Safe and Secure Autonomy by Machine Learning Techniques: A Systematic Literature Review

Afshin Hassani[a], Mehran Alidoost Nia[b], Reza Ebrahimi Atani[a]

[a]*Department of Computer Engineering, Faculty of Engineering, University of Guilan, Rasht, Iran,*
[b]*Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran,*

Abstract

This systematic literature review investigates recent advancements in machine learning techniques aimed at ensuring safety and security in autonomous systems. By analyzing 129 scholarly articles published between 2018 and 2024, we identify dominant methodologies, prominent machine learning techniques, and key application domains. As the paper illustrates, Reinforcement Learning (RL) and Deep Learning (DL) have emerged as leading approaches, especially in contexts requiring real-time adaptation such as autonomous vehicles, Unmanned Aerial Vehicles (UAVs) and robotics. Our analysis reveals significant research trends, including a strong emphasis on safe motion, prediction, vulnerability detection, and security assurance. We highlight current research gaps, notably the need for standardized benchmarks and improved model robustness under adversarial conditions. This review provides valuable insights and future research directions, serving as a comprehensive guide for young researchers entering the field of safe and secure autonomy.

Keywords: Safe Autonomy, Autonomous Vehicle, Machine Learning, Security

1. Introduction

In recent years, the benefits of autonomous systems in critical domains, such as transportation, healthcare, robotic systems, and industrial automation, necessitate the need for robust safety and security measures. Autonomous systems operate with minimal human intervention, relying heavily on sensor data, machine learning, and real-time decision-making processes, which introduces unique vulnerabilities. Ensuring safety means preventing hazards from humans and the environment by guaranteeing predictable, reliable operation even in uncertain conditions (1). Security, on the other hand, is crucial to protect these systems from malicious attacks, which can lead to system failures, compromised data, or unintended behaviors (2). The potential consequences of these system failures can be severe. Safety and security are fundamental to assure public trust and facilitating the widespread adoption of autonomous technologies. These make them pivots in the development of resilient and trustworthy autonomous systems.

Safety hazards in autonomous systems are diverse and can lead to significant risks for users and the surrounding environment (3). In autonomous vehicles, for instance, sensor malfunctions or poor data interpretation can fail to detect obstacles, pedestrians, or road signals, which may lead to accidents (4). A high-profile example is misinterpreting objects or environmental conditions, which may lead to severe collisions. Similarly, robotic-assisted surgical systems in healthcare rely on precise control and feedback (5); however, if a system experiences delays or errors in the data processing stage, it may result in incorrect incisions or other critical mistakes that endanger patient safety. Industrial robots also pose safety risks: unexpected movements or malfunctions can endanger human workers nearby (6). These examples illustrate the varied and potentially severe hazards associated with autonomous systems, underscoring the urgent need for rigorous safety mechanisms to detect and mitigate failures in real-world conditions.

Security hazards in autonomous systems pose significant risks as they can be exploited to manipulate system behavior or compromise sensitive data. The vulnerability of these systems is an open challenge (7). In autonomous vehicles, for instance, researchers have demonstrated that cyber attackers can remotely take control of steering or braking functions, as seen in cases where hackers accessed the control systems of cars via their internet-connected interfaces. Such incidents affect passengers, other vehicles, and pedestrians on the road (8). Similarly, autonomous robotic systems used in surgeries and patient monitoring are vulnerable to unauthorized access in healthcare. These systems breaches could lead to harmful manipulations or unauthorized data collection, jeopardizing patient safety and privacy. Industrial autonomous systems also face threats, as attackers could gain control of robotic arms or automated machinery, leading to operational disruptions or physical harm to employees (9). These cases illustrate the critical need for robust cybersecurity measures in autonomous systems to prevent malicious exploitation (11) and safeguard users (10) and the broader environment.

Without safety, an autonomous system's unpredictable behavior could lead to accidents, equipment damage, or even human injury. Similarly, these systems become vulnerable to cyber-attacks without security, which can cause intentional disruptions, data breaches, and unauthorized control, undermining safety and trust. However, the potential of machine

learning techniques to adapt to new threats in real time is a game-changer tool (12). By enabling autonomous systems to recognize patterns, predict risks, and adapt to new threats in real time, machine learning plays a decisive role in reinforcing these properties. For example, anomaly detection algorithms can identify unusual sensor readings to preemptively signal safety issues, while reinforcement learning can help systems autonomously optimize safety measures under dynamic conditions (13). On the security front, machine learning-based intrusion detection systems can recognize and respond to sophisticated cyber threats as they evolve. Through such capabilities, machine learning enhances safety and security and contributes to the reliability and resilience of autonomous systems in complex environments (14).

In this research, we aim to investigate the recent efforts in the field of autonomous vehicles in depth to guarantee safety and security under different scenarios and domains. We reveal the current research trends, open problems and gaps using a systematic literature review approach. This paper is highly recommended to young researchers in autonomous systems who want to solve a security and safety problem using a machine learning technique. We analyze the tools, approaches, techniques and solutions used to address autonomous systems' safety and security issues. The main contributions of the research are as follows:

- We systematically review the recent advancements in safe and secure autonomous systems using machine learning techniques. To the best of our knowledge, this is the first systematic review focused on this filed of study.
- The research highlights current gaps and challenges in machine learning approaches for autonomous system safety and security.
- We provide practical insights from real-world case studies on machine learning applications in autonomous vehicles.
- We study the relationship between research purposes, methodologies, and machine learning techniques, and reveal interesting facts for continuing the research in the filed of safe and secure autonomy by machine learning techniques.

The structure of the paper is as follows. In the next section, we review the technical background of the paper. In Section 3, we present the research questions and the criteria of the research. In Section 4, we analyze the data extracted from the papers and give answers to each research question. Section 5 is dedicated to discussing the results and giving insight into the research trends. Finally, the paper concludes in Section 6.

2. Background and Review Focus

In this section, we provide an overview of several interrelated concepts fundamental to our research. We begin by briefly introducing Machine Learning (ML) and its various branches, crucial for understanding their application in complex systems. We then transition to discussing self-adaptive systems, particularly focusing on the Monitor-Analyze-Plan-Execute and Knowledge (MAPE-K) framework essential for runtime decision-making (15). Following this, we define safety in the context of autonomous systems and explain what constitutes *safe autonomy*. Lastly, we highlight the critical integration of machine learning in self-adaptive systems, highlighting its importance for enhancing safety and reliability.

2.1. Machine Learning

ML is a branch of artificial intelligence (AI) and computer science focusing on developing methods for computers to learn and improve performance. It aims to replicate human learning processes, leading to gradual improvements in accuracy for specific tasks. The main goals of ML include classifying data based on developed models (like detecting spam emails) and making predictions about future outcomes (like predicting house prices).

Supervised learning, as a branch of ML, is based on the principle of learning from labeled datasets. In this approach, the algorithm is exposed to input data paired with the correct output, which could be a category label in classification tasks or a continuous value in regression tasks (16). This setup allows the algorithm to learn a mapping from inputs to outputs through repeated exposure to various examples. The primary goal is for the algorithm to make accurate predictions or decisions when presented with new, unseen data. This learning methodology is fundamental in applications like image and speech recognition, where the model needs to accurately identify and classify input based on learned patterns from the training data.

In practical scenarios, supervised learning algorithms undergo a training phase where they adjust their parameters (such as weights in neural networks) to minimize the error between their predictions and actual outcomes. After training, the performance of these models is evaluated using a separate test dataset, which helps gauge their ability to generalize to new data. The efficiency and accuracy of supervised learning models heavily rely on the quality, volume, and representativeness of the training data. Challenges in supervised learning include overfitting, where a model learns the training data too well but performs poorly on new data, and the need for extensive labeled datasets, which can be resource-intensive to compile (85).

Unsupervised learning diverges from its supervised counterpart by learning from data that is not labeled. In this approach, algorithms sift through data to discover underlying patterns and groupings without explicit output guidance, focusing on the intrinsic structure of the data (17). This type of learning is pivotal in scenarios where the relationships within data are not known in advance. Techniques like clustering, where data points with similar characteristics are grouped, and dimensionality reduction, where complex data is simplified while preserving essential features, are quintessential examples of unsupervised learning. These methods are invaluable in exploratory data analysis, helping to reveal natural groupings, detect anomalies, or reduce the complexity of datasets for further analysis.

Unsupervised learning is particularly useful in understanding complex, unstructured data where direct supervision is infeasible. For instance, in market segmentation, unsupervised learn-

ing can identify customer groups with similar buying patterns without predefined categories. Similarly, in image processing, it can detect patterns or features without reference to labeled outcomes. The challenge in unsupervised learning lies in determining the effectiveness of the model, as there is no straightforward way to measure accuracy without labeled outcomes. Despite this, its ability to uncover hidden structures in data makes it a powerful tool in the ML toolkit, providing insights that can guide further data exploration and decision-making (125).

Within the realm of modern Reinforcement Learning (RL) (18), which is intricately modeled as a Markov Decision Process (MDP) (19), the dynamics of learning are characterized by the interaction of an agent with its environment, denoted as E, in discrete time steps. At each time step, denoted as t, the agent perceives the state of the environment, St, which is a part of the set S representing all possible states within E. Critical to this interaction is the function P(s, r — St, At), which signifies the conditional joint probability of transitioning into a new state s and receiving a reward r, given the agent's action At in the current state (33).

Reinforcement Learning, in this framework, emphasizes the agent's role in learning from the consequences of its actions. Unlike other machine learning paradigms where learning is driven by data, RL is driven by the agent's experience obtained from interacting with the environment. This experience is quantified through rewards (or penalties), which the agent seeks to maximize (or minimize) over time. The agent's decision-making strategy, known as a policy, is shaped by these experiences, leading to a learned behavior that is aimed at achieving the best possible outcome in a given situation. The agent's learning process involves exploration, where it tries out different actions to discover their effects, and exploitation, where it uses its knowledge of the effects of actions to make better decisions. This balance between exploration and exploitation is crucial in RL and is often a key area of research and development.

Deep Learning, a subfield of machine learning, involves algorithms inspired by the structure and function of the brain called artificial neural networks. It is characterized by its use of deep neural networks with multiple layers; each layer processes a different aspect of the data, leading to an intricate hierarchy of learned features. The term *deep* usually refers to the number of layers in the neural networks. Traditional neural networks might contain 2-3 layers, while deep networks can have hundreds (20).

Deep Learning automates much of the feature extraction process, eliminating the need for manual intervention. This is particularly advantageous in fields such as image and speech recognition, where conventional machine learning techniques struggle with the complexity and variability of the data. In these applications, deep learning models can learn to identify patterns and features that are predictive of certain outcomes or classifications (108). Training deep learning models typically requires a large amount of labeled data and significant computational power. The training process involves adjusting the parameters of the neural networks based on the feedback from a performance metric, using algorithms like backpropagation and optimization techniques like gradient descent.

2.2. Self-Adaptive Systems

In design of autonomous systems, the evolution towards more dynamic and flexible architectures has led to the emergence of self-adaptive systems, based on the principle of autonomic computing. Inspiring by the human body's autonomic nervous system, which operates without conscious oversight, autonomic computing introduces the concept of self-managing computing resources (21). These systems are designed to handle complex, distributed environments with minimal human intervention, aligning with the administrator's objectives and adapting online to changing conditions (22). These systems are designed to automatically adjust their operations to meet certain objectives, which may include improving performance, conserving resources, or maintaining reliability. Self-adaptation is particularly important in dynamic environments, where conditions can change unpredictably or beyond the system's initial design parameters (139).

The concept of self-adaptation in systems draws from various fields, including control theory, artificial intelligence, and software engineering. These systems typically involve a feedback loop where the system continuously monitors its environment and its own performance, analyzes this data to detect changes or anomalies, plans or decides the next course of action, executes these adaptations over a shared Knowledge base. This cycle, often referred to as the Monitor-Analyze-Plan-Execute over a shared Knowledge base (MAPE-K) loop, allows the system to be resilient and maintain optimal functioning despite external or internal changes.

The loop begins with the *Monitoring* phase, where systems continuously collect data about their performance and environment. This is crucial in contexts like cybersecurity (e.g., (62; 64)) where network traffic is monitored for potential threats. The *Analysis* phase then processes this data, often using advanced algorithms, to extract insights. This phase is essential in autonomous vehicles (56) for understanding environmental conditions.

The *Planning* phase involves formulating a strategy based on the analysis, crucial for dynamic decision-making in autonomous systems (30; 31). The *Execution* phase implements this strategy, adapting the system's behavior in real-time to maintain safety and efficiency.

A significant part of these phases is the *Knowledge* component, representing the system's repository of historical data, policies, and models. Knowledge informs all other phases of the loop (23). Applications of self-adaptive systems are found in various domains, from autonomous vehicles and smart grids to software applications and network security, where the ability to adapt in real-time is crucial.

The relationship between ML and the Monitor-Analyze-Plan-Execute over a shared Knowledge base (MAPE-K) framework in self-adaptive systems is very close. ML techniques can enhance each component of the MAPE-K loop. For instance, in the Monitoring phase, ML can be used for efficient data collection and anomaly detection. In the Analyzing phase,

ML algorithms can interpret complex data to identify patterns or predict trends. During Planning, ML can inform decision-making for system adaptations, and in Execution, ML can optimize the implementation of these adaptations. The shared Knowledge base benefits from ML by becoming more robust and informed, enhancing the system's overall adaptability and responsiveness (20).

Runtime decision-making in self-adaptive systems is a process where the system autonomously evaluates and responds to changes in its operational environment while it is running (24). This decision-making is crucial for maintaining optimal performance and reliability, especially in dynamic or unpredictable environments. It involves the system continuously monitoring its own performance and the external environment, analyzing this data to identify any necessary changes, and then making decisions about how to adapt in real-time.

The effectiveness of runtime decision-making is enhanced by integrating advanced technologies such as ML (25; 26). ML algorithms can analyze complex and large datasets quickly, predict potential future states, and make informed decisions to adapt the system's behavior accordingly. This ability to make intelligent, real-time decisions allows self-adaptive systems to operate efficiently and effectively, even under changing conditions, and reduces the need for human intervention in system operations.

2.3. Safe-Autonomy

Safety in engineered systems refers to the assurance that engineered systems, particularly cyber-physical systems, operate without causing unacceptable risk of harm to people or damage to the environment. This involves ensuring that these systems reliably and consistently perform their intended functions under designated conditions, while avoiding unintended consequences that could lead to accidents or failures (28).

Safe autonomy in the context of self-adaptive systems and ML, refers to the absence of unacceptable risks of physical injury or damage to the health of individuals and damage to property or the environment (27). Safe autonomy implies that autonomous systems are designed and operated in ways that mitigate risks, ensure reliability, and prevent failures that could lead to unsafe situations.

The integration of ML in self-adaptive systems is central to achieving safe autonomy (20). ML algorithms contribute by enhancing the system's ability to perceive, understand, and adapt to complex and dynamic environments. This adaptability is crucial for maintaining safety standards, especially in scenarios where operational conditions are unpredictable or rapidly changing. For instance, in autonomous vehicles, ML-driven systems continuously process vast amounts of sensor data to make runtime decisions, ensuring safe navigation in varying traffic conditions.

Imagine an autonomous vehicle starts a journey along a track that presents three distinct road conditions: dry, wet, and icy. This scenario illustrates the challenges of maintaining safety in the face of environmental uncertainties and the critical role of safe autonomy in advanced vehicular technology. Initially, the vehicle traverses a dry section, where traction and visibility are best. Here, the vehicle's control systems are set to standard parameters, suitable for dry conditions. As the vehicle progresses, it encounters a wet surface, a condition that reduces tire grip and increases the likelihood of skidding. The car's advanced sensor array detects the change in road condition. In response, the autonomous system dynamically adjusts the vehicle's speed, braking sensitivity, and steering responsiveness to adapt to the wet conditions. The most challenging part of the journey comes when the vehicle approaches an icy section of the track. This condition presents the highest degree of uncertainty and demands precise adaptations. The vehicle's control system, equipped with sophisticated algorithms, detects the icy surface and implements a series of critical adjustments. It further reduces speed, modifies traction control settings, and adjusts braking patterns to counteract the reduced friction. This case study illustrates the imperative of safe autonomy in autonomous vehicles, especially when confronted with the unpredictability of varying road conditions. It necessitates for these vehicles to be equipped with advanced, context-aware systems capable of real-time adaptations. Such capabilities are essential for ensuring passenger safety and vehicle stability across a spectrum of environmental conditions (56).

3. Research Criteria

This SLR is structured according to the guidelines proposed by Kitchenham (29), tailored to explore the domain of safe and secure autonomy via machine learning techniques. The primary goal is to investigate recent research in this field, identify current challenges, trends, and case studies, and highlight insights for future research in autonomous computing.

3.1. Research Questions

We formulated the goal of the study using the classic Goal-Question-Metric (GQM) approach:

- Purpose: analyze and characterize.
- Issue: the use of machine learning.
- Object: to support self-adaptation based on MAPE-K feedback loops.
- Viewpoint: from a researcher's viewpoint.

We translated the overall goal of this SLR to seven concrete research questions:

- **RQ1.** What are the major research purposes, methodologies and findings in safe autonomy by machine learning techniques?
- **RQ2.** What are the main machine learning techniques used in developing safe autonomy where research is being made?
- **RQ3.** Which case-studies are concerned in the research papers?
- **RQ4.** Which types of requirements are addressed in the studies?

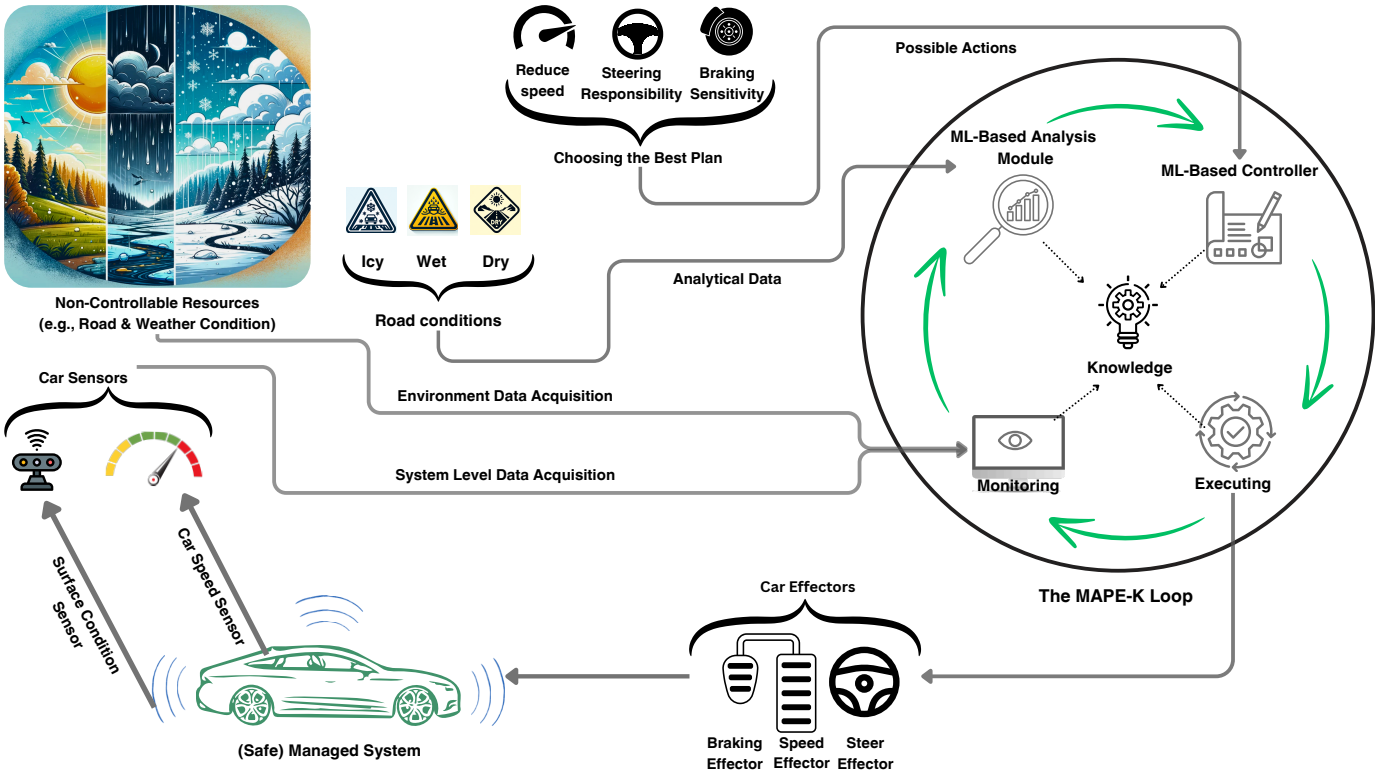

Figure 1: A general perspective on assuring autonomy by machine learning.

- **RQ5.** What are the limitations of current research or main problems that should have been addressed?
- **RQ6.** What are the future research directions that are worth investigating?
- **RQ7.** What are the highly-cited articles in safe autonomy?

With RQ1, we want to understand the primary goals or objectives of existing research in the field, the methods used to achieve these goals, and the key outcomes or findings reported. We aim to provide a comprehensive overview of the current state of research in safe autonomy as achieved through machine learning. With RQ2, we want to focuses on identifying the specific machine learning algorithms, models, or techniques that are most commonly employed in the development of safe autonomous systems. To answer it, we categorize and understand the technical approaches in use. With RQ3, we target uncover real-world applications or instances where the principles of safe autonomy via machine learning have been applied. Answering RQ2 elicits from specific case studies or examples cited in the literature. With RQ4, we need to understand the different requirements (safety or security) that are considered essential in the research and development of safe autonomous systems using machine learning techniques. With RQ5, we identify the gaps, challenges, or limitations in the current body of research. It's about understanding what's lacking, what problems are yet to be solved, and where the research may have fallen short. With RQ6, we want to identify areas for future research. It's about understanding where the research is heading, what unanswered questions remain, and what new or emerging topics should be considered for exploration. By answering RQ7, we identify key influential studies in the field of safe autonomy. This involves looking at articles that have been cited frequently, indicating their impact and importance in shaping the research landscape.

3.2. *Searched Sources and Search Query*

Our search strategy for the systematic literature review is a combination of automated and manual searches. The primary objective is to comprehensively cover the field of self-adaptation in machine learning within the specified time frame. Our search is conducted across four major digital libraries, known for their extensive collections in computer science and engineering research: IEEE Xplore, ACM Digital Library, Springer Link, and Elsevier. The search string used is:

```
"Safe autonomous systems by machine learning" OR
"Safe self-adaptive systems by machine learning" OR
"Secure autonomous systems by machine learning" OR
"Software-controlled machine learning"
```

This search string was applied across the selected databases to retrieve articles from 1 May 2018 to 31 December 2024, ensuring the focus on recent developments in the field.

3.3. *Inclusion and Exclusion Criteria*

We use the following inclusion criteria to select papers:

- All papers must be published between 1 May 2018 and 31 December 2024.
- Papers must be published in a journal or conference proceeding.
- The focus must be on safe autonomy by machine learning.
- The papers have to address a few issues regarding safe autonomy by machine learning and propose a few solutions to that problem.
- The papers indicate future research directions to be investigated.

We use the following exclusion criteria:

- Book chapters.
- Patents.
- Papers that do not address some issue regarding safe autonomy.
- The papers do not indicate future directions or problems to be addressed.
- Survey papers.

A paper is selected if it meets all inclusion criteria and does not meet any exclusion criterion.

3.4. Data Items

To answer the research questions, we define a set of data items to extract basic materials from the papers. Table 1 gives an overview of the data items that we briefly discuss now. The concrete options for each data item are further discussed in the next section.

V1: This column refers to the database or repository where the article is accessed, providing insight into the origin and accessibility of the research.
V2: It lists the official title of the article, which gives a quick overview of the study's focus.
V3: This column contains the names of the article's authors, indicating the contributors and their affiliations.
V4: The country or countries associated with the authors' affiliations are mentioned, which reveals the geographical context of the research.
V5: This indicates whether the publication is a journal article or a conference paper, differentiating the type of publication.
V6: It specifies the name of the journal or conference where the article is published, providing context about the publication's reach and significance.
V7: The year of publication is listed, helping to map the timeline of the research development.
V8: This column shows the citation count, a key indicator of the article's impact in the field.
V9: It mentions the platform or tool where the citation count was obtained, ensuring transparency in citation analysis.
V10: Lists the total number of references in each article, indicating the depth of research involved.
V11: This differentiates between research-focused articles and reviews, providing clarity on the nature of the content.
V12: Highlights the primary subject or field of study discussed in each article, providing clarity on the central theme of the research.
V13: Summarizes the primary objectives and goals of each article, offering a clear understanding of the research's main intentions and focus.
V14: Provides an overview of the research methods, approaches, or frameworks employed in each article, offering insights into the study's investigative techniques.
V15: Specifies the machine learning methods, algorithms, or approaches used in each article, which gives insights into the specific techniques applied in the research.
V16: Lists the machine learning software, tools, or frameworks mentioned in each article, providing an overview of the specific technologies utilized in each research paper.
V17: It indicates whether the article includes a case study or practical application. It gives insights into real-world examples or implementations discussed in the research.
V18: It identifies the specific requirements discussed in each article, focusing on aspects related to safety, security, or both. It provides insight into the core prerequisites considered in the research.
V19: This column identifies the limitations and challenges noted in each study, such as methodological constraints, scope boundaries, or data limitations. It provides insights into the potential impacts on the study's validity and areas for further research.
V20: It highlights the areas for future research or further investigation as suggested in each article, pointing towards potential advancements and open questions in the field.
V21: In terms of quality score, it reflects the quality of the research reporting in the paper, which is critical for analyzing and interpreting the data. In line with Reference (20), our assessment focused on the following quality items: (1) Main topic area, Research purpose, and Methodology, (2) Machine Learning Technique(s), (3) Case-Study, (4) Requirements, (5) Limitations of the study, (6) Future work, (7) Number of citations. In order to assess the quality of each item, we consider three levels: explicit description (2 points), general description (1 point), and no description (0 points). Based on all items scored in a study, we calculate a quality assessment score (max 14), which is dicsussed in the next section.

4. Analytical Results

We now present the results of our systematic literature review (SLR). We begin with an overview of general findings, followed by detailed results organized according to each research question addressed in this study.

4.1. Demographics

Our systematic search strategy initially yields a total of 420 papers. After a meticulous application of our inclusion and exclusion criteria (screening process), we distilled this number down to 129 articles that met the SLR's requirements for quality and relevance. These articles were disseminated across 75

Table 1: Data Extraction Items.

Item ID	Item	Use	Description
V1	Source	Documentation	Scientific Databases
V2	Title	Documentation	Article Title
V3	Authors	Documentation	Name of Article Authors
V4	Country	Documentation	Authors Countries
V5	Journal/Conference	Documentation	Type of Publication
V6	Journal/Conference Name	Documentation	Publication Name/Venue
V7	Year	Documentation	Publication Year
V8	Number of cites	RQ7	Citation Count
V9	Source of cites	RQ7	Citation Source
V10	Number of References	Documentation	Reference Count
V11	Type of Paper	Documentation	Research or Review
V12	Main topic area	RQ1	Research Subject
V13	Research purpose	RQ1	Purpose of Article
V14	Methodology	RQ1	Article Methodology
V15	Machine Learning Technique(s)	RQ2	ML Techniques used in the article
V16	Machine Learning Tools	RQ2	ML tools used in the article
V17	Case-Study	RQ3	Case study in the article
V18	Requirements	RQ4	Indicates article is about safety, security, or both of them
V19	Limitations	RQ5	Shows the limitations of the article
V20	Future work	RQ6	Future Work Suggestions
V21	Quality score	RQ1-6	Measuring quality of the reporting

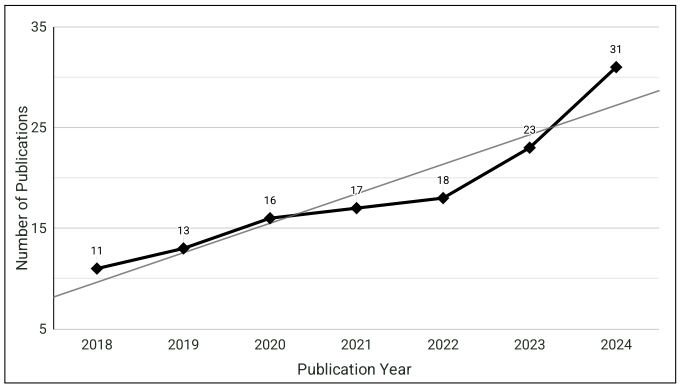

Figure 2: Distribution of the published papers over the years. As denoted, the trend line illustrates a certain demand in assuring autonomy by machine learning techniques over the recent years.

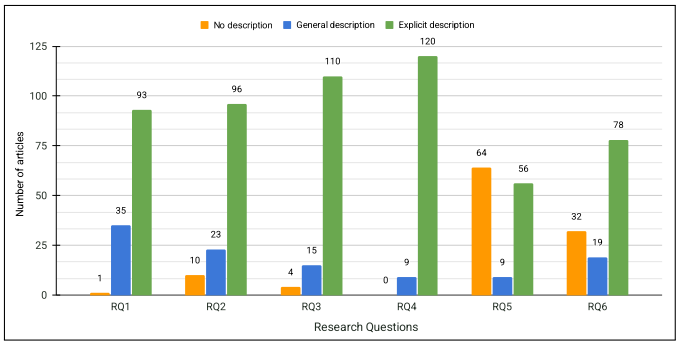

Figure 3: Quality scores of the papers are evaluated by three categories: "no description", "general description" and "explicit description". We calculate the quality scores for each research question, separately, to show how much the papers explicitly share the scientific contents with the readers.

distinct publication venues, comprising both journals and conferences, which reflects the interdisciplinary interest and academic engagement in the field.

The temporal spread of the selected articles, as depicted in Figure 2, indicates a notable concentration of research activity between the years 2018 to 2024. This period shows a dynamic phase of development and heightened interest in the application of machine learning techniques to improve the safety mechanisms within autonomous systems. In addition, Figure 4 denotes the distribution of the countries that have major contributions on this filed of study. This can be good for finding new researchers and institutions to follow the relevant research topics.

The overall quality of the research within the selected articles is summarized by Figure 3. Our analysis indicates that most articles offer a clear problem statement and contextualize their research within the existing literature effectively. They typically present their contributions with clarity and discuss the implications of their findings, though some variations exist in how explicitly these discussions are conducted. A recurring observation is the insufficient reporting on the limitations of the studies. Totally, we have observed 76 conference papers and 53 journal articles, with average quality score of 9 and 10.1, respectively. It shows the strength and higher quality of the journal papers published in this field of study, where the total average quality score is about 9.4 for all 129 collected papers with standard deviation of 1.8.

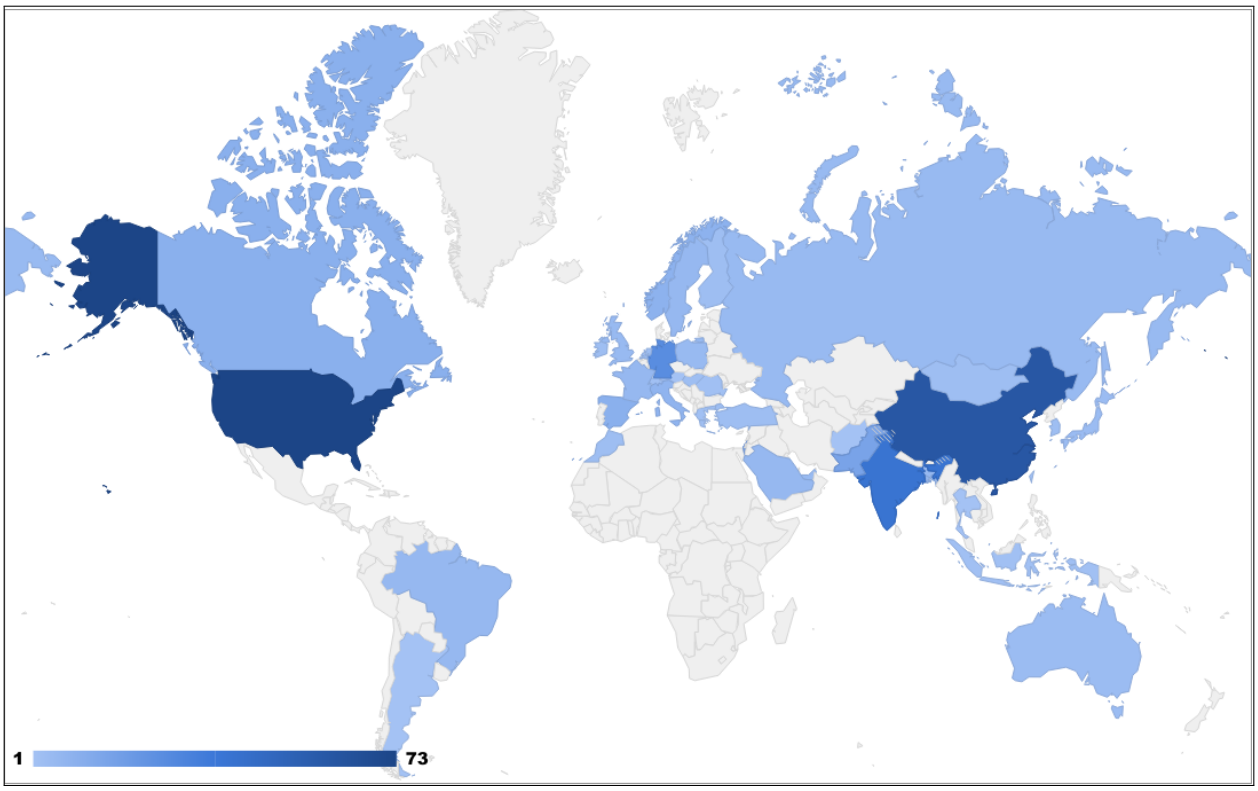

Figure 4: Distribution of authors' country.

4.2. RQ1: What are the major research purposes, methodologies and findings in safe autonomy by machine learning techniques?

In order to answer this research question, the following data items are analyzed: main topic area (V12), research purpose (V13), and methodology (V14). We review each item separately.

Main Topic Areas. In the context of safe autonomy using machine learning techniques, main topic areas refer to the specific domains and sub-fields that researchers focus on. These topics encompass various aspects of ensuring autonomous systems operate safely. Table 2 denotes the main topic areas extracted from the papers, illustrated with examples. The last column shows the number of papers of each type.

Research Purpose. The research purposes refer to the primary goals and objectives that studies aim to achieve within the field of safe autonomy using machine learning techniques. These purposes address various aspects such as enhancing system reliability, improving decision-making accuracy, and ensuring compliance with safety regulations. Table 3 lists the research purposes identified from the papers, illustrated with examples. The right most column shows the number of papers of each type.

Methodologies. Methodologies in the field of safe autonomy using machine learning techniques refer to the approaches and techniques used to conduct research and achieve the set objectives. These methodologies include various machine learning algorithms, validation techniques, and safety assurance frameworks. Table 4 presents the methodologies elicited from the papers, illustrated with examples. The most right column shows the number of papers of each type. As shown by Figure 5, there is an undeniable emphasize by the community on the various aspects of safe autonomous system design, when we compare research purposes vs. research topics of the selected papers. The results highlights that safe motion, safety assurance, safe prediction and self-adaptation are the most frequent research purposes. In addition, vulnerability detection is the most focused research purpose in security topic.

Methodology vs. Research purpose. The mapping between research purposes and methodologies is essential for understanding the current trends in the field of autonomous systems. This mapping, illustrated by Figure 6, enables us to pinpoint the preferred methodologies for each research purpose and gives more details on how different approaches are being applied across various research objectives.

Several observations are made from this mapping. First, the field of safe motion prominently employs a mix of methodologies, with a significant focus on machine learning models. This suggests a balanced reliance on data-driven insights and theoretical models to ensure the safety and precision of motion in autonomous systems. The combination of these methodologies is crucial for developing systems that can operate reliably in complex and dynamic environments.

Second, in the context of vulnerability detection, the heatmap reveals a strong preference for machine learning models. This indicates an emphasis on using sophisticated algorithms and models to identify and mitigate potential vulnerabilities in autonomous systems. The integration of these methodologies re-

	Safe Autonomous Systems	Performance Engineering in Autonomous Systems	Security issues in Autonomous Systems
Safety Assurance	20	1	1
Self Adaptation	13	0	0
Optimization	4	12	0
Safe Motion	28	0	0
Safe Prediction	17	0	0
Vulnerability Detection	1	0	20
Attack Mitigation	1	0	3
Runtime Monitoring	3	0	0
Security Assurance	0	0	5

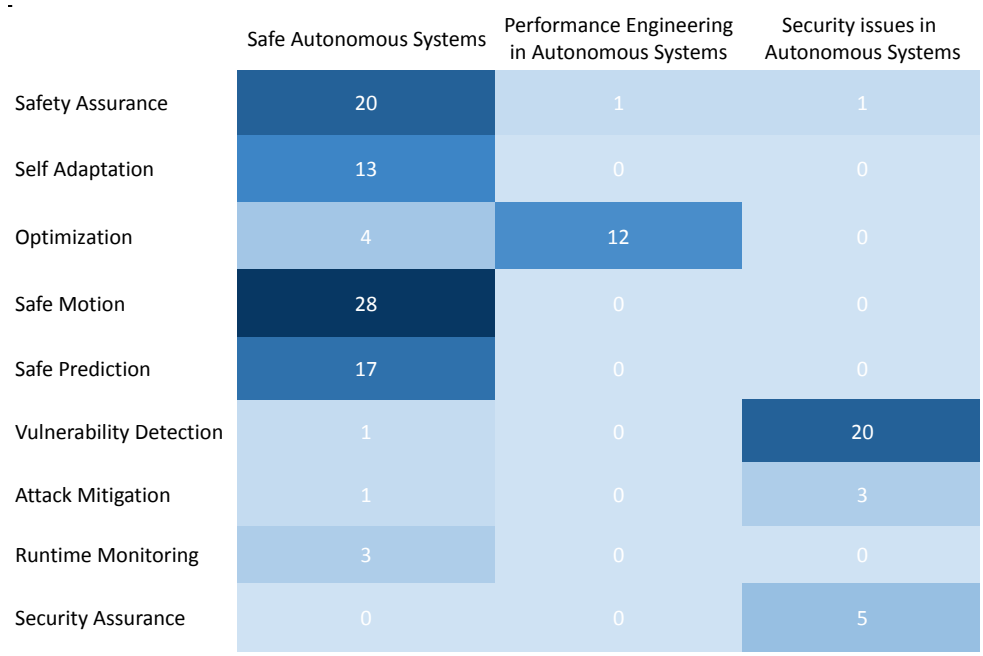

Figure 5: The heatmap for extracting the relationship between Research purpose and Research Topics

	Data Driven Analysis	Model-Based	Machine Learning Model	LPV Control Design	Model-Based/Machine Learning Model	Data Driven Analysis/ Machine Learning Model
Safety Assurance	3	4	10	0	3	2
Self Adaptation	3	4	4	1	1	0
Optimization	3	0	11	0	2	0
Safe Motion	4	4	16	1	1	1
Safe Prediction	3	3	10	0	1	0
Vulnerability Detection	2	2	13	1	1	1
Attack Mitigation	1	0	3	0	0	0
Runtime Monitoring	0	0	3	0	0	0
Security Assurance	0	1	3	0	0	0

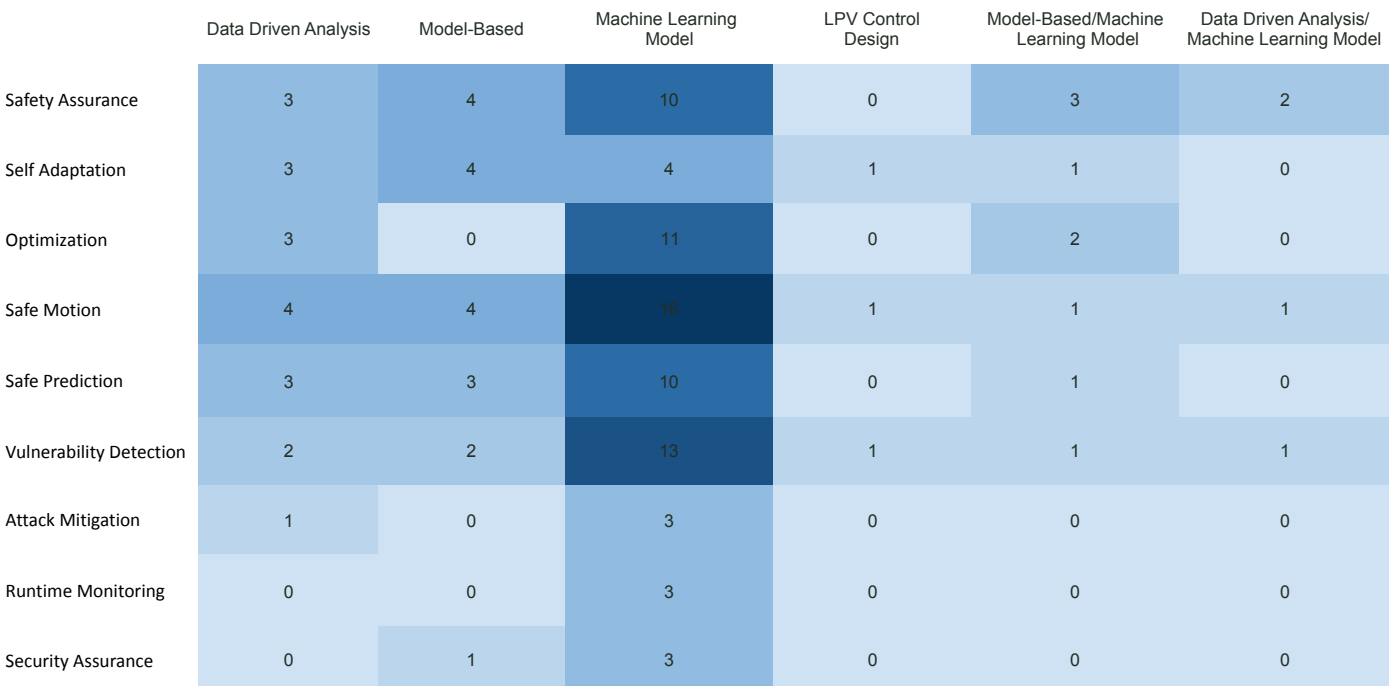

Figure 6: The heatmap for extracting the relationship between Methodology and Research purpose

Table 2: Main topic areas illustrated with examples, with the frequencies of papers (the # column).

Main topic area	Brief Description and Example	#
Safe Autonomous System	This topic area involves managing self-adaptation in learning-enabled systems (LES) to ensure safety in safety-critical systems with artificial intelligence and machine learning components. An example is (88) managing self-adaptation in learning-enabled systems to ensure safety in safety-critical systems with AI and ML components.	70
Security Issues in Autonomous Systems	This topic area addresses the security concerns in autonomous systems. An example is (92) the real-time detection of malware attacks in autonomous vehicles for secure smart transportation.	23
Performance Engineering in Autonomous Systems	This topic area focuses on improving the performance of autonomous systems. An example is (43) improving traffic flow through a bottleneck using decentralized multi-agent reinforcement learning in autonomous vehicles.	7

flects the importance of a comprehensive approach to security, where data analysis and theoretical modeling work together to anticipate and counteract security threats.

Third, the category of safe prediction similarly shows a diverse application of methodologies, with notable emphasis on machine learning models. This trend highlights the critical role of predictive adaptation in autonomous systems, where anticipating future states and behaviors is key to maintaining safety and reliability.

The heatmap also highlights areas with emerging methodologies, such as the use of LPV (Linear Parameter-Varying) (131) control design in a variety of research purposes. While its application is less prevalent compared to more traditional methodologies, its presence signals a growing interest in robust control strategies within the research community.

Answer to RQ1—What are the major research purposes, methodologies and findings in safe autonomy by machine learning techniques?

We identified three primary areas where machine learning is applied to autonomous systems: *Safe Autonomous Systems, Performance Engineering in Autonomous Systems, and Security Issues in Autonomous Systems.* These areas encompass a variety of research purposes: *Safety Assurance, Self-Adaptation, Optimization, Safe Motion, Safe Prediction, Vulnerability Detection, Attack Mitigation, Runtime Monitoring, and Security Assurance.* To address these research purposes, various methodologies are employed: *Data-Driven Analysis, Model-Based Approaches, Machine Learning Models, LPV Control Design, and Hybrid Approaches (Model-Based and Machine Learning Models, or Data-Driven Analysis with Machine Learning).* The dominant case where machine learning is used to address research purposes in autonomous systems is in Safe Motion. This is primarily achieved through the applications of Machine Learning Models, which are employed to enhance the system's ability to make real-time decisions that ensure safe navigation and operation. These models are critical in enabling autonomous systems to predict and react to dynamic environments, thus maintaining safety while performing complex tasks.

4.3. RQ2: What are the main machine learning techniques used in developing safe autonomy where research is being made?

In order to answer this research question, we analyze the following data items: Machine Learning Technique(s) (V15) and Machine Learning Tools (V16).

Machine Learning Techniques. Table 5 provides an overview of the machine learning techniques identified in the reviewed papers, including examples of their applications in safe autonomous systems. The last column indicates the number of papers employing each technique.

Machine Learning Tools. The machine learning tools include software frameworks, libraries, and platforms that support development, training, and deployment of machine learning models in the context of autonomous systems. These tools facilitate various tasks, including data preprocessing, model building, and performance evaluation, thereby enabling researchers and practitioners to implement machine learning solutions more efficiently.

Machine Learning Techniques vs. Methodology. Figure 7 provides a detailed comparison of the frequency with which different machine learning techniques are utilized across various methodologies in the field of autonomous systems. Several notable patterns emerge:

- Dominance of Reinforcement Learning in Machine Learning Models: Reinforcement Learning is the most frequently used technique within the Machine Learning Model methodology, with 20 instances. This highlights its critical role in enabling autonomous systems to make decisions and adapt to dynamic environments through trial and error learning.

- Widespread Use of Supervised Learning: Supervised Learning is extensively applied, particularly within the Machine Learning Model methodology (18 instances). This indicates its importance in training models with labeled data for tasks such as classification, prediction, and safety assurance in autonomous systems.

- Deep Learning's Strong Presence: Deep Learning is another key technique, with significant use in the Machine Learning Model methodology (10 instances). This reflects its capability to handle complex, high-dimensional data, making it a valuable tool for solving challenging problems in autonomous systems.

Answer to RQ2— What are the main machine learning techniques used in developing safe autonomy where research is being made?

The research studies employ a variety of machine learning techniques: *Supervised Learning, Reinforcement Learning, Supervised/Deep Learning, Deep Learning, Unsupervised Learning, Supervised/Reinforcement Learning, and Unsupervised/Deep Learning.* The dominant technique applied across these studies is Reinforcement Learning. These techniques are crucial in enabling autonomous systems to learn from vast amounts of data, adapt to new and unseen environments, and ensure safety while performing complex, real-time tasks in dynamic settings. These models are essential for the development of robust and reliable autonomous systems capable of operating safely in a variety of challenging scenarios.

4.4. RQ3: Which case-studies are concerned in the research works?

Case Studies. Case studies refer to in-depth investigations of specific instances where machine learning techniques are applied to autonomous systems. These studies provide detailed insights into real-world applications, challenges, and outcomes, offering valuable lessons for both researchers and practitioners. By focusing on particular scenarios, case studies illustrate the practical implications of implementing machine learning in safe autonomous systems, highlighting successes, limitations, and areas for improvement. Figure 8 summarizes the case studies identified in the literature, with examples showcasing diverse applications. The last column indicates the number of papers that include each type of case study. The figure reveals that Autonomous Cars are the most extensively studied category, with a frequency of 65, indicating a strong research focus on this type of autonomous system. Autonomous cars are a specific subset of autonomous vehicles designed primarily for personal transportation. The high frequency reflects the significant attention given to the challenges and advancements needed to safely deploy these vehicles in everyday road environments.

In contrast, the Autonomous Vehicle category, which is distinct from cars and has a much lower frequency of 3, likely represents other types of autonomous vehicles that are not cars like autonomous train.

Research Purpose versus Case Study. Figure 9 analysis of research purposes versus case studies reveals critical insights into the focal areas of machine learning applications in autonomous systems. Several notable patterns emerge:

- Safe Motion is Predominantly Studied in Autonomous Cars: Safe Motion is the most heavily studied research purpose within the context of Autonomous Cars, with 22 instances. This highlights the critical focus on ensuring safe navigation and movement in autonomous vehicles, where precise control and safety are paramount.

- Safe Prediction is Strongly Associated with Autonomous Cars: Safe Prediction also has a significant presence in Autonomous Cars, with 13 instances. This indicates a strong emphasis on predicting potential risks, obstacles, or system behaviors to maintain safety during the operation of autonomous vehicles.

- Vulnerability Detection is Spread Across Multiple Domains: Vulnerability Detection appears in various case studies, including Autonomous Systems, IOT, and Unmanned Aerial Vehicles, with 6, 2, and 4 instances respectively. This spread shows the importance of identifying and mitigating security risks across a range of autonomous applications, reflecting the broader concern for cybersecurity in these systems.

	Data Driven Analysis	Model-Based	Machine Learning Model	LPV Control Design	Model-Based/ Machine Learning Model	Data Driven Analysis/ Machine Learning Model
NA	2	3	2	0	0	0
Supervised learning	7	5	24	2	2	0
Reinforcement learning	3	3	28	1	3	2
Supervised/Deep learning	1	1	2	0	0	0
Deep learning	6	3	11	0	2	1
Unsupervised learning	0	2	2	0	0	1
Supervised/Reinforcement learning	0	1	0	0	1	0
Unsupervised/Deep Learning	0	0	1	0	0	0

Figure 7: The heatmap for extracting the relationship between Methodology and Machine Learning.

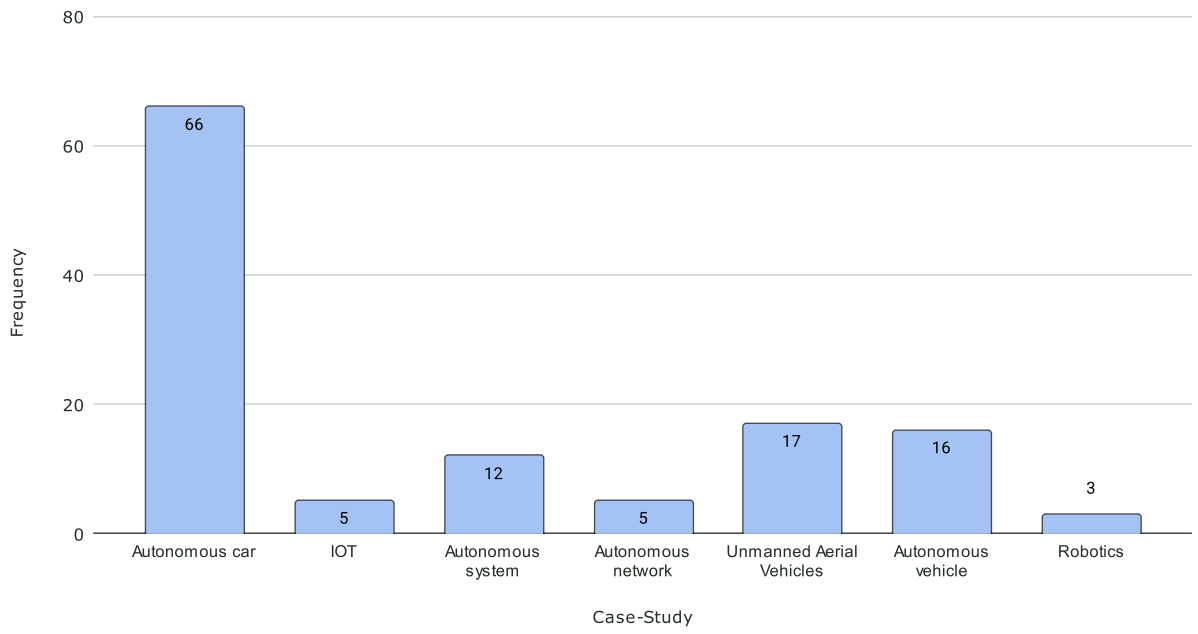

Figure 8: Frequency of Case-studies

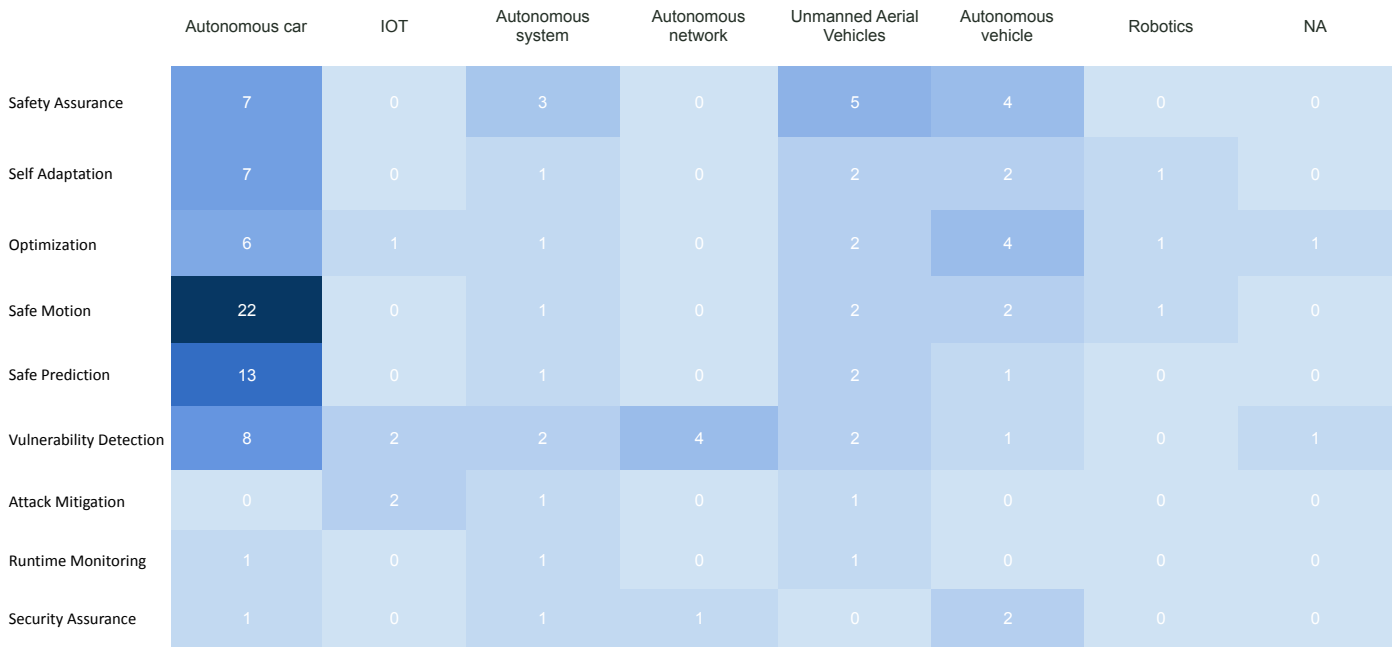

Figure 9: Research purpose versus Case-study

Answer to RQ3— Which case-studies are concerned in the research works?

These case studies include a variety of application areas: *Autonomous Car, Internet of Things (IoT), Autonomous System, Autonomous Network, Unmanned Aerial Vehicles, Autonomous Vehicle (non-car), and Robotics.* The dominant case where machine learning is used to address research purposes is in Autonomous Cars. This focus is primarily on achieving Safe Motion, where machine learning models are applied to enhance the system's ability to make real-time decisions for safe navigation and operation. These models are crucial in enabling autonomous cars to predict and adapt to dynamic driving environments, ensuring both safety and performance while conducting complex driving tasks.

4.5. RQ4: Which types of requirements are addressed in the studies?

Safety and security are two critical requirements in the development and deployment of autonomous systems. Both concepts, while closely related, address different aspects of system reliability and robustness.

Safety refers to the system's ability to operate without causing harm to users, the environment, or other systems. It involves the creation of mechanisms that ensure the system behaves as expected, even in the presence of faults, and that any failure is predictable and manageable. In autonomous systems, safety is paramount, as these systems often operate in dynamic, real-world environments where unforeseen circumstances can arise. Safety measures typically include rigorous testing, adherence to standards, and the implementation of fail-safe mechanisms.

Security, on the other hand, focuses on protecting the system from malicious attacks and unauthorized access. Security is about safeguarding the system's data and operations from external threats that could compromise its functionality or lead to harmful outcomes. In autonomous systems, security is particularly challenging due to the system's complexity and the potential for vulnerabilities in both hardware and software. Ensuring security involves implementing encryption, access controls, and intrusion detection systems, among other measures.

Figure 10 shows that there is a significantly higher emphasis on safety, with 76 requirements listed under this category, compared to 22 requirements under security. This disparity highlights the current prioritization within the field, where ensuring the safe operation of autonomous systems is a primary concern. The relatively lower number of security requirements suggests that while security is essential, it may be either less emphasized or the challenges are being addressed with a smaller but possibly more robust set of security measures. The chart serves as a visual representation of how research and development efforts are distributed between these two critical aspects of autonomous systems.

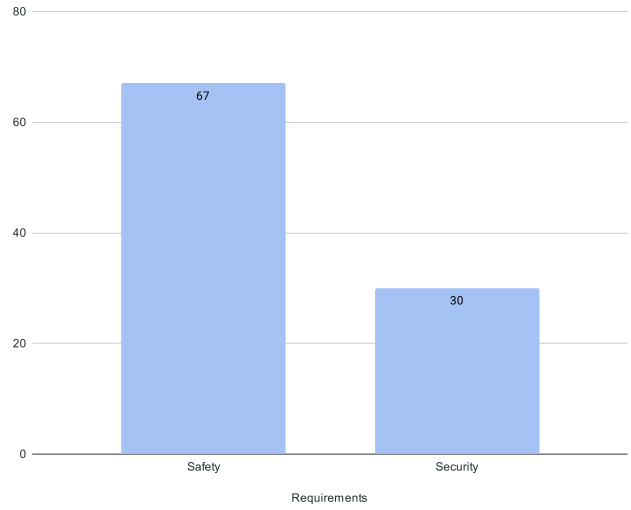

Figure 10: Distribution of requirements

Answer to RQ4— Which types of requirements are addressed in the studies?

We identified two primary requirements in the studies: *Safety, and Security.* Based on the analysis depicted in the provided bar chart, there is a clear emphasis on safety requirements, with 76 instances of safety being addressed across the studies. This indicates a strong focus on ensuring that autonomous systems operate without causing harm and behave predictably in dynamic environments. In contrast, security requirements are addressed less frequently, with only 22 instances noted in the studies. While security remains a critical concern—focused on protecting systems from malicious attacks and unauthorized access—it appears that the primary concern in the current body of research is ensuring the safety of these systems.

4.6. RQ5: What are the limitations of current research or main problems that should have been addressed?

In SLRs, identifying and categorizing limitations is crucial for understanding the challenges and gaps in the research field. These limitations help in framing the scope and boundaries of the study, ensuring that the conclusions drawn are contextualized within the identified constraints.

Data Constraints. This category refers to limitations related to the availability, quality, or relevance of data used in the studies. Data constraints can affect the generalizability of the results and may lead to biased outcomes if the data is not representative of the broader context. In the context of autonomous systems, data constraints could mean insufficient data for training machine learning models or outdated datasets that do not reflect current scenarios.

Validation and Applicability. Limitations under this category pertain to the challenges in validating the results of the studies or the applicability of the findings to real-world scenarios. For autonomous systems, validation might be constrained by the lack of standardized benchmarks or the difficulty in replicating real-world conditions in a controlled environment. The applicability of certain theoretical models or simulation results to actual systems might also be questioned.

Technical and Implementation Challenges. This category addresses the difficulties encountered during the implementation or technical execution of the studies. These challenges could include issues related to software, hardware, or the integration of complex systems. In autonomous systems, technical challenges might involve the integration of different subsystems, handling real-time data processing, or ensuring system robustness under diverse operational conditions.

Methodological and Design Limitations. This category covers the inherent limitations in the study design or methodology. These could include sample size limitations, the use of specific analytical techniques that may not be generalizable, or the design of experiments that do not fully capture the complexity of the systems being studied. For instance, a study on autonomous vehicles might be limited by the choice of simulation tools or the assumptions made during model development.

Figure 11 highlights the distribution of these limitations across the reviewed articles. The chart shows that the majority of limitations fall under the 'NA' category, which might indicate articles where limitations were not explicitly discussed or categorized. 'Validation and Applicability' is the second most common limitation, reflecting ongoing challenges in ensuring that theoretical advancements can be reliably implemented in practical, real-world systems. 'Technical and Implementation Challenges' also feature prominently, underscoring the technical difficulties that researchers often face. Finally, 'Methodological and Design Limitations' and 'Data Constraints' are less frequent but still significant, pointing to areas where research design could be improved for more robust findings and data-related issues.

This distribution of limitations provides valuable insight into the current state of research in the field, emphasizing the areas that require further attention and development. The chart serves as a visual representation of how these challenges are distributed across different studies, guiding future research efforts towards addressing these critical limitations.

4.7. RQ6: What are the future research directions that are worth investigating?

Future research directions in the field of autonomous systems are critical for guiding the next steps in innovation and ensuring the continuous evolution of these technologies. The analysis of future work categories from the studies reviewed provides insights into where researchers are focusing their efforts for further exploration.

Application and Generalization. This category, which appears 10 times in the reviewed studies, emphasizes the importance of ensuring that research findings and developed systems are not only applicable in specific contexts but can also be generalized across a variety of scenarios. The goal is to create solutions that can be adapted to different environments and use

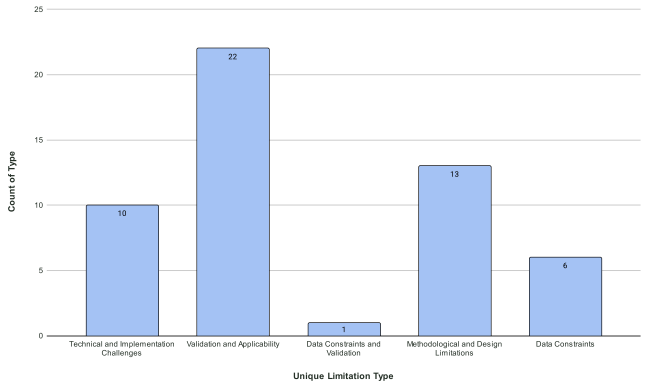

Figure 11: Distribution of limitations

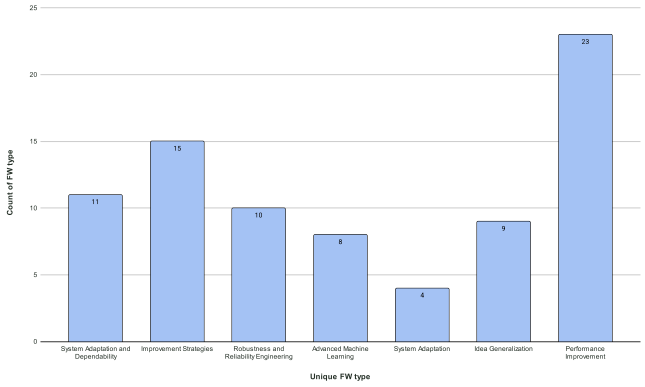

Figure 12: Distribution of future work

cases, thereby increasing the versatility and robustness of autonomous systems.

Advanced Machine Learning. Mentioned 7 times, this category highlights the ongoing need to push the boundaries of machine learning techniques. Future work in this area is likely focused on developing more sophisticated algorithms that can handle the complex data processing and decision-making required by autonomous systems, particularly in dynamic and unpredictable environments.

Robustness and Reliability Engineering. With 7 mentions, this category underscores the critical nature of developing systems that are not only effective but also reliable and resilient under varying conditions. Research in this area seeks to ensure that autonomous systems can maintain their functionality despite unforeseen challenges, which is essential for their safe deployment in real-world scenarios.

Improvement Strategies. Identified in 9 instances, improvement strategies involve developing new approaches, methodologies, or tools that enhance the performance and reliability of autonomous systems. This could include refining current technologies or introducing innovative solutions to address existing limitations or challenges.

Improvement Performance. This category, which is as prominent as the NA category with 32 mentions, points to a significant focus on optimizing the performance of autonomous systems. Researchers are interested in finding ways to make these systems more efficient, faster, and more reliable, ensuring that they meet the increasing demands of modern applications.

System Adaptation and Dependability. Although only mentioned once, this category is crucial as it addresses the ability of autonomous systems to adapt to changes and maintain dependable operations over time. This includes ensuring that systems can respond to new or unexpected conditions while continuing to operate safely and effectively.

Figure 12 visually represents these future research directions, highlighting the areas that researchers deem most critical for the continued development of autonomous systems. The large number of mentions under the NA category, however, suggests that many studies either do not explicitly outline their future work or that this is an area where more detailed discussions could be beneficial.

This analysis of future work not only provides a roadmap for where the field is headed but also underscores the importance of addressing both the technical challenges and broader application potentials of autonomous systems.

4.8. RQ7: What are the highly-cited articles in safe autonomy?

Among the body of literature analyzed in this study, the five most highly cited papers reflect the growing emphasis on deep reinforcement learning and trajectory prediction within the domain of autonomous driving. These works have attracted significant academic attention, indicating their influence and foundational role in shaping current research trends. Their contributions span critical areas such as decision-making frameworks for safe lane changing, navigation in pedestrian-dense environments, and the integration of neural networks for trajectory forecasting. The most cited papers in our analysis are: High-level Decision Making for Safe and Reasonable Autonomous Lane Changing using Reinforcement Learning (105), Collision Avoidance in Pedestrian-Rich Environments With Deep Reinforcement Learning (134), Motion trajectory prediction based on a CNN-LSTM sequential model (83), Automated Lane Change Strategy using Proximal Policy Optimization-based Deep Reinforcement Learning (123), and Attention-Based Hierarchical Deep Reinforcement Learning for Lane Change Behaviors in Autonomous Driving (110). All the mentioned paper had citation number above 100 in the Google Scholar platform at the time of our analysis.

5. Discussion

In this section, we briefly discuss the outcome of the paper and its impact on the community, new research directions and young reseachers in the field of safe autonomous systems.

5.1. Emerging Trends and Techniques in Safe Autonomy

The systematic literature review reveals several trends in the research of safe autonomy supported by machine learning techniques. Reinforcement Learning and Deep Learning are among

the most frequently adopted approaches, illustrating their effectiveness in dynamic and unpredictable environments. These ML techniques significantly contribute to developing adaptive control systems capable of real-time decision-making, vital for maintaining safety standards in autonomous systems.

5.2. Research Focus and Application Domains

Analysis of the reviewed literature highlights that autonomous vehicles, especially self-driving cars, constitute the predominant domain where safe autonomy research is concentrated. This reflects the critical importance of safety and security in transportation, a domain directly impacting human life. Additionally, other emerging areas such as Unmanned Aerial Vehicles, robotics and IoT systems are progressively attracting attention. This expansion indicates a broadening scope and increased applicability of machine learning-based safety mechanisms.

5.3. Identification of Research Gaps and Challenges

The reviewed papers consistently underline several challenges and limitations within current research methodologies. Notable among these are the robustness and reliability of ML models under adversarial conditions and their adaptability in real-time scenarios. There is also a necessity for standardized benchmarks and datasets to facilitate comparative studies and validation of new safety mechanisms. Addressing these gaps will be crucial for advancing the field.

5.4. Implications for Young Researchers

The outcomes of this analysis are particularly beneficial for young researchers. It provides a comprehensive view of current research directions and gaps. Emerging scholars can utilize these insights to focus their studies strategically, addressing unresolved challenges and adopting effective methodologies highlighted in recent literature. In addition, understanding current trends such as the prominent role of reinforcement and deep learning techniques can guide researchers in acquiring relevant competencies.

5.5. Future Research Directions

The systematic literature review identifies several promising directions for future research. Further exploration of hybrid models combining model-based and ML-based approaches appears particularly promising. Additionally, enhancing the transparency of ML decisions in safety-critical contexts will be essential for gaining public trust and meeting regulatory standards. Security aspects, including real-time intrusion detection and response in autonomous systems, represent another fertile area for research, urging the necessity of developing robust and proactive cybersecurity frameworks.

6. Summary and conclusions

This systematic literature review has highlighted significant progress and persistent challenges in applying machine learning to safe and secure autonomous systems. It reveals the major use of Reinforcement Learning and Deep Learning techniques, especially in autonomous vehicles, to achieve adaptive safety mechanisms. While significant advancements have been made, critical gaps remain, including the necessity for robustness, real-time adaptability, standardized datasets and benchmarks. Addressing these issues will not only enhance system reliability but also facilitate broader acceptance and deployment of autonomous technologies. Future research in this field should focus on developing interpretable and hybrid machine learning models, proactive security solutions, and enhanced model validation frameworks, ultimately driving forward the safe integration of machine learning into critical autonomous applications.

A potential future direction for this SLR is to periodically update the analysis to capture emerging trends, novel methodologies, and recently identified challenges within the rapidly evolving field of autonomous systems. Further, extending the review to incorporate interdisciplinary studies could enrich understanding of complex interactions between technical solutions and socio-ethical considerations.

Appendix A: List of Selected Papers (Primary Studies)

We have collected, reviewed and analyzed 129 papers for this study. The list of these papers is denoted by Table 6.

Table 3: Research purpose illustrated with examples, with the frequencies of papers (the # column).

Research purpose	Brief Description and Example	#
Safety Assurance	This research purpose focuses on developing methods and systems to ensure the safety and reliability of autonomous operations in various environments. An example is (89) developing a robust system for the autonomous landing of UAVs that integrates depth completion and semantic segmentation to ensure safe and reliable landing in dynamically changing and complex environments.	22
Self Adaptation	This purpose involves creating systems that can adapt themselves in response to changes in their environment or internal state. An example is (88) introducing a model-driven approach for self-adaptation of learning-enabled systems (LES) to address uncertainties in learned behavior when real-world phenomena are not fully captured during training.	13
Optimization	This area aims to enhance the performance and efficiency of systems through improved data processing and decision-making techniques. An example is (67) using data mining and machine learning methods to more accurately classify and identify traffic flows of cloud applications and network services.	16
Safe Motion	This research focuses on developing control mechanisms for autonomous vehicles to navigate safely in complex scenarios. An example is (148) addressing the challenges associated with controlling autonomous vehicles at pedestrian crossings, particularly when the action space is hybrid (simultaneously continuous and discrete) and the reward is sparse.	28
Safe Prediction	This purpose explores predictive methods to enhance the safety and responsiveness of autonomous systems. An example is (121) exploring ways of utilizing emergency ambulance drones to tackle the problem of inaccessibility and delay of assistance to patients in emergency situations.	17
Vulnerability Detection	This area aims to identify and mitigate vulnerabilities within autonomous systems to prevent security breaches. An example is (149) exploring the use of deep learning, specifically a convolutional neural network (CNN) linked with variational autoencoders (VAEs), to create a model for preventing cyber security threats.	21
Attack Mitigation	This research purpose involves developing strategies to defend against malicious attacks on autonomous systems. An example is (64) proposing a new approach to mitigating link-flooding attacks (LFA) in IoT environments.	4
Runtime Monitoring	This area focuses on real-time monitoring of systems to ensure ongoing safe operation. An example is (116) presenting a system for monitoring an aircraft operator's physiological state in real-time using sensors like eye tracking and electrocardiogram.	3
Security Assurance	This research aims to provide methods to ensure that autonomous systems are protected against security threats. An example is (92) ensuring security assurance in autonomous vehicles against malware by machine learning.	5

Table 4: Methodology illustrated with examples, with the frequencies of papers (the right most column).

Methodology	Brief Description and Example	#
Data Driven Analysis	This methodology involves collecting and analyzing data to train models and make predictions. An example is (133) collecting driving data and corresponding steering commands from an experienced driver. This data was then used to train a seventeen-layer Convolutional Neural Network (CNN) called BCNet using behavioral cloning, with the network optimized using the Adam's optimization algorithm, ultimately leading to successful replication of the driving behavior through extensive simulations.	22
Model-Based	This approach uses mathematical models to represent the behavior of systems. An example is (102) a navigation system based on three main blocks: (i) Perception module, (ii) Control laws, and (iii) Machine Learning module. The Perception module extracts useful information from sensor data to generate state and environment representations. The Control laws generate online control commands for vehicle motion control, based on both geometric and machine learning-based methods, while the Machine Learning module improves obstacle avoidance by learning from past scenarios.	18
Machine Learning Model	This methodology involves using machine learning techniques to build predictive models. An example is (124) a new methodology that uses a Recurrent Neural Network (RNN) to predict the driving intention of surrounding vehicles. The RNN takes as input a sequence of data describing the motion of the surrounding vehicles and produces as output a probability distribution over all possible driving intentions of these vehicles. The RNN is trained using a large dataset of labeled driving scenarios and is designed to learn the complex spatio-temporal patterns characteristic of these scenarios. The article also describes the training process for the RNN, which includes several steps such as data preprocessing, weight initialization, and error back-propagation.	73
LPV Control Design	This approach combines big data and machine learning with Linear Parameter-Varying (LPV) control techniques to enhance system performance. An example is (131) utilizing big data and machine learning to plan optimal velocities combined with advanced LPV control techniques for autonomous vehicle stability and trajectory tracking. The data-driven reachability analysis is a novel aspect.	3
Model-Based/Machine Learning Model	This methodology integrates model-based approaches with machine learning techniques. An example is (43) applying multi-agent reinforcement learning algorithms to model and optimize traffic flow through a bottleneck, specifically using a scaled model of the San Francisco–Oakland Bay Bridge. The study compares the performance of these algorithms with a hand-designed feedback controller and explores the feasibility of decentralized strategies using local sensing.	9
Data Driven Analysis/Machine Learning Model	This approach involves a combination of data-driven analysis and machine learning models. An example is (39) designing a custom gridworld environment for a Markov Decision Process (MDP) using Python, where an intelligent agent learns to identify and navigate around hidden rewards. The agent's movements are monitored, and a heatmap is constructed to help the agent avoid unsafe conditions by learning the positions of hidden rewards effectively.	4

Table 5: Machine learning techniques illustrated with examples, with the frequencies of papers (the # column).

ML technique	Brief Description and Example	#
Supervised learning	A technique where models are trained on labeled datasets to predict outcomes or classify data based on input-output pairs. An example from the article (47) is collecting labeled network traffic data (normal or attack) from the UNSW-NB15 dataset. This data was used to train a Decision Tree classifier, a Multilayer Perceptron (MLP) neural network with ReLU activation and hidden layers, and an XGBoost ensemble model. The models were trained using supervised learning, with the MLP optimized using the Adam algorithm from the scikit-learn library. The trained models achieved accuracies of 96.5% (Decision Tree), 89.83% (MLP), and 89.9% (XGBoost) in classifying network intrusions, demonstrating their effectiveness in replicating human-defined threat detection patterns.	41
Unsupervised learning	A machine learning paradigm where algorithms identify patterns or structures in unlabeled data without explicit guidance. A large dataset of lateral dynamics measurements—such as yaw rate, side-slip angle, and longitudinal velocity—is collected from high-fidelity simulations using CarSim (131). This dataset is then fed into an unsupervised clustering algorithm, for example, K-means clustering, to automatically partition the data into distinct groups based solely on inherent similarities in vehicle behavior. The resulting clusters reveal natural groupings within the data; one of these clusters corresponds to the operating conditions where vehicle states remain in the linear (stable) region of dynamics. The boundaries of this cluster are then used as the approximation of the vehicle's reachability set, which serves as a foundation for optimizing the longitudinal velocity profile. By leveraging unsupervised clustering, the method can extract the stability regions directly from the data without relying on manually crafted labels, ultimately supporting the design of a robust LPV-based controller that ensures safe autonomous vehicle operation under varying driving conditions.	5
Reinforcement learning	A method where an agent learns optimal behaviors through iterative interactions with an environment, guided by reward signals. An example (42) of reinforcement learning from the article is the hierarchical deep reinforcement learning (DRL) framework used by ScanBot for autonomous 3D scene reconstruction	42
Deep learning	A subset of machine learning using multi-layered neural networks to model high-level abstractions in data, often for complex tasks. An example (137) of deep learning from the article is the DNN-based intrusion detection system used in the hybrid IDS framework for securing in-vehicle CAN networks.	24
Supervised/Deep learning	The application of deep neural networks within a supervised framework, leveraging labeled data for tasks like classification or regression (65).	4
Supervised/Reinforcement learning	A hybrid approach combining labeled data training with reward-driven optimization, often used for imitation or transfer learning (129).	2
Unsupervised/Deep Learning	The use of deep neural networks to discover latent patterns or representations in unlabeled data, such as clustering or generative modeling (68).	1

Title of the Papers	Journal/Conference Name	Year
Human–machine Teaming with Small Unmanned Aerial Systems in a MAPE-K Environment (48)	ACM Transactions on Autonomous and Adaptive Systems	2024
Dynamic Planning of Optimally Safe Lane-change Trajectory for Autonomous Driving on Multi-lane Highways Using a Fuzzy Logic–based Collision Estimator (49)	ACM Journal on Autonomous Transportation Systems	2024
Fairness-Driven Private Collaborative Machine Learning (50)	ACM Transactions on Intelligent Systems and Technology	2024
RAMPART: Reinforcing Autonomous Multi-Agent Protection through Adversarial Resistance in Transportation (51)	ACM Journal on Autonomous Transportation Systems	2024
CASTNet: A Context-Aware, Spatio-Temporal Dynamic Motion Prediction Ensemble for Autonomous Driving (52)	ACM Transactions on Cyber-Physical Systems	2024
Hybrid Algorithm Selection and Hyperparameter Tuning on Distributed Machine Learning Resources: Hierarchical Agent-Based Approach (53)	ACM Transactions on Internet Technology	2024
MarMot: Metamorphic Runtime Monitoring of Autonomous Driving Systems (54)	ACM Transactions on Software Engineering and Methodology	2024
Using Reinforcement Learning and Error Models for Drone Precise Landing (55)	ACM Transactions on Internet Technology	2024
A power prediction approach for a solar-powered aerial vehicle enhanced by stacked machine learning technique (72)	Computers and Electrical Engineering	2024
End-to-end neural network based optimal quadcopter control (73)	Robotics and Autonomous Systems	2024
Safe multi-agent drone control using control barrier functions and acceleration fields (74)	Robotics and Autonomous Systems	2024
Enhancing lane detection with a lightweight collaborative late fusion model (75)	Robotics and Autonomous Systems	2024
Sim-to-real transfer of active suspension control using deep reinforcement learning (76)	Robotics and Autonomous Systems	2024
Human-robot interactions in autonomous hospital transports (77)	Robotics and Autonomous Systems	2024
Dynamic Global/Local Multi-Layer Motion Planner Architecture for Autonomous Cognitive Surgical Robots (78)	Robotics and Autonomous Systems	2024
Applications of AI/ML in Maritime Cyber Supply Chains (79)	5th International Conference on Industry 4.0 and Smart Manufacturing	2024
Enhancing Accessible Navigation: A Fusion of Speech, Gesture, and Sonification for the Visually Impaired (80)	KES 2024	2024
Cybersecurity of Autonomous Ships in Offshore Wind Farms (81)	KES 2024	2024
Reinforcement Imitation Learning for Reliable and Efficient Autonomous Navigation in Complex Environments (97)	Neural Computing and Applications	2024
Traffic Signal Optimization Framework Using Interpretable Machine Learning Technique Under Heterogeneous-Autonomy Traffic Environment (96)	Neural Computing and Applications	2024
A Decision-Making Model for Self-Driving Vehicles Based on GPT-4V, Federated Reinforcement Learning, and Blockchain (95)	Neural Computing and Applications	2024
A Machine Learning-Based Ensemble Model for Securing the IoT Network (94)	Cluster Computing	2024
Local Planner Selection for Autonomous Robots Using Machine Learning (93)	International Journal of Intelligent Robotics and Applications	2024
Learning a Safe Driving Policy for Urban Area from Real-World Data (158)	6th International Conference on Reconfigurable Mechanisms and Robots (ReMAR)	2024
Gradient Monitored Reinforcement Learning for Jamming Attack Detection in FANETs (157)	IEEE Access	2024
FusedWSS: Water Surface Segmentation Fusing Machine Learning and Geometric Cues (156)	Not explicitly stated	2024
Safeguarding Autonomous UAV Navigation: Agent Design using Evolving Behavior Trees (155)	IEEE International Systems Conference (SysCon)	2024
Enhancing Longitudinal Velocity Control With Attention Mechanism-Based Deep Deterministic Policy Gradient (DDPG) for Safety and Comfort (154)	IEEE Access	2024
Enhancing Safety for Autonomous Agents in Partly Concealed Urban Traffic Environments Through Representation-Based Shielding (153)	IEEE Intelligent Vehicles Symposium (IV)	2024
A Simulation Tool for Space Traffic Management (152)	AIAA/IEEE Digital Avionics Systems Conference (DASC)	2024
Assuring Safe Navigation and Network Operations of Autonomous Ships (79)	IEEE Annual Computing and Communication Workshop and Conference (CCWC)	2024
ThirdEye: Attention Maps for Safe Autonomous Driving Systems (41)	ASE '22	2023
ScanBot: Autonomous Reconstruction via Deep Reinforcement Learning (42)	ACM Transactions on Graphics (TOG)	2023
Optimizing Mixed Autonomy Traffic Flow with Decentralized Autonomous Vehicles and Multi-Agent Reinforcement Learning (43)	ACM Transactions on Cyber-Physical Systems	2023
Test Scenario Generation for Autonomous Driving Systems with Reinforcement Learning (44)	ICSE '23	2023
Resource Efficient Intrusion Detection Systems for Internet of Things Using Online Machine-Learning Models (45)	SIET '23	2023
A Multi-layered Collaborative Framework for Evidence-driven Data Requirements Engineering for Machine Learning-based Safety-critical Systems (46)	SAC '23	2023
Machine Learning based Intrusion Detection System for IoT Applications using Explainable AI (47)	AIMLR '23	2023
Multi-sensor based strategy learning with deep reinforcement learning for unmanned ground vehicle (70)	International Journal of Intelligent Networks	2023
Reliable and Efficient Lane Changing Behaviour for Connected Autonomous Vehicle through Deep Reinforcement Learning (71)	Procedia Computer Science	2023
MoDALAS: addressing assurance for learning-enabled autonomous systems in the face of uncertainty (88)	Software and Systems Modeling	2023
A study on multi-agent reinforcement learning for autonomous distribution vehicles (90)	Iran Journal of Computer Science	2023
Enhanced detection and recognition system for vehicles and drivers using multi-scale retinex guided filter and machine learning (91)	Multimedia Tools and Applications	2023
Cybersecurity for autonomous vehicles against malware attacks in smart-cities (92)	Cluster Computing	2023
Enhancing Pedestrian Safety in Autonomous Vehicles through Machine Learning (142)	2023 International Conference on Sustainable Communication Networks and Application (ICSCNA)	2023
Safe Reinforcement Learning for Model-Reference Trajectory Tracking of Uncertain Autonomous Vehicles With Model-Based Acceleration (143)	IEEE Transactions on Intelligent Vehicles	2023
Path Tracking and Obstacle Avoidance Control Using Deep Reinforcement Learning for Autonomous Vehicles (144)	2023 17th International Conference on Open Source Systems and Technologies (ICOSST)	2023
An Anomaly Behavior Analysis Framework for Securing Autonomous Vehicle Perception (145)	2023 20th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)	2023
Data-driven Deep Reinforcement Learning for Automated Driving (146)	26th International Conference on Intelligent Transportation Systems (ITSC)	2023
Deep Reinforcement Learning for Autonomous Vehicle Intersection Navigation (147)	Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)	2023
Hybrid Deep Reinforcement Learning Model for Safe and Efficient Autonomous Vehicles at Pedestrian Crossings (148)	Intelligent Systems (IWIS)	2023
Machine Learning with Deep Learning Approach for Cyber Security Threats Prevention Model (149)	Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)	2023
Safe Reinforcement Learning for Lane-changing with Comprehensive Analysis of Safety Detection (150)	Intelligent Transportation Systems (ITSC)	2023
Securing Cyber-Physical Systems Against GPS Spoofing Attacks Using Confidence Attribution (151)	Software, Telecommunications and Computer Networks (SoftCOM)	2023
IoT Based Traffic Management, Accident Detection, and Accident Prevention System Using Machine Learning Method (37)	ICCA '22	2022
Safe Reinforcement Learning in Simulated Environment of Self-Driving Laboratory (39)	ESSE '21	2022
The Security Concerns On Cyber-Physical Systems And Potential Risks Analysis Using Machine Learning (57)	Procedia Computer Science	2022
Instance-based defense against adversarial attacks in Deep Reinforcement Learning (58)	Engineering Applications of Artificial Intelligence	2022
A hybrid method of entropy and SSAE-SVM based DDoS detection and mitigation mechanism in SDN (60)	Computers & Security	2022
Cluster-based attacks prevention algorithm for autonomous vehicles using machine learning algorithms (68)	Computers and Electrical Engineering	2022
RAMARL: Robustness Analysis with Multi-Agent Reinforcement Learning - Robust Reasoning in Autonomous Cyber-Physical Systems (69)	Procedia Computer Science	2022
Integrating Safety Guarantees into the Learning Classifier System XCS (86)	Applications of Evolutionary Computation	2022
Intrusion-Based Attack Detection Using Machine Learning Techniques for Connected Autonomous Vehicle (87)	Advances and Trends in Artificial Intelligence. Theory and Practices in Artificial Intelligence	2022
Robust autonomous landing of UAVs in non-cooperative environments based on comprehensive terrain understanding (89)	Science China Information Sciences	2022
Transfer Learning based GPS Spoofing Detection for Cellular-Connected UAVs (99)	International Wireless Communications and Mobile Computing (IWCMC)	2022
Securing CAN Bus in Connected and Autonomous Vehicles Using Supervised Machine Learning Approaches (100)	IEEE International Conference on Electro Information Technology (eIT)	2022
Using Machine Learning and Regression Analysis to Classify and Predict Danger Levels in Burning Sites (114)	2022 IEEE World AI IoT Congress (AIIoT)	2022
Emergency Medical Assistance by Ambulance Drone Using Machine Learning, Light-Weight Cryptography and Variable Image Steganography (121)	2022 IEEE VLSI Device Circuit and System (VLSI DCS)	2022
GNSS Spoofing Attack Detection By IMU Measurements Through A Neural Network (126)	2022 10th Workshop on Satellite Navigation Technology (NAVITEC)	2022
A Hybrid Approach Toward Efficient and Accurate Intrusion Detection for In-Vehicle Networks (137)	IEEE Access	2022
Safe Robot Navigation Using Constrained Hierarchical Reinforcement Learning (140)	2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)	2022
Collision Avoidance Predictive Motion Planning Based on Integrated Perception and V2V Communication (141)	IEEE Transactions on Intelligent Transportation Systems	2022
ConAML: Constrained Adversarial Machine Learning for Cyber-Physical Systems (35)	ASIA CCS '21	2021
Effectiveness of Transfer Learning in Autonomous Driving using Model Car (38)	ICMLC 2021	2021
An integrated methodology for real-time driving risk status prediction using naturalistic driving data (65)	Accident Analysis & Prevention	2021
Safe Adaptive Deep Reinforcement Learning for Autonomous Driving in Urban Environments. Additional Filter? How and Where? (98)	IEEE Access	2021
Autonomous vehicle navigation based in a hybrid methodology: model based and machine learning based (102)	IEEE International Conference on Mechatronics (ICM)	2021
An Automated Machine Learning (AutoML) Method of Risk Prediction for Decision-Making of Autonomous Vehicles (104)	IEEE Transactions on Intelligent Transportation Systems	2021
Intrusion Detection System for CAN Bus In-Vehicle Network based on Machine Learning Algorithms (111)	2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)	2021
End-to-End Autonomous Driving Risk Analysis: A Behavioural Anomaly Detection Approach (115)	IEEE Transactions on Intelligent Transportation Systems	2021
Collaborative Communications Between A Human And A Resilient Safety Support System (116)	2021 IEEE International Conference on Autonomous Systems (ICAS)	2021
Railroad Near-Miss Occurrence Detection and Risk Estimation System with Data from Camera Using Deep Learning (117)	2021 5th International Conference on Imaging, Signal Processing and Communications (ICISPC)	2021
Prediction of Human Intention in Vehicles, Pedestrians and Bicyclists Interactions (119)	2021 IEEE International Intelligent Transportation Systems Conference (ITSC)	2021

Lane-Change Intention Inference Based on RNN for Autonomous Driving on Highways (124)	IEEE Transactions on Vehicular Technology	2021
Real-Time Prediction of Multi-Class Lane-Changing Intentions based on Highway Vehicle Trajectories (125)	2021 IEEE International Intelligent Transportation Systems Conference (ITSC)	2021
Collision Avoidance in Pedestrian-Rich Environments With Deep Reinforcement Learning (134)	IEEE Access	2021
Attack-Detection Architectural Framework Based on Anomalous Patterns of System Performance and Resource Utilization—Part II (136)	IEEE Access	2021
Automatic Recognition of Traffic Signs Based on Visual Inspection (138)	IEEE Access	2021
Self-Adaptive Traffic Control Model With Behavior Trees and Reinforcement Learning for AGV in Industry 4.0 (139)	IEEE Transactions on Industrial Informatics	2021
Machine Learning-Based Intrusion Detection System for Controller Area Networks (111)	DIVANet '20	2020
Modelling Human Driving Behavior for Constrained Model Predictive Control in Mixed Traffic at Intersections (59)	IFAC-PapersOnLine	2020
Sensor Defense In-Software (SDI): Practical software based detection of spoofing attacks on position sensors (61)	Engineering Applications of Artificial Intelligence	2020
Network attacks classification using Long Short-term memory based neural networks in Software-Defined Networks (62)	Procedia Computer Science	2020
LPV-based control for automated driving using data-driven methods (63)	IFAC-PapersOnLine	2020
Mitigating LFA through segment rerouting in IoT environment with traceroute flow abnormality detection (64)	Journal of Network and Computer Applications	2020
Motion trajectory prediction based on a CNN-LSTM sequential model (83)	Science China Information Sciences	2020
Deep Reinforcement Learning for Safe Local Planning of a Ground Vehicle in Unknown Rough Terrain (101)	IEEE Robotics and Automation Letters	2020
Assuring the Safety of End-to-End Learning-Based Autonomous Driving through Runtime Monitoring (103)	23rd Euromicro Conference on Digital System Design (DSD)	2020
Safe Reinforcement Learning for Autonomous Lane Changing Using Set-Based Prediction (106)	2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)	2020
Towards Autonomous Driving: A Machine Learning-based Pedestrian Detection System using 16-Layer LiDAR (108)	2020 13th International Conference on Communications (COMM)	2020
Safe and Computational Efficient Imitation Learning for Autonomous Vehicle Driving (109)	2020 American Control Conference (ACC)	2020
Estimation of Collision Priority on Traffic Videos using Deep Learning (112)	2020 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)	2020
ReachFlow: An Online Safety Assurance Framework for Waypoint-Following of Self-driving Cars (118)	2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)	2020
Automated Lane Change Strategy using Proximal Policy Optimization-based Deep Reinforcement Learning (123)	2020 IEEE Intelligent Vehicles Symposium (IV)	2020
LPV-based autonomous vehicle control using the results of big data analysis on lateral dynamics (131)	2020 American Control Conference (ACC)	2020
Machine learning algorithm of immune neuro-fuzzy anti-collision embedded system for autonomous unmanned aerial vehicles team (31)	APPIS '19	2019
Simulated autonomous driving in a realistic driving environment using deep reinforcement learning and a deterministic finite state machine (33)	APPIS '19	2019
Efficient analysis of large adaptation spaces in self-adaptive systems using machine learning (34)	SEAMS '19	2019
Learning-based adversarial agent detection and identification in cyber physical systems applied to autonomous vehicular platoon (36)	SEsCPS '19	2019
Road surface estimation based LPV control design for autonomous vehicles (56)	IFAC-PapersOnLine	2019
UCRLF: unified constrained reinforcement learning framework for phase-aware architectures for autonomous vehicle signaling and trajectory optimization (84)	Evolutionary Intelligence	2019
Novelty Detection via Network Saliency in Visual-Based Deep Learning (107)	2019 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)	2019
Attention-Based Hierarchical Deep Reinforcement Learning for Lane Change Behaviors in Autonomous Driving (110)	2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)	2019
Object detection and object tracking in front of the vehicle using front view camera (120)	2019 Zooming Innovation in Consumer Technologies Conference (ZINC)	2019
Adaptive Vehicle Detection for Real-time Autonomous Driving System (122)	2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)	2019
A Lane Merge Coordination Model for a V2X Scenario (128)	2019 European Conference on Networks and Communications (EuCNC)	2019
Lane Change Intention Awareness for Assisted and Automated Driving on Highways (129)	IEEE Transactions on Intelligent Vehicles	2019
Multi-Reward Architecture based Reinforcement Learning for Highway Driving Policies (132)	2019 IEEE Intelligent Transportation Systems Conference (ITSC)	2019
Towards a holistic software systems engineering approach for dependable autonomous systems (30)	SEFAIS '18	2018
Towards integrating undependable self-adaptive systems in safety-critical environments (40)	SEAMS '18	2018
Data-Driven Reachability Analysis for the Reconfiguration of Vehicle Control Systems (66)	IFAC-PapersOnLine	2018
Comprehensive approach for optimization traffic routing and using network resources in a virtual data center (67)	Procedia Computer Science	2018
A Self-adaptive Approach for Autonomous UAV Navigation via Computer Vision (85)	Computational Science and Its Applications – ICCSA 2018	2018
High-level Decision Making for Safe and Reasonable Autonomous Lane Changing using Reinforcement Learning (105)	2018 21st International Conference on Intelligent Transportation Systems (ITSC)	2018
An Advanced Driver Assistance System Using Computer Vision and Deep-Learning (113)	2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)	2018
Using Optimization, Learning, and Drone Reflexes to Maximize Safety of Swarms of Drones (127)	2018 IEEE Congress on Evolutionary Computation (CEC)	2018
Deep Q Learning Based High Level Driving Policy Determination (130)	2018 IEEE Intelligent Vehicles Symposium (IV)	2018
Behavior Cloning for Autonomous Driving using Convolutional Neural Networks (133)	2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)	2018
Learning How to Drive in Blind Intersections from Human Data (135)	IEEE International Conference on Systems, Man, and Cybernetics (SMC)	2018

References

[1] M.L. Cummings, David Britton, Chapter 6 - Regulating safety-critical autonomous systems: past, present, and future perspectives, Editor(s): Richard Pak, Ewart J. de Visser, Ericka Rovira, Living with Robots, Academic Press, 2020, Pages 119-140, ISBN 9780128153673, https://doi.org/10.1016/B978-0-12-815367-3.00006-2.

[2] H. He, J. Gray, A. Cangelosi, Q. Meng, T. M. McGinnity and J. Mehnen, "The Challenges and Opportunities of Human-Centered AI for Trustworthy Robots and Autonomous Systems," in IEEE Transactions on Cognitive and Developmental Systems, vol. 14, no. 4, pp. 1398-1412, Dec. 2022, doi: 10.1109/TCDS.2021.3132282.

[3] Hsu, Kai-Chieh, Haimin Hu, and Jaime F. Fisac. "The safety filter: A unified view of safety-critical control in autonomous systems." Annual Review of Control, Robotics, and Autonomous Systems 7 (2023).

[4] A. Hasani, M. A. Nia and R. E. Atani, "Balancing Safety and Security in Autonomous Driving Systems: A Machine Learning Approach with Safety-First Prioritization," 2024 Annual Computer Security Applications Conference Workshops (ACSAC Workshops), Honolulu, HI, USA, 2024, pp. 30-41, doi: 10.1109/ACSACW65225.2024.00012.

[5] Attanasio, Aleks, Bruno Scaglioni, Elena De Momi, Paolo Fiorini, and Pietro Valdastri. "Autonomy in surgical robotics." Annual Review of Control, Robotics, and Autonomous Systems 4, no. 1 (2021): 651-679.

[6] Mehran Alidoost Nia, Radu Calinescu, Mehdi Kargahi, and Alessandro Abate, "Efficient Model Verification at Runtime through Adaptive Dynamic Approximation," ACM Trans. Auton. Adapt. Syst. 20, 1, Article 3 (March 2025), 45 pages. https://doi.org/10.1145/3708560

[7] C. Gao, G. Wang, W. Shi, Z. Wang and Y. Chen, "Autonomous Driving Security: State of the Art and Challenges," in IEEE Internet of Things Journal, vol. 9, no. 10, pp. 7572-7595, 15 May15, 2022, doi: 10.1109/JIOT.2021.3130054.

[8] X. Sun, F. R. Yu and P. Zhang, "A Survey on Cyber-Security of Connected and Autonomous Vehicles (CAVs)," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 7, pp. 6240-6259, July 2022, doi: 10.1109/TITS.2021.3085297.

[9] H. Pu, L. He, P. Cheng, M. Sun and J. Chen, "Security of Industrial Robots: Vulnerabilities, Attacks, and Mitigations," in IEEE Network, vol. 37, no. 1, pp. 111-117, January/February 2023, doi: 10.1109/MNET.116.2200034.

[10] Mehran Alidoost Nia, Behnam Bahrak, Mehdi Kargahi, and Benjamin Fabian, "Detecting new generations of threats using attribute-based attack graphs," IET Inf. Secur., 13: 293-303, 2019, https://doi.org/10.1049/iet-ifs.2018.5409

[11] Mehran Alidoost Nia, A. Ruiz-Martínez, Systematic literature review on the state of the art and future research work in anonymous communications systems, Computers & Electrical Engineering, Volume 69, 2018, Pages 497-520, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2017.11.027.

[12] Hawkins, Richard, Colin Paterson, Chiara Picardi, Yan Jia, Radu Calinescu, and Ibrahim Habli. "Guidance on the assurance of machine learning in autonomous systems (AMLAS)." arXiv preprint arXiv:2102.01564 (2021).

[13] B. Lian et al., "Anomaly Detection and Correction of Optimizing Autonomous Systems With Inverse Reinforcement Learning," in IEEE Transactions on Cybernetics, vol. 53, no. 7, pp. 4555-4566, July 2023, doi: 10.1109/TCYB.2022.3213526.

[14] 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," in IEEE Transactions on Industrial Informatics, vol. 17, no. 12, pp. 7897-7912, Dec. 2021, doi: 10.1109/TII.2021.3071405.

[15] Mehran Alidoost Nia, Mehdi Kargahi, Alessandro Abate, Resilient monitoring in self-adaptive systems through behavioral parameter estimation, Journal of Systems Architecture, Volume 118, 2021, 102177, ISSN 1383-7621, https://doi.org/10.1016/j.sysarc.2021.102177.

[16] G. Kahn, P. Abbeel and S. Levine, "BADGR: An Autonomous Self-Supervised Learning-Based Navigation System," in IEEE Robotics and Automation Letters, vol. 6, no. 2, pp. 1312-1319, April 2021, doi: 10.1109/LRA.2021.3057023.

[17] Y. Tang et al., "Perception and Navigation in Autonomous Systems in the Era of Learning: A Survey," in IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 12, pp. 9604-9624, Dec. 2023, doi: 10.1109/TNNLS.2022.3167688.

[18] B. Kiumarsi, K. G. Vamvoudakis, H. Modares and F. L. Lewis, "Optimal and Autonomous Control Using Reinforcement Learning: A Survey," in IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 6, pp. 2042-2062, June 2018, doi: 10.1109/TNNLS.2017.2773458.

[19] M. A. Nia, M. Kargahi and A. Abate, "Self-Adaptation with Imperfect Monitoring in Solar Energy Harvesting Systems," 2020 CSI/CPSSI International Symposium on Real-Time and Embedded Systems and Technologies (RTEST), Tehran, Iran, 2020, pp. 1-8, doi: 10.1109/RTEST49666.2020.9140148.

[20] Omid Gheibi, Danny Weyns, and Federico Quin, "Applying Machine Learning in Self-adaptive Systems: A Systematic Literature Review," ACM Trans. Auton. Adapt. Syst. 15, 3, Article 9 (September 2020), 37 pages, 2021. https://doi.org/10.1145/3469440

[21] J. O. Kephart and D. M. Chess, "The vision of autonomic computing," in Computer, vol. 36, no. 1, pp. 41-50, Jan. 2003, doi: 10.1109/MC.2003.1160055.

[22] Mehran Alidoost Nia, Mehdi Kargahi, Fathiyeh Faghih, "Probabilistic approximation of runtime quantitative verification in self-adaptive systems," Microprocessors and Microsystems, Volume 72, 2020, 102943.

[23] Jane Cleland-Huang, Theodore Chambers, Sebastian Zudaire, Muhammed Tawfiq Chowdhury, Ankit Agrawal, and Michael Vierhauser. 2024. Human–machine Teaming with Small Unmanned Aerial Systems in a MAPE-K Environment. ACM Trans. Auton. Adapt. Syst. 19, 1, Article 3 (March 2024), 35 pages. https://doi.org/10.1145/3618001

[24] Mehran Alidoost Nia, "Runtime Probabilistic Analysis of Self-Adaptive Systems via Formal Approximation Techniques," PhD diss., University of Tehran (Iran), 2022.

[25] Dongdong An, Jing Liu, Min Zhang, Xiaohong Chen, Mingsong Chen, Haiying Sun, Uncertainty modeling and runtime verification for autonomous vehicles driving control: A machine learning-based approach, Journal of Systems and Software, Volume 167, 2020, 110617, ISSN 0164-1212, https://doi.org/10.1016/j.jss.2020.110617.

[26] M. Dastranj, M. A. Nia and M. Kargahi, "Deploying Reinforcement Learning for Efficient Runtime Decision-Making in Autonomous Systems," 2022 CPSSI 4th International Symposium on Real-Time and Embedded Systems and Technologies (RTEST), Tehran, Iran, Islamic Republic of, 2022, pp. 1-9, doi: 10.1109/RTEST56034.2022.9850141.

[27] Gudela Grote, Safety and autonomy: A contradiction forever?, Safety Science, Volume 127, 2020, 104709, ISSN 0925-7535, https://doi.org/10.1016/j.ssci.2020.104709.

[28] S. Jha, V. Raman, D. Sadigh, S. A. Seshia, "Safe Autonomy Under Perception Uncertainty Using Chance-Constrained Temporal Logic." J Autom Reasoning 60, 2018, 43–62.

[29] Keele, Staffs, "Guidelines for performing systematic literature reviews in software engineering." (2007).

[30] Adina Aniculaesei, Jörg Grieser, Andreas Rausch, Karina Rehfeldt, and Tim Warnecke, "Towards a holistic software systems engineering approach for dependable autonomous systems" In Proceedings of the 1st International Workshop on Software Engineering for AI in Autonomous Systems (SEFAIS '18). Association for Computing Machinery, New York, NY, USA, 23–30, 2018, https://doi.org/10.1145/3194085.3194091

[31] M. Gorobetz, L. Ribickis, A. Levchenkov, and A. Beinarovica, "Machine learning algorithm of immune neuro-fuzzy anti-collision embedded system for autonomous unmanned aerial vehicles team," in Proceedings of the 2nd International Conference on Applications of Intelligent Systems, 2019.

[32] O. Minawi, J. Whelan, A. Almehmadi, and K. El-Khatib, "Machine learning-based intrusion detection system for controller area networks," in Proceedings of the 10th ACM Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications, 2020.

[33] P. Klose and R. Mester, "Simulated autonomous driving in a realistic driving environment using deep reinforcement learning and a deterministic finite state machine," in Proceedings of the 2nd International Conference on Applications of Intelligent Systems, 2019.

[34] F. Quin, D. Weyns, T. Bamelis, S. Singh Buttar, and S. Michiels, "Efficient analysis of large adaptation spaces in self-adaptive systems using machine learning," in 2019 IEEE/ACM 14th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), 2019.

[35] J. Li, Y. Yang, J. S. Sun, K. Tomsovic, and H. Qi, "ConAML: Constrained adversarial machine learning for cyber-physical systems," in Proceedings of the 2021 ACM Asia Conference on Computer and Communications Se-

curity, 2021.
[36] E. Khanapuri, T. Chintalapati, R. Sharma, and R. Gerdes, "Learning-based adversarial agent detection and identification in cyber physical systems applied to autonomous vehicular platoon," in 2019 IEEE/ACM 5th International Workshop on Software Engineering for Smart Cyber-Physical Systems (SEsCPS), 2019.
[37] F. Hasan, W. I. Khan Sarat, S. Khan, M. L. Hossain, M. T. Islam, and M. Ibrahim, "IoT based traffic management, accident detection, and accident prevention system using machine learning method," in Proceedings of the 2nd International Conference on Computing Advancements, 2022.
[38] S. Chiba and H. Sasaoka, "Effectiveness of transfer learning in autonomous driving using model car," in 2021 13th International Conference on Machine Learning and Computing, 2021.
[39] A. V. Chernov, I. K. Savvas, M. A. Butakova, and O. O. Kartashov, "Safe reinforcement learning in simulated environment of self-driving laboratory," in 2021 2nd European Symposium on Software Engineering, 2021.
[40] G. Weiss, P. Schleiss, D. Schneider, and M. Trapp, "Towards integrating undependable self-adaptive systems in safety-critical environments," in Proceedings of the 13th International Conference on Software Engineering for Adaptive and Self-Managing Systems, 2018.
[41] A. Stocco, P. J. Nunes, M. D'Amorim, and P. Tonella, "ThirdEye: Attention maps for safe autonomous driving systems," in Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering, 2022.
[42] H. Cao, X. Xi, G. Wu, R. Hu, and L. Liu, "ScanBot: Autonomous reconstruction via deep reinforcement learning," ACM Trans. Graph., vol. 42, no. 4, pp. 1–16, 2023.
[43] E. Vinitsky, N. Lichtlé, K. Parvate, and A. Bayen, "Optimizing mixed autonomy traffic flow with decentralized autonomous vehicles and multi-agent reinforcement learning," ACM Trans. Cyber-phys. Syst., vol. 7, no. 2, pp. 1–22, 2023.
[44] C. Lu, "Test scenario generation for autonomous driving systems with reinforcement learning," in 2023 IEEE/ACM 45th International Conference on Software Engineering: Companion Proceedings (ICSE-Companion), 2023.
[45] M. Data and F. A. Bakhtiar, "Resource efficient intrusion detection systems for internet of things using online machine-learning models," in Proceedings of the 8th International Conference on Sustainable Information Engineering and Technology, 2023.
[46] S. Dey and S.-W. Lee, "A multi-layered collaborative framework for evidence-driven data requirements engineering for machine learning-based safety-critical systems," in Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, 2023.
[47] M. A. Mukhtar Bhatti, M. Awais, and A. Iqtidar, "Machine Learning based Intrusion Detection System for IoT Applications using Explainable AI," in Proceedings of the 2023 Asia Conference on Artificial Intelligence, Machine Learning and Robotics, 2023.
[48] J. Cleland-Huang, T. Chambers, S. Zudaire, M. T. Chowdhury, A. Agrawal, and M. Vierhauser, "Human–machine Teaming with small unmanned aerial systems in a MAPE-K environment," ACM Trans. Auton. Adapt. Syst., vol. 19, no. 1, pp. 1–35, 2024.
[49] O. Sharma, N. C. Sahoo, and N. B. Puhan, "Dynamic planning of optimally safe Lane-change trajectory for autonomous driving on multi-lane highways using a fuzzy logic–based collision estimator," ACM J. Auton. Transport. Syst., vol. 1, no. 1, pp. 1–50, 2024.
[50] D. Pessach, T. Tassa, and E. Shmueli, "Fairness-driven private collaborative machine learning," ACM Trans. Intell. Syst. Technol., vol. 15, no. 2, pp. 1–30, 2024.
[51] M. T. Hossain, H. La, and S. Badsha, "RAMPART: Reinforcing Autonomous Multi-agent Protection through Adversarial Resistance in transportation," ACM J. Auton. Transport. Syst., vol. 1, no. 4, pp. 1–25, 2024.
[52] T. Mortlock, A. Malawade, K. Tsujio, and M. Al Faruque, "CASTNet: A context-aware, spatio-temporal dynamic motion prediction ensemble for autonomous driving," ACM Trans. Cyber-phys. Syst., vol. 8, no. 2, pp. 1–20, 2024.
[53] A. Esmaeili, J. Rayz, and E. Matson, "Hybrid algorithm selection and hyperparameter tuning on distribute machine learning resources: Hierarchical agent-based approach," ACM Trans. Internet Technol., vol. 24, no. 4, pp. 1–30, 2024.
[54] J. Ayerdi, A. Iriarte, P. Valle, I. Roman, M. Illarramendi, and A. Arrieta, "MarMot: Metamorphic runtime monitoring of autonomous Driving Systems," ACM Trans. Softw. Eng. Methodol., 2024.
[55] S. Saryazdi, B. Alkouz, A. Bouguettaya, and A. Lakhdari, "Using reinforcement learning and error models for drone precise landing," ACM Trans. Internet Technol., vol. 24, no. 3, pp. 1–30, 2024.
[56] D. Fényes, B. Németh, P. Gáspár, and Z. Szabó, "Road surface estimation based LPV control design for autonomous vehicles," IFAC-PapersOnLine, vol. 52, no. 28, pp. 120–125, 2019.
[57] M. E. Sahin, L. Tawalbeh, and F. Muheidat, "The security concerns on cyber-physical systems and potential risks analysis using machine learning," Procedia Comput. Sci., vol. 201, pp. 527–534, 2022.
[58] J. García and I. Sagredo, "Instance-based defense against adversarial attacks in Deep Reinforcement Learning," Eng. Appl. Artif. Intell., vol. 107, no. 104514, p. 104514, 2022.
[59] J. Bethge, B. Morabito, H. Rewald, A. Ahsan, S. Sorgatz, and R. Findeisen, "Modelling human driving behavior for constrained model predictive control in mixed traffic at intersections," IFAC-PapersOnLine, vol. 53, no. 2, pp. 14356–14362, 2020.
[60] Z. Long and W. Jinsong, "A hybrid method of entropy and SSAE-SVM based DDoS detection and mitigation mechanism in SDN," Comput. Secur., vol. 115, no. 102604, p. 102604, 2022.
[61] K. S. Tharayil et al., "Sensor Defense In-Software (SDI): Practical software based detection of spoofing attacks on position sensors," Eng. Appl. Artif. Intell., vol. 95, no. 103904, p. 103904, 2020.
[62] S. S. Volkov and I. I. Kurochkin, "Network attacks classification using Long Short-term memory based neural networks in Software-Defined Networks," Procedia Comput. Sci., vol. 178, pp. 394–403, 2020.
[63] D. Fényes, B. Németh, and P. Gáspár, "LPV-based control for automated driving using data-driven methods," IFAC-PapersOnLine, vol. 53, no. 2, pp. 13898–13903, 2020.
[64] L. Xie, Y. Ding, H. Yang, and Z. Hu, "Mitigating LFA through segment rerouting in IoT environment with traceroute flow abnormality detection," J. Netw. Comput. Appl., vol. 164, no. 102690, p. 102690, 2020.
[65] Q. Shangguan, T. Fu, J. Wang, T. Luo, and S. Fang, "An integrated methodology for real-time driving risk status prediction using naturalistic driving data," Accid. Anal. Prev., vol. 156, no. 106122, p. 106122, 2021.
[66] D. Fényes, B. Németh, and P. Gáspár, "Data-driven reachability analysis for the reconfiguration of vehicle control systems," IFAC-PapersOnLine, vol. 51, no. 24, pp. 831–836, 2018.
[67] I. Bolodurina and D. Parfenov, "Comprehensive approach for optimization traffic routing and using network resources in a virtual data center," Procedia Comput. Sci., vol. 136, pp. 62–71, 2018.
[68] R. Sankaranarayanan, K. S. Umadevi, N. P. G. Bhavani, B. M. Jos, A. Haldorai, and D. V. Babu, "Cluster-based attacks prevention algorithm for autonomous vehicles using machine learning algorithms," Comput. Electr. Eng., vol. 101, no. 108088, p. 108088, 2022.
[69] A. Saad and A. Håkansson, "RAMARL: Robustness analysis with multi-agent reinforcement learning - robust reasoning in autonomous cyber-physical systems," Procedia Comput. Sci., vol. 207, pp. 3662–3671, 2022.
[70] M. Luo, "Multi-sensor based strategy learning with deep reinforcement learning for unmanned ground vehicle," International Journal of Intelligent Networks, vol. 4, pp. 325–336, 2023.
[71] S. Alagumuthukrishnan, S. Deepajothi, R. Vani, and S. Velliangiri, "Reliable and efficient Lane changing behaviour for connected autonomous vehicle through deep reinforcement learning," Procedia Comput. Sci., vol. 218, pp. 1112–1121, 2023.
[72] N. Sehrawat, S. Vashisht, A. Singh, G. Dhiman, W. Viriyasitavat, and N. S. Alghamdi, "A power prediction approach for a solar-powered aerial vehicle enhanced by stacked machine learning technique," Comput. Electr. Eng., vol. 115, no. 109128, p. 109128, 2024.
[73] R. Ferede, G. de Croon, C. De Wagter, and D. Izzo, "End-to-end neural network based optimal quadcopter control," Rob. Auton. Syst., vol. 172, no. 104588, p. 104588, 2024.
[74] V. M. Gonçalves, D. Chaikalis, A. Tzes, and F. Khorrami, "Safe multi-agent drone control using control barrier functions and acceleration fields," Rob. Auton. Syst., vol. 172, no. 104601, p. 104601, 2024.
[75] L. L. F. Jahn, S. Park, Y. Lim, J. An, and G. Choi, "Enhancing lane detection with a lightweight collaborative late fusion model," Rob. Auton. Syst., vol. 175, no. 104680, p. 104680, 2024.
[76] V. Wiberg et al., "Sim-to-real transfer of active suspension control using deep reinforcement learning," Rob. Auton. Syst., vol. 179, no. 104731, p. 104731, 2024.

[77] A. Zachariae, F. Plahl, Y. Tang, I. Mamaev, B. Hein, and C. Wurll, "Human-robot interactions in autonomous hospital transports," Rob. Auton. Syst., vol. 179, no. 104755, p. 104755, 2024.
[78] N. Sayols et al., "Dynamic Global/Local multi-layer motion planner architecture for autonomous Cognitive Surgical Robots," Rob. Auton. Syst., vol. 180, no. 104758, p. 104758, 2024.
[79] R. Diaz, R. Ungo, K. Smith, L. Haghnegahdar, B. Singh, and T. Phuong, "Applications of AI/ML in maritime cyber supply chains," Procedia Comput. Sci., vol. 232, pp. 3247–3257, 2024.
[80] M. Matei and L. Alboaie, "Enhancing accessible navigation: A fusion of speech, gesture, and sonification for the visually impaired," Procedia Comput. Sci., vol. 246, pp. 2558–2567, 2024.
[81] A. Smacki, M. Kaup, M. Kaup, A. Wiktorowska-Jasik, and M. Baszak, "Cybersecurity of autonomous ships in offshore wind farms," Procedia Comput. Sci., vol. 246, pp. 5525–5534, 2024.
[82] A. Biondi, F. Nesti, G. Cicero, D. Casini and G. Buttazzo, "A Safe, Secure, and Predictable Software Architecture for Deep Learning in Safety-Critical Systems," in IEEE Embedded Systems Letters, vol. 12, no. 3, pp. 78-82, Sept. 2020, doi: 10.1109/LES.2019.2953253.
[83] Xie, Guo, Anqi Shangguan, Rong Fei, Wenjiang Ji, Weigang Ma, and Xinhong Hei, "Motion trajectory prediction based on a CNN-LSTM sequential model." Science China Information Sciences 63, 2020, 1-21.
[84] C. Sur, "UCRLF: unified constrained reinforcement learning framework for phase-aware architectures for autonomous vehicle signaling and trajectory optimization," Evol. Intell., vol. 12, no. 4, pp. 689–712, 2019.
[85] G. Fornari, V. A. de Santiago Júnior, and E. H. Shiguemori, "A self-adaptive approach for autonomous UAV navigation via computer vision," in Lecture Notes in Computer Science, Cham: Springer International Publishing, 2018, pp. 268–280.
[86] T. Hansmeier and M. Platzner, "Integrating safety guarantees into the learning classifier system XCS," in Lecture Notes in Computer Science, Cham: Springer International Publishing, 2022, pp. 386–401.
[87] M. Bhavsar, K. Roy, Z. Liu, J. Kelly, and B. Gokaraju, "Intrusion-based attack detection using machine learning techniques for connected autonomous vehicle," in Lecture Notes in Computer Science, Cham: Springer International Publishing, 2022, pp. 505–515.
[88] M. A. Langford, K. H. Chan, J. E. Fleck, P. K. McKinley, and B. H. C. Cheng, "MoDALAS: addressing assurance for learning-enabled autonomous systems in the face of uncertainty," Softw. Syst. Model., vol. 22, no. 5, pp. 1543–1563, 2023.
[89] L. Chen, Y. Xiao, X. Yuan, Y. Zhang, and J. Zhu, "Robust autonomous landing of UAVs in non-cooperative environments based on comprehensive terrain understanding," Sci. China Inf. Sci., vol. 65, no. 11, 2022.
[90] S. Ergün, "A study on multi-agent reinforcement learning for autonomous distribution vehicles," Iran J. Comput. Sci., vol. 6, no. 4, pp. 297–305, 2023.
[91] Z. Mahmood, K. Khan, M. Shahzad, A. Fayyaz, and U. Khan, "Enhanced detection and recognition system for vehicles and drivers using multi-scale retinex guided filter and machine learning," Multimed. Tools Appl., vol. 83, no. 6, pp. 15785–15824, 2023.
[92] S. Aurangzeb, M. Aleem, M. T. Khan, H. Anwar, and M. S. Siddique, "Cybersecurity for autonomous vehicles against malware attacks in smart-cities," Cluster Comput., vol. 27, no. 3, pp. 3363–3378, 2023.
[93] H. K. Öner and V. Sezer, "Local planner selection for autonomous robots using machine learning," Int. J. Intell. Robot. Appl., 2024.
[94] R. Singh, K. P. Sharma, and L. K. Awasthi, "A machine learning-based ensemble model for securing the IoT network," Cluster Comput., vol. 27, no. 8, pp. 10883–10897, 2024.
[95] T. Alam, R. Gupta, N. N. Ahamed, and A. Ullah, "A decision-making model for self-driving vehicles based on GPT-4V, federated reinforcement learning, and blockchain," Neural Comput. Appl., 2024.
[96] M. Al-Turki, M. T. Kashifi, N. T. Ratrout, and S. M. Rahman, "Traffic signal optimization framework using interpretable machine learning technique under heterogeneous-autonomy traffic environment," Neural Comput. Appl., vol. 36, no. 22, pp. 13761–13781, 2024.
[97] D. Kumar, "Reinforcement imitation learning for reliable and efficient autonomous navigation in complex environments," Neural Comput. Appl., vol. 36, no. 20, pp. 11945–11961, 2024.
[98] S. Alighanbari and N. L. Azad, "Safe adaptive deep reinforcement learning for autonomous driving in urban environments. Additional filter? How and where?," IEEE Access, vol. 9, pp. 141347–141359, 2021.
[99] Y. Dang, C. Benzaid, T. Taleb, B. Yang, and Y. Shen, "Transfer learning based GPS spoofing detection for cellular-connected UAVs," in 2022 International Wireless Communications and Mobile Computing (IWCMC), 2022, pp. 629–634.
[100] R. Gundu and M. Maleki, "Securing CAN bus in connected and autonomous vehicles using supervised machine learning approaches," in 2022 IEEE International Conference on Electro Information Technology (eIT), 2022, pp. 042–046.
[101] S. Josef and A. Degani, "Deep reinforcement learning for safe local planning of a ground vehicle in unknown rough terrain," IEEE Robot. Autom. Lett., vol. 5, no. 4, pp. 6748–6755, 2020.
[102] Santos, M. F., & Victorino, A. C. (2021). Autonomous vehicle navigation based in a hybrid methodology: model based and machine learning based. 2021 IEEE International Conference on Mechatronics (ICM), 1–6.
[103] J. Grieser, M. Zhang, T. Warnecke, and A. Rausch, "Assuring the safety of end-to-end learning-based autonomous driving through runtime monitoring," in 2020 23rd Euromicro Conference on Digital System Design (DSD), 2020, pp. 476–483.
[104] X. Shi, Y. D. Wong, C. Chai, and M. Z.-F. Li, "An automated machine learning (AutoML) method of risk prediction for decision-making of autonomous vehicles," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 11, pp. 7145–7154, 2021.
[105] B. Mirchevska, C. Pek, M. Werling, M. Althoff, and J. Boedecker, "High-level decision making for safe and reasonable autonomous Lane changing using reinforcement learning," in 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 2018.
[106] H. Krasowski, X. Wang, and M. Althoff, "Safe reinforcement learning for autonomous Lane changing using set-based prediction," in 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 2020.
[107] V. Chen, M.-K. Yoon, and Z. Shao, "Novelty detection via network saliency in visual-based deep learning," in 2019 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W), 2019.
[108] S. Mihai, P. Shah, G. Mapp, H. Nguyen, and R. Trestian, "Towards autonomous driving: A machine learning-based pedestrian detection system using 16-layer LiDAR," in 2020 13th International Conference on Communications (COMM), 2020.
[109] F. S. Acerbo, H. Van der Auweraer, and T. Duy Son, "Safe and computational efficient imitation learning for autonomous vehicle driving," in 2020 American Control Conference (ACC), 2020.
[110] Y. Chen, C. Dong, P. Palanisamy, P. Mudalige, K. Muelling, and J. M. Dolan, "Attention-based hierarchical deep reinforcement learning for Lane change behaviors in autonomous driving," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019.
[111] A. Alfardus and D. B. Rawat, "Intrusion detection system for CAN bus in-vehicle network based on machine learning algorithms," in 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2021.
[112] G. Madhumitha, R. Senthilnathan, K. M. Ayaz, J. Vignesh, and K. Madhu, "Estimation of collision priority on traffic videos using deep learning," in 2020 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), 2020.
[113] Y. Trivedi and P. Negandhi, "An advanced driver assistance system using computer vision and deep-learning," in 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018.
[114] A. A. Ishola and D. Valles, "Using machine learning and regression analysis to classify and predict danger levels in burning sites," in 2022 IEEE World AI IoT Congress (AIIoT), 2022.
[115] C. Ryan, F. Murphy, and M. Mullins, "End-to-end autonomous driving risk analysis: A behavioural anomaly detection approach," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 3, pp. 1650–1662, 2021.
[116] S. Samani, R. Jessop, and A. Harrivel, "Collaborative communications between A human and A resilient safety support system," in 2021 IEEE International Conference on Autonomous Systems (ICAS), 2021, pp. 1–5.
[117] A. Dagvasumberel, B. Myagmardulam, B. Myagmar, B. Luvsankhuu, and T. Nakayama, "Railroad near-miss occurrence detection and risk estimation system with data from camera using deep learning," in 2021 5th International Conference on Imaging, Signal Processing and Communications (ICISPC), 2021.
[118] Q. Lin, X. Chen, A. Khurana, and J. M. Dolan, "ReachFlow: An online

safety assurance framework for waypoint-following of self-driving cars," in 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020.
[119] T.-Y. Chen, H.-J. Lin, C.-S. Shih, K.-T. Kuo, Q. Liu, and R. H. M. Chan, "Prediction of human intention in vehicles, pedestrians and bicyclists interactions," in 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), 2021.
[120] J. Ciberlin, R. Grbic, N. Teslic, and M. Pilipovic, "Object detection and object tracking in front of the vehicle using front view camera," in 2019 Zooming Innovation in Consumer Technologies Conference (ZINC), 2019.
[121] I. Das, S. Roy, S. Dey, T. Dey, and P. Roy, "Emergency medical assistance by ambulance drone using machine learning, light-weight cryptography and variable image steganography," in 2022 IEEE VLSI Device Circuit and System (VLSI DCS), 2022, pp. 134–138.
[122] M. Hemmati, M. Biglari-Abhari, and S. Niar, "Adaptive vehicle detection for real-time autonomous driving system," in 2019 Design, Automation & Test in Europe Conference & Exhibition (DATE), 2019.
[123] F. Ye, X. Cheng, P. Wang, C.-Y. Chan, and J. Zhang, "Automated Lane change strategy using proximal policy optimization-based deep reinforcement learning," in 2020 IEEE Intelligent Vehicles Symposium (IV), 2020.
[124] Li, L., Zhao, W., Xu, C., Wang, C., Chen, Q., & Dai, S. (2021). Lane-change intention inference based on RNN for autonomous driving on highways. IEEE Transactions on Vehicular Technology, 70(6), 5499–5510. https://doi.org/10.1109/tvt.2021.3079263
[125] A. Abraham, Y. Zhang, and S. Prasad, "Real-time prediction of multi-class Lane-changing intentions based on highway vehicle trajectories," in 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), 2021.
[126] C. Guizzaro, F. Formaggio, and S. Tomasin, "GNSS spoofing attack detection by IMU measurements through A neural network," in 2022 10th Workshop on Satellite Navigation Technology (NAVITEC), 2022.
[127] A. Majd, A. Ashraf, E. Troubitsyna, and M. Daneshtalab, "Using optimization, learning, and drone reflexes to maximize safety of swarms of drones," in 2018 IEEE Congress on Evolutionary Computation (CEC), 2018.
[128] L. Sequeira, A. Szefer, J. Slome, and T. Mahmoodi, "A Lane merge coordination model for a V2X scenario," in 2019 European Conference on Networks and Communications (EuCNC), 2019.
[129] T. Rehder, A. Koenig, M. Goehl, L. Louis, and D. Schramm, "Lane change intention awareness for assisted and automated driving on highways," IEEE Trans. Intell. Veh., vol. 4, no. 2, pp. 265–276, 2019.
[130] K. Min, H. Kim, and K. Huh, "Deep Q learning based High Level driving policy determination," in 2018 IEEE Intelligent Vehicles Symposium (IV), 2018.
[131] Fenyes, D., Nemeth, B., & Gaspar, P. (2020). LPV-based autonomous vehicle control using the results of big data analysis on lateral dynamics. 2020 American Control Conference (ACC), 2250–2255.
[132] W. Yuan, M. Yang, Y. He, C. Wang, and B. Wang, "Multi-reward architecture based reinforcement learning for highway driving policies," in 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 2019.
[133] Farag, W., & Saleh, Z. (2018). Behavior cloning for autonomous driving using convolutional neural networks. 2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), 1–7.
[134] M. Everett, Y. F. Chen, and J. P. How, "Collision avoidance in pedestrian-rich environments with deep reinforcement learning," IEEE Access, vol. 9, pp. 10357–10377, 2021.
[135] K. Sama, Y. Morales, N. Akai, E. Takeuchi, and K. Takeda, "Learning how to drive in blind intersections from human data," in 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2018.
[136] A. Aloseel, S. Al-Rubaye, A. Zolotas, and C. Shaw, "Attack-detection architectural framework based on anomalous patterns of system performance and resource utilization—part II," IEEE Access, vol. 9, pp. 87611–87629, 2021.
[137] L. Zhang and D. Ma, "A hybrid approach toward efficient and accurate intrusion detection for in-vehicle networks," IEEE Access, vol. 10, pp. 10852–10866, 2022.
[138] S. He et al., "Automatic recognition of traffic signs based on visual inspection," IEEE Access, vol. 9, pp. 43253–43261, 2021.
[139] H. Hu, X. Jia, K. Liu, and B. Sun, "Self-adaptive traffic control model with behavior trees and reinforcement learning for AGV in industry 4.0," IEEE Trans. Industr. Inform., vol. 17, no. 12, pp. 7968–7979, 2021.
[140] F. Schmoeller Roza, H. Rasheed, K. Roscher, X. Ning, and S. Günnemann, "Safe robot navigation using constrained hierarchical reinforcement learning," in 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA), 2022.
[141] S. Zhang, S. Wang, S. Yu, J. J. Q. Yu, and M. Wen, "Collision avoidance predictive motion planning based on integrated perception and V2V communication," IEEE Trans. Intell. Transp. Syst., vol. 23, no. 7, pp. 9640–9653, 2022.
[142] S. R. Kothuri, V. Nivedita, Padmaloshani, M. Dharmateja, A. Chinnaraj, and K. B. Sachidananda, "Enhancing pedestrian safety in autonomous vehicles through machine learning," in 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA), 2023, pp. 1587–1592.
[143] Y. Hu, J. Fu, and G. Wen, "Safe reinforcement learning for model-reference trajectory tracking of uncertain autonomous vehicles with model-based acceleration," IEEE Trans. Intell. Veh., vol. 8, no. 3, pp. 2332–2344, 2023.
[144] F. Shafique, T. Naeem, and A. H. Farooqi, "Path tracking and obstacle avoidance control using deep reinforcement learning for autonomous vehicles," in 2023 17th International Conference on Open Source Systems and Technologies (ICOSST), 2023.
[145] M. M. Abrar and S. Hariri, "An Anomaly Behavior Analysis framework for securing autonomous vehicle perception," arXiv [cs.RO], 2023.
[146] A. Prabu, L. Li, Y. Chen, and B. King, "Data-driven deep reinforcement learning for automated driving," in 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), 2023.
[147] B. B. Elallid, H. E. Alaoui, and N. Benamar, "Deep reinforcement learning for autonomous vehicle intersection navigation," arXiv [cs.RO], 2023.
[148] A. Brunoud, A. Lombard, A. Abbas-Turki, N. Gaud, and J. Kang-Hyun, "Hybrid deep reinforcement learning model for safe and efficient autonomous vehicles at pedestrian crossings," in 2023 International Workshop on Intelligent Systems (IWIS), 2023, pp. 1–6.
[149] S. Jagan, R. Pokhariyal, K. Mahajan, C. L. N. Deepika, P. D. Sudha, and A. Dutta, "Machine learning with deep learning approach for cyber security threats prevention model," in 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), 2023, pp. 1–5.
[150] S. Li, S. Yang, L. Wang, and Y. Huang, "Safe reinforcement learning for Lane-changing with comprehensive analysis of safety detection," in 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), 2023.
[151] M. Wagner and A. A. Fröhlich, "Securing cyber-physical systems against GPS spoofing attacks using confidence attribution," in 2023 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2023, pp. 1–6.
[152] K. Thangavel, N. E.-D. Safwat, K. F. Hussain, A. Gardi, and R. Sabatini, "A simulation tool for space traffic management," in 2024 AIAA DATC/IEEE 43rd Digital Avionics Systems Conference (DASC), 2024, pp. 1–6.
[153] P. Haritz, D. Wanke, and T. Liebig, "Enhancing safety for autonomous agents in partly concealed urban traffic environments through representation-based shielding," in 2024 IEEE Intelligent Vehicles Symposium (IV), 2024, pp. 1758–1763.
[154] F. Islam, J. E. Ball, and C. T. Goodin, "Enhancing longitudinal velocity control with attention mechanism-based deep deterministic policy gradient (DDPG) for safety and comfort," IEEE Access, vol. 12, pp. 30765–30780, 2024.
[155] N. Potteiger and X. Koutsoukos, "Safeguarding autonomous UAV navigation: Agent design using evolving behavior trees," in 2024 IEEE International Systems Conference (SysCon), 2024, pp. 1–8.
[156] J. T. Grini, R. Mester, T. A. Nygård, N. Dalhaug, E. F. Brekke, and A. Stahl, "FusedWSS: Water surface segmentation fusing machine learning and geometric cues," in 2024 27th International Conference on Information Fusion (FUSION), 2024, pp. 1–8.
[157] J. Ghelani, P. Gharia, and H. El-Ocla, "Gradient monitored reinforcement learning for jamming attack detection in FANETs," IEEE Access, vol. 12, pp. 23081–23095, 2024.
[158] A. Mahmud and C. Huang, "Learning a safe driving policy for urban area from real-world data," in 2024 6th International Conference on Reconfigurable Mechanisms and Robots (ReMAR), 2024, vol. 22, pp. 493–499.