IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 January 2024, accepted 28 January 2024, date of publication 2 February 2024, date of current version 8 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3361829

SURVEY

Anomaly Detection in Connected and Autonomous Vehicles: A Survey, Analysis, and Research Challenges

SIHEM BACCARI[1], MOHAMED HADDED[2], HAKIM GHAZZAI[3], (Senior Member, IEEE), HAIFA TOUATI[1], AND MOURAD ELHADEF[2], (Member, IEEE)

[1]Laboratoire Hatem Bettahar IResCoMath, Universite de Gabés, Gabes 6072, Tunisia
[2]Abu Dhabi University, Abu Dhabi, United Arab Emirates
[3]King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia

Corresponding author: Mohamed Hadded (mohamed.elhadad@adu.ac.ae)

This work was supported by the Abu Dhabi University Office of Research and Sponsored Programs under Grant 19300788.

ABSTRACT In Intelligent Transportation Systems (ITS), ensuring road safety has paved the way for innovative advancements such as autonomous driving. These self-driving vehicles, with their variety of sensors, harness the potential to minimize human driving errors and enhance transportation efficiency via sophisticated AI modules. However, the reliability of these sensors remains challenging, especially as they can be vulnerable to anomalies resulting from adverse weather, technical issues, and cyber-attacks. Such inconsistencies can lead to imprecise or erroneous navigation decisions for autonomous vehicles that can result in fatal consequences, e.g., failure in recognizing obstacles. This survey delivers a comprehensive review of the latest research on solutions for detecting anomalies in sensor data. After laying the foundation on the workings of the connected and autonomous vehicles, we categorize anomaly detection methods into three groups: statistical, classical machine learning, and deep learning techniques. We provide a qualitative assessment of these methods to underline existing research limitations. We conclude by spotlighting key research questions to enhance the dependability of autonomous driving in forthcoming studies.

INDEX TERMS Connected vehicles, autonomous vehicles, vehicular networks, artificial intelligence, sensors, anomaly detection, outlier detection.

ABBREVIATIONS

ADS	Anomaly Detection System.
ADAS	Advanced Driver Assistance System.
AE	Auto-Encoder.
AEV	Autonomous Electric Vehicles.
AI	Artificial Intelligence.
AUC	Area Under the Curve.
AV	Autonomous Vehicles.
CAVs	Connected and Autonomous Vehicles.
CNN	Convolutional Neural Network.
DL	Deep Learning.
FDI	False Data Injection.
GNSS	Global Navigation Satellite System.
HLF	High Level Fusion.
ITS	Intelligent Transport Systems.
IoT	Internet of Things.
IMU	Inertia Measurement Unit.
KF	Kalman Filter.
LSTM	Long Short-Term Memory.
LLF	Low Level Fusion.
LiDAR	Light Detection and Ranging.
ML	Machine Learning.
MLF	Mid-Level Fusion.
MAE	Mean Absolute Error.
MSE	Mean Squared Error.
ROC	Receiver Operating Characteristic.

The associate editor coordinating the review of this manuscript and approving it for publication was Tiago Cruz.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

RNN	Recurrent Neural Networks.
SAE	Society of Automotive Engineers.
SSL	Solid-State LiDAR.
STPA − Sec	System-Theoretic Process Analysis for Security.
STRIDE	Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, Elevation of privilege.
SVM	Support Vector Machine.
TVRA	Threat Vulnerability Risk Analysis.
ToF	Time of Flight.
V2I	Vehicle to Infrastructure.
V2V	Vehicle to Vehicle.
V2X	Vehicle to Everything.

I. INTRODUCTION

During the last few years, with the increase in population and rapid urbanization, the rate of vehicles around the world has grown extremely rapidly. This increase was implicitly among the primary causes of millions of road crashes yearly, as the high density of vehicles on the roads created an unstable congested traffic environment with more complex and hazardous driving conditions. For this reason, among others, Intelligent Transportation Systems have been established as smart technological solutions to improve road safety and promote smart mobility [1]. ITS offers several benefits including smart traffic management and monitoring, enhanced safety services, user-oriented mobility services, etc. [2]. Moreover, they rely on a connected infrastructure via inter-vehicle (V2V) and infrastructure (V2I) communication links interconnecting road sensors, road users, and vehicles. This allows them to exchange real-time notifications on their statuses as well as report road conditions, which helps avoid dangerous situations and improve traffic flow management [3]. Therefore, due to these significant advancements in ITS technologies, autonomous driving systems are becoming more and more reliable to implement and, hence revolutionizing the global transportation system. They consist of a variety of components that combine both Internet of Things (IoT) and Artificial Intelligence (AI) technologies, which helps reduce dependence on human intervention [4], [5]. Thus, a self-driving system allows vehicles to navigate freely with minimum, not to say without human guidance, using sophisticated algorithms and machine learning models to interpret data collected from multiple onboard sensors and make real-time self-driving decisions [6].

The implementation of Autonomous Vehicles (AVs) paves the way for a new era of transportation, yielding many positive implications for road safety and transport efficiency [7]. They hold the potential to significantly enhance road safety by mitigating human errors, which are the primary cause of most road accidents [8]. In addition, they have the capability to streamline traffic flow and reduce congestion by employing improved coordination and more efficient driving techniques [9]. They also contribute to increased accessibility for the elderly or people with physical or visual impairments [10], [11], [12]. Furthermore, AVs could help protect the environment by reducing carbon dioxide emissions, optimizing energy consumption, and facilitating the adoption of electric vehicles [13].

Since 2010, many technology and automotive companies have invested in the development of AVs, such as Tesla, Google's Waymo, AutoX, Baidu's Apollo, BMW, and many other companies [14]. Hence, many self-driving vehicle tests are currently underway in carefully selected cities and regions. Nevertheless, despite all the efforts that have been made in this sector, there are still a number of significant obstacles that prevent the widespread use of this technology. One of the primary barriers to its wider adoption is anomalies and unanticipated happenings [15]. Currently, Connected and Autonomous Vehicles (CAVs) rely heavily on multiple sensors data for making maneuver decisions (e.g., lane changes, acceleration, deceleration, etc.) as well as navigating the environment. In recent autonomous driving applications, three types of sensors, including cameras, LiDARs, and Radars, are commonly used for sensing the surrounding environment [16], [17], [18], [19]. However, due to sensor data uncertainties caused by either cyberattacks or other external factors, such as sensor malfunctions, environmental anomalies and weather conditions, autonomous driving systems require a sophisticated design to capture those abnormalities and eventually mitigate their impacts. The presence of outliers can result from various factors such as the presence of multiple objects with different profiles, especially in urban scenarios, the low accuracy of perception sensor measurements, the loss of data after the fusion process, etc. [20]. These outliers can cause errors in the sensor data, which can lead to erroneous navigation decisions, resulting in accidents and fatalities. Therefore, it is crucial to identify and detect these outliers in the collected data as they may have a fatal impact on autonomous driving systems.

In order to mitigate the effects of these anomalies on the operation of autonomous driving systems and AVs in general, several methods have been investigated in the literature. Each one of them provides a different detection approach and deals with different types of data. Moreover, these methods are implemented at different levels of the autonomous driving system framework and present different impacts on the system's performance. In this context, this paper presents an in-depth review of the literature on recent advances in anomaly detection in sensor data for CAVs and provides a comprehensive guide for researchers and practitioners in the field.

A. RELATED SURVEYS

Anomaly detection plays an important role in maintaining the safe operation of AVs. Hence, many efforts have been made in this area to study the topic from different perspectives. Thus, a significant volume of literature has been generated over the last decade. In this section, we discuss

TABLE 1. Summary of prior existing surveys on anomaly detection solutions.

Reference	Year of Publication	Covered Years	Research issues	Qualitative Analysis	Objective	Advantages	Limitations
[24] [25]	2020 / 2021	2007-2020	No	No	Present a comprehensive categorization of corner cases encountered in visual perception of autonomous driving and the associated detection methods	Corner cases are classified according to the abstraction level in five category. A good number of detection solutions have been discussed in each level	Anomaly detection methods are limited only to camera data
[23]	2021	2016-2020	Yes	No	Present a systematic outlook of AI-based anomaly detection techniques designed for autonomous electric vehicles	Provides a detailed taxonomy classifying anomaly detection techniques considering network, security and AI-based solutions in the IoV environment	Focuses only on anomalies caused by malicious AEVs behavior
[26]	2021	2016-2021	Yes	Yes	Provide a structured overview of recent studies focused on anomaly detection solutions used for IoT data streams	Highlight several issues and challenges facing the design of an anomaly detection solution	Does not focus on anomaly detection methods for IoT-based autonomous driving
[21]	2022	2015-2022	No	Yes	Provide a detailed review of anomaly detection methods designed for autonomous driving	Detection techniques are classified according to a variety of sensor modalities including camera, radar, lidar, multimodal, and abstract object level data	Type of anomalies that methods are designed to detect (e.g. sensor malfunction, environmental anomaly, weather conditions...) are not clearly discussed
[22]	2022	2010-2021	No	No	Provide a comprehensive literature review of current methods used to identify outliers in IoT systems, including a structured taxonomy of these techniques	Study in depth the fundamentals of outlier detection in IoT with an emphasis on their sources, the different detection methods and the challenges facing their design.	Does not focus exclusively on outlier detection in IoT systems dedicated for autonomous vehicles
Our work	2023	2019-2023	Yes	Yes	Present a detailed survey of recent techniques used for anomaly detection in autonomous vehicle sensor data and detail all aspects related to anomalies: their types, causes and mitigation techniques	Present a qualitative analysis of the presented solutions based on several criteria and discuss several open research issues related to anomaly detection	

recent comprehensive surveys that review existing literature focusing on the anomaly detection field. We provide a summary and comparison of these surveys in Table 1.

The survey in [21] offers a comprehensive examination of various anomaly detection methods that rely on Radar, LiDAR, camera, multimodal, and abstract object-level data. It presents a systematic analysis that encompasses several criteria such as detection approaches, corner cases (i.e., anomalies) level and simulation datasets. Thus, detection approaches are further classified into five categories: reconstruction, prediction, generative, confidence scores, and feature extraction. Moreover, in [21], the authors discussed the latest advancements in the field and identified areas where further research is needed.

In [22], the authors tackled the issue of outliers in IoT applications by examining its different sources, existing detection approaches, how to assess detection techniques, and the difficulties encountered in designing such a solution. As a next step, the authors provided a literature review of the most recent existing outlier detection techniques by classifying them into seven categories: statistics-based, cluster-based, nearest-neighbor-based, classification-based, artificial intelligence-based, spectral decomposition-based and hybrid methods. Within each category, a set of solutions are discussed and analyzed according to several criteria such as the nature of test data and the dedicated approach (online, offline, distributed, centralized, etc.).

The survey presented in [23] investigated the anomaly detection solutions for Autonomous Electric Vehicles (AEVs) through AI-based approaches. The review fills the gaps in existing surveys through a detailed study of associated security vulnerabilities and corresponding AI methods to classify irregular behaviors. Additionally, this survey provides a detailed classification that categorizes anomaly detection techniques, considering network, security, and AI-based approaches. On the other hand, in [24], the authors addressed corner cases for visual perception in autonomous driving and categorized them into five levels based on their complexity of detection: Pixel Level, Domain

Level, Object Level, Scene Level and Scenario Level. For each level, a set of detection solutions have been highlighted. The authors pursued their investigation in [25] by covering more detection approaches in addition to their respective categories (as in [21]) and linking them to the corner case levels.

Finally, the authors of [26] focused on anomaly detection techniques based on Machine Learning (ML) and Deep Learning (DL) approaches designed to assess irregular behaviors in the IoT data stream. The authors have also presented a detailed taxonomy that defines the current literature based on various elements including the nature of the data, the types of anomalies studied, the learning modes employed, the window model, the data set and the criteria used to evaluate such a detection solution. The paper additionally proposes some future research directions that may aid in the advancement of innovative anomaly detection techniques.

B. CONTRIBUTIONS

The aforementioned surveys provide valuable insights into certain aspects of anomaly detection. Nevertheless, a comprehensive survey focusing on outlier detection in autonomous driving, encompassing diverse sensors such as cameras, Radar, and LiDAR, along with various anomaly categories and appropriate detection techniques, is currently unavailable. The main contributions of our work compared to the existing reviews are:

- We provide an overview of how connected and autonomous vehicles work, their key components and the benefits arising from the implementation of this technology.
- We present a comprehensive taxonomy that encompasses several elements related to anomaly detection, which helps better understand this research area.
- We conduct a systematic literature review on anomaly detection techniques for CAVs. We classify existing solutions into Statistical, classical Machine Learning, and Deep Learning methods, which help to discern recent trends in anomaly detection for CAVs and to identify emerging techniques within this field.
- We conduct an in-depth qualitative evaluation of the existing proposals to highlight the strengths and weaknesses of each solution. This serves as a foundation to provide insights for enhancing the security of emerging CAVs applications.
- We identify open research issues related to self-driving vehicles. These challenges form a roadmap for future research efforts by recognizing where further research is needed to develop secure solutions for autonomous driving.

C. ORGANIZATION

As shown in Fig. 1, we organize the rest of the paper as follows: We provide a background on autonomous vehicles in Section II. In Section III, we present a detailed taxonomy around anomaly detection. We discuss, in Section IV, an overview of the recent proposed solutions from the literature. Section V provides a qualitative analysis of the solutions discussed while also highlighting some open research issues. Finally, we close this investigation with a conclusion. We further provide the list of acronyms used in this survey.

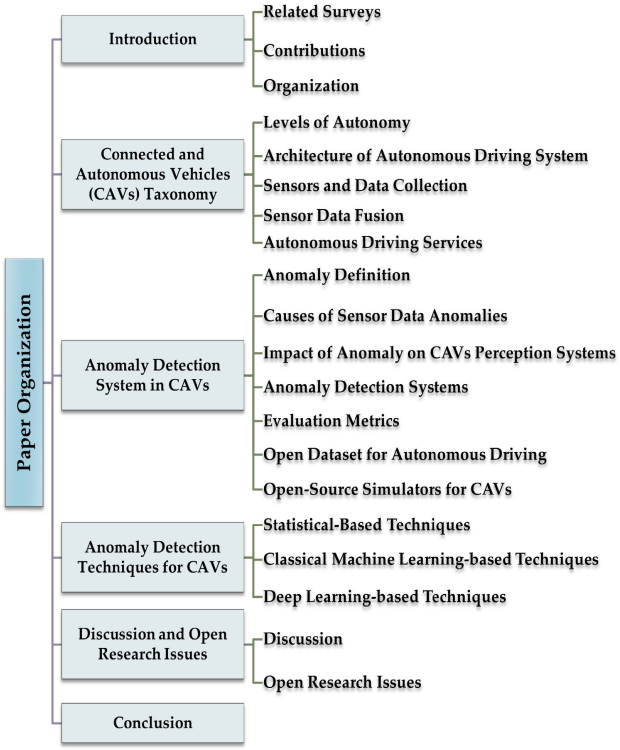

FIGURE 1. Survey structure and main sections.

II. CONNECTED AND AUTONOMOUS VEHICLES (CAVS) TAXONOMY

Prior to delving into the core of the subject, it is necessary to introduce several key concepts related to self-driving technology. In this section, we detail the six levels of autonomy of CAVs, the types of sensors included in a CAV how data are collected through them, as well as how decisions are generated.

A. LEVELS OF AUTONOMY

With the aim of providing a common basis for understanding the different levels of automation, the Society of Automotive Engineers (SAE) [27] defined the J3016 standard in 2014 [28], [29].

As shown in Fig. 2, this standard divides vehicle autonomy into six levels, ranging from zero to full automation depending on system capabilities. At Level 0, all driving tasks are fully managed by the driver (human) without any assistance or automated features. More and more, new automated features are added at Level 1 to assist the human driver. At Level 2, an Advanced Driver Assistance System

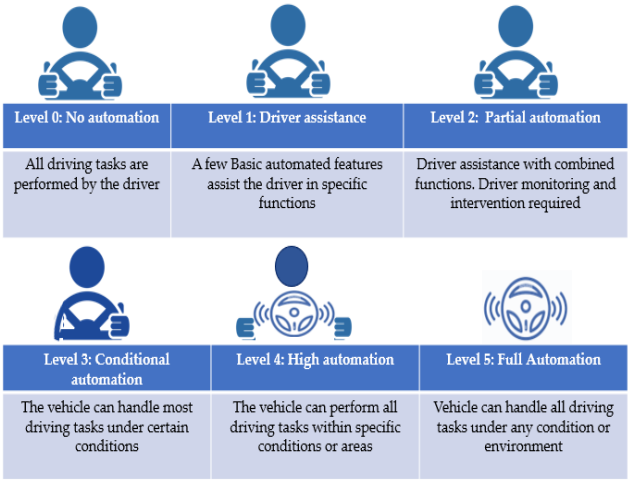

FIGURE 2. Driving automation levels.

(ADAS) is built into the vehicle, which can provide steering and braking/acceleration features. Upon reaching Level 3, the vehicle will be able to perform the majority of driving tasks as long as specific conditions are met. Furthermore, the driver is not required to constantly monitor the driving process, unless prompted by the system. At the highest degree of automation, the vehicle achieves full autonomy making it capable of independently managing all driving responsibilities in all situations and environments without requiring human intervention. Hence, the evolution of CAVs towards the full level of driving automation, indicates that futuristic vehicles will be very dependent on sensors and that navigation decisions will be dependent on the quality of the collected data.

B. ARCHITECTURE LAYERS OF CAVS

The ability of self-driving vehicles to navigate freely lies in the use of diverse embedded sensor technologies. As illustrated in Fig. 4, which represents the conventional architecture of an autonomous driving system, an AV consists of several key components that can be organized into three layers: sensing, perception, and decision layers [30]. These components operate collaboratively to allow the vehicle to sense and understand its environment, make appropriate decisions, and navigate smoothly on the roads. Thus, as a first step, the sensors take care of collecting data from the environment. These data are then processed in the second layer in order to extract relevant information such as recognizing objects, identifying obstacles, and determining their positions. The extracted information is subsequently used to generate commands. Finally, the decision layer takes responsibility for translating orders into mechanical actions such as braking, acceleration, and steering [31], [32]. These layers are summarized as follows:

- **Sensing layer:** Sensing is the crucial first step in enabling the self-driving car to understand its environment and make informed decisions. Thus, a variety of sensor types are used to collect essential data, such as camera, LiDAR, Radar, ultrasonic, etc.

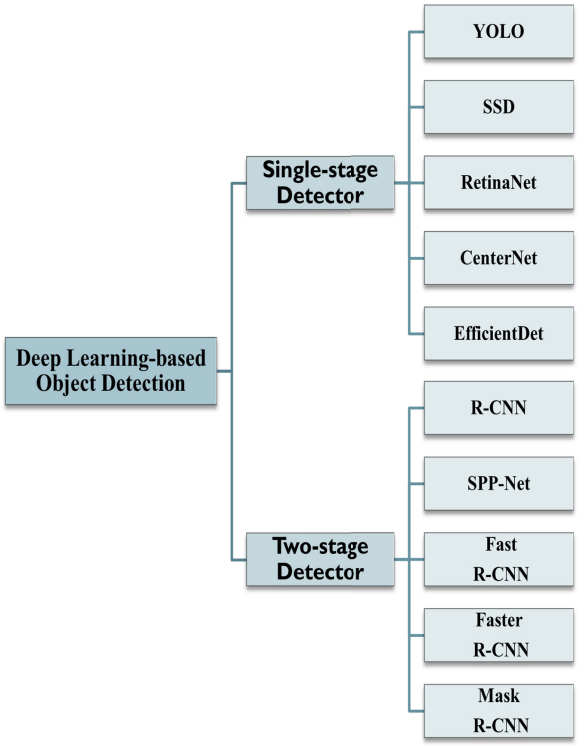

FIGURE 3. Deep learning-based object detection methods.

- **Perception layer:** This layer plays a central role in the architecture of CAVs. Thus, the raw data collected by the different sensors will be processed in this layer in order to transform them into a meaningful representation of the environment and to make decisions accordingly. This layer uses several sophisticated Machine Learning algorithms for the interpretation of the extracted information and the object detection and classification tasks. In addition, these techniques often rely on Deep Learning algorithms using neural networks. These methods are trained using extensive datasets to enhance their capability in recognizing various object classes, including pedestrians, cyclists, cars, traffic signs, and others. These algorithms are mainly divided into single-stage detectors and two-stage detectors. The main difference lies in the number of steps involved in the object detection/classification process. Single-stage detectors are generally very fast, as they perform detection and classification in a single step whereas two-stage detectors do it in two separate steps [33]. In Fig. 3, we have presented a summary of the most popular object detection algorithms. These algorithms include You Only Look Once (YOLO) [34] and its later versions (currently YOLOv8), Single Shot MultiBox Detector (SSD) [35], Region-Based Convolutional Neural Network (R-CNN) [36] and its variants Fast R-CNN [37], Faster R-CNN [38] and Mask R-CNN [39].

TABLE 2. Comparative analysis of machine learning applications for autonomous driving.

Aspect	Object Detection/Classification	Traffic Sign Recognition
Precision	Capable of detecting and classifying objects with high precision	High precision in recognizing standard traffic signs
Adaptability	Adaptability to diverse environments and object types	Adaptability to various traffic signs
Data Annotation Requirements	Depend on a large quantity of annotated data	Depend on a large quantity of annotated data
Complex Object Interactions	Challenges in handling complex scenes with multiple interacting objects	Difficulty in handling complex traffic scenarios
Small Object Recognition	Can be less accurate with small objects	Challenges in recognizing small or partially occluded signs
Lighting Conditions	Challenges in detecting poorly illuminated objects	Signs visibility can be affected by lighting variations
Weather Sensitivity	Adverse weather can impact objects visibility	Adverse weather conditions may prevent visibility of signs
Sensitivity to anomaly	Vulnerable to errors with objects of unknown classes	Can be fooled by adversary attacks (e.g. modified signs

Table 2 provides a comparison between two major ML applications used by self-driving systems to perceive their surroundings: object detection/classification and traffic sign recognition. As mentioned, object detection techniques provide a high accuracy and excel in the ability to handle diverse object categories, although they may be sensitive to small or poorly lit objects. Likewise, traffic sign recognition algorithms demonstrate high accuracy in identifying different road signs, but they can encounter certain challenges disrupting their performance. The capabilities and limitations are explored in the following table, providing insight into the challenges unique to these two key applications.

- **Decision layer:** This layer is responsible for making decisions based on data received from the perception layer. It serves as the core of the autonomous driving system, as it determines how the vehicle must react to changing conditions in its surroundings. Thus, the obtained data is analyzed to understand the current state of the vehicle's environment. This can include identifying nearby vehicles and predicting their movements, understanding traffic signals, detecting obstacles and other road users (pedestrians, cyclists, animals, etc.). Using this information, the decision layer generates actions including commands for steering control, speed and braking. Once the decision is made, a trajectory is generated to plan the vehicle's route while considering driving rules, detected obstacles and anticipated movements of other vehicles.

C. SENSORS AND DATA COLLECTION

The integrated sensors in a CAV can be classified into two groups, proprioceptive and exteroceptive, depending on whether they measure the internal state (e.g., Global Navigation Satellite System (GNSS) or Inertia Measurement Unit (IMU)) of a vehicle system or collect data from the external environment (e.g., camera, LiDAR, Radar and ultrasonic sensors). They can also be classified as passive or active sensors depending on whether they depend on the energy emitted by the environment (e.g., camera) or whether they emit energy themselves to gather information (e.g., LiDAR and Radar) [40]. These various sensors, characterized by different functionalities, collaborate harmoniously to create a holistic perception system for self-driving vehicles. By combining data from these sensors, the vehicle can examine and comprehend its environment, thereby facilitating secure and efficient autonomous navigation. In the following, we elaborate on the functioning of sensors and the process of data collection through them. Our primary focus will be on camera, LiDAR, Radar, and ultrasonic sensors.

1) MAIN SENSORS

- **Camera:** This type of sensor is essential for autonomous vehicles as it can allow them to perceive their environment by detecting both stationary and moving objects with different profiles. Thus, cameras offer a significant advantage over other types of sensors because they can differentiate between colors and textures. In addition, the installation of several cameras in different positions around the vehicle provides a 360° view as well as a bird's eye view allowing the vehicle to identify nearby objects, such as neighboring vehicles, pedestrians, road lines, and traffic signs. Currently, modern high-definition cameras have the capability to generate millions of pixels in each frame, achieving a frame rate ranging from 30 to 60 frames per second [41]. Generally, cameras can be divided into two categories: monocular and binocular. Monocular cameras can detect objects and capture two-dimensional images of the environment close to the vehicle (short to medium-range perception). On the other hand, binocular cameras can provide a three-dimensional (3D) representation of the scene that simulates human eyes [42]. They are more suitable for medium to long range perception allowing for more accurate depth perception compared to monocular cameras.

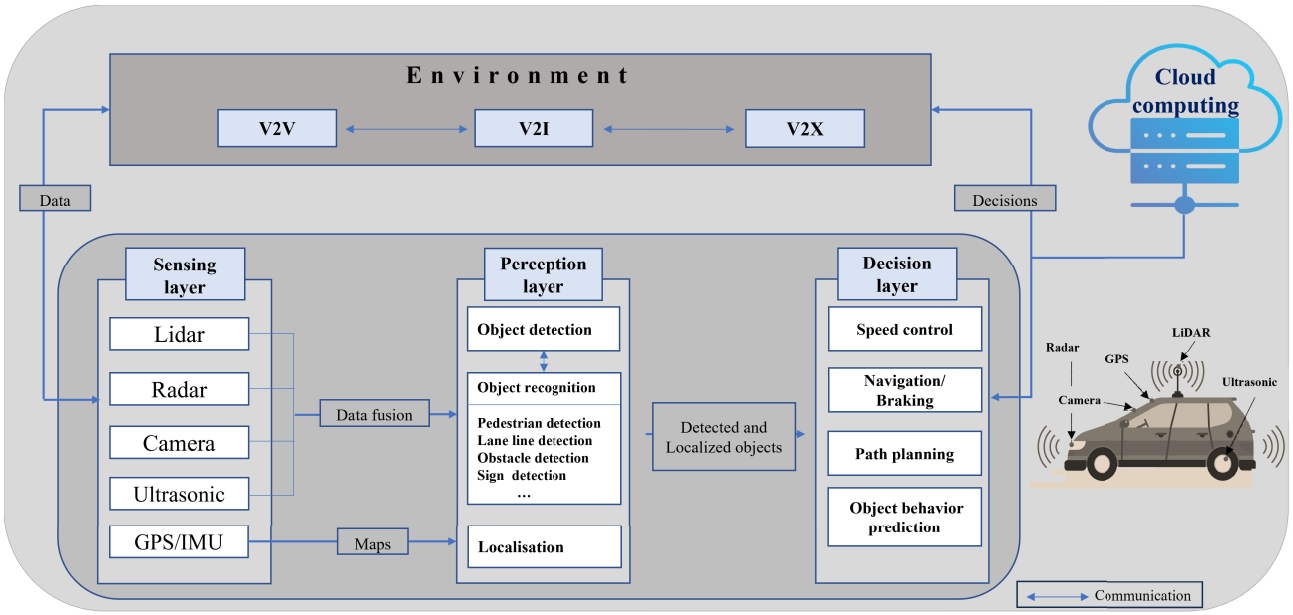

FIGURE 4. Conventional architecture of an automated driving system.

- **LiDAR:** This is a remote sensing system that employs laser pulses to precisely determine the distance of objects. When applied to CAVs, these particular sensors are crucial as they enable a more comprehensive and efficient understanding of the surroundings, surpassing the capabilities of cameras. LiDAR devices can be classified as medium to long-range sensors, offering a measurement range of over 200 meters [43]. A LiDAR sensor is mainly composed of an infrared transmitter and a receiver for the reflected signals by the objects that contain a timer calculating the time elapsed between the emission and the reception of laser (*Time of Flight (ToF)*) [44]. To determine the distance of an object, initially, the sensor emits laser pulses in various directions. These beams come into contact with objects along their path, causing them to bounce back. Then, the receiving device detects these reflected signals and computes the distance by measuring the time it takes for the signal to travel. This calculation is performed for several points which creates a 3D representation of the vehicle environment's geometry called *point cloud* [45]. LiDAR sensors can be classified into two main types: mechanical LiDAR and Solid-State LiDAR (SSL) (see [40] for more details). The main difference between these two categories lies in the technique employed to direct the laser beams. The first one uses a rotating mechanical mirror while SSLs utilize semiconductors, such as SSL diodes, to generate and emit laser beams.
- **Radar:** The term *Radar* stands for Radio Detection and Ranging. It refers to a remote detection system using radio waves to locate, detect, and track objects. The operation of a radar sensor is based on the principles of broadcasting radio waves and analyzing their reflection on objects in order to gather valuable data such as the object's distance, its speed, its direction, and more. Autonomous vehicles rely on radars as one of their initial sensor options due to their ability to withstand various weather conditions. Similar to a LiDAR system, the radar antenna sends radio waves, in the form of electromagnetic signals at a given frequency. Once these waves have encountered a solid object, some of the wave's energy will be reflected back to the radar (the reflected signal strength depends on the size of the object), which will process the signal into a usable form. Calculating the speed and position of an object is based on the Doppler property of EM waves [46]. Thus, radar sensors employ various frequency bands for different purposes. This includes 24 GHz, 76 GHz, 77 GHz, and 79 GHz where the higher frequencies, such as those in the 77 GHz and 79 GHz ranges, provide higher resolution, enabling better real-time differentiation of objects [47].
- **Ultrasonic Sensors:** These measurement devices operate without physical contact and utilize high-frequency sound waves to determine the distance to a surrounding object. In the context of CAVs, these types of sensors are often employed for the detection of objects close to the vehicle. Hence, in the same way as a LiDAR or Radar sensors, the operation of an ultrasonic sensor is based on the emission of sound waves and the measurement of the duration it takes for these waves to bounce off objects and return to the sensor [48]. By using the speed of sound in air, which depends on some factors like the

TABLE 3. Comparison of sensors' characteristics.

Criteria	Camera	LiDAR	Radar	Ultrasonic
Perception technology	Light	Laser pulses	Radio waves	Sound waves
Range	200 m	200 m	250 m	10 m
Resolution	High	High	Med.	Low
Cost	Low	High	Med.	Low
Size	Small	Med.-Large	Small	Small
Object detection	Low	High	High	High
Object classification	High	Med.	Low	Low
Speed estimation	No	Yes	Yes	No
Color detection	Yes	No	No	No
Distance estimation	No	Yes	Yes	Yes
Resistance to bad weather	Low	Low	High	High
Low light performance	Low	High	High	High

temperature and the humidity of the air (e.g., at 20°C the speed of sound is equal to about 343 m/s), the sensor can calculate the distance between itself and an object. Moreover, ultrasonic sensors rely on sonic transducers to transmit sound waves falling between 40 kHz to 70 kHz for automotive applications. This range of frequencies exceeds the range of human hearing making it safer for use [44]. These sensors are generally useful for the detection of objects at short distances (approximately 10 meters), enabling tasks like parking maneuvers and low-speed obstacle avoidance [49].

2) SENSORS' CHARACTERISTICS COMPARISON

In this section, we summarize the characteristics previously discussed for each sensor (camera, LiDAR, Radar, and ultrasonic sensors). In Table 3, we provide a comparative analysis of these technologies. This analysis encompasses various criteria including the perception technology employed, range, sensitivity to weather conditions, cost, and other relevant aspects. Our evaluation draws from multiple studies, including [18], [44], [46], [50], [51].

D. SENSOR DATA FUSION

The data fusion refers to the process of combining data from various sensors. The objective of this stage is to achieve a more precise and complete representation of the vehicle's environment, in order to make smarter decisions and improve the overall performance of the system. Indeed, each sensor possesses its own strengths and weaknesses regarding to resolution, precision, range, and so on, data fusion aims to take the benefits of each data source while mitigating their limitations. Generally, to enhance the perception capabilities of self-driving vehicles, three possible sensor combinations are adopted for sensors fusion:

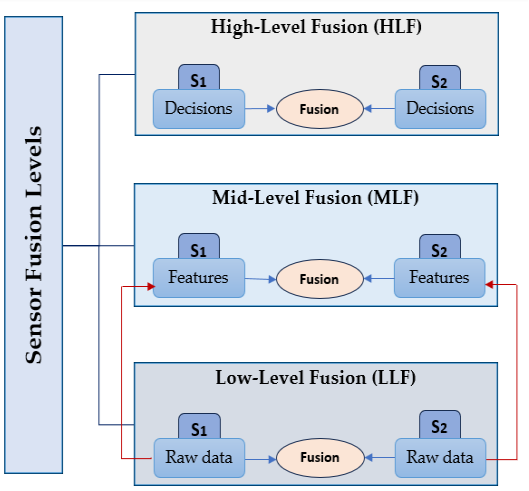

FIGURE 5. Classification of sensor fusion techniques according to the level of abstraction of fused data.

Camera-Radar Camera-LiDAR, and Camera-LiDAR-Radar [52]. In fact, camera sensors can provide detailed appearance data on objects, however, they are sensitive to some lighting conditions which can cause data anomalies. To address this limitation, combining camera images with Radar or LiDAR outputs provides more details about the positions of objects and their speeds with the possibility of tracking them.

As shown in Fig. 5, to combine information from two or more types of sensors, three approaches can be adapted: High-Level Fusion (HLF), Low-Level Fusion (LLF) and Mid-Level Fusion (MLF) [53]. By using HLF methods, each sensor interprets its data independently and the data fusion is done using the results of each one of them. These techniques have the advantage of lower complexity compared to other techniques, however, they may generate inadequate or insufficient information resulting in data anomalies. On the other hand, LLF approaches aim to merge the raw data from each sensor together without prior filtration with the objective of improving the process of object recognition. In practical implementation, these methods necessitate intricate sensor pre-configuration to achieve a clear perception [40]. Likewise, MLF approaches are techniques that combine HLF and LLF strategies. Their objectives are to derive features from the initial readings obtained from each sensor and merge them in a subsequent stage to generate a combined signal that can be used for further analysis [54].

E. AUTONOMOUS DRIVING SERVICES

Autonomous vehicles can provide several benefits and services. These advantages can be divided as follows:

1) DIRECT BENEFITS

- **Comfortable and safe driving:** By eliminating more and more human involvement in the driving process, CAVs can significantly reduce the risk of accidents and make decisions in an efficient way. Thus, with the

use of a combination of multiple types of sensors and intelligent algorithms, CAVs can anticipate potential collisions and take preventive measures more quickly than a human driver such as braking or changing lanes [55]. Moreover, by means of V2V and V2I communications, vehicles are able to collaborate and make decisions based on a more complete understanding of their environment. As a result, this facilitates a decrease in both the frequency of accidents and traffic congestion on the roadways for around 90% using full autonomous vehicles [56].

- **Accessible driving to everyone:** Autonomous vehicles can make driving accessible to everyone by reducing mobility constraints for people with special needs. In addition, the use of self-driving vehicles eliminates the need for advanced driving skills and people who have not learned to drive or who still have difficulty driving can benefit from transport services.
- **Optimized driving time:** As a result of applying the advanced features provided by self-driving vehicles, driving time can be minimized and managed in an efficient way. Thus, CAVs are coded to use enhanced driving methods in comparison to those used by humans [57]. These improvements result in better decision-making regarding route selection, faster navigation, and reduced time spent on parking.

2) INDIRECT BENEFITS

- **Reduced air pollution:** In the higher levels of automation, particularly levels 4 and 5, driving will be smoother and more efficient so that fuel consumption and carbon dioxide emissions will be reduced remarkably [58]. As well as, by minimizing waiting times and improving traffic flow, the release of polluting gases caused by traffic congestion and frequent stopping and starting of vehicles will be considerably reduced [59], [60].
- **Moderate energy consumption:** By using self-driving systems, energy consumption is expected to be more moderate compared to traditional driving practices. Since CAVs allow to reduce the time spent on the road, the energy consumption necessary to cover a given distance will be decreased as a result of this. Thus, a number of studies, reported in [9] and [61], have shown that the application of CAVs can save up to 40% of fuel.
- **Less expenses for maintenance:** A very important advantage results from the use of autonomous driving systems which is the reduction of maintenance cost. Since CAVs are equipped with advanced technologies, they can quickly detect any problem or potential malfunction. Furthermore, by avoiding collisions and road damage, maintenance costs will be significantly reduced.

III. ANOMALY DETECTION SYSTEMS IN CAVS

Anomaly detection is an essential task to guarantee the safety of autonomous vehicles and the certainty of their decisions. Above all, it is also important to understand what an anomaly is, its forms/types and the potential causes that can increase the risks of producing erroneous readings. To clearly illustrate the Anomaly Detection System (ADS) in CAVs, we present, in Fig. 6, a detailed taxonomy which highlights several important elements, mainly: the types of anomalies and their sources, the different categories of techniques used for detection and the diversity of datasets used for evaluation. We present, in more detail, all of these elements in the remainder of this section.

A. ANOMALY DEFINITION

Explaining an anomaly is somewhat challenging as there is no precise definition for it. Nevertheless, various works have suggested approximate explanations for it, such as in [62] and [63]. In general, an anomaly, commonly referred to as an *outlier* or *corner case*, occurs when a measurement or reading significantly diverges from the typical values generated by a sensor. In simpler terms, it represents data that deviate from the rule and shows unexpected behavior compared to what the sensor usually produces.

As shown in Fig. 6, there are typically three categories of outliers that can occur: point anomaly, contextual anomaly, and collective anomaly. Point anomalies are the easiest to identify and refer to a reading that significantly differs from the rest of the captured data. An instance of this anomaly could be a false detection of an object's distance caused by a temporary radar failure. Contextual anomalies, on the other hand, depend on the context of perception. They might be considered normal in one context but abnormal in another. For example, if there is a sensor malfunction, the distance measurement, which is considered normal at a given speed X, turns into an anomalous value when moving to a different speed Y. Finally, collective anomalies represent a group of observations that diverge together from the expected values, e.g., when several autonomous vehicles simultaneously encounter and report unusual situations such as detecting an unknown object on the road that does not conform to the typical objects.

B. CAUSES OF SENSOR DATA ANOMALIES

Several factors can cause anomalies in the data collected by sensors. However, they can be categorized into three classes: sensor malfunction, measurement errors due to changing environmental conditions, and security threats. We categorize them as follows:

- **Sensor malfunction:** IoT devices are highly susceptible to hardware and software errors. These faults are caused by several sources, including sudden sensor failure, damage to internal components due to prolonged sensor usage, wrong calibration, software errors, or connectivity problems. As a result, these factors can contribute to disrupting the proper functioning of the sensor by producing low-precision readings [64], [65].

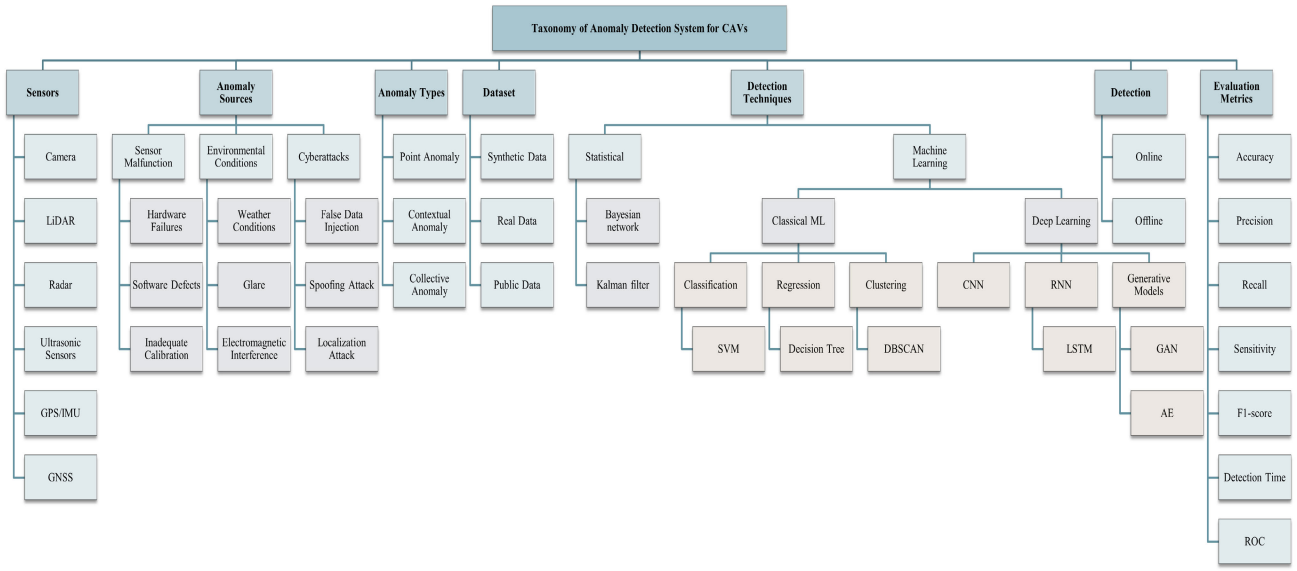

FIGURE 6. Taxonomy of anomaly detection system in CAVs.

- **Environmental conditions:** One of the major challenges facing CAVs is the changing environmental conditions. Thus, the performance of the sensors is significantly affected by climate conditions such as temperature, fog, rain, humidity, strong light, etc. For example, when it rains, water droplets can interfere with the laser beam trajectory of a LiDAR sensor, causing unwanted reflections as a result. LiDAR can interpret these reflections as additional obstacles leading to inaccurate detection of the number of obstacles [49], [50].
- **Cyberattacks:** The network of CAVs is very exposed to several cyberattacks which can have serious impacts for the safety of CAVs. Thus, as sensors are fundamental elements for autonomous driving, attackers can manipulate the various captured data, misleading the CAVs decision systems [66]. These types of attacks, such as jamming and spoofing attacks, primarily target the availability and integrity of sensor data [67], [68]. In addition, the control components of self-driving vehicles can also be targeted by attacks aimed at disrupting the stability of their trajectories or taking full control of them remotely. Furthermore, since CAVs rely on cloud services to primarily manage data storage and software updates, attackers can affect the availability of these cloud infrastructures by launching attacks like denial of services (DoS).

C. IMPACT OF ANOMALIES ON CAVS' OPERATION

As mentioned previously, several factors such as enrionmental conditions and cybersecurity threats, can drastically affect and disrupt the safe operation of CAVs, including their perception systems. These disruptions can make decision systems less precise and harm the safety of road users [49], [69]. In addition, frequent perceptual errors can lead to loss of trust in autonomous driving systems. Thus, it is necessary to implement mechanisms for identifying anomalies and preventing their adverse effects on the security and performance of CAVs. In the following, we summarize the potential impacts of anomalies on self-driving vehicles perception systems.

- **Incorrect interpretation of the environment:** the presence of anomalies of various types can lead to errors in the readings provided by the autonomous vehicle's sensors, resulting in an incomplete or erroneous perception of the surrounding. These errors mainly include misclassification of objects and obstacles.
- **Inaccurate decisions:** Once the data received from the perception layer is incomplete or incorrect, the decision system will generate erroneous or inadequate actions. For instance, it could incorrectly anticipate the movements of other vehicles on the road or react incorrectly to the presence of an unforeseen obstacle.
- **Navigation Issues:** As a result of the presence of anomalies, self-driving vehicles may have trouble following the lane and adapting to changing complex traffic conditions and understanding signs.
- **Increased risk of collision:** the inaccurate decisions generated can increase the risk of collisions, as the CAVs cannot correctly perceive their environments. This can be of particular concern in congested traffic situations or challenging weather conditions.

D. ANOMALY DETECTION SYSTEMS

Since autonomous vehicles rely on sensory data and AI-based algorithms to make decisions and navigate their environment,

it is very likely that some anomalies will be produced. These anomalies can have serious consequences for the safety of CAVs [70]. Hence, it is mandatory to implement ADS capable of mitigating the negative impacts that anomalies can cause on the navigation decisions of the CAVs. An ADS for CAVs is a collection of mechanisms/algorithms that makes it possible to identify, isolate, and prevent any deviation from the normal state of the CAV system towards an abnormal situation due to several causes discussed previously. The ADS is characterized by several tasks, primarily monitoring the system state and collecting data against anomalies using advanced algorithms. As soon as anything abnormal is detected, the ADS can take safety measures by generating instant notifications to the vehicle's control system, which in turn, for example, will isolate the affected sensor.

In general, an ADS is composed of several modules and stages and its operation differs depending on the types of learning, i.e., supervised, unsupervised, and semi-supervised. In supervised learning mode, the ADS goes through two phases: the training phase, using examples of labeled samples (normal and abnormal) and the effective online detection. In the first phase, the ADS model is trained on a base model representing the normal and abnormal behavior of the data. Then, in a second phase, the ADS uses the base model as a reference to compare the data in real time, using a detection/classification algorithm, in order to identify anomalies. Additionally, the raw data streams from the different sensor types are passed to a pre-processing stage for normalization before being used [71], [72]. Detection models under this learning mode provide high performance with reduced false positive rates, however they are not reliable against unexpected or rare anomalies not presented in the training data. On the other hand, in unsupervised learning, the algorithm does not require labeled data for training, and it is based, essentially, on the available data to identify and learn regular behaviors and through which it can detect any deviations. These types of models have the advantage of detecting a variety of anomalies including those that are not present in the training data, unlike supervised learning. However, these models suffer from a higher number of false positives/negatives, since they do not have an annotated schema to differentiate clearly between normal and abnormal data [73], [74]. In semi-supervised learning, a combination of labeled and unlabeled data is used to take the advantages of two previous modes (supervised and unsupervised). Labeled data, in this mode, is used to train the model to identify anomalies in the unannotated data [75]. Nevertheless, these models are generally more complex in their implementations, as they handle both varieties of data types.

Several algorithms and methods have been introduced in the literature for anomaly detection. In general, these methods can be classified into algorithms that are either based on Statistical approaches or ML approaches. As for the statistical approaches, they are based on mathematical models, and they are designed to handle sensor uncertainties and errors reliably due to their strong theoretical basis. However, they may have certain limitations in terms of processing non-linear or complex data. In contrast, classical ML and DL approaches refer to artificial intelligence models that use advanced algorithms to tackle complex tasks [76], [77].

In Fig. 6, we have presented a summary of anomaly detection approaches. For each category, we have included an example of the most frequently used algorithms. For statistical techniques, we can cite the Bayesian network, Kalman Filter (KF) [78], etc. As for classical ML models, supervised and unsupervised ML are very commonly used for anomaly detection. For example, clustering algorithms like Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [79], regression models like Decision Tree (DT) [80], and classification models like Support Vector Machine (SVM) [81]. As for methods based on DL, they include Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), such as long short-term memory (LSTM) and Generative models like Generative Adversarial Networks (GANs) [82], [83].

E. EVALUATION METRICS

Evaluation metrics help improve the reliability and safety of driverless vehicles by ensuring high-quality solutions in preventing potential incidents and detecting anomalies. Thus, several metrics are used to measure the technique's performance and they are varied depending on the type of approach used (i.e., ML, DL or Statistical). These metrics mainly include Accuracy, Precision, Recall, F1-score, Mean Squared Error (MSE), ROC (Receiver Operating Characteristic) curve [84], Area Under the Curve (AUC) and more [85]. Accuracy measures the rate of correct results among all predictions. It explains a model's ability to correctly differentiate between anomalies and normal data. Precision aims to assess the number of normal data correctly identified among positive predictions. On the other hand, Recall measures the number of anomalies correctly detected among all real anomalies. F1-score combines Precision and Recall by giving an overall measure of the detection efficiency [86]. The use of these metrics is mainly linked to the type of task adopted (i.e., Regression, Classification or Clustering). Thus, for classification, algorithms are evaluated using Accuracy, Precision, Recall, F1-score, ROC or AUC. As for Regression tasks, we can use the MSE metric, which calculates the difference between the predictions and the true values, and the Mean Absolute Error (MAE) which calculates the average of the absolute deviations between the predictions and the correct values.

F. OPEN DATASETS FOR CAVs

To reliably assess an anomaly detection solution, it is crucial to choose a diverse and representative dataset including a wide range of scenarios. This allows for a comprehensive examination of the solution's performance and limitations.

In this subsection, we present some of the most well-known open datasets.

Generally, three categories of databases are widely employed for evaluating anomaly detection methods: synthetic, public, and real-world databases [87]. Synthetic datasets are artificially generated by injecting various scenarios of anomalies using a simulation environment. In contrast, real-world databases are composed of real data gathered from the onboard sensors of autonomous vehicles while navigating. Many of these datasets either synthetic or real-world are made public and can be used by practitioners to evaluate their anomaly detection methods in CAVs. The open datasets have a variety of benefits for autonomous vehicle research and development. They provide a large amount of data with diverse and realistic scenarios. In addition, the process of collecting real-world data for AVs is generally expensive and complex, and therefore public datasets offer a cost-effective alternative for researchers [88]. Since public datasets provide a large volume of real data coming from several types of sensors, anomaly detection models are trained on a large amount of normal data, which makes the detection of abnormal situations, therefore, more efficient and accurate. Additionally, the variety of scenarios and conditions makes it possible to create more complex and unusual scenarios.

As shown in Table 4, a variety of datasets are available for use and testing, among which we find datasets for 2D annotations for RGB (Red-Green-Blue) cameras and multimodal datasets containing several types of sensors. Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) dataset launched in 2012, is one of the first widely used public multimodal datasets whose task of interest includes optical flow, stereo, visual odometry and 3D object detection, and tracking. Since then, research in this field has been intensively increased and many open datasets have been released, for example, Apolloscape in 2018, Open Waymo, A2D2 (Audi Autonomous Driving Dataset) and nuScenes in 2019. On the other hand, there exist other datasets that are dedicated mainly to semantic segmentation, which aims to identify objects and obstacles based on images collected through camera sensors by assigning a class label (pedestrian, car, tree, etc.) to each pixel of the image. These datasets provide a rich base of thousands of captured real-world scenes that are ready to use in various tasks such as anomaly detection. Cityscapes and BDD100K (Berkeley DeepDrive 100K) datasets are two examples of databases used to gain a semantic understanding of complex urban scenes. In Table 4, we present and compare nine public datasets according to several criteria such as the types of sensors used for data collection, the number of classes, and the studied environment.

G. OPEN-SOURCE SIMULATORS FOR CAVs

Dedicated simulators for autonomous vehicles play an important role in testing and validating AV-related solutions. They provide virtual test environments that are very close to reality. These environments are flexible and they allow researchers to experiment and evaluate the performance of their algorithms on complex scenarios and improve them before real deployment [98], [99]. Thus, they make it possible to reduce the costs related to developments and tests compared to those in real conditions. In addition, these environments enable the study of autonomous vehicles' behavior when faced with unusual scenarios and situations, i.e. in the presence of anomalies, by providing means to integrate abnormal behavior, environmental disturbances or hardware failures.

Currently, there are several simulators for autonomous driving, the majority of which are open-source. For instance, the ''Car Learning to Act'' tool known as CARLA is an open-source simulator launched in 2017 by Computer Vision Center (CVC) and Intel Labs, which is widely used and known by its detailed and realistic environments [100]. Similarly, the LGSVL simulator is a free tool developed and launched by LG Electronics in 2018, which provides a virtual test environment for autonomous driving. It offers an organized architecture allowing to simulate a variety of sensor types [101]. Also, Gazebo [102] is a well-known simulator launched in the early 2000s. It is characterized by its versatility as it allows to simulate multiple types of robots in addition to autonomous systems. However, it does not have an integrated support allowing to simulate weather conditions contrary to the other simulators dedicated to CAVs. As for Apollo, developed by the company Baidu in 2016, is a simulator dedicated mainly to autonomous driving. It provides functionality for creating and testing complex driving scenarios in a realistic 3D environment [103]. AirSim is a Microsoft product launched in 2017, which offers diversified scenarios and it also offers the ability to generate synthetic data [104].

IV. ANOMALY DETECTION TECHNIQUES FOR CAVs

As we mentioned previously, anomaly detection techniques fall into three main groups: purely statistical techniques, techniques based on classical Machine Learning, and others based on deep artificial neural networks. In this section, we discuss the advances in each category and investigate the most recent proposed solutions in the literature.

A. STATISTICAL-BASED TECHNIQUES

Anomaly detection techniques based on statistical approaches seek to identify observations that exhibit behaviors diverging remarkably from the normal data distribution. In this section, we discuss some statistical anomaly detection methods used in the context of CAVs.

In [105], the authors have sought to increase the resilience of autonomous vehicles in the face of sensor faults and adverse attacks. To this end, they have proposed a detection technique exploiting the redundancy of the data coming from several sensors that measure the same physical variable (e.g., distance calculation). Thus, the redundant data is

TABLE 4. Comparison of selected open datasets for autonomous driving.

Ref.	Dataset	Year	Source	Sensors	Number of Classes	Last Update	Environment
[89]	KITTI	2012	Karlsruhe Institute of Technology and Toyota Technological Institute	4xCamera, 1xLiDAR	11	2015	Mid-size city of Karlsruhe
[90]	CityScapes	2016	Daimler AG, TU Darmstadt, MPI Informatics, TU Dresden	Camera	30	2020	50 cities
[91]	BDD 100K	2017	Berkeley Artificial Intelligence Research Lab (BAIR) and DeepDrive Industry Consortium	Camera	19	2020	New York, San Francisco
[92]	ApolloScape	2018	Baidu Apollo	2xCamera, 2xLiDAR	28	2020	China
[93]	Open Waymo	2019	Waymo LLC, Google LLC	5xCamera, 5xLiDAR	23	2023	USA
[94]	Lyft Level5	2019	Lyft Inc.	7xCamera, 3xLiDAR	9	2021	Palo Alto, California
[95]	nuScenes	2019	nuTonomy	6xCamera, 1xLiDAR, 5xRadar	23	2021	Boston, Singapore
[96]	A2D2	2019	Audi Electronics Venture GmbH	6xCamera, 5xLiDAR	14	2020	Germany
[97]	Argoverse	2019	Argo AI	9xCamera, 2xLiDAR	15	2021	Pittsburgh, Miami

used to feed a sensor fusion algorithm to estimate the correct information in the presence of attacks. Subsequently, detection and isolation of corrupted sensors are performed based on the estimation results. In this work, the authors have also designed an H∞ controller dedicated to CAVs with an integrated Cooperative Adaptive Cruise Control (CACC). The role of this controller is to stabilize the closed-loop dynamics of each CAV in a platoon.

In [106], Wang et al. have proposed a lightweight anomaly detection method for CAVs based on a nonlinear car tracking motion model minimizing false positives/negatives. The proposed technique considers both data from embedded sensors and data received via V2Vs and V2I communications to improve detection performance. In this work, the authors have addressed the problem of potential delay in the communication channel when using information from the lead vehicle, which can make the use of conventional fault detection methods, such as the X^2 detector, is not suitable. In order to overcome this challenge, they implemented an *Enhanced Extended Kalman Filter (AEKF)* that takes into consideration environmental information about the trajectory of the lead vehicle to optimize detection accuracy. To integrate this information, the authors have used the Intelligent Driver Model (IDM) car tracking model [107]. In parallel with the AEKF, they applied a classic X^2 detector whose role is to identify several varieties of anomalies mainly: short anomaly, noise, bias, gradual drift and miss.

Collective Awareness (CA) in intelligent agent networks within CAVs was the focus of the study presented in [108]. Kanapram et al. have proposed an approach to establish an initial level of CA. Thus, they considered a specific functionality of collective self-awareness named agent-centred detection of abnormal situations occurring in the environment. In this approach, the authors have used for the detection and prevention of anomalies, *Dynamic Bayesian network (DBN) models* which take into account time series of sensor data collected during sensing. Each DBN is linked to an agent in the network, which allows all agents to be informed of potential anomalies occurring. To train node variables and conditional probabilities linking nodes in DBN models, the authors relied on a Growing Neural Gas (GNG) algorithm. Therefore, each agent will have a model representing the normal behaviors of all agents in the same network. In addition, each agent uses, for state estimation and anomaly detection, a Markov jump particle filter (MJPF). To simulate this solution, the authors have used a dataset collected from AVs in a real environment.

Mori et al. [109], have emphasized the complexity of fault detection and isolation for sensor systems in AVs. To overcome this problem, the authors have presented a novel strategy for detecting and isolating defects using a *Student's t-distribution based adaptive unscented Kalman filter* (T-AUKF). This filter is used to evaluate the behavior of each sensor with the T^2 Hotelling test based on the predicted output of the sensor and its covariance. This method allows precise detection of faults/anomalies through the evaluation of the correlation between the data generated in the same sensor. Moreover, with the identification of the covariance and degree of freedom of the robust Student's t-distribution of outliers, the measurement noise will be updated adaptively. To validate their solution in terms of location accuracy and measurement noise estimation, the authors used the CarSim simulator and an experiment on a highway scenario.

B. CLASSICAL MACHINE LEARNING-BASED TECHNIQUES

The use of traditional ML-based anomaly detection techniques can offer a variety of advantages making them still usable even in the era of deep learning. In fact, traditional

TABLE 5. Summary of selected statistical-based techniques.

Ref.	Year	Technique	Dataset	Source Sensors	Dedicated For	Simulated Anomalies	Evaluation Criteria
[105]	2021	Secure Sensor Fusion Framework	N/A	Multiple sensor types	Cyber attacks Sensor faults	3 spoofing attacks examples	Distance estimation, Detection and isolation rate, Tracking error at startup, Measured Velocity Response at Startup
[106]	2020	Augmented Extended Kalman Filter	SPMD dataset	Multiple sensor types	Cyber attacks Sensor faults	Short anomaly, Noise, Bias, Gradual drift, Miss	AUC value, Mean Squared Error
[108]	2020	Dynamic Bayesian Network	Real data sets	Multiple sensor types	Cyber attacks Sensor faults	N/A	Accuracy, ROC Area under the curve(AUC)
[109]	2019	Fault detection and isolation (FDI) for sensor systems	CarSim simulator, Experiment on a highway scenario	GNSS	Sensor faults	Bias, Noise, Fault data	Localization accuracy, Measurement noise estimation

ML techniques are robust against outliers/anomalies when the data size is small or medium. In this section, we present some proposed solutions and summarize the most representative ones in Table 6.

Han et al., in [110], have proposed a novel collaborative approach for protecting autonomous driving systems with lifelong anomaly detection. This approach aims to protect CAVs against time series anomalies, i.e., GPS spoofing threat, and adversarial image examples, primarily road sign and lane recognition attacks. The proposed method workflow is divided into two stages: offline training and online prediction. During the first stage, the one-class model is trained to learn normal data (labeled as benign) collected from AVs. As a result, the model will be able to predict outliers (labeled as malicious) that deviate significantly from normal samples. During online prediction, the perception and localization modules are monitored by the anomaly detector module. If a suspicious event is reported, the control module will be notified to take mitigating actions.

In [111], the authors have presented a new observer-based method to improve the security of AVs against sensor faults and false injection attacks. The proposed framework combines a signal filtering model, using an Adaptive Extended Kalman filter (AEKF), and a detection and recovery method based on One-Class Support Vector Machine (OCSVM) models. At each instant, the AEKF generates the innovation value, which measures the deviation between the readings, coming from sensors, and the prediction, and then sends it to the fault detector module for anomaly detection. The detection model is composed of several OCSVM models and it can dynamically choose which one to use depending on the average innovation. To incorporate anomalous behavior, the authors have randomly injected anomalies into the normal trajectory data.

The authors of [112] have addressed the security issues caused by GPS spoofing attacks facing the CV/AV localization system. In this work, after collecting a sufficient number of historical trajectories as a demonstration, maximum entropy inverse reinforcement learning will be adopted to derive the optimal driving policy that will be used to generate a predicted optimal trajectory. In addition, a statistical method is developed to compare the optimal trajectory with the observed one. Finally, a decision tree classifier is adopted to differentiate between normal trajectories and attacked trajectories. To evaluate the performance of the proposed technique, two attack patterns are simulated. The first model aimed to generate lateral deviations from the original AV trajectory. On the other hand, the second model aims to attack and disrupt the operation of Basic Safety Messages (BSMs) in CVs.

Liu et al. [113], have discussed the effect of Perception Error Attacks (PEAs) on AV functioning and proposed a detection technique called LIFE. This technique relies on the fusion of point cloud LiDAR and camera images to detect PEAs and determine which sensor is under attack. The operation of LIFE is mainly divided into two stages: verification of consistency between LiDAR and camera data and the evaluation of the sensor reliability. To check the consistency, the 3D LiDAR points are extracted and clustered using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, each cluster of which represents a detected object. Then, each group will be projected on 2D image by calculating their positions. Finally, the positions of detected objects on the images will be compared with those calculated from LiDAR. Once an inconsistency is detected, LIFE can detect which sensor is attacked.

In [114], the authors have presented a new solution dedicated to radar-type sensors aimed at identifying moving

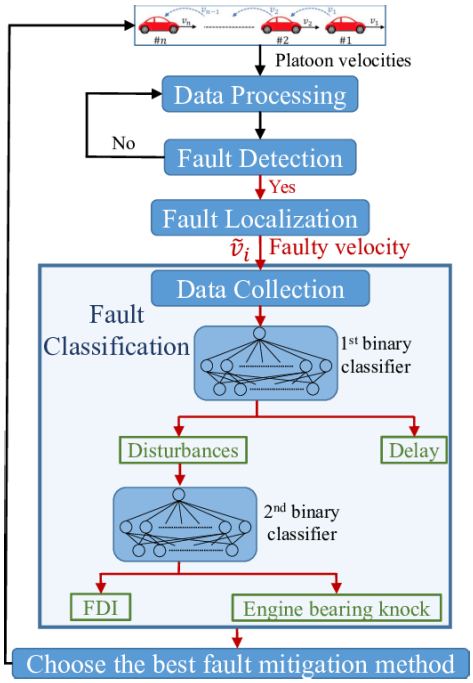

FIGURE 7. Fault classification scheme [115].

objects called *ghosts*. To collect data for the test, several experiments were performed on a 77 GHz automotive radar. Then, a list of detection features is computed and extracted for use as input for a classifier based on the Random Forests (RF) algorithm. The classifier then divides the detected data into three classes: real moving detection, infrastructure, and ghost Detection. To predict in which class a new observation is included, it is necessary to follow the decisions in the tree from the root to one of the leaf nodes containing the class type that is consistent with its characteristics.

The authors of [115] have addressed the safety and stability of CAVs platoons in the presence of faults. To this end, they proposed a supervised classifier capable of identifying and classifying anomalies into three categories: Engine Bearing Knock (EBK), False Data Injection (FDI) attack, or communication time delay. As shown in Fig. 7, to properly identify the anomaly class, the authors have relied on two classifiers, the first of which has the role of checking whether the fault is a limited disturbance or a communication delay, while the second determines if the disturbances are of EBK type or FDI attacks. Also, for each classifier, the authors have tested several techniques from the literature (mainly, SVM, Naive Bayes (NB), K-Nearest Neighbors (KNN), and Quadratic Discriminant (QD)) in order to validate the best combination. Simulation results proved that using Q-SVM as a first classifier and a QD for the second gives a high accuracy rate of 97.7% and 96.2%, respectively.

C. DEEP LEARNING-BASED TECHNIQUES

The adoption of DL offers numerous benefits compared to classical ML methods, mainly in terms of processing non-linear or complex data and the accuracy of anomaly detection. This explains the substantial amount of ongoing research centered around this approach. In this section, we present various suggested techniques for identifying anomalies in sensor data within Connected and Autonomous Vehicles. A summary of these solutions is presented in Table 7.

In [116], the authors have proposed an anomaly detection technique for CAVs based on the Long Short-Term Memory (LSTM) deep network model. The solution proposed in this study aims to detect FDI attacks on the control system of autonomous vehicles, where attackers manipulate sensor data to compromise vehicle behavior. The dataset used for the simulation was generated by injecting FDI attacks into an AV simulation-based system, developed by MathWorks Inc, to create anomalies. After the preprocessing step, the data will be passed to the LSTM layer. Then, the prediction process (normal/abnormal) is performed in the fully connected Softmax layers. To demonstrate the effectiveness of this approach, its performance was evaluated against existing methods in the literature in terms of accuracy where it achieved an average of 99.95%.

In [117], Wyk et al. have proposed an approach for detecting anomalies, caused by false injection attacks and sensor failures, which combines two methods: DL using convolutional neural network (CNN) and Kalman Filtering (KF) with a X^2 detector. In the CNN-KF framework, the initial step involves processing the raw data using the CNN layer to detect and eliminate abnormal readings. Subsequently, the remaining normal data is then forwarded to the KF model for additional identification of any other anomalies. To simulate the proposed solution, the authors have injected different types of anomalies into the data set extracted from the RDE database.

The study in [118] proposed an anomaly detection method that integrates a combination of a multistage attention mechanism with a CNN based on the LSTM model. The aim is to detect anomalies resulting from various factors such as sensor malfunctions, errors, or cyberattacks. As shown in Fig. 8, the data in the MSALSTM-CNN model is first reformed into 3D sequences to feed the CNN model. The features extracted by the CNN are converted into vectors and forwarded to the LSTM layer. In addition, the authors also have designed a method called Weight-Adjusted fine-tuned Ensemble (WAVED), comprising a set of distinct classifiers that are adjusted according to weights. This method is intended to detect anomalies in multi-sensor data. As in previous works, anomalies were injected by modifying the original dataset obtained from the Safety Pilot Model Deployment (SPMD) dataset [119]. Simulation results show that the MSALSTM-CNN method can provide high anomaly detection rates regardless of whether the anomaly rates in the dataset are low or high.

Another DL-based anomaly detection solution is presented in [120], which uses a modified CNN, denoted by M-CNN, to identify a variety of anomaly types such as sensor faults

TABLE 6. Summary of selected ML-based techniques.

Ref.	Year	Technique	Dataset	Source Sensors	Dedicated For	Simulated Anomalies	Evaluation Criteria
[110]	2023	ADS-Lead	Baidu Apollo platform, GTSRB and Tumsimple public datasets	GPS, IMU, Camera	Time series anomalies Adversarial image	Localization attacks, Lane detection attacks, Traffic sign recognition attacks	Accuracy, Precision Recall, F1-Score
[112]	2023	Detection model using learning from demonstration	KAIST, Michigan roundabout datasets	IMU, GPS, LiDAR	GPS spoofing attack	MSF-based localization attack, Falsified BSMs	Average Displacement Error, Detection time False Positive, False Negative
[115]	2021	Supervised fault classifiers	Experiment data	Multiple sensor types	Faults	False Data Injection attack, Communication time delay, Engine Bearing Knock	Accuracy, Delay
[113]	2021	LIFE	Modified KITTI dataset	LIDAR and camera data	Perception error attacks	Camera blinding attack LIDAR spoofing attack LIDAR saturation attack LIDAR distance error attack LIDAR rotation error attack	Detection ratio False alarms
[111]	2020	AEKF with OCSVM model	RDE database	Multiple sensor types	False injection attacks Sensor faults	Short, Noise, Bias, Gradual drift	Accuracy, ROC
[114]	2019	Random Forest classifier	Real world experiments	Long range front Radar	Ghost detections	N/A	Prediction success rate

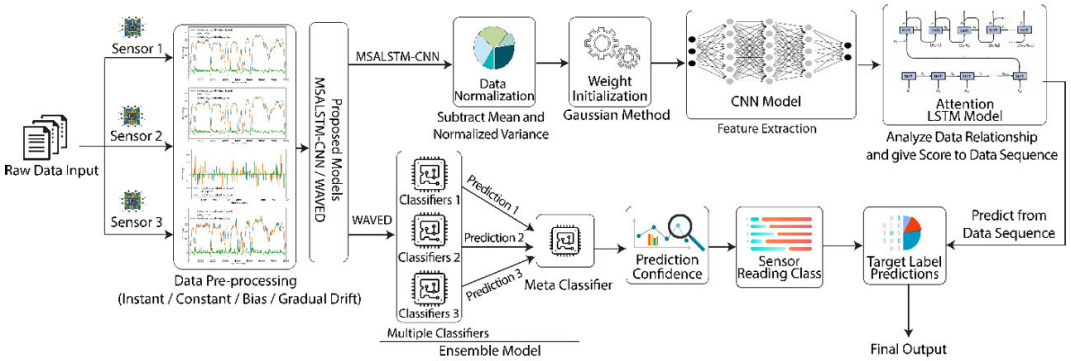

FIGURE 8. Overview of the proposed MSALSTM-CNN framework [118].

or cyberattacks. As shown in Fig. 10, as a first step, the raw data, obtained from the SMDP database, is pre-processed to eliminate redundant data and irrelevant null values. These data are then used to extract from them the most relevant features for anomaly detection. To extract these features, convolutional and pooling layers are used for this task. Thus, the M-CNN architecture consists of five convolutional layers, and after each convolutional layer, a maximum pooling layer is applied. Finally, anomaly detection is performed in the fully connected layer of CNN. To highlight the detection efficiency of this model, the authors compared it with two algorithms widely used in the literature, SVM and Isolation Forest (IF) algorithms. The presented simulation results show that the MCNN model provides a high accuracy rate of 99%, outperforming the other two models.

The work of [121] introduced a new anomaly detection model, called CWT-CNN, which combines Continuous Wavelet Transform (CWT) and CNN to identify anomalies caused by malicious behaviors. By transforming the in-vehicle sensor signals extracted from a real-world dataset into a CWT scalogram (CWTS), the model captures both time and frequency domain information. Then, the CNN uses the CWTS generated to learn and differentiate between normal and abnormal vehicle sensor behaviors. The dataset for the

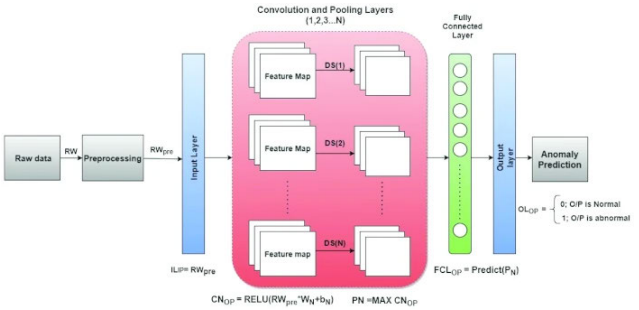

FIGURE 9. General architecture of the MCNN-based anomaly detection technique [120].

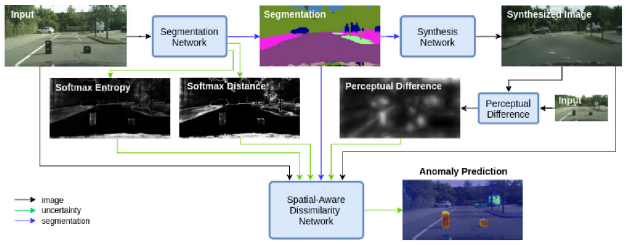

FIGURE 10. Anomaly segmentation framework proposed in [129].

simulation is taken from Open Source Driving data [122]. These data were collected using a Lincoln MKZ car driving for 70 minutes in various weather conditions.

In [123], the authors have proposed a DL model that enables CAVs to perform real-time anomaly detection on the data collected from both onboard and external sensors. To achieve this goal, an LSTM Auto-Encoder (AE) is employed to extract relevant features from the input signals. Once these features are extracted, they will be taken as input for the CNN classifier which is composed of three one-dimensional convolutional layers with 32 filters and kernels of different sizes and thus the data can be treated at different resolutions. To evaluate the performance of the AE-CNN model, the authors have employed real-world data from the Multi-Modal Intelligent Transportation Signal Systems (MMITSS) dataset. Additionally, they investigated the impact of tuning model parameters on anomaly detection across three scenarios. The proposed technique achieves a precision rate of 94.2% and an accuracy of 95.5%.

To protect CAVs against sensor errors and security threats, Rezaei et al. [124], have proposed a detection technique, based on GAN, called GAN-enabled Autoencoder for Anomaly Detection (GAAD). Thus, the authors have used a GAN-based method for detecting anomalies, building on a framework already proposed in literature called GANomaly. They further extended the GANomaly architecture by incorporating an extra AE. The hypothesis is that the AE can refine the generator's reconstructed outputs and correct errors introduced during the reconstruction of anomalous behavior. To evaluate this technique, the authors have used a dataset extracted from real data collected by a fleet of 20 autonomous vehicles.

Watts et al. [125], have presented another technique for detecting anomalies caused by anomalous/faulty information by integrating a CNN classification model into a Bayesian framework comprising a Partially Observable Markov Decision Process (POMDP) model. In this method, the CNN model first analyzes past sensor readings and provides probabilities of anomalies at each epoch. These probabilities, along with additional features, serve as 'imperfect observations' for the POMDP model. In the next step, the POMDP model determines the optimal anomaly classification threshold based on the system's belief state. Furthermore, they employed the Asynchronous Advantage Actor-Critic (A3C) algorithm to approximate the optimal policy, enabling a dynamic adjustment of the threshold in real time.

The authors in [126] have tackled the problem of corner cases (anomalies) resulting from the identification of instances outside the learning distribution (unknown classes). To this end, a novel pipeline is proposed for detecting unknown objects by leveraging the strengths of both LiDAR and camera data. To start the process, the input image undergoes semantic segmentation, resulting in a mask that outlines the road's coordinates. This mask is subsequently projected into the 3D LiDAR space. When certain objects cannot be detected by the 3D object detector, their corresponding points are remapped back into the 2D image space, which contains more detailed information. Afterwards, an image classifier is utilized to classify these objects. An object is deemed anomalous only if the image classifier fails to assign it to a specific class. For the evaluation, the authors presented a qualitative evaluation only by defining typical scenarios and presenting their outputs.

Similar to the previously mentioned study, Jin et al. [127], have exploited the advantages of fusing camera data with LiDAR point cloud to provide an anomaly detection framework that facilitates robust autonomous navigation against disturbances. The authors have defined three main components in this framework: (i) joint representation learning for camera and LiDAR data fusion using variational auto-encoder (VAE), (ii) anomaly detection learning to identify anomalous readings, and (iii) anomaly reconstruction and navigation policy learning that helps reduce the anomaly effects. For the evaluation, the dataset is synthetically generated using the CARLA platform.

Unlike previous approaches that target global anomaly detection and work across various sensor types, the solution presented in [128] focused specifically on radar sensor readings. Thus, the authors have addressed the problem of ghost targets (false targets), which can interfere with radar operations. They have designed an anomaly detection technique based on the PointNet++ architecture in which they have extended the Multi-Scale Grouping (MSG) module with a new module called Multi-Form Grouping (MFG), considering anomalous radar targets in a ring-shaped region around the radar sensor origin. The MFG module combines both circular and ring querying forms to capture neighbors information at multiple scales.

TABLE 7. Summary of selected DL-based techniques.

Ref.	Year	Technique	Dataset	Source Sensors	Dedicated For	Simulated Anomalies	Evaluation Criteria
[127]	2023	Robust autonomous navigation framework	CARLA platform	Camera and LiDAR data	Occlusion, sensor noise, challenging weather illumination conditions	Semantic Disturbance, Occlusion, Luminance Contrast Interference	Route Completion Driving Score
[121]	2023	CWT-CNN	Real-world vehicle dataset	Multiple sensor types	Cyberattacks	Hijack, Bias, Injection DOS, Replay	Accuracy, Precision Recall, F1-Score
[116]	2022	Intelligent symmetrical LSTM neural network	Dataset generated using MathWorks	Multiple sensor types	False Data Injection (FDI) attacks	FDI attacks	Accuracy, Precision Recall, F1-Score
[120]	2022	M-CNN	SPMD dataset	Multiple sensor types	Fault, errors, cyberattacks	Instant	Accuracy, Precision Recall, ROC
[126]	2022	Multimodal detection of unknown objects	Waymo Open Dataset	LiDAR and camera data	Unknown classes	N/A	N/A
[125]	2022	POMDP	RDE database	Multiple sensor types	Anomalous/faulty information	Bias, Gradual drift	Accuracy, Recall, Specificity, PPV, F1-score, Improve
[128]	2021	PointNet++ MFG	hand-labeled dataset	ARS 408-21 radar data	Ghost targets	Anomalous radar targets	F1-score, Inference time
[130]	2021	Multi-stage intrusion detection framework	UNSWNB-15 dataset car hacking dataset	Multiple sensor types	Cyberattacks Zero-day attacks	A variety of attacks (DoS, Gear, Fuzzy,...	Accuracy, Recall, Precision,F-score, kappa
[124]	2021	GAAD	Lyft Level 5 dataset	Multiple sensor types	False injection attacks Sensor faults	Bias, Gradual drift	Accuracy, Recall Specificity, G-Mean
[129]	2021	Pixel-wise anomaly detection framework	FS Lost & Found (L&F), FS Static, FS Web	Camera	Pixel anomaly	Void class object anomaly, Synthetic Segmentation Anomaly	Average precision (AP), False Positive Rate at 95%, True positive rate (FPR95)
[118]	2020	MSALSTM-CNN	A modified SPMD dataset	Multiple sensor types	Fault, errors, cyberattacks	Instant, Constant, Gradual drift, Bias	Accuracy, Precision Recall, F1-Score
[123]	2020	AE-CNN	MMITSS dataset	Multiple sensor types	Fault, errors, cyberattacks	Abrupt, Intermittent gradual	Accuracy, Precision Recall, F1-Score
[117]	2019	CNN-KF	RDE database	Multiple sensor types	False injection attacks sensor faults	Instant, Constant, Gradual drift, Bias, Miss	Accuracy, Precision Recall, F1-Score

In addition to detecting multipath anomalies such as ghost targets, this approach also provides the ability to deal with single-target anomalies resulting from errors in the direction of arrival estimation or Doppler velocity ambiguities.

In [129], the authors have presented a new solution to improve the reliability of camera data (captured images), against instances of anomalies encountered during semantic segmentation. The authors have proposed a pixel-wise anomaly detection framework combining both the advantages

of resynthesis approaches with that of the uncertainty estimation methods proposed in the literature. At first, the input image will be passed through a segmentation network which will produce a semantic map and two uncertainty maps (softmax entropy and softmax distance). Then, the generated semantic map will be processed to produce a photo-realistic image using the synthesis network. Subsequently, the characteristics of the input and generated images will be compared in order to verify the perceptual difference. Finally, all the images are sent to the spatial dissimilarity module to generate the anomaly prediction.

In [130], a multi-stage intrusion detection technique is presented to mitigate the effects of intrusions on CAVs In-Vehicle Network (IVN). In this framework, the first step consists in filtering and cleaning the extracted data (for training and testing) with the aim of standardizing it. After their pre-processing, the resulted features are passed to a bi-directional normal state-based LSTM classifier for attack identification. To evaluate and improve the performance of the proposed technique, the authors have used two data sources: the UNSWNB-15 database [132], which is intended for exterior network communications, and Car Hacking database [131], which is intended for in-vehicles communication. Moreover, each database encompasses a variety of attack types such as DoS and fuzzy. Simulations indicated that the proposed method provides high performance reaching an accuracy rate of 99.11% for the Car Hacking dataset and 98.88% for the UNSWNB-15 dataset.

V. DISCUSSION AND OPEN RESEARCH ISSUES

After investigating a number of anomaly detection techniques for CAVs existing in the literature, we propose in this section to discuss the main findings of this review and summarize the most important developments in this field. Afterwards, we suggest several open research issues that require additional efforts to devise high-standard ADS for futuristic CAVs.

A. DISCUSSION

It is important to perform a qualitative analysis of these techniques in order to better understand the strengths and weaknesses of the existing methods and guide practitioners in the field about the best strategies to follow for ADS. In this section, we compare the previously discussed solutions based on five criteria: complexity, accuracy, scalability, detection time, and number of anomaly sources investigated. These criteria are defined as follows:

- **Complexity:** describes how difficult the implementation of the solution is, by taking into consideration several factors such as the adopted architecture, the used methods, and the simulation requirements.
- **Accuracy:** This criterion gives an idea of the performance of the solution and to what extent it is precise.
- **Scalability:** determines a solution's ability to adapt to increasing volume of data without degrading performance.
- **Detection time:** refers to the computational complexity of the proposed method and indicates whether a solution is able to rapidly detect anomalies or not.
- **Addressed anomaly sources:** as we have classified, in Section III-B, the potential causes of anomalies into three categories (i.e., Sensor faults, environmental conditions and security threats), we determine the number of causes of anomalies addressed by each solution. For example, 1 means that only one anomaly cause is addressed by the solution, while 3 indicates that the solution addresses the three causes simultaneously.

Table 8 shows the response of each solution to these different criteria. The term "N/A" indicates that it is not clear whether a criterion is satisfied by the solution or not. Through the results of this analysis, we can conclude the following points:

- The majority of these solutions can provide high performance in terms of accuracy and detection time. Thus, as shown in Fig. 11, approximately 83% of these solutions have a high accuracy rate. However, these results are very sensitive to simulation parameters and estimations.
  As a result, a change in these parameters can degrade solution performance.
- Most of these techniques are evaluated generally using low or medium amounts of data. In addition, tests on anomalies are carried out by generating them artificially. All of these factors can influence the scalability of a solution. For instance, only about 39% of solutions can provide a high scalability.
- It is difficult to design a generic anomaly detection solution that deals with all types of anomalies and their potential causes. For instance, the evaluation shows that only about 9% of these solutions have addressed the three categories of anomaly sources. This can be explained by the lack of datasets rich in ready-made labeled anomalous scenarios, especially environmental disturbances such as adverse weather conditions.
- The evaluation shows that the tendency in designing anomaly detection techniques is to focus, generally, on accurate and fast solutions without significantly focusing on the scalability and complexity of the algorithm, seeing that rapidity and safety are top priority for CAVs.
- It can be clearly noticed that DL-based anomaly detection methods show their dominance over other methods. As shown in Fig. 12, about 57% of these solutions are based on DL. This is explained by their ability to manage massive and non-linear data while ensuring a high accuracy in anomaly detection.

B. OPEN RESEARCH ISSUES

Over the last decade, a lot of efforts have been made to develop effective anomaly detection solutions for CAVs. Nevertheless, there are many challenges that need to be further addressed to design ADSs to meet the rapid evolution

TABLE 8. Evaluation of anomaly detection solutions for CAVs according to a variety of criteria.

Ref.	Technique	Addressed Anomaly Sources	Complexity	Accuracy	Scalability	Detection time
[109]	Fault detection and isolation (FDI) for sensor systems	1 (Sensor faults)	Medium	High	Medium	Fast
[110]	ADS-Lead	1 (Cyber attacks)	Medium	Medium	High	Fast
[112]	Detection model using learning from demonstration	1 (Cyber attacks)	Low	High	High	Fast
[115]	Supervised fault classifiers	1 (Cyber attacks)	Medium	High	High	Fast
[113]	LIFE	1 (Cyber attacks)	Medium	High	Medium	Medium
[114]	Random Forest classifier	1 (Environmental conditions)	Low	High	High	N/A
[127]	Robust autonomous navigation framework	1 (Environmental conditions)	Medium	High	High	N/A
[121]	CWT-CNN	1 (Cyber attacks)	Medium	High	Medium	Fast
[116]	Intelligent symmetrical LSTM neural network	1 (Cyber attacks)	Medium	High	Medium	Fast
[128]	PointNet++ MFG	1 (Environmental conditions)	High	Medium	Low	Slow
[130]	Multi-stage intrusion detection framework	1 (Cyber attacks)	Medium	High	Medium	Fast
[126]	Multimodal detection of unknown objects	2 (Sensor faults, Environmental conditions)	Medium	N/A	Medium	N/A
[125]	POMDP	2 (Cyber attacks, Sensor faults)	Medium	High	Medium	Fast
[111]	OCSVM with the IDM model	2 (Cyber attacks, Sensor faults)	Medium	High	Medium	Fast
[105]	Secure Sensor Fusion Framework	2 (Cyber attacks, Sensor faults)	Medium	N/A	Medium	N/A
[106]	Augmented Extended Kalman Filter	2 (Cyber attacks, Sensor faults)	Low	High	High	Fast
[108]	Dynamic Bayesian Network	2 (Cyber attacks, Sensor faults)	High	High	Medium	Medium
[124]	GAAD	2 (Cyber attacks, Sensor faults)	Medium	High	Medium	N/A
[129]	Pixel-wise anomaly detection framework	2 (Sensor faults, Environmental conditions)	Low	High	High	Fast
[123]	AE-CNN	2 (Cyber attacks, Sensor faults)	Medium	High	Medium	Fast
[117]	CNN-KF	2 (Cyber attacks, Sensor faults)	Medium	High	High	Fast
[118]	MSALSTM-CNN	3 sources	Medium	High	Medium	N/A
[120]	M-CNN	3 sources	Medium	High	High	Fast

in the automotive industry. These challenges can be identified as follows:

1) ULTRA FAST REAL-TIME PROCESSING

Autonomous vehicles primarily rely on sensory data and notifications from the environment in their operation to make real-time decisions such as navigation, braking, and acceleration. These data are usually massive and require very rapid processing to identify anomalies. In real driving scenarios, particularly on highways, autonomous vehicles navigate at high speed in a high-density environment, where decisions must be made very quickly. A reliable anomaly

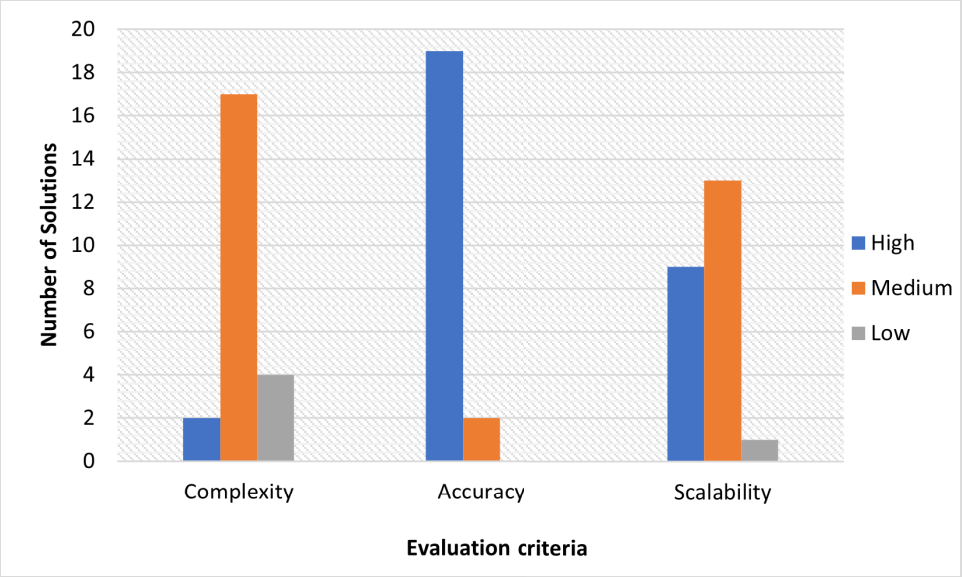

FIGURE 11. Evaluation criteria versus number of solutions.

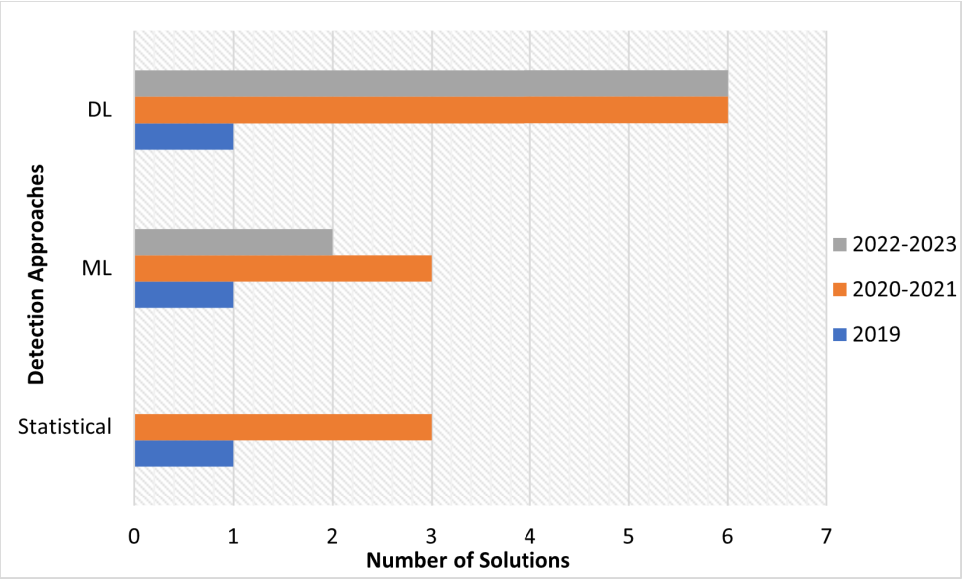

FIGURE 12. Detection approaches per years versus number of solutions.

detection technique must therefore take these constraints into account in order to guarantee ultra-rapid anomaly detection in real-time.

2) DATASET AVAILABILITY

Currently, the existing datasets used for validating and testing anomaly detection solutions for AVs have quite a few limitations hampering the design of efficient and resistible models for different scenarios. Mainly, the lack of labeled data that allows to properly differentiate between normal samples and abnormal ones, can slow down the development of robust models for CAVs. Moreover, most of the proposed solutions are trained only on the normal data and any diverging behavior will be considered as an anomaly. This is essentially due to the absence of datasets rich in abnormal scenarios and attacks, which makes the search for an adequate dataset a difficult task. Therefore, it is a priority to consider existing weaknesses when creating new datasets. Thus, an appropriate dataset must be clearly labeled, contain a variety of scenarios and above all constantly updated to enrich the knowledge of the model with other new anomalies.

3) OPEN-SOURCE SIMULATORS LIMITATIONS

Open-source platforms used for simulating anomaly detection algorithms for autonomous vehicles, despite their numerous advantages, can have several limitations making simulated solutions less reliable in their performance. Thus, existing simulators are generally unable to handle the

complexity of real-world scenarios and rarely occurring situations. Furthermore, the lack of periodic updates can make these platforms unsuitable for rapid changes in CAV technology. For this, it is important to face these limits to guarantee high-precision solutions.

4) SCALABILITY

Many concerns revolve around the use of CAVs and how much this technology will help us ensure road safety. In other words, there are always risks associated with the reliability of sensors and the effectiveness of the decision-making system. For this, many questions are being raised about the ability of AVs to make safe decisions in complex or unpredictable conditions where anomalies are highly likely to occur. Thus, most anomaly detection techniques are trained on less complex and well-structured datasets. However, in real scenarios, CAVs sensors will collect a huge and diverse amount of data, making anomalies more frequent and unpredictable in terms of type and source. In this case, it is very likely that the detection method will lose its efficiency and its performance will deteriorate. For this, it is important to guarantee the detection technique's capability to manage and process a large volume of data.

5) COLLECTIVE ANOMALY DETECTION

The heterogeneous types of anomalies and their varied sources represent a major challenge facing the decision systems of driverless vehicles since it is very difficult to manage all types of anomalies at the same time. In addition, the higher the complexity of the algorithm, the greater the risk of associated latency. Therefore, it is a priority to think of new mechanisms to optimize the accuracy and detection time at once, notably, cooperative anomaly detection where vehicles collaborate to detect all types of new anomalies encountered by sharing data collected from several sources. However, these techniques pose several challenges, especially security and adaptability problems due to the heterogeneity of data. Therefore, future research must focus on these issues in order to create reliable and secure cooperative detection solutions.

6) CONTEXT-AWARENESS FOR AUTONOMOUS DRIVING

Contextual awareness is a crucial aspect of autonomous driving, due to its numerous advantages over CAVs decision systems. Increased contextual awareness allows self-driving vehicles to better react and quickly adapt to changing driving conditions, detecting and avoiding potential risks. Active inference models, as a good example, aim not only to passively perceive the environment but also to actively predict and adapt to unanticipated changes, which is crucial for safe and efficient autonomous driving. However, there are still some concerns that require further investigation in future research, essentially, investigating how these models can better handle uncertainty and make robust decisions in less than perfect conditions.

7) INFRASTRUCTURE AND EMERGENCY SITUATIONS

Undoubtedly, the present Infrastructure faces numerous limitations concerning equipment and connectivity. Hence, to guarantee permanent cooperation and exchange between CAVs, reliable high-speed connectivity is essential to allow the different forms of communications within a network of vehicles including V2V, V2I, and V2X. This will, consequently, allow vehicles to improve their decisions and increase their knowledge of their surrounding. In addition, until now, tests on AVs have been carried out in well-selected areas and less complex driving conditions. However, during the effective use of CAVs, several unpredictable events will be produced and will put the vehicle in unusual scenarios. Thus, several events can disrupt the normal operation of the AV, such as the case of an accident on the road, the sudden sight of a pedestrian, a sudden braking failure or damaged roads. In these cases, it is important to ensure that the vehicle reacts quickly to avoid collisions and to protect the driver and other road users.

8) PRIVACY PRESERVATION

Autonomous vehicles are vulnerable to a wide range of potential attacks, highlighting the critical importance of enhancing their cybersecurity. However, it is also important to point out the security issues related to protecting user privacy and ensuring the integrity of sensitive information. Since CAVs sensors collect several types of data from their surroundings, it is possible that some of this data is sensitive, for example camera sensors can capture the faces of pedestrians and vehicle license plates. To address these challenges, it is essential to pay close attention to the variety of threats in future research. This will enable the development of new security solutions that consider the specificities of this technology.

9) THREAT ANALYSIS AND RISK ASSESSMENT MODELS INTEGRATION

The security of CAVs can be improved through the use of threat analysis frameworks. These approaches make it possible to comprehensively identify potential risks, thereby helping to enhance the resilience and reliability of autonomous systems. Thus, there are several systematic threat analysis and risk assessment models that can be applied to the security of self-driving vehicles, such as Threat Vulnerability Risk Analysis (TVRA) [133], Spoofing, Tampering, Repudiation, Information disclosure, DoS, Elevation of privilege (STRIDE) [134] and more [135]. In this context, adopting the System-Theoretic Process Analysis for Security (STPA-Sec) methodology can effectively contribute to the creation of resilient and secure autonomous driving environments. STPA-Sec is a systematic methodology widely used in the safety assessment and design of complex systems such as autonomous driving systems. This framework excels in identifying risks such as cyberattacks against AI algorithms, unauthorized access

to vehicle control commands and its sensors. STPA-Sec provides a comprehensive hazard assessment, addressing concerns related to incorrect decisions, system failures and malicious interventions [136]. Additionally, their focus on developing and implementing security controls ensures a strong defense against identified risks. In the following, how the integration of STPA-sec approach can benefit the safety of autonomous driving systems:

- A holistic risk analysis: STPA-Sec allows a complete understanding by considering the entire self-driving system, including hardware or software components, the various embedded sensors, in addition to interactions with the environment. This ensures a complete overview of possible safety hazards.
- Threat Anticipation: Thanks to its systemic approach, STPA-Sec allows for the early identification of dangers that could compromise the operation of autonomous vehicles. This enables proactive intervention before risks materialize.
- In-Depth Problem Analysis: STPA-Sec goes beyond superficial problem identification, by analyzing deeply the causes of vulnerabilities. This allows for more effective corrective measures, helping also to understand the reasons why certain situations may turn into dangers.

VI. CONCLUSION

Autonomous vehicles represent a very promising future alternative for road safety and the efficiency of transportation systems. However, several challenges require further attention to ensure the successful implementation of this technology. Anomalies are one of those major issues that can threaten the functioning of a driverless vehicle system. Through the study presented in this paper, we have drawn up an overview of CAVs, the types, causes and impacts of anomalies that can affect its sensors and decision systems as well as the varieties of detection methods.

First, we have presented the key components of a CAV system and the main types of integrated sensors, with a focus on the services and benefits offered by autonomous driving. We have also provided a detailed taxonomy of anomaly detection systems, where a variety of important elements in anomaly analysis and identification are discussed, including potential sources of anomaly, different categories of detection techniques, open datasets for anomaly research and testing and available open-source simulators. This can help researchers interested in the field of autonomous driving and anomaly detection specifically to better understand this issue.

The second step of this study consists in discussing a number of recent anomaly detection solutions dedicated to CAVs. We have classified these techniques mainly into three categories: statistical, classical machine learning, and deep learning. In addition, for each technique, we have provided information about the used algorithms for the implementation, and the simulation environment such as the dataset used and the type of addressed anomalies. Then, we have presented a qualitative evaluation of these solutions based on several criteria. The aim of this evaluation is to focus on the strengths and weaknesses of each solution as well as the open areas for research.

Finally, we have identified some open research issues that require more attention in order to design reliable anomaly detection solutions. In future work, we plan to design an anomaly detection solution taking into consideration the challenges presented in this study.

REFERENCES

[1] M. Alam, J. Ferreira, and J. Fonseca, "Introduction to intelligent transportation systems," in *Intelligent Transportation Systems*. Cham, Switzerland: Springer, 2016, pp. 1–17.

[2] M. Hadded, P. Muhlethaler, A. Laouiti, R. Zagrouba, and L. A. Saidane, "TDMA-based MAC protocols for vehicular ad hoc networks: A survey, qualitative analysis, and open research issues," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2461–2492, 4th Quart., 2015.

[3] S. Baccari, H. Touati, M. Hadded, and P. Muhlethaler, "Performance impact analysis of security attacks on cross-layer routing protocols in vehicular ad hoc networks," in *Proc. Int. Conf. Softw., Telecommun. Comput. Netw. (SoftCOM)*, Split, Croatia, Sep. 2020, pp. 1–6, doi: 10.23919/SoftCOM50211.2020.9238259.

[4] M. N. Ahangar, Q. Z. Ahmed, F. A. Khan, and M. Hafeez, "A survey of autonomous vehicles: Enabling communication technologies and challenges," *Sensors*, vol. 21, no. 3, p. 706, Jan. 2021, doi: 10.3390/s21030706.

[5] I. W. Damaj, J. K. Yousafzai, and H. T. Mouftah, "Future trends in connected and autonomous vehicles: Enabling communications and processing technologies," *IEEE Access*, vol. 10, pp. 42334–42345, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9759292

[6] *Waymo—Self-Driving Cars—Autonomous Vehicles*. [Online]. Available: https://waymo.com/waymo-one/

[7] M. Hadded, P. Merdrignac, S. Duhamel, and O. Shagdar, "Security attacks impact for collective perception based roadside assistance: A study of a highway on-ramp merging case," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, Limassol, Cyprus, Jun. 2020, pp. 1284–1289, doi: 10.1109/IWCMC48107.2020.9148235.

[8] *Global Status Report on Road Safety 2018*, World Health Org., Geneva, Switzerland, Dec. 2018.

[9] M. Taiebat, A. L. Brown, H. R. Safford, S. Qu, and M. Xu, "A review on energy, environmental, and sustainability implications of connected and automated vehicles," *Environ. Sci. Technol.*, vol. 52, pp. 11449–11465, Sep. 2018.

[10] Y.-C. Lee and J. H. Mirman, "Parents' perspectives on using autonomous vehicles to enhance children's mobility," *Transp. Res. C, Emerg. Technol.*, vol. 96, pp. 415–431, Nov. 2018.

[11] R. Bennett, R. Vijaygopal, and R. Kottasz, "Willingness of people who are blind to accept autonomous vehicles: An empirical investigation," *Transp. Res. F, Traffic Psychol. Behaviour*, vol. 69, pp. 13–27, Feb. 2020.

[12] B. E. Dicianno, S. Sivakanthan, S. A. Sundaram, S. Satpute, H. Kulich, E. Powers, N. Deepak, R. Russell, R. Cooper, and R. A. Cooper, "Systematic review: Automated vehicles and services for people with disabilities," *Neurosci. Lett.*, vol. 761, Sep. 2021, Art. no. 136103.

[13] D. Rojas-Rueda, M. J. Nieuwenhuijsen, H. Khreis, and H. Frumkin, "Autonomous vehicles and public health," *Annu. Rev. Public Health*, vol. 41, no. 1, pp. 329–345, Apr. 2020.

[14] H. Khayyam, B. Javadi, M. Jalili, and R. N. Jazar, "Artificial intelligence and Internet of Things for autonomous vehicles," in *Nonlinear Approaches in Engineering Applications*. Cham, Switzerland: Springer, 2020, pp. 39–68.

[15] A. Qayyum, M. Usama, J. Qadir, and A. Al-Fuqaha, "Securing connected & autonomous vehicles: Challenges posed by adversarial machine learning and the way forward," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 998–1026, 2nd Quart., 2020, doi: 10.1109/comst.2020.2975048.

[16] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A survey of autonomous driving: Common practices and emerging technologies," *IEEE Access*, vol. 8, pp. 58443–58469, 2020, doi: 10.1109/ACCESS.2020.2983149.

[17] J. Van Brummelen, M. O'Brien, D. Gruyer, and H. Najjaran, "Autonomous vehicle perception: The technology of today and tomorrow," *Transp. Res. C, Emerg. Technol.*, vol. 89, pp. 384–406, Apr. 2018.

[18] L. Liu, S. Lu, R. Zhong, B. Wu, Y. Yao, Q. Zhang, and W. Shi, "Computing systems for autonomous driving: State of the art and challenges," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6469–6486, Apr. 2021, doi: 10.1109/JIOT.2020.3043716.

[19] J. Z. Varghese and R. G. Boone, "Overview of autonomous vehicle sensors and systems," in *Proc. Int. Conf. Oper. Excellence Service Eng.*, 2015, pp. 178–191.

[20] J. Wang, L. Zhang, Y. Huang, and J. Zhao, "Safety of autonomous vehicles," *J. Adv. Transp.*, vol. 2020, Oct. 2020, Art. no. 8867757.

[21] D. Bogdoll, M. Nitsche, and J. M. Zollner, "Anomaly detection in autonomous driving: A survey," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 4488–4499.

[22] M. A. Samara, I. Bennis, A. Abouaissa, and P. Lorenz, "A survey of outlier detection techniques in IoT: Review and classification," *J. Sensor Actuator Netw.*, vol. 11, no. 1, p. 4, Jan. 2022.

[23] P. Dixit, P. Bhattacharya, S. Tanwar, and R. Gupta, "Anomaly detection in autonomous electric vehicles using AI techniques: A comprehensive survey," *Expert Syst.*, vol. 39, no. 5, p. e12754, Jun. 2022.

[24] J. Breitenstein, J.-A. Termöhlen, D. Lipinski, and T. Fingscheidt, "Systematization of corner cases for visual perception in automated driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Las Vegas, NV, USA, Oct. 2020, pp. 1257–1264, doi: 10.1109/IV47402.2020.9304789.

[25] J. Breitenstein, J.-A. Termöhlen, D. Lipinski, and T. Fingscheidt, "Corner cases for visual perception in automated driving: Some guidance on detection approaches," 2021, *arXiv:2102.05897*.

[26] R. Al-amri, R. K. Murugesan, M. Man, A. F. Abdulateef, M. A. Al-Sharafi, and A. A. Alkahtani, "A review of machine learning and deep learning techniques for anomaly detection in IoT data," *Appl. Sci.*, vol. 11, no. 12, p. 5320, Jun. 2021.

[27] Society of Automotive Engineers (SAE). *Website*. [Online]. Available: https://www.sae.org

[28] *Taxonomy and Definitions for Terms Related to On-road Motor Vehicle Automated Driving Systems*, SAE Committee, Warrendale, PA, USA, 2014.

[29] (2021). *SAE: SAE Levels of Driving Automation*. [Online]. Available: https://www.sae.org/site/blog/sae-j3016-update

[30] Y. Deng, T. Zhang, G. Lou, X. Zheng, J. Jin, and Q. -L. Han, "Deep learning-based autonomous driving systems: A survey of attacks and defenses," *IEEE Trans. Ind. Informat.*, vol. 17, no. 12, pp. 7897–7912, Dec. 2021, doi: 10.1109/TII.2021.3071405.

[31] S. Malik, M. A. Khan, H. El-Sayed, J. Khan, and O. Ullah, "How do autonomous vehicles decide?" *Sensors*, vol. 23, no. 1, p. 317, Dec. 2022.

[32] X. Yi, H. Ghazzai, and Y. Massoud, "End-to-end neural network for autonomous steering using LiDAR point cloud data," in *Proc. IEEE 65th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Fukuoka, Japan, Aug. 2022, pp. 1–4, doi: 10.1109/MWSCAS54063.2022.9859277.

[33] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," *Digit. Signal Process.*, vol. 126, Jun. 2022, Art. no. 103514.

[34] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[35] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. 14th Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands. Cham, Switzerland: Springer, 2016, pp. 3–21.

[36] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2014, pp. 580–587.

[37] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[38] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," 2015, *arXiv:1506.01497v3*.

[39] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.

[40] D. J. Yeong, G. Velasco-Hernandez, J. Barry, and J. Walsh, "Sensor and sensor fusion technology in autonomous vehicles: A review," *Sensors*, vol. 21, no. 6, p. 2140, Mar. 2021.

[41] J. Kocic, N. Jovicic, and V. Drndarevic, "Sensors and sensor fusion in autonomous vehicles," in *Proc. 26th Telecommun. Forum (TELFOR)*, Belgrade, Serbia, Nov. 2018, pp. 420–425, doi: 10.1109/TELFOR.2018.8612054.

[42] P. Wang, "Research on comparison of LiDAR and camera in autonomous driving," *J. Phys., Conf. Ser.*, vol. 2093, no. 1, Nov. 2021, Art. no. 012032.

[43] S. Campbell, N. O'Mahony, L. Krpalcova, D. Riordan, J. Walsh, A. Murphy, and C. Ryan, "Sensor technology in autonomous vehicles: A review," in *Proc. 29th Irish Signals Syst. Conf.*, Jun. 2018, pp. 1–4.

[44] J. Vargas, S. Alsweiss, O. Toker, R. Razdan, and J. Santos, "An overview of autonomous vehicles sensors and their vulnerability to weather conditions," *Sensors*, vol. 21, no. 16, p. 5397, Aug. 2021.

[45] X. Yi, H. Ghazzai, and Y. Massoud, "A LiDAR-assisted smart car-following framework for autonomous vehicles," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Monterey, CA, USA, May 2023, pp. 1–5, doi: 10.1109/iscas46773.2023.10181437.

[46] H. A. Ignatious, H. Sayed, and M. Khan, "An overview of sensors in autonomous vehicles," *Proc. Comput. Sci.*, vol. 198, pp. 736–741, Jan. 2022.

[47] D. J. Yeong, J. Barry, and J. Walsh, "A review of multi-sensor fusion system for large heavy vehicles off road in industrial environments," in *Proc. 31st Irish Signals Syst. Conf. (ISSC)*, Jun. 2020, pp. 1–6.

[48] R. Stiawan, A. Kusumadjati, N. S. Aminah, M. Djamal, and S. Viridi, "An ultrasonic sensor system for vehicle detection application," *J. Phys., Conf. Ser.*, vol. 1204, Apr. 2019, Art. no. 012017.

[49] Y. Zhang, A. Carballo, H. Yang, and K. Takeda, "Perception and sensing for autonomous vehicles under adverse weather conditions: A survey," *ISPRS J. Photogramm. Remote Sens.*, vol. 196, pp. 146–177, Feb. 2023.

[50] A. S. Mohammed, A. Amamou, F. K. Ayevide, S. Kelouwani, K. Agbossou, and N. Zioui, "The perception system of intelligent ground vehicles in all weather conditions: A systematic literature review," *Sensors*, vol. 20, no. 22, p. 6532, Nov. 2020.

[51] M. Hirz and B. Walzel, "Sensor and object recognition technologies for self-driving cars," *Comput.-Aided Design Appl.*, vol. 15, no. 4, pp. 501–508, Jul. 2018.

[52] Z. Wang, Y. Wu, and Q. Niu, "Multi-sensor fusion in automated driving: A survey," *IEEE Access*, vol. 8, pp. 2847–2868, 2020, doi: 10.1109/ACCESS.2019.2962554.

[53] K. Banerjee, D. Notz, J. Windelen, S. Gavarraju, and M. He, "Online camera LiDAR fusion and object detection on hybrid data for autonomous driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Changshu, China, Jun. 2018, pp. 1632–1638, doi: 10.1109/IVS.2018.8500699.

[54] F. Garcia, D. Martin, A. de la Escalera, and J. M. Armingol, "Sensor fusion methodology for vehicle detection," *IEEE Intell. Transp. Syst. Mag.*, vol. 9, no. 1, pp. 123–133, Jan. 2017, doi: 10.1109/MITS.2016.2620398.

[55] R. Arvin, A. J. Khattak, M. Kamrani, and J. Rio-Torres, "Safety evaluation of connected and automated vehicles in mixed traffic with conventional vehicles at intersections," *J. Intell. Transp. Syst.*, vol. 25, no. 2, pp. 170–187, Mar. 2021, doi: 10.1080/15472450.2020.1834392.

[56] J. Rios-Torres and A. A. Malikopoulos, "Impact of connected and automated vehicles on traffic flow," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Yokohama, Japan, Oct. 2017, pp. 1–6, doi: 10.1109/ITSC.2017.8317654.

[57] H. U. Ahmed, Y. Huang, P. Lu, and R. Bridgelall, "Technology developments and impacts of connected and autonomous vehicles: An overview," *Smart Cities*, vol. 5, no. 1, pp. 382–404, Mar. 2022.

[58] J. M. Bandeira, E. Macedo, P. Fernandes, M. Rodrigues, M. Andrade, and M. C. Coelho, "Potential pollutant emission effects of connected and automated vehicles in a mixed traffic flow context for different road types," *IEEE Open J. Intell. Transp. Syst.*, vol. 2, pp. 364–383, 2021, doi: 10.1109/OJITS.2021.3112904.

[59] R. E. Stern, Y. Chen, M. Churchill, F. Wu, M. L. D. Monache, B. Piccoli, B. Seibold, J. Sprinkle, and D. B. Work, "Quantifying air quality benefits resulting from few autonomous vehicles stabilizing traffic," *Transp. Res. D, Transp. Environ.*, vol. 67, pp. 351–365, Feb. 2019, doi: 10.1016/j.trd.2018.12.008.

[60] S. Rafael, L. P. Correia, D. Lopes, J. Bandeira, M. C. Coelho, M. Andrade, C. Borrego, and A. I. Miranda, "Autonomous vehicles opportunities for cities air quality," *Sci. Total Environ.*, vol. 712, Apr. 2020, Art. no. 136546, doi: 10.1016/j.scitotenv.2020.136546.

[61] P. Kopelias, E. Demiridi, K. Vogiatzis, A. Skabardonis, and V. Zafiropoulou, "Connected & autonomous vehicles—Environmental impacts—A review," vol. 712, Apr. 2020, Art. no. 135237.

[62] A. A. Cook, G. Mısırlı, and Z. Fan, "Anomaly detection for IoT time-series data: A survey," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6481–6494, Jul. 2020, doi: 10.1109/JIOT.2019.2958185.

[63] M. Fahim and A. Sillitti, "Anomaly detection, analysis and prediction techniques in IoT environment: A systematic literature review," *IEEE Access*, vol. 7, pp. 81664–81681, 2019.

[64] M. Realpe, B. X. Vintimilla, and L. Vlacic, "A fault tolerant perception system for autonomous vehicles," in *Proc. 35th Chin. Control Conf. (CCC)*, Chengdu, China, Jul. 2016, pp. 6531–6536, doi: 10.1109/ChiCC.2016.7554385.

[65] A. B. Sharma, L. Golubchik, and R. Govindan, "Sensor faults: Detection methods and prevalence in real-world datasets," *ACM Trans. Sensor Netw.*, vol. 6, no. 3, pp. 1–39, Jun. 2010.

[66] M. Hadded, O. Shagdar, and P. Merdrignac, "Augmented perception by V2X cooperation (PAC-V2X): Security issues and misbehavior detection solutions," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Tangier, Morocco, Jun. 2019, pp. 907–912, doi: 10.1109/IWCMC.2019.8766683.

[67] S. Parkinson, P. Ward, K. Wilson, and J. Miller, "Cyber threats facing autonomous and connected vehicles: Future challenges," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 2898–2915, Nov. 2017, doi: 10.1109/TITS.2017.2665968.

[68] X. Sun, F. R. Yu, and P. Zhang, "A survey on cyber-security of connected and autonomous vehicles (CAVs)," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6240–6259, Jul. 2022, doi: 10.1109/TITS.2021.3085297.

[69] M. Pham and K. Xiong, "A survey on security attacks and defense techniques for connected and autonomous vehicles," *Comput. Secur.*, vol. 109, Oct. 2021, Art. no. 102269.

[70] M. Masmoudi, H. Ghazzai, M. Frikha, and Y. Massoud, "Object detection learning techniques for autonomous vehicle applications," in *Proc. IEEE Int. Conf. Veh. Electron. Saf. (ICVES)*, Cairo, Egypt, Sep. 2019, pp. 1–5, doi: 10.1109/ICVES.2019.8906437.

[71] W. Jia, R. M. Shukla, and S. Sengupta, "Anomaly detection using supervised learning and multiple statistical methods," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Boca Raton, FL, USA, Dec. 2019, pp. 1291–1297, doi: 10.1109/ICMLA.2019.00211.

[72] H. Tyagi and R. Kumar, "Attack and anomaly detection in IoT networks using supervised machine learning approaches," *Revue d'Intelligence Artificielle*, vol. 35, no. 1, pp. 11–21, Feb. 2021.

[73] K. Leung and C. Leckie, "Unsupervised anomaly detection in network intrusion detection using clusters," in *Proc. 28th Australas. Conf. Comput. Sci.*, vol. 38, 2005, pp. 333–342.

[74] J. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. A. Zuluaga, "USAD: Unsupervised anomaly detection on multivariate time series," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2020, pp. 3395–3404.

[75] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "GANomaly: Semi-supervised anomaly detection via adversarial training," in *Proc. 14th Asian Conf. Comput. Vis.*, Perth, WA, Australia. Cham, Switzerland: Springer, Dec. 2018, pp. 622–637.

[76] K. Demestichas, T. Alexakis, N. Peppes, and E. Adamopoulou, "Comparative analysis of machine learning-based approaches for anomaly detection in vehicular data," *Vehicles*, vol. 3, no. 2, pp. 171–186, Apr. 2021.

[77] Y. Dong, K. Chen, Y. Peng, and Z. Ma, "Comparative study on supervised versus semi-supervised machine learning for anomaly detection of in-vehicle CAN network," in *Proc. IEEE 25th Int. Conf. Intell. Transp. Syst. (ITSC)*, Macau, China, Oct. 2022, pp. 2914–2919, doi: 10.1109/ITSC55140.2022.9922235.

[78] Q. Li, R. Li, K. Ji, and W. Dai, "Kalman filter and its application," in *Proc. 8th Int. Conf. Intell. Netw. Intell. Syst. (ICINIS)*, Tianjin, China, Nov. 2015, pp. 74–77, doi: 10.1109/ICINIS.2015.35.

[79] M. Çelik, F. Dadaşer-Çelik, and A. Ş. Dokuz, "Anomaly detection in temperature data using DBSCAN algorithm," in *Proc. Int. Symp. Innov. Intell. Syst. Appl.*, Istanbul, Turkey, 2011, pp. 91–95, doi: 10.1109/INISTA.2011.5946052.

[80] S. D. Jadhav and H. P. Channe, "Comparative study of K-NN, naive Bayes and decision tree classification techniques," *Int. J. Sci. Res.*, vol. 5, no. 1, pp. 1842–1845, 2016.

[81] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. their Appl.*, vol. 13, no. 4, pp. 18–28, Jul. 1998, doi: 10.1109/5254.708428.

[82] Y. Cui, R. Chen, W. Chu, L. Chen, D. Tian, Y. Li, and D. Cao, "Deep learning for image and point cloud fusion in autonomous driving: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 722–739, Feb. 2022, doi: 10.1109/TITS.2020.3023541.

[83] H.-H. Jebamikyous and R. Kashef, "Autonomous vehicles perception (AVP) using deep learning: Modeling, assessment, and challenges," *IEEE Access*, vol. 10, pp. 10523–10535, 2022, doi: 10.1109/ACCESS.2022.3144407.

[84] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[85] T. Limbasiya, K. Z. Teng, S. Chattopadhyay, and J. Zhou, "A systematic survey of attack detection and prevention in connected and autonomous vehicles," *Veh. Commun.*, vol. 37, Oct. 2022, Art. no. 100515.

[86] H. Wu and F. J. Meng, "Review on evaluation criteria of machine learning based on big data," *J. Phys., Conf. Ser.*, vol. 1486, no. 5, Apr. 2020, Art. no. 052026.

[87] D. Bogdoll, S. Uhlemeyer, K. Kowol, and J. M. Zöllner, "Perception datasets for anomaly detection in autonomous driving: A survey," 2023, *arXiv:2302.02790*.

[88] H. Yin and C. Berger, "When to use what data set for your self-driving car algorithm: An overview of publicly available driving datasets," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Yokohama, Japan, Oct. 2017, pp. 1–8, doi: 10.1109/ITSC.2017.8317828.

[89] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.

[90] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3213–3223.

[91] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving dataset for heterogeneous multitask learning," 2018, *arXiv:1805.04687*.

[92] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, "The ApolloScape dataset for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 954–960.

[93] P. Sun et al., "Scalability in perception for autonomous driving: Waymo open dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 2446–2454.

[94] J. Houston, G. Zuidhof, L. Bergamini, Y. Ye, L. Chen, A. Jain, S. Omari, V. Iglovikov, and P. Ondruska, "One thousand and one hours: Self-driving motion prediction dataset," in *Proc. Conf. Robot Learn.*, vol. 155, Nov. 2021, pp. 409–418.

[95] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A multimodal dataset for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11621–11631.

[96] J. Geyer, Y. Kassahun, M. Mahmudi, X. Ricou, R. Durgesh, A. S. Chung, L. Hauswald, V. H. Pham, M. Mühlegg, S. Dorn, T. Fernandez, M. Jänicke, S. Mirashi, C. Savani, M. Sturm, O. Vorobiov, M. Oelker, S. Garreis, and P. Schuberth, "A2D2: Audi autonomous driving dataset," 2020, *arXiv:2004.06320*.

[97] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, and J. Hays, "Argoverse: 3D tracking and forecasting with rich maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 8748–8757.

[98] P. Kaur, S. Taghavi, Z. Tian, and W. Shi, "A survey on simulators for testing self-driving cars," in *Proc. 4th Int. Conf. Connected Auto. Driving (MetroCAD)*, Detroit, MI, USA, Apr. 2021, pp. 62–70, doi: 10.1109/MetroCAD51599.2021.00018.

[99] M. Masmoudi, H. Friji, H. Ghazzai, and Y. Massoud, "A reinforcement learning framework for video frame-based autonomous car-following," *IEEE Open J. Intell. Transp. Syst.*, vol. 2, pp. 111–127, 2021.

[100] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proc. Conf. Robot Learn.*, 2017, pp. 1–16.

[101] G. Rong, B. H. Shin, H. Tabatabaee, Q. Lu, S. Lemke, M. Možeiko, E. Boise, G. Uhm, M. Gerow, S. Mehta, E. Agafonov, T. H. Kim, E. Sterner, K. Ushiroda, M. Reyes, D. Zelenkovsky, and S. Kim, "LGSVL simulator: A high fidelity simulator for autonomous driving," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–6.

[102] N. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sendai, Japan, Feb. 2004, pp. 2149–2154, doi: 10.1109/IROS.2004.1389727.

[103] Baidu. (2019). *Apollo Auto*. [Online]. Available: https://github.com/ApolloAuto/apollo

[104] S. Shah, D. Dey, C. Lovett, and A. Kapoor, "AirSim: High-fidelity visual and physical simulation for autonomous vehicles," in *Proc. Field Service Robot., Results 11th Int. Conf.* Cham, Switzerland: Springer, 2018, pp. 621–635.

[105] T. Yang and C. Lv, "A secure sensor fusion framework for connected and automated vehicles under sensor attacks," *IEEE Internet Things J.*, vol. 9, no. 22, pp. 22357–22365, Nov. 2022, doi: 10.1109/JIOT.2021.3101502.

[106] Y. Wang, N. Masoud, and A. Khojandi, "Anomaly detection in connected and automated vehicles using an augmented state formulation," in *Proc. Forum Integr. Sustain. Transp. Syst. (FISTS)*, Delft, The Netherlands, Nov. 2020, pp. 156–161, doi: 10.1109/FISTS46898.2020.9264885.

[107] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 62, no. 2, p. 1805, 2000.

[108] D. Thekke Kanapram, F. Patrone, P. Marin-Plaza, M. Marchese, E. L. Bodanese, L. Marcenaro, D. Martín Gómez, and C. Regazzoni, "Collective awareness for abnormality detection in connected autonomous vehicles," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3774–3789, May 2020, doi: 10.1109/JIOT.2020.2974680.

[109] D. Mori, H. Sugiura, and Y. Hattori, "Adaptive sensor fault detection and isolation using unscented Kalman filter for vehicle positioning," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 1298–1304.

[110] X. Han, Y. Zhou, K. Chen, H. Qiu, M. Qiu, Y. Liu, and T. Zhang, "ADS-lead: Lifelong anomaly detection in autonomous driving systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 1, pp. 1039–1051, Jan. 2023, doi: 10.1109/TITS.2021.3122906.

[111] Y. Wang, N. Masoud, and A. Khojandi, "Real-time sensor anomaly detection and recovery in connected automated vehicle sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1411–1421, Mar. 2021, doi: 10.1109/TITS.2020.2970295.

[112] Z. Yang, J. Ying, J. Shen, Y. Feng, Q. A. Chen, Z. M. Mao, and H. X. Liu, "Anomaly detection against GPS spoofing attacks on connected and autonomous vehicles using learning from demonstration," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 9, pp. 9462–9475, Sep. 2023, doi: 10.1109/TITS.2023.3269029.

[113] J. Liu and J.-M. Park, "'Seeing is not always believing': Detecting perception error attacks against autonomous vehicles," *IEEE Trans. Dependable Secure Comput.*, vol. 18, no. 5, pp. 2209–2223, Sep./Oct. 2021.

[114] R. Prophet, J. Martinez, J. F. Michel, R. Ebelt, I. Weber, and M. Vossiek, "Instantaneous ghost detection identification in automotive scenarios," in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2019, pp. 1–6.

[115] A. Khalil and M. Al Janaideh, "On fault classification in connected autonomous vehicles using supervised machine learning," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Prague, Czech Republic, Sep. 2021, pp. 1198–1204, doi: 10.1109/IROS51168.2021.9636741.

[116] A. A. Alsulami, Q. A. Al-Haija, A. Alqahtani, and R. Alsini, "Symmetrical simulation scheme for anomaly detection in autonomous vehicles based on LSTM model," *Symmetry*, vol. 14, no. 7, p. 1450, Jul. 2022.

[117] F. van Wyk, Y. Wang, A. Khojandi, and N. Masoud, "Real-time sensor anomaly detection and identification in automated vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1264–1276, Mar. 2020, doi: 10.1109/TITS.2019.2906038.

[118] A. R. Javed, M. Usman, S. U. Rehman, M. U. Khan, and M. S. Haghighi, "Anomaly detection in automated vehicles using multistage attention-based convolutional neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4291–4300, Jul. 2021, doi: 10.1109/TITS.2020.3025875.

[119] D. Bezzina and J. Sayer, "Safety pilot model deployment: Test conductor team report," Dept. U.S. Dept. Transp., Univ. Michigan Library, Washington, DC, USA, Tech. Rep. DOT HS 812 171, 2014.

[120] S. Rajendar and V. K. Kaliappan, "Sensor data based anomaly detection in autonomous vehicles using modified convolutional neural network," *Intell. Autom. Soft Comput.*, vol. 32, no. 2, pp. 859–875, 2022.

[121] L. Wang and X. Zhang, "Anomaly detection for automated vehicles integrating continuous wavelet transform and convolutional neural network," *Appl. Sci.*, vol. 13, no. 9, p. 5525, Apr. 2023.

[122] *Open Sourcing 223gb of Driving Data*. [Online]. Available: https://medium.com/udacity/open-sourcing-223gb-of-mountain-view-driving-data-f6b5593fbfa5

[123] R. Oucheikh, M. Fri, F. Fedouaki, and M. Hain, "Deep real-time anomaly detection for connected autonomous vehicles," *Proc. Comput. Sci.*, vol. 177, pp. 456–461, Jan. 2020.

[124] S. Rezaei, A. Khojandi, and N. Masoud, "GAAD: GAN-enabled autoencoder for real-time sensor anomaly detection and recovery in autonomous driving," 2021, doi: 10.13140/RG.2.2.20707.68646/3.

[125] J. Watts, F. van Wyk, S. Rezaei, Y. Wang, N. Masoud, and A. Khojandi, "A dynamic deep reinforcement learning-Bayesian framework for anomaly detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 22884–22894, Dec. 2022, doi: 10.1109/TITS.2022.3200906.

[126] D. Bogdoll, E. Eisen, M. Nitsche, C. Scheib, and J. M. Zöllner, "Multimodal detection of unknown objects on roads for autonomous driving," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Prague, Czech Republic, Oct. 2022, pp. 325–332, doi: 10.1109/SMC53654.2022.9945211.

[127] K. Jin, F. Mu, X. Han, G. Wang, and Z. Liu, "Anomaly detection for robust autonomous navigation," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, London, U.K., May 2023, pp. 10026–10032, doi: 10.1109/ICRA48891.2023.10161507.

[128] T. Griebel, D. Authaler, M. Horn, M. Henning, M. Buchholz, and K. Dietmayer, "Anomaly detection in radar data using PointNets," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, Sep. 2021, pp. 2667–2673, doi: 10.1109/ITSC48978.2021.9564730.

[129] G. Di Biase, H. Blum, R. Siegwart, and C. Cadena, "Pixel-wise anomaly detection in complex driving scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16918–16927.

[130] I. A. Khan, N. Moustafa, D. Pi, W. Haider, B. Li, and A. Jolfaei, "An enhanced multi-stage deep learning framework for detecting malicious activities from autonomous vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 25469–25478, Dec. 2022, doi: 10.1109/TITS.2021.3105834.

[131] E. Seo, H. M. Song, and H. K. Kim, "GIDS: GAN based intrusion detection system for in-vehicle network," in *Proc. 16th Annu. Conf. Privacy, Secur. Trust (PST)*, Belfast, Ireland, Aug. 2018, pp. 1–6, doi: 10.1109/PST.2018.8514157.

[132] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *Proc. Mil. Commun. Inf. Syst. Conf. (MilCIS)*, Canberra, ACT, Australia, Nov. 2015, pp. 1–6, doi: 10.1109/MilCIS.2015.7348942.

[133] J. Hao and G. Han, "On the modeling of automotive security: A survey of methods and perspectives," *Future Internet*, vol. 12, no. 11, p. 198, Nov. 2020. [Online]. Available: https://www.mdpi.com/1999-5903/12/11/198

[134] Z. Abuabed, A. Alsadeh, and A. Taweel, "STRIDE threat model-based framework for assessing the vulnerabilities of modern vehicles," *Comput. Secur.*, vol. 133, Oct. 2023, Art. no. 103391.

[135] K. Yang, X. Tang, J. Li, H. Wang, G. Zhong, J. Chen, and D. Cao, "Uncertainties in onboard algorithms for autonomous vehicles: Challenges, mitigation, and perspectives," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 9, pp. 8963–8987, Sep. 2023, doi: 10.1109/TITS.2023.3270887.

[136] S. Ghosh, A. Zaboli, J. Hong, and J. Kwon, "An integrated approach of threat analysis for autonomous vehicles perception system," *IEEE Access*, vol. 11, pp. 14752–14777, 2023, doi: 10.1109/ACCESS.2023.3243906.

SIHEM BACCARI received the bachelor's degree in computer sciences and the master's (Research) degree in computer sciences and networks from the Faculty of Sciences, University of Gabes, Tunisia, in 2018 and December 2020, respectively. Since June 2021, she has been a Blockchain Developer with T+ Company, France. She is currently a Research Consultant in collaboration with Abu Dhabi University. Her research interests include vehicular networks, including autonomous vehicles, cybersecurity, and blockchain technology and its applications.

MOHAMED HADDED received the joint Ph.D. degree in computer science engineering from the Telecom SudParis College in co-accreditation with Pierre and Marie Curie Campus (Sorbonne University), in 2016. He is currently with the CSIT Department, College of Engineering, Abu Dhabi University, as an Assistant Professor in cybersecurity engineering. Prior to joining ADU, he was a Senior Cybersecurity Researcher in intelligent transportation systems with IRT SystemX Paris, France, from 2021 to 2022. From 2018 to 2021, he was a Cybersecurity Research Engineer with the VEDECOM Institute, Versailles, France. Before that, he was a Postdoctoral Research Fellow with the National Institute for Research in Digital Science and Technology (INRIA), France, from 2017 to 2018. From 2015 to 2017, he was a Teaching and Research Assistant (ATER) with Paris Descartes University and University of Franche-Comté (UFC).

HAKIM GHAZZAI (Senior Member, IEEE) received the Diplome d'Ingenieur degree (Hons.) in telecommunication engineering and the master's degree in high-rate transmission systems from École Supérieure des Communications de Tunis (SUP'COM), Tunis, Tunisia, in 2010 and 2011, respectively, and the Ph.D. degree in electrical engineering from the King Abdullah University of Science and Technology (KAUST), Saudi Arabia, in 2015. He was a Researcher Scholar with the Qatar Mobility Innovations Center (QMIC), Qatar, Karlstad University, Sweden; and Stevens Institute of Technology, NJ, USA. He is currently a Research Scientist with KAUST. He is the author or coauthor of more than 170 publications. His general research interests include artificial intelligence enabled applications, the Internet of Things, intelligent transportation systems (ITS), and mobile and wireless networks. Since 2019, he has been on the editorial board of the IEEE COMMUNICATIONS LETTERS and the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY. Since 2020, he has been joining the Board of IoT and Sensor Networks (specialty section of *Frontiers in Communications and Networks*) as an Associate Editor. He was a recipient of appreciation for an Exemplary Reviewer of IEEE WIRELESS COMMUNICATIONS LETTERS, in 2016, and IEEE COMMUNICATIONS LETTERS, in 2017.

HAIFA TOUATI received the Engineering and master's degrees in network and the Ph.D. degree in computer science from the National School of Computer Science (ENSI-Tunisia), in 2011, and the H.D.R. degree, in 2021. She is currently an Associate Professor in computer science with the Faculty of Sciences of Gabes (FSG-Tunisia), where she was a Supervisor with the Master's Degree Program in Computer Science and Networking. Additionally, she is also the Director of the IReSCoMath Research Laboratory. Her research interests include named data networking, vehicular communications, security, machine learning, and blockchain technology.

MOURAD ELHADEF (Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science from L'Institut Supérieur de Gestion, Tunis, Tunisia, and the Ph.D. degree in computer science from the University of Sherbrooke, Québec, Canada. He has over 100 publications in international refereed journals and conference proceedings. His current research interests include distributed and parallel computing, wireless mobile ad-hoc, mesh, sensor networks, fault tolerance and fault diagnosis, artificial intelligence, security, and intelligent vehicle-based ad-hoc networks (inVANETs). He is an active Reviewer of various international conferences and journals, such as IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS and *Journal of Parallel and Distributed Computing*.