

Machine Learning for Intrusion Detection and Trust Management in Vehicular Ad-Hoc Networks: A Focused Review (2023-2025)

Mohammad Abbas Alkifae

College of Computer Science and Mathematics
University of Kufa
Najaf, Iraq

mohammada.alkifae@student.uokufa.edu.iq
<https://Orcid.org/0009-0004-3731-0886>

Fahad Ghalib Abdulkadhim

College of Computer Science and Mathematics
University of Kufa
Najaf, Iraq

fahadg.abdulkadhim@uokufa.edu.iq
<https://Orcid.org/0000-0002-4922-0878>

DOI: <http://dx.doi.org/10.31642/JoKMC/2018/130202>

Received Nov. 10, 2025. Accepted for publication Apr. 14, 2026

Abstract—Vehicular Ad-Hoc Networks (VANETs) provide the ultra-low-latency Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) links that underpin next-generation intelligent transportation systems. The same open wireless medium, however, renders VANETs vulnerable to a broad spectrum of cyber-attacks, from denial-of-service floods and replay injections to Sybil-based trust manipulation. While several surveys published before 2023 examined either intrusion detection or trust evaluation in isolation, none has jointly analysed both domains within the rapidly evolving machine-learning (ML) landscape. Addressing this gap, the present review offers the first integrated synthesis (2023-2025) of ML-driven intrusion detection and trust management for VANET security. State-of-the-art ML contributions are then organised into three focal tracks: (i) supervised and unsupervised intrusion-detection systems for V2X links; (ii) misbehaviour-, reputation-, and Sybil-node detection schemes that quantify inter-vehicle trust; and (iii) collaborative and federated learning approaches that distribute model training across vehicles and RSUs to conserve bandwidth and protect privacy. For each track, the review compares leading techniques with respect to datasets (e.g., VeReMi, SUMO/OMNeT++ traces), feature engineering pipelines, model families (CNN/LSTM, graph neural networks, hybrid ensembles), resource footprint, and reported performance, highlighting that many recent models exceed 95 % detection accuracy while maintaining real-time inference. A consolidated comparison table distils trade-offs among detection rate, false-positive rate, computational overhead, and explainability, thereby guiding practitioners in architecture selection. The analysis further exposes persistent gaps: scarcity of large, standardised datasets; susceptibility of ML models to physics-informed adversarial examples; limited frameworks for fusing intrusion evidence with trust scores; and the absence of lightweight yet adaptive models suitable for resource-constrained on-board units.

Keywords—Vehicular Ad-Hoc Networks (VANETs), Intrusion Detection, Trust Management, Sybil Attack Detection, Machine Learning, Deep Learning, V2X Security.

I. INTRODUCTION

Modern autonomous vehicles increasingly rely on vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications together forming vehicular ad-hoc networks (VANETs) to power advanced driver-assistance functions, real-time navigation, and over-the-air updates. By exchanging data cooperatively, connected vehicles benefit from augmented processing and situational awareness; yet this same connectivity widens the attack surface, exposing the network to cyber-attacks and privacy breaches. A typical VANET architecture consists of on-board units (OBUs) inside vehicles, roadside units (RSUs), and a back-end traffic-management centre, as depicted in Figure 1. While this multi-tier design boosts computational capability, it simultaneously multiplies the points at which adversaries can intervene.



Fig. 1: Representative Vehicular Ad-Hoc Network showing V2V links, RSU-mediated V2I links

High-profile incidents have already underscored the stakes: researchers remotely hacked a Jeep in 2015, seizing control of steering and braking and triggering a recall of 1.4 million vehicles; similar exploits against Tesla and BMW models have exposed critical vulnerabilities [1]. These events confirm that strong cybersecurity and privacy safeguards are prerequisites for the safe roll-out of autonomous vehicles. Traditional defences struggle to keep pace with VANETs' dynamic, high-mobility environment [2]. For example, the in-vehicle Controller Area Network (CAN) lacks native encryption or authentication, making it susceptible to spoofing and injection attacks. Likewise, rule-based intrusion-detection systems (IDS) dependent on fixed attack signatures cannot generalise to novel threats in time.

Machine learning (ML) provides a data-driven alternative: ML models, especially deep-learning architectures, excel at parsing the large, spatio-temporal datasets generated by vehicles, learning patterns that distinguish legitimate from malicious behaviour. With online learning and periodic retraining, these ML-based defences can adapt to emerging threats over time [3], [4]. For example, Yang et al. [1] conduct a semi-systematic review of 85 studies on ML for cybersecurity and privacy in connected vehicles. Alalwany and Mahgoub highlight that ML has "emerged as a potent solution" to IoV security and trust issues [4].

Building on these foundations, our survey synthesises state-of-the-art (2023 - 2025) ML techniques for intrusion detection and trust/misbehaviour management in VANETs. We examine security threats across the VANET spanning V2V, V2I, and broader V2X links and review ML-based defences against malware, denial-of-service, spoofing, and Sybil attacks. Lightweight models suitable for on-board units and federated-learning frameworks that preserve data privacy are emphasised. Our contributions are: (I) an overview of VANET communication models (V2V, V2I, RSU-assisted) and their security challenges (Section 2); (II) a structured review of ML-based VANET security solutions covering IDS, misbehaviour detection, trust management, and collaborative learning (Section 3); (III) a comparative analysis table contrasting datasets, accuracy, false-positive rates, and computational cost (Section 4); and (IV) a discussion of open issues such as adversarial robustness, real-time constraints, evaluation standardisation, and explainability (Section 5). By integrating insights from recent literature, this survey provides a reference for researchers and practitioners seeking to deploy robust ML-driven defences within VANETs, thereby enhancing the resilience of connected autonomous vehicles against evolving cyber-threats. Unlike broader prior surveys, this review focuses specifically on VANET studies published from 2023 to 2025 and jointly analyses intrusion detection, trust/misbehaviour detection, and collaborative/federated learning, while concluding with the main open problems and future research directions.

II. METHODOLOGY OF THE REVIEW

A. Scope and Research Questions

This survey investigates machine-learning-based intrusion detection and trust/misbehaviour management in vehicular ad-hoc networks (VANETs). It answers three research questions:

- RQ1: Which ML techniques have been applied to detect intrusions and evaluate trust in VANETs?
- RQ2: How are these techniques evaluated (datasets, metrics, testbeds) and how comparable are the reported results?
- RQ3: What gaps, open challenges, and future research directions remain?

B. Search Strategy and Data Sources

Works published 01 Jan 2023 - 30 Jun 2025 were retrieved from IEEE Xplore, ACM DL, Scopus, Web of Science, and arXiv. Leading journals (e.g., IEEE T-ITS, IEEE IoT-J, Computers & Security) and specialised workshops were hand-searched.

C. Inclusion and Exclusion Criteria

Inclusion criteria:

1. Peer-reviewed studies on VANET security (intrusion or misbehaviour).
2. ML forms the core of the approach.
3. Empirical evaluation with quantitative metrics.

Exclusion criteria:

1. Pure cryptographic protocols without ML.
2. Non-vehicular domains lacking clear transferability.
3. Editorials/opinion pieces.
4. Duplicates (retain final version).

D. Screening and Selection Process

The initial search retrieved 312 records. After deduplication, 257 unique items remained. Title and abstract screening against the inclusion/exclusion criteria reduced this set to 94 candidates. Full-text assessment yielded 25 studies for final inclusion.

E. Data Extraction and Coding Scheme

For each included study, we extracted: publication year and venue; targeted layer (in-vehicle vs. VANET/cloud); threat/attack model; ML technique (e.g., autoencoder, LSTM, GNN, FL); dataset or data source; evaluation metrics (Accuracy, F1, AUC, etc.); computational footprint (parameters, latency, memory); and key strengths and limitations. All fields were coded into a structured spreadsheet to support qualitative synthesis and basic quantitative aggregation.

III. BACKGROUND AND SECURITY CHALLENGES IN VEHICULAR NETWORKS

A. VANET Communication Architecture

A vehicular ad-hoc network (VANET) is a decentralised, short-range wireless system in which on-board units (OBUs) installed in vehicles exchange safety and service messages over vehicle-to-vehicle (V2V) links and interact with roadside units (RSUs) over vehicle-to-infrastructure (V2I) links. RSUs forward aggregated data to a traffic-management centre (TMC) or other back-office services, enabling applications such as collision avoidance, cooperative perception, and real-time traffic management.

A typical VANET communication model shows V2V broadcast beacons between vehicles and RSU-mediated V2I links toward a traffic-management centre. While this three-tier architecture (vehicle $\leftrightarrow$ RSU $\leftrightarrow$ TMC) delivers low-latency coordination, each wireless hop introduces new attack vectors that adversaries may exploit. The openness of V2V/V2I channels enlarges the attack surface: adversaries may broadcast forged safety messages (spoofing), create phantom vehicles (Sybil), jam the wireless spectrum (DoS), or tamper with RSUs to inject malicious data [1].

B. Threat Landscape in VANETs

The open, mobile, and peer-to-peer nature of VANETs substantially enlarges the attack surface, because safety beacons are broadcast frequently over a shared wireless medium and are consumed by nearby vehicles with limited time for verification. Key external threats include:

1. Message Spoofing / Fabrication. An attacker injects forged safety beacons (e.g., false hazard, speed, or position messages) to mislead neighbouring vehicles and manipulate their decisions.
2. Replay Attacks. Legitimate packets are captured and rebroadcast at a later time, producing stale or contradictory state information that can disrupt cooperative awareness.
3. Sybil Attacks. A single physical node forges and transmits under multiple identities, creating phantom vehicles that can bias density- and voting-based applications and distort trust/reputation aggregation.
4. Denial-of-Service (DoS / DDoS). Adversaries jam the wireless channel or flood RSUs/back-end services, degrading availability and delaying safety-critical messages.
5. Eavesdropping and Traffic Analysis. Weakly protected broadcasts allow passive observers to infer routes and identities, enabling tracking and privacy breaches.

High-profile demonstrations, such as the remote compromise of a Jeep Cherokee's steering and braking systems [1], confirm that practical exploits can translate into real safety risks. Conventional defences, static firewalls, signature-based intrusion rules, and standalone cryptographic checks often struggle to keep pace with VANETs' high mobility and rapidly changing topology [2].

C. Intrusion and Misbehaviour Detection in VANETs

To mitigate these threats, Intrusion Detection Systems (IDS) monitor V2X traffic for signs of attacks, while Misbehaviour Detection Systems (MDS) assign scores to participating nodes, filtering entities that disseminate false or malicious information. Traditional signature-based IDS cannot recognise previously unseen attacks in real time; hence, machine-learning techniques have become prominent.

Recent studies deploy convolutional neural networks (CNNs), recurrent networks such as LSTMs, graph neural networks (GNNs), and ensemble models to classify packets or aggregate reputation evidence, regularly exceeding 95 % detection accuracy on benchmark datasets like VeReMi [3]. Lightweight model-compression and federated learning further allow OBUs and RSUs to train models collaboratively without sharing raw data, preserving privacy and bandwidth while adapting to evolving attack patterns [4]. Subsequent sections of this survey analyse these ML-based IDS/MDS solutions in detail, compare their evaluation protocols, and highlight open research challenges such as adversarial robustness, standardised benchmarking, and on-board resource constraints.

IV. ML-BASED SECURITY SOLUTIONS FOR VEHICULAR NETWORKS

The 2023-2025 literature on vehicular security shows a decisive shift from static, rule-centric defences toward data-driven, machine-learning (ML) pipelines that sense, decide, and react within milliseconds even while vehicles hurtle across cell boundaries at highway speed. To present a coherent picture, we organise the state of the art into four tightly coupled threads:

- Intrusion-detection for V2X traffic - packet-level or flow-level models that spot spoofing, replay, DoS, and advanced multi-stage campaigns on the air interface.
- Misbehaviour and trust management - behavioural analytics that expose Sybil identities, selfish relays, and reputation fraud that cryptography alone cannot stop.
- Federated and fully-distributed learning - privacy-preserving training topologies that exploit RSUs or peer-to-peer gossip to overcome intermittent links and non-IID data.
- Performance-oriented design - compression, feature pruning, and adaptive thresholds that make sophisticated models viable on resource-starved on-board units (OBUs) and roadside units (RSUs).

Table 1 later in this section distils the salient traits of each cited work, including the dataset, model family, peak metric, deployment target, and the main contribution and one concise limitation.

A. Intrusion Detection for V2X Links

The canonical VANET IDS pipeline now begins with in situ feature engineering (raw IEEE 802.11p beacons, C-V2X sidelink logs, or IP flows from infotainment hotspots), progresses through a sequence-aware model, and ends with

thresholding or score fusion. Jayasri et al. [2] fuse Discrete-Fourier spectra with an LSTM whose hyper-parameters are tuned by the Crocodile-Optimisation Algorithm; the resulting ILSTM-COA posts 98.9 % accuracy on UNSW-NB15 while slashing false alarms compared with untuned baselines [2]. Ababsa et al. [9] escalate temporal learning one step further, interleaving LSTM (long-horizon) and GRU (short-burst) channels under an attention gate; training is hardened with adversarial examples so the detector retains ≈ 96.6 % accuracy under real-time DDoS in VeReMi-derived traces [9].

Speed and interpretability remain non-negotiable. Elsadig et al. [10] marry information-gain pruning with an 80-tree Random Forest, squeezing inference into sub-millisecond windows on an ARM RSU while still hitting 99.8 % accuracy on a re-balanced slice of CICIDS-2017 [10]. When the traffic is wildly imbalanced, pre-processing sometimes matters more than network depth. Some studies show that NearMiss and Tomek-Link resampling can push a plain k-NN to a perfect detection score on minority car-hacking attacks that defeat deeper models trained on skewed data [11].

Dedicated DoS/DDoS studies echo the same pattern: compress features with Random Projection and Randomised Matrix Factorisation before feeding an ExtraTrees ensemble, realising ≈ 98 % accuracy on IoV-specific DDoS bursts [12]. Babbar et al. [14] operate inside a software-defined VANET, where entropy-driven Random Forests flag reconnaissance, single-source, and distributed floods at ≈ 91 % accuracy / 92 % F1 while adding negligible edge latency [14]. A rising design philosophy is the cascaded hybrid. Some approaches front-load a lightweight CNN that filters the obvious benign flows, then unleash a self-attention Bi-LSTM for the ambiguous residue. A multi-variant meta-heuristic tunes both stages, tallying ≈ 99 % accuracy across VeReMi, KDD-Cup 99, and ToN-IoT with 40 % less inference cost than a monolithic deep stack [13]. Taken together, these works confirm the value of sequence-aware architectures.

B. Misbehaviour Detection, Trust, and Sybil Mitigation

Whereas IDS looks for illicit bits, misbehaviour analytics hunt for illicit behaviours. The archetypal threat is the Sybil attack, in which one vehicle spawns dozens of phantom IDs to sway majority-vote protocols or falsify congestion maps. Zhu et al. [16] embed vehicles, RSUs, and beacon links into a time-evolving heterogeneous graph; an attention kernel spots abnormal co-movement, achieving ≈ 97 % Sybil detection and even back-tracing the physical perpetrator for revocation [16].

Trust scores can also emerge bottom-up. One study introduces TASER, where every car maintains an on-board trust ledger and cross-checks neighbours via directional-antenna challenges; simulations with a 30 % Sybil population show TASER shortens detection delay by 66 % all without leaning on RSU infrastructure [17]. On the ledger side, another study entwines a Schnorr/PUF key-agreement with a light consortium blockchain, yielding BASF-ITS that blocks replay, impersonation, and Sybil in < 2 ms crypto overhead while continuously adjusting reputations [18]. Collectively, these efforts pivot from packet syntax to entity semantics, a prerequisite for truly autonomous, self-healing fleets.

C. Federated and Distributed Learning for VANET Security

With tens of millions of vehicles generating proprietary telemetry, no single data repository can see the whole threat surface, making Federated Learning (FL) an architectural necessity rather than a research novelty. However, the effectiveness of any FL implementation in a vehicular context is critically dependent on the underlying physical infrastructure, specifically the layout of Roadside Units (RSUs). An optimised RSU deployment is essential for ensuring the reliable connectivity required for model aggregation. Abdulkadhim et al. [19] focus on creating systematic deployment mechanisms that balance network coverage with deployment costs (CAPEX), thereby mitigating data-loss risks and establishing a stable foundation for scalable FL.

Building upon such an optimised infrastructure, advanced architectures like Hierarchical FL have been proposed to enhance efficiency and robustness. One prominent approach introduces a two-tier aggregation loop where On-Board Units (OBUs) first push model updates to their local RSUs. These RSUs perform a preliminary merge before forwarding the aggregated models to a global cloud orchestrator [20]. This hierarchical structure was shown to elevate the global model's F1 score by approximately 10 percentage points under non-IID data conditions and removed the single-point-of-failure risk inherent in traditional flat FL architectures [20].

Not all optimisation must live in the gradient. Rashid et al. [21] blend four shallow classifiers (LR, SVM, RF, GBT) behind density-aware threshold tuning; on bursty DDoS they sustain 97-99 % accuracy while shedding 40 % of false alarms relative to static cut-offs [21]. At the frontier, some studies drop the server entirely: neighbouring cars swap model updates hop-by-hop in a gossip routine dubbed Distributed FL (DFL). On VeReMi-Extension, low-data vehicles gain > 15 percentage points, but poisoning and jamming reveal the need for robust consensus filters [22]. The message is clear: hierarchical, adaptive, and server-less FL variants are converging toward a privacy-preserving, bandwidth-savvy learning fabric.

D. Performance Optimisation and Lightweight Design

Accuracy alone cannot keep a safety-critical IDS alive if it starves the ECU of cycles or drains the 12-V rail. Three engineering levers dominate recent work. Feature / model compression: some studies apply correlation filtering + PCA, then quantise a DNN, shrinking memory from 28 kB to 9 kB while raising F1 from 96.5 % to 98.4 % [23].

- End-to-end AutoML. Alotaibi et al. co-optimize a Conv-BiLSTM and its feature selector in two cascaded meta-heuristics, peaking at 98.5 % accuracy and converging faster than one-shot tuning [24].
- Real-time RSU inference. Haydari and Yilmaz install an adaptive-threshold LSTM autoencoder on an RSU ARM core; live tests sustain 97-99 % detection without perceptible V2V latency, matching CAN-bus rates [25].

The collective evidence from Sections 4.1 through 4.4 paints a vivid trajectory for ML-driven vehicular security. First,

the research community now recognises that temporal context is king: virtually every high-performing IDS or MDS injects some form of sequence awareness, whether through ILSTM, hybrid LSTM-GRU stacks, or temporal graph attention. Second, trust is no longer a bolt-on afterthought. Graph-based reasoning, reputation ledgers, and Sybil-traceability modules are being woven directly into detection pipelines, converting one-off alerts into richer behavioural verdicts that can drive cooperative counter-measures at fleet scale. Third, privacy-preserving learning topologies have left the prototype stage: hierarchical FL and server-less DFL not only safeguard proprietary telemetry, but demonstrably raise F1 versus monolithic training by mitigating non-IID drift an insight that will matter even more once OEMs begin geo-fencing OTA model updates.

E. Adversarial Robustness

Despite their high detection accuracy under controlled conditions, many ML-based IDSs remain vulnerable to adversarial inputs, small, crafted perturbations that preserve semantic plausibility while misleading the detector. A particularly concerning class involves physics-informed adversarial examples, in which the attacker subtly manipulates input features without violating physical constraints.

Consider a V2V beacon containing GPS coordinates, speed, and heading. A malicious vehicle could inject a location offset of +2.4 meters, carefully calibrated to mimic typical GNSS drift. This perturbation is too small to trigger map-based rejection, yet large enough to shift a spatio-temporal graph model's embedding vector, leading the IDS to classify the node as benign even if its trajectory contradicts legitimate traffic flows. Similarly, in CAN-bus scenarios, attackers can modulate throttle sensor values within the accepted operational range, e.g., oscillating between 41-43 % to poison LSTM models trained on tightly regulated drive cycles. These changes remain undetectable by rules-based filters or even cryptographic authentication, as the packet payloads are syntactically valid.

Empirical studies report up to 42 % degradation in detection accuracy when such adversarial samples are introduced during live inference [9]. While adversarial training mitigates this effect, it comes at the cost of model complexity and generalisation. Furthermore, real-time adversarial defence remains challenging in OBU environments due to latency and compute constraints. Addressing this threat surface demands more than data augmentation: it requires embedding domain constraints directly into the model architecture (e.g., physics-informed priors), real-time adversarial sample detectors, and robust uncertainty quantification to flag out-of-distribution patterns.

On the engineering front, the literature shows a pragmatic shift from accuracy at all costs toward multi-objective optimisation: kilobytes per percentage-point, joules per inference, and milliseconds per ROC-AUC. Techniques such as depthwise separable convolutions, PCA-assisted feature trimming, quantised weights, and AutoML-guided co-design routinely slash memory footprints by 50-70 % while improving detection metrics, a reminder that algorithmic elegance and

hardware realism can coexist. Yet several gaps threaten to stall real-world adoption:

1. Adversarial robustness - Only a handful of works inject adversarial traffic during training; few test black-box transfer or poisoning at FL aggregators.
2. Benchmark fragmentation - UNSW-NB15, CICIDS, VeReMi, and bespoke CAN traces are rarely cross-validated, making headline accuracies hard to compare. A unifying, annotated, open-source VANET corpus is overdue.
3. End-to-end field trials - Simulated DSRC channels and synthetic DDoS bursts do not capture cellular hand-offs, urban canyon multipath, or driver evasive reactions. Multi-OEM testbeds and long-haul validation remain necessary.

V. COMPARATIVE ANALYSIS OF ML-BASED IDS APPROACHES

Table 1 summarises representative ML-based security solutions for vehicular networks published between 2023 and 2025. For each included study, we report the protected layer (in-vehicle CAN, VANET/V2X, or SD-VANET edge), the primary attack focus, the evaluation dataset or testbed, the core methodology/model family, the reported performance metrics (e.g., Accuracy, Precision/Recall, F1, AUC), and the main contribution and one concise limitation of each approach. This consolidated view enables direct comparison across studies and highlights practical trade-offs between detection effectiveness, false-alarm behaviour, and computational footprint relevant to real-world deployment. Results obtained on proxy IT datasets (e.g., UNSW-NB15 and similar non-vehicular traffic corpora) are reported for completeness but should not be used as direct evidence of performance under VANET dynamics, since such datasets do not capture core vehicular properties such as high mobility, rapidly changing topology, and beacon semantics. Accordingly, we prioritise conclusions drawn from VANET-specific datasets and testbeds (e.g., VeReMi, SUMO/OMNeT++ traces, and real vehicular traces), and we interpret proxy-dataset results with explicitly limited transferability to real VANET settings.

TABLE 1. SUMMARY OF SELECTED ML-BASED INTRUSION DETECTION SYSTEMS.

<i>Ref</i>	<i>Security Layer</i>	<i>Primary Attack Focus</i>	<i>Dataset / Testbed</i>	<i>Core Model Family</i>	<i>Best Reported Metric*</i>	<i>Contributions & Limitations</i>
[2] Jayasri 2025	VANET / V2X	Mixed (DoS, Probe, etc.)	UNSW-NB15 traffic flows	DFT-transformed features + LSTM classifier; Crocodile Optimisation Algorithm (COA) for hyperparameter tuning	Accuracy 98.9 %	Frequency-domain pre-processing; nature-inspired optimiser boosts detection and reduces false alarms. Limitation: Proxy IT dataset (UNSW-NB15).
[6] Yang 2025	In-vehicle (CAN)	DoS, Spoofing	CAN injection dataset (lab vehicle)	Triple-attention deep network for CAN intrusion detection; Grey-lag Goose Optimisation (GGO) for parameter/architecture tuning	Accuracy 96.3 %	Three-stage attention highlights salient bytes; grey-lag goose optimisation improves generalisation. Limitation: Lab-vehicle CAN dataset only.
[5] Zhao 2025	In-vehicle (CAN)	DoS, Spoofing, Fuzzy	Real car traces; attack classes balanced via SN-GAN	DSC-KAN (depthwise-separable CNN + Kolmogorov-Arnold Network); SN-GAN-based class balancing/augmentation	Accuracy ≥ 99 %	Ultra-light CNN layers; synthetic minority over-sampling keeps the model tiny yet precise. Limitation: Relies on SN-GAN balancing.
[8] Jeong 2024	In-vehicle (CAN)	Multi-class intrusion	X-CANIDS benchmark	CNN-based CAN IDS + signal-aware knowledge graph (KG) for explainable detection	F1 ≈ 98 %	Explainable IDS; graph filters impossible signal relations, lowering false positives. Limitation: In-vehicle CAN benchmark only.
[14] Babbar 2024	SD-VANET (edge)	Random Forest + Entropy flow stats	BoT-IoT (vehicular flows)	Entropy/statistical flow-feature extraction + Random Forest; SDN-integrated detection/mitigation pipeline	Accuracy 91 %; F1 92 %	Statistical-flow pre-filtering; full SDN loop including mitigation logic. Limitation: Limited real-road VANET validation.
[10] Elsadig 2025	VANET	Composite VANET attacks	CICIDS-2017 (re-balanced)	Information-gain feature selection + oversampling/re-balancing + lightweight Random Forest for VANET attack detection	Accuracy 99.8 %	Information-gain feature pruning + oversampling; millisecond inference on OBUs/RSUs. Limitation: Re-balanced CICIDS-2017 proxy data.
[16] Zhu 2024	VANET	Sybil attacks & traceability	Beacon-packet simulation	Temporal heterogeneous graph attention network (Hetero-GAT) over beacon interactions for Sybil detection/traceability	Detection ≈ 97 %; F1 ≈ 98 %	Locates the physical source of Sybil behaviour via time-aware graph embeddings. Limitation: Simulation-based evaluation.
[20] Althunayyan 2024	In-vehicle (CAN) + VANET (mixed)	Multi-class intrusion	CAN & CIC-IDS splits (non-IID)	Hierarchical Federated Learning (OBU $\rightarrow$ RSU $\rightarrow$ cloud aggregation) with ANN + LSTM Autoencoder under non-IID splits	-	RSU-level aggregation reduces non-IID drift and eliminates the single cloud bottleneck. Limitation: No single headline metric reported.
[21] Rashid 2023	VANET	Malicious node / Misbehaviour	Custom VANET simulator	Shallow ML ensemble (LR/SVM/RF/GBT) with adaptive, density-aware thresholding for misbehaviour / malicious-node detection	Accuracy 97 - 99 %	Density-aware thresholding cuts false alarms by 40 % versus fixed thresholds. Limitation: Custom simulator only.
[9] Ababsá 2025	VANET / IoV	Real-time DDoS	Extended VeReMi beacon traces (multi-modal)	Multimodal LSTM+GRU fusion with attention; adversarial training for real-time DDoS robustness	Accuracy 96.6 %	Fusion of long- and short-term RNN paths; adversarial robustness for on-line detection. Limitation: Limited field-scale evidence.

Analysis. As Table 1 and the heatmap in Figure 2 jointly suggest, there is no single “one-size-fits-all” design for vehicular cyber-defence because deployment constraints differ sharply across layers. At the in-vehicle level, CAN-bus monitoring runs on safety-critical ECUs where intrusion-detection workloads must coexist with real-time control tasks (e.g., braking and airbag logic). Consequently, effective in-vehicle IDS designs tend to prioritise lightweight architectures and efficiency-oriented techniques such as compact CNN variants (including depthwise-separable designs), autoencoder-based detectors, and model/feature compression to reduce runtime overhead while maintaining detection performance and acceptable false-alarm behaviour. By contrast, VANET/V2X monitoring at RSUs and edge servers typically offers more computational headroom, enabling richer pipelines such as multimodal LSTM-GRU fusions with attention, entropy-guided ensemble methods, and hierarchical federated-learning schemes that aggregate models across regions to better handle non-IID data. This additional capacity broadens coverage to more complex threats (e.g., Sybil swarms, reconnaissance activity, and mixed DDoS patterns) that are harder to capture with strictly resource-minimal in-vehicle models.

What unifies both tiers is a hybrid, defence-in-depth philosophy. Cryptographic certificates still police basic admission, but ML layers continue monitoring after authentication: forged keys may borrow an identity, yet they rarely reproduce realistic beacon timing, trajectory continuity, or cross-sensor correlations for long. Together, the crypto + ML stack closes both the who-are-you? and the what-are-you-doing? gaps. Explainability is emerging as an equally decisive factor. Systems such as X-CANIDS use signal-aware knowledge graphs; graph-attention models trace Sybil origins; SHAP-attributed gradient boosting reveals which bytes or intervals fire an alarm. Regulators and OEM safety boards increasingly insist on why an alert triggers, not merely that it triggers both for ISO 26262/21434 audits and for driver trust.

The field’s Achilles’ heel remains benchmark fragmentation. Figure 2 shows five different headline metrics Accuracy, Detection Rate, F1, AUC, even raw Inference Time reported on mutually exclusive datasets (UNSW-NB15, VeReMi, bespoke CAN captures). Without a unified corpus and metric suite, 99 % on Dataset A can be less informative than 92 % on a harder, real-world trace. The ongoing push toward integrated mobility-and-network simulators, open CAN/FlexRay dumps, and city-scale V2X recordings is therefore critical: only standardised, cross-layer benchmarks will let the community separate truly generalisable defences from laboratory curiosities and speed the journey toward resilient, self-defending transportation networks.

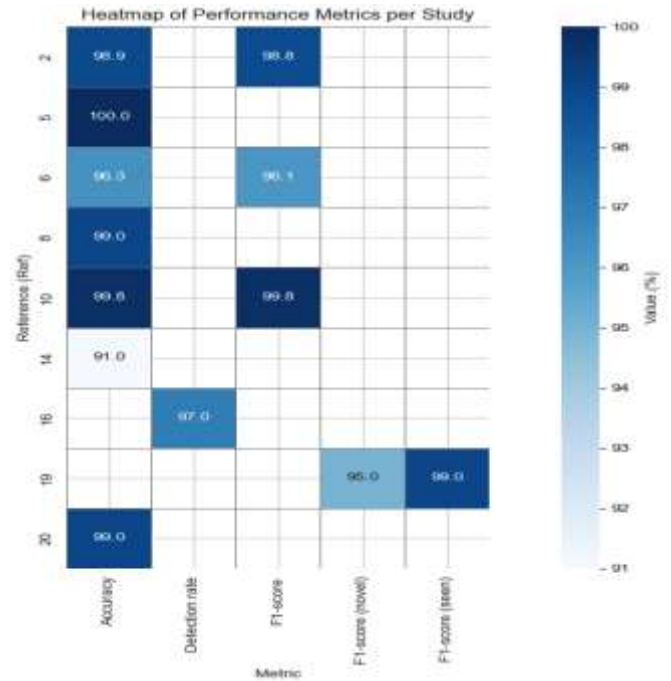


Fig 2. Heatmap of reported performance metrics.

VI. FUTURE RESEARCH DIRECTIONS

1. Adversarial Robustness - Even today’s best-performing detectors, whether depth-wise separable CNNs such as DSC-KAN or triple-attention networks, are evaluated almost exclusively under benign test conditions. Future work must harden models against carefully crafted CAN frames that exploit gradient weaknesses and poisoning of federated gradients. Promising directions include adversarial training with physics-informed constraints, ensemble diversity, and on-line outlier re-weighting.

2. Hard Real-Time Scalability - A spoofed brake beacon must be flagged within ≤ 50 ms (ETSI EN 302 637-2 safety window). While sub-megabyte models protect a single ECU, gigabit-scale 5G/6G V2X telemetry from millions of vehicles will require multi-tier pipelines in which OBUs raise instant alerts, RSUs correlate regionally, and cloud services mine long-horizon trends. Systematic latency budgets and queue-aware scheduling are still rare in the literature.

3. Benchmark and Interoperability - Results span UNSW-NB15, CIC-IDS2017, VeReMi, bespoke CAN traces, and SDN flow captures, rendering headline accuracies non-comparable. A cross-dataset leaderboard that covers message spoofing, Sybil, DoS/DDoS, and generative attacks and uses a common alert schema would provide the reproducibility baseline the field lacks.

4. Privacy-Preserving Learning - Hierarchical and server-less federated schemes lower raw-data exposure but can still leak information via model updates. Differentially-private optimisation, secure Shamir secret-sharing aggregation, homomorphic inference, and token- or blockchain-based

incentive mechanisms are needed to encourage OEM and driver participation without compromising privacy.

5. Holistic, Multi-Layer Fusion - True defence-in-depth couples in-vehicle anomaly scores with VANET misbehaviour analytics and ledger-based reputation. Research should move beyond simple score averaging toward evidence-level fusion: if a CAN auto-encoder and a graph-based Sybil tracker both raise low-grade suspicion, their joint evidence ought to escalate automatically to a high-confidence alert.

6. Mitigation and Safe Response - Detection is only half the battle. Reinforcement-learning agents could orchestrate tiered responses, rate-limiting, certificate revocation, or lane-level re-routing while respecting ISO/PAS 21448 constraints. Few studies quantify collateral safety risks or recovery latencies after an automated quarantine.

7. Digital-Twin and Field Validation - Simulators accelerate prototyping, yet they omit cellular hand-offs, driver reactions, and multi-vendor interoperability. Edge-to-cloud digital twins should precede large-scale road pilots, replaying city-scale attack traces in deterministic fashion. Joint academia-industry-municipality efforts are needed to collect, anonymise, and share such traces.

8. Multi-Stakeholder Governance - Technical advances intersect policy and liability. IDS outputs must remain explainable and auditable to satisfy UNECE WP.29 and ISO 21434 cybersecurity regulations. Collaboration with legal and regulatory experts will ensure that blockchain-assisted key management, trust scoring, and automated mitigation align with data-protection and road-safety mandates.

Progress along these eight axes, robustness, latency-aware scalability, privacy, benchmark standardisation, layered fusion, safe mitigation, field validation, and governance, will accelerate the transition from isolated prototypes to cooperative, trustworthy V2X ecosystems capable of safeguarding next-generation connected and autonomous vehicles.

VII. CONCLUSION

Connected and autonomous vehicles promise transformative gains in road safety, traffic efficiency, and environmental impact, yet their pervasive connectivity to vehicle-to-everything (V2X) networks, cloud back-ends, and over-the-air update channels simultaneously creates an expansive and rapidly evolving cyber-attack surface. As documented throughout this survey, machine-learning techniques have progressed from isolated academic prototypes to increasingly practical components of the emerging vehicular-security stack.

Across the in-vehicle and VANET/V2X layers, recent work consistently leverages temporal deep models and graph-based reasoning for behaviour-level threats such as Sybil attacks, while federated and collaborative learning topologies reduce raw-data exposure and help address non-IID training conditions. Complementary trust and authentication mechanisms further reinforce a defence-in-depth posture by

constraining misbehaving entities and supporting accountability.

Despite these advances, several barriers still limit fleet-scale deployment. Adversarial robustness is under-explored, with relatively few studies evaluating physics-plausible perturbations or poisoning risks in collaborative training. Evaluation practices remain fragmented across datasets and metrics, and results obtained on proxy IT datasets do not reliably translate to VANET dynamics, highlighting the need for VANET-native, standardised benchmarks and clearer reporting of deployment-oriented constraints for safety-critical settings. Moreover, explainable AI is increasingly necessary to support operational debugging and auditability. Looking forward, progress will depend on aligning benchmark suites with field-representative testbeds, strengthening robustness under realistic threat models, and enabling cross-layer fusion that combines intrusion evidence with trust signals. Addressing these gaps will accelerate the transition from laboratory demonstrations to deployable, trustworthy security mechanisms for next-generation vehicular networks.

REFERENCES

- [1] T. Yang, R. Sun, R. S. Rathore, and I. Baig, "Enhancing Cybersecurity and Privacy Protection for Cloud Computing-Assisted Vehicular Network of Autonomous Electric Vehicles: Applications of Machine Learning," *World Electric Vehicle Journal*, vol. 16, no. 1, p. 14, Dec. 2024, doi: <https://doi.org/10.3390/wevj16010014>.
- [2] C. Jayasri, V. Balaji, C. M. Nalayani, and S. Pradeep, "Detecting cyber attacks in vehicle networks using improved LSTM based optimization methodology," *Scientific Reports*, vol. 15, no. 1, May 2025, doi: <https://doi.org/10.1038/s41598-025-04643-8>.
- [3] R. Rai, J. Grover, P. Sharma, and A. Pareek, "Securing the CAN bus using deep learning for intrusion detection in vehicles," *Scientific Reports*, vol. 15, no. 1, Apr. 2025, doi: <https://doi.org/10.1038/s41598-025-98433-x>.
- [4] E. Alalwany and I. Mahgoub, "Security and Trust Management in the Internet of Vehicles (IoV): Challenges and Machine Learning Solutions," *Sensors*, vol. 24, no. 2, p. 368, Jan. 2024, doi: <https://doi.org/10.3390/s24020368>.
- [5] W. Zhao, Y. Yang, H. Hu, Y. Chen, and F. Yu, "A lightweight intrusion detection approach for CAN bus using depthwise separable convolutional Kolmogorov Arnold network," *Scientific Reports*, vol. 15, no. 1, May 2025, doi: <https://doi.org/10.1038/s41598-025-02474-1>.
- [6] H. Yang and M. Effatparvar, "A deep learning based intrusion detection system for CAN vehicle based on combination of triple attention mechanism and GGO algorithm," *Scientific Reports*, vol. 15, no. 1, Jun. 2025, doi: <https://doi.org/10.1038/s41598-025-04720-y>.
- [7] Meghana R, Sowmyashree Sakrepatna Ramesha, and A. Mukhopadhyay, "QCAE-QOC-SVM: A Hybrid Quantum Machine Learning Model for DoS and Fuzzy Attack Detection on Autonomous Vehicle CAN Bus," *MethodsX*, vol. 15, no. 103471, pp. 103471-103471, Jun. 2025, doi: <https://doi.org/10.1016/j.mex.2025.103471>.
- [8] S. Jeong, S. Lee, H. Lee, and H. K. Kim, "X-CANIDS: Signal-Aware Explainable Intrusion Detection System for Controller Area Network-Based In-Vehicle Network," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 3, pp. 3230-3246, Mar. 2024, doi: <https://doi.org/10.1109/tvt.2023.3327275>.
- [9] M. Ababsá, S. Ribouh, A. Malki, and L. Khoukhi, "Deep Multimodal Learning for Real-Time DDoS Attacks Detection in

- Internet of Vehicles,” ICC 2025, pp. 3039-3044, doi: <https://doi.org/10.1109/ICC52391.2025.11160909>.
- [10] M. A. Elsadig et al., “Connected Vehicles Security: A Lightweight Machine Learning Model to Detect VANET Attacks,” World Electric Vehicle Journal, vol. 16, no. 6, p. 324, Jun. 2025, doi: <https://doi.org/10.3390/wevj16060324>.
- [11] S. Alshathri, A. Sayed, and E. E.-D. Hemdan, “An Intelligent Attack Detection Framework for the Internet of Autonomous Vehicles with Imbalanced Car Hacking Data,” World Electric Vehicle Journal, vol. 15, no. 8, p. 356, Aug. 2024, doi: <https://doi.org/10.3390/wevj15080356>.
- [12] N. Ahmed, F. Hassan, K. Aurangzeb, A. H. Magsi, and M. Alhussein, “Advanced machine learning approach for DoS attack resilience in internet of vehicles security,” Heliyon, vol. 10, no. 8, p. e28844, Apr. 2024, doi: <https://doi.org/10.1016/j.heliyon.2024.e28844>.
- [13] A. Manderna, S. Kumar, U. Dohare, M. Aljaidi, O. Kaiwartya, and J. Lloret, “Vehicular Network Intrusion Detection Using a Cascaded Deep Learning Approach with Multi-Variant Metaheuristic,” Sensors, vol. 23, no. 21, p. 8772, Oct. 2023, doi: <https://doi.org/10.3390/s23218772>.
- [14] H. Babbar, S. Rani, and M. Driss, “Effective DDoS attack detection in software-defined vehicular networks using statistical flow analysis and machine learning,” PLOS ONE, vol. 19, no. 12, p. e0314695, Dec. 2024, doi: <https://doi.org/10.1371/journal.pone.0314695>.
- [15] I. Mahmoudi, D. Boubiche, S. Athmani, H. Toral-Cruz, and F. Chan-Puc, “Academic Editors: Olusola Tolulope Odeyomi and Temitayo Olowu Toward Generative AI-Based Intrusion Detection Systems for the Internet of Vehicles (IoV),” Future Internet, vol. 17, no. 310, 2025, doi: <https://doi.org/10.3390/fi17070310>.
- [16] Y. Zhu et al., “Sybil Attacks Detection and Traceability Mechanism Based on Beacon Packets in Connected Automobile Vehicles,” Sensors, vol. 24, no. 7, p. 2153, Jan. 2024, doi: <https://doi.org/10.3390/s24072153>.
- [17] T. L. Morton, A. Borah, and A. Paranjothi, “Trust-Aware Sybil Attack Detection for Resilient Vehicular Communication,” Internet Technology Letters, Nov. 2024, doi: <https://doi.org/10.1002/itl2.617>.
- [18] A. Badshah et al., “Blockchain-Assisted Lightweight Authenticated Key Agreement Security Framework for Smart Vehicles-Enabled Intelligent Transportation System,” IEEE Transactions on Automation Science and Engineering, vol. 21, no. 3, pp. 2425-2439, Jul. 2024, doi: <https://doi.org/10.1109/tase.2024.3381068>.
- [19] F. G. Abdulkadhim, Z. Yi, C. Tang, A. N. Onaizah and A. H. Radie, “Optimizing Roadside via Unit Deployment Mechanism in V ANET,” 2020 3rd International Conference on Engineering Technology and its Applications (IICETA), Najaf, Iraq, 2020, pp. 144-149, doi: <https://doi.org/10.1109/IICETA50496.2020.9318821>.
- [20] Muzun Althunayyan, A. Javed, and O. Rana, “A robust multi-stage intrusion detection system for in-vehicle network security using hierarchical federated learning,” Vehicular Communications, vol. 49, pp. 100837-100837, Aug. 2024, doi: <https://doi.org/10.1016/j.vehcom.2024.100837>.
- [21] K. Rashid, Y. Saeed, A. Ali, F. Jamil, R. Alkanhel, and A. Muthanna, “An Adaptive Real-Time Malicious Node Detection Framework Using Machine Learning in Vehicular Ad-Hoc Networks (VANETs),” Sensors, vol. 23, no. 5, p. 2594, Feb. 2023, doi: <https://doi.org/10.3390/s23052594>.
- [22] U. Demir, Y. Sagduyu, T. Erpek, H. Jafari, S. Kompella, and M. Xue, “Distributed Federated Learning for Vehicular Network Security: Anomaly Detection Benefits and Multi-Domain Attack Threats,” 2025, doi: <https://doi.org/10.1145/3733965.3733967>.
- [23] F. Hassan et al., “A hybrid approach for intrusion detection in vehicular networks using feature selection and dimensionality reduction with optimized deep learning,” PLOS ONE, vol. 20, no. 2, p. e0312752, Feb. 2025, doi: <https://doi.org/10.1371/journal.pone.0312752>.
- [24] M. Alotaibi et al., “Integrating Two-Tier Optimization Algorithm with Convolutional Bi-LSTM Model for Robust Anomaly Detection in Autonomous Vehicles,” IEEE Access, pp. 1-1, Jan. 2024, doi: <https://doi.org/10.1109/access.2024.3523539>.
- [25] A. Haydari and Y. Yilmaz, “RSU-Based Online Intrusion Detection and Mitigation for VANET,” Sensors, vol. 22, no. 19, p. 7612, Oct. 2022, doi: <https://doi.org/10.3390/s22197612>.