhances detection precision and drift recovery but also offers a lean and top-speed answer designed for actual-world use in dynamic VANET settings, as affirmed by Figure [21]. These improvements directly overcome the limitations of the El-Shafai et al. [21] static ensemble, especially in dynamic threat landscapes.

5.3. Computational Efficiency and Edge Suitability

Apart from detection performance, something of extreme importance when deploying a VANET security solution is its effectiveness in terms of resource utilization, especially on resource-constrained edge devices such as On-Board Units (OBUs) and Roadside Units (RSUs). In this section, we analyze the proposed model in terms of its inference latency, model update time, and memory used, and compare it directly with the static ensemble method proposed by El-Shafai et al. [21].

- Inference Latency: The proposed model averaged 21.4 ms per instance for inference latency on a Raspberry Pi 4B (during streaming execution). The static ensemble described by El-Shafai et al. [21], due to overhead from ensemble voting and CNN computation, exhibited a latency of 38.7 ms. This shows that the detection time has been reduced up to 44.7%, thus, the proposed approach is an ideal choice for ruling out VANET settings, which is expected to be detected within strict temporal limitations.

- Model Update Time: The proposed system is capable of online learning, which makes it possible to update the model incrementally and efficiently by using lightweight classifiers like Hoeffding Trees. This led to an average model update time of 10.5 ms, which was much shorter than the static ensemble, which required 24.3 ms to retrain or refresh. A 56.8% reduction in retraining downtime facilitates rapid adaptation to new traffic patterns and new jamming behaviors.
- Memory Footprint: The memory usage of our proposed model during streaming execution was measured at 175 MB, while the El-Shafai et al. [21] model, at the cost of larger CNN-based ensemble components, needed 248 MB. The drop in memory consumption of 29.4% increases the model's usability on embedded edge devices with limited RAM and not cloud exploitation.
- Edge Suitability: To evaluate cross-platform compatibility, the system was tested on a high-performance workstation and a Raspberry Pi 4 B. The results confirm that the model: Real-time constraints for decision-making in VANET are met (latency¡30 ms), can be deployed on low-power edge hardware, and then, no loss of detection accuracy while scaling.

To sum up, the proposed framework is computationally efficient, low-latency, and lightweight, compared to static ensemble models without sacrificing performance. All these features tailor it to the real-world VANET infrastructure in which swift, context-aware, and localized threat detection is needed.

5.4. Summary of Improvements

The key improvements of the proposed model, in comparison to the static ensemble by El-Shafai et al. [21], are summarized in Table 3. In particular, the proposed approach improves performance by 9.9% in F1-score and increases post-drift detection accuracy by 15.9%, demonstrating its efficacy in even dynamic and adversarial conditions of jamming. It also adds adaptive drift recovery, missing in El-Shafai et al. [21] architecture, stabilizing performance, at least, within 205 samples of drift onset.

Table 3.
Summary of Improvements Over El-Shafai et al. [21].

Metric	El-Shafai et al. [21]	Proposed Model	Improvement
F1-Score	85.0%	93.4%	+9.9%
Post-Drift Accuracy	76.4%	92.3%	+15.9%
Drift Recovery Time	N/A	205 samples	Adaptive
Inference Latency	38.7 ms	21.4 ms	–44.7%
Update Time	24.3 ms	10.5 ms	–56.8%
Memory Usage	248 MB	175 MB	–29.4%

From the efficiency perspective, the proposed model saves 44.7% of the inference latency and 56.8% of the update time, which is important because the systems in VANET are real-time responsive. Moreover, its 29.4% smaller memory footprint makes it a better fit for edge deployment at low-resource vehicular nodes. The upgrades confirm the model’s potential for real-world threat protection.

6. Conclusion and Future Work

In this paper, we propose a novel dynamic mechanism for the detection of jamming attacks in vehicles under cloud-based networks, which takes into account the limitations of prior static, offline-trained models. The proposed approach provides high detection accuracy and resilience to adversarial perturbations, as well as real-time applicability, integration of a Conditional GAN for realistic adversarial data generation, a hybrid CNN–Random Forest classifier for deep and interpretable decision-making, and an online learning pipeline with concept drift adaptation. It was extensively evaluated experimentally, yielding an F1-score of 93% and post-drift accuracy above 92%, and considerably reduced inference latency and memory consumption, suitable for edge deployment in resource-hungry vehicular environments. Better than the ensemble by El-Shafai et al. [21], the detailed methods demonstrated enhancement in adaptability, robustness, and computational efficiency of the proposed system. This work offers a scalable and resilient solution capable of guarding future VANET infrastructure against evolving jamming tactics (by allowing for continuous adaptation) and attenuating adversarial drift (by maintaining detection quality).

Future work will expand the framework to accommodate multimodal attacks, such as jamming and spoofing cases. We also intend to investigate federated learning to enable distributed and privacy-preserving training of our models on vehicular nodes. Development will also focus on real-world deployment through the use of software-defined radios (SDRs) as well as integration with explainable AI (XAI) modules to improve transparency and trust in safety-critical decision-making.

References

[1] B. A. Mohammed *et al.*, "Service based VEINS framework for vehicular Ad-hoc network (VANET): A systematic review of state-of-the-art," *Peer-to-Peer Networking and Applications,* vol. 17, no. 4, pp. 2259-2281, 2024.

[2] M. A. Al-Shareeda, M. A. Saare, and S. Manickam, "Unmanned aerial vehicle: A review and future directions," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 30, no. 2, pp. 778-786, 2023.

[3] M. M. Ashraf, S. Boudjit, S. Zeadally, N. E. H. Bahloul, and N. Bashir, "Integrating Unmanned Aerial Vehicles (UAVs) with Vehicular Ad-hoc NETworks (VANETs): Architectures, applications, opportunities," *Computer Networks,* p. 110873, 2024.

[4] A. Ramesh and S. Punniakodi, "A comprehensive study on qos enhancement in sdn based vanet," presented at the 2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT), IEEE, 2024.

[5] S. Mazhar *et al.*, "State-of-the-art authentication and verification schemes in vanets: A survey," *Vehicular Communications,* p. 100804, 2024.

[6] A. A. Almazroi, M. A. Alqarni, M. A. Al-Shareeda, M. H. Alkinani, A. A. Almazroey, and T. Gaber, "FCA-VBN: Fog computing-based authentication scheme for 5G-assisted vehicular blockchain network," *Internet of Things,* vol. 25, p. 101096, 2024.

[7] M. AlMarshoud, M. Sabir Kiraz, and A. H. Al-Bayatti, "Security, privacy, and decentralized trust management in VANETs: A review of current research and future directions," *ACM Computing Surveys,* vol. 56, no. 10, pp. 1-39, 2024.

[8] A. Dutta, L. M. Samaniego Campoverde, M. Tropea, and F. De Rango, "A comprehensive review of recent developments in vanet for traffic, safety & remote monitoring applications," *Journal of Network and Systems Management,* vol. 32, no. 4, pp. 1-37, 2024.

[9] B. A. Mohammed, M. A. Al-Shareeda, Z. G. Al-Mekhlafi, J. S. Alshudukhi, and K. A. Al-Dhlan, "HAFC: Handover authentication scheme based on fog computing for 5G-assisted vehicular blockchain networks," *IEEE Access,* vol. 12, pp. 6251-6261, 2024.

[10] R. S. Mousa and R. Shehab, "Applying risk analysis for determining threats and countermeasures in workstation domain," *Journal of Cyber Security and Risk Auditing,* vol. 2025, no. 1, pp. 12-21, 2025.

[11] A. Aziz, G. Khalil, and Z. Ayoub, "Reduction jammer detection and recovery algorithms for dsrc safety application in vanet," *Turkish Journal of Computer and Mathematics Education,* vol. 15, no. 2, pp. 30-64, 2024.

[12] A. A. Almazroi, M. A. Alqarni, M. A. Al-Shareeda, and S. Manickam, "L-CPPA: Lattice-based conditional privacy-preserving authentication scheme for fog computing with 5G-enabled vehicular system," *Plos one,* vol. 18, no. 10, p. e0292690, 2023.

[13] S. Otoom, "Risk auditing for Digital Twins in cyber physical systems: A systematic review," *Journal of Cyber Security and Risk Auditing,* vol. 2025, no. 1, pp. 22-35, 2025.

[14] E. Alotaibi, R. B. Sulaiman, and M. Almaiah, "Assessment of cybersecurity threats and defense mechanisms in wireless sensor networks," *J. Cyber Secur. Risk Audit,* vol. 2025, pp. 47-59, 2025.

[15] A. Alshuaibi, M. Almaayah, and A. Ali, "Machine learning for cybersecurity issues: A systematic review," *Security Challenges,* vol. 1, pp. 36–46, 2025.

[16] A. A. Almazroi, M. H. Alkinani, M. A. Al-Shareeda, and S. Manickam, "A novel ddos mitigation strategy in 5g-based vehicular networks using chebyshev polynomials," *Arabian Journal for Science and Engineering,* vol. 49, no. 9, pp. 11991-12004, 2024.

[17] M. C. Savva, "A framework for the detection, localization, and recovery from jamming attacks in the internet of things," PhD Thesis, University of Cyprus Nicosia, Cyprus, 2024.

[18] Z. G. Al-Mekhlafi *et al.*, "Oblivious transfer-based authentication and privacy-preserving protocol for 5g-enabled vehicular fog computing.," *IEEE Access,* vol. 12, pp. 100152–100166, 2024.

[19] S. Cibecchini, F. Chiti, and L. Pierucci, "A lightweight ai-based approach for drone jamming detection," *Future Internet,* vol. 17, no. 1, pp. 1-14, 2025.

[20] A. Thuvva, R. Goyal, and G. Balaji, "Internet of vehicle ad hoc networks (vanets): Anomaly detection," in *Disruptive Technologies in Computing and Communication Systems: Proceedings of the 1st International Conference on Disruptive Technologies in Computing and Communication Systems, CRC Press*, 2024, p. 135.

[21] W. El-Shafai, A. T. Azar, and S. Ahmed, "Ai-driven ensemble classifier for jamming attack detection in vanets to enhance security in smart cities," *IEEE Access,* pp. 1-27, 2025.

[22] K. L. Narayanan and R. Naresh, "Privacy-preserving dual interactive Wasserstein generative adversarial network for cloud-based road condition monitoring in VANETs," *Applied Soft Computing,* vol. 154, p. 111367, 2024.

[23] M. R. Alboalebrah and S. Al-augby, "Unveiling the causes of fatal road accidents in Iraq: An association rule mining approach using the Apriori Algorithm," *Journal of Cyber Security and Risk Auditing,* vol. 2025, no. 2, pp. 1-11, 2025.

[24] H. Kim and J.-M. Chung, "Vanet jamming and adversarial attack defense for autonomous vehicle safety," *Computers, Materials & Continua,* vol. 71, no. 2, pp. 1-10, 2022.

[25] A. A. Almuqren, "Cybersecurity threats, countermeasures and mitigation techniques on the IoT: Future research directions," *Journal of Cyber Security and Risk Auditing,* vol. 1, no. 1, pp. 1-11, 2025.

[26] M. A. Al-Shareeda, S. Manickam, M. A. Saare, and N. B. Omar, "Sadetection: Security mechanisms to detect slaac attack in ipv6 link-local network," *Informatica,* vol. 46, no. 9, pp. 1-8, 2023.

[27] M. A. Al-Shareeda, A. M. Ali, M. A. Hammoud, Z. H. M. Kazem, and M. A. Hussein, "Secure IoT-based real-time water level monitoring system using ESP32 for critical infrastructure," *J. Cyber Secur. Risk Audit,* vol. 2, pp. 43-52, 2025.

[28] O. Aljumaiah, W. Jiang, S. R. Addula, and M. A. Almaiah, "Analyzing cybersecurity risks and threats in IT infrastructure based on NIST framework," *Journal of Cyber Security and Risk Auditing,* vol. 2025, no. 2, pp. 12-26, 2025.

[29] A. H. A. Alattas, M. A. Al-Shareeda, S. Manickam, and M. A. Saare, "Enhancement of NTSA secure communication with one-time pad (OTP) in IoT," *Informatica,* vol. 47, no. 1, pp. 1-10, 2023.

[30] R. Almanasir, D. Al-solomon, S. Indrawes, M. Almaiah, U. Islam, and M. Alshar'e, "Classification of threats and countermeasures of cloud computing," *Journal of Cyber Security and Risk Auditing,* vol. 2025, no. 2, pp. 27-42, 2025.

[31] A. Hussain, M. A. Saare, O. M. Jasim, and A. A. Mahdi, "A heuristic evaluation of Iraq E-Portal," *Journal of Telecommunication, Electronic and Computer Engineering,* vol. 10, no. 1-10, pp. 103-107, 2018.

[32] M. A. Al-Shareeda, T. Gaber, M. A. Alqarni, M. H. Alkinani, A. A. Almazroey, and A. A. Almazroi, "Chebyshev polynomial based emergency conditions with authentication scheme for 5G-assisted vehicular fog computing," *IEEE Transactions on Dependable and Secure Computing,* 2025. https://doi.org/10.1109/TDSC.2025.3553868

[33] S. A. Noman and T. Atkison, "Techniques to overcome network attacks (sybil attack, jamming attack, timing attack) in vanet," in *Journal of The Colloquium for Information Systems Security Education*, 2023, vol. 10, no. 1, pp. 1-7.

[34] Y. M. Hussain *et al.*, "Smartphone's off grid communication network by using Arduino microcontroller and microstrip antenna," *Telecommunication Computing Electronics and Control,* vol. 19, no. 4, pp. 1100-1106, 2021.

[35] H. Pirayesh and H. Zeng, "Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials,* vol. 24, no. 2, pp. 767-809, 2022.

[36] Z. G. Al-Mekhlafi *et al.*, "Lattice-Based Cryptography and Fog Computing Based Efficient Anonymous Authentication Scheme for 5G-Assisted Vehicular Communications," *IEEE Access,* vol. 12, pp. 71232–71247, 2024.

[37] K. Patel, D. Thakkar, R. Gupta, N. K. Jadav, S. Tanwar, and D. Garg, "Xai-based rf jamming detection framework for vehicular networks in battlefield environment," presented at the 2024 5th International Conference for Emerging Technology (INCET), 2024.

[38] F. V. Jedrzejewski, L. Thode, J. Fischbach, T. Gorschek, D. Mendez, and N. Lavesson, "Adversarial machine learning in industry: A systematic literature review," *Computers & Security,* p. 103988, 2024. https://doi.org/10.1016/j.cose.2024.103988

[39] B. A. Mohammed *et al.*, "Efficient blockchain-based pseudonym authentication scheme supporting revocation for 5G-assisted vehicular fog computing," *IEEE Access,* vol. 12, pp. 33089–33099, 2024. https://doi.org/10.1109/ACCESS.2024.3372390

[40] M. Khan and L. Ghafoor, "Adversarial machine learning in the context of network security: Challenges and solutions," *Journal of Computational Intelligence and Robotics,* vol. 4, no. 1, pp. 51–63, 2024.

[41] H. Waghela, J. Sen, and S. Rakshit, "Enhancing adversarial text attacks on bert models with projected gradient descent," *arXiv preprint arXiv:2407.21073,* 2024. https://doi.org/10.48550/arXiv.2407.21073

[42] Z. G. Al-Mekhlafi *et al.*, "Coherent taxonomy of vehicular ad hoc networks (vanets)-enabled by fog computing: A review," *IEEE Sensors Journal,* 2024. https://doi.org/10.1109/JSEN.2024.3436612

[43] H. Waghela, J. Sen, and S. Rakshit, "Robust image classification: Defensive strategies against FGSM and PGD adversarial attacks," *arXiv preprint arXiv:2408.13274,* 2024. https://doi.org/10.1109/ACOIT62457.2024.10941671

[44] B. Biggio, G. Fumera, and F. Roli, "Security evaluation of pattern classifiers under attack," *IEEE Transactions on Knowledge and Data Engineering,* vol. 26, no. 4, pp. 984-996, 2013. https://doi.org/10.1109/TKDE.2013.57

[45] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial machine learning at scale," *arXiv preprint arXiv:1611.01236,* 2016. https://doi.org/10.48550/arXiv.1611.01236

[46] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Computing Surveys (CSUR),* vol. 46, no. 4, pp. 1-37, 2014.

[47] P. Sehrawat and M. Chawla, "Prediction and analysis of machine learning models for efficient routing protocol in vanet using feature information," *Wireless Personal Communications,* vol. 136, no. 2, pp. 735-758, 2024.

[48] M. Al-Qatf, Y. Lasheng, M. Al-Habib, and K. Al-Sabahi, "Deep learning approach combining sparse autoencoder with SVM for network intrusion detection," *IEEE Access,* vol. 6, pp. 52843-52856, 2018.

[49] S. Rahman, S. Pal, S. Mittal, T. Chawla, and C. Karmakar, "SYN-GAN: A robust intrusion detection system using GAN-based synthetic data for IoT security," *Internet of Things,* vol. 26, p. 101212, 2024.