applied sciences

MDPI

Review

Explainable Artificial Intelligence (XAI) for Intrusion Detection and Mitigation in Intelligent Connected Vehicles: A Review

Cosmas Ifeanyi Nwakanma [1], Love Allen Chijioke Ahakonye [2], Judith Nkechinyere Njoku [2], Jacinta Chioma Odirichukwu [3], Stanley Adiele Okolie [3], Chinebuli Uzondu [4], Christiana Chidimma Ndubuisi Nweke [5] and Dong-Seong Kim [2,*]

1 ICT-Convergence Research Center, Kumoh National Institute of Technology, Gumi 39177, Republic of Korea
2 Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi 39177, Republic of Korea
3 Department of Computer Science, Federal University of Technology, Owerri 340110, Nigeria
4 Department of Transport Management Technology, Federal University of Technology, Owerri 340110, Nigeria
5 Department of Big Data Solution Architecture, Conestoga College, Kitchener, ON N2G 4M4, Canada
* Correspondence: dskim@kumoh.ac.kr

Abstract: The potential for an intelligent transportation system (ITS) has been made possible by the growth of the Internet of things (IoT) and artificial intelligence (AI), resulting in the integration of IoT and ITS—known as the Internet of vehicles (IoV). To achieve the goal of automatic driving and efficient mobility, IoV is now combined with modern communication technologies (such as 5G) to achieve intelligent connected vehicles (ICVs). However, IoV is challenged with security risks in the following five (5) domains: ICV security, intelligent device security, service platform security, V2X communication security, and data security. Numerous AI models have been developed to mitigate the impact of intrusion threats on ICVs. On the other hand, the rise in explainable AI (XAI) results from the requirement to inject confidence, transparency, and repeatability into the development of AI for the security of ICV and to provide a safe ITS. As a result, the scope of this review covered the XAI models used in ICV intrusion detection systems (IDSs), their taxonomies, and outstanding research problems. The results of the study show that XAI though in its infancy of application to ICV, is a promising research direction in the quest for improving the network efficiency of ICVs. The paper further reveals that XAI increased transparency will foster its acceptability in the automobile industry.

Keywords: intelligent connected vehicle; intrusion detection; safety; security; XAI

Citation: Nwakanma, C.I.; Ahakonye, L.A.C.; Njoku, J.N.; Odirichukwu, J.C.; Okolie, S.A.; Uzondu, C.; Ndubuisi Nweke, C.C.; Kim, D.-S. Explainable Artificial Intelligence (XAI) for Intrusion Detection and Mitigation in Intelligent Connected Vehicles: A Review. *Appl. Sci.* **2023**, *13*, 1252. https://doi.org/10.3390/app13031252

Academic Editors: Agostino Marcello Mangini and Michele Roccotelli

Received: 14 December 2022
Revised: 13 January 2023
Accepted: 16 January 2023
Published: 17 January 2023

1. Introduction

Most major cities across the world are faced with transportation, traffic, logistic, and environmental sustainability problems as a result of the rapid development of the human population and the rise in the number of vehicles on the road [1,2]. Technology has been shown to be a huge help in managing transportation systems sustainably. For instance, the fifth-generation (5G) communication network is an enabler for intelligent transportation systems (ITSs) in smart cities [3]. A recent approach to solving the growing demand for sustainable transportation is the introduction of intelligent connected vehicles (ICVs) reliable for managing road capacity and fuel consumption [1,2]. Another 5G-enabled technology is the single-circuit-board user interface solutions that can collect real-time data from moving cars for traffic control, routing, and trip planning are used [3,4]. This system aids in the effective and efficient delivery of products and services both inside and outside the nation and can address the problem plaguing transportation systems [5]. Some 5G technology and communication protocols such as wireless access in vehicular environment (WAVE) and dedicated short-range communication (DSRC) allow for data collection and sharing between vehicles and in vehicle [6]. Between- and in-vehicle data sharing is known as "cooperative sensing" or "collective perception", thus "cooperative ITS" (C-ITS) [6,7].

The CAR 2 CAR Communication Consortium members build their C-ITS deployment plans on the cooperative vehicle to "other things" (V2X) short-range communications that perform everywhere at any time via local ad hoc networks in the 5.9 GHz band. This cooperative V2X (C-V2X) communication uses the European standard ETSI ITS G5 [7] which is based on the US market IEEE 802.11p WLAN standard specially designed for automotive applications in addition to the WAVE/DSRC, and long-term evolution-advanced (LTE-A) delivering quality of service for most V2X applications [8]. The successor standard, IEEE 802.11 bd, offers improved performance and seamless evolution of the radio technology which ensures efficient use of the allocated spectrum, and continuous operation of implemented services [4,9]. Cooperative V2X systems in vehicles analyze the data received and warn the driver against dangers. By this principle, critical road safety situations and resulting accidents can be avoided [7]. The performance requirements for basic road safety and advanced V2X services are adapted from [10] and summarized in Table 1.

Table 1. Performance Requirements for the Basic Road Safety and Advanced V2X Services [10].

Use Case Group	Transmission Mode	Latency (ms)	Reliability (%)	Maximum Data Rate (Mbps)	Communication Range (m)
Basic road safety services supported by 3GPP Rel-14/Rel-15	Broadcast	10–100	90	31.7	100–300
Vehicles Platooning	Broadcast, groupcast and unicast	10-25	90–[99.99]	[65]	less than 100; [5–10] s max relative speed
Advanced driving	Broadcast	[3–100]	[99.99]–[99.999]	[50]	[5–10] s max relative speed
Extended sensor	Broadcast	3–100	[90–99.999]	1000	[5–1000]
Remote driving	Unicast	[5–20]	[99.999]	Uplink: 25 Downlink: 1	Same as cellular Uplink and Downlink

To meet the demands of efficient connectivity of ICVs (awareness driving, sensing driving, and cooperative driving), the authors in [11] proposed the introduction of the sixth generation (6G) to compensate for the inability of the existing 5G to guarantee the best connectivity for fully autonomous connected vehicles as corroborated by [10]. Moreover, manually driven vehicles profit from cooperative V2X as well as all levels of assistance and automation up to fully self-driving cooperative automated vehicles [12]. Services such as providing information about traffic light signal phases and their predicted changes or barriers on the route in real-time furthermore support smooth and comfortable traveling [12]. By avoiding strong accelerations/decelerations, the fuel/energy consumption of vehicles can be reduced with favored effects on lowering noise and emissions. Sophisticated sensors of vehicles and in road infrastructure are able to detect other road participants [5]. However, common issues of ITSs include but are not limited to: (a) vehicle routing, (b) demand forecasting, (c) traffic prediction for flow and location, (d) processes optimization, (e) arrival time forecasting, and (f) anomaly or intrusion detection [13]. This survey is delimited to the state-of-the-art progress of XAI applications in solving the intrusion detection and mitigation issues of ICV networks. This is important as trust is critical to the C-ITS security as enumerated by the European Telecommunications Standards Institute (ETSI) [14].

It was difficult to determine what caused the accident that killed a woman crossing a road on 18 March 2018, in Arizona, when a self-driving car hit and killed her [15]. As a result, questions about the function of artificial intelligence (AI) and the safety of its integration into the ITS, which is made up of driverless or autonomous vehicles, ICVs,

and the Internet of vehicles (IoVs), were raised [16]. While it is a debate to determine the causes for the misbehavior of ICVs, one of the potential causes is the possibility of intrusion and the potential harm that can occur from compromising ICVs/IoVs. In addition, the use of the "black box" approach in the design of traditional AI solutions makes investigation difficult [17].

AI is an innovative technology used for developing sophisticated systems that can understand and learn things effectively, just like humans do. Real-world problems such as securing traditional transportation and ITS can be solved with more accuracy and speed with AI and big data algorithms [4,9,13,18]. AI systems are capable of learning new things and making accurate decisions through the help of machine learning algorithms, and artificial neural networks (ANN) [19]. The introduction of AI has redefined the transport sector resulting in the ITS. ITS is recognized as a type of AI-based mobility technology that can comprehend and provide satisfaction to end users, markets, and society at large [20]. ANNs, genetic algorithms, simulated annealing, fuzzy logic models, and ant colony optimizers are some examples of AI technologies that support transportation that is deemed innovative, disruptive, and emerging by the World Economic Forum. These technologies are used to address problems with transportation management such as congestion, guaranteeing that journeying times are reasonable for passengers, and boosting the efficiency of the entire transportation network [21]. In addition, AI helps to resolve some of the problems affecting the transportation industry's sub-systems, including traffic management, public transportation, safety management, manufacturing, and logistics [22].

AI growth has resulted in enormous models being used to meet the daily needs of mankind, including safe ITS [23]. However, AI models are challenged by the need for transparency, "simulability", fidelity, and compactness, despite their benefits. The drawback is a result of the systems' "black-box" design, which enables accurate decision-making but leaves out the justification for the decision being made. This then led to the development of "explainable AI" (XAI), a revolutionary technology. This novel idea enhances the dependability and openness of AI-based systems [24]. The concept of XAI was put up as a means of increasing AI's transparency and fostering its acceptability in some important industries [25]. XAI reduces the complexity of AI while increasing the application of AI systems in sectors such as security, healthcare, and transportation [26]. Consequently, several research works have been published on XAI. One common challenge of the plethora of publications on XAI is the difficulty to situate growth in XAI with specific sector unique demands. To address this, several review papers on XAI have been published. However, a general review of XAI still leaves stakeholders with the problem to grapple with their sector-specific needs. Consequently, in this review, we have articulated all XAI models and how they have addressed the unique demand of securing ITS.

Intrusion detection is the process of monitoring and analyzing events in a system or network [27] for indications of potential events, which are breaches or vulnerabilities of security regulations, and acceptable use of standard security practices [28]. Intrusion detection techniques are usually grouped based on the identified activities and the approach used to identify intrusions. Intrusion detection systems (IDS) can be host-based or network-based. Specification-based, signature-based, and anomaly-based detection are the focal intrusion detection approaches. Aside from event monitoring and assessment, IDS normally collects activity information, alerts of critical occurrences via warnings and alarms, and provides relevant reports [23]. IDS is an essential security tool to safeguard networks against adversarial and non-adversarial attacks from malicious traffic [29]. In recent times, machine learning (ML) has aided the development of IDS research. According to the IoT Analytics [30], in August 2022, Nozomi researchers identified three security holes (CVE-2022-29831, CVE-2022-29832, and CVE-2022-29833) that could allow an attacker to obtain information from Mitsubishi GX Works3 (configuration and programming software for certain Mitsubishi PLCs) project files to compromise connected safety CPU modules. Guan et al. [31] also revealed that researchers found 14 vulnerabilities in the infotainment system of several BMW series. A critical part of modern ITSs is the possibility

of interconnected vehicles, known as vehicle-to-vehicle (V2V) connection, IoV in general, the connection of vehicles to other road infrastructures known as (V2I), the connection of vehicles to other "things" known as (V2X), or simply ICVs [1]. The prevailing challenge of connected things is vulnerability to attacks and falsification of data as they get transmitted in the ICVs [32,33]. It is thus critical to detect and mitigate against these attacks in ICVs [2,31].

The past decade has witnessed a significant increase in the research and innovation of IDS for secured IoVs/ICVs [33]. The design and implementation of dedicated, advanced IDS are mandated by demanding controls for real-time operation and data integrity, regular traffic patterns, and a limited choice of telecommunication protocols [34]. Although there are several survey studies on security threats and key management schemes, this article provides an exhaustive contemporary systemic survey of the XAI-based IDS for secured ICVs/IoVs approaches with recourse to confirming their adherence to the key properties of XAI such as fidelity, completeness, simulability, and compactness. Put in perspective, the main contributions of this review are as follows:

1. This study employed the PRISMA article selection approach to acquire articles focused on ITS, IDS, and XAI with a focus on the trends, challenges, and open research issues in ICV security IDS and designs and dynamics.
2. This study reviewed articles published within a five-year duration between 2017 and 2022. This is to obtain recent information, trends in the design of AI-based IDS, and open issues.
3. This study assessed the performance of various XAI techniques, with fidelity, completeness, simuability, and compactness as focus.
4. This study investigated issues of ethics and policy concerns of ICV and safety of road users [35].
5. This study gave an evidence-based technology strategy for evaluating the performance of different datasets, collection methods, and how close to reality they are. We highlight data gathering issues, how some researchers tackled the problem, and testbed-based research.

The organization of this work is as follows: Section 2 gives background information on ITS, vulnerabilities, and justification for securing the IoV network with XAI. It also presented a review of related works while emphasizing the uniqueness of our paper. Section 3 elaborates the employed methodology for the study, highlighting IDS techniques and benchmark public datasets. Section 4 discusses the state-of-the-art frameworks and performance evaluation. Section 5 concludes the study with open issues and future direction. A list of abbreviations in this work is listed at the end of this paper.

2. Background and Review of Related Works

2.1. ITS as an Emerging Transportation Solution

Ensuring sustainable and efficient transport systems include safe travel for every road user is fundamental to economic growth. Roads are a major means of transportation in most cities across the world and provide easy access to jobs, schools, and markets. Therefore, a safe, efficient, and sustainable road transport system is imperative for economic growth. However, there are serious concerns about the negative impacts of transport on human health and the environment including road traffic crashes (RTCs). According to [36], 1.35 million people are thought to die from RTCs every year, and 50 million suffer injuries as a result of RTCs. In addition to the deaths of those who are involved in the accidents, their families and society as a whole are also affected by the aftereffects of traffic accidents, which may include long- and short-term physical injuries for those who survive the collisions [37]. Furthermore, according to WHO [38], the whole cost of RTCs, which includes the economic worth of lost quality of life, costs governments roughly 3% of their gross domestic product. The problem is worse in developing countries because of the lack of suitable and integrated policies and approaches guiding transportation, including failure

to give sufficient consideration to road safety features in the design and construction of roads [39].

More people are becoming aware of the significance of promoting safety in order to achieve successful policy goals. Over the years, various measures aimed at reducing RTCs have been adopted in different countries across the world, most of which have been targeted toward improving behavior, infrastructure, and vehicles. In the past decade, the traditional methods of improving traffic safety by merely deploying traffic lights and signs, using traffic police, etc., have become less efficient and not achieving the intended results as could be seen in increasing crash numbers and associated injury rates. However, recent advances in information and communication technology have contributed to providing other new possibilities, ways, and effective solutions to the transport safety problem. One of these is to use ITS technology, which can be applied in different transport modes and encompasses a very wide range of technologies to deal with different transport issues including transport safety. One of the several advantages when applying ITS technology to transportation projects is to prioritize the safety of all road users, bolster transportation infrastructure, and give road users vital information on safety [40]. This could be done through the integration of advanced communications technologies which focus on both the infrastructure and vehicles including integrated applications between them. It contributes to modifying the way motorists drive by providing information on safety and travel time needed to make informed decisions. It can gather the information required, for instance, to estimate the probability of a collision, identify and confirm accidents, speed up the reaction to traffic incidents, and send out safety messages to road users if an incident happened on their route.

In developed nations such as Japan, South Korea, Singapore, the United States, and the United Kingdom, ITSs are being used and implemented more frequently to increase the efficacy, efficiency, and safety of road transportation systems [41]. However, this is not the case in developing countries where most governments are yet to adopt and implement policies to readily integrate ITSs into the current transport system. Considering the increasing rate of RTCs and associated injury rates recorded in developing countries annually, it is very important and urgent that ITS-based measures are adopted and must be designed to suit the traffic safety situation of these countries by considering their unique socio-economic and environmental conditions [41].

2.2. The Internet of Vehicles Structure and Need for XAI-IDS

The astounding development in ITS has greatly spurred the invention of smart cars, resulting in the concept of "The IoV" which allows vehicles to initiate communication with accessible networks and the environment. Vehicles can exchange and collect data about other vehicles and roads in real-time, a concept known as the ICV [33]. It is an extended application of the IoT in intelligent transportation, designed as a data sensing and processing platform for the ITS [42]. A comprehensive report on the evolution and future prospects of the ICV can be found in [43], while the authors in [44] presented the global progress made on ICV, especially with a focus on China being one of the leading countries in ICV implementation [45]. The IoV architecture consists of three layers, as shown in Figure 1 [46,47], which includes the application (service platform), network (channel for V2X communication), and perception layers (device layer ensuring the incorporation of intelligent devices into the connected vehicles giving rise to the ICV).

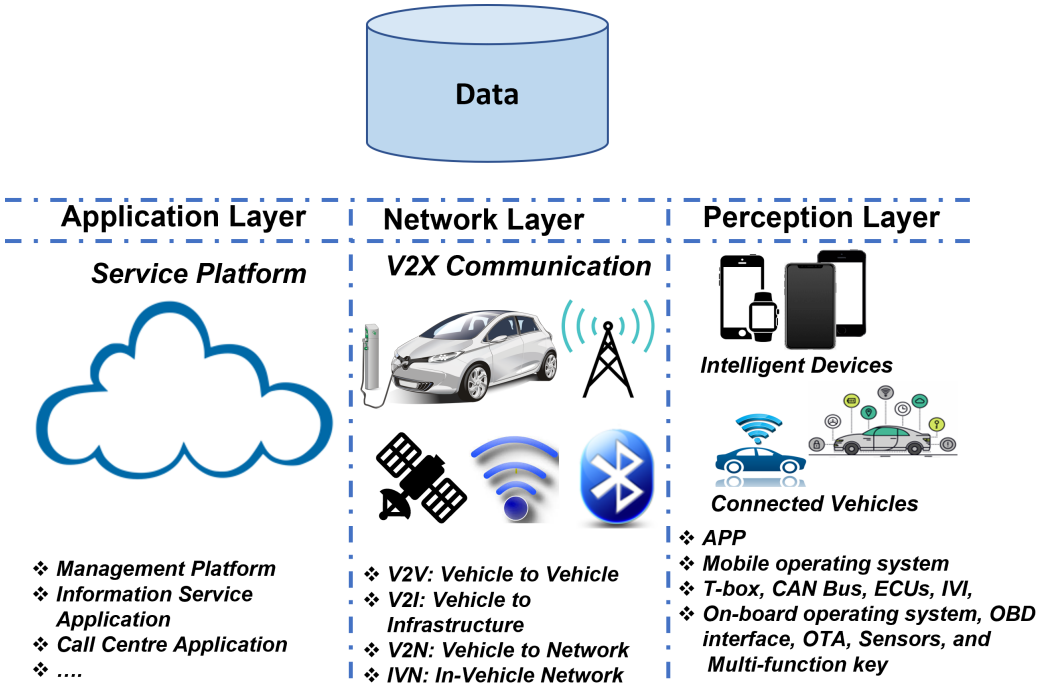

Figure 1. Three-Layer Structure of the IoV and the Associated Functionalities Enabling the Emergence of Connected Vehilces.

Since the IoV is a form of the IoT, it is faced with challenges related to efficient communication between thousands of integrated vehicles, the decision in data processing at the onboard unit, big data integration with IoV, intrusion detection and mitigation, etc. [48]. Additionally, the ability to process the vast amount of data to reduce road congestion, improve the management of traffic, and ensure road safety are parts of the issues in the future IoV trends [49,50]. Hence, AI technology coupled with ML algorithms is capable of improving the network efficiency of the IoV [23]. With ML algorithms, data processing at the onboard units (OBUs), fog level, or cloud level can be resolved. Further issues such as the rapid topology of the IoV, channel modeling, optimization quality of experience, energy, and time can be solved with ML algorithms [50]. To resolve the intrusion challenges, several authors have developed AI and ML models [23,32,47,49]. However, recent works have become devoted to the concept of XAI to resolve the inadequacies of traditional AI. To classify an AI as XAI, the following requirements should be present:

1. *Fidelity:* For instance, it is not enough that AI-IDS for securing an ITS performed with high accuracy, it is now a research concern to know the details of the datasets and how it affected the systems. Is there any element of bias? How are individuals or persons affected by the decisions of AI-based decisions in response to the EU general data protection regulation [35]?
2. *Simulability:* Common questions that are now asked are "can a third party check the correctness of the model?" and "Is it possible to repeat the simulation and arrive at similar results?"
3. *Completeness:* "Explainability" is not enough. It is encouraging to have proper documentation of the model development for a sustainable enhancement of the system [17].
4. *Compactness:* Give human users the knowledge they need to comprehend, properly trust, and successfully manage the new generation of AI partners.

2.3. IoV/ICV Vulnerabilities and Attacks

Security is a critical concern in ITS due to the sensitivity and safety implications of compromised ICVs/IoV. An ICV is constantly exposed and vulnerable due to protocol limitations and limited study mitigation approach, and real-time monitoring [28,51–53].

These attacks occur at the different layers of the IoV architecture and the open systems interconnection (OSI) layers [51,52]. The most common ICV/IoV attacks [54] are as follows:

1. Man-in-the-middle (MiM): This attack occurs when the intruder intercepts network traffic information by gaining access between communication units. It is carried out by monitoring the network, injecting anomalies in the transmission, and forwarding the same to the recipient. A successful attempt assumes the session maintains the connection while the spoofing keeps the attacker unrecognized. This attack can be with SSLStrip, Evilgrade, and Ettercap [28,55,56].
2. Denial of service and distributed denial of service (DoS/DDoS): In this attack scenario, an authorized user is denied access to resources by attacking the availability requirement of network resources [57]. A compromised RTU sends arbitrary packets to the MTU, thereby depleting the network bandwidth and constraining resource availability to users. It disrupts the communication link between the RTU and MTU, making control and process monitoring difficult. It can be with attack tools known as Low Orbit Ion Cannon (LOIC), Slowloris, and GoldenEye [28,58,59].
3. Eavesdrop: This attack comes in two ways, namely, active and passive eavesdropping. An eavesdropping device assesses the wired or wireless network with the aid of tcpdump, dsniff, or Wireshark [28,55,56].
4. Reconnaissance: These attacks seek information about a network and its equipment features. As a result, it is critical to safeguard the sensor measurements from the physical process. Response injection attacks inject misleading inputs into a control system, causing control algorithms to make wrong choices. Fake control commands enter the control system in a command injection attack. It can occur as a consequence of human interference, which results in incorrect control action, or as a result of the injection of false commands, which results in the overwriting of RTU software and field device register values [60].

These IoV vulnerabilities can be grouped into the following five (5) domains: ICV security, intelligent device security, service platform security, V2X communication security, and data security. The ICV intrusion can also be classified based on attacks on the network, software, and hardware connections [52] as follows:

1. Network connection attack: Intrusion on the IoV communication transverse transport, network, and application layers. It targets the exploitation of the OSI model and violates the security goals such as availability, authentication, integrity, and confidentiality.
2. Hardware attack: In this case, the intruder gains unauthorized entry to the IoV system units and violates their operations. Access control is the most difficult aspect of securing hardware.
3. Software attack: The ICV/IoV system uses a range of software to improve its efficiency by satisfying operational demands. Nevertheless, it is prone to trojan horse, SQL injection, and buffer overflow attacks due to inadequate implementation. Since the mobile application is gradually becoming an essential part of the IoV, it has become a hot spot of attack for attackers [61].

Consequently, the need for IDSs, lightweight firewall, hardware encryption, and trusted execution environment have been canvased by stakeholders, as shown in Figure 2 [61].

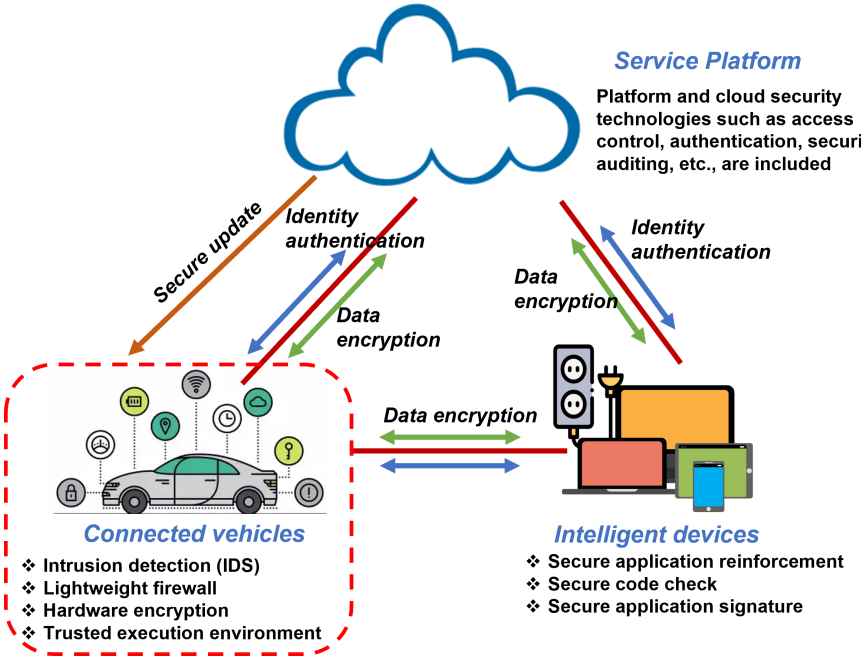

Figure 2. Various approaches to mitigating ICVs' vulnerability to attacks highlighting the role of IDSs, among others [61].

2.4. Securing the ICV with XAI: Background Information

2.4.1. Existing Approaches to Combating Security Breaches on ICVs

Security has remained a big concern among ICV stakeholders due to their reliance on the internet connection, thus exposing them to cyber criminals/hackers. Hackers can take control of a connected car, forcing unintended braking, acceleration, or steering. Privacy invasion can also be a security concern as hackers can gain access to the personal information of users. To solve this, the report from the connected vehicle 2022 summit by Mandal [62] opined that "Escrypt", a cyber threat protection company, was of the view that vehicle data safety could be ensured through blockchain-based communication systems, smart gateways, cyber digital twin, AI-based detectors and other encryption systems [62]. Similarly, the authors of [63] proposed what they considered an advanced AI and ML technique to protect connected vehicles from vulnerabilities related to automated driving, smart charging of electric vehicles, and communication among vehicles or between vehicles and roadside infrastructure. The advanced AI/ML known as "CARAMEl" was demonstrated to secure ICVs from spoofing attacks [63]. This work shows the promising role of AI in securing ICVs. However, XAI promises even more benefits to the ICVs dues to the trustworthiness introduced in addition to the known gains of AI. This paper focuses on the adoption of AI with an attention on the XAI for securing ICVs.

2.4.2. Explainable-AI (XAI) Frameworks and Result Evaluation

The ICV network is confronted with security issues that necessitate more adaptive, automated, and integrated IDS. Explainable AI (XAI) is necessary for mitigating vulnerability and intrusion detection, with explanation and interpretability of classification model decisions [64–67], as well as managing evolving threats and attack mechanisms efficiently. XAI aims to develop ML algorithms that generate more explainable techniques while preserving efficient learning performance, as shown in Figure 3. It also allows human comprehension of the models' actions and decisions. As a rapidly expanding area, it aids in extracting information and the visualization of the results generated with maximum transparency [68].

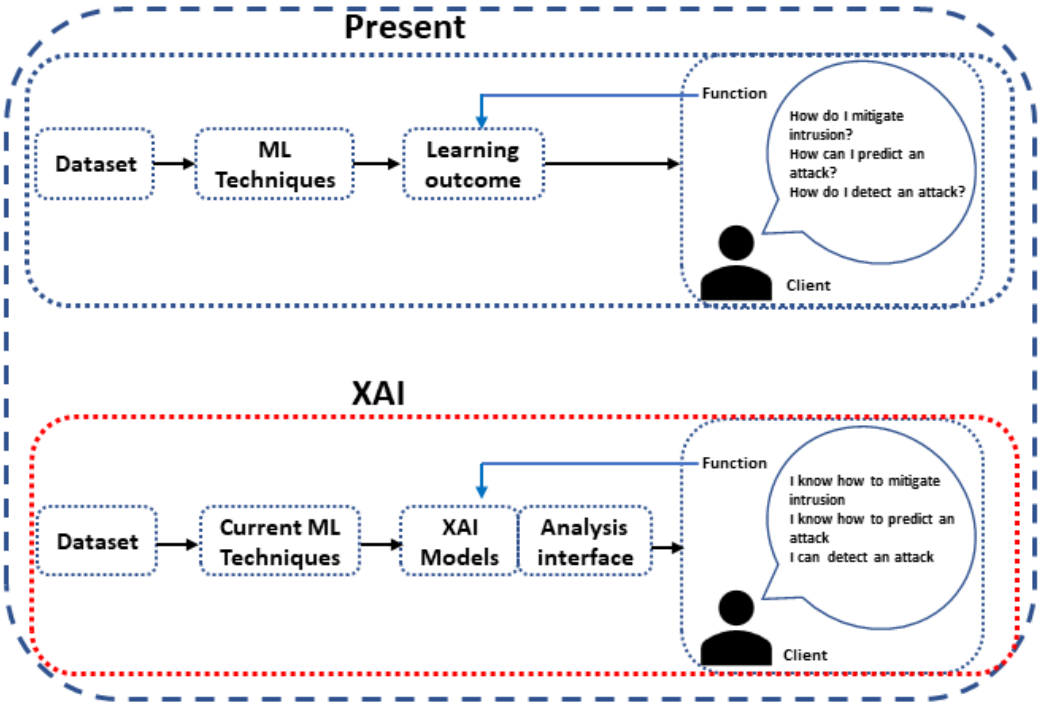

Figure 3. The concept of XAI.

The national institute of standards (NIST), part of the U.S. Department of Commerce, enumerated five types of "explanations" for AI. These were enumerated in [69] as follows:

1. Inform the subject of an algorithm: In the case of the IDS for ITS, an explanation of how the AI model guarantees the security of the system from intruders is critical.
2. Comply with compliance or regulatory requirements: As AI algorithms gain importance in regulated industries, they must be able to show that they follow rules. For instance, self-driving AI algorithms should detail how they adhere to any applicable traffic laws.
3. Build social trust in AI systems by using explanations that support the model and approach rather than focusing on specific outputs. This could involve detailing the algorithm's goals, development process, data used, and sources, as well as its advantages and disadvantages [70].
4. Help with future system development: In order to improve an AI system, technical employees must comprehend where and why a system produces incorrect results.
5. Benefit the owner of the algorithm: Businesses are implementing AI across all sectors in the hope of reaping considerable rewards. For instance, a streaming service benefits from recommendations that are easy to understand and keep people subscribed.

2.4.3. Practical Implementation of ML and XAI-Based Models

The authors of [71] employed layer-wise relevance propagation (LRP) to decompose input significance ratings in gait classification using convolutional neural networks (CNNs). LRP is a later technique for generating interpretations for machine learning model predictions [71,72]. It operates in the input space, where clinical practitioners typically analyze signals. LRP divides a function's prediction $f(y)$ for an input data variable y into element and time input significance values S_i for each input data y_i. It allows the interpretation of the prediction of an ML model as the individual roles of each input data. LRP enables the comprehension of the data a model employs for prediction. It verifies the significance of the ground reaction force characteristics correlating to the areas with the highest input relevance scores. A highly skilled diagnostic expert examined the experimental results of the derived relevance scores of LRP from a medical point of view. Another study [73] demonstrates how the combination of deep neural networks and SHapley Additive exPlanations (SHAP)

or "DeepSHAP" can interpret the impacts on ML for spoofing identification. According to the experimental results, SHAP analysis highlights the consideration adapted by a given classifier at reduced spectro-temporal ranges. In [74], the authors investigate side-channel website fingerprinting attacks by employing XAI methods. Because the dataset includes both side-channel measurements and captured network traces, the XAI technique was to determine the most prevalent network request types that significantly impact website detection. The ROAR metric was used in the study, which demonstrated that LIME and saliency maps accurately detect the most dominant features in side-channel measurements. A study that interprets the performance of mobile traffic classification using DL methods [75] gives specific insights for performance enhancement. It involves the analysis and evolution of a multimodal-DL approach called mimetic-enhanced using explainable artificial intelligence tools. The strategy was evaluated based on the predictability of trust and confidence using calibration and explainability using SHAP-based methods. Meanwhile, adjustment by focal loss attains a 6X reduction over the uncalibrated case, resulting in a substantial increase in trustworthiness. Considering interpretability, a global interpretation for each modality was obtained, measuring the relevance of each and emphasizing the importance of retaining high payload, despite the enormous majority of encrypted traffic.

2.5. Review of Related Works

2.5.1. Review of Survey on XAI Related Works

The concept of XAI has penetrated several fields such as healthcare [76–78], regression modeling [79], anomaly detection [80], power system emergency control [81], fault detection [82], landslide prediction [83], measuring uranium enrichment [84], online game prediction [85], civil engineering [86], and no-teardown vehicle component cost estimation [87]. Thus, the application of XAI to the ITS is attracting a lot of attention from researchers.

Although some review papers on XAI have been written, to the best of our knowledge, this is the first attempt to conduct an extensive review of XAI's applicability to ITS with a focus on intrusion mitigation. The authors in [88] presented a review on XAI without delimiting any field. They focused on the general concepts, applications, emerging issues, and approaches of XAI designs. On the other hand, the authors in [17] reviewed XAI with a focus on the major requirements and agenda setting for the future implementation of XAI. Specifically, they explained the provenance information concept and its applicability to XAI. Gossen et al. [89], as a follow-up to their previous work [90], posited that algebraic aggregation [90] combined with semantic analyses solving for infeasible path reduction, has an impact on both explainability and velocity. Specifically, the work indicated their impact concerning running time [89].

Similarly, the authors in [77] reviewed the applicability of XAI to the healthcare sector while focusing on data analysis and interpretation of models. The mention of XAI's applicability to the transportation sector was only done in passing. Focusing on the emergence of XAI for Industry 4.0, the authors in [91] provided a comprehensive survey of XAI-based approaches adopted to meet their demands for automatic and real-time implementation. In [79], the authors present a review aimed at establishing the unique features of XAI for regression models known as XAIR. The review provided a brief overview of the general concept of XAI before narrowing it down to XAIR, emphasizing the important specificity of the regression problem that necessitates XAI adaptation. Moreover, authors in [92] reviewed XAI from the perspective of data and knowledge engineering, enumerating the existing works of XAI for data-driven and knowledge-aware scenarios.

The closest review of XAI for security is presented in [93]. The authors of this study provided an overview of XAI concepts, trends, enabling technologies, and applications for system security. However, the reviews confirm the hypothesis that review works available on XAI and the security of ITS is limited. Due to the scarcity of research on XAI for ITS security, this study provides an exhaustive contemporary study with a systematic and comprehensive review of the extensive research on XAI-IDS for ITS. This study differs from

previous XAI reviews in that it is the first to systematically survey research works and directions in the use of XAI for the security of ITS limited to IDS.

2.5.2. Review of Survey on Security Issues of ICVs

Guan et al. [31] gave a comprehensive survey on the impending cybersecurity for ICVs detailing the challenges and constraints of attack mitigation, three common approaches to IDS in the vehicle network as well as three emerging technologies to solve the network security problems of ICVs. The identified IDS approaches are (1) feature-based detection, (2) information theory and statistical analysis detection, and (3) detection based on machine learning. Although the authors highlighted the need for a robust machine learning-based IDS, they did not clarify what constitutes a robust ML. The XAI in the view of this paper is a promising and robust solution for IDSs for ICVs.

Mitigating the vulnerabilities and threats to ICVs at the early stage of vehicle development has been acknowledged by automakers to be a veritable means of securing ICVs with minimal cost. Luo et al. in [94] presented a comprehensive survey on threat analysis and risk assessment (TARA) for connected vehicles. They [94] concluded among other issues that the large amount of data collected by ICVs can bring many possibilities for TARA as it could guarantee the accuracy of threat model training. This data-driven TARA process is a new research direction and it will be promising to see the possibility of incorporating TARA and XAI to secure ICVs.

Dibaei et al. [95] provided existing defense mechanisms against attacks on ICVs while preferring research directions. The defense mechanisms listed corroborated with other researchers and included cryptography, software vulnerability detection, malware detection, and network security. The paper recommended the use of deep learning over existing ML. However, the issue of computational complexity, huge data traffic, and latency will necessitate the need for deep learning on edge computing. In all, no position was taken on the important role of XAI.

Following the same pattern as other reviewed works, Banafshehvaragh and Rahmani [96] exhaustively discussed all detection approaches for smart vehicles where they proposed high-performance intrusion, anomaly, and attack detection methods for all communication dimensions of smart vehicles, particularly external and internal network vehicles. Their statistical study of the detection evaluation criteria further reaffirms the position of this paper that XAI is yet to be studied and explored by most authors. One argument in favor of this is the issue of how to manage the practical implementation of XAI while handling the challenge of computational complexity.

In conclusion, Wang et al. [97] noted that the application of AI algorithms for in-vehicle communication has become a new research hotspot. They suggested improving the quality of the data set, reduction in the interference caused by emergencies, improving and optimizing the AI algorithm, and improving the ride comfort of the road.

2.5.3. Summary of Related Works and Research Gaps

From the above review of surveys both in XAI and ICV intrusion detection schemes, it is evident that the XAI applicability to ICV is yet to be fully studied. This review work makes a bold attempt to draw the nexus between XAI and the security of ICVs. Although there are survey works on XAI, to the best of our knowledge, none have considered the applicability to the security of ICVs. On the other hand, despite the numerous surveys on the security of ICVs and recommendation for AI, there is a limited survey available on XAI applicability to the security of ICVs. Thus, this is the first attempt to conduct a comprehensive review of the use of XAI for the security of ICVs.

3. Review Methodology

This section offers a systematic description used for the detailed review. The innovative reviewing methodologies in this study are inspired by the meta-analysis (PRISMA) [98,99] and the "mentefacto conceptual design" [100]. Articles published between 2017 and 2022

were given priority during selection. Nonetheless, the year of publication is irrelevant when the need arises for a historical perspective in the review. According to [101], computer science and engineering-related studies can preferably come from the following databases: IEEE Xplore, ScienceDirect, Wiley, Springer, Taylor and Francis, and select social sites such as Academia and Researchgate, Sage, and Google Scholar. In addition, only papers written in the English language we considered for final analysis. Using the key search "ITS", "IoV", "ICV", "IDS", "AI", and "XAI", a summary of papers according to the database source is listed in Table 2. Similarly, the PRISMA flow diagram for the systematic reviews and the basis for the final selection of articles are shown in Figure 4. Quantitative analysis was done using only XAI papers for IDS. Therefore, the results of the screened literature were systematically summarized and reported by a narrative comprehensive analysis. Specifically, the steps of this analysis include determining the review problem, sorting out and comparing the data, and drawing conclusions [97]. The inclusion criteria for papers used in the survey are:

1. Articles must be original articles published in journals, arXiv, or conference proceedings.
2. Except for the purpose of history or background, only papers published between 2017 and 2023 were considered for final inclusion for discussion.
3. For qualitative analysis, only papers that addressed the issues and concerns of ITS and ICV security using AI/XAI were considered.
4. In comparing this review paper with recent review works, ICVs, security, and AI must be covered to qualify for comparison.
5. The papers have to be written entirely in English.
6. Finally, the papers whose databases had access restrictions were excluded because the authors could not access them.

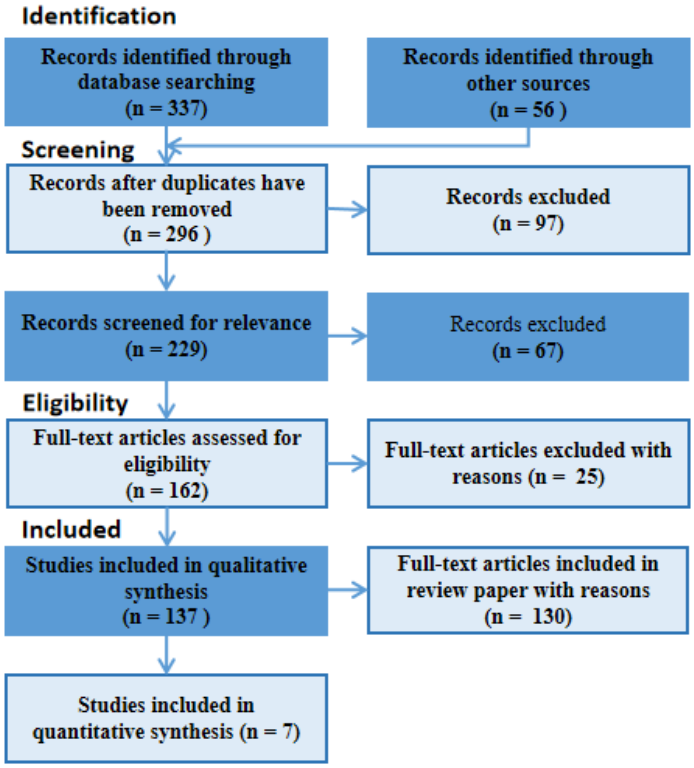

Figure 4. PRISMA flow showing the process of the final selection of the 137 papers at the reference and the 7 specific articles that focused on XAI for IDS in ICV.

Table 2. Publications used in this review.

Database Source	No. of Documents	% Freq
IEEE Xplore (Journals)	52	37.96
IEEE Xplore (Conferences)	12	8.76
Taylor and Francis	4	2.92
Wiley	4	2.92
MDPI	8	5.84
Springer	11	8.03
ACM	4	2.92
ArXiv Pre-print	9	6.57
Google Scholar	7	5.11
ScienceDirect (Elsevier)	15	10.95
Other sources (Blogs, Reports, and Websites)	11	8.02
Total	137	100.00

The summary of document search and usage are summarized in Table 2 and Figure 4, respectively. A total of 393 (337 + 56) documents were identified by the search. A total of 97 documents were excluded due to duplication, leaving a total of 296 for screening. Sixty-seven were further excluded after screening for relevance and elimination of papers with open abstracts but restricted access to full paper content. Of the 162 documents remaining, 25 were excluded using the inclusion criteria above. Thus, a total of 137 documents were used for the survey. A total of 130 of these were used for qualitative analysis, while the other 7 (see Table 3) were used for the specific review since they were strictly papers on XAI for IDSs, while 10 papers (see Table 4) served the purpose of the XAI framework details for background studies.

Table 3. Publications on XAI-based IDS solutions for IoV/ICV.

Author	Approach	Aim	Performance	Year
[54]	Proposed explainable deep learning to secure IoV using the SHAP mechanism	To increase the DL-based IDS' transparency and resilience in IoT networks	The experimental findings demonstrated the proposed framework's strong performance with a 99.15% accuracy and a 98.83% F1 score, highlighting its capacity to defend IoV networks from complex cyber-attacks.	2022
[92]	Extensive review of XAI approaches to data-driven and knowledge-aware scenarios such as ITS	To provide state-of-the-art evaluation metrics and deployment applications in industrial practice	The knowledge of taxonomies and trends in XAI for data-driven applications will enrich the designs of future XAI systems	2022
[102]	Introduced a novel VisExp approach for IDS for in-vehicle networks	Aims to detect and mitigate CAN bus attack in in-vehicle networks	The proposed approach gave a promising result, making the VisExp a potential candidate	2022

Table 3. *Cont.*

Author	Approach	Aim	Performance	Year
[103]	Integration of ANFIS and human–computer interface platforms to enhance the understanding of UAV behavior	The approach is to translate the ANFIS output to linguistic value easily understood by human	The proposed approach shows the potential application of ANFIS for the development of XAI	2019
[104]	A novel XAI applicable to IoT generally including IoV	To ensure a robust explanation of IDS decisions in detecting and mitigating attacks in IoT	Developed multiple XAI models such as SHAP, and RuleFit to aid the deep neural network for transparency and trust	2022
[105]	A majority vote ensemble approach combined with recursive feature elimination-extreme gradient boosting	Intended to provide a more accurate solution by combining the most viable features and prediction from various classifiers	The experimental result shows that the proposed approach improved accuracy, F1-score, and recall while reducing miss rate, compared to previous techniques	2021
[106]	A multilayer, data-driven cyber-attack system	To enhance ICS cyber-security by covering a wider attack scope utilizing the defense-in-depth concept	Experimental results show that the proposed approach had a high detection accuracy	2019

Table 4. Selected Publications Showing Usage of XAI Frameworks.

Author	XAI Framework	Specific Usage	Detail	Year
[54]	SHAP mechanism	To increase the DL-based IDS' transparency and resilience in IoT networks	Provided global and local explanations to XAI models using SHAP plots	2022
[102]	Novel VisExp	Enhance the trustworthiness of the XAI-powered IV-IDS	Based on SHAP, compared the proposed knowledge-based VisExp with a rule-based explanation	2022
[103]	Integration of ANFIS	Approach is to translate the ANFIS output to linguistic value easily understood by humans to enhance the understanding of UAV behavior	Proposed approach shows the potential application of ANFIS for the development of XAI	2019
[104]	RuleFit, SHapley, and SHAP	Multiple XAI frameworks built and integrated to ML/DL-based IDS	Efficient, transparent and trustworthy IDS for IoT applications	2022
[72]	SHAP	A unified framework for prediction analysis	Conceptual identification of a new class of additive feature importance measures that possesses desirable features	2017

Table 4. *Cont.*

Author	XAI Framework	Specific Usage	Detail	Year
[107]	LIME	Explains the prediction of any classifier model with ease while ensuring local fidelity with speed	Has the flexibility to be applicable to different models such as random forests and image classification	2016
[108]	TRUST	To ensure fast and accurate XAI numerical applications using factor analysis to transform input features	Model-agnostic, and high-performing, applicable to numerical applications	2021
[109]	LORE	Blackbox outcome explanation	Agnostic approach to learning the local interpretable predictor for decision rule-based explanations	2018
[110]	GRAD-CAM	Heat map for each class of a single image	Uses the feature maps produced by the CNN. It is model-specific	2019
[111]	CEM	Leverages the features needed to predict that input instance are of same class	Agnostic model, local and for post hoc explanations	2018

4. Findings and Discussion

4.1. ITS/IoV Intrusion Detection Systems (IDS)

An intrusion is any unauthorized activity that disrupts or damages a network. This means that any incursion that violates the integrity, confidentiality, or availability of the ITS is an intrusion. The objective is to detect any abnormal activity in an IoV network traffic that a standard firewall cannot detect. It is critical to attain high levels of security against actions that jeopardize system availability, integrity, or confidentiality. IDSs can be classified based on the detection or deployment approach, and Figure 5 depicts the classification of an IDS.

From the standpoint of the deployment-based IDS approach, the classification of IDS can be either host or network based. The host-based system monitors all activity on the single host and examines security vulnerabilities and intrusions. The major limitation with this system is that it must be on all hosts that need intrusion protection, resulting in increased computational complexity for each node and eventually degrading IDS performance [112]. On the other hand, a network-based system is to mitigate intrusions on all devices and the entire network. The network-based system constantly monitors network traffic and investigates for intrusions.

On the other hand, the detection-based IDS approach can be either signature or anomaly-based. The signature-based system (also known as knowledge-based IDS) defines a signature for attack behaviors derived from footprints left by each intrusion. These signatures are in the signature database, and data patterns are matched with these signatures to detect attacks. One advantage of this system is the high detection rate due to the database's availability of attack signatures residents. However, the approach is unsuitable in real-time due to the lack of signature patterns; therefore, it is incapable of detecting new attacks. However, the approach considered in this study is the network and anomaly-based IDS, which focuses on various AI approaches.

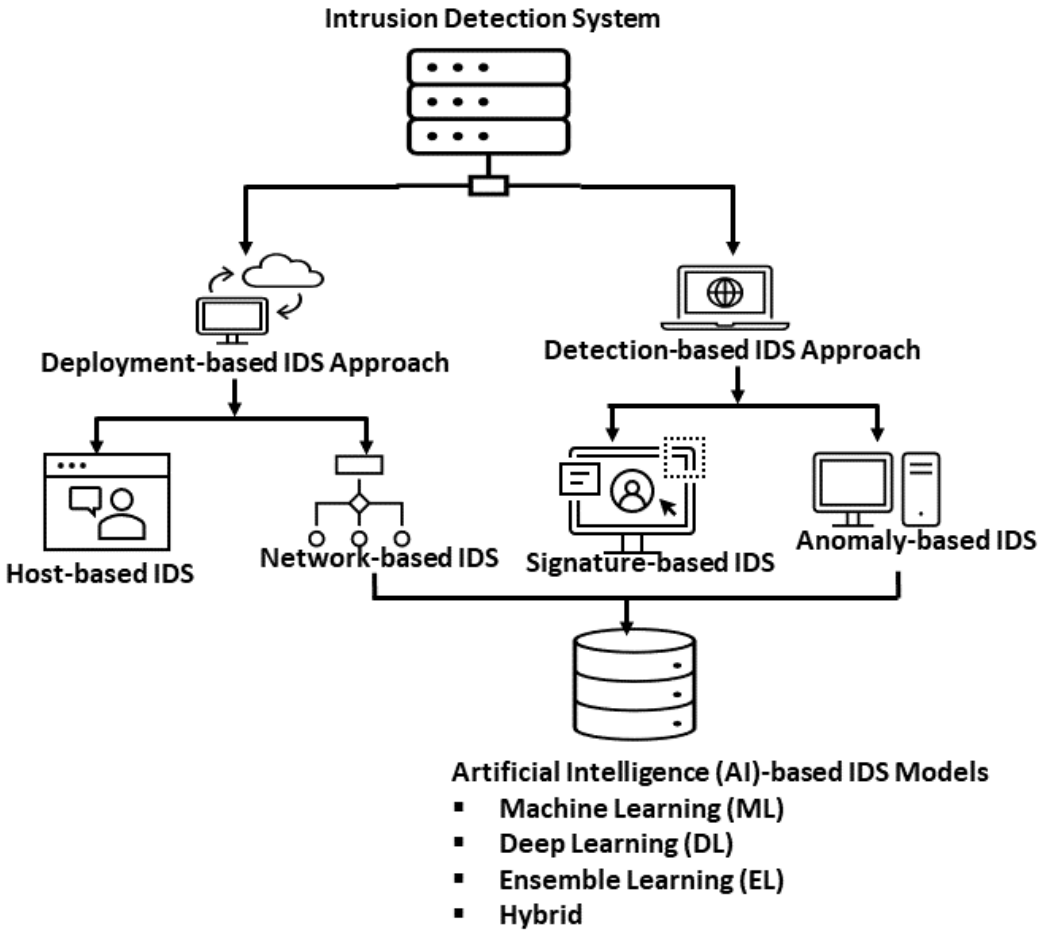

Figure 5. Classification of Intrusion Detection Systems.

AI-enabled IDSs are an effective security measure to combat invasions. Conventional IDSs include rule-based, signature-based, flow-based, and traffic-based techniques. Because most connections and traffic in IoV networks were previously specified, these IDS detect aberrant activity. For example, when the intruder establishes new connections to the victim or delivers a different type of traffic, the network will see anomalous data flows [113,114]. Unfortunately, due to the frequent network, which results in constant topology changes, conventional IDSs do not perform adequately. In addition, intelligent IDSs are necessary to counter the emergence of new forms of attacks or wisely planned intrusions such as man-in-the-middle (MiM) attacks.

AI-based IDSs are often helpful if an intruder disrupts network traffic communication. It is difficult to detect the breach if the attacker does not connect with any network components. To compromise network operations, the intruder must disrupt the network somehow. AI techniques can detect anomalies that are difficult to detect by a human. The capacity of AI techniques to detect even minor anomalies distinguishes them from other types of IDSs. AI-based IDSs can be constructed with a shifting target to create a secure network. Because attacks are continually developing and new vulnerabilities occur daily, the capacity of AI-based models to adapt and grow is precious. It is why signature-based IDSs are becoming outdated, thereby giving way to anomaly-based ML IDS as an emerging trend [113].

AI-based techniques are apt for provisioning authentication and authorization of security control and network traffic patterns, which require high learning sensitivity and robustness. They have been helpful as detection and mitigation tools against data integrity threats and snooping intrusion by targeting the availability of specific attack features. Their

superiority to network analyzers is in automation and disinclination to human error [115]. Below are some AI-based IDS frameworks for ITS/IoV:

4.1.1. Artificial Intelligence (AI) Frameworks and Result Evaluation

Vehicular networks or IoV constitute a critical part of the ITS. Its dynamic nature, as well as the exponential increase in the number of connected devices as a result of B5G technologies, has resulted in cyberattacks and the need for security. AI has played a critical role in securing connected devices/things such as the IoV, and vehicular ad hoc networks (VANETs). In [23], the authors proposed an optimized support vector machine for the protection of VANET from DDoS attacks. Although the proposed model achieved an accuracy of 99.33%, it cannot be said to be an XAI as no evidence of the XAI principles was presented by the authors.

Additionally, the authors in [32] proposed a random-forest-algorithm-based IDS for detecting false basic messages in IoV. The accuracy of the proposed model was 99.6%. The work was based on the vehicular reference misbehavior (VeReMi) dataset. Using the same VeReMi dataset, the authors in [116] developed a distributed IDS for the detection of misbehavior and position falsification of VANETs using an ensemble of k-nearest neighbor (KNN) and random forest. Although the trio of [117–119] proposed ML and deep learning (DL) solutions for the security of ITS (such as autonomous vehicles (AVs), and IoV), their choice of the dataset can be challenged as there could have been more suitable ITS-specific datasets such as the VeReMi [120], BursTADMA [121], or VeReMi extensions, as used by the authors in [115].

Moreover, excessive demands for handcrafted feature engineering are a common challenge of ML research, particularly intrusion detection [122]. Moreover, traditional ML schemes are not well suited for multi-classification problems. However, studies have shown the adaptability and flexibility of ML-based IDS over DL-based IDS. ML-based IDS has shown tremendous potential in improving detection accuracy with minimal computational complexity and time [123,124]. It has led to the adoption of deep learning approaches [125].

ML algorithms can detect seemingly tricky attack patterns. An ML-based IDS has the operational aim of delivering a secure network. Because attacks and attack methods are continually developing and new vulnerabilities occur daily, the capacity of ML models to adapt and grow is precious. It is another reason signature-based IDS are declining while anomaly-based IDS based on ML are advancing.

4.1.2. Application of XAI-Based IDSs for ICVs

The explainability and call for the trustworthiness of the ML have motivated the need for XAI. Table 3 is a summary of some current studies on XAI-based IDS for ITS, showing their performance, approach, and year of publication.

The works of [54] proposed the use of SHapley Additive exPlanations (SHAP) to enhance the transparency and resilience of deep learning-based IDS for IoV networks. This approach gave an accuracy of 99.15% regarding its ability to protect IoV networks against attacks. However, the major obstacle is the challenge of computational complexity.

Recently (September 2022), the authors in [102] employed the XAI for trustworthy In-vehicle IDS leveraging CAN bus data. They further enriched the body of XAI by proposing a novel "VisExp" approach for visualizing the explanations necessary for gaining trust in the system. However, the limitation of the approach is the reliance on a survey formatted using Google forms. This has the potential of skewing the results.

In [103], the authors proposed a rule-based XAI leveraging the adaptive neuro-fuzzy inference system (ANFIS) for understanding unmanned aerial vehicles (UAV) maneuvering decisions. In addition, they developed a human–computer interface (HCI) where the outputs of the ANFIS are translated to linguistic values. This is a promising approach to aiding the monitoring and mitigation of possible intrusion into UAV networks.

Interestingly, the authors of [104] proposed novel XAI-based deep learning solutions for IDSs in the IoT generally not limited to the IoV. However, the approach is similar to other authors as they used SHAP and RuleFit. The use of SHAP has enjoyed dominance due to its ease of use and flexibility in generating the basis for the visualization of explainable results [86].

4.2. Overview of Datasets

This section gives an overview of the datasets used by the researchers to test their presented techniques' performance.

1. The VeReMi dataset: This dataset [120] is an ITS-specific dataset that captures malicious messages intended to trigger falsification, and hence serves as a model for evaluating falsification detection models in an IoV network. The dataset consists of vehicle onboard message logs, including a ground truth labeled, generated from the simulation environment of the Luxembourg City Vehicle Network. The Luxembourg City Network is a smart city that greatly represents an ecosystem enabled by interaction among technologies such as IoT, AI, open data platforms, autonomous cars, smart lights, and wearable devices. Using LuST and VEINS, the simulated network generated GPS data on local vehicles, as well as sets of BSMs, received from other vehicles through DSRC. The classification process was carried out using *Maat* to execute multiple parallel detectors. The initial dataset contains a number of simple attacks with 614,940 observations and 17 features. A detailed description of the dataset can be found in [120].
2. The BurST-ADMA: The BursTADMA dataset [121] was released in March 2022. The BurST-ADMA is an Australian motorway dataset made up of onboard vehicle message logs generated on a Burwood road map that connects Melbourne's city to the suburbs. Vehicle trajectory information was extracted from the Simulation of Urban Mobility (SUMO) road traffic simulation framework and false data were injected into the retrieved trajectory information for a total duration of 1000 s. The BurST-ADMA dataset contains 207,315 observations with 179,126 normal BSM data and 28,189 false data. Seven (7) different falsifications are labeled on the BSMs and one normal vehicle data type. Each of these data points contains the time-step, the vehicle's ID, its X and Y location coordinates, and differential information such as heading, speed, acceleration, and labels [121].
3. V2X falsification dataset: The scenario created here mimics a malicious program or a malfunctioning sensor by injecting false data into the vehicle's positional and speed readings. To model the transmission of CAM in the vehicle network, two of the most well-known simulators SUMO and Network Simulator-3 (NS-3) were used on a VSimRTI platform, a Java-based platform that can seemingly couple the operations of both simulators. The falsification attack is divided into attacks with several parameters: speed, acceleration, and heading. Hence, four (4) different scenarios were explored where vehicles broadcast messages using different strategies and service platforms capturing all simulated false data injection scenarios [126].
4. Car-hacking dataset: The car-hacking dataset [127] was used in conjunction with traditional IDS datasets by the authors in [118] to evaluate the performance of using rule extraction methods from deep learning neural networks to implement a two-stage IDS for ITS. On the other hand, the authors in [117] used the same dataset to validate their proposed enhanced multi-stage deep learning framework for detecting malicious activities from autonomous vehicles. The dataset comprised DoS, fuzzy, and RPM/GEAR attacks, respectively, as detailed in [127].

4.3. Explanable Frameworks in XAI

This section describes the approaches used for model explanation in XAI enumerated in Table 4.

4.3.1. What Is SHAP and What Is Its Applicability to IDSs for ICVs?

SHAP is a method for explaining the output of any machine learning model [54,72]. The concept of Shapley values from cooperative game theory [72,85] allows a fair distribution of values among a group of individuals. The method assigns each feature an importance value for a particular prediction. The values can be for both global and local interpretation of a model's predictions, and it is considered a state-of-the-art method for feature importance in XAI. It is also model-agnostic, meaning that it can explain the output of any machine learning model, regardless of its architecture. SHAP aids the perception of the contribution of each data feature to the particular prediction made by a machine learning model. Intrusion detection in ICV is a critical task, as it helps ensure the vehicle's security and safety and its passengers. SHAP explains the decisions made by an IDS in ICVs, which can help to improve transparency and trust in the system. Applying SHAP aids the understanding of the feature importance and contribution of each feature to the predictions made by the model for detecting intrusion attempts.

Based on the SHAP values approach, the following criteria are used for providing both global and local explanations for XAI models [54]: SHAP Force Plots, SHAP Importance Plots, SHAP Summary Plots, and SHAP Dependence Plots. SHAP Summary Plots give a global summary of the features of the SHAP values distribution in the dataset. They demonstrate how each feature's relevance is distributed. This makes it easier to see the dataset's features.

Dependence plots, according to [54], are deemed to be the most straightforward global interpretation plots. The link between the value of a feature and the related Shapley values for each instance in the dataset is displayed in the dependence charts. The SHAP dependence plots display the precise structure of the relationship, whereas summary plots only display the relationship between a feature value and its influence on the prediction.

The SHAP Force Plots aid in the visualization of feature attributions of specific occurrences, where each feature value is a force that either confirms or refutes a prediction. The Force Plots were utilized by the authors of [54] to depict the packet flow characteristics that support or refute the hypothesis made by our IDS model. The size or effect that each feature has on the forecast is also displayed through Force Plots.

Additionally, SHAP Importance Plots for Shapley values are employed to determine the level of importance or significance of instances in the dataset. Large absolute Shapley values in this plot are classified as key features since they have a greater average impact on the model output [54].

4.3.2. What Are TRUST and LIME and What Is Their Applicability to IDSs for ICVs?

TRUSTvs. LIME: TRUST is a framework for assessing AI models' transparency, robustness, and fairness [108]. It evaluates the comprehensibility and interpretability of intrusion detection models for ICV, considered black boxes because they are difficult to interpret [108]. Applying TRUST to the trained IDS model facilitates the transparency of the model's decision-making process and identifies how well the model can detect intrusion attempts. TRUST's feature importance and decision path analysis aid in understanding the most important features and how it makes its predictions. The framework includes several robustness assessment techniques for evaluating IDS in ICVs, such as robustness testing and adversarial example generation [93,108]. It enables the interpretability of the model's generalization to new and unseen data and its resistance to adversarial examples [93,108]. TRUST can also evaluate the model's fairness by assessing the demographic parity and equalized odds. It assists in clarifying the model's fair treatment of various features and identifies potential bias in the model [93,108]. It is helpful for intrusion detection in ICVs as it provides more interpretable and transparent explanations of the decision-making process of the IDS, which can help to improve the security and safety of the vehicle and its occupants by ensuring the robustness and fairness of the model.

Similarly, LIME is a technique for explaining the predictions of any machine learning model by approximating the model locally with an interpretable model. It is model-agnostic, following the ability to explain the output of any machine learning model, regardless of its architecture [107]. It is a common approach for interpretability because it provides human-understandable explanations for complex models [107]. LIME enables a more intuitive and interpretable comprehension of the complex decision-making process of IDS models, which is vital for building trust in the model and its predictions [107]. Its utilization enhances the understanding of the significant features relevant to intrusion detection predictions, which improves the security and safety of the ICV and its user.

The benchmark XAI explainer model is known as the Local Interpretable Model-agnostic Explanations (LIME), introduced by [107]. This model, however, was outperformed by the Transparency Relying Upon Statistical Theory TRUST) explainer introduced by [108]. It achieved 25 times faster results and is a promising candidate for real-time and critical applications such as the ITS. In addition, the TRUST model took care of the computational complexity deficiency of SHAP.

4.3.3. What Is LORE and What Is Its Applicability to IDSs for ICVs?

For a given black box case, LOcal Rule-based Explanations (LORE) introduced by Guidotti et al. [109] generates an interpretable prediction. On a dense set of synthetic examples, a decision tree is utilized to train the local interpretable predictor. A local explanation that consists of a single-choice rule and a number of counterfactual rules for the reversed decision can be extracted from the decision tree. This framework is local-based, post hoc, and model-agnostic [128]. The originators of the LORE in [109] based their work on a local interpretable predictor, a decision tree. The decision tree allows for the extraction of a local explanation, which consists of a single choice rule and a collection of counterfactual rules for the reversed decision [128]. The identified challenge that LORE's applicability to XAI for ICV could be the need for human comprehensibility of the explanations provided by LORE [109]. To achieve a flexible framework for LORE interpretability, Rajapaksha et al. [129] proposed the use of k-optimal associations, known as Local Rule-based Model Interpretability with k-optimal associations (LoRMIkA). LoRMIkA provides a flexible way to obtain predictive rules needed for explanations. The argument here is that the most predictive rules are not necessarily the rules that provide the best explanations. Since the ICV situation is expectedly dynamic, LoTMIkA shows a promising adoption owing to its ability to provide multiple rules capable of explaining predictions in various scenarios.

4.3.4. What Is GRAD-CAM and What Is Its Applicability to IDSs for ICVs?

A method called Gradient-weighted Class Activation Mapping (GRAD-CAM) [110] creates a heat map for each class from a single image. Grad-CAM then generates a localization map that discriminates classes. The framework uses the feature maps produced by the final convolutional layer of a CNN. This is model-specific, local-based, and post hoc [128]. GRAD-CAM finds usage in the visualization of deep learning (such as CNN) outputs. GRAD-CAM achieves this by using the features obtained from the last convolutional layer of a CNN, for example. The weighted combination of feature maps followed by the ReLU activation function is used to obtain a precise heat map. Although GRAD-CAM is one of the most popular methods for explaining deep neural network decisions, it violates key axiomatic properties, such as sensitivity and completeness. Integrated gradients are an axiomatic attribution method that aims to cover this gap [130]. In 2020, GRAD-CAM was reported to have a challenge in its ability to reflect the model's computation because of the gradient average steps. These steps are thus unreliable in prediction. To solve this, a modified CAM was proposed, known as High-Resolution Class Activation Mapping (HiResCAM), for faithful explanations of CNN [131].

4.3.5. What Is CEM and What Is Its Applicability to IDSs for ICVs?

The contrastive explanation method (CEM) is a method for generating explanations for AI models to understand the decision-making process of a model by comparing the decision made for a specific input to the decision made for a similar but different input [111]. Its classification model explanations are provided in [111]. Furthermore, it retrieves the characteristics that must be sufficiently present to predict the input instance will belong to the same class. The minimal attributes that must be altered in order to associate the input instance with a different class are also identified. This is local, post hoc, and model-neutral [128]. According to [128], CEM, when used for neural networks, highlights not only the pertinent positives but also the pertinent negatives in its explanation, making it a preferred choice over LRP. The ability of CEM to extract both positive and negative pertinent features can help in reducing error during the diagnosis of a target. It is, however, yet to be seen how CEM could be applied effectively to the IDS for ICVs where non-image data is dominant. This is important if the CEM is to be used for intrusion detection in ICVs in real-time by monitoring and analyzing network traffic, system logs, and other data sources to identify unusual or suspicious activity. CEM is vital to interpreting the AI model predictions.

4.4. Discussion

4.4.1. Computational Complexity Challenge of XAI Implementation

The computationally demanding and costly operation of SHAP is one of its disadvantages. Additionally, it has been discovered to be open to hostile attacks. In a subsequent study, the authors in [54] seek to assess the degree to which SHAP is susceptible to malicious attacks and look into potential new or improved defenses that might be utilized to enhance its robustness in IoT systems.

4.4.2. Concept Misrepresentation

Although the phrases "interpretability" and "explainability" are sometimes used synonymously in the literature, they are not the same. A model's active quality of explainability defines the steps it takes to reveal to people its frequently intricate internal workings [132]. The level at which a model's mechanics make sense to people, in contrast, is referred to as a model's interpretability [70,132]. Thus, there is a need for universally acceptable terminology, and elimination of ambiguity in definitions [93].

4.4.3. Need for ICV-Based Dataset

Although, XAI adoption for ICV is still at the infant stage, the AI-based approaches witnessed the use of inappropriate datasets for model evaluation. Since the ICV or IoV has unique features, datasets reflecting the reality of the connected vehicles should be advocated if the goal of XAI will be achieved.

4.4.4. Reliability

Reliability analysis determines a system's level of assurance in terms of probability density [133]. They utilized the XGBoost to calibrate the resistance reduction factors in reaching the stipulated objective system. The suggested ML-based models are proven to efficiently determine the shear resistance of fabric-reinforced cementitious matrix-strengthened reinforced concrete beams, yielding the most consistent, precise, and reliable predictions while meeting the specified target reliability of the system. In another study outlining some intriguing potential prospects for utilizing machine learning to increase safety and reliability concerns, the authors of [134] examined the use of machine learning in quality engineering and security solutions. Since the LIME is based on Gaussian assumptions, the reliability of the output becomes doubtful once the assumptions do not fit into any target system. The challenge of overfitting also depletes the reliability of XAI models as some researchers do not scrutinize the potential flaws of "explainers" [135].

One reliability case of concern to researchers is the effect of adversarial perturbations (APs) on deep learning. APs are the procedures intended to mislead a target model by injecting imperceptible noise (attacks). Galli et al. [136] proposed the use of the disc similarity coefficient (DSC) and correlation coefficient (CC) as measures for the reliability of XAI methods. The reliability of deep learning models was found to be affected by the huge number of pixels in the target image, and AP. However, the number of classes used in datasets had no significant effect on the reliability of XAI. However, most of the existing reliability experiments were conducted using datasets too little in comparison to real ICV expectations.

4.4.5. Future Direction Based on Research Gaps

The authors in [70] opined that beyond XAI, there is a call for responsible AI (RAI). RAI is an additional constraint on the XAI to guarantee fairness, ethical deployment, transparency, security, safe usage, accountability, and privacy awareness. This review shows that although XAI and RAI are gaining adoption for various target domains, they are yet to be extensively deployed for IDSs in ITSs/ICVs. It will be a promising research direction to explore LIME, SHAP, and TRUST for IDSs in ITSs while scrutinizing all "explainer" options [135]. In addition, the computational complexity introduced by the use of XAI needs to be addressed since the ICVs are real-time networks where the need for latency can not be ignored. Furthermore, for XAI to be effectively deployed in ICVs, lightweight models and user-friendly practical implementation is desired. From the case studies, it is evident that some leading AI companies are already developing XAI interfaces in this direction. Additionally, issues of real-scenario experiments/testing, the complexity of security XAI-based solutions, and the need for embedded/portable solutions have been listed as additional future directions in [137].

5. Conclusions

This paper presented a comprehensive review of explainable artificial intelligence (XAI) for intrusion detection and mitigation inintelligent connected vehicles (ICVs). ICVs/IoV as an extended application of the Internet of things (IoT) in intelligent transportation systems (ITSs) requires effective security from attackers due to vulnerabilities inherent in connected devices. However, due to the "black-box'" nature of artificial intelligence applied to most detection systems, transparency or interpretability becomes a problem; thus the need for explainable AI. In this survey, a comprehensive background of the existing XAI frameworks, their capabilities and use cases, and their applicability to ICV security is discussed. In particular, the need for reliability, low computational complexity, and incorporation of user-friendly XAI modules is identified as a promising research issue. Finally, XAI developers must address the issue of bias introduced by rule-based XAI to foster its acceptability in essential industries such as the automobile industry.

Author Contributions: Conceptualization, C.I.N., L.A.C.A., J.N.N. and C.U.; methodology, C.I.N., L.A.C.A., J.N.N., C.U., C.C.N.N. and D.-S.K.; software, C.I.N., L.A.C.A., J.N.N., J.C.O. and S.A.O.; validation, C.I.N., L.A.C.A., J.N.N., C.U. and D.-S.K.; formal analysis, C.I.N.; investigation, C.I.N., L.A.C.A., J.N.N., C.U., C.C.N.N. and D.-S.K.; resources, C.I.N. and D.-S.K.; data curation, C.I.N.; writing—original draft preparation, C.I.N., L.A.C.A., J.N.N., C.U. and D.-S.K.; writing—review and editing, C.I.N., L.A.C.A., J.N.N., J.C.O., S.A.O. and C.U.; visualization, L.A.C.A. and J.N.N.; supervision, D.-S.K.; project administration, C.I.N. and D.-S.K.; funding acquisition, C.I.N. and D.-S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research work was supported by Priority Research Centers Program through the national research foundation of Korea (NRF) funded by the ministry of education, science and technology (MEST) (2018R1A6A1A03024003) and the ministry of science and information & communication technology (MSIT), Korea, under grand information technology research center support program (IITP-2023-2020-0-01612) supervised by the Institute for Information & communications Technology Planning & Evaluation (IITP), Korea.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This research work was supported by Priority Research Centers Program through the national research foundation of Korea (NRF) funded by the ministry of education, science and technology (MEST) (2018R1A6A1A03024003) and the ministry of science and information & communication technology (MSIT), Korea, under grand information technology research center support program (IITP-2023-2020-0-01612) supervised by the Institute for Information & communications Technology Planning & Evaluation (IITP), Korea.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANFIS	Adaptive Neuro Fuzzy Inference System
ANN	Artificial Neural Networks
AP	Adversarial Perturbation
CAM	Class Activation Mapping
CAN	Controller Area Network
CC	Correlation Coefficient
CEM	Contrastive Explanation Method
C-ITS	Cooperative ITS
CNN	Convolutional Neural Network
C-V2X	Cooperative V2X
DDoS	Distributed Denial of Service
DL	Deep Learning
DoS	Denial of Service
DSC	Disc Similarity Coefficient
DSRC	Dedicated Short Range communications
ETSI	European Telecommunications Standards Institute
GPS	Global Positioning Systems
GRAD-CAM	Gradient-weighted Class Activation Mapping
HCI	Human–Computer Interface
HiResCAM	High-Resolution Class Activation Mapping
ICS	Industrial Control System
ICV	Intelligent Connected Vehicle
IDS	Intrusion Detection Systems
IoT	Internet of Things
IoV	Internet of Vehicles
ITS	Intelligent Transportation System
KNN	K-Nearest Neighbor
LIME	Local Interpretable Model-agnostic Explanations
LOIC	Low Orbit Ion Cannon
LORE	Local Rule-based Explanations
LoRMIkA	Local Rule-based Model Interpretability with k-optimal Associations
LRP	Layer-wise Relevance Propagation
LTE-A	Long-Term Evolution-Advanced
MiM	Man-in-the-Middle
ML	Machine Learning
MTU	Mobile Terminal Unit
NIST	National Institute of Standards
NS-3	Network Simulator-3

OBUs	Onboard Units
OSI	Open Systems Interconnection
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ReLU	Rectifier Linear Unit
RTC	Road Traffic Crashes
RTU	Remote Terminal Unit
SHAP	SHapley Additive exPlanations
SUMO	Simulation of Urban Mobility
TARA	Threat Analysis and Risk Assessment
TRUST	Transparency Relying Upon Statistical Theory
UAV	Unmanned Aerial Vehicles
V2V	Vehicle-to-Vehicle
V2I	Vehicles-to-Infrastructures
V2X	Vehicles-to-Things
VANET	Vehicular Ad Hoc Networks
VeReMi	Vehicular Reference Misbehavior
WAVE	Wireless Access in Vehicular Environment
WHO	World Health Organization
XAI	Explainable AI
XAIR	Explainable AI for Regression
5G	Fifth Generation

References

1. Wang, P.; Ye, R.; Zhang, J.; Wang, T. An Eco-Driving Controller Based on Intelligent Connected Vehicles for Sustainable Transportation. *Appl. Sci.* **2022**, *12*, 4533. [CrossRef]
2. Zhi, P.; Zhao, R.; Zhou, H.; Zhou, Y.; Ling, N.; Zhou, Q. Analysis on the development status of intelligent and connected vehicle test site. *Intell. Converg. Netw.* **2021**, *2*, 320–333. [CrossRef]
3. Yu, M. Construction of Regional Intelligent Transportation System in Smart City Road Network via 5G Network. *IEEE Trans. Intell. Transp. Syst.* **2022**, 1–9. [CrossRef]
4. Montoya-Torres, J.R.; Moreno, S.; Guerrero, W.J.; Mejía, G. Big Data Analytics and Intelligent Transportation Systems. *IFAC-PapersOnLine* **2021**, *54*, 216–220. [CrossRef]
5. Garg, T.; Kaur, G. A Systematic Review on Intelligent Transport Systems. *J. Comput. Cogn. Eng.* **2022**. [CrossRef]
6. Anyanwu, G.O.; Nwakanma, C.I.; Lee, J.M.; Kim, D.S. Falsification Detection System for IoV Using Randomized Search Optimization Ensemble Algorithm. *IEEE Trans. Intell. Transp. Syst.* **2023**, 1–15. [CrossRef]
7. Mütze, F. *Cooperative Intelligent Transport Systems (C-ITS)*; European Transport Safety Council: Brussels, Belgium, November 2017; pp. 1–12. Available online: https://etsc.eu/briefing-cooperative-intelligent-transport-systems-c-its/ (accessed on 8 January 2023).
8. Chekkouri, A.S.; Ezzouhairi, A.; Pierre, S. Connected vehicles in an intelligent transport system. *VEhicular Commun. Netw.* **2015**, 193–221. [CrossRef]
9. Kaffash, S.; Nguyen, A.T.; Zhu, J. Big Data Algorithms and Applications in Intelligent Transportation System: A Review and Bibliometric Analysis. *Int. J. Prod. Econ.* **2021**, *231*, 107868. [CrossRef]
10. Chen, S.; Hu, J.; Shi, Y.; Zhao, L.; Li, W. A Vision of C-V2X: Technologies, Field Testing, and Challenges with Chinese Development. *IEEE Internet Things J.* **2020**, *7*, 3872–3881. [CrossRef]
11. Nguyen, V.L.; Hwang, R.H.; Lin, P.C.; Vyas, A.; Nguyen, V.T. Towards the Age of Intelligent Vehicular Networks for Connected and Autonomous Vehicles in 6G. *IEEE Netw.* **2022**, 1–8. [CrossRef]
12. Consortium, C.C. C-ITS: Cooperative Intelligent Transport Systems and Services. In Proceedings of the 16th Car 2 CAr Forum. Car 2 CAr Communication Consortium, Moses Lake, WA, USA, 28–29 September 2022. Available online: https://www.car-2-car.org/about-c-its/ (accessed on 11 January 2023).
13. Tsolaki, K.; Vafeiadis, T.; Nizamis, A.; Ioannidis, D.; Tzovaras, D. Utilizing Machine Learning on Freight Transportation and Logistics Applications: A review. *ICT Express* **2022**, *in press*. [CrossRef]
14. ETSI. Automotive Intelligent Transport Systems (ITS). 2021. Available online: https://www.etsi.org/technologies/automotive-intelligent-transport (accessed on 6 January 2023).
15. Wakabayashi, D. Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam. *The New York Times*, 19 March 2018.
16. Van Brummelen, J.; O'Brien, M.; Gruyer, D.; Najjaran, H. Autonomous Vehicle Perception: The Technology of Today and Tomorrow. *Transp. Res. Part C Emerg. Technol.* **2018**, *89*, 384–406. [CrossRef]
17. Jaigirdar, F.T.; Rudolph, C.; Oliver, G.; Watts, D.; Bain, C. What Information is Required for Explainable AI?: A Provenance-based Research Agenda and Future Challenges. In Proceedings of the 2020 IEEE 6th International Conference on Collaboration and Internet Computing (CIC), Los Angeles, CA, USA, 12–14 December 2020; pp. 177–183. [CrossRef]
18. Boukerche, A.; Tao, Y.; Sun, P. Artificial Intelligence-Based Vehicular Traffic Flow Prediction Methods for Supporting Intelligent Transportation Systems. *Comput. Netw.* **2020**, *182*, 107484. [CrossRef]

19. Garg, P.K. Overview of Artificial Intelligence. In *Artificial Intelligence: Technologies, Applications, and Challenges*, 1st ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2022; Chapter 1, pp. 1–16. [CrossRef]
20. Nikitas, A.; Michalakopoulou, K.; Njoya, E.T.; Karampatzakis, D. Artificial Intelligence, Transport, and the Smart City: Definitions and Dimensions of a New Mobility Era. *Sustainability* **2020**, *12*, 2789. [CrossRef]
21. Abduljabbar, R.; Dia, H.; Liyanage, S.; Bagloee, S.A. Applications of Artificial Intelligence in Transport: An Overview. *Sustainability* **2019**, *11*, 189. [CrossRef]
22. Iyer, L.S. AI Enabled Applications Towards Intelligent Transportation. *Transp. Eng.* **2021**, *5*, 100083. [CrossRef]
23. Anyanwu, G.O.; Nwakanma, C.I.; Lee, J.M.; Kim, D.S. Optimization of RBF-SVM Kernel using Grid Search Algorithm for DDoS Attack Detection in SDN-based VANET. *IEEE Internet Things J.* **2022**. [CrossRef]
24. Le, T.-T.-H.; Kim, H.; Kang, H.; Kim, H. Classification and Explanation for Intrusion Detection System Based on Ensemble Trees and SHAP Method. *Sensors* **2022**, *22*, 1154. [CrossRef]
25. Speith, T. A Review of Taxonomies of Explainable Artificial Intelligence (XAI) Methods. In Proceedings of the FAccT'22: 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, 21–24 June 2022; pp. 2239–2250. [CrossRef]
26. Saeed, W.; Omlin, C. Explainable AI (XAI): A Systematic Meta-Survey of Current Challenges and Future Opportunities. *arXiv* **2021**, arXiv:2111.06420.
27. Tran-Dang, H.; Bhardwaj, S.; Rahim, T.; Musaddiq, A.; Kim, D.S. Reinforcement Learning Based Resource Management for Fog Computing Environment: Literature Review, Challenges, and Open Issues. *J. Commun. Netw.* **2022**. [CrossRef]
28. Ghosh, S.; Sampalli, S. A Survey of Security in SCADA Networks: Current Issues and Future Challenges. *IEEE Access* **2019**, *7*, 135812–135831. [CrossRef]
29. Lin, Z.; Shi, Y.; Xue, Z. IDSGAN: Generative Adversarial Networks for Attack Generation Against Intrusion Detection. In *Advances in Knowledge Discovery and Data Mining*; Gama, J., Li, T., Yu, Y., Chen, E., Zheng, Y., Teng, F., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 79–91. [CrossRef]
30. Taparia, A. IoT 2022 in reveiew: The 10 most relevant IoT developments of the year. In Proceedings of the IoT Analytics: Market Insights for the Internet of Things, Hamburg, Germany, 9 January 2023. Available online: https://iot-analytics.com/iot-2022-in-review/ (accessed on 9 January 2023).
31. Guan, T.; Han, Y.; Kang, N.; Tang, N.; Chen, X.; Wang, S. An Overview of Vehicular Cybersecurity for Intelligent Connected Vehicles. *Sustainability* **2022**, *14*, 5211. [CrossRef]
32. Anyanwu, G.O.; Nwakanma, C.I.; Lee, J.M.; Kim, D.S. Novel Hyper-Tuned Ensemble Random Forest Algorithm for the Detection of False Basic Safety Messages in Internet of Vehicles. *ICT Express* **2022**. [CrossRef]
33. Xiong, S.; Qi, X. The Research on Security Model Algorithm in Intelligent Connected Vehicles. In Proceedings of the 2022 IEEE International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), Hongqing, China, 5–7 August 2022; pp. 372–376. [CrossRef]
34. Duan, X.; Yan, H.; Tian, D.; Zhou, J.; Su, J.; Hao, W. In-Vehicle CAN Bus Tampering Attacks Detection for Connected and Autonomous Vehicles Using an Improved Isolation Forest Method. *IEEE Trans. Intell. Transp. Syst.* **2021**, 1–13. [CrossRef]
35. Goddard, M. The EU General Data Protection Regulation (GDPR): European Regulation that has a Global Impact. *Int. J. Mark. Res.* **2017**, *59*, 703–705. [CrossRef]
36. *Global Status Report on Road Safety 2018*; World Health Organization: Luxembourg, 2018.
37. *Global Status Report on Road Traffic*; Time for Action; World Health Organization: Luxembourg, 2020.
38. *Global Status Report on Road Safety 2015*; World Health Organization: Luxembourg, 2015.
39. Uzondu, C.; Jamson, S.; Marsden, G. Road Safety in Nigeria: Unravelling the Challenges, Measures, and Strategies for Improvement. *Int. J. Inj. Control Saf. Promot.* **2022**, 1–11. [CrossRef]
40. Bie, J.; Roelofsen, M.; Jin, L.; van Arem, B. Lane Change and Overtaking Collisions: Causes and Avoidance Techniques. In *Wireless Vehicular Networks for Car Collision Avoidance*; Springer: New York, NY, USA, 2013; pp. 143–187. [CrossRef]
41. Singh, G.; Bansal, D.; Sofat, S. Intelligent Transportation System for Developing Countries:A Survey. *Int. J. Comput. Appl.* **2014**, *85*, 34–38. [CrossRef]
42. Kang, J.; Yu, R.; Huang, X.; Zhang, Y. Privacy-Preserved Pseudonym Scheme for Fog Computing Supported Internet of Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 2627–2637. [CrossRef]
43. Yang, D.; Jiang, K.; Zhao, D.; Yu, C.; CAO, Z.; XIE, S.; XIAO, Z.; JIAO, X.; Wang, S.; Zhang, K. Intelligent and connected vehicles: Current status and future perspectives. *Sci. China Technol. Sci.* **2018**, *61*, 1446–1471. [CrossRef]
44. Li, Y.; Cao, Y.; Qiu, H.; Gao, L.; Du, Z.; Chen, S. Big wave of the intelligent connected vehicles. *China Commun.* **2016**, *13*, 27–41. [CrossRef]
45. Kuang, X.; Zhao, F.; Hao, H.; Liu, Z. Intelligent connected vehicles: The industrial practices and impacts on automotive value-chains in China. *Asia Pac. Bus. Rev.* **2018**, *24*, 1–21. [CrossRef]
46. Huang, J.M. Research on Internet of Vehicles and its Application in Intelligent Transportation. In *Mechatronics and Industrial Informatics*; Applied Mechanics and Materials; Trans Tech Publications Ltd.: Zurich, Switzerland, 2013; Volume 321, pp. 2818–2821. [CrossRef]
47. Contreras-Castillo, J.; Zeadally, S.; Guerrero-Ibañez, J.A. Internet of Vehicles: Architecture, Protocols, and Security. *IEEE Internet Things J.* **2018**, *5*, 3701–3709. [CrossRef]

48. Xu, C.; Wu, H.; Liu, H.; Gu, W.; Li, Y.; Cao, D. Blockchain-Oriented Privacy Protection of Sensitive Data in the Internet of Vehicles. *IEEE Trans. Intell. Veh.* **2022**. [CrossRef]
49. Hamid, U.Z.A.; Zamzuri, H.; Limbu, D.K. Internet of Vehicle (IoV) Applications in Expediting the Implementation of Smart Highway of Autonomous Vehicle: A Survey. In *Performability in Internet of Things*; Springer International Publishing: Cham, Switzerland, 2019; pp. 137–157. [CrossRef]
50. Ali, E.S.; Hassan, M.B.; Saeed, R.A. Machine Learning Technologies in Internet of Vehicles. In *Intelligent Technologies for Internet of Vehicles*; Springer International Publishing: Cham, Swtizerland, 2021; pp. 225–252. [CrossRef]
51. Hindy, H.; Brosset, D.; Bayne, E.; Seeam, A.K.; Tachtatzis, C.; Atkinson, R.; Bellekens, X. A Taxonomy of Network Threats and the Effect of Current Datasets on Intrusion Detection Systems. *IEEE Access* **2020**, *8*, 104650–104675. [CrossRef]
52. Zhu, B.; Joseph, A.; Sastry, S. A Taxonomy of Cyber Attacks on SCADA Systems. In Proceedings of the 2011 International Conference on Internet of Things and 4th International Conference on Cyber, Physical and Social Computing, Dalian, China, 19–22 October 2011; pp. 380–388. [CrossRef]
53. Hilal, H.; Nangim, A. Network Security Analysis SCADA System Automation on Industrial Process. In Proceedings of the 2017 International Conference on Broadband Communication, Wireless Sensors and Powering (BCWSP), Jakarta, Indonesia, 21–23 November 2017; pp. 1–6. [CrossRef]
54. Oseni, A.; Moustafa, N.; Creech, G.; Sohrabi, N.; Strelzoff, A.; Tari, Z.; Linkov, I. An Explainable Deep Learning Framework for Resilient Intrusion Detection in IoT-Enabled Transportation Networks. *IEEE Trans. Intell. Transp. Syst.* **2022**, 1–15. [CrossRef]
55. Melnick, J. Top 10 Most Common Types of Cyber Attacks. Netwrix Blog. 15 May 2018. Available online: https://blog.netwrix.com/2018/05/15/top-10-most-common-types-of-cyber-attacks/ (accessed on 7 January 2023).
56. Zhang, Y.; Xiang, Y.; Wang, L. Reliability Analysis of Power Grids with Cyber Vulnerability in SCADA System. In Proceedings of the 2014 IEEE PES General Meeting | Conference & Exposition, National Harbor, MD, USA, 27–31 July 2014; pp. 1–5. [CrossRef]
57. Anyanwu, G.O.; Nwakanma, C.I.; Lee, J.M.; Kim, D.S. RBF-SVM Kernel-based Model for Detecting DDoS Attacks in SDN Integrated Vehicular Network. *Ad Hoc Netw.* **2023**, *140*, 103026. [CrossRef]
58. Kalluri, R.; Mahendra, L.; Kumar, R.S.; Prasad, G.G. Simulation and Impact Analysis of Denial-of-Service Attacks on Power SCADA. In Proceedings of the 2016 National Power Systems Conference (NPSC), Bhubaneswar, India, 19–21 December 2016; pp. 1–5. [CrossRef]
59. Amaizu, G.; Nwakanma, C.; Bhardwaj, S.; Lee, J.; Kim, D. Composite and Efficient DDoS Attack Detection Framework for B5G Networks. *Comput. Netw.* **2021**, *188*, 107871. [CrossRef]
60. Rakas, S.V.B.; Stojanović, M.D.; Marković-Petrović, J.D. A Review of Research Work on Network-Based SCADA Intrusion Detection Systems. *IEEE Access* **2020**, *8*, 93083–93108. [CrossRef]
61. Tian, H. *Introduction of IoV Security*; China Academy of Information and Communication Technology (CAICT): Beijing, China, 2017; pp. 1–59.
62. Mandal, S. Protecting Software-defined Vehicles with Cybersecurity Solutions. In Proceedings of the Connected Vehicle 2022 Summit: From ADAS to Autonomous Mobility, Bengaluru, India, 4–6 May 2022.
63. Vitale, C.; Piperigkos, N.; Laoudias, C.; Ellinas, G.; Casademont, J.; Escrig, J.; Kloukiniotis, A.; Lalos, A.S.; Moustakas, K.; Diaz Rodriguez, R.; et al. CARAMEL: Results on a secure architecture for connected and autonomous vehicles detecting GPS spoofing attacks. *J. Wirel. Commun. Netw.* **2021**, *2021*, 115. [CrossRef]
64. Sarhan, M.; Layeghy, S.; Portmann, M. An Explainable Machine Learning-Based Network Intrusion Detection System for Enabling Generalisability in Securing IoT Networks. *arXiv* **2021**. [CrossRef]
65. Zebin, T.; Rezvy, S.; Luo, Y. An Explainable AI-Based Intrusion Detection System for DNS Over HTTPS (DoH) Attacks. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 2339–2349. [CrossRef]
66. Marino, D.L.; Wickramasinghe, C.S.; Manic, M. An Adversarial Approach for Explainable AI in Intrusion Detection Systems. In Proceedings of the IECON 2018—44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, USA, 21–23 October 2018; pp. 3237–3243. [CrossRef]
67. Das, S.; Agarwal, N.; Shiva, S. DDoS Explainer using Interpretable Machine Learning. In Proceedings of the 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 27–30 October 2021; pp. 0001–0007. [CrossRef]
68. Srivastava, G.; Jhaveri, R.H.; Bhattacharya, S.; Pandya, S.; Maddikunta, P.K.R.; Yenduri, G.; Hall, J.G.; Alazab, M.; Gadekallu, T.R. XAI for Cybersecurity: State of the Art, Challenges, Open Issues and Future Directions. *arXiv* **2022**, arXiv:2206.03585.
69. McNamara, M. Explainable AI: What Is It? How Does It Work? And What Role Does Data Play? NetApp. Available online: https://www.netapp.com/blog/explainable-ai/ (accessed on 22 February 2022).
70. Arrieta, A.B.; Díaz-Rodríguez, N.; Ser, J.D.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
71. Slijepcevic, D.; Horst, F.; Lapuschkin, S.; Raberger, A.M.; Zeppelzauer, M.; Samek, W.; Breiteneder, C.; Schöllhorn, W.I.; Horsak, B. On the Explanation of Machine Learning Predictions in Clinical Gait Analysis. *arXiv* **2019**, arXiv:1912.07737.
72. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, arXiv:1705.07874.

73. Ge, W.; Patino, J.; Todisco, M.; Evans, N. Explaining Deep Learning Models for Spoofing and Deepfake Detection with SHapley Additive ExPlanations. In Proceedings of the ICASSP 2022—2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 23–27 May 2022; pp. 6387–6391. [CrossRef]
74. Gulmezoglu, B. XAI-based Microarchitectural Side-Channel Analysis for Website Fingerprinting Attacks and Defenses. *IEEE Trans. Dependable Secur. Comput.* **2021**. [CrossRef]
75. Nascita, A.; Montieri, A.; Aceto, G.; Ciuonzo, D.; Persico, V.; Pescapé, A. XAI Meets Mobile Traffic Classification: Understanding and Improving Multimodal Deep Learning Architectures. *IEEE Trans. Netw. Serv. Manag.* **2021**, *18*, 4225–4246. [CrossRef]
76. Zou, L.; Goh, H.L.; Liew, C.J.Y.; Quah, J.L.; Gu, G.T.; Chew, J.J.; Prem Kumar, M.; Ang, C.G.L.; Ta, A. Ensemble Image Explainable AI (XAI) Algorithm for Severe Community-Acquired Pneumonia and COVID-19 Respiratory Infections. *IEEE Trans. Artif. Intell.* **2022**. [CrossRef]
77. Saraswat, D.; Bhattacharya, P.; Verma, A.; Prasad, V.K.; Tanwar, S.; Sharma, G.; Bokoro, P.N.; Sharma, R. Explainable AI for Healthcare 5.0: Opportunities and Challenges. *IEEE Access* **2022**, *10*, 84486–84517. [CrossRef]
78. Narteni, S.; Orani, V.; Cambiaso, E.; Rucco, M.; Mongelli, M. On the Intersection of Explainable and Reliable AI for Physical Fatigue Prediction. *IEEE Access* **2022**, *10*, 76243–76260. [CrossRef]
79. Letzgus, S.; Wagner, P.; Lederer, J.; Samek, W.; Müller, K.R.; Montavon, G. Toward Explainable Artificial Intelligence for Regression Models: A Methodological Perspective. *IEEE Signal Process. Mag.* **2022**, *39*, 40–58. [CrossRef]
80. Hwang, C.; Lee, T. E-SFD: Explainable Sensor Fault Detection in the ICS Anomaly Detection System. *IEEE Access* **2021**, *9*, 140470–140486. [CrossRef]
81. Zhang, K.; Zhang, J.; Xu, P.D.; Gao, T.; Gao, D.W. Explainable AI in Deep Reinforcement Learning Models for Power System Emergency Control. *IEEE Trans. Comput. Soc. Syst.* **2022**, *9*, 419–427. [CrossRef]
82. Srinivasan, S.; Arjunan, P.; Jin, B.; Sangiovanni-Vincentelli, A.L.; Sultan, Z.; Poolla, K. Explainable AI for Chiller Fault-Detection Systems: Gaining Human Trust. *Computer* **2021**, *54*, 60–68. [CrossRef]
83. Collini, E.; Palesi, L.A.I.; Nesi, P.; Pantaleo, G.; Nocentini, N.; Rosi, A. Predicting and Understanding Landslide Events with Explainable AI. *IEEE Access* **2022**, *10*, 31175–31189. [CrossRef]
84. Ryu, J.; Park, C.; Park, J.; Cho, N.; Park, J.; Cho, G. Development of Neural Network Model with Explainable AI for Measuring Uranium Enrichment. *IEEE Trans. Nucl. Sci.* **2021**, *68*, 2670–2681. [CrossRef]
85. Tao, J.; Xiong, Y.; Zhao, S.; Wu, R.; Shen, X.; Lyu, T.; Fan, C.; Hu, Z.; Zhao, S.; Pan, G. Explainable AI for Cheating Detection and Churn Prediction in Online Games. *IEEE Trans. Games* **2022**. [CrossRef]
86. Wakjira, T.G.; Ibrahim, M.; Ebead, U.; Alam, M.S. Explainable Machine Learning Model and Reliability Analysis for Flexural Capacity Prediction of RC Beams Strengthened in Flexure with FRCM. *Eng. Struct.* **2022**, *255*, 113903. [CrossRef]
87. Moawad, A.; Islam, E.; Kim, N.; Vijayagopal, R.; Rousseau, A.; Wu, W.B. Explainable AI for a No-Teardown Vehicle Component Cost Estimation: A Top-Down Approach. *IEEE Trans. Artif. Intell.* **2021**, *2*, 185–199. [CrossRef]
88. Adadi, A.; Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160. [CrossRef]
89. Gossen, F.; Margaria, T.; Steffen, B. Formal Methods Boost Experimental Performance for Explainable AI. *IT Prof.* **2021**, *23*, 8–12. [CrossRef]
90. Gossen, F.; Steffen, B. Algebraic Aggregation Random Forests: Towards Explainability and Rapid Evaluation. *Int. J. Softw. Tools Technol. Transf.* **2020**, *22*, 8–12. [CrossRef]
91. Ahmed, I.; Jeon, G.; Piccialli, F. From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, How, and Where. *IEEE Trans. Ind. Inform.* **2022**, *18*, 5031–5042. [CrossRef]
92. Li, X.H.; Cao, C.C.; Shi, Y.; Bai, W.; Gao, H.; Qiu, L.; Wang, C.; Gao, Y.; Zhang, S.; Xue, X.; et al. A Survey of Data-Driven and Knowledge-Aware EXplainable AI. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 29–49. [CrossRef]
93. Rawal, A.; Mccoy, J.; Rawat, D.B.; Sadler, B.; Amant, R. Recent Advances in Trustworthy Explainable Artificial Intelligence: Status, Challenges and Perspectives. *IEEE Trans. Artif. Intell.* **2021**. [CrossRef]
94. Luo, F.; Jiang, Y.; Zhang, Z.; Ren, Y.; Hou, S. Threat Analysis and Risk Assessment for Connected Vehicles: A Survey. *Secur. Commun. Netw.* **2021**, *2021*, 1263820. [CrossRef]
95. Dibaei, M.; Zheng, X.; Jiang, K.; Abbas, R.; Liu, S.; Zhang, Y.; Xiang, Y.; Yu, S. Attacks and defences on intelligent connected vehicles: A survey. *Digit. Commun. Netw.* **2020**, *6*, 399–421. [CrossRef]
96. Banafshehvaragh, S.T.; Rahmani, A.M. Intrusion, anomaly, and attack detection in smart vehicles. *Microprocess. Microsyst.* **2023**, *96*, 104726. [CrossRef]
97. Wang, B.; Han, Y.; Wang, S.; Tian, D.; Cai, M.; Liu, M.; Wang, L. A Review of Intelligent Connected Vehicle Cooperative Driving Development. *Mathematics* **2022**, *10*, 3635. [CrossRef]
98. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred Reporting Items for Systematic Reviews and Meta-analyses: The PRISMA Statement. *Ann. Intern. Med.* **2009**, *151*, 264–269. [CrossRef]
99. Njoku, J.N.; Nwakanma, C.I.; Amaizu, G.C.; Kim, D.S. Prospects and Challenges of Metaverse Application in Data-Driven Intelligent Transportation Systems. *IET Intell. Transp. Syst.* **2023**, *17*, 1–21. [CrossRef]
100. Torres-Carrión, P.V.; González-González, C.S.; Aciar, S.; Rodríguez-Morales, G. Methodology for Systematic Literature Review Applied to Engineering and Education. In Proceedings of the 2018 IEEE Global Engineering Education Conference (EDUCON), Santa Cruz de Tenerife, Spain, 17–20 April 2018; pp. 1364–1373. [CrossRef]

101. Misra, S. A Step by Step Guide for Choosing Project Topics and Writing Research Papers in ICT Related Disciplines. In *Proceedings of the International Conference on Information and Communication Technology and Applications*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 727–744. [CrossRef]
102. Lundberg, H.; Mowla, N.I.; Abedin, S.F.; Thar, K.; Mahmood, A.; Gidlund, M.; Raza, S. Experimental Analysis of Trustworthy In-Vehicle Intrusion Detection System using eXplainable Artificial Intelligence (XAI). *IEEE Access* **2022**. [CrossRef]
103. Keneni, B.M.; Kaur, D.; Al Bataineh, A.; Devabhaktuni, V.K.; Javaid, A.Y.; Zaientz, J.D.; Marinier, R.P. Evolving Rule-Based Explainable Artificial Intelligence for Unmanned Aerial Vehicles. *IEEE Access* **2019**, *7*, 17001–17016. [CrossRef]
104. El Houda, Z.A.; Brik, B.; Senouci, S.M. A Novel IoT-Based Explainable Deep Learning Framework for Intrusion Detection Systems. *IEEE Internet Things Mag.* **2022**, *5*, 20–23. [CrossRef]
105. Upadhyay, D.; Manero, J.; Zaman, M.; Sampalli, S. Intrusion Detection in SCADA Based Power Grids: Recursive Feature Elimination Model with Majority Vote Ensemble Algorithm. *IEEE Trans. Netw. Sci. Eng.* **2021**, *8*, 2559–2574. [CrossRef]
106. Zhang, F.; Kodituwakku, H.A.D.E.; Hines, J.W.; Coble, J. Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network, System, and Process Data. *IEEE Trans. Ind. Inform.* **2019**, *15*, 4362–4369. [CrossRef]
107. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *arXiv* **2016**, arXiv:1602.04938.
108. Zolanvari, M.; Yang, Z.; Khan, K.; Jain, R.; Meskin, N. TRUST XAI: Model-Agnostic Explanations for AI with a Case Study on IIoT Security. *IEEE Internet Things J.* **2021**. [CrossRef]
109. Guidotti, R.; Monreale, A.; Ruggieri, S.; Pedreschi, D.; Turini, F.; Giannotti, F. Local Rule-Based Explanations of Black Box Decision Systems. *arXiv* **2018**, arXiv:1805.10820.
110. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2019**, *128*, 336–359. [CrossRef]
111. Dhurandhar, A.; Chen, P.Y.; Luss, R.; Tu, C.C.; Ting, P.; Shanmugam, K. Explanations based on the missing: Towards contrastive explanations with pertinent negatives. In Proceedings of the NIPS'18: 32nd International Conference on Neural Information Processing Systems, Montréal, QC, Canada, 3–8 December 2018. [CrossRef]
112. Ahmad, Z.; Shahid Khan, A.; Wai Shiang, C.; Abdullah, J.; Ahmad, F. Network Intrusion Detection System: A Systematic Study of Machine Learning and Deep Learning Approaches. *Trans. Emerg. Telecommun. Technol.* **2021**, *32*, e4150. [CrossRef]
113. Zolanvari, M.; Teixeira, M.A.; Gupta, L.; Khan, K.M.; Jain, R. Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things. *IEEE Internet Things J.* **2019**, *6*, 6822–6834. [CrossRef]
114. Khraisat, A.; Gondal, I.; Vamplew, P.; Kamruzzaman, J. Survey of Intrusion Detection Systems: Techniques, Datasets and Challenges. *Cybersecurity* **2019**, *2*, 1–22. [CrossRef]
115. Alladi, T.; Kohli, V.; Chamola, V.; Yu, F.R.; Guizani, M. Artificial Intelligence (AI)-Empowered Intrusion Detection Architecture for the Internet of Vehicles. *IEEE Wirel. Commun.* **2021**, *28*, 144–149. [CrossRef]
116. Ercan, S.; Ayaida, M.; Messai, N. Misbehavior Detection for Position Falsification Attacks in VANETs Using Machine Learning. *IEEE Access* **2022**, *10*, 1893–1904. [CrossRef]
117. Khan, I.A.; Moustafa, N.; Pi, D.; Haider, W.; Li, B.; Jolfaei, A. An Enhanced Multi-Stage Deep Learning Framework for Detecting Malicious Activities From Autonomous Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2021**, 1–10. [CrossRef]
118. Almutlaq, S.; Derhab, A.; Hassan, M.M.; Kaur, K. Two-Stage Intrusion Detection System in Intelligent Transportation Systems Using Rule Extraction Methods From Deep Neural Networks. *IEEE Trans. Intell. Transp. Syst.* **2022**, 1–15. [CrossRef]
119. Ashraf, J.; Bakhshi, A.D.; Moustafa, N.; Khurshid, H.; Javed, A.; Beheshti, A. Novel Deep Learning-Enabled LSTM Autoencoder Architecture for Discovering Anomalous Events From Intelligent Transportation Systems. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 4507–4518. [CrossRef]
120. Heijden, R.W.V.D.; Lukaseder, T.; Kargl, F. VeReMi: A Dataset for Comparable Evaluation of Misbehavior Detection in VANETs. *arXiv* **2018**, arXiv:1804.06701.
121. Amanullah, M.A.; Baruwal Chhetri, M.; Loke, S.W.; Doss, R. BurST-ADMA: Towards an Australian Dataset for Misbehaviour Detection in the Internet of Vehicles. In Proceedings of the 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops), Pisa, Italy, 21–25 May 2022; pp. 624–629. [CrossRef]
122. Panda, M.; Mousa, A.A.A.; Hassanien, A.E. Developing an Efficient Feature Engineering and Machine Learning Model for Detecting IoT-Botnet Cyber Attacks. *IEEE Access* **2021**, *9*, 91038–91052. [CrossRef]
123. Ahakonye, L.A.C.; Nwakanma, C.I.; Lee, J.M.; Kim, D.S. Efficient Classification of Enciphered SCADA Network Traffic in Smart Factory Using Decision Tree Algorithm. *IEEE Access* **2021**, *9*, 154892–154901. [CrossRef]
124. Upadhyay, D.; Manero, J.; Zaman, M.; Sampalli, S. Gradient Boosting Feature Selection with Machine Learning Classifiers for Intrusion Detection on Power Grids. *IEEE Trans. Netw. Serv. Manag.* **2020**, *18*, 1104–1116. [CrossRef]
125. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: http://www.deeplearningbook.org (accessed on 5 January 2023).
126. Gonçalves, F.; Ribeiro, B.; Gama, O.; Santos, J.; Costa, A.; Dias, B.; Nicolau, M.J.; Macedo, J.; Santos, A. V2X Security Threats. 2020. Available online: https://doi.org/10.5281/zenodo.4304411 (accessed on 5 January 2023).
127. Seo, E.; Song, H.M.; Kim, H.K. GIDS: GAN Based Intrusion Detection System for In-Vehicle Network. *arXiv* **2019**, arXiv:1907.07377.

128. Capuano, N.; Fenza, G.; Loia, V.; Stanzione, C. Explainable Artificial Intelligence in CyberSecurity: A Survey. *IEEE Access* **2022**, *10*, 93575–93600. [CrossRef]
129. Rajapaksha, D.; Bergmeir, C.; Buntine, W. LoRMIkA: Local rule-based model interpretability with k-optimal associations. *Inf. Sci.* **2020**, *540*, 221–241. [CrossRef]
130. Das, P.; Ortega, A. Gradient-Weighted Class Activation Mapping for Spatio Temporal Graph Convolutional Network. In Proceedings of the ICASSP 2022—2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 18 May 2022; pp. 4043–4047. [CrossRef]
131. Draelos, R.L.; Carin, L. Use HiResCAM instead of Grad-CAM for faithful explanations of convolutional neural networks. *arXiv* **2020**, arXiv:2011.08891.
132. Mendel, J.M.; Bonissone, P.P. Critical Thinking About Explainable AI (XAI) for Rule-Based Fuzzy Systems. *IEEE Trans. Fuzzy Syst.* **2021**, *29*, 3579–3593. [CrossRef]
133. Wakjira, T.G.; Ebead, U.; Alam, M.S. Machine Learning-Based Shear Capacity Prediction and Reliability Analysis of Shear-Critical RC Beams Strengthened with Inorganic Composites. *Case Stud. Constr. Mater.* **2022**, *16*, e01008. [CrossRef]
134. Xu, Z.; Saleh, J.H. Machine Learning for Reliability Engineering and Safety Applications: Review of Current Status and Future Opportunities. *Reliab. Eng. Syst. Saf.* **2021**, *211*, 107530. [CrossRef]
135. Fryer, D.; Strümke, I.; Nguyen, H. Shapley Values for Feature Selection: The Good, the Bad, and the Axioms. *IEEE Access* **2021**, *9*, 144352–144360. [CrossRef]
136. Galli, A.; Marrone, S.; Moscato, V.; Sansone, C. Reliability of eXplainable Artificial Intelligence in Adversarial Perturbation Scenarios. In *Proceedings of the Pattern Recognition. ICPR International Workshops and Challenges*; Del Bimbo, A., Cucchiara, R., Sclaroff, S., Farinella, G.M., Mei, T., Bertini, M., Escalante, H.J., Vezzani, R., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 243–256. [CrossRef]
137. Rathore, R.; Hewage, C.; Kaiwartya, O.; Lloret, J. In-Vehicle Communication Cyber Security: Challenges and Solutions. *Sensors* **2022**, *22*, 6679. [CrossRef] [PubMed]