Received: 29 October 2020 | Revised: 17 May 2021 | Accepted: 18 May 2021

DOI: 10.1002/ett.4325

SURVEY ARTICLE

WILEY

A survey of new orientations in the field of vehicular cybersecurity, applying artificial intelligence based methods

Zsombor Pethő | **Árpád Török** | **Zsolt Szalay**

Department of Automotive Technologies, Budapest University of Technology and Economics, Budapest, Hungary

Correspondence to:
Zsombor Pethő, Department of Automotive Technologies, Budapest University of Technology and Economics, J Building, Stoczek u. 6., 1111 Budapest, Hungary.
Email: zsombor.petho@edu.bme.hu

Funding information
Innovative Mobility Program of KTI; National Research Development and Innovation Office in the field of Artificial Intelligence

Abstract

Nowadays, cybersecurity is an emerging research area in the automotive industry, and it is investigated by many different perspectives. Our article is a review of existing vehicular security solutions that covers the state-of-the-art and future research directions. This article is a new contribution to tutorials/surveys related to the vehicular cybersecurity domain with the latest details. We developed a database from 140 articles from the field of automotive security. In the database, we assigned specific attributes to every article (such as Web of Science Impact Factor or the number of citations). The data set was analyzed by the K-means clustering and decision tree analysis methods to identify and characterize the generated groups of papers. Following this, the article highlights the research areas that might receive more attention in the future. Accordingly, the result of the current research can be applied by the decision-makers, researchers, and Original Equipment Manufacturers to allocate additional resources to those domains, which is expected to shape the future of vehicular security.

1 | INTRODUCTION

Technologies in the automotive industry are continually changing.[1] Nevertheless, there is no doubt that the most evolving and leading research areas are different communication technologies. With the integration of diverse wireless communication-technologies (cellular, Wi-Fi, DSRC, ZigBee, etc. comprehensive study in Reference 2) and the interconnection of electronic controller units (ECUs), sensors, and actuators, the number of security leaks and vulnerabilities is increasing proportionately. The detection of potential vulnerabilities is of primary importance. In the past years, several different types of intrusion detection systems (IDSs) and other countermeasures have been proposed for automotive security leaks. However, a considerably large proportion of them have severe limitations in operating as a comprehensive protection mechanism. Unlike in the case of classical computer networks,[3] in the field of vehicular communication networks, there are still numerous serious challenges related to user privacy and information sensitivity, especially when the rapidly growing number of Internet of Vehicles (IoV) devices is considered.

It should also be emphasized that currently, the main vulnerabilities of vehicles can still be exploited through the controller area network (CAN) bus. In general, the most important reason for the mentioned vulnerability is the lack of encryptions in standard implementations. Besides, the field of default device authentication processes related to in-vehicle networks (IVNs) is still an under-developed domain.[4]

Abbreviation: OEM, Original Equipment Manufacturer.

One of the most critical vulnerabilities of IVNs is the possibility of connecting externally to them through the on-board diagnostic (OBD) port. However, the OBD-II port has limited access control with different security access levels. Beyond this, the efficiency of the usage of security algorithms for the OBD interface should still be improved. These algorithms mostly protect the diagnostic services.[4]

Following the above-mentioned circumstances, the recent article aims to contribute to the development of automotive security by identifying the most relevant domains of future development processes.

In this article, a comprehensive literature review is presented, focusing on the most relevant research articles investigating the field of automotive security. In the light of the performed general evaluation, the most relevant research domains were identified, considering such indicators as the impact factor or the citation index of the analyzed articles.

The motivation for this article is to identify those vehicular security domains that will have a critical influence on the future of connected cars. The security of future mobility is strongly affected by the communication systems' protection level, therefore we would like to identify those critical security areas, of which vulnerabilities have to be reasonably mitigated.

After the introduction, we present the methodology (Section 2) of the performed investigations from three perspectives: (i) the description of each automotive security research topic and the selection process; (ii) the database creation process; and (iii) the development of the analytical model, where we present the applied non-classical analytical methods. After the methodology section, we introduce and discuss the results of the performed analysis.

Finally, each article is evaluated based on the generated classification structure, the discussed topic of the articles, and their previously mentioned typical indicators. To perform a thorough analysis, we implemented a task-oriented data mining approach. Numerous research studies have already applied data mining and statistical analysis based methodology to estimate the expected research orientations through considering the distribution and the changes of the investigated attributes in time.[5-7] However, in the field of automotive security, non-classical analytical method-based literature reviews and the forecast of the emerging scientific topics are still underrepresented.

Accordingly, this article applies supervised and unsupervised learning techniques to describe the actual research trends characterizing the automotive security domain. Therefore, this article aims to support decision-makers and researchers to allocate additional resources to those domains, which is expected to shape the future of vehicular security. Beyond this, the study also contributes to automotive Original Equipment Manufacturers (OEMs) in identifying those development orientations that can be applied to mitigate certain critical vulnerabilities even in the design phase taking into account the feedback from the outcomes of the scientific research.

On the other hand, researchers and academic experts can also benefit from the results of this study, since our article concludes the expected future research orientations based on the trends of the past and the actual tendencies following a scientific methodological framework. The flow diagram of the applied certain consecutive procedures is presented in Section 2.

2 | METHODOLOGY

In this section, we aim to introduce the most relevant considerations and approaches applied during the investigation. In the first part of the chapter, we introduce the categories used to classify the articles and the papers. In the second step, we present the development process of the database. In the final section, we describe the developed analytical model.

During the literature review process, we have identified numerous detailed surveys/review papers and in-depth research articles related to automotive security. Among others, this section also includes the classification of the studied and processed reviews and research papers and the representation of the evaluated articles and topics by introducing some of the most substantial papers in that particular topic.

Based on the considerations mentioned above, we analyzed the experimental results and the impact of the articles as the function of different journal article indicators, like citation or impact factor. Based on the in-depth literature review, we distinguished seven specific topics in the automotive security sector:

- Security and privacy design (SNP).
- Defense mechanism (DFM).
- Intrusion detection system (IDS).
- Security analysis (SA).
- In-vehicle network (IVN).

- Inter-vehicular communication systems and architecture (V2X).
- Artificial intelligence (AI).

The identification of the classes mentioned above was started by defining the most relevant keywords, on which the article selection process should be based. The keyword specification process led to a considerably extended list, which had to be contracted. In other words, the amount of overlap among the different domains, represented by the identified keywords, should have been reduced. As a result of the merging process of the classes, we concluded that the introduced seven topics should be used to classify the reviewed papers. The definitions of the discussed topics are explained in more detail in Section 2.

2.1 | Flow diagram of the research procedures

During the literature review phase, we searched for highly cited articles on the topic of automotive security. In the selection phase, based on the keywords of the selected relevant articles, we defined the most frequently used keywords. Following this, in the labeling phase, we assigned automotive security sub-domains to the selected keywords. To avoid overlapping between topics, in the contraction phase, we merged closely related categories to generate disjunctive sets. In the collection phase, we gathered the investigated articles based on the identified requirements (see Section 2.3). We generated the database based on the selected article parameters (see in Section 2.3). To investigate the normality of the database parameters, we applied the Kolmogorov-Smirnov method. We classified the articles based on the selected article parameters by applying K-means clustering in the next step. Following this, we implemented a chi-squared test to verify the results of the clustering process. Finally, we validated the clustering model by performing a decision tree analysis (see in section 2). The review process is illustrated in the flow diagram (Figure 1).

2.2 | Description of the applied security categories/topics

In this subsection, we would like to give a brief descriptive explanation of each vehicle cybersecurity topic mentioned above and introduce some of the relevant research studies related closely to the particular topic.

2.2.1 | Security and privacy design

In this group, we evaluated the research papers related to the core security/user privacy issues. In accordance with the investigated articles, the core security and privacy requirements are defined as follows:

- Authentication. In the case of communication processes, components of electronic systems have to prove their assertions included in the messages through identifying themselves.
- Integrity. This term refers to the requirements of consequent data transfer processes and the testing methods of data consistency. The concept of integrity makes the system capable of detecting whether the stored or exchanged data packets between components are consistent or not. This approach can support the system in identifying manipulated or injected messages in the communication data flow.

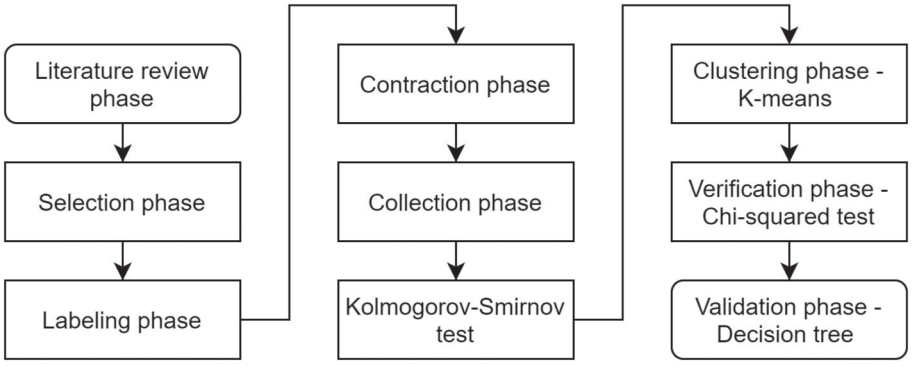

FIGURE 1 Flow diagram of the research procedure presented in this study

- Access control. Accessibility of vehicle components and electronic units from the external environment has to be controlled and, if necessary, restricted since the lack of control can lead to unreasonable vulnerabilities. In a communication system, access control increases reliability by assigning access rights to authenticated nodes within the system, thereby marking nodes outside the system as invalid/rogue nodes. Access control policy is perhaps one of the most helpful security techniques as it thoroughly narrows down the success of a malicious attack. This technique should be used in the intra-vehicular system, as there can be assigned access control to each node, thus minimizing the risk of a node sending a message whose messages are not relevant to the destination node.
- Privacy. This term refers to the requirements of protecting sensitive user information and data (eg, geographical position of the car, personal identity information of the driver.)
- Intellectual property protection. In the field of cybersecurity, this requirement represents the objective of preventing the unauthorized access and usage of intellectual property, especially considering domains threatened by reverse engineering, cloning, or data theft.
- Availability. Due to the required safety and functional integrity level of the highly automated vehicles, the availability of the in-vehicle control modules and ECUs has to be provided continuously during the operation process. Therefore, the robustness and reliability requirements of in-vehicle communication networks (CAN, LIN, MOST, Flexray, Ethernet) are crucial to be satisfied at all times. This is specifically relevant for safety-critical functions since data exchange between safety-critical ECUs and sensors of the vehicle should be continuously provided.
- Message freshness. If valid data transmission is recorded and later maliciously repeated, then the structure and the format of the data may suit the requirements of the system. In such a case, the detection of the attack can be difficult. If the freshness of the messages and the related requirements are investigated, then the system might have the ability to detect and prevent a replay attack.
- Confidentiality. Access to specific types of information primarily related to the identity of the vehicle or the user has to be limited. These types of information shall be identified as confidential. Accordingly, confidentiality can be defined as the requirements for placing restrictions on data and information. In the case of road transportation,[8] the private data and identity of the passengers and users have to be handled confidentially. This issue is especially challenging when wireless channels are involved in the communication process.[9]

After the introduction of the most critical requirements of security, it seems to be reasonable to discuss the relationship between safety and security. Safety and security are very closely related disciplines, and they could benefit from each other if their interactions are properly considered. Both areas should be integrated into the development process from the initial phases. From the safety side, for example, failure mode and effects analysis (FMEA) or fault tree analysis (FTA) is standardized techniques, but security testing methods have not yet formed an integral part of the automotive development process. Safety and security co-engineering could be an adequate solution for this problem, which is getting more and more accepted by the professional society.

When we are talking about promising techniques in automotive security, we should also highlight the blockchain technology. The decentralization offered by this novel approach can significantly contribute to the development of vehicular ad-hoc network domain in the automotive industry. The application of the blockchain method in the control of transport processes can help us to prevent any single party from viewing all pieces of information (especially considering sensitive data), and from spoofing previously recorded data.[10-12]

Modern cars have to meet a set of requirements during the development, design, and testing phase. These requirements intend to provide the safety, the security, and the privacy of the users and the whole society with regard to the processes related to the vehicular operation.[13]

In the survey of Bernardini et al,[9] these overall aspects were investigated. Bernardini et al identified three main requirement categories for modern cars: security, safety, and standardization/architectural requirements.

Besides, they also call attention to research challenges that primarily highlight weaknesses in the intra-vehicle network (authentication of ECUs) and the obligatory use of network monitoring systems with real-time constraints.

Another CAV security (connected and autonomous vehicles) related emerging technology is the Over-the-Air (OTA) remote software and firmware updates, which require significant attention nowadays and in the future. One of the approaches to secure the OTA remote updates is the appliance of hardware security modules (HSMs) alongside with safety-critical ECUs, to store the cryptographic keys in a specific secured memory and perform some cryptographic operations.[14]

2.2.2 | Defense mechanism

This category contains countermeasures and response methods to specific malicious interventions targeting the intra- and inter vehicular network. The topic of DFMs is closely related to the IDS solutions; however, we would like to emphasize the importance of this field. The reason for this is the relatively little amount of research performed in this domain compared to its importance. The main difference between IDS and DFMs is the response provided by the system. If an IDS is not able to respond to detected intrusions, then it acts like a "sensor," and it is classified in the IDS category. Articles in this topic propose various defense systems like firewall, hybrid security system (HSS), intrusion detection and prevention system (IDPS).

According to a defense taxonomy proposed by Thing and Wu,[15] we can classify defense strategies in four main categories: preventive, passive, active, and collaborative defense.

Preventive defense is a countermeasure that stops an attack from happening and includes the following techniques to enhance security:

- Secure communication using encryption.
- In-vehicle device authentication (accredited OEMs).
- User authentication (biometric authentication).
- Firewall.

Passive defense provides another layer of security against attackers. This category of defense consists of attack detection, attack response, and attack recovery. The different types of IDSs are the central pillars of the attack detection models, but anti-malware solutions also belong to this group. Since the importance of IDSs becomes more and more significant in the automotive industry, this review classifies the articles investigating this domain in a separate category (IDS). For responding to an attack, nullification (ability to neutralize a cyber-attack) and isolation could be a DFM. The third type of passive defense is the ability to recover after a security attack, which is strongly related to the availability of the vehicle components, which is a safety-critical requirement also.

Active defense is among the most complicated defense types because, in the case of active defense, data traffic is needed to be continuously monitored, which requires significant resources.

Since the vehicle communication system is not a static state system but a dynamic process, the usage of traditional security countermeasures like IDS and IPS is not sufficient, therefore there is a severe need for continuous monitoring systems like adaptive security and honeypots. These dynamic security systems log and trace the activities of a potential attacker. A well-designed adaptive security architecture should be able to recognize ongoing security breaches and custom attacks. Accordingly, adaptive security mechanisms/algorithms allow vehicles to use cheaper CPUs with lower clock frequencies.[16] Experiments have shown that adaptive security can respond 10 to 15 milliseconds faster while using traditional security techniques; a better-performing processor achieves worse results. Basically, honeypots are computer programs/systems that give the attacker the illusion that they are valid and legitimate targets. In contrast, the attacker is constantly monitored, analyzed, and the honeypot is completely isolated from the original attack surface, thus preventing any damage to the target system. There are basically two types: static and dynamic honeypot. The algorithms and parameters of a static honeypot do not change at runtime. In contrast, the parameters and system properties of a dynamic version can change in real-time and adapt to the attack strategy. It works on a similar principle to machine learning algorithms that change their operation based on input and empirical data. Unfortunately, zero-day attacks and sophisticated multi-layer attacks are still a significant challenge for defensive security algorithms. Based on our literature review, we found that there are not too many studies in this field; however, there are some rather promising achievements.[4,16-20]

2.2.3 | Intrusion detection system

We classify IDS related articles in a separate category since, currently, IDS is among the central areas of security-science, and fortunately, many different intrusion detection techniques have been developed over the last few years. The main concept of IDS can be summarized as monitoring the traffic of the network in order to detect malicious and unexpected data exchange and, in some cases, policy violations. Besides this, we have to mention that IDS should not be applied in itself, but in cooperation with other security layers. The system can react properly to an attack, if the following security

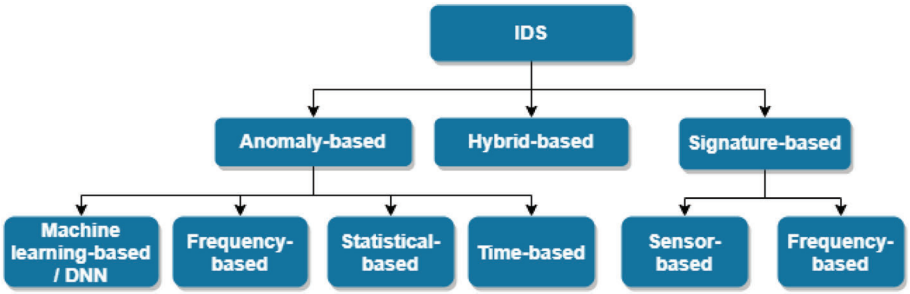

FIGURE 2 IDS taxonomy organized by detection approach

modules can implement an effective cooperation: message authentication for preventing unauthorized access, logging outgoing and incoming messages for intrusion detection, reacting to the attacks via intrusion prevention system (IPS), and traceability for the recovery procedure.[21,22]

IDSs can be categorized by the following aspects: deployment strategy, detection approach (see in Figure 2), attacking techniques, and technical challenges. The deployment strategy of an IDS consists of choosing the system component where it needs to be applied.

In general, we can differentiate two types of automotive IDS models. In the case of the host-based IDS, we install an IDS module to each vehicle ECU, and in this way, we get an overview of the internal activities. The advantage of this configuration is the ability of the system to detect suspicious traffic in real-time. In the case of the network-based IDS (NIDS), the IDS is installed to the CAN network or to the central gateways. The advantage of this approach is the more favorable resource-efficiency.

By detection approach, the literature distinguishes IDSs by working principle. The most common and promising IDS type is anomaly based IDS, which detects any deviation from normal behavior, using pattern recognition.[21]

Sharma et al investigated and evaluated numerous different IDS solutions like anomaly detection, signature-based IDS, and honeypot-based proactive security mechanisms.[20] Their survey focused primarily on VANETs and VANET Clouds, and concerning these domains, the paper revealed the advantages and limitations of different IDS techniques.

Lokman et al[21] examined various kinds of possible attack surfaces related to the CAN bus in their review article, from direct physical access to remote wireless access.

They introduced a novel taxonomy in classifying IDS related research articles according to the following aspects: IDS detection approaches, attacking techniques, deployment strategies, and technical challenges. Their review paper provides a considerably comprehensive overview regarding IDSs; however, it does still not cover all the possibly applicable approaches. Dupont et al[23] presented an evaluation framework for different CAN Network Intrusion Detection System (NIDS) collected from previous literature.

Their experimental results proved that the capability of the investigated NIDSs is very limited in detecting various attack types, and the false positive rate is usually too high.

2.2.4 | Security analysis

The extremely rapid development of new technologies results in a dramatic increase in the number of possible attack paths and in the complexity of attack methods and cases. Therefore, it is highly essential to pay particular attention to attack surfaces, attack vectors with more and more sophisticated attack scenarios. The fast pace of technological changes can make it quite challenging to be prepared for the different kinds of attacks and to predict all possible attack surfaces (such as OBD-II port, Bluetooth, WiFi, cellular communication, or DSRC radio channels). Despite every effort made by security researchers, there will always be security vulnerabilities and loopholes that have not yet been discovered and documented, like zero-day attacks.

Based on our literature review, we can say that this topic deserves a separate category, especially as we move toward connected and autonomous vehicles (CAV). When we talk about malicious interventions targeting vehicles or vehicular networks, we should handle the investigation of attack surfaces and threat analysis separately. On the one hand, the investigation of attack surfaces focuses on the identification of vulnerabilities. On the other hand, threat analysis primarily evaluates the relationship of:

- the exploitability of a specific vulnerability,
- the expected significance or severity possibly caused by the exploitation of the investigated vulnerability, and
- the level of the assumed motivations and ambitions behind the exploitation of the given vulnerability.

The main concept of this threat modeling approach is to analyze each component of the whole system for susceptibility to threats and mitigation for all threats to each component in order to claim that the system is secure.[24,25]

Accordingly, threat modeling aims to analyze:

- each component of the whole system for susceptibility to attacks, and
- possible mitigation of the attacks to be detected and treated in case of each component,

in order to estimate and improve the security level of the system.[24,25]

Threat modeling is a base stone of cybersecurity, an indispensable activity that should be performed in application development and system evaluation processes. It helps developers to make proactive architectural decisions that reduce threats from the start. Threat modeling has different objectives in each development phase and has different input concerning the details of the system. The threat model describes the security considerations for a particular type of system by associating a range of potential vulnerabilities, attacks, and threats, considering the potential set of tools associated with specific features or use cases.[26]

One of the most prevalent threat modeling methodologies is STRIDE (Spoofing, Tampering, Repudiation, Information disclosure, Denial of Service, and Elevation of privilege). This software-centric threat model was developed by Microsoft, especially for enterprise services, but it was adapted successfully by the automotive industry, and serves as a basis for other threat modeling methodologies. DREAD (Damage Potential, Reproducibility, Exploitability, Affected Users, Discoverability) is a classification scheme to quantify, compare the amount of risk presented by each evaluated threat. OCTAVE (Operationally Critical Threat, Asset, and Vulnerability Evaluation) focuses mainly on organizational risks rather than technological risks, making it difficult to be applied in complex vehicle systems.

EVITA method originates from a research project EVITA (E-Safety Vehicle Intrusion Protected Applications), which focuses on the on-board network protection methods. The EVITA method consists of three main activities: threat identification, threat classification, and risk analysis.[27] TVRA (Threat Vulnerability and Risk Analysis) was proposed by ETSI, which is a systematic identification process. Its goal is to determine the risk related to the system based upon the product of the likelihood of an attack and its expected impacts.[28]

HEAVENS (HEAling Vulnerabilities to Enhance Software Security and Safety) project aims to identify state-of-the-art technologies and security threats and to investigate together all the aspects of threat modeling (assets, vulnerabilities, risks, threat agents, countermeasures) to construct a complex security model. The HEAVENS security model investigates safety and security co-engineering approaches related to the E/E architecture of the vehicle and also maps IT software issues to the automotive domain.[29]

SAHARA (Safety aware hazard analysis and risk assessment) integrates the automotive HARA approach with the STRIDE method. SAHARA approach is compatible with the base requirements of the HARA analysis (from ISO26262, the automotive safety standard for road vehicles) but extends the impact by taking security threats into account.[30]

The SINA (Security In Networked Automotive) methodology is a security-oriented interpretation of the ISO 26262 process, which targets the design phase of automotive products. It is the security equivalent of HARA to determine and prioritize potential security threats. It consists of three main phases: DFD (Data Flow Diagram) modeling, keyword-based threat identification, and risk assessment with the help of attack trees to identify the most severe threats.[31]

The Common Vulnerability Scoring System (CVSS) is an open framework and numerical method that offers a scoring system for vulnerability evaluation. A CVSS score can be computed based on a predefined formula combining all possible metrics, which could be obtained from a vulnerability database. The CVSS was developed for analyzing severity of software vulnerabilities and now it has been adapted for cyber-physical and connected vehicle systems.[32] Graphical security models such as attack trees and attack graphs have been used for SA of automotive systems because they are well constructed to support risk assessment studies and to visualize potential attack surfaces and scenarios. The threat models mentioned above, as well as their main application context and focus, can be found in Table 1.

From a cryptography point of view, an interesting threat model approach is the Dolev Yao (DY) threat model, which idealizes the attacking party and assumes that an attacker with certain characteristics exists in a communication system

TABLE 1 Threat modeling methods, security analysis, and risk assessment frameworks in the automotive security domain

Methods	Focus	Main application context
STRIDE	Software-centric	Cyber and cyber-physical systems
DREAD	Asset-centric	Enterprise systems
OCTAVE	Asset-centric	Enterprise systems
EVITA	Attacker centric	Vehicular IT systems
TVRA	Asset-centric	Communication services in ITS
TARA	Attacker centric	Connected vehicle systems
HEAVENS	Threat-centric	Automotive E/E systems
SAHARA	Threat-centric	Automotive embedded systems
SINA	Threat-centric	Connected vehicle systems
CVSS	Software-centric	Connected vehicle systems
ATA	Threat-centric	Vehicular IT systems

with a specific probability. These properties are generalized as follows: a malicious attacker can access any message, be a valid member of the system, and thus this malicious actor can send and receive messages from any node.[33]

Finally, in this SA topic, it is worth mentioning (although it could also belong to security design) that one of the formal security analyses is the Canetti-Krawczyk (CK) model. The CK model is a modular model suitable for examining and designing key-exchange protocols. According to the CK model, the following three sub-cases can be assumed.

According to the unauthenticated links adversarial model (UM), an attacker can control not just the communication links but also has access to private information (eg, secret keys). In the authenticated-links models (AM) model, the adversary can only forward messages from the original parties but cannot modify or add to their content. An attacker could deploy a test session query against the key agreement protocol. In this case, the CK model classifies the type of attack based on the information to be obtained (session-key query, session-state reveal).[34]

Automotive attack surfaces can be classified by different aspects (such as the assumed motivation, the applied tool, or the exploited vulnerability), as presented by Sommer et al.[35] Following their taxonomy, attack surfaces and attack paths substantially determine the attack classes (such as spoofing or tampering). On-board sensor spoofing is a popular attack type, and it can be easily achieved by software defined radios (SDR), but with the help of other IMU sensors and environmental sensors—like RADAR, LiDAR, ultrasonic sensor, camera—the actual GPS position can be verified and validated so the GPS spoofing attack can be detected in time. Sensor jamming is another safety critical attack type, especially for autonomous vehicles that rely mainly on environment perception sensors mentioned before.

2.2.5 | In-vehicle network

This topic includes in-vehicle domain (CAN, LIN, FlexRay, MOST, Automotive Ethernet) related research articles. The investigated research subtopics are the following: security frameworks for IVN, privacy enhancing technologies and methods to improve security of the CAN bus, authorization and authentication methods, proposed solutions in order to secure gateways and anomaly detection methods on the CAN bus.

The currently applied CAN bus protocol focuses on communication reliability and functional safety, it is not designed for protection and prevention against severe cybersecurity attacks, meanwhile, the number of ECUs in a vehicle is rapidly growing and increases the number of exploitable opportunities and potential attack surfaces. The CAN network is the car's nervous system, and in many cases, it is responsible for transmitting the information measured by the sensors to the control units. Based on the forwarded information, the ECUs control the actuators. There are many proposals by researchers and security experts for enhancing the security of CAN and CAN-FD (CAN Flexible Data Rate) protocols, like lightweight authentication methods, various integrity/plausibility checks and full-frame encryptions in the transmission process. An essential factor is the latency, which is an important condition for reliability and availability on which safety-critical functions are based. Latency can be manipulated by DoS-like attacks, which results in bandwidth congestion.

Since CAN is a message-based protocol there are set of in-built methodologies to support robust communication. An important feature is the carrier sense multiple access with collision avoidance (CSMA/CA), which senses when the bus

is idle and handles the message transmission process. There are other bit-level error checking mechanisms in the CAN message frames, such as acknowledgment (ACK), cyclic redundancy check (CRC), and end of frame (EOF) bits. The ACK slot is for indicating transmission errors if the destination node did not receive the transmitted message. The CRC slot has a reserved space of up to 21 bit (in CAN-FD protocol) in a frame, and it serves integrity checks. Although these are not security methodologies, in a case of a basic integrity/modification attack, the CRC can identify the threat and return error frames. The EOF bit indicates the end of a message frame and it has to be a recessive bit, otherwise (in the case of a dominant bit) an error is generated, which is able to improve security level as well.[4,20,21]

Since automotive CAN networks still have numerous vulnerabilities, such as the lack of sender authentication or message encryption, researchers investigated a lot of different IDS solutions to protect the network from direct physical and remote cyber-attacks.[36]

The first successful remote attack targeting a vehicle was made in July 2015 by Charlie Miller and Chris Valasek, which, among other experimental attacks, also exploited the vulnerabilities of the automotive CAN. Attacks before this incident were accomplished mainly with direct physical access to the CAN bus through OBD-II diagnostic port. The security vulnerability of the in-vehicle communication network is highly dependent on the E/E (Electronic/Electrical) Architecture of the vehicles. We distinguish three main types of topological design based on the EEA of the vehicles, which fundamentally influence the security level of the vehicle:

- *Gateway-based E/E architecture*. This is the traditional approach of sharing CAN messages on the same bus. The main disadvantage of this traditional architecture is that the ECUs have to share the same bandwidth, which limits the data processing ability, and there is no space for high performance computation, which is a crucial requirement for intelligent driving functions.
- *Domain-based E/E architecture*. The core concept of this topology is that the autonomous driving functions are divided into domains, and it means that every domain has a separate computation platform, which is called domain ECU. This topology is more flexible than the traditional ones, and the data flow within the domain does not reduce the shared bandwidth, which is a considerable advantage of this layout.
- *Centralized E/E architecture*. Centralized EEAs represent the next generation software-driven EEA approach, where a central computation unit/entity is present. This component is responsible for high performance computation tasks. The main benefit of this centralized layout is the ability to effectively realize the complete sensor fusion process, which supports more reliable real-time decisions. There is a need for faster and more reliable communication-technology, for example, automotive Ethernet with higher bandwidth (100 Mbit/s-50 Gbit/s).[37]

2.2.6 | Inter-vehicular communication systems and architecture

This topic is a central pillar of cybersecurity, and it covers all kinds of communication types used in the automotive industry, including their applications and architecture. Generally, it has four sub-domains: in-vehicle domain (human-machine interface (HMI), on-board unit (OBU)), ad-hoc domain (V2V—Vehicle-to-Vehicle), infrastructure domain (V2I—Vehicle-to-Infrastructure, road-side unit (RSU)) and service domain (Cloud, Internet Service Providers, Certificate Authorities).[2] In most of the research articles, this topic includes V2X communications and protocols at the same time. However, in this review, we handle these two concepts separately since the domain of V2X communication-technology is changing from a supportive sub-domain to a high-technology leading-domain. Accordingly, the topic of "vehicular communication systems and architecture" covers architecture- and protocol-related issues, while the topic of V2X (VANET) discusses the communication technology-related issues.[38]

The discipline focusing on the communication among the vehicles and the other network components (such as road-side units, pedestrians, or cyclists) is a vast and complex topic in the domain of automotive security that consists of the following essential parts:

- V2V: Vehicle-to-Vehicle communication, which is associated with the VANET concept,
- V2I: Vehicle-to-Infrastructure communication, when communication takes place between vehicle and the roadside infrastructure units (RSUs),
- V2N: Vehicle-to-Network communication, which is similar to the V2I mentioned above, but here the network infrastructure could be cloud, fog, edge, grid networks, or servers at big service providers,

- V2P: Vehicle-to-Pedestrian/Device communication is another possible mode of communication between a vehicle and a device used and carried by a pedestrian or other road user (eg, a cyclist).

If we categorize communication applications by radio access technology (RAT), it could be DSRC (Dedicated Short-Range Communication), WiFi, or cellular. There are some existing potential candidates for V2V, for example, Visible Light Communication (VLC). DSRC is based on IEEE802.11p and Wireless Access in Vehicular Environment (WAVE) internet protocols. Attack scenarios against V2X communication systems are usually exploiting vulnerabilities of standard WiFi protocols (IEEE802.11a/b/g/n/ac), since it was not designed for vehicular mobility.[2]

Based on the performed analysis of the related studies, we see that the V2X communication system will be among the most relevant research topics of the automotive communication domain. It can be explained by its significant impact on road safety, energy efficiency, and traffic control.

Regarding current communication-technologies and architectures, we found that the investigation of Sing et al[2] was the most comprehensive and well-structured study. Also, this review paper gives a good high-level overview of the V2X (Vehicle-to-Everything) communication system, which is the foundation stone of connected cars. Sharma and Kaushik[18] performed an in-depth survey focusing on IoV, which concerns to the emerging concept of intelligent vehicles connected by the cyberspace. The paper also provides a detailed description of IoV and VANETs, emphasizing security requirements, future challenges, and critical attack vectors with an outstandingly high-security risk.[39]

One of the most comprehensive up-to-date surveys in this research field was introduced by Singh et al.[2] This article includes a state-of-the-art survey with the latest details regarding vehicular communication technologies and network architectures and their applications for CAVs.

2.2.7 | Artificial intelligence

AI-based methods can support the operation of highly automated vehicles in many ways. The most important reason for applying AI-based models is the huge amount of data generated and processed during the transport processes, which requires an outstandingly fast, real-time decision-making process. And that is what AI-based approaches are good at. Numerous useful applications focus in this field on Big Data handling, especially in real-time processes, which is a significant problem in the age of connected cars. Therefore, the automotive industry can make good use of these methods in the field of security; since in the case of transport processes extensive real-time data traffic has to be continuously analyzed, evaluated and compared to isolate malicious interventions and detect contradictions, such as:

- malicious software and codes,
- anomaly in the regular operation of the system, or
- segregating possibly adverse processes.

In recent years, machine learning methods and algorithms are among the most investigated fields of AI. Especially deep neural networks, which are primarily applied to detect malicious traffic on the network based on the perceived anomaly.

Commonly used AI algorithms in the vehicle security domain:

- anomaly detection with neural networks (CNN, DNN, RNN and their combinations),
- sequence context anomaly detection with long-short term memory cells (LSTM),
- Support vector machine (SVM) based intrusion detection,
- distributed anomaly detection with hierarchical temporal memory (HTM).

In the light of the performed literature review, Song et al[36] developed an AI-based intrusion detection method focusing on the automotive IVNs. In this article, the authors designed a DCNN (Deep Convolutional Neural Network), which was optimized for the data traffic of the CAN bus. This method has significantly low false rates compared to conventional machine learning algorithms.[2]

The general application possibilities of cybersecurity-related DL (Deep Learning) models were summarized in detail by Mahdavifar and Ghorbani.[40]

2.3 | Constructing a research paper database

There are a huge number of articles written in the field of automotive security. Therefore, one of the most challenging tasks of a review is to identify a general methodology supporting the selection of the reviewed papers. In the first step, we defined a wider range of relevant keywords in the field of automotive security, which, in our case, functioned as a search word. It is important to emphasize that we chose search words based on the reviewed literature.

Accordingly, we did not aim to identify perfectly supplementing categories, but to define the relevant research topics of the automotive security domain. Following this, some of the chosen categories may have overlap with other topics, which means that these research categories may be partly laid on common scientific and professional basics. However, all the considered categories can be defined as separate, scientifically significant, and considerably relevant research orientations.

We performed searches in the most widely used academic search engines and research tools like Google Scholar, Web of Science, Science Direct. Based on our first experiences, we could have concluded that some of the identified categories (keywords) can be merged. Besides this, we also learned that a part of the articles does not contain considerably new scientific or professional results but summarize the previously achieved results or adapt the available knowledge to a specific practical problem. Accordingly, if an article summarized the already achieved results or adapted the available knowledge, we did not include that particular article in the database (Appendix 1). In the next step of the research, we built up a comprehensive data-structure to evaluate the selected articles (see the available dataset). Therefore, we recorded each article's title (with a hyperlink), containing publication-media or -work (such as conference or journal), the number of pages, impact factor according to the Web of Science database, quartile-ranking, and the year of publication. In the following analysis, the investigation focused on the developed database to conclude the relationship between the attributes and the identified categories. During the development of the database, we intended to restrict the set of selected articles to the papers included by the Web of Science database.

2.4 | Development of the analytical model

In the next step, we would like to identify the proper method to classify and evaluate the papers based on their publication attributes. Therefore, we use cluster analysis to assign data to homogeneous classes. The identification of the proper clustering method was primarily fitted to the goals of the current investigation, since cluster analysis does not have a generally accepted methodological guideline referring to the applicability conditions of the different clustering methods.[41,42] Based on the results of Rodriguez et al,[43] K-means clustering can be well applied in our case. K-means clustering aims to minimize the sum of squared errors (SSE) by minimizing the below presented objective function.

$$min \rightarrow SSE = \sum_{i=1}^{k} \sum_{P \epsilon C_i} (P - m_i)^2, \tag{1}$$

where

- k is the number of clusters,
- i is the index of the clusters,
- C_i is the set of objects in the ith cluster,
- m_i is the centroid of the ith cluster, and
- P values are the objects of the ith clusters.

Based on the Elbow method, the number of classes in our case is four.[44] As a programming language, we used Python, and essential data mining libraries: NumPy for multi-dimensional numerical computations, Pandas for organizing data into data frames, Matplotlib for visualization, and Scikit-learn for K-means clustering and for decision tree algorithm.

During the analysis, we considered the following attributes of the developed database: the current IF (Impact Factor) of the journal, the number of citations, and the year of publication.

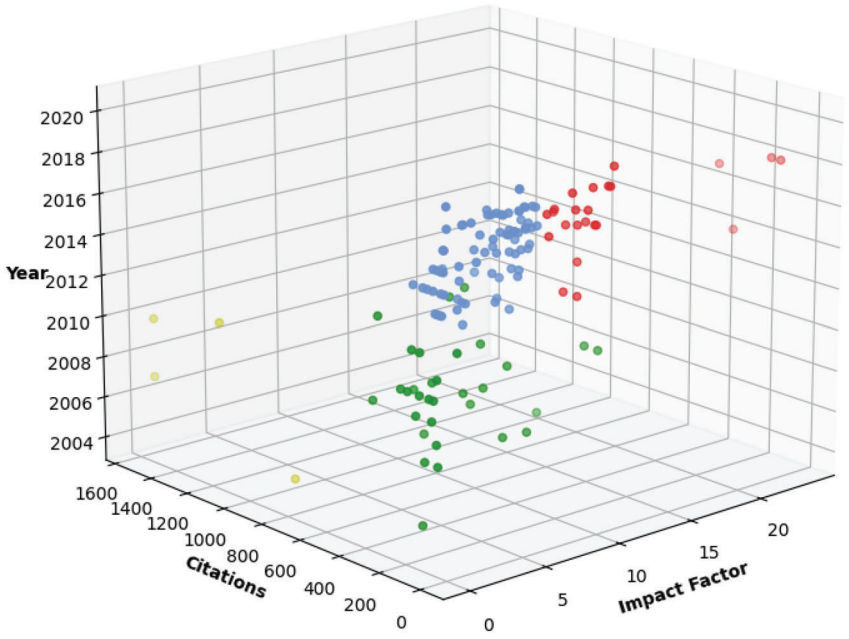

FIGURE 3 Results of the cluster analysis in 3D

We used these three indicators (features) as inputs in the K-means clustering method, which resulted in four separate groups. Figure 3 shows the data points within the clusters, in 3D space with different colors: C0: red, C1: light blue, C2: yellow, C3: green. We will use these cluster annotations in the later discussions.

In the next step, we performed a chi-squared test to verify the significance of the clustering variables. A chi-squared test can be applied to analyze whether two properties of a statistical population are associated with each other or not. The investigated sample is classified into the contingency table's structure by describing the frequency values of the cross-relation table's joint classes related to the two investigated properties.

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\left(k_{ij} - \frac{k_{i.}k_{.j}}{N}\right)^2}{\frac{k_{i.}k_{.j}}{N}}, \tag{2}$$

where

- χ^2 test function,
- $k_{i,j}$ frequency of $A_i \cap B_j$,
- A_i classes of population's first property,
- B_j classes of population's second property,
- $k_{i.} = \sum_{j=1}^{s} k_{ij}$ frequency of A_i event,
- $k_{.j} = \sum_{i=1}^{r} k_{ij}$ frequency of B_j event,
- N size of the sample.

Besides the K-means clustering method, we applied another data mining method, namely the *decision tree analysis* to support the obtained results. The outcomes of the decision tree method led to the probability of the sample subsets meeting the feature conditions, identified by the algorithm.[45] Based on the approach applied by the Python library, split points can be identified based on the Gini method.[46]

$$\text{Gini(D)} = 1 - \sum_{i=1}^{m} p_i^2, \tag{3}$$

where p_i is the likelihood that an arbitrary feature combination D belongs to class Ci.

We investigated the normality of the distributions by applying the Kolmogorov-Smirnov test. In the case of the Kolmogorov-Smirnov test, we assume that X is the observed probability variable, and x_i represents our observations (where $i = 1, \ldots, n$). Based on our observations, we can identify the relative frequency values, $S(x_i)$. Following this, we can evaluate the relationship of our empirical distribution, $S(x_i)$ and the normal distribution, $F_n(x_i)$. So, if X can be characterized by the normal distribution, then the values of the two functions must be close to each other. Accordingly, we identified the maximum of the indicated absolute differences in the case of every i:

$$d_{max} = sup_x|S(x_i) - F(x_i)|. \tag{4}$$

If d_{max} is larger than D_{krit_α}, then we need to ignore the null hypothesis on an α significance level, which assumed that the two distributions are identical.

3 | RESULTS AND DISCUSSION

In this section, we aim to introduce the most relevant results of the performed investigations. In the first step, we briefly introduce the results of the performed statistical analysis. The database contains 38 conference papers and 102 journal articles. In the first step, it is necessary to investigate whether the evaluated attributes of the database can be characterized by normal distribution or not. In the case of the investigated articles, we found that none of considered attributes can be characterized by normal distribution on a significance level of 5%.

Based on the results of the Kolmogorov-Smirnov test, we can conclude that the arithmetic mean value and the SD (σ) parameters cannot be applied to characterize the sample. Accordingly, we did not perform a classical descriptive statistical analysis.

Following the chi-squared test, it can be stated that there is a relationship between the clustering variables (such as IF or YEAR) and the resulted clusters. Accordingly, the test-function values were larger than the critical values in all the investigated cases for DOF = 15 (degree of freedom) (see Table 2).

3.1 | Evaluation of the clusters based on the performed analytical methods

In this section, we introduce the outcomes of the cluster analysis (see the available dataset). Based on the characteristics of the generated groups (see Table 3), we describe the different types of papers and articles. According to our expectations, this approach can help us to identify the relevant research topics of the future in the field of automotive security.

C0 cluster contains highly cited publications with relatively high impact factor (with an average of 12.1) within this field of research. All the papers of this group were published in international scientific journals. In the case of this cluster, the average year of publication is 2017, which indicates that the topics discussed by the papers of this cluster are actual and up to date. V2X and VANET oriented research made up the representative part of this cluster (45%). If we examine the

TABLE 2 Results of the chi-square test, cluster analysis validation

Analyzed cluster variable	Critical value	Test value	H0 hypothesis
IF	24.9	131.2	Dependent/Reject H0
CIT	24.9	86.6	Dependent/Reject H0
YEAR	24.9	137.8	Dependent/Reject H0

TABLE 3 Description of the generated clusters

Clusters	Average IF	Average CIT	Average YEAR
C0	12.1	115	2017
C1	4.23	53	2017
C2	0.94	1273	2008
C3	5.77	185	2010

distribution of the evaluation categories in this cluster, we can draw the conclusion that the most widely investigated topics were V2X, SA, and security and privacy. At this point, it is also reasonable to investigate the changes in the distribution values of the most representative categories. The basis of the comparison should be the year of publication considering the differences in the distribution of the related clusters, containing earlier or newer papers. In the case of V2X, we can observe that this category is similarly dominant in all the four clusters; however, regarding newer papers with high impact factor, there is a significant increase in the ratio of this category. The representation of SA topic is continuously dominant in the clusters 0, 2, 3, independently of the publication year.

This can be explained by the fact that threat modeling is a crucial pillar of cybersecurity. As introduced previously, this domain has developed continuously and numerous general methods have been adopted by the automotive industry (STRIDE, DREAD, EVITA, TVRA, HEAVENS, SAHARA, SINA, CVSS).[27-32]

Comparing the newly published and early published papers, it can be observed that the representation of IVN security topic decreases significantly. In parallel, the research interests regarding the use of AI methods in the vehicle security domain can be characterized by a growing trend. This could suggest that the field of intra-vehicular security vulnerabilities was researched considerably deeply in the past. Therefore, we can expect that it might receive moderate attention in the future. Instead of this, in the case of AI and V2X, we can expect increasing interest. In accordance with this, as a result of the decision tree analysis (Figure 5), we can also conclude that more than 80% of the V2X related articles has a higher impact factor than 1.594 and a higher citation index than 66.5. Almost a third of the articles was published after 2018. Comparing the correlation between the results of the training set and the test set, it can be concluded that the accuracy of the performed estimation is 78%. The high number of references in a short time can be explained by introducing cutting edge technologies like blockchain-based certificate management schemes, edge, fog and cloud computing concepts, the fifth generation telecommunication technology (5G), the concept of C-ITS and connected vehicles, and security issues related to connected vehicles. Based on the performed literature review, keywords of C0 (see the distribution of topics within C0 in Figure 4) could be summarized as blockchain technology, 5G, VANET, V2X, IoV, fog computing, edge computing, cloud computing, C-ITS, machine learning models.

C1 cluster contains the most articles in our database, most of the articles are journal papers and only 25% were published in international symposiums/conferences. The average year of publication is the same as in the first cluster, mentioned above (2017), but the average Impact Factor is not that high (4.23). V2X, AI, and IDS oriented research made up the representative part of this cluster (57%). Beyond the trends described in section focusing on the evaluation of cluster 0, it is reasonable to highlight the dominant presence of AI and IDS topics. Since IDS has traditionally been closely associated with the IVN topic in the field of automotive security, we could expect that the dominance of IDS will also decrease as the ratio of IVN is reduced. Contrarily, the analysis suggests that in the given cluster, this declining trend cannot be observed. Presumably, the reason for this is the increasing use of AI models in field of IVN security. Therefore, the results of analyzing cluster 1 are in accordance with the output of the evaluation introduced in the previous chapter.

In accordance with this, as a result of the decision tree analysis (Figure 5), we can also conclude that 80% of the AI related articles has less citations than 51.5. This also supports our previously obtained results that the topic of AI is more frequently discussed in the recently published articles. Comparing the correlation between the results of the training set and the test set, it can be concluded that the accuracy of the performed estimation is 80%.

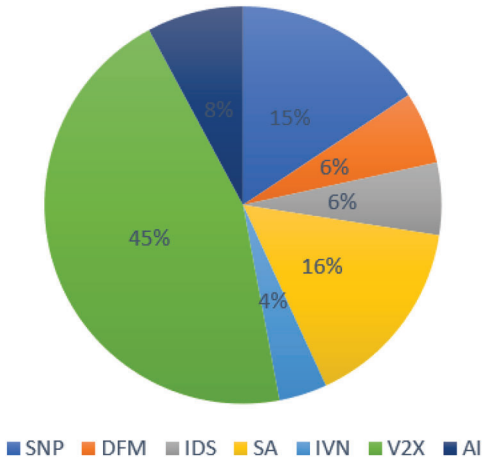

FIGURE 4 Distribution of topics within the first cluster C0

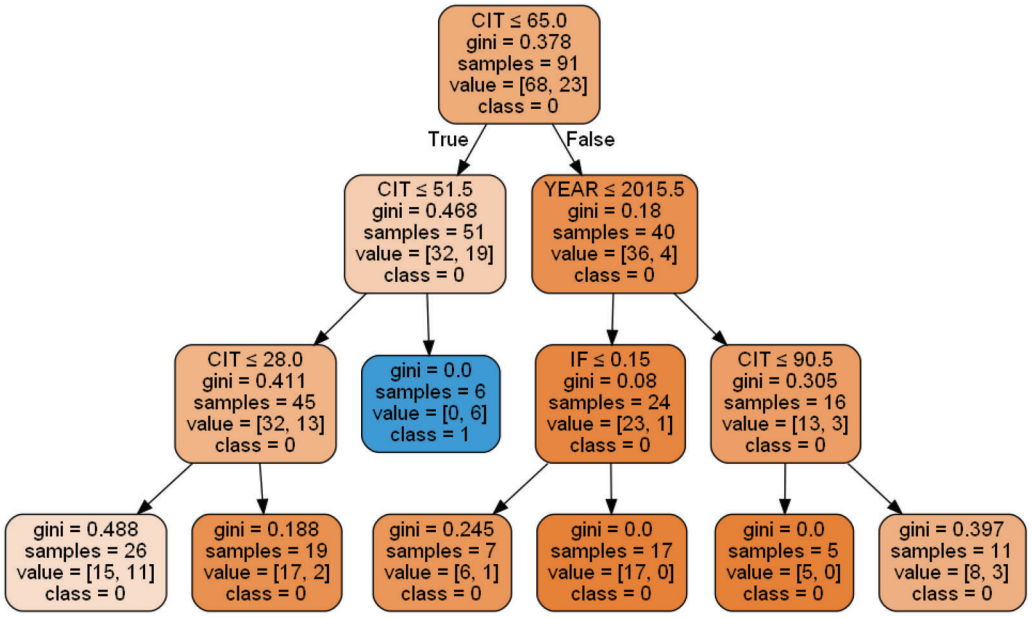

FIGURE 5 Decision tree analysis of the AI topic

A representative part of the articles within this cluster focus on developing new IDS, IDPS systems, mainly for the protection of internal communication networks. Based on the performed literature review, keywords of C1 could be summarized as VANET, machine learning, IDSs, IVN security.

C2 cluster contains the lowest number of articles in our database. Half of the articles are conference papers, and the remaining part of this group was published in scientific journals. The value of average publication year is 2008, but the average number of citations is outstandingly high (1273). Accordingly, we consider these papers as basic articles, referred by numerous newly published studies. Since the number of samples are relatively low in this cluster, it does not seem reasonable to evaluate the distribution of the topics in a detailed way.

Based on the performed literature review, keywords of C2 could be summarized as SPN, SA, attack surfaces.

C3 cluster contains the earlier published articles, half of the articles, are conference papers and the remaining part of this group was published in scientific journals. The value of average publication year is 2010, and the average number of citations is relatively high (185), partly resulting from the earlier average publication year. The average Impact Factor is 5.77. V2X, SNP, IVN, and SA oriented research made up the representative part of this cluster (78%). Based on the performed literature review, keywords of C3 could be summarized as VANET, privacy enhancing technologies, authentication schemes and protocols, IVN.

If we investigate the distribution of the evaluation categories in the first cluster, we can observe that the most frequently investigated topics were attack surfaces, IVN, and general security. This could suggest that these domains were researched considerably deeply in the past. Therefore, we can expect that they might receive moderate attention in the future. Instead of this, the field of AI and DFMs is underrepresented. Hence, we can expect a broad interest in these domains.

If we analyze the distribution of the evaluation categories in the second cluster, we can conclude that the most often investigated topics were IDS, IVN, and attack surfaces. These results should be evaluated in the light of the introduction of the third cluster, which contains most of the newer, high prestigious journal papers. If an evaluation category is strongly represented in the second cluster but underrepresented in the third cluster, that might suggest that research from this certain scientific domain can hardly produce significant new scientific results in recent times, which could be published in a prestigious scientific journal. Furthermore, if these scientific fields are also strongly represented in the first cluster, we can assume that these fields were researched in the past thoroughly, which supports the previously drawn conclusion. The characteristics of IVN and attack surfaces fit the introduced conditions. Therefore, these findings verify the assumption that we can expect less attention to these domains in the future. Instead of this, the field of IDS appears both in the second and third clusters as a strongly represented domain, which shows that this field is in the focus of many scientifically significant research studies.

If we examine the distribution of the evaluation categories in the third cluster, we can draw the conclusion that the most widely investigated topics were IDS, V2X, and attack surfaces. At this point, it is also reasonable to investigate the changes in the distribution values of the most representative categories. In the case of IDS, we can observe that this category is similarly dominant in the first and the third clusters as well; however, there is a significant increase in the ratio of this category. The baseline representation of V2X was moderate (included by the first cluster). However, in the third cluster, V2X becomes a dominant category, which means a relevant development. In the first cluster, the category of attack surfaces is the leading domain. However, in the third cluster, the representation of attack surfaces decreases significantly. This could suggest that the field of attack surfaces was researched considerably deeply in the past. Therefore, we can expect that it might receive moderate attention in the future. Instead of this, in the case of IDS and V2X, we can expect increasing interest.

Taking into consideration the security requirements mentioned in Section 2.2, the following measures can be taken to mitigate security vulnerabilities and prevent their exploitation:

- layered approach: to reduce probability of an attack success and to defend safety-critical functions,
- vehicle development process with cybersecurity considerations: software development approaches like security by design and privacy by design,
- information sharing (transparency with respect to the user privacy): shared attack and vulnerability reports,
- documented process for responding to incidents, exploits, and vulnerabilities,
- self-auditing (penetration test results, risk assessment): risk-based approach to assessing vulnerabilities and potential impacts,
- limiting developer/debugging access to ECUs,
- controlling cryptographic keys,
- controlling diagnostic access,
- control access to firmware: firmware extracting is extremely risky, because this is the first stage of discovering some vulnerability in a vehicular network,
- segmentation and isolation techniques in vehicle architecture design: logical or physical isolation to separate sub-networks and wired/wireless external connections,
- controlling internal communication network: CAN bus spoofing can be eliminated with dedicated ECU sensor inputs,
- continuous event logging: traceability of the occurred events,
- controlling wireless interfaces,
- controlling communication with back-end servers: privacy enhanced PKI certificate management schemes,
- preparing security measures for using aftermarket devices (eg, phones, Bluetooth devices, USB, OBD-II dongles, etc.).[47,48]

4 | CONCLUSION

For the background work for this article, we developed a database from 140 research and review papers from the field of automotive security. In the database, we assigned to every record the following attributes: Impact Factor based on WoS database (in the case of journal papers), year of publication, and the current number of citations at the moment of selection. Following this, we assigned binary values to research papers to indicate the presence of a given topic in every article. Since the relatively high number of categories reduced the transparency of the analysis, we merged the closer topics to facilitate the evaluation. In accordance with this, we finally identified seven categories. It is reasonable to emphasize that, in some cases, these categories:

- use similar theoretical considerations or techniques,
- are developed from a common methodological basis, or
- aim to solve similar problems from a different perspective.

In the next step, we briefly introduced the performed statistical analysis of the data set, which was followed by the identification of the proper method to classify the papers. Based on the presented literature review, two non-classical methods were applied, the K-means clustering and decision tree analysis. In accordance with this, we aimed to identify the most emerging future research topics and orientations of the automotive security domain.

Following the results of the performed analysis, the field of AI and DFMs were previously underrepresented in the research studies. Hence, we can expect broad interest in these domains. The baseline representation of V2X was moderate. However, in recent times, V2X becomes a dominant category, which means a relevant development. Accordingly, in the case of V2X, we can expect increasing interest. The field of IDS, related to field of inter-vehicular communication, appears to be a strongly represented domain in the recent research papers.

Following our evaluation, machine learning methods and V2X related security solutions are expected to receive even more attention in the future, both in the industry domain as well as in the research domain.

ACKNOWLEDGEMENTS

The research reported in this article was supported by the Hungarian Academy of Science (HAS) for providing the Janos BOLYAI Scholarship. Moreover, the authors are grateful for the support of New National Excellence Programme Bolyai+ scholarship.

The research was supported by the Ministry of Innovation and Technology NRDI Office within the framework of the Autonomous Systems National Laboratory Program.

The presented work was carried out within the MASPOV Project (KTI_KVIG_4-1_2021), which has been implemented with support provided by the Government of Hungary in the context of the Innovative Mobility Program of KTI.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Mendeley Data at https://doi.org/10.17632/z4744w5ptv.3.

REFERENCES

1. Zöldy M. Legal barriers of utilization of autonomous vehicles as part of green mobility. Paper presented at: Proceedings of the International Congress of Automotive and Transport Engineering. Springer, New York, NY; 2018:243-248.
2. Singh P, Nandi S, Nandi S. A tutorial survey on vehicular communication state of the art, and future research directions. *Veh Commun*. 2019;18:100164. https://doi.org/10.1016/j.vehcom.2019.100164.
3. Ládi G, Buttyán L, Holczer T. Message format and field semantics inference for binary protocols using recorded network traffic. Paper presented at: Proceedings of the 2018 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia; 2018:1-6; IEEE.
4. Rizvi S, Willet J, Perino D, Marasco S, Condo C. A threat to vehicular cyber security and the urgency for correction. *Proc Comput Sci*. 2017;114:100-105. https://doi.org/10.1016/j.procs.2017.09.021.
5. Sergis A, Hardalupas Y. Anomalous heat transfer modes of nanofluids: a review based on statistical analysis. *Nanoscale Res Lett*. 2011;6(1):391.
6. Maleki MR, Amiri A, Castagliola P. Measurement errors in statistical process monitoring: a literature review. *Comput Ind Eng*. 2017;103:316-329.
7. Tseng GC, Ghosh D, Feingold E. Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Res*. 2012;40(9):3785-3799.
8. Obaid M, Szalay Z. A novel model representation framework for cooperative intelligent transport systems. *Period Polytech Transp Eng*. 2020;48(1):39-44.
9. Bernardini C, Asghar MR, Crispo B. Security and privacy in vehicular communications: challenges and opportunities. *Veh Commun*. 2017;10:13-28. https://doi.org/10.1016/j.vehcom.2017.10.002.
10. Axon L, Goldsmith M, Creese S. Privacy requirements in cybersecurity applications of blockchain. *Advances in Computers*. Vol 111. San Diego, CA: Elsevier; 2018:229-278.
11. Romanou A. The necessity of the implementation of privacy by design in sectors where data protection concerns arise. *Comput Law Secur Rev*. 2018;34(1):99-110.
12. Pohrmen FH, Das RK, Saha G. Blockchain-based security aspects in heterogeneous Internet-of-Things networks: a survey. *Trans Emerg Telecommun Technol*. 2019;30(10):e3741.
13. Koschuch M, Sebron W, Szalay Z, Török Á, Tschiürtz H, Wahl I. Safety & security in the context of autonomous driving. Paper presented at: Proceedings of the 2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE), Graz, Austria; 2019:1-7; IEEE.
14. Halder S, Ghosal A, Conti M. Secure over-the-air software updates in connected vehicles: a survey. *Comput Netw*. 2020;107343.

15. Thing VLL, Wu J. Autonomous vehicle security: a taxonomy of attacks and defences. Paper presented at: Proceedings of the 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Chengdu, China; 2016: 164-170; IEEE.
16. Javed MA, Hamida EB, Al-Fuqaha A, Bhargava B. Adaptive security for intelligent transport system applications. *IEEE Intell Transp Syst Mag*. 2018;10(2):110-120.
17. Rupareliya J, Vithlani S, Gohel C. Securing VANET by preventing attacker node using watchdog and Bayesian network theory. *Proc Comput Sci*. 2016;79:649-656. https://doi.org/10.1016/j.procs.2016.03.082.
18. Sharma S, Kaushik B. A survey on internet of vehicles: applications, security issues & solutions. *Veh Commun*. 2019;20:100182. https://doi.org/10.1016/j.vehcom.2019.100182.
19. Shi L, Li Y, Liu T, Liu J, Shan B, Chen H. Dynamic distributed honeypot based on blockchain. *IEEE Access*. 2019;7:72234-72246.
20. Sharma S, Kaul A. A survey on intrusion detection systems and honeypot based proactive security mechanisms in VANETs and VANET cloud. *Veh Commun*. 2018;12:138-164. https://doi.org/10.1016/j.vehcom.2018.04.005.
21. Lokman S-F, Othman AT, Abu-Bakar M-H. Intrusion detection system for automotive controller area network (CAN) bus system: a review. *EURASIP J Wirel Commun Netw*. 2019;2019(1):1–17. https://doi.org/10.1186/s13638-019-1484-3.
22. Meidan Y, Bohadana M, Mathov Y, et al. N-baiot–network-based detection of IoT botnet attacks using deep autoencoders. *IEEE Pervasive Comput*. 2018;17(3):12-22.
23. Dupont G, Hartog J, Etalle S, Lekidis A. Network intrusion detection systems for in-vehicle network. Technical report; 2019.
24. Macher G, Armengaud E, Brenner E, Kreiner C. Threat and risk assessment methodologies in the automotive domain. *Proc Comput Sci*. 2016;83:1288-1294.
25. Török Á, Szalay Z, Sághi B. Development of a novel automotive cybersecurity, integrity level, framework. *Acta Polytechnica Hungarica*. 2020;17(1).
26. Kadhirvelan SP, Söderberg-Rivkin A. *Threat Modelling and Risk Assessment within Vehicular Systems* [Master's thesis]; 2014.
27. Cui J, Sabaliauskaite G. *On the alignment of safety and security for autonomous vehicles*; 2017.
28. TR ETSI ETSI 102 893 Intelligent transport systems (ITS); security; threat, vulnerability and risk analysis (TVRA). Technical report; March 2017.
29. Lautenbach A, Islam M. HEAVENS—healing vulnerabilities to enhance software security and safety. The HEAVENS Consortium (Borås SE); 2016.
30. Macher G, Sporer H, Berlach R, Armengaud E, Kreiner C. SAHARA: a security-aware hazard and risk analysis method; 2015.
31. Schmidt K, Tröger P, Kroll H-M, Bünger T, Krueger F, Neuhaus C. Adapted development process for security in networked automotive systems. *SAE Int J Passenger Cars-Electron Electr Syst*. 2014;7(2014-01-0334):516-526.
32. Mell P, Scarfone K, Romanosky S. Common vulnerability scoring system. *IEEE Secur Priv*. 2006;4(6):85-89.
33. Ponikwar C, Hof HJ, Gopinath S, Wischhof L. Beyond the Dolev-Yao model: realistic application-specific attacker models for applications using vehicular communication; 2016. arXiv preprint arXiv:1607.08277.
34. Li X, Ma J, Moon SJ. On the security of the Canetti-Krawczyk model. Paper presented at: Proceedings of the International Conference on Computational and Information Science. Springer, New York, NY; 2005:356-363.
35. Sommer F, Dürrwang J, Kriesten R. Survey and classification of automotive security attacks. *Information*. 2019;10(4). https://doi.org/10.3390/info10040148.
36. Song H, Woo J, Kim HK. In-vehicle network intrusion detection using deep convolutional neural network. *Veh Commun*. 2019;21:100198. https://doi.org/10.1016/j.vehcom.2019.100198.
37. Dibaei M, Zheng X, Jiang K, et al. An overview of attacks and defences on intelligent connected vehicles; 2019. arXiv preprint arXiv:1907.07455.
38. Paul A, Chilamkurti N, Daniel A, Rho S. *Intelligent Vehicular Networks and Communications: Fundamentals, Architectures and Solutions*. San Diego, CA: Elsevier; 2016.
39. Fang W, Zhang W, Liu Y, Yang W, Gao Z. BTDS: Bayesian-based trust decision scheme for intelligent connected vehicles in VANETs. *Trans Emerg Telecommun Technol*. 2020;31(12):e3879.
40. Mahdavifar S, Ghorbani AA. Application of deep learning to cybersecurity: a survey. *Neurocomputing*. 2019;347:149-176. https://doi.org/10.1016/j.neucom.2019.02.056.
41. Hennig C, Liao TF. Comparing latent class and dissimilarity based clustering for mixed type variables with application to social stratification; 2010.
42. Von Luxburg U, Williamson RC, Guyon I. Clustering: science or art. Paper presented at: Proceedings of ICML Workshop on Unsupervised and Transfer Learning, JMLR Workshop and Conference, Bellevue, Washington, USA; 2012:65-79.
43. Rodriguez MZ, Comin CH, Casanova D, et al. Clustering algorithms: a comparative approach. *PLoS One*. 2019;14(1):2–18.
44. Dabbura I. K-means clustering—algorithm, applications, evaluation methods, and drawbacks; 2020. https://imaddabbura.github.io/post/kmeans-clustering/. Accessed January 20, 2020.
45. Bhargava N, Sharma G, Bhargava R, Mathuria M. Decision tree analysis on j48 algorithm for data mining. *Proc Int J Adv Res Comput Sci Softw Eng*. 2013;3(6):1114–1119.
46. Wang X. Decision-tree-based relay selection in dualhop wireless communications. *IEEE Trans Veh Technol*. 2019;68(6):6212-6216.

47. Administration National Highway Traffic Safety. *Cybersecurity Best Practices for Modern Vehicles (DOT HS 812 333)*. Washington, DC: National Highway Traffic Safety Administration; 2016:2.
48. Kim S, Shrestha R. In-vehicle communication and cyber security. *Automotive Cyber Security*. New York, NY: Springer; 2020:67-96.

How to cite this article: Pethő Z, Török Á, Szalay Z. A survey of new orientations in the field of vehicular cybersecurity, applying artificial intelligence based methods. *Trans Emerging Tel Tech*. 2021;e4325. https://doi.org/10.1002/ett.4325