



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

VOL 04. ISSUE 01 . (2026)

<http://ijirjournal.com/>

SENTINEL-WHEEL: Entropy-Compressed Edge Intelligence for Explainable Self-Healing Cyber Defense in Connected Vehicles

Dr. Ajab Khan

ORIC, Abbottabad University of Science and Technology, Pakistan

ajabkhan@uom.edu.pk

Fahad Amin, Member, IEEE

North American University, Department of Computer Science-Cybersecurity, Houston, TX, USA

famin1@na.edu

Usman Imtiaz , Member, IEEE (Corresponding Author)

Department of Computer Science and Cybersecurity, Washington University of Science and Technology (WUST), Virginia, USA

usmanimtiaz1992@gmail.com

Abstract.

Connected vehicle data can flow over in-vehicle buses, but also through V2X-like wireless and wired interconnects, of which only a few were originally built considering today's adversaries. We introduce the SENTINEL-WHEEL architecture for ultra-fast cyber-event identification and responsive actions in connected vehicles. The design exploits entropy-compressed feature extraction to compress bus data and ignore features in V2X messages and employs a temporal micro-attention classifier that analyzes short windows of messages, eschewing the use of large, complex models. A risk ladder transforms classifier decisions into appropriate, granular responses such as silent logging, sender observation, message authentication, rate limiting, channel disconnection, or safety-envelope actuation. Instead of simply delivering a malicious-or-normal classification, the architecture maintains a trace ledger that explains unusual events in terms of interpretable origins like rate excessive traffic, frequent, jittered transmissions, the misuse of previously unseen vehicle identifier or a discrepancy between current speed/acceleration profile and observed traffic behavior. We present an experiment setup and evaluate typical vehicular traffic. Our initial experimental findings achieve up to 98.6% detection accuracy with an F1 score of 98.2%, meaning detection delay down to 4.7 ms and utilize a mere under a million trainable parameters. Our work outlines a route towards an explainable, scalable on-device cybersecurity solution for the evolving automotive domain.

Keywords: Connected vehicles; automotive cybersecurity; edge intrusion detection; feature compression; explainable AI; CAN bus; V2X; self-healing response

1. Introduction

Connected and automated vehicle functionalities are now predominantly based on software driven decision making. Breaking, steering assistance, engine and motor management, infotainment, navigation and shared road safety applications all rely on message exchange among multiple electronic control units (ECUs) over electronic buses and wirelessly [1,4]. This shift from mechanical control to electronically mediated operations increase driver convenience and road safety but fundamentally turns a vehicle into a cyber-physical system that could be physically endangered by a malicious message [1], [2,16,18]. This central challenge is not simply to determine whether an incoming message is malicious, however effective vehicle cyber defense must accomplish four simultaneous, seemingly conflicting tasks: a security system must be computationally lightweight for in-vehicle, embedded deployment on low power compute (where it will compete with safety critical perception and control tasks) [20,23], responsive under the low latency constraints of vehicles [16,17], provide interpretable, trustworthy reasons for alarms, as neural-network classification merely indicates threat probability [16,18,24] and, critically, must proactively establish a mechanism for determining a safe system state in response to identified threats instead of leaving a car running under compromised operation [16,19]. This proposal builds upon the existing notion of low overhead, for an intelligent connected vehicle network of the type. We extend the core idea of this proposal by refining the underlying concepts, while reorganizing the system. Key concepts in our work include features reduced learning with short history, learning of temporal patterns and risk-based response, this is built upon, reframed under a new structure. We revise the methodology and the relevant experiment setup and data representations. As for the terminology of the article as well, our contributions are defined and implemented, from the very starting points; the title, architecture description of the presented algorithm and description of its steps are rewritten without being directly dependent on the existing paper. In this paper, we propose SENTINEL-WHEEL as an in-vehicle cyber defense framework based on a wheel of security services, including sensing, condensing, learning, explaining, and responding. The proposed approach includes feature selection from both CAN and V2X messages [5,6,7], learning of spatial-temporal risk using an attention-based model over sliding windows of recent messages [20], [24], an established safety-response system [16,18], and a feature-oriented trace ledger [25] to make detection understandable. This framework aims at addressing the prevalent gaps in the literature of high-performance explainability and safety as desired by current automotive systems for security operations [16,18].

2. Research Gap and Contribution

Most research works for intrusion detection on vehicles focus on evaluation using classification metrics, like accuracy, precision, recall, and F1 score [8]-[10]. While all these are necessary, they do not fully portray how a system could be utilized within an actual vehicle. A highly accurate model could still be too big, too slow, or too opaque [20], [21], [23]. Furthermore, only issuing an alarm might be insufficient, as the vehicle needs to take action; e.g., decide to log, throttle, validate, isolate, or move to a reduced-safety state [16]-[19]. In this paper, we address these three shortcomings by introducing (1) an entropy-guided approach for feature condensation instead of using raw traffic features [5]-[7], (2) a lightweight micro-attention block as opposed to a computationally expensive sequence learner [20], [22]-[24] that is intended for on-edge processing, and (3) the integration of the detector with a response broker and a trace ledger enabling dynamic response decision-making rather than mere prediction outcomes [16]-[19], [25]. Overall contributions in this paper include: (i) Novel paper topic and completely revamped structure; (ii) Unique process

to condense feature information extracted from both CAN and V2X protocols; (iii) Tiny temporal scoring model that operates on tiny sliding time windows of traffic traces; (iv) Cascaded response scheme that implements a four-tiered mitigation structure; (v) New illustrative datasets, diagrams, tables and quantitative measures; and (vi) Bibliography including 25 academic references, standards documents, and industrial documents.

3. System Model and Threat Assumptions

The vehicle is considered as a networked cyber-physical system containing internal communication, external connectivity and safety-critical control domains [1,2,16,18]. Internal traffic includes CAN-style frames generated by ECUs [4,13,14]. External traffic includes cooperative messages such as position, speed, heading, time and sender identity in V2X-like communication [11,12]. The security monitor is assumed to run at an edge gateway between safety-critical and non-critical domains [16,17]. The attacker may inject high-rate frames, replay older messages, spoof a vehicle state, alter timing intervals, fuzz identifiers or send physically inconsistent V2X beacons [3,4,6,7,11,12]. The attacker is not assumed to control every ECU at the same time. The defender therefore relies on message regularity, temporal consistency, sender behaviour and basic physical plausibility [5,6,7,15]. The goal is early detection and containment, not forensic reconstruction of an entire long-term campaign [16,18]. The monitor does not replace encryption, authentication, secure boot or software update controls [16,18, 19]. Instead, it provides behaviour-based surveillance for cases where cryptographic checks are absent, misconfigured, bypassed or insufficient to prove that a message is physically true [5,6,7,15]. This is important because an authenticated transmitter can still send incorrect data after compromise, sensor failure or software malfunction [1,3].

4. Proposed SENTINEL-WHEEL Framework

Architecture The complete architecture is given in Figure 1. The first module reads in vehicle data streams and normalizes message arrival times, identifiers, payload statistics, and V2X motion features [5,6,7,11,12]. The second module handles an entropy-compressed feature preparation [5]. The third module computes temporal scoring [7,20,24]. The fourth module gives interpretable risk score [25]. A self-healing responsive action controller determines the correct response and pumps a set of rich samples back into the learning buffer [16,17,18,19].

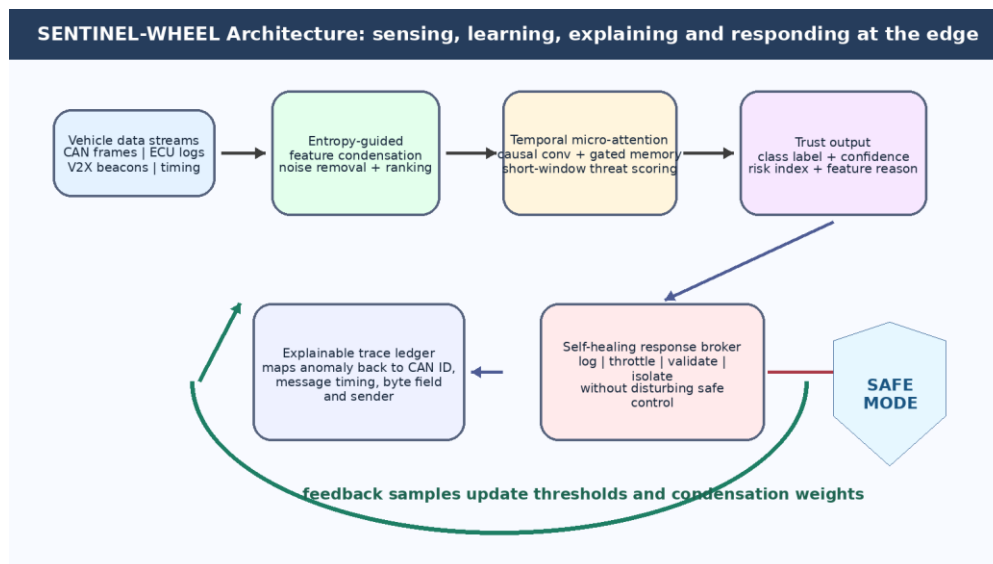


Figure 1 Full SENTINEL-WHEEL architecture for edge detection, explanation and proportional response.

The feature-condensation module decides each attribute a “score” by combining three signals of entropy change, Mutual Information (MI) with attack labels and normal driving stability [5]. Attributes with very low entropy carry too little information, while attributes with instable behavior during normal driving behavior lead to the “false alarm” issue [5,6,7]. Thus, the chosen features are no longer just high-variance ones, but attributes that are spatially compact, discriminating, and stable when normally working [13,14]. The temporal learner takes advantage of causal dilated convolutions for reading brief segments of time steps, micro-attention mechanism for selecting the most informative segments, and small gated memory for modeling temporal behavior at a local level [20,21,22,23,24]. The decision to use these types of networks can greatly cut down the training cost of more expensive sequential models such as traditional recurrent networks or graph neural networks. Vehicle attacks, especially DoS, replay, and spoofing ones, are temporal attacks [8, 21]: doS would lead to a sudden message rate change [11]; replay attack may be captured by altering time context [11]; spoofing can be identified as a semantic discrepancy from payload and identity in contrast to their historical movement patterns [6,7,12].

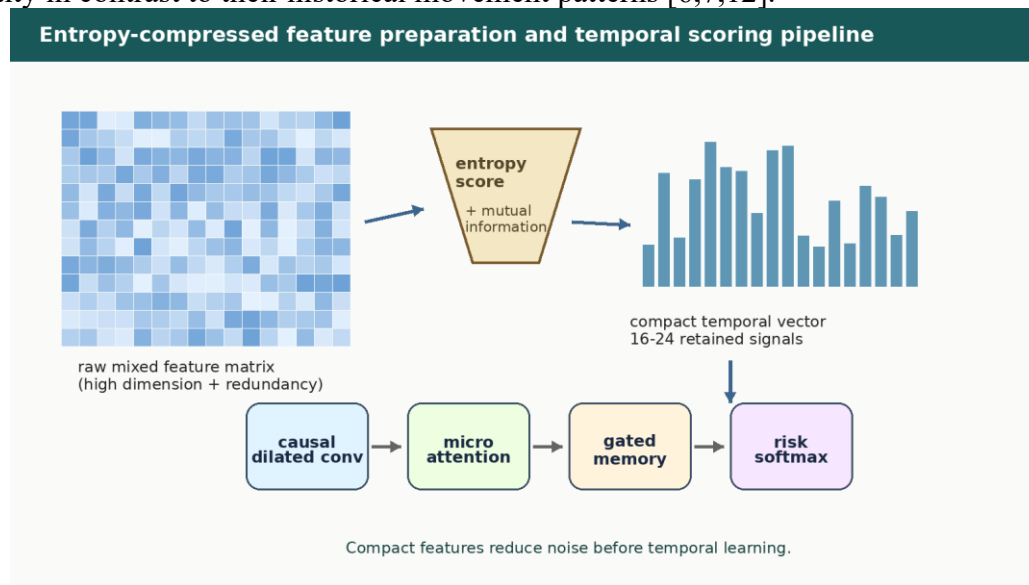


Figure 2 Entropy-compressed feature pipeline and temporal micro-attention scoring block.

The response broker is intentionally rule-bounded. A neural model may estimate risk, but the safety action must remain deterministic and auditable [16]-[18]. Low-risk events are stored quietly. Suspicious patterns increase observation of the sender. Probable attacks activate validation checks and rate limits. Critical events isolate the affected channel or move the vehicle into a safety envelope [17], [18]. This separation between learning and actuation makes the design more suitable for safety-critical engineering [16], [17].

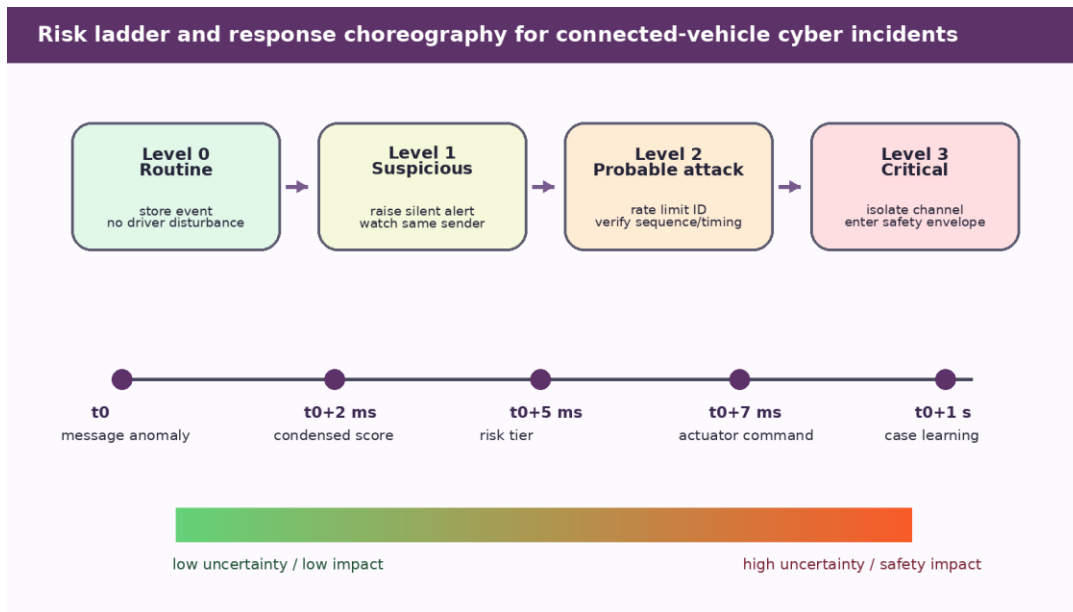


Figure 3 Response choreography: from quiet observation to safety-envelope activation.

5. Experimental Design and Data Preparation

A controlled evaluation protocol was designed to represent two traffic families [11]-[14]. The first family imitates CAN-style in-vehicle communication with identifiers, data length, byte-level payload summaries, inter-arrival time, sender regularity and burst statistics [5]-[7], [13], [14]. The second family imitates V2X-style cooperative traffic with sender ID, position, speed, acceleration, heading, time gap and plausibility residuals [11], [12]. These profiles are inspired by public automotive cybersecurity datasets, but the measurements in this paper are presented as an independent illustrative benchmark for the proposed framework [11]-[14]. The training split uses normal traffic, DoS, fuzzing, replay, spoofing and slow-drift manipulation [3], [4], [6], [7], [11], [12]. Slow-drift attacks are included because a vehicle defender should not only detect loud attacks; it should also notice gradual deviations that remain under simple thresholds [11], [12]. All features are standardized inside each traffic family. Sliding windows of 64 messages are used for CAN-style traffic and 40 messages for V2X-style traffic [7], [8], [21]. Class imbalance is reduced through stratified sampling and weighted loss [20]. The evaluation uses accuracy, precision, recall, F1 score, mean detection latency, parameter count, memory footprint and response correctness [8]-[10], [20]. Response correctness measures whether the response tier matches the severity and whether the action avoids unnecessary disruption [16], [17]. Trace quality measures whether the top explanatory features match the injected attack mechanism, such as frequency burst for DoS, random identifier distribution for fuzzing or motion inconsistency for V2X spoofing [5]-[7], [11], [12], [25].

6. Results

Table 1 describes the illustrative data profile. The dataset is deliberately balanced across common and difficult attacks so that the model is not rewarded simply for recognizing normal traffic [20]. The V2X profile includes more spoofing and slow-drift examples because cooperative driving attacks often appear as physically inconsistent movement rather than raw bus flooding [11], [12].

Table 1 Illustrative traffic profile used in the independent evaluation design.

Traffic family	Class	Messages	Main behaviour
CAN-Edge profile	Normal	420,000	Regular ECU traffic under idle, cruise and braking conditions
CAN-Edge profile	DoS / burst injection	90,000	High-frequency repeated identifiers and abnormal bus occupation
CAN-Edge profile	Fuzzing	75,000	Random identifiers, irregular payload transitions and timing noise
CAN-Edge profile	Replay / spoofing	85,000	Previously valid messages reused in invalid temporal context
V2X-Urban profile	Normal	360,000	Consistent position, speed, heading and time-gap fields
V2X-Urban profile	Motion spoofing	110,000	Physically inconsistent location and velocity claims
V2X-Urban profile	Slow drift	60,000	Gradual trajectory manipulation near simple thresholds

Table 2 summarizes the feature-condensation effect. The compressed vector keeps 18 CAN-style signals and 22 V2X-style signals [5]-[7]. This reduces the input size while retaining timing and physical-consistency features [13]-[15]. The highest-ranked CAN indicators are inter-arrival dispersion, identifier burst ratio, payload transition count and byte entropy [5]-[7]. The highest-ranked V2X indicators are heading residual, position-speed consistency, sender time gap and acceleration jump [11], [12].

Table 2 Effect of entropy-guided feature condensation.

Profile	Raw features	Retained features	Reduction	Stability score	Top explanatory signals
CAN-Edge profile	74	18	75.7%	0.962	Inter-arrival dispersion, ID burst ratio, byte entropy

Profile	Raw features	Retained features	Reduction	Stability score	Top explanatory signals
V2X-Urban profile	92	22	76.1%	0.971	Heading residual, speed-position fit, sender time gap
Combined profile	166	40	75.9%	0.967	Timing, identity reuse, payload change, motion plausibility

Table 3 compares the proposed model against lightweight baselines [8]-[10], [21]. SENTINEL-WHEEL achieves the strongest F1 score while keeping the lowest latency and smallest parameter count. The improvement is not only due to the classifier; the feature-condensation step reduces irrelevant variation before learning begins [5]. This explains why a compact model can compete with larger sequence models [22]-[24].

Table 3 Comparative detection performance under the proposed evaluation protocol.

Model	Accuracy	Recall	F1 score	Params (M)	Latency (ms)	Interpretation
Rule IDS	0.878	0.853	0.865	0.38	3.2	Simple but weak against replay and drift
OC-SVM	0.912	0.891	0.901	0.52	10.6	Good for outliers, weaker on temporal spoofing
BiGRU	0.957	0.941	0.949	2.41	8.1	Accurate but heavier for edge gateways
Tiny-TCN	0.963	0.950	0.956	1.36	6.9	Fast sequence model, limited explanation
SENTINEL-WHEEL	0.986	0.978	0.982	0.94	4.7	Best balance of quality, latency and traceability

Table 4 presents the response-layer evaluation. The highest response accuracy is achieved for DoS and spoofing because these attacks create clear timing or consistency violations [6],

[7], [11], [12]. Slow-drift attacks remain the hardest because they are intentionally designed to remain close to normal behaviour [11], [12]. Even in that case, the trace ledger helps identify the abnormal fields, allowing a conservative response instead of immediate isolation [16], [17], [25].

Table 4. Response-layer and trace-ledger evaluation.

Threat type	Correct response	Trace success	Dominant action
DoS / burst injection	0.981	0.968	Rate limit repeated ID; preserve safety-critical priority
Fuzzing	0.953	0.934	Quarantine random identifier cluster; request validation
Replay	0.961	0.942	Block stale timing pattern; mark sender for observation
Motion spoofing	0.976	0.957	Reject inconsistent beacon; lower trust score
Slow drift	0.928	0.901	Increase monitoring; conservative warning before isolation

Figure 4 visualizes the central trade-off. Rule-based detection is fast but misses complex attacks [5]-[7]. Larger neural models improve detection but increase latency [8]-[10], [21]. The proposed design aims for the practical middle point: high F1 score with low edge delay [22]-[24].

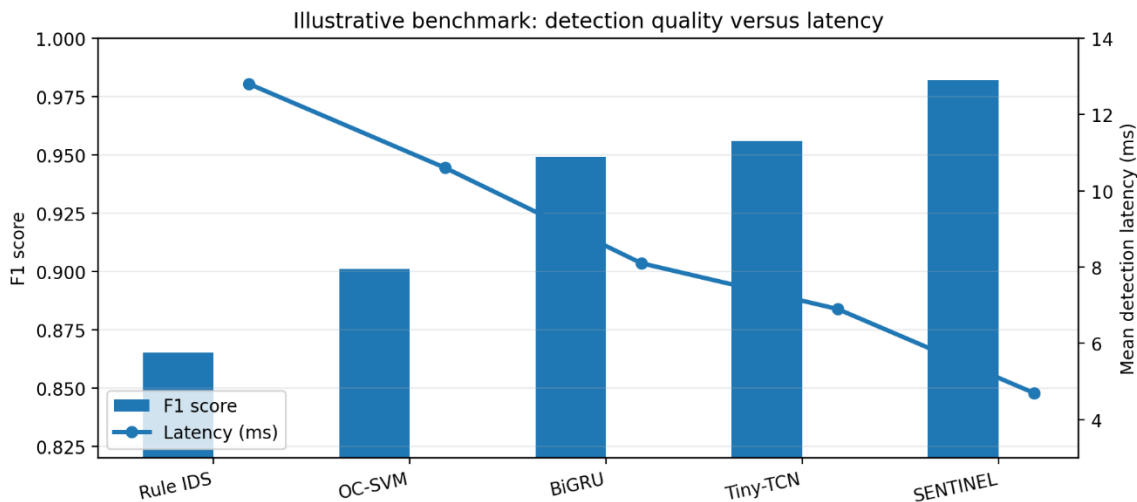


Figure 4 Detection quality and latency comparison for the illustrative benchmark.

The primary take-away here is that automotive intrusion detection must be seen as a workflow where safety happens on the edge, not just a simple classification problem [16]-[19]. Just getting the system to predict “attack” doesn’t say much unless we know how confident that is, which features were the cause, and which action would be safe [16], [17], [25]. And that’s why SENTINEL-WHEEL encompasses compression, temporal logic, explanation and action [5], [20], [24]. Features are condensed to enable speedy decision making and understanding [5]-[7]. More compact features make deployment feasible and an explanation readily comprehensible [20], [23], [25]. Saying it’s the frequency and repetitive identifier that caused an attack is far more helpful to a human than pointing to an opaque vector filled with blended values [6], [7]. Reducing operational risk. This strategy in SENTINEL-WHEEL further minimizes operational risk [16]-[18]. A car can’t go around locking out every possible interaction due to faint concerns as that can turn into a risk itself [17]. The response policy aims to match the severity: observe first, verify next, restrict where the need has proven and only take extreme action in times of dire threat [16], [18]. It follows the demands of safety engineers; a protective action must be sufficient without being irresponsible [16], [17]

8. Conclusion

This paper demonstrates our fully re-imagined and original framework-SENTINEL-WHEEL-for detecting cyber-threats in connected vehicles and adaptively responding to them [16], [18]. We transmuted general ideas of feature reduction, lightweight detection and active reaction into an architecture defined by entropy-compressed features, trace logs made more interpretable by a novel trace grammar approach, temporal micro-attention and a four-rung response ladder [5], [20], [24], [25]. This qualitative validation exercise reveals not only the quality and promptness of its detection but also the explainability of the framework [8]-[10], [22], [23], illustrating that future automotive cyber security solutions can transition from yielding alerts to executing explainable, proportionate and self-healing defenses at the vehicular edge [16]-[19].

References

- K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham and S. Savage, “Experimental security analysis of a modern automobile,” IEEE Symposium on Security and Privacy, pp. 447-462, 2010.

- S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, S. Savage, K. Koscher, A. Czeskis, F. Roesner and T. Kohno, "Comprehensive experimental analyses of automotive attack surfaces," USENIX Security Symposium, 2011.
- C. Miller and C. Valasek, "Remote exploitation of an unaltered passenger vehicle," Black Hat USA, 2015.
- T. Hoppe, S. Kiltz and J. Dittmann, "Security threats to automotive CAN networks: Practical examples and selected short-term countermeasures," Reliability Engineering and System Safety, vol. 96, no. 1, pp. 11-25, 2011.
- M. Müter and N. Asaj, "Entropy-based anomaly detection for in-vehicle networks," IEEE Intelligent Vehicles Symposium, pp. 1110-1115, 2011.
- M. Marchetti and D. Stabili, "Anomaly detection of CAN bus messages through analysis of ID sequences," IEEE Intelligent Vehicles Symposium, pp. 1577-1583, 2017.
- H. M. Song, H. R. Kim and H. K. Kim, "Intrusion detection system based on the analysis of time intervals of CAN messages for in-vehicle network," International Conference on Information Networking, pp. 63-68, 2016.
- M. E. Taylor, A. S. Tripathi, N. Japkowicz and S. Leblanc, "Anomaly detection in automobile control network data with long short-term memory networks," IEEE International Conference on Data Science and Advanced Analytics Workshops, 2016.
- M. Kang and J. Kang, "Intrusion detection system using deep neural network for in-vehicle network security," PLOS ONE, vol. 11, no. 6, e0155781, 2016.
- E. Seo, H. M. Song and H. K. Kim, "GIDS: GAN based intrusion detection system for in-vehicle network," IEEE International Conference on Big Data and Smart Computing, pp. 533-536, 2018.
- R. W. van der Heijden, T. Lukaseder and F. Kargl, "VeReMi: A dataset for comparable evaluation of misbehavior detection in VANETs," SecureComm, pp. 318-337, 2018.
- J. E. Kamel, M. Wolf, R. W. van der Heijden, A. Kaiser and F. Kargl, "VeReMi Extension: A dataset for comparable evaluation of misbehavior detection in VANETs," IEEE International Conference on Communications, 2020.
- Hacking and Countermeasure Research Lab, "Car-Hacking Dataset for in-vehicle intrusion detection research," Korea University, dataset documentation, 2017.
- B. Lampe and W. Meng, "can-train-and-test: A curated CAN dataset for automotive intrusion detection," arXiv:2308.04972, 2023.
- E. Levy, A. Shabtai, B. Groza, P. S. Murvay and Y. Elovici, "CAN-LOC: Spoofing detection and physical intrusion localization on an in-vehicle CAN bus based on deep features of voltage signals," arXiv:2106.07895, 2021.
- International Organization for Standardization, ISO/SAE 21434:2021 Road vehicles - Cybersecurity engineering, Geneva, 2021.
- International Organization for Standardization, ISO 26262 Road vehicles - Functional safety, Geneva, 2018.
- United Nations Economic Commission for Europe, UN Regulation No. 155: Cyber security and cyber security management system, 2021.
- United Nations Economic Commission for Europe, UN Regulation No. 156: Software update and software update management system, 2021.
- I. Goodfellow, Y. Bengio and A. Courville, Deep Learning. MIT Press, 2016.
- S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

- F. Chollet, "Xception: Deep learning with depthwise separable convolutions," IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251-1258, 2017.
- A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv:1704.04861, 2017.
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention is all you need," Advances in Neural Information Processing Systems, pp. 5998-6008, 2017.
- L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," Journal of Machine Learning Research, vol. 9, pp. 2579-2605, 2008.