Explainable AI for Real-Time Threat Analysis in Autonomous Vehicle Networks

By Dr. Javad Salehi

Professor of Electrical Engineering, University of Tehran, Iran

1. Introduction

The paper will present a ML-based approach that aims to improve the robustness and reliability of AVs through on-device real-time traffic threat assessment. In particular, we will investigate how real-time traffic analysis and prediction can estimate near-future traffic situations in an unstructured and non-instrumented urban environment. We will delve into how a feedforward neural network is deemed as efficient to achieve the aforementioned objective. As the essential novelty, our approach is mostly based on real-world data, which might mitigate overfitting and data-scraping profoundly, if training data is not accurately selected, processed and probabilistically analyzed. We summarize the contributions as follows: (a) implementing a methodology to track vehicle trajectory data on urban networks; (b) real-time traffic origin-destination identification as well as time delay traffic map generation which represent the baseline ground-truth problems in our study; (c) local and fast intersection-based traffic light prediction; (d) global real-time parking lot prediction; (e) local and real-time traffic speed prediction; and (f) local and fast traffic density prediction for high and low traffic intensity areas with a prioritization for high-delay scenarios, unlike prior works.

[1] [2]Machine learning (ML) and related Artificial Intelligence (AI) technologies are swiftly becoming the backbone of modern autonomous vehicles (AVs). Their effectiveness and efficiency stem from their capability to seamlessly learn from the vast data generated by sensors and real-time operations of AVs themselves. However, being black-boxes, such complex models make it impossible for users to comprehend which signals trigger which reactions. Due to ethical reasons, the black-box problem might turn out to be undesirable and cost-ineffective for AVs, where trustworthiness and safety are critical.

1.1. Background and Significance

There are several concepts around which an XAI model can be understood, such as traceability, decomposability, comprehensibility, reliability, and trustworthiness. Each concept comes with its own particular nature and challenges, along with different combinations of components and purposes. For example, traceability, or interpretability, is the ability to draw back from a given conclusion at any moment, step by step, to a given source in the data (either initial or later feedback). Compressing this capability results in the structure "here and up" (to the output of the final layers) and "here and down" (till the input layer). An interpretable model that is traceable at this level is a more interpretable model, but since the data is huge and spans lower layers, it is practically infeasible to perfectly compress a given conclusion into easily interpretable decisions.

Explainable AI (XAI) has recently emerged as a critical capability of intelligent systems such as autonomous vehicles, which are expected to operate in various environmental conditions and to make real-time decisions [3]. In mission-critical scenarios, it is essential that human decision-makers understand the cause-effect relations based on explainable decision-making criteria. This is vital not only when regulation and law enforcement compliance are at stake [4], but also to win public trust. In other words, the more transparent and understandable the decisions of an AI system are, the more trustworthy they will be perceived to be [5].

1.2. Scope and Objectives

The mapping relationship between the driving behavior of unmanned vehicles and human drivers is usually very difficult to explain. To solve this problem, it is indispensable to make the working condition of unmanned vehicles known to users physically or psychologically. No interpretability can be generalized that the same degradation of trust would occur in every automated system, and so in the cur-rent work, we define interpretation as presenting control strategy output in a distinctively human-perceptible manner consistent with human senses to render system transparency amidst automation. Relevant human driver models are established in the experiment to simulate visual tracking and gaze prediction actions in steering tasks. The accuracy of an untrained CNN to predict the human driver's response is only lowered slightly on the PROGNET data set compared with the VIRAT data set, which means no significant loss of mapping relationship between human vision and steering can be caused as long as the efficient amount of necessary frames are used to predict driving actions. What' s more, interpretability is also achieved through the timeline visualization display, because the attention area sequence makes public, weak physical explanations are also given, together with high cognitive interpretability [6].

The ultimate function of the mapping relationship between the driving behavior of unmanned vehicles and human drivers is to establish the advanced driver-assistance systems, which is influential to its safety and comfort, inspire better human–machine interaction, and largely compress the gap from the unmanned vehicles to ultimate successful driverless driving [1]. However, the complex principle of data processing and decision-making in the above process are hidden and are also not easy to explain, which may tremendously limit the development of unmanned vehicles in a large range. At the present stage, the safety and reliability study of unmanned vehicles are mainly carried out from the perspective of two levels, namely micro level and macro level, specifically, road lane marker recognition, vehicle detection, and pedestrian detection in the micro level and energy saving, traffic fluency, and real-time control in the macro level. But in fact, it still lies some asymptotic space, because the issue of driver-vehicle interaction has not been stated in the same level [5].

2. Autonomous Vehicle Networks

On the other hand, all XAI mechanisms have limitations in terms of their effectiveness when applied to complex, distributed, or large system models. Recently, too little attention has been paid to how systems' explainable designs and components could be revealed to determine their level of efficiency or to share this information between difference explainable components. Consequently, performance evaluation of XAI needs to be investigated, and metrics to assess the effectiveness and efficiency of mapping from XAI components to simplified performable representations in terms of system behaviors should be developed. This is particularly important considering the deployment of intelligent and autonomous systems in real-world applications such as autonomous vehicle networks [7].

Explainable AI (XAI), formally known as interpretability or transparent AI, has been defined as "any form of AI whose actions can be meaningfully understood by humans." XAI has several unique characteristics: (i) to bridge the gap between humans and AI, black box AI systems are replaced with explainable or transparent AI systems; (ii) the ability of users to trust their AI systems depends on their level of understanding of the model; (iii) if clear explanation of decision making is not possible, it becomes impossible to anticipate what AI may do; and (iv) XAI is expected to contribute to the addition of legal accountability for AI models [8].

2.1. Overview of Autonomous Vehicles

One challenging aspect of realizing future mobility solutions based on vehicle-to-vivhicle communication is that trust between systems (vehicles) is often computed based using artificial intelligence. Recognising different and emerging security attacks is essential for achieving trust within the system and safe and secure operations of traffic. The proposed AURENE project consists of a cellular network-Ethereum one-2-way communication, where ground owners (own the road infrastructure) are incentivised to cooperate with each other for contributing to a direct long-range vehicle-to-vehicle communication between CAVs. CAVs at the front of the vector generate efficiently observediteall for the road owner first and the road owner queries the cellular network for the CAVs in his region for transferring the message and initiating the observed one-to-one communication. This is an advantage to an attacker as between each transmission of a sender, different route of senders and receiver are considered which hides the incoming transmissions of te attacker.

The creation of distributed autonomous vehicle systems is of vital importance for the near future as it stands to be a corner stone for sensing, data collection and processing within traffic management and smart city systems. Within the CAV ecosystem vehicle-to-everything (V2X) communication will facilitate mobility while guaranteeing safety and security. However, shifting to such vehicle-to-vehicle communication will potentially increase attack surface at the vehicle level. The Automated Sharing UtilizAtion and RecognitioN in rEaltime (AURENE) project has been proposed as a decentralized solution to cope with threats to the autonomous vehicle networks. Recognition of cyber attacks in autonomous vehicles has been proposed in [9]. In-van traffic distribution has been suggested as an alternate communication method that is sufficiently decentralized and removes the attractiveness of becoming an easy target for cyber attacks.

2.2. Communication and Networking in Autonomous Vehicle Systems

CAVs are able to connect to enable cooperative driving and communication [10]. Cooperation between CAVs begins based on the technologies like vehicle-to vehicle (V2V) communication or external signals with vehicle to infrastructure (V2I) communication (e.g., from traffic lights or traffic signs) [9]. It is necessary to incorporate this connection with a warning system in the dynamics of the autonomous driving system (ADS). CAVs are able to process the encrypted communication with the classical sophisticated security systems. The deep learning algorithms are implemented into the software of the CAVs and therefore, high-performance hardware can be conservatively used in the industry projects, e.g., high-quality radar, LiDAR.

V2V is vehicle-to-vehicle communication, or in other words, wireless transmission of driving data between vehicles. V2I is vehicle-to-infrastructure communication, i.e., wireless transmission of driving data between vehicles and smart road infrastructure. V2N, standing for vehicle-to- networking devices, is the wireless communication for collecting the driving information of vehicles, and it can be set up within a specific geographical area, which is quite similar to V2I information exchange. V2X (vehicle-to-everything) is a system where vehicles can communicate with other systems built into the traffic environment. V2V, V2I, and V2X are the most important communication types of CAVs (connected and autonomous vehicles) [11]. During the communication between CAVs and the external interfaces, security and privacy are important topics to be considered. Thus, intruders can have the possibility of just listening to the exchange of information, or indeed, they remotely try to access them. Further to this, it can be seen that fake Kalman filter (FKF) schemes are used to disturb the performance of secure vehicles. Furthermore, this filter can delay the signal processing of the secure vehicles. The use of cryptographic algorithms can add severe overhead to the processing time and message length, which is not suitable for real-time applications like the remote update or in-car multimedia (ICM).

3. Artificial Intelligence in Autonomous Vehicles

Most of the classification methods leverage machine learning (ML). The ML model has proven to be very successful at data classification. Moreover, long short-term memory (LSTM) networks were utilized in order to detect biometric leaks and to protect user data, creating an advanced approach for real time driver-in-the-loop security. However, another challenge lies beyond data collection and model training, and it encompasses the real time and integrated security system. In fact, this step relies upon a system integration designing capable of taking into account autonomous vehicle security necessity from scratch. This research presents advanced work on threat usage and abuse cases identification into the system modeling and software architecture. [12]

Security and safety are crucial in the domain of autonomous vehicles. This is a completely new research area, with all the associated technology facing significant security threats and risks. As hardware and software tools have started to advance, they have brought alongside them a new set of substantial security vulnerabilities. Evolving real-world threat scenarios make it particularly difficult to support the safety and security of autonomous vehicles. Artificial Intelligence (AI) can offer new opportunities in developing the security infrastructure for autonomous vehicles. This is especially relevant in considering solutions which can support the capability of real-time risk analysis and threat detection. These new systems now face the challenge of real time and accurate classification, as well as a need for continuous monitoring and debugging. [11]

3.1. Machine Learning and Deep Learning in AVs

Explainable AI (XAI) is a branch of AI, where one of the scientists tries to develop a specific model for an AI system that shows the underlying reason for taking a decision whenever something goes wrong in the form of explainability. For computer vision, the traditional data flow for action kernels or first order linear convolutional layers is inspired from human vision and this is very intuitive. Beyond this point of view, it is not possible for even trained specialists to understand about their decisions. This gives the idea for incorporation of intelligence like human's that will lead to adaptive learning or to use another approach called explainable AI for an expert system development.

Autonomous vehicles (AVs) are vehicles that can operate without human intervention. So, these vehicles use machine learning or deep learning to take driving decisions [13]. Autonomous vehicles are primarily connected with vehicle-to-vehicle or vehicle-to-infrastructure communication technology. These connected devices share their data among themselves to makes driving decisions. This shared data could be useful for any attacker to attempt an attack on any of the AV nodes and then this attack might disturb the whole AV network [11]. It is agreed in recent research to implement the explainable AI more in future AVs, where it will become very difficult for the attacker to bypass it by controlled data adversarial attack [14].

3.2. Explainable AI in AVs

For autonomous vehicles, undesirable or dangerous information can result in alert fatigue, inevitably affecting driving safety. Nevertheless, of critical concern is understanding how an AV made a risky decision after it has caused an accident. In a black-box model, such an investigation would require replicating the environment conditions that existed at the time accident occurred. However, an explainable decision-making model would provide insights with respect to the model features and the prediction scores, allowing for accurate post-hoc analysis. Driver safety could be improved in the long-run with the decision-making process and the rationale behind these decisions being made more transparent. Unlike expert drivers, AI systems do not have an inductive bias, an implicit 'sense of consideration'. There is no assurance that life-threatening decisions will not be made during training or production. In this sense, there is a lack of trust for AI-driven vehicles by some. In this vein, the provision of clearing rational decisions, those that fall within commonsense, would potentially build confidence in the driver or an autonomous passenger [15].

The recent advances in artificial intelligence (AI), particularly in machine learning and deep-learning, have led to the growing presence of software-driven decision-making algorithms in complex domains. In the realm of autonomous vehicles, traditionally open-loop control systems are providing tarmac to closed-loop systems. In the latter, vehicle control systems observe feedback and make real-time decisions, akin to humans [3]. Recent studies have shown that these AI-driven vehicles make intelligent decisions by exploiting large volumes of training data. Even though explainability and transparency are critical for autonomous vehicles, most state-of-the-art AI-driven vehicles are opaque and hard to understand by end-users [6]. To address this issue, the field of AI has suggested the development of Explainable AI (XAI), emphasizing that intelligent behaviors are generated by intelligible subsystems.

4. Real-Time Threat Analysis

Further, Autonomous vehicles, like any other interconnected systems, are subject to security attacks and should be safeguarded from being compromised. In recent years, securing autonomous and connected vehicles has received increased attention. Mentioned security attacks and countermeasures methods can be classified into three categories: Physical attacks, protocols and cryptosystems and Data and Operation Systems. Though immense works have been done in these directions, host-based attacks (e.g., physical modifications, obsfucated binary code execution, AI-based blade attacks) are still under-focused though AI-driven self composed attack testcases are provoke. But, it is significant to reason about the training of various models on the host based dta of self driving cars, for example, how to prevent the ultimate stop of the car or how to trick the decision making system for access and gain control of the cars, all those tactics are envisioned. Therefore, a systematic view on the current self driving vehicle ready attack vectors and techniques, as well as possible countermeasures would fill this hole [16].

Decentralized real-time threat analysis services are a necessity for the future of autonomous vehicles. The promise of vehicles being able to self-heal by sending and receiving warnings about attacks is an appealing proposition; however, real-time threat analysis services that determine, in a decentralized and timely manner, if a vehicle is subject to an attack that prevents it from providing safety and reliability, are crucial for the safety of current and future autonomous vehicles. Recent attacks on self-driving cars have demonstrated that the sole trust in learning-based models, in particular computer vision models, poses a severe threat to the safety and security of self-driving cars, creating the necessity for real-time and explainable threat detection systems [2]. It is impossible for learning-based models to learn all possible actions in all possible scenarios. In addition, small shifts in the outcome of the last layer of the model could have disproportionate (and poorly understood) effects on the final decision of an end-to-end learning model. For that reason, computer vision models are especially vulnerable to adversarial attacks [17].

4.1. Types of Threats in AV Networks

The resulting attack landscape is rather diverse. There are attacks that first disrupt the communications between a vehicle and the cloud or other vehicles to perform a bunch of other adverse actions on a vehicle. For example, when a vehicle, under fake signal respect to cloud, receives a "stop command" from the cloud, the vehicle needs to be stopped immediately or risk collision. If the stop command is fake, indeed, and it is a phishing attack, a severe risk of a potential collision will occur. Secondly, the adversaries still block vehicle to cloud flow of information to hide themselves and make vehicle-to-cloud parameter malicious value alteration. As a result, vehicle to cloud properties may be affected very badly. Car-attacks of then blocked vehicle-to-vehicle(view blocking attack) properties will be performed maliciously. Attackers can make two-friend cars to collision with a hacking strategy in the view blocking attack [18].

Defending vehicles and networks from attacks is a critical emerging problem. A growing number of studies have suggested that various autonomous systems including CPSs and vehicles are now equipped with more advanced hardware and larger software stacks making them more vulnerable to attacks. These attacks become a bigger challenge when the systems start running autonomously. Since connected and autonomous vehicles (CAVs) are sensitive to cyber-attacks, we categorize a variety of threat scenarios that autonomous vehicles could be exposed if vehicles are being attacked; we consider three threat models [11].

4.2. Challenges in Real-Time Threat Analysis

Although some strategies for real-time threat analysis and real-time intrusion detection and prevention systems (IDS/IPS) have been proposed to mitigate the security risks and attacks in autonomous vehicle systems (from edge nodes to vehicular networks) through (1) learning-based predictive methods and models; (2) the design of a data-driven IDS using the network intrusion prediction model (NIPM); (3) a hybrid IDS framework using multiple neural networks and deep learning mechanism; and (4) an LSTM-based autoencoder model to provide comprehensive solutions. They were viewed as insufficient to provide continuous protection and adjustability in response to changing environments, evolving new security vulnerabilities, and advanced diverse attack patterns in real time [16]. An ML-based predictor and a data-driven IDS are used for real-time threat analysis. An LSTM-based autoencoder model is used in conjunction with the IDS to detect new or unknown attacks and assist in their feature extraction. An attack graph model is employed to visualize and compute the threat severity level of new or unknown attacks within the target network. An anomaly-based ML model is used as a final prevention response mechanism based on a safety and reliability analysis to prevent their impacts on the target network. The results of computational experiments with the target real data of KSU show that this comprehensive system in real time effectively detects autonomous vehicle network attacks and predicts the outputs of future intrusions in heterogeneous networks.

The survey in [19] showed that some key challenges regarding security in autonomous vehicle networks are diverse and complex. The following challenges pertain specifically to real-time threat analysis: (i) Due to the highly dynamic and potentially wide-range networks of vehicles and infrastructure in autonomous vehicle environments, it is difficult to respond effectively to new and unknown security threats and attacks; (ii) the diverse security threats and attacks can vary in the number of attackers and location, attack paths, and attacker motivations, resulting in great challenges in risk analysis and formulating appropriate threat hunting strategies; and (iii) different connected devices in these vehicle networks, such as 5G base stations, edge and cloud servers, and vehicles, may significantly vary in computing and storage resources, processing speed, and energy demands, which will impact real-time prediction algorithms, decision-making and follow-up response decisions. The survey results further revealed that by continuously incorporating new information and minimizing the amount of historical and training data required, the need to pragmatically collect and process various real-time log data in autonomous vehicle networks in the presence of these challenges.

5. Explainable AI Techniques

The requirement for AI models improved by the advent of deep learning software libraries with increased focus on explainability eventually forged the path for the development of explainable neural networks (XNs) such as LRP based approach to deep networks. XNs work by assigning saliency scores to the different neurons across the different layers of a neural network, which are then processed for generating relevance maps providing transparency to black box operators. XNs are computationally minimally invasive as they require executing strictly the forward pass of the primary predictive network only once to provide explainable predictions compare to e.g. the backpropagation-based variant of the LRP technique [2].

Developing algorithms with explainable AI has become very important for the development of safety-critical systems, such as autonomous vehicle networks. Black-box algorithms like deep neural networks require interpretability and explainability. This has led to the development of explanatory neural networks [20] and the use of techniques like creating proxy models and salience maps. Interpretability and explainability are essential for understanding how data is processed and for assessing hazard and risk analysis. In the AI-Tooling framework, users can observe activations of different layers in the neural network architecture to interpret decisions. In complex networks, reducing parameters or focusing on the most important parts can enhance explainability. Concentration on the areas in the data or model that are most relevant to the problem being tackled through localized modifications using saliency methods can also be a key for explainability of the predictions and decisions relating to model outputs to their input features [21].

5.1. Interpretable Machine Learning Models

The future of optimized EV charging and driving solutions lies in the implementation of real-time data from the AVN and intersecting it with real-time data from the grid, electric distribution networks and generation. In the best scenario case, the data themselves will not be subject to any attack and will only be sensitive enough to aid researchers in creating trend-impacted models in increasingly time intensive simulations. Considering many open-source projects and unavoidable protocol intricacies in future-proof smart electric transportation systems, one can assume that the operational pneumonia stage might remain between researchers, grid operators and attackers. It is essential to keep it that way and even make it harder for the latter by considering newly discovered attacks and simulating foreseeable future risks in laboratory settings. The local implementation of robust encrypting transformers or increasingly reliable differentiation strategies in Torch, Keras and TensorFlow2-based privacy tools should not allow attackers to guess the exact step when and where the line shall break.

The emergence of Explainable AI (XAI) has seen significant debate and ongoing research. Many real life applications and projects benefit greatly from increasing AI interpretability and transparency [22]. In power systems of future cities, the tension between energy management, green technological installations and flexibility still makes it impractical to rely solely on intricate optimization solutions. As a result, decentralized explanations will be essential; especially if smart scheduling MaaS tools and complex energy algorithms cloud these local flexibility prediction algorithms. A bright side of this would be visibility at all stages: optimal, near-optimal and heuristic deployment of data science, especially when vehicle users can become micropower RES managers. A vain, fatalistic hope would be to expect the N-1 future to deliver super-interpretable models, compatible in vision and action [23].

5.2. Model Explanation Methods

5.2.1. Saliency map Saliency maps provide a white box explanation method, showing the contribution of relevant pixels in the input image to the classification of the network. For a supervised learning task, the saliency map can be obtained by calculating the gradient of the output of the network with respect to each pixel of the input image. Typically, then, the saliency map is derived by back-propagating the gradient to the input space for image classification models. This provides more cognitive learning, but the solution is specific to only images and does not work for textual data. When applied to a sequential data classification task, the gradient becomes the sum of gradient of each time step and these gradients are used to calculate the importance of each time step in the decision of the AI model.

Explainable AI (XAI) incorporates model explanation strategies, supports domain expert interpretation of model outputs, and provides confidence information about model predictions [24]. In practice, gap evaluation between confidence and users' perception influences the model's overall trustworthiness [25]. This section summarises XAI methods, ranging from local explanations like saliency detection, which identifies the pixels playing a key role in the classification, to global explanation methods like SHapley Additive exPlanations (SHAP), which summarise the model's prediction function using cooperative game theory. For autonomous vehicles, understanding the models' reasoning becomes essential for humans to trust, control, and manage them.

6. Case Studies and Applications

The explainability paradigm for real-time autonomous vehicle network threat analysis is receiving significant attention these days due to the non-explainable nature of state-of-the-art threat monitoring and analysis tools. Yet, due to the stringent latency constraints in the network, real-time deep learning paradigms cannot be very readily adopted for the task unless several improvements are made in the architecture and are explicitly studied [4]. Limitations of the state-of-the-art research lauding the development of ML and DL-based threat monitoring models have been suggested in. While the problem of adversarial attacks primarily suits the situation of evolution in the distribution due to adversarial signs that may get added to the actual distribution expressed by the model, the need to perform complex calculations due to a potential increase in network resource usage and the utter lagging of the deployment process seemed to be the most bothering drawback that was pointed out. To address these and other related concerning issues in the designs of threat analysis tools for autonomous and intelligent vehicle networks (AIVNs) with explainable artificial intelligence.

The ultimate aim of "Autonomous AI," AI-driven processes and AI-enabled applications, is for machines to replicate human judgment. The key focus of this chapter is Explainable AI in Depth, covering a feasible state of the art technology for Real-Time Threat Analytics [26]. This is especially critical in autonomous vehicle networks where a detailed understanding of the operations can set the basis for growth and development of smart vehicles and the associated infrastructure. Vehicular network Intrusion Detection Systems (IDSs) traditionally use machine learning (ML)-based models for detection where the traditional Convolutional Neural Network (CNN) has also been employed [27]. In general, deep learning (DL) and ML based IDSs constitute the majority of IoV research. Moreover, static datasets with constant distributions are generally employed in research limiting the application of such approaches to the real-world domain. This problem is exacerbated due to the latency and throughput considerations for real-time IoV networks. To cover the real-time requirements, the Memetic III3, DL and ML based RSU model for real-time IoV network Intrusion Detection (ID) with varying traffic distributions.

6.1. Existing Systems and Technologies

The Controller Area Network (CAN) in-vehicle network has been effectively used in for veichle for the purpose of communication in the in-vehicle network on the time and is no longer reliable enough. Taking into an account of the CAN, more powerful and complex security system metrics have been defined and used in research works. The in-vehicle network driving system is also calculated and decided by the controller at run time. So, lack of security measurements in CAN cause tremendous accidents. To detect the attacks CAN network-based manipulation, security measurements must be determined. For security considerations, artificial intelligence (AI) and machine learning (ML) have also been effectively used in recent research works. In machine learning, the large set of network traffic-based features have been used such as Standard deviation, Mean of network traffic bytes, and number of packets. These features are used to classify genuine and manipulated packets. In combination with related research works on machine learning security requirements, classification-based security measurement, decision-making and detection of manipulation-based network traffic are explained [28].

Vehicular ad hoc networks (VANETs) have been recognized as an effective cooperative architecture for road safety, traffic efficiency, and environmental protection [29]. However, the security of VANETs is a major concern, and attacks are found through the major focus on intelligent transport systems (ITS). These threats include the launching point intrusion, denial-of-service (DoS), and message alteration. Furthermore, the self-driving cars (Autos) need to satisfy the broad challenge of interstate protection, distinct from all ITSs' security problems. Therefore, the security mechanism has a lot of challenges for both ITS and self-sufficient cars.

Analogously, alike to conventional intrusion detection systems (IDS), network IDS and host IDS (H-IDS) need to be applied to ensure network entry and exit security in all-inclusive VANETs. In general, vehicular cloud networking investigates the transaction between all vehicles, the roadside unit, vehicles, and back-end servers.

6.2. Use Cases in Real-World Scenarios

XAI system with overlay, divided into target functions and vehicular network edge and cloud infrastructure use cases for domain-specific and cross-domain threat analysis, including their attack, defence and detection with possible hybrid Learning-based model of attack and use case-specific defense model on the left. Two practical examples; omitted services and SOS emergency counterfeit usage by attackers are depicted on the application layer with different sign representation of detected and responding potential anomalies while the explanations for the phenomena can be transferred through all higher layers by using the extended simple analytical threat description in the form of text, rule-based system rules, the mapping performances of domain-specific classifiers, and I 2 A 1 X A S edges to help potential human assessors in detecting, understanding and judging different situations and network events in predefined simpler to more comprehensive epicentre views for cluster vitals marking in route, leaving spoilage marks related to P2P and SDN-SCN relations through an optimal network topology; civil vehicle and non-vehicle interactions and payload attacks effectiveness are given [8]. Not only local but also distributed ML-based and rule-based monitors and detectors have the capability to work coherently to aggregate results of intelligent results to selected segments of IoT structures and supportive networks from Vehicular Ad hoc Networks on the Physical layer in cross-layer service with biomedical and RFID technologies, cellular and edge transmissions to deployed protection switch sufficiently reliable and efficient to integrate multiple sources of data in Vehicular Blockchain, strengthen interwoven-related coherence with the vehicular population, and represent attacked interlayer connections by using a unified simulation system and perform Edge Intelligence on centralized IoT and autonomous vehicles at intermediate edge SCNs with additional edge layers to real-time react on the scenario graph martyred as the general knowledge sorter available on a classical global Internet scale [11].

In the context of connected vehicles, systematic classification from an endpoint (vehicle) to the cloud infrastructures can be performed to enhance explainability, (X)AI trust, and, therefore, overall safety. Categories can also include altered/trained sensitivity to relevant input parameters (e.g., vehicle vehicle distance and relative speed differentiation in models like they can predict a wide range of critical scenarios and assign a higher priority to more likely states. In the same vein, synthetic data for simulating very rare events, as well as adversarial examples that aim at activating or suppressing some certain features in a careless manner and to monitor any abnormal predictions and responses, are crucial. These types of explanations can produce different insights into an AI model to the logic, according to which the model operates, and to the intention of the model because the ultimate goal must be engaging creative control achieved through sensitivity analyses. The common techniques related to white-box models are related to explainable deep-learning approaches like hierarchical representation learning and decision paths while optimising on a training set also called RuleFit. The final appealing solution is an integrated unprecedented XAI system with complementary functionalities and services for vehicle holistic threat analysis in real-time across all OSI layers [30].

7. Evaluation and Performance Metrics

A secure and compliant AI-driven digital infrastructure is crucial for automotive manufacturers to efficiently and safely operate their business models, produce vehicles conforming to all the applicable laws and standards, and ensure modern-day consumers' data privacy [31]. When operating a network for hosting IA applications and services, in compliance with a principal of GDPR and data protection, users who access the network's privacy-sensitive functions are anteceded to provide the AI/ML model with evidence of confirmed identity of themselves [8]. It is proposed that the monitoring of the Received Power Indicator (RPI) information that is transmitted from the full payload-to-compliant vehicles could be employed as a form of payload integrity validation.

As can be seen from the results above, the attack injection system would have the potential to successfully attack a target low-level control system over the FlexRay network interface should a vulnerability be found that would allow the attack injection system to successfully locate and access said target. Fortunately, most low-level control systems already have very strong and robust protection against such attacks by age-old design, ensuring the overall trustworthiness of the hardware layout in autonomous vehicles.

7.1. Key Performance Indicators

Another driver KPI discussed in the context of a vehicle network and its current application layers is driving comfort. The action execution, such as lane change requests or inter-vehicle coordination on the longitudinal and lateral vehicle control KPIs, should aim to providing the most comfortable and consistent maneuvers [32]. The lateral control KPIs describe the target behavior of the vehicle's operation in a longitudinal direction. Furthermore, there are global comfort KPIs that the vehicle has to respect, even if a strong evasive maneuver is necessary to avoid a dangerous collision or misunderstanding, especially in heavily occupied traffic during traffic jam evaluations. The behavior of the vehicle may include its own driving comfort and the comfort of other partners on the road. In many of the controlled traffic scenarios, especially in scenarios without prior coordination and agreement about the intended maneuvers, the margin for potential actions becomes much more restricted. For example, unrestrained lateral maneuvers for an AV are not possible, for instance, during dense traffic due to potential losses for other partners on the road by higher amplified control commands. An overall introduction and further details about the vehicle and system KPIs for traffic applications and collision avoidance from a control perspective can be found in other publications [ref: 578deea2-91d7-48ce-b3be-7e78c0bb8fe2; ref: f9475fc0-1c24-4d24-be8c-0acbd01a7270].

[31] Key performance indicators (KPIs) provide an overview of our vehicle's functional aspects. The demands for diverse vehicle collaboration and collision avoidance scenarios define a number of driving requirements. One of these KPIs is the control effort and energy involved when coordinating and arranging movement tasks to avoid accidents. The executing tasks are lane keeping and speed control, lane change or overtaking, coordination on merging and weaving, cooperative-adaptive cruise control, cooperative route planning, and collision avoidance among members of a vehicle network. The performance of distributed cooperative collision avoidance is influenced by all these vehicle–environment behavioral interaction KPIs. The actions decided and adopted by the higher-control levels should focus on eliminating dangerous collisions and managing collision avoidance geometries.

7.2. Comparative Analysis of Techniques

Laser et al. have presented "Feature Scattering as a Defense Against Adversarial Examples," another defense mechanism sought from capturing conscious local perturbations aimed at changing classification labels [33]. In this study, authors explore if spatially-correlated noise can be similarly helpful for making deep networks non-differentiable around input images. It has been empirically observed that adding random noise to input features can make adversarial attack difficult by taking the random Gaussian noise as inputs regaining smoothness and differentiability in the decision surface. The proposed defense strategy does not require retraining after receiving adversarial attacks, and when applied with conventional adversarial training, they observed the proposed feature scattering improves test natural and adversarial accuracy, adversarial diversity with multiple attack techniques, and attack iterations. Their method has improved accuracy with respect to noisy data; secure predictions with and without adversarial training models are provided when training with feature scattering regularized models.

Conventional security metrics, such as classification accuracy, are not well-suited for evaluating deep learning models' performance against black-box attacks [17]. Adversarial training, where models' weights are optimized using adversarially perturbed samples as both ground truth and perturbed inputs, is a straightforward technique to train adversary-tolerant models. Even though adversarial training has shown impressive success in improving models' accuracy and stability against adversarial samples, it is accompanied by a few concerns, such as additional computational overhead, the decrease in natural image accuracy, and inherent first-order approximation non-Gaussian likelihood of the adversarial example. Moin et al. have quantitatively shown that adversarial training minimizes a surrogate bound of the adversarial risk at the cost of larger natural risk. They propose a new algorithm called bad-child re-train and show that it reliably reduces the natural risk at the expense of the adversarial risk. Brendel et al. have presented a black-box adaptive attack based on zeroth-order optimization to fool machine learning models [34]. They release 65 thoroughly evaluated evasion attacks across 12 classifiers and show adaptability in a score-based, decision-based, and score squeeze reduced setting.

8. Future Directions and Research Challenges

Moreover, the paper discussed the real-time threat assessment and remediation in the IoT/I2T sensor-based AI systems in the emerging scenario of the ecological war for the dominance, including vehicle-to-everything and internet for vehicles autonomous vehicle networks. Post-facto threat assessment is very well served by AI. However, explainability and reliability of the instant real-time threat assessment window AI tools and techniques in the current I2T sensor based AI network is still under research [35]. Future research must also explore the fireball scenario of sensor data based AI engine reasoning during very rare and unseen, unpopular events or scenarios, which could have a macroscopic impact on the society, partially or in whole. For selected future research scholars, we have also suggested a few of the unsolved cyberphysical hardware level vulnerabilities which need to be dealt with, to transform privacy of the data into an art of proverbial science.

Because connected vehicles and autonomous vehicles period, relying on several intelligent and IoT sensors, are threatened by various security attacks such as injection or spoofing of sensory data, unfair use of shared spectrum, privacy erosion through inference attacks, and so forth [26]. Sensor data-based artificial intelligence (AI) engines will be often attacked in the future, prime facie. They too will require AI techniques to detect and defend themselves, which will slowly morph into an ecological war for dominance and control. It is essential for this upcoming battle to not transform into a “mutually assured destruction” phase, like the current scenario of the cyber security domain, which is heavily dependent on the hacker, not on the defender. Safety assessment and management of these connected and autonomous vehicle I2T networks and the evolving physical layer data science domain should be augmented with the explainable AI techniques, in order to smoothen the public acceptance of this peak of the human intellect [2]. This paper has explored the fundamental need to develop the explainable AI (XAI) techniques for all the future domains of the internet of things (IoT), the internet for vehicles (IFV), and the internet for things (I2T) or automotive Internet of things.

8.1. Emerging Trends in Explainable AI

To this point, an exploration paper which surveys the essentials of logical AI in AV organizations gives rich outlines of AI classifications, including AI depictions for traffic checking from a self-driving perspective, normal AI classes for self-rule and Demonstration of the AI disattach will be publicized [8].Returning to scientific AI applied into the authority and checking of a few different AI-controlled vehicles, this article recommends and talks about arising evolving approaches for AI techniques, however doesn't expound on confirming foundation portrayals happening. Both logical philosophy and traditions related adapting to the different comparing swelling approaches being pluralism— AI as a non-algorithmic case.

Nowadays, society is becoming increasingly dependent on AI techniques, including machine learning (ML) and deep learning (DL) techniques for driver help and independent applications. Traffic incidents, their details, and the causes are followed through these applications. Despite the fact that these methods are exceptionally powerful, it is hard to deliver persuading critical thinking clarifications for the discovered issues [1]. In such a manner, actualizing and utilizing an interpretability methodology for AI calculations is of extraordinary significance today because of lawful prerequisites, to let individuals look and validate the expected conduct of the AI-controlled independent frameworks and to improve client trust [36]. Clarifying the critical thinking of cutting edge AI-controlled AVs would even encourage constant observing of their expected activities, expanding security and unquestionably adding to the efficient usage of the AI calculations.

8.2. Unsolved Problems and Open Research Questions

More research is required to understand how human drivers behave. The vehicle AI should be able to understand the driver by interpreting the driver's behavior of operating the vehicle (accelerating, decelerating, turning) and the driver's intention by predicting how the driver will drive [1]. In the short term, the design of driver monitoring module in vehicles should combine the design of the monitoring system with the prediction of driver behavior. In the long term, the development trend of vehicles driving towards L5 will continue to increase the degree of automation, and L5 vehicles do not require the driver to perform manual driving tasks. By then, the driver no longer needs to be monitored, and the development direction of the module in the car is the driver's behavior state recognition.

Regardless of the substantial advances, there are still major unsolved problems and open research questions [3]. An important unsolved problem is dealing with rare and unexpected events in which explainability needs to be augmented with reasoning that is capable of representing various sources of information (e.g., expected vehicle behaviors and human drivers) instead of just focusing on explainability of the local decision disparities. Indeed, interdisciplinary research projects should address the AI vulnerabilities, ensuring the robustness and resilience of the system against failures, so regaining the trust of users when failures occur. These convergence research endeavors call for interdisciplinary collaborations, involving both experts from AI and transportation research domains and stakeholders in developing and validating research scenarios.

9. Conclusion

For iSENSE architecture to prioritize threat detection management engines, the unsupervised im- age preprocessing engine could not only help enhance the anomaly segmentation detection mechanisms. But, it could also help design a manufacturing-specific attack modality for adversary modeling of the defense system [37]. Perceptively, an artificial neural system learning to efficiently categorize the chemical surround- ings of the vehicular environment has to learn instantly in face of an rejecting adversary in the car bot system. In mathematical terms, a multi-layered neural network has to adapt to the new data distribution generated by a modifying adversary coming from the mighty iSENSE [38].

The goal of this study is to guide the development of a threat identification system for real-time analysis of security outliers in vehicle networks. These concepts can potentially be extended to Level 4 and 5 autonomous vehicle applications. Using large-scale neural network architectures with domain knowledge integrations facilitating explainable AI, we have endeavored to develop an AI-powered defense system called iSENSE (intuitive Security NEural network System for Autonomous driving and Vehicle Networks). At the very beginning, iSENSE employed various unsupervised methods to understand the opened up decision map towards the AI-specific threat analysis detection in vehicular networks. The proposed ideas and hints in this study walk through several numerical demonstrations and routes in exchanging threat vectors between AI development and these data-driven defense functions.

9.1. Summary of Key Findings

[12] In order to address the limitations of NIDS in providing a complete, real-time threat awareness, we proposed an end-to-end Explainable AI system for threat detection in on-board networks of autonomous vehicles, which uses SHAP to provide actors as well as threat explanations. In order to gain the necessary understanding of how, when and why attackers perform these manipulations we also investigated corresponding real-world manipulations on a test vehicle of a German domestic OEM [39]. As an input for actor explanations in the scenario, a SHAP value is passed to IDA to reason about the trust level of received threats and whether the attackers attempts were successful in the respective scenario.[40] Based on the presented use cases the advantages of the approach should be evident. It is possible to understand, counteract and prevent sophisticated, realistic threat scans on V2I and V2V levels in a just-in-time manner as well as detect and validate the point in time when a scan was

performed via analysis in real time. The study indicates that it is a suitable approach for the higher-order ICS networks and reasons about the actors, their states as well as executed threat scenarios and provides necessary explanations.

9.2. Implications for Future Research

Another technique that is discussed to detect zero-day attacks for RL and block them is to influence the network input factors, by using a new feature selection method, inspired by multiple model predictability theory. The new interference method can be used to enforce a multi-layer machine learning (ML) model to build human-like decision making models, and hardware Others-of-a-kind (OOK). OOK-based interference approach with diversities, can reduce adversaries' chance to gain unified inputs. The impaired network reaches the adversarial decision space (ADS) to extend the explain presupposition, and this approach protect any endorsing attack scenarios, with less probability. Pre calibrated real-time driving on smart city real environment only for OOK gained human-likeness-boosted agents. These systems reach an explainable real-time MDP that interacts with adversaries anymore. These studies provide an answer for the first and the second research question and thus affected the associated long-term implication of AV networks.

[41] [38] [2] A crucial future research direction is to develop intelligent methods and algorithms to detect and block adversarial zero-day attacks in the network. Machine-learning algorithms without any consideration to the adversarial samples and scenarios are vulnerable, due to their excessive real-time updates. Legislation issues could come into play, if these algorithms produce a significant number of accidents or fatalities in smart transportation systems. How people or governments hold the blame factors impacts the result of court decisions, which play a crucial role in safe AI flow assurance. A new method is proposed to detect and block adversarial zero-day attacks using the structural dependency graph, which helps to convert reinforcement learning-generated real-time signal triples, to the related explainable graph pattern.

Reference:

1. Tatineni, Sumanth, and Anjali Rodwal. "Leveraging AI for Seamless Integration of DevOps and MLOps: Techniques for Automated Testing, Continuous Delivery, and

Model Governance". Journal of Machine Learning in Pharmaceutical Research, vol. 2, no. 2, Sept. 2022, pp. 9-
41, https://pharmapub.org/index.php/jmlpr/article/view/17.

2. Prabhod, Kummaragunta Joel. "Advanced Machine Learning Techniques for Predictive Maintenance in Industrial IoT: Integrating Generative AI and Deep Learning for Real-Time Monitoring." Journal of AI-Assisted Scientific Discovery 1.1 (2021): 1-29.
3. Tatineni, Sumanth, and Venkat Raviteja Boppana. "AI-Powered DevOps and MLOps Frameworks: Enhancing Collaboration, Automation, and Scalability in Machine Learning Pipelines." *Journal of Artificial Intelligence Research and Applications* 1.2 (2021): 58-88.