

bradscholars

A Framework to Handle Uncertainties of Machine Learning Models in Compliance with ISO 26262

Item Type	Book chapter
Authors	Vasudevan, Vinod;Abdullatif, Amr R.A.;Kabir, Sohag;Campean, Felician
Citation	Vasudevan V, Abdullatif ARA, Kabir S et al (2022) A Framework to Handle Uncertainties of Machine Learning Models in Compliance with ISO 26262. In: Jansen T, Jensen R, Mac Parthaláin N et al (Eds) Advances in Computational Intelligence Systems. UKCI 2021. Advances in Intelligent Systems and Computing. Vol 1409: 508-518. Springer, Cham.
DOI	https://doi.org/10.1007/978-3-030-87094-2_45
Rights	(c) 2022 Springer Cham. Full-text reproduced in accordance with the publisher's self-archiving policy.
Download date	2025-05-15 05:02:32
Link to Item	http://hdl.handle.net/10454/18707

A framework to handle uncertainties of Machine Learning Models in compliance with ISO 26262

Vinod Vasudevan, Amr Abdullatif, Sohag Kabir, and Felician Campean

University of Bradford, Bradford, BD7 1DP, United Kingdom
{v.vasudevan, a.r.a.abdullatif, s.kabir2, f.campean}@bradford.ac.uk

Abstract. Assuring safety and thereby certifying is a key challenge of many kinds of Machine Learning (ML) Models. ML is one of the most widely used technological solutions to automate complex tasks such as autonomous driving, traffic sign recognition, lane keep assist etc. The application of ML is making a significant contributions in the automotive industry, it introduces concerns related to the safety and security of these systems. ML models should be robust and reliable throughout and prove their trustworthiness in all use cases associated with vehicle operation. Proving confidence in the safety and security of ML-based systems and there by giving assurance to regulators, the certification authorities, and other stakeholders is an important task. This paper proposes a framework to handle uncertainties of ML model to improve the safety level and thereby certify the ML Models in the automotive industry.

Keywords: Artificial Intelligence, Uncertainty, Robustness, Certification, Machine Learning, Evidential Deep Learning

1 Introduction

Every year there are around 1.35 million lives lost due to traffic crashes around the world mentioned in waymo safety report [1]. There for safe and reliable autonomous vehicles can help to save lives by reducing the accidents involved by human driving. Many researchers have used Artificial Intelligence (AI) techniques in these safety-critical systems [2]. AI in these systems helps to solve complex problems and to improve the performance of the systems. However, this comes with many challenges as well as opportunities. Automotive Electronic Control Units (ECU) are increasingly given decision-making power to take actions with minimal human intervention. ML-based approaches, systems continuously learn from their operation and dynamically reconfigure in response to changes such as unexpected failures of components/subsystems. In practice, ML technology raises various challenges that could prevent them from being used in a system that requires formal certification.

The International Organisation of Standardization (ISO) introduced ISO 26262 to regulate the functional safety of automobiles E/E Components [3]. It provides requirements and recommendations for the entire life cycle of vehicle

manufacturing. ISO 26262 defines the safety standard for Automotive Electrical/Electronics components which defines vehicle safety. Hazard Analysis and Risk Assessment(HARA) is required to determine hazard levels. These safety requirements are then used to guide the system(software & hardware) development process. ISO 26262 part-6 defines the V-model (see Fig. 1) for the software development process. The objective of this model is to make sure the software safety requirements are covered in design and verified completely.

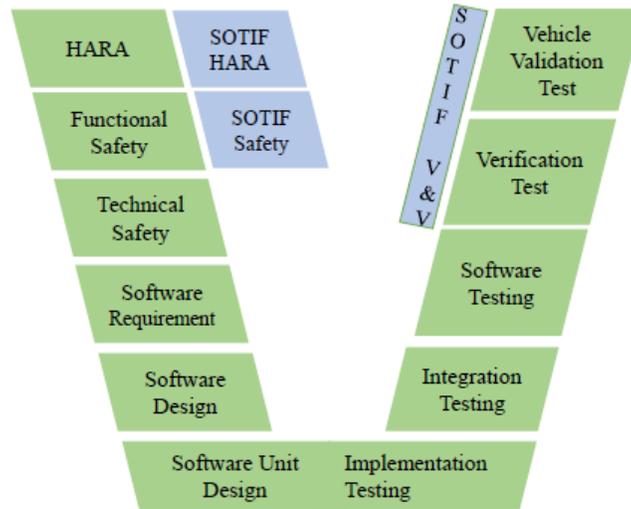


Fig. 1. Comparison between V-Model in ISO 26262 and ISO/PAS 21448

The main aim of ISO 26262 is to help the automotive industry to address functional safety issues in a more systematic approach. ISO did not consider ML as part of the ISO 26262 standard until ML become an essential part of the automotive industry. Therefore, conventional safety assurance methods suggested by the ISO 26262 standard are insufficient or inapplicable for the assurance of ML [4].

There are several interesting papers published in this area of safety assurance of ML systems. The papers in the first such group attempt to analyze the possibility of adaptation and extension of existing functional safety standards such as ISO 26262. In [5], Salay *et al.* presented an analysis of ISO-26262 part-6 methods with respect to ML models safety. The authors found that 40% of software safety methods do not apply to ML models. In these works, the authors discussed five topics that should be addressed in development of ML-based components such as the following: new types of hazards specific to ML, new types of faults and failure modes, usage of incomplete training datasets, the level on which ML algorithms should be used, and which software techniques should be required in these algorithms verification. However, their analysis does not cover

the effectiveness level of these software techniques in covering newly identified failure modes and hazards. As a result, automotive manufacturers and suppliers are faced with different challenges when incorporating ML/Deep Neural Network (DNN) in autonomous cars. The contribution of the current paper is complementary to above research specifically we consider the impact in the areas of hazard analysis during runtime.

ISO/PAS 21448 or Safety of the Intended Function (SOTIF) describes an iterative process that includes design, development, V&V (verification & validation) phases. SOTIF standard recognizes the performance limitation of the software and expects that the scenarios that belong to unsafe-unknown and unsafe-known situations should be reduced so that residual risk is acceptable [6]. This approach allows but manages the performance limitations inherent to ML. It explicitly allows unknown situations for which no learning data will be provided during the learning phase but provides a methodological framework to handle them in a safe manner [6].

Since 1990, deep learning has improved the state-of-art in many ML tasks such as image classification, object detection, speech recognition, vehicle control [7]. One of the most popular variations of deep learning architecture is Convolutional Neural Networks (CNN) which is widely used in computer vision applications such as object detection, image segmentation, recognition, motion tracking, etc. DNN algorithms are currently being applied in safety and security-critical applications such as self-driving cars [2], face detection, robotics, etc. However, despite the power of deep learning-based models in precision classification, we still face problem of making them more cautious by allowing them to assign highly uncertain samples to set of classes. Hence its challenging to assure safety when it is predicated on correct output of these algorithms. Traditional safety measures against systematic software failures, like code review or white box testing are not effective or applicable to ML models [8].

The Dempster-Shafer(DS) theory of belief functions, also referred as evidence theory [9], can be harnessed to provide a solutions to this problem. DS theory is a well-established formalism for reasoning and making decisions with uncertainty . It is based on representing independent pieces of evidence by completely monotone capacities and combining them using a generic operator called Dempster's rule . In Sensoy *et al.* [10] presented evidential theory used for uncertainty estimation and high quality uncertainty modeling is critically important in predicting uncertainty. A model should not only take care of about the accuracy, but also about how certain the prediction is. This is important factor in decision making of safety critical algorithm.

In this paper we try to propose a method to improve the safety/certification process of ML Models.

- The use of Evidential Deep Learning as an alternate mechanism to provide the means of uncertainty of ML models and accounting for uncertainty in the certification process.
- Our framework also proposes recommended actions to applied to minimize the severity of hazard during operation. The modified hazard analysis pro-

poses actions to be applied to minimise the severity of risk during the operation.

2 Safety Critical Systems:Background and Literature

ML do really well in solving the problems that are difficult to specify in the traditional way. However its challenging to assure safety and there by certify when it is predicated on correct outputs of the model. So traditional safety measures against failures like code review (white box testing) are applicable to ML models. Despite highly sophisticated in learning and decision-making, ML systems are prone to many attacks. These lead to harmful consequences in the field of safety-critical systems. ML models are fragile to the domain shift [11], data corruption, and natural perturbations [12]. In the current scenario, there is a necessity for building a safe and secure system. A model is said to be robust and reliable if it does not changes its output or behaviour due to environmental conditions and deployment [13].

Automotive Safety Integrity level(ASIL)is a risk classification defined by ISO 26262 functional safety of road vehicles. The determination of ASIL is the result of hazard analysis and risk assessment. Each hazard is assessed in terms of severity of possible injuries within the context how much of the time a vehicle is exposed to the possibility of the hazard happening as well as the relative likelihood that a typical driver can act to prevent the injury. ASIL refers both to risk and to risk-dependent requirements (standard minimal risk treatment for a given risk). Whereas risk may be generally expressed as:

$$Risk = Severity * (Exposure * Likelihood) \quad (1)$$

The systems or subsystems with in a vehicles can be classified as ‘A’, ‘B’, ‘C’ or ‘D’ gives view on how critical a subsystem is. ASIL D represents the highest degree of safety in automotive (see Fig. 2). All safety critical application should meet the ASIL D requirements. In ASIL D the probability of catastrophic event of the automotive shall be lower than 10^{-8} per driving hour.

$$ASIL = Severity * (Exposure * Controllability) \quad (2)$$

- Severity is the type of injuries to drivers/passengers
- Exposure is how often vehicle is exposed to hazard
- Controllability is how much driver can do to prevent the injury

Each of these parameters are broken down into different levels.

- Severity has four levels ranging from “no injuries” (S0) to “life-threatening/fatal injuries” (S3).
- Exposure has five level covering the “incredibly unlikely” (E0) to the “highly probable” (E4).
- Controllability has four levels ranging from “controllable in general” (C0) to “uncontrollable” (C3).

Severity Class	Probability Class		Controllability Class		
		Class	C1	C2	C3
S1		E1	QM	QM	QM
		E2	QM	QM	QM
		E3	QM	QM	A
		E4	QM	A	B
S2		E1	QM	QM	QM
		E2	QM	QM	A
		E3	QM	A	B
		E4	A	B	C
S3		E1	QM	QM	A
		E2	QM	A	B
		E3	A	B	C
		E4	B	C	D

Fig. 2. Automotive Safety Integrity Level

All variables and sub-classes are analyzed and combined to determine the required ASIL. For example, a combination of the highest hazards (S3 + E4+ C3) would result in an ASIL D classification.

If the decision is taken in the absence of human driver means the controllability will always be C3.

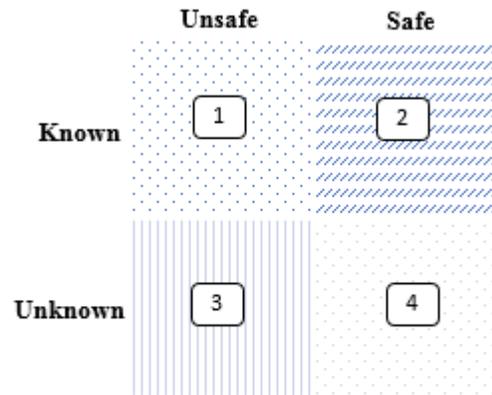


Fig. 3. Known/Unknown and Safe/Unsafe scenarios

SOTIF plays a important role in ML modes as its based on the Intended features. In SOTIF any hazardous event uses cases are classified in to four cate-

gories(see Fig. 3). The aim of SOTIF activities to evaluate Known/Unknown of unsafe area (area 1 and area 2) and there by maximising the safe areas.

3 Framework

The aim of the framework is to handle the uncertainty- related failures in these safety critical application and reduce the severity due to ML failures is acceptably low. Several classification have been proposed for the sources of uncertainty. Aleatoric means uncertainty caused by the noise in the input and epistemic refers to the uncertainty that is systematic that is not sufficiently addressed by a given model [14]. The motivation behind this separation is that aleatoric part of uncertainty has to be accepted, but the epistemic part should be reduced as much as possible by collecting more data Fig. 4. There are several approaches for estimating epistemic uncertainty, such as Bayesian Neural Network, Evidential Deep learning etc. However Bayesian Neural Network face several limitations. Evidential Neural Network are very fast and memory efficient and do not require any sampling to estimate their uncertainty [15].

Aleatoric Uncertainty	Epistemic Uncertainty
Data Uncertainty	Model Uncertainty
Describes confidence in the input data	Describes the confidence of the prediction
High when input data is noisy	High when missing training data
Cannot be reduced by adding data	Can be reduced by adding more data

Fig. 4. Aleatoric vs Epistemic Uncertainty

The certification process involves two stage process. In the first stage, a manufacturer needs to demonstrate to relevant authority that the designed end product behaves as per high level requirements. The first level of certifying a product is to certify the safety requirements(compliance with safety requirements) and the second level is compliance with the legal requirements. Here we look in to the first scenario where we certify a model which gives confidence to the authority. The proposed model is transferring the control to the human(or fail safe) when the ML safety monitoring model is unable to take decision based on the uncertainty and the confidence level fig.5.

The proposed ML safety model in comparison with the modified HARA of ISO 26262 in Fig.6. During the operation,safety critical systems react to different

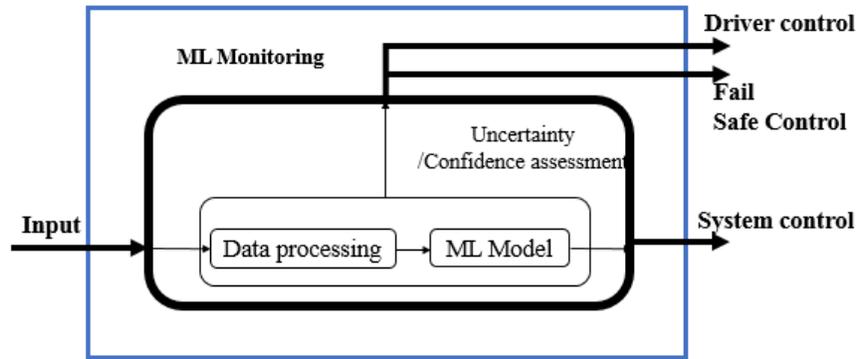


Fig. 5. ML safety monitoring model

known and unknown event and monitoring system calculated the uncertainty/-confidence level. Depends on the impact of the confidence level and uncertainty the safety model switch between driver control, system control and fail safe control. In doing ML monitoring system determine variable in safety constraint more precisely and there by avoid worst case assumptions that affect performance.

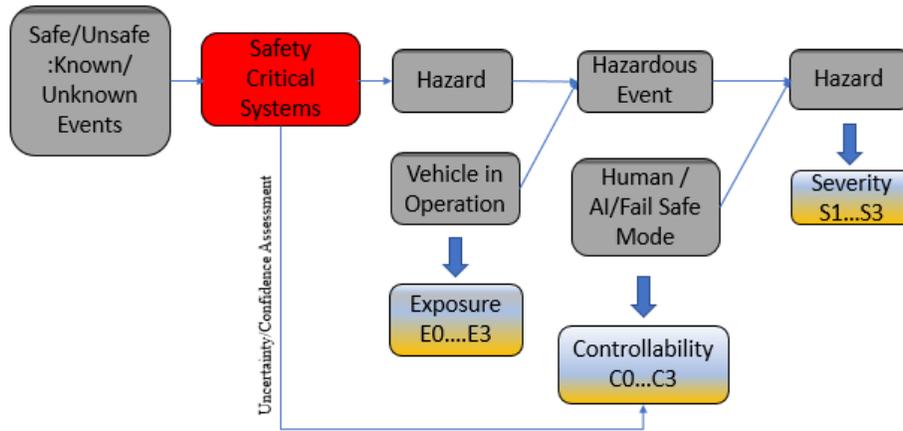


Fig. 6. modified hazard analysis of ISO 26262

4 Certification:Illustrative Example

In 2018, 85% of the road accidents in United Kingdom are due to the human error. So automated vehicle technologies can reduce the human error and there by reduce the collisions. These technology also has the potential to improve driving experience. The automated technologies introduces new risk and challenges, which places significant responsibility on the driver and which may require changes to ensure the safety for this new technology. In this section, we use an Automated Lane Keep Assist System (ALKS) to demonstrate the proposed approach. For instance level 3 of of ALKS (restrictive) was already introduced in June 2020 at UNECE [16].

An ALKS is designed to control the lateral and longitudinal movement of vehicle for certain period with out further driver command. The system is in primary control of the vehicle and perform the driving task instead of driver. The condition of use are still restrictive with the below regulations.

- ALKS available on the roads where pedestrian and cyclists are prohibited
- Operational speed of ALKS systems are limited to a maximum of 60km/hr

It is therefore not designed for situations of heavy, slow moving traffic on a motor way. This new regulations makes a concrete need for dependable and robust ML and certification process. ALKS feature uses video camera to detect the lane marking ahead of the vehicle. If the vehicle is too close to the side of it lane then system will take action by applying corresponding torque to the steering control module.

In the Automotive domain SOTIF address the problem in another way. SOTIF classifies the scenarios according to their impact of safety, ie the proportion of operation scenario leading to a safe situation and minimise the unsafe area.

In this section proposes a framework to find out the uncertainty estimates and confidence level to improve the safety level and there by certify a ML model Fig.7.

A typical representation of ML model development based on the current software development process is shown as part of offline training. An iterative training and testing approach is required as part of ML software development , which is required as part of certification process.

Mohseni *et al.* [6]in reviewed and categorized several techniques that can be used to enhance the dependability and safety of ML algorithms. They analyzed different error detection mechanisms such as uncertainty estimation methods, in/out distribution error detector etc. Willers *et al.* [17]in defined safety concerns are related to the following issues such as failure of data distributions to adequately approximate real-world distributions, distributional shifts in data over time, incomprehensible behavior, unknown behavior in rare & critical situations, unreliable confidence information, brittleness of deep neural networks (DNNs), inadequate separation of test and training data, dependence on labeling quality. There are several other research works in these area. In this paper we follow the V model and ISO 26262 standard as part of offline procedure. In the run time evidential deep learning for uncertainty measurement and follow

5 Conclusion

In this paper, we present a framework for addressing issue of safety/certification in ML models. Autonomous driving and other applications such as image recognition task requires ML models. One of the key challenge in this context is handling the uncertainty of the prediction of ML models. In order to ensure the safety certified of these ML models, we cannot completely rely on the output of these models. If the ML safety model is able to determine the uncertainty of ML using evidential deep learning further enables and improve the safety of the system.

The proposed approach for solving this challenge by determining the uncertainty/confidence assessment during the operation and which used to improve the safety of the system. This approach is always better than working with worst case assumptions. Minimising the uncertainty and improving the confidence in the design time will help to reduce the Known/Unsafe & Unknown/safe areas which is part of SOTIF activity. This is application specific and always better than working with worst case assumptions. We believe that our work will contribute to future progress in applying safety certification of ML with in the automotive industry.

References

1. Waymo LLC, “Waymo safety report,” pp. 1–48, 2020.
2. M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang *et al.*, “End to end learning for self-driving cars,” *arXiv preprint arXiv:1604.07316*, 2016.
3. I. O. for Standardization, “International organization for standardization, 2011, iso 26262 road vehicles functional safety.”
4. Q. Rao and J. Frtunikj, “Deep learning for self-driving cars: chances and challenges,” in *Proceedings of the 1st International Workshop on Software Engineering for AI in Autonomous Systems*, 2018, pp. 35–38.
5. R. Salay, R. Queiroz, and K. Czarnecki, “An analysis of ISO 26262: Using machine learning safely in automotive software,” *arXiv preprint arXiv:1709.02435*, pp. 1–6, 2017.
6. S. Mohseni, M. Pitale, V. Singh, and Z. Wang, “Practical solutions for machine learning safety in autonomous vehicles,” *arXiv preprint arXiv:1912.09630*, 2019.
7. Y. LeCun, Y. Bengio *et al.*, “Convolutional networks for images, speech, and time series,” *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.
8. R. Salay and K. Czarnecki, “Using machine learning safely in automotive software: An assessment and adaption of software process requirements in iso 26262,” *arXiv preprint arXiv:1808.01614*, 2018.
9. A. P. Dempster, “Upper and lower probabilities induced by a multivalued mapping,” in *Classic works of the Dempster-Shafer theory of belief functions*. Springer, 2008, pp. 57–72.
10. M. Sensoy, L. Kaplan, and M. Kandemir, “Evidential deep learning to quantify classification uncertainty,” 2018.
11. S. Mohseni, N. Zarei, and E. D. Ragan, “A multidisciplinary survey and framework for design and evaluation of explainable ai systems,” *arXiv preprint arXiv:1811.11839*, 2018.
12. D. Hendrycks and T. Dietterich, “Benchmarking neural network robustness to common corruptions and perturbations,” 2019.
13. M. Shafique, M. Naseer, T. Theocharides, C. Kyrkou, O. Mutlu, L. Orosa, and J. Choi, “Robust machine learning systems: Challenges, current trends, perspectives, and the road ahead,” *IEEE Design & Test*, vol. 37, no. 2, pp. 30–57, 2020.
14. A. Der Kiureghian and O. Ditlevsen, “Aleatory or epistemic? does it matter?” *Structural safety*, vol. 31, no. 2, pp. 105–112, 2009.
15. A. Amini, W. Schwarting, A. Soleimany, and D. Rus, “Deep evidential regression,” 2020.
16. D. f. Transport, “Safe use of automated lane keeping system on gb motorways: call for evidence,” Apr 2021. [Online]. Available: <https://www.gov.uk/government/consultations/safe-use-of-automated-lane-keeping-system-on-gb-motorways-call-for-evidence>
17. O. Willers, S. Sudholt, S. Raafatnia, and S. Abrecht, “Safety concerns and mitigation approaches regarding the use of deep learning in safety-critical perception tasks,” in *International Conference on Computer Safety, Reliability, and Security*. Springer, 2020, pp. 336–350.