

A Review Study on Functional Safety Verification and Validation for Autonomous Vehicles

Cuneyt Haspolat¹, Ecem Celik¹, Mert Aygun¹, Isil Cosar¹, and Melike Karaduman¹

¹ AVL Research and Engineering Turkey, Istanbul, Turkey
cuneyt.haspolat@avl.com, ecem.celik@avl.com, mert.aygun@avl.com, isil.cosar@avl.com, and
melike.karaduman@avl.com

Abstract

In recent years, autonomous vehicle studies have been rapidly increasing with technological developments in the automotive industry. While improving autonomous driving performances and increasing the diversity of autonomous vehicles, these studies bring new challenges as well. In this study, a review study on functional safety verification and validation for autonomous vehicles, which are developed using machine learning, neural networks, and artificial intelligence structures, is presented considering as one of the main challenges in this domain. Within the scope of this study, “What kind of difficulties exist in safety verification and validation processes”, “Which theoretical and experimental methods are proposed against the stated difficulties”, “Which software and hardware tools are used”, “Which approaches are recommended for ISO 26262 compliance” and “how autonomy level changes the existing verification and validation methods” are investigated. A comprehensive literature review including highlights from papers is presented with respect to the key research questions.

1. Introduction

Nowadays, Autonomous Vehicles (AVs) with different autonomy levels based on the SAE J3016 (Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles [1]) have been increasingly integrated with the automotive industry. The SAE J3016 defines six levels of driving automation from no automation (level 0) to full vehicle autonomy (level 5). The vehicle functionality becomes more complex as the level of autonomy increases. Especially for level 4 and level 5 AVs, Autonomous Driving (AD) and Advanced Driver Assistance System (ADAS) functions become more complex compared to their conventional and partially autonomous counterparts. To manage this complexity, researchers and manufacturers have given attention to Machine Learning (ML), Neural Networks (NNs) and Artificial Intelligence (AI) based applications with increasing dominance in the automotive world [16]. It is obvious that this worldwide increasing interest in AI-based systems produces new application areas, especially for which come from the automotive industry, and researchers and manufacturers are competing with each other to showcase originality by solving challenges. As shown in numerous applications in non-automotive domains, AI-based solutions have the capability to work with complex systems and to learn even when there is insufficient information [2-7].

Although it has been shown that AI-based solutions are powerful in solving challenging AV problems [14, 15, 31, 37, 41] by making remarkable advancements possible in today's AD

technologies, they also bring new challenges in terms of robust system design, wide-range testing, and confirmation for public use. For example, AD decision-making processes require a variety of interactions with the vehicle environment in various known and unknown scenarios and it is quite a challenging task to evaluate all edge cases that AI-based system that might fail or react in a safe manner in dangerous situations. In this context, AI-based applications in safety-relevant systems need to be analysed deeply since these systems may lead to a crash which may cause injury or death to the humans that interact with them [3]. As stated in [7], an AV could have an acceptable error rate as its own; however, this error rate becomes inadmissible when considering the total error rates of millions of AVs. Therefore, apart from the fact that it is an essential part of certification processes, Verification and Validation (V&V) is one of the core challenges of AVs that requires consideration of potential traffic scenarios and influencing factors. As stated in [6], there are many studies conducted focusing on V&V methods of software developed with AI. But these studies primarily focus on the reliability or robustness of the automotive control systems, rather than their safety. There has been far less written about the engineering safety of such algorithms. Although various methods have been put forth on how to construct AI-based systems, there is no consensus yet. For the sake of easier analysis and clear interpretation of AI-based systems, contemporary research initiatives frequently target the term opaqueness [17].

Automotive authorities have also highlighted another key aspect of AVs, which is the V&V of AI-based safety-critical systems in the scope of ISO 26262 [6]. The international standard, ISO 26262 Road Vehicles – Functional Safety [8], defines functional safety as “Absence of unreasonable risk due to hazards caused by malfunctioning behaviour of Electrical/Electronic systems”. From this point of view, an AV (e.g., AI-based ADAS/AD software) can be proven as robust or reliable up to a certain point, but it might be unsafe due to the lack of required V&V procedures listed in the scope of functional safety. ISO 26262 seems still deficient in interpretability, verification, validation, and performance limitation against the challenges of AI, as it is designed to cover systematic system and random hardware failures which are as likely deterministic. Accordingly, structural concepts like ML and NN seem to have an impact on well-established deterministic V&V processes. Especially concerning today's technology in AV development, it is still an open research topic how V&V methods could cover various aspects of AI-based systems according to ISO 26262. The major challenge on V&V of AI-based systems is that these systems come up with a lack of design information such that predefined pass/fail criteria cannot be evaluated in formal common ways. Besides, these systems can be rated as black box systems, for which formal safety analysis (e.g., FMEA) cannot be performed

easily. As stated in numerous studies, even though V&V is a crucial step in the realization of functional safety activities in compliance with ISO 26262, the well-known procedures cannot be applied directly in AV applications [4, 5].

In this paper, we presented a review study on functional safety V&V for AVs in order to investigate recent concerns and practices of automotive community in the area of safety V&V of AI-based systems of AVs. For this purpose, after interpreting concerning V&V topics by Research Questions (RQs), literature papers are examined to find out answers for following key aspects in the direction of RQs; What kind of difficulties exist in safety V&V processes? Which theoretical and experimental methods are suggested against these difficulties? Which software and hardware tools are used to handle these difficulties? Which V&V approaches are recommended for ISO 26262 compliance? How autonomy level changes the existing V&V methods. The relevant papers in the literature are analysed in a systematic manner and the main highlights of this literature review are presented.

The remainder of the paper is structured in the following manner: subjects are given in Section II, the research methodology is presented in Section III, the literature review is examined in Section IV, the conclusion is given in Section V.

2. Research Questions

As V&V, the key factor of automotive safety applications, becomes one of the biggest challenges to prove the novelty of the new AI-based components. Numerous researchers have also worked on and contributed to V&V with regard to various automotive systems and performance traits. Although V&V is a crucial component of safety certification, it is currently unclear how to create safety cases for specific applications which are utilizing AI-based methods [16]. That is why one of the backbone questions, in other words the Research Questions (RQs), to be answered is "What are the difficulties in safety V&V processes arising from AI-based systems?" so that the safety aspect of such applications is still ambiguous. The main challenging characteristics of AI-based systems can be summarized in terms of uncertainty about the system's behaviour, unpredictable systematic failures, quality of processed data, and vulnerabilities against untrained inputs. The second RQ is linked to best practices throughout the V&V processes seeking an answer to "What sort of theoretical and experimental methods and tools have been proposed or can be used for the safety V&V of AI-based algorithms?". Apart from these V&V-oriented concerns, the challenges in higher autonomy levels are selected as the third RQ of this paper.

In this context, the following RQs are considered in this paper:

RQ-1: What are the difficulties in safety V&V processes arising from AI-based systems in AVs?

RQ-2: What are the theoretical or experimental methods and tools proposed for the Safety V&V processes of AI-based systems in AVs?

RQ-3: How will greater autonomy change the approaches to verify and validate AI-based algorithms in AVs?

3. Research Methodology

3.1. Review Policy

A comprehensive pool of related papers is collected with a systematic literature research method by searching various databases such as Google Scholar and IEEE Xplore. For this purpose, literature papers are selected according to keywords and

a combination of these keywords to find all papers related to V&V and ML/NN/AI. The Research Keywords (RKs) are chosen to find the best answers to the RQs from the closest related studies. The RKs are given in Table 1.

Table 1. Research Keywords

	Keyword Combination
RK-1	Validation OR Verification OR Ver&Val OR V&V OR Scenario OR Situation
RK-2	Functional Safety OR ISO 26262 OR Safety Assurance
RK-3	Autonomous Vehicle OR AV OR Unmanned OR Self Driving Car OR Driverless Vehicle
RK-4	Machine Learning OR ML OR Artificial Intelligence OR AI OR Deep Learning OR Neural Network OR NN

The RK-1 collects studies that mainly focus on V&V area including scenario generation and situation generation (of AVs), while the RK-2 finds studies on a general functional safety scope by indexing safety assurance, functional safety, and ISO 26262 keywords. The RK-3 lists AV studies (which include V&V scope). Finally, the RK-4 is intended to collect remaining studies on ML, NN and AI from the V&V perspective.

3.2. Inclusion and Exclusion Criteria

Once all related papers are listed, they are processed according to Inclusion Criteria (IC) and Exclusion Criteria (EC) to find out which study addresses which RQs, before starting paper evaluation. In this context, the papers that fulfil IC are included in the evaluation, while the papers that fulfil EC are excluded from the evaluation. The executed IC and EC are given in Table 2.

Table 2. Inclusion and Exclusion Criteria

	Keyword Combination
IC-1	Does the paper examine the latest status of AVs (including ADAS/AD features) regarding V&V safety?
IC-2	Does the paper make recommendations for standardization (methods, tools, procedures) of V&V methods of AVs?
IC-3	Does the paper provide information on software environments or test procedures related to the V&V of AVs?
IC-4	Does the paper give information about V&V of safety-critical systems along with the levels of AVs?
EC-1	Is the paper language different than English?
EC-2	Is the paper not peer-reviewed (conference, journal)?
EC-3	Is the paper's publishing date before 2016?

4. Literature Review

In this section, research questions related to Safety V&V processes of AVs are discussed. First, the difficulties that are encountered by AI-based systems are examined, and then the papers on theoretical or experimental methods/tools are presented. Finally, considering the effects of autonomy levels on this topic, relevant papers are examined. The research papers related to defined RQs are listed in Table 3.

Table 3. Research Papers

	Related Papers
RQ-1	[9], [10], [11], [12], [13], [14]
RQ-2	[11], [13], [37], [38]
RQ-3	[39], [40], [41], [42], [43]

4.1. Research Question 1

AI-based algorithms are not yet approved by industry practice for use in safety-critical applications of AVs on public roadways. The reviewed papers for this study demonstrate that engineering professionals from various organizations lack the necessary skills to approach regulatory authorities and certification processes with AI-based solutions. Since the existing gap does not appear to be closable soon through incremental research alone, it is most likely that automobile manufacturers and safety regulations will need to change significantly to meet the new paradigm that comes with AI-based systems in AVs. To begin with, new challenges that come with AI-based solutions for V&V processes, which are essential to the safety concept, are to be identified as a starting point. The main highlights of the reviewed papers in this given context are presented below.

Zhao et al. [9] studied the formal methods for the safety assessment from the perspective of V&V. They emphasized that considering all the possible challenges in the V&V area for AV applications there are four main challenges which should be paid attention to. These can be generalized as: the resolution of the AI systems responsibility, the economic concerns of verification for legitimate scenario coverage and the expense of re-verification. They state that the general approach regarding V&V is to apply scenario coverage, but it brings challenges which is hard to evaluate enough samples so that it will cover the risks up to an acceptable level. Therefore, they suggest a method which is called a unified safe scenario coverage framework. That methodology can be used for formal methods and sample-based methods.

Koopman and Wagner [10] analyse that there are five important challenges in the V&V of ISO 26262 when considering AI-based components in the automotive industry which are driver out of the loop, dealing with complex requirements, handling non-deterministic algorithms, incorporating inductive learning algorithms, and addressing fail-operational systems. They also mention stand-out approaches to the challenges as the use of relaxed operational scenarios for phased deployment, monitor/actuator pair architecture usage to distinguish between the complex autonomy functions and simpler safety functions. Additionally, employing fault injection proves effective in conducting more efficient edge case testing.

Schwalbe and Schels [11] discuss the topic in a more generalized way and refer to every aspect of ISO 26262 from the perspective of AI systems. They address that even though the approaches which are used regarding verification of the AI-based applications are developed enough, they offer that rule checkers can be implemented to use in the first-order logics such as Satisfiability Module Theory-based solvers to achieve certain performance levels in a more effective way. They also review open challenges from a validation point of view and identify two conditions which should be focused. These are the inner logic and representation of the ML components and testing and training the data.

Gharib et al. [12] describes challenges that emerge from ML-based component usage in the automotive industry after they analyse the constraints of ISO 26262 and Safety of The Intended

Functionality (SOTIF) ISO/PAS 21448 from the perspective of ML applications. They explain the possible results of using ML algorithms in safety-critical systems in the following: ML algorithms complication if the safety requirements are traceable or not, the ability to generalize to novel situations. All the mentioned aspects have the possibility to create major issues from the V&V perspective when one tries to adapt the ML-based components in functional safety applications.

Mohseni et al. [13] highlights that although ML is a hot topic in engineering, the automotive industry and the safety standard are new to it. The application of ML-based systems to the safety standard is under academic scrutiny, and the public introduction of self-driving cars in the automotive industry is still a matter of debate. They also state that V&V and that robustness is a challenging issue for ML-based components in the automotive field and makes recommendations for compliance with the standard.

Haeri et al. [14] underlines that with the difficulty of understanding the effects of AI on safety and security, even if it has success in applications such as machine reasoning and control. In ML, the system is dynamically formed by selecting the best-performing model from the tested systems with many inputs and outputs. However, the inadequacy of the ISO 26262 standard is emphasized due to the dynamic learning methods of ML and line and branch coverage in the V&V processes.

4.2. Research Question 2

Specific tools and methods are required to properly adapt the V&V processes to AI-based systems. As discussed throughout the paper, this brings many challenges. Therefore, it is of significant importance to create appropriate and purposeful tools and methods in order to manage these processes correctly. This paper, it is aimed to make a summary of the tools and methods obtained at the experimental or theoretical level by asking RQ-2. This chapter is divided into two folds; first, the V&V methods are examined then the V&V tools are focused on in the second part. It is clearly seen that although we have a substantial number of tools and methods, the literature is still open to development considering the advantages and disadvantages of tools and methods. The main highlights of the reviewed papers in this context are presented below.

Schwalbe and Schels [11] address solvers, output bound estimation and search algorithms in the verification section and data validation, qualitative analysis, and quantitative analysis in the validation section as generally preferred approaches for the V&V of the ISO 26262, respectively. Satisfiability Modulo Theory [18] or mixed integer linear programming [19] can be an example to the solvers. ReluVal [20] is used to optimize networks when considering the output-bound estimation approach. They are referring to the work that is conducted by Pei et al. [21] which they implement a methodology called VERIVIS to verify the safety aspects of the computer vision systems when it comes to giving examples of such search algorithms which can be used for verification of ML-based components for further research. [11] is referring to three main approaches if the focus is on the validation part of ISO 26262. In the data validation part, NN-specific metrics with a test case generation framework [22] and tools which is mature enough for counterexample generation using efficient fuzzy testing [23] are given as examples. Local linear approximations [24] can be counted as one of the examples of the qualitative analysis which is used for validation of the ML applications. Feature visualization [25] and additional explanatory output [26] approaches can be further examples for

the qualitative analysis part of the paper. They pointed out that techniques for the quantitative analysis particularly are limited at this point. Local or global extraction of rules using inductive logic programming can be taken into consideration as an example of the topic. Performance analysis on sub-tasks [27] and distributed representation of semantic domain concepts [28] can be other techniques which can be used as quantitative analysis methods. Net2Vec [29] and TCAV [30] are given as further examples of the representation of semantic domain concepts by the writers of the paper.

Mohseni et al. [13] addresses practical safety techniques to implement ML and AI-based systems such as design specification [31], implementation transparency [32], formal verification, uncertainty estimation, in-distribution error detection [33], out-of-distribution error detection [34], domain generalization [35], perturbation and corruption robustness [36].

Wishart et al. [37] evaluates a literature review regarding V&V aspects of the AV's world. They collect simulation platforms through detailed research that has been done on the complete literature. The findings of the driving simulation platforms can be found as follows: Apollo by Baidu, Applied Intuition, AutoVi-Sim by UNC Gamma, Autoware, Carcraft (Waymo), CARLA, CarSim, Cesium, Cognata, Dactle, EB Robinos, Metamoto, NVIDIA DriveWorks, PreScan, Pro-SiVIC by ESI Group, rFpro, and finally VRXPERIENCE by ANSYS. All these simulation platforms are commercially available regarding AV applications in the automotive industry and offer different advantages for the development and testing of the systems.

Tahir Z. et al. [38] presents the AV-test framework and defines state hyperspace for situation coverage. State hyperspace includes different axes such as environmental conditions. They used the CARLA simulation program and investigated vehicle-vehicle interactions. The fact that CARLA has some advantages over other simulation programs has been a key factor in the authors' use of CARLA. Some of these advantages contributed to the V&V processes are the capability to test different environmental conditions and scenarios and to create scenarios.

4.3. Research Question 3

In the first autonomy levels according to SAE J3016, it is expected that the driver fulfils some responsibilities, such as keeping the steering wheel constantly, while ADAS functions are active. At the last levels, actions such as minimum risk manoeuvre and safe stop are expected to be performed by fully AD features. AI-based algorithms are also included in the software for these features that gradually take place in the automotive domain. Considering the changes in the level of autonomy and the necessity of extensive testing processes for AI-based algorithms, significant changes are required in the common well-established V&V processes of functional safety. V&V processes aim to provide different adaptations according to autonomy levels of AVs due to geographical, climatic, and infrastructural differences and road participants' behaviour. Accordingly, the creation/scale/content/quality/quantity of V&V scenarios are interpreted as key concepts in the V&V processes. The main highlights of the reviewed papers are presented below.

Menzel et al. [39] states that there are differences in the use of scenarios throughout the entire V-cycle in a project made according to the ISO 26262 standard. They make recommendations for creating a systematic approach to scenarios used in different phases of the process. They examined the scenarios as scenarios in the concept phase, the scenarios in the

system development phase, and the scenarios used in the V&V phase and suggested certain concepts for a systematic approach. These are representation scenarios in a semi-formal way, scenarios formulation in natural language by experts, representation scenarios with state value's parameter ranges and formal notation of parameter range and representation scenarios model without inconsistency by ensuring an automated test execution and reproducibility. In addition, they stated with examples that it would be advantageous to separate the scenarios into three levels as functional, logical, and concrete scenarios in the design and test processes.

Sadigh et.al [40] emphasized the importance of the trajectories and behaviours of other vehicles in the training of AVs. Likewise, the trajectory and behaviour of the AV affect other vehicles. They emphasize that the interaction becomes more critical as the level of autonomy increases. In order to examine the effects of human and AVs' behaviour, they created a dynamic model and performed the training with reverse reinforcement learning. To better examine this two-sided interaction, they arranged the test scenarios accordingly and carried out design and verification tests in a human-in-the-loop simulation environment.

Feng et al. [41] and Feng et al. [42] studied connected automated vehicles as Part 1 methodology and Part 2 Case Studies. In these studies, their aim is to create a generalizable testing scenario library generation for a connected automated vehicle. For the given operational design domains, a methodology is proposed to find the test set and critical test scenarios suitable for connected AVs. In addition to finding the criticality of the scenario with certain parameters and multi-objective optimization methods, reinforcement learning is also used for high-dimensional scenarios. The proposed method was tested with three common scenarios identified as Case Studies. Achieving similar accuracy rates to on-road tests with fewer test sets is theoretically presented and it is emphasized that it may speed up the process.

Bach et.al [43] presented a new methodology, stating that it is not sufficient for the testing and validation of complex driving scenarios. In this method, the determination of domain model-based understandable driving scenarios is carried out by abstracting time and spatial information. With the characterization and parameterization of the scenarios with the proposed method, the automatic generation of test scenarios ready to be assessed in the simulator has been ensured.

6. Conclusions

In this study, it has been examined the challenges in the field of V&V of safety-critical AI-based applications consumed by AVs, how such applications are in compliance with ISO 26262 and what kind of studies have been presented in the literature to keep them in line with ISO 26262. In addition, it investigated the solutions and approaches that are proposed in the context of V&V processes in the safety of AI-based applications, and whether these suggestions are theoretical or experimental. In the suggestions presented with experimental setups, which software tools are used, if any, and what kind of hardware is used are also examined. Finally, the effects of increasing autonomy on safety applications are discussed from a V&V perspective. All in all, considering the answers to the RQs and the associated literature together, it is obvious that there are concrete steps towards to safety concepts of such cutting-edge technologies. Even though we found a lot of ideas and how-to-dos around this topic, there is not much work that we can categorize as a real safety application

that is applied to a real system. Therefore, the realization of these proposed ideas remains a big challenge ahead.

7. References

- [1] SAE International, "SAE J3016 202104: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," 2021.
- [2] Z. Kurd, T. Kelly, and J. Austin, "Developing artificial neural networks for safety critical systems," *Neural Computing and Applications*, vol. 16, no. 1, pp. 11–19, 2007.
- [3] M. Bozzano, & A. Villafiorita, "Design and safety assessment of critical systems." *CRC press*, 2010.
- [4] R. Rana, M. Staron, C. Berger, J. Hansson, M. Nilsson, & F. Törner, "Early verification and validation according to iso 26262 by combining fault injection and mutation testing." In *International Conference on Software-Technologies* (pp. 164-179). Springer, Berlin, Heidelberg. 2013.
- [5] K. R. Varshney, "Engineering safety in machine learning," *Information Theory and Applications Workshop, ITA 2016*, no. 1601.04126v1, 2016.
- [6] P. Koopman and M. Wagner, "Autonomous vehicle safety: An interdisciplinary challenge," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 1, pp. 90–96, 2017.
- [7] J. Wang, L. Zhang, Y. Huang, and J. Zhao, "Safety of autonomous vehicles," *Journal of advanced transportation*, vol. 2020.
- [8] ISO, 26262:2018 - road vehicles – functional safety," International Organization for Standardization in ISO 26262, 2018.
- [9] T. Zhao, E. Yurtsever, J. Paulson, & G. Rizzoni, "Formal Certification Methods for Automated Vehicle Safety Assessment.," *IEEE Transactions on Intelligent Vehicles*. 2022
- [10] P. Koopman, & M. Wagner, "Challenges in autonomous vehicle testing and validation." *SAE International Journal of Transportation Safety*, 4(1), 15-24. 2016
- [11] G. Schwalbe, & M. Schels, "A survey on methods for the safety assurance of machine learning based systems." In *10th European Congress on Embedded Real Time Software and Systems (ERTS 2020)*. 2020
- [12] M. Gharib, P. Lollini, M. Botta, E. Amparore, S. Donatelli, & A. Bondavalli, "On the safety of automotive systems incorporating machine learning based components: a position paper." In 2018 *48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)* (pp. 271-274). IEEE. 2018
- [13] S. Mohseni, M. Pitale, V. Singh, & Z. Wang, "Practical solutions for machine learning safety in autonomous vehicles." *arXiv preprint arXiv:1912.09630*. 2019.
- [14] S. Haeri, B. Pfeufer, & K. Zhu, "Functionally Safe Artificial Intelligence Components for Autonomous Driving." *ATZelectronics worldwide*, 15(7), 54-57. 2020.
- [15] A. König, A. Dengel, K. Hinkelmann, K. Kise, R. J. Howlett, & L. C. Jain, (Eds.). "Knowledge-Based and Intelligent Information and Engineering Systems, Part II." *15th International Conference, KES 2011*, Kaiserslautern, Germany, September 12-14, 2011, Proceedings, Part II (Vol. 6882). Springer. 2011.
- [16] R. Salay, R. Queiroz, & K. Czarnecki, "An analysis of ISO 26262: Machine learning and safety in automotive software." *SAE Technical Papers*, 2018
- [17] M. Borg, C. Englund, K. Wnuk, B. Duran, C. Levandowski, S. Gao, ... & J. Törnqvist, "Safely entering the deep: A review of verification and validation for machine learning and a challenge elicitation in the automotive industry." *arXiv preprint arXiv:1812.05389*. 2018.
- [18] R.R. Bunel, I. Turkaslan, P. Torr, P. Kohli, P.K. Mudigonda: "A unified view of piecewise linear neural network verification." In: *Advances in Neural Information Processing Systems* 31, pp. 4790–4799, 2018
- [19] S. Dutta, S. Jha, S. Sankaranarayanan, A. Tiwari: "Output range analysis for deep feedforward neural networks." In: *Proc. 10th Int. Symp. NASA Formal Methods*, vol. 10811, pp. 121–138, 2018.
- [20] S. Wang, K. Pei, J. Whitehouse, J. Yang, S. Jana: "Formal security analysis of neural networks using symbolic intervals." In: *Proc. 27th USENIX Security Symp.*, pp. 1599–1614, 2018.
- [21] Y. Sun, M. Wu, W. Ruan, X. Huang, M. Kwiatkowska, D. Kroening: "Concolic testing for deep neural networks." In: *Proc. 33rd ACM/IEEE Int. Conf. Automated Software Engineering*, pp. 109–119, 2018.
- [22] T. Dreossi, S. Ghosh, X. Yue, K. Keutzer, A. Sangiovanni-Vincentelli, & S. A. Seshia, "Counterexample-guided data augmentation." *arXiv preprint arXiv:1805.06962*, 2018
- [23] J. Guo, Y. Jiang, Y. Zhao, Q. Chen, J. Sun: "DLFuzz: Differential fuzzing testing of deep learning systems." In: *Proc. ACM Joint Meeting on European Software Engineering Conf. and Symp. Foundations of Software Engineering*, pp. 739–743, 2018
- [24] M.T. Ribeiro, S. Singh, C. Guestrin: "Why should I trust you?": Explaining the predictions of any classifier. In: *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016
- [25] C. Olah, A. Mordvintsev, L. Schubert: Feature visualization. *Distill* 2(11), e7, 2017.
- [26] J. Kim, A. Rohrbach, T. Darrell, J.F. Canny, Z. Akata: "Textual explanations for self-driving vehicles." In: *Proc. 15th European Conf. Comput. Vision, Part II*, vol. 11206, pp. 577–593, 2018.
- [27] J. Rabold, M. Siebers, U. Schmid: "Explaining black-box classifiers with ILP – empowering LIME with Aleph to approximate non-linear decisions with relational rules." In: *Proc. Int. Conf. Inductive Logic Programming*, pp. 105–117, 2018.
- [28] F. B. Fuchs, O. Groth, A. R. Kosiosek, A. Bewley, M. Wulfmeier, A. Vedaldi & H. I. Posner, "Neural Stethoscopes: Unifying analytic, auxiliary and adversarial network probing." *arXiv*. 2018.
- [29] R. Fong, A. Vedaldi: "Net2Vec: Quantifying and explaining how concepts are encoded by filters in deep neural networks." In: *Proc. 2018 IEEE Conf. Comput. Vision and Pattern Recognition*, pp. 8730–8738, 2018.
- [30] B. Kim, M. Wattenberg, J. Gilmer, C. Cai, J. Wexler, F. Viegas, R. Sayres, "Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (TCAV)." In: *Proc. 35th Int. Conf. Machine Learning*, vol. 80, pp. 2668–2677, 2018.
- [31] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mane, Concrete problems in ai safety. *arXiv preprint arXiv:1606.06565*. 2016.
- [32] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, & B. Kim, Sanity checks for saliency maps. *Advances in neural information processing systems*, 31. 2018.
- [33] C. Guo, G. Pleiss, Y. Sun, & K. Q. Weinberger, On calibration of modern neural networks. In *International conference on machine learning* (pp. 1321-1330). PMLR. 2017.
- [34] I. Golan, & R. El-Yaniv, Deep anomaly detection using geometric transformations. *Advances in neural information processing systems*, 31. 2018.
- [35] Y. Ganin, & V. Lempitsky, Unsupervised domain adaptation by backpropagation. In *International conference on machine learning* (pp. 1180-1189). PMLR., 2015.

- [36] G. Huang, D. Chen, T. Li, F. Wu, L. Van Der Maaten, & K. Q. Weinberger, Multi-scale dense networks for resource efficient image classification. *arXiv preprint arXiv:1703.09844*, 2017.
- [37] J. Wishart, S. Como, U. Forgiione, & J. Weast, Literature review of verification and validation activities of automated driving systems. *SAE Int. J. Connect. Automat. Veh*, 3(4), 267-323. 2020.
- [38] Z. Tahir, & R. Alexander, Intersection Focused Situation Coverage-Based Verification and Validation Framework for Autonomous Vehicles Implemented in CARLA. In *International Conference on Modelling and Simulation for Autonomous Systems* (pp. 191-212). Springer, Cham. 2022
- [39] T. Menzel, G. Bagschik, & M. Maurer, (2018, June). Scenarios for development, test and validation of automated vehicles. In 2018 IEEE Intelligent Vehicles Symposium (IV) (pp. 1821-1827). IEEE.
- [40] D. Sadigh, S. Sastry, S. A. Seshia, & A. D. Dragan, (2016, June). Planning for autonomous cars that leverage effects on human actions. In *Robotics: Science and systems* (Vol. 2, pp. 1-9).
- [41] S. Feng, Y. Feng, C. Yu, Y. Zhang, & H. X. Liu, (2019). Testing Scenario Library Generation for Connected and Automated Vehicles, Part I: Methodology. <https://doi.org/10.1109/TITS.2020.2972211>
- [42] S. Feng, Y. Feng, H. Sun, S. Bao, Y. Zhang, & H. X. Liu, (2019). Testing Scenario Library Generation for Connected and Automated Vehicles, Part II: Case Studies. <https://doi.org/10.1109/TITS.2020.2988309>
- [43] J. Bach, S. Otten, & E. Sax, (2016, June). Model based scenario specification for development and test of automated driving functions. In 2016 IEEE Intelligent Vehicles Symposium (IV) (pp. 1149-1155). IEEE.