

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/364416900>

Standardisation Considerations for Autonomous Train Control

Chapter · October 2022

DOI: 10.1007/978-3-031-19762-8_22

CITATIONS

2

READS

64

3 authors:



Jan Peleska

Universität Bremen

147 PUBLICATIONS 1,999 CITATIONS

SEE PROFILE



Anne Haxthausen

Technical University of Denmark

97 PUBLICATIONS 1,396 CITATIONS

SEE PROFILE



Thierry Lecomte

ClearSy System Engineering

54 PUBLICATIONS 418 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:






LCHIP: Low Cost, High Integrity Platform [View project](#)



AMASS - Architecture-driven, Multi-concern and Seamless Assurance and Certification of Cyber-Physical Systems [View project](#)



Standardisation Considerations for Autonomous Train Control

Jan Peleska¹(✉) , Anne E. Haxthausen² , and Thierry Lecomte³ 

¹ Department of Mathematics and Computer Science, University of Bremen,
Bremen, Germany

peleska@uni-bremen.de

² DTU Compute, Technical University of Denmark, Kongens Lyngby, Denmark
aeha@dtu.dk

³ CLEARSY, Aix en Provence, France

thierry.lecomte@clearsy.com

Abstract. In this paper, we review software-based technologies already known to be, or expected to become essential for autonomous train control systems with grade of automation GoA 4 (unattended train operation) in existing open railway environments. It is discussed which types of technology can be developed and certified already today on the basis of existing railway standards. Other essential technologies, however, require modifications or extensions of existing standards, in order to provide a certification basis for introducing these technologies into non-experimental “real-world” rail operation. Regarding these, we check the novel pre-standard ANSI/UL 4600 with respect to suitability as a certification basis for safety-critical autonomous train control functions based on methods from artificial intelligence. As a thought experiment, we propose a novel autonomous train controller design and perform an evaluation according to ANSI/UL 4600. This results in the insight that autonomous freight trains and metro trains using this design could be evaluated and certified on the basis of ANSI/UL 4600.

Keywords: Autonomous train control · Standards · Certification · Verification · Validation

1 Introduction

Motivation. Recently, the investigation of autonomous trains has received increasing attention, following the achievements of research and development for autonomous vehicles in the automotive domain. The business cases for autonomous train control are very attractive, in particular for autonomous rolling stock and metro trains [23].

However, several essential characteristics of autonomous transportation systems are not addressed in the standards serving today as the certification basis

J. Peleska—Partially funded by the German Ministry of Economics, Grant Agreement 20X1908E.

© The Author(s) 2022

T. Margaria and B. Steffen (Eds.): ISoLA 2022, LNCS 13704, pp. 286–307, 2022.

https://doi.org/10.1007/978-3-031-19762-8_22

for train control systems. (1) For modules using machine learning, the *safety of the intended functionality* no longer just depends on correctness of a specification and its software implementation, but also on the completeness and unbiasedness of the training data used [12] (Flammini et al. [8] call this “*the opaque nature of underlying techniques and algorithms*”). (2) Agent behaviour based on belief databases and plans cannot be fully specified at type certification time, since the behaviour can change in a significant way later on, due to machine learning effects, updates of the belief database, and changes of plans during runtime [3]. (3) Laws, rules applying to the transportation domain, as well as ethical rules, that were delegated to the responsible humans (e.g. train engine drivers) in conventional transportation systems, are now under the responsibility of the autonomous system controllers. Therefore, the correct implementation of the applicable rule bases needs to be validated [7].

In this light, we analyse the pre-standard ANSI/UL 4600 [24] that addresses the safety assurance of autonomous systems at the system level. Together with several sub-ordinate layers of complementary standards, it has been approved by the US-American Department of Transportation for application to autonomous road vehicles.¹ While examples and checklists contained in this document focus on the automotive domain, the authors of the standard state that it should be applicable to *any* autonomous system, potentially with a preceding system-specific revision of the checklists therein [24, Sect. 1.2.1]. To the best of our knowledge, the ANSI/UL 4600 pre-standard is the first “fairly complete” document addressing system-level safety of autonomous vehicles, and its applicability to the railway domain has not yet been investigated.

Observe that driverless metro trains, people movers and similar rail transportation systems with *Grade of Automation GoA 4 (Unattended train operation, neither the driver nor the staff are required)* [8] have been operable for years², but in *segregated environments* [8]. In these environments, the track sections are protected from unauthorised access, and ubiquitous comprehensive automation technology is available, such as line transmission or radio communication for signalling, precise positioning information, as well as platform screen doors supporting safe boarding and deboarding of passengers between trains and platforms.

In contrast to this, we investigate the certifiability of autonomous train control systems with GoA 4 in *open railway environments*, where unauthorised access to track sections, absence of platform screen doors, and less advanced technology (e.g. visual signalling) have to be taken into account. This scenario is of high economic interest, and first prototype solutions have recently become available [19], but none of them has yet achieved GoA 4 with full type certification.

¹ <https://www.youtube.com/watch?app=desktop&v=xCIjxiVO48Q&feature=youtu.be>.

² The driverless Paris metro METEOR, for example, is operative since 1998 [2]. A list of automated train systems is available under https://en.wikipedia.org/wiki/List_of_automated_train_systems.

Flammini et al. [8] emphasise the distinction between automatic and autonomous systems. The latter should be “...*capable of taking autonomous decisions, learning from experience, and adapting to changes in the environment*”. The train protection systems considered in this paper exhibit a “*moderate*” degree of autonomy, as described below in Sect. 4: they react, for example, to the occurrence of obstacles and degradation of position information by slowing down the train’s speed and decide to go back to normal velocity as soon as obstacles have been removed or precise positioning information is available. These reactions, however, are based on pre-defined deterministic behavioural models and do not depend on AI functionality or on-the-fly learning effects. Some data providers for the train protection system, as, for example, the obstacle detection module, use AI-based technology, such as image classification based on neural networks. We think that this moderation with respect to truly autonomous behaviour is essential for enabling certifiability for train operation in the current European railway infrastructure.

Main Contributions. We propose a novel design for an autonomous train control system architecture covering the functions *automatic train protection (ATP)* and *automatic train operation (ATO)*. This architecture is suitable for GoA 4 in an open environment. This operational environment is assumed to be heterogeneous, with diverse track-side equipment, as can be expected in Europe today. Furthermore, we assume the availability of controlled allocation and assignment of movement authorities, as is performed by today’s interlocking systems (IXL, potentially supported by radio block centres (RBC)). Apart from the communication between train and RBC/IXL, no further “vehicle-to-infrastructure” communication channels are assumed. Moreover, the design does not require “vehicle-to-vehicle” communication, since this is not considered as standard in European railways today. As a further design restriction, we advocate the strict separation between conventional control subsystems, and novel, AI-based subsystems that are needed to enable autonomy. It turns out that the latter are only needed in the perception part of the so-called *autonomy pipeline*

$$\textit{sensing} \rightarrow \textit{perception} \rightarrow \textit{planning} \rightarrow \textit{prediction} \rightarrow \textit{control} \rightarrow \textit{actuation},$$

which is considered as the standard paradigm for building autonomous systems today [13]. Fail-safe perception results are achieved by means of a sensor→perceptor design with redundant, stochastically independent channels.

This deliberately conservative architecture serves as the setting for a thought experiment analysing whether such a GoA 4 system could (and should) be certified. The conventional subsystems can be certified on the basis of today’s CENELEC standards [4–6]. For the AI-based portion of the design, however, the CENELEC standards cannot be applied. Instead, we use the ANSI/UL 4600 pre-standard [24] and investigate, whether this part can be certified according to this standard with a convincing safety case.

We demonstrate that this architecture for autonomous train control will be certifiable for freight trains and metro trains. In contrast to this, we deem the

trustworthy safety assurance of autonomous high-speed passenger trains with GoA 4 to be infeasible today – regardless of the underlying ATP/ATO design. This assessment is justified by the fact that existing obstacle detection functions can only be executed to operate with sufficient reliability for trains with speed up to 120 km/h.

Related Work and Distinction from Other Approaches. The terminology in this paper is in line with terms and definitions introduced by Flammini et al. [8], where a wide range of existing and potential future technologies are discussed and classified.

It is important to point out that visions of autonomous train control far beyond the “fairly moderate” concepts considered in this paper exist. Trentesaux et al. [23] point out the attractiveness of business cases based on trains autonomously negotiating their way across a railway network in an open, uncontrolled (i.e. not fully secured) environment. To this end, they suggest a train control architecture whose behaviour is based on plans that are continuously adapted to increase safety and efficiency. A typical software implementation paradigm for this type of behaviour would be *belief-desire-intention (BDI) agents* [3]. Unsurprisingly, the authors come to the conclusion that the safety assurance and certification of such systems will be quite difficult. Indeed, we will point out below that exactly this type of train control is the one with the least prospects of becoming certifiable in the future.

Flammini et al. [8] discuss the certifiability issues of a variety of ATP/ATO concepts, including the “more futuristic” ones, in a more systematic manner. For all variants, the authors advocate a strict separation between automated ATP and ATO, because the former is safety critical and requires certification according to the highest safety integrity level SIL-4, while the latter could be certified according to a lower SIL, since ATP will ensure that the train will remain safe, even in presence of ATO malfunctions. This distinction between ATP and ATO has influenced the design decisions presented in Sect. 4.

It is interesting to note that the advantages of vehicle-to-vehicle communication deemed to be promising for future train control variants for various purposes [8, 23] has already been investigated during 1990s, with the objective to abolish centralised interlocking systems [10]. For the architectural train control concept presented here, however, it is crucial that the safety of allocated train routes is performed by “conventional” IXLs/RBCs, so that these tasks are not contained in the trains’ autonomy pipelines.

The paper presented here is inspired by the work of Koopman et al. discussing certification issues of road vehicles [14–16]. It will become clear in the remainder of this paper, however, that their results cannot be “translated in one-to-one fashion” for the railway domain.

The material presented here is complemented by a technical report containing behavioural specification models for some of the ATP/ATO aspects discussed in this paper [11, Appendix A].

Overview. In Sect. 2, the standards and pre-standards of interest in the context of this paper are briefly reviewed. In Sect. 3, we describe existing technology that is needed to realise autonomous train control systems. Up to now, most of these technologies have been used in proof-of-concept projects, so that conformance to standards and certification was not yet an issue. In Sect. 4, we present a new reference architecture for autonomous train control systems that we advocate, due to having fair chances of becoming certifiable in the near future. In Sect. 5, we perform an evaluation of certifiability according to ANSI/UL 4600 for the reference architecture introduced before. Lastly, Sect. 6 contains some concluding remarks.

2 Standardisation and Certification

In the railway domain, safety-critical track-side and on-board systems in Europe must be designed, verified and validated according to the CENELEC standards EN50126, EN50128, and EN50129, in order to pass type certification. None of these documents provides guidance for V&V of AI-based sub-functions involving machine learning, classification techniques, or agent-based autonomous planning and plan execution. Since, as outlined in Sect. 3, autonomous train control depends on such AI-based techniques, this automatically prevents the certification of autonomous train control systems on the basis of these standards alone.

To the best of our knowledge, the ANSI/UL 4600 pre-standard for the evaluation of autonomous products [24] is the first document that is sufficiently comprehensive to serve (in modified and extended form) as a certification basis for operational safety aspects of autonomous products in the automotive, railway, and aviation domains. The standard is structured into 17 sections and 4 annexes. Section 5 addresses the elaboration of safety cases and supporting arguments in general, and Sect. 6 covers general risk assessment. For the context of the paper presented here, Sect. 7 and Sect. 8 of the ANSI/UL 4600 standard are the most relevant parts.

The focus of Sect. 7 is on interaction between humans, animals and other systems and the autonomous system under evaluation (denoted as the *item* in the standard). While this section needs extensive cover for autonomous road vehicles in urban environments, its application is more restricted for the railway domain: here, the pre-planned interaction between humans and autonomous trains takes place in train stations on platforms, during boarding and deboarding. The safety of these situations is handled by the passenger transfer supervision subsystem discussed below. On the track, humans are expected on railway construction sites and level crossings, otherwise their occurrence is illegal. For both legal and illegal occurrences, the on-track interaction between humans and the train is handled by the obstacle detection subsystem described in Sect. 4.

Section 8 of the standard explicitly addresses the autonomy functions of a system, as well as auxiliary functions supporting autonomy. It explains how the impact of autonomy-related system functions on safety should be addressed by means of hazard analyses. For the non-negligible risks induced by these functions,

it has to be explained how mitigating functions have been incorporated into the system design. The operational design domain with its different situations and changing environmental conditions needs to be specified, and it has to be shown how the hazards induced by each situation paired with environmental conditions are controlled by the safety mechanisms of the target system. To present hazards caused by autonomy functions, associated design decisions, and mitigations in a well-structured manner, the section is structured according to the autonomy pipeline introduced in Sect. 1.

The other sections of ANSI/UL 4600 cover the underlying software and systems engineering process and life cycle aspects, dependability, data, networking, V&V, testing, tool qualification, safety performance indicators, and assessment of conformance to the standard. These aspects are beyond the scope of this paper.

3 Technology

A number of technologies are required to implement autonomous train control on existing railway networks. The non-modification of existing infrastructure, in particular track-side signalling equipment, is sought in order to facilitate their deployment at lower cost.

We agree with the recommendations of the Federal Railroad Administration of the U.S. Department of Transportation [28] who envision a *sensor platform* combining several different technologies to identify *objects of interest (OOI)* (obstacles, landmarks enabling the improvement of position calculation, train stations, ...) and *conditions of interest (COI)* (“*track is free of obstacles up to location ...*”, “*the train location has distance n meters to its end of movement authority*”, ...). The perception of the immediate train environment is mandatory to ensure a correct navigation regarding signalling equipment, but also to avoid catastrophic collisions with obstacles (trains, objects, animals) by perceiving the scene up to its braking distance. The use of different types of sensing techniques and technologies (radar, laser, LiDAR, camera time-of-flight, camera IR) is necessary to obtain a functional capacity for a wide variety of environmental situations. By using different wavelengths or physical principles (or combination of), it is possible to avoid receiving incorrect information (from radar secondary lobe) or becoming completely blind under certain situations. Indeed, weather conditions (precipitation, snow, humidity, high light levels, mist, dust etc.) have a direct impact on the quality and accuracy of the perceived information, which can strongly alter the representation of the observed scene. For example, an occlusion (spot on an optic) could hide an obstacle; a low sun on the horizon in the axis of the rails could prevent the detection of a light due to sensor saturation.

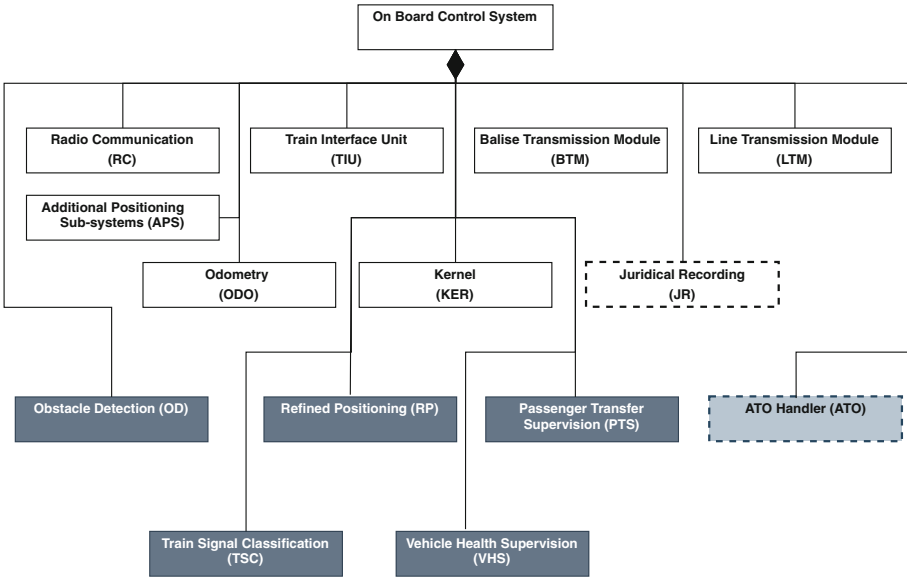


Fig. 1. Reference architecture of autonomous train to be considered for certification.

4 A Reference Architecture for Autonomous Train Controllers

Architecture – Functional Decomposition. In the subsequent paragraphs, we will investigate an autonomous on-board train controller, whose functional decomposition is shown in Fig. 1. The grey boxes are functions required for autonomous trains only. They cannot be certified on the basis of the CENELEC standards alone, because they rely on AI-based functionality.

The white boxes represent components already present in modern conventional on-board units supporting partially automated train control according to GoA 2³, as suggested by the UNISIG recommendations for ETCS [25]. This structuring into conventional modules is re-used for the autonomous train architecture introduced here. Even existing GoA 2 module implementations could be re-used, but the kernel module has to be significantly extended, as described below. All “white-box modules” in Fig. 1 – even the kernel in its extended form – can be certified on the basis of the CENELEC standards, because no AI-based functionality is deployed on these modules.

In the detailed description below, it will turn out that the kernel in Fig. 1 realises the ATP functionality and the other solid-line boxes provide safety-relevant data to the kernel. Therefore, they need to be certified according to the highest safety integrity level SIL-4. The ATO handler, however, could be certified

³ *Semi-automatic train operation.* ATO and ATP systems automatically manage train operations and protection while supervised by the driver [8].

according to lower integrity levels, because the automatic train operation is always supervised and restricted by the ATP functions. The same applies to juridical recording, since this has no impact on the train's dynamic behaviour. With this approach, the strict segregation between ATP and ATO advocated by Flammini et al. [8] has been realised.

Conventionally Certifiable On-Board Modules. The central module is the *kernel* which executes the essential ATP operations in various operational modes described below. All decisions about interventions of the normal train operation are taken in the kernel. Based on the status information provided by the other subsystems, the kernel controls the transitions between operational modes (Fig. 2 below). Interventions are executed by the kernel through access to the *train interface unit*, for activating or releasing the service brakes or emergency brakes. The decisions about interventions are taken by the kernel based on the information provided by peripheral modules: (1) The *odometry module* and *balise transmission module* provide information for extracting trustworthy values for the actual train positions. As known from modern high-speed trains, *additional positioning subsystems* provide satellite positioning information in combination with radar sensor information to improve the precision and the reliability of the estimated train location. (2) The *radio communication module* provides information about movement authority and admissible speed profiles, as sent to the train from interlocking systems via radio block centres. In the train-to-trackside transmission direction, the train communicates its actual position to radio block centre/interlocking system. (3) The *line transmission module* provides signal status information provided by trackside equipment for the train. (4) The *juridical recording module* stores safety-relevant kernel decisions and associated data.

Note that, depending on the availability of track-side equipment, not all the data providers listed above will be available. In the non-autonomous case, the missing information is compensated by the train engine driver who, for example, visually interprets signals if trackside line transmission equipment is unavailable. For the autonomous case, additional support modules as described below are required.

Operational Modes. The operational design domain and its associated hazard analyses regarding operational safety (this is further discussed in Sect. 5) induce different operational modes for the train protection component realised by the kernel, providing suitable hazard mitigations. These modes and the transitions between them are depicted in Fig. 2.

In the *autonomous normal operation (ANO)* mode, the train is fully functional and controlled with full autonomy within the range of its current position and the end of movement authority (MA) obtained from the interlocking system (IXL) via radio block controller (RBC). The ANO-(sub-)controller supervises the observation of movement authorities, ceiling speed and braking to target (e.g. the next train station or a level crossing). Its design and implementation

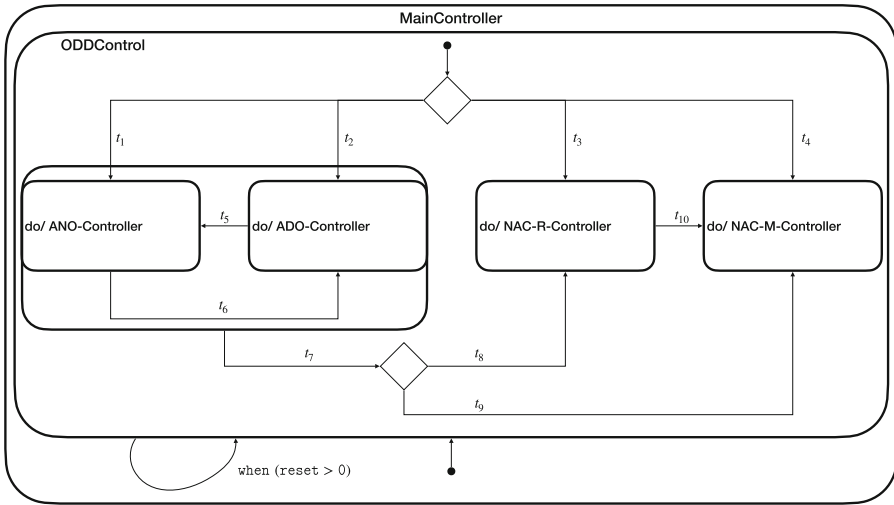


Fig. 2. Operational modes for train protection in autonomous trains.

is “conventional” in the sense that the complete functional behaviour is pre-determined by formal models (e.g. state machines) available at type certification time. Indeed, the design of the ANO-controller can be based on that already used for (non-autonomous) ETCS trains today. The only difference is that the interface to the train engine driver is no longer used. Instead, acceleration and braking commands to be executed within the safety limits supervised by the ANO-controller are provided by the ATO-handler described below.

In *autonomous degraded operation (ADO)* mode, the train is still protected autonomously by the ADO-controller and operated by the ATO module, but with degraded performance (e.g., with lower speed). Mode ADO is entered from ANO, for example, if the available position information is not sufficiently precise, so that the train needs to be slowed down until trustworthy position information is available again (e.g. because the train passed a balise with precise location data). Also, the occurrence of an unexpected obstacle (e.g. animals on the track) leads to a transition to the ADO mode. It is possible to transit back from degraded mode to autonomous normal operation, if the sensor platform signals sufficiently precise location information (e.g. provided by a balise that has been passed) and absence of obstacles. Again, the ADO-controller can be modelled, validated and certified conventionally according to EN 50128 [4]. The difference to non-autonomous operation consists in the fact that the transition from ANO to ADO is triggered by events provided by the sensor and perceptor platform, since no train engine driver is available.

In case of a loss of vital autonomous sub-functions (see description of these functions below), the train enters one of the *non-autonomous control (NAC)* modes. In NAC-R, the train can still be remotely controlled by a human from some centralised facility. The operational safety of remotely controlled trains has

been discussed by Tonk et al. [22]. If no remote control facility is available, the train enters mode NAC-M and has to be manually controlled by a train engine driver boarding the train.

Modules Supporting Autonomous Trains Operation. The *obstacle detection module (OD)* has the task to identify objects on the track, like persons, fallen trees, or cars illegally occupying a level crossing. Note that the absence of other trains on the track is already guaranteed by the IXL, so OD can focus on unexpected objects alone. OD uses a variety of sensors (cameras, LiDAR, radar, infrared etc.) [28] to determine whether obstacles are on the track ahead. In case an obstacle is detected, it would be required to estimate its distance from the train in order to decide (in the kernel) whether an activation of emergency brakes is required or if the service brakes suffice. A further essential functional feature is the distinction between obstacles on the train’s track and obstacles of approaching trains on neighbouring tracks, where no braking intervention is necessary. Camera-based obstacle detection can be performed by conventional computer vision algorithms or by means of image classification techniques based on neural networks and machine learning [18, 29]. None of the available technologies are sufficiently precise and reliable to be used alone for obstacle detection [28]. Instead, a redundant sensor combination based on several technologies is required, as described below. In any case, experimental evidence is only available for train speeds up to 120 km/h [18]; this induces our restriction to autonomous freight trains and metro trains. From the perspective of the autonomy pipeline described in Sect. 1, the obstacle detection module performs sensing and perception. It provides the “*obstacle present in distance d* ” information to the kernel which operates on a state space aggregating all situational awareness data.

The *refined positioning module (RP)* provides additional train location information, with the objective to compensate for the train engine driver’s awareness of the current location that is no longer available in the autonomous case. A typical use case for refined positioning information is the train’s entry into a station, where it has to stop exactly at a halt sign. To achieve the positioning precision required for such situations, signposts and other landmarks with known map positions have to be evaluated. This requires image classification, typically based on trained neural networks [20]. Again, conventional image recognition based on templates for signs and landmarks to expect can be used [17] to allow for fusion of conventional and AI-based sub-sensors. The *train signal classification module (TSC)* is needed on tracks without line transmission facilities. Signals and other signs need to be recognised and classified. Summarising, the OD, RP, and TSC modules represent perception functions helping the kernel to update its situational awareness status. All three modules can be realised by means of sensor combination techniques involving both conventional image recognition methods and trained neural networks. These observations become important in the sample evaluation performed in Sect. 5.

The *passenger transfer supervision module (PTS)* is needed to ensure safe boarding and disembarking of passengers. It applies to the fully autonomous case of passenger trains being operated without any personnel and in absence of screen doors on the platform. This module requires sophisticated image classification techniques, for example, to distinguish between moving adults, children, and other moving objects (e.g. baggage carts on the platform). Again, PTS is a sensing and perception function providing the kernel with the “*passengers still boarding/deboarding at door . . .*” and “*passengers or animals dangerously close to train*” information that shall prevent the train from starting to move and leave the station. Sensor combination with conventional technology could be provided by various sorts of light-sensors, in particular, safety light curtains⁴.

The *vehicle health supervision module (VHS)* is needed to replace the train engine drivers’ and the on-board personnel’s awareness of changes in the vehicle health status. Indications for such a change can be detected by observing acoustic, electrical, and temperature values. The conclusion about the actual health status, however, strongly relies on the experience of the personnel involved. This knowledge needs to be transferred to the health supervision in the autonomous case [23]. Since the effect of human experience on the train’s safety is very hard to assess, it is quite unclear how “sufficient performance” of module VHS should be specified, and how it should be evaluated. Therefore, we do not consider this component anymore in the sequel.

The handler for automated train operation (*ATO handler*) acts within the restrictions enforced by the ATP functionality. The kernel defines the actual operational level (ANO, ADO, NAC-R, NAC-M), and the ATO handler realises automated operation accordingly. In autonomous normal operation mode ANO, the ATO handler could, for example, optimise energy consumption by using AI-based strategies for efficient acceleration and braking [19]. After a trip situation leading to an emergency stop (this is controlled by the kernel, including the transition into autonomous degraded operation ADO), the ATO handler controls re-start of the train and negotiates with the IXL/RBC the location and time from where ANO can be resumed. The train movements involved are again within the limits of the actual movement authority provided by the IXL/RBC, so the essential safety assurance is provided by ATP. In the degraded mode NAC-R, the ATO handler performs the protocol for remotely controlled train operation. If remote control is unavailable, a switch to NAC-R is performed by the kernel, and the ATO handler becomes passive, since train operation is switched to manual.

Dual Channel Plus Voting Design Pattern. As a further design decision, we introduce a two-channel design pattern for the modules OD, TSC, RP, and PTS, as shown in Fig. 3. The objective of this design is to produce a fail-safe sensor→perceptor component, such that it can be assumed with high probability that *either the perception results transmitted to the kernel are correct, or the component will signal ‘failure’ to the kernel*. In the ‘failure’ case, the kernel will transit into one of the degraded modes ADO, NAC-R, NAC-M, depending on

⁴ https://en.wikipedia.org/wiki/Light_curtain.

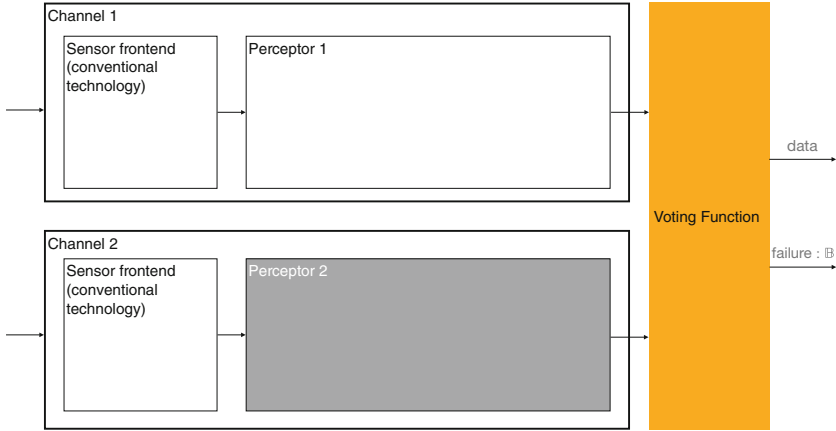


Fig. 3. Two-channel design pattern used for modules OD, TSC, RP, and PTS.

the seriousness of the fault. A *reliable* sensor→perceptor subsystem can then be constructed from three or more fail-safe components using complementary technologies (e.g. one component is based on radar technology, while the other uses cameras), so that a deterministic sensor fusion by means of m -out-of- n voting decisions can be made in the kernel.

Each channel of a fail-safe component has a sensor frontend (camera, radar etc.) for receiving environment information. The sensor frontends use redundant hardware, so that they can be assumed to be stochastically independent with respect to hardware faults. The remaining common cause faults for the sensors (like sand storms blinding all camera lenses) can be detected with high probability, because both sensor data degrade nearly simultaneously.

The sensor frontends pass their raw data to the perceptor submodules: each perceptor processes a sequence of sensor readings to obtain a classification result such as ‘obstacle detected’ or ‘halt signal detected’. We require perceptors 1 and 2 to use ‘orthogonal’ technology, so that their classification results (e.g. ‘obstacle present’) are achieved in stochastically independent ways. For example, a pair of vision-based perceptors could be realised by neural networks with different layering structure and trained with different data sets. Alternatively, one perceptor could be based on trained neural networks, while the other uses conventional image recognition technology [18]. A third option is to combine two orthogonal sensor→perceptor technologies that are a priori independent, such as one channel based on camera vision, and another on radar.

Note that in this context, stochastic independence does not mean that the two perceptors are very likely to produce different classification results, but that they have obtained these results *for different reasons*. For example, one perceptor detects a vehicle standing on the track by recognising its wheels, while the other detects the same obstacle by recognising an aspect of the vehicle body (e.g. the radiator grill). This type of independence will allow us to conclude that

the probability for the perceptrons to produce an unanimous misclassification is the product of the individual misclassification probabilities. We have devised a method to verify the stochastic independence of perceptrons by means of ‘explainable AI’ approaches [20] and statistical tests; this, however, is beyond the scope of this paper (see Sect. 6). Both perceptrons pass their result data and possibly failure information from the sensor frontends to a joint voting function that compares the results of both channels and relays the voting result or a failure flag to the kernel.

Design of Voting Functions. For the OD module, the voting function raises the failure flag if both channels provided contradictory “*no obstacle/obstacle present*” information over a longer time period. For unanimous “*obstacle present*” information with differing distance estimates, the function “falls to the safe side” and relays the shorter distance to the kernel. Similar voters can be designed for RP, TSC, and PTS.

Table 1. Mapping of architectural components to SIL and autonomy pipeline.

	Sensing	Perception	Planning	Prediction	Control	Actuation
SIL-4	OD, TSC, RP, PTS, VHS	RC, ODO, APS, BTM, LTM	KER	KER	KER	TIU
SIL-4+AI		OD, TSC, RP, PTS, VHS				
Lower SIL+AI			ATO	ATO	ATO	

Mapping Modules to the Autonomy Pipeline. The architectural components discussed above can be mapped to the autonomy pipeline as shown in Table 1. The abbreviations used have been defined in Fig. 1. The table also shows the required safety integrity levels. These are derived from the existing CENELEC standards and their requirements regarding functional safety. The marker “+AI” in column 1 indicates that AI-based implementations are required for the respective components. For integrity level SIL-4, which is the main concern of this paper, AI-based methods are strictly confined to the perception part of the pipeline. As discussed above, the ATO module can be certified according to a lower SIL. It could contain both conventional sub-functions and AI-based functions. In the latter case (not discussed in this paper), the evaluation and certification would be performed according to ANSI/UL 4600.

5 A Sample Evaluation According to ANSI/UL 4600

Evaluation Procedure. In this section, Sect. 8 (*Autonomy Functions and Support*) of ANSI/UL 4600 is applied to analyse whether a safety case for the autonomous train control architecture described in Sect. 4 conforming to this standard could be constructed. The procedure required is as follows [24, 8.1]. (Step 1) Identify all hazards related to autonomy and specify suitable mitigations. (Step 2) Specify the autonomy-related implications on the operational design domain. (Step 3) Specify how each part of the autonomy pipeline contributes to the identified hazards and specify the mitigations designed to reduce the risks involved to an acceptable level.

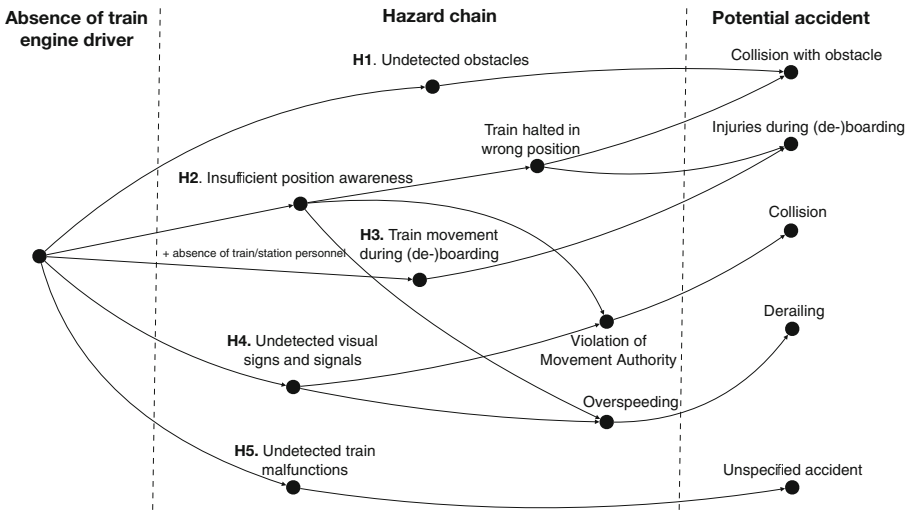


Fig. 4. Hazards caused by absence of train engine driver and personnel.

Step 1. Autonomy Functions, Related Hazards, and Mitigations. The absence of a train engine driver and other train service personnel induces the hazard chains shown in Fig. 4, together with the resulting potential accidents. In this diagram, the hazards from H1 to H5 have been identified as suitable for mitigation and thereby preventing each of the hazard chains from leading to an accident. The hazards marked from H1 to H5 are mitigated by the autonomic function pipelines listed in Table 2 as follows.

Table 2. Hazard mitigations to enable autonomy.

Id.	Hazard	Mitigations by pipeline
H1	Undetected obstacles	OD \rightarrow KER \rightarrow TIU
H2	Insufficient position awareness	{ODO,APS,BTM,RP} \rightarrow KER \rightarrow TIU
H3	Train movement during (de-)boarding	PTS \rightarrow KER \rightarrow TIU
H4	Undetected visual signs and signals	{LTM,TSC} \rightarrow KER \rightarrow TIU
H5	Undetected train malfunctions	VHS \rightarrow KER \rightarrow TIU

H1 (unidentified obstacles) is prevented by the pipeline OD \rightarrow KER \rightarrow TIU covering sensing and perception (OD), planning, prediction, and control (KER), and actuation via train interface unit TUI. The OD indicates detected obstacles to the kernel. The kernel first performs a hard-coded planning task covering three alternatives: (a) if the train is still far from the obstacle, it shall be decelerated by means of the service brakes, in the expectation that the obstacle will disappear in time, and re-acceleration to normal speed can be performed. (b) If the obstacle is not removed in time, the train shall brake to a stop. (c) if the obstacle is too close for the service brakes, the train shall be stopped by means of the emergency brakes. The prediction part of the pipeline is likewise hard-coded. The kernel calculates the stopping positions depending on current position, actual speed and selection of the brake type.⁵ The control part triggers planning variant (a), (b) or (c) according to the prediction results and the obstacle position estimate and acts on the brakes by means of the train interface unit TIU. Since obstacle handling requires a deviation from normal behaviour by braking the train, the planning-prediction-control part is implemented in the ADO-handler for degraded autonomous operation inside the kernel.

The autonomy function pipelines for mitigating hazards from H2 to H5 operate in analogy to the pipeline mitigating H1.

These considerations show that the hazards are adequately mitigated, *provided that the associated mitigation pipelines from Table 2 fulfil their intended functionality* in the sense of ISO 21448 [12]. Therefore, each of the pipelines listed in Table 2 needs to be evaluated according to Sect. 8 (Autonomy Functions and Support) of the ANSI/UL 4600 standard, as described below.

Step 2. Operational Design Domain and Autonomy-Related Implications. The *operational design domain (ODD)* is defined in ANSI/UL 4600 as “*The set of environments and situations the item is to operate within.*” [24, 4.2.30]. Safety cases conforming to this standard need to refer to the applicable ODD subdomains, when presenting safety arguments for autonomous system functions. Originally introduced for autonomous road vehicles [21], systematic

⁵ This calculation is based on well-known braking models [27].

approaches to ODD elaboration in the railway domain exist [22]. For a comprehensive safety case, it has to be shown that system operation within the limits of the ODD and its subdomains is safe, and that transitions leaving the ODD are prevented or at least detected and associated with safe reactions (e.g. transitions to a safe state).

The attributes of an ODD are structured into three categories: (1) scenery, (2) environmental conditions, and (3) dynamic elements. In the context of this paper, one class of scenery attributes describes the railway network characteristics the train might visit or travel through: train stations, maintenance depots, tunnels, level crossings, “ordinary” track sections between stations. Note that it is not necessary to differentiate between network characteristics controlled by the interlocking, such as different kinds of flank protection or the availability of shunts in a given network location: since the safety of IXLs is demonstrated separately, and since our investigation is based on current IXL technology that can be certified by conventional means, these aspects can be abstracted away for the type of autonomous trains discussed here.

Regarding environmental conditions, weather and illumination conditions are critical for the sensors and perceptors enabling automated train protection. Moreover, the availability of supporting infrastructure (e.g. GPS, line transmission, balises) varies with the train’s location in the railway network, and with exceptional conditions (e.g. unavailability of GPS).

Dynamic elements to be considered apart from the train itself are just illegally occurring obstacles, like vehicles or persons on closed level crossings or variants of obstacles on the track. There is no need to consider other trains, since their absence is controlled by the IXL.

Observe that large portions of the ODD can be created from existing knowledge compiled before to satisfy the reliability, availability, maintainability, and safety requirements for non-autonomous trains according to EN 50126 [5]. The new ODD aspects to be considered for the architecture advocated in this paper are related to the novel sensor and perceptor platform needed for OD, RP, PTS, TSC, and VHS.

As discussed next, the ODD induces V&V objectives that need to be fulfilled in order to guarantee that the train will operate safely under *all* scenarios, environmental conditions, and dynamic situations covered by the ODD. Note that for road vehicles, it is usually necessary to consider states outside the ODD (e.g. a car transported into uncharted terrain and started there), where safe fall-back operation has to be verified. For the railway domain as considered here, the ODD is complete, since the IXL ensures that the train will only receive movement authorities to travel over admissible track sections of the European railway network.

Step 3. Evaluation of the Autonomy Pipeline. Each of the hazard mitigation pipelines listed in Table 2 needs to be evaluated according to ANSI/UL 4600, Sect. 8 to show that they really mitigate their associated hazards from H1 to H5

with acceptable performance under all conditions covered by the ODD. The pre-standard suggests to structure the evaluation according to the autonomy pipeline and address the specific operational safety aspects of every pipeline element separately.

Sensor Evaluation. Until today, cameras have been used on trains for obstacle detection and refinement of positioning information only in experiments. Evaluation results already obtained for cameras in autonomous road vehicles cannot easily be re-used, since the train sensor platform requires cameras detecting obstacles and landmarks in greater distances than cars. Also, adequate operation in presence of higher vibrations need to be considered. Experiments have shown, however, that raw image information of cameras can be provided with acceptable performance under the lighting and weather conditions specified in the ODD [18].

ANSI/UL 4600 requires a detailed evaluation of the sensor redundancy management. As described above, we exploit sensor redundancy to detect the (temporary) failure of the two-channel sensor→perceptor subsystem due to adverse weather conditions. Moreover, the sensor redundancy contributes to achieving stochastic independence between the two sensor→perceptor channels. Both redundancy objectives need to be validated separately at design level and in field tests. The ANSI/UL 4600 requirement to identify and mitigate risks associated with sensor performance degradation is fulfilled by the design proposed here in the following ways: (a) total (2-out-of-2) sensor failures are detected, communicated via voting unit to the kernel and lead to a switch into non-autonomous mode which is always accompanied by an emergency stop until manual train operation takes over. (b) 1-out-of-2 sensor failure is tolerated over a limited time period. If recovery cannot be achieved, a transition into non-autonomous mode becomes necessary, since the redundancy is needed to ensure the fail-safe property of the complete sensor→perceptor component. (c) Performance degradation in one sensor leads to discrepancies in the two perceptor channels. If the voter can “fall to the safe side” (e.g. by voting for ‘HALT’ if one TSC channel perceives ‘HALT’ while the other perceives ‘GO’), the autonomous operation can continue. If no such safe results can be extracted from the differing channel data, a transition into non-autonomous mode is required.

Further sensor types (e.g. radar and GPS antennae) already exist on today’s high-speed trains, and the certification credit obtained there can be re-used in the context of autonomous trains.

Perceptor Evaluation. The first evaluation goal consists in the demonstration that the perceptor’s functional performance is acceptable. The main task to achieve this goal is to demonstrate that both the false negative rate and the false positive rate are acceptable. For the sensor→perceptor sub-pipelines mitigating hazards from H1 to H5, false negatives have the following meanings.

Id.	Definition of false negative
H1	indication ‘no obstacle’ though an obstacle is present
H2	indication ‘no position error’ though estimate is wrong
H3	indication ‘no (de-)boarding passengers’ though passengers are still present at doors or close to train
H4	indication ‘no restrictive signal present’ (e.g. HALT, speed restriction) though such a signal can be observed
H5	indication ‘no malfunction’ though malfunction is present

With these definitions, the false negative rates impair safety, while the false positive rates only impair availability. With the stochastic independence between the two perceptor channels and the voter principle to fall to the safe side, the false negative rates can be controlled.

For each perceptor, an ontology has to be created, capturing the events or states to be perceived (e.g. “obstacle on my track” or “obstacle on neighbouring track”). During the validation process, it has to be shown that the sensor data received is mapped by the perceptor to the correct ontology objects. The ontology needs to be sufficiently detailed to cover all relevant aspects of the ODD (e.g. “obstacle on my track in tunnel” and “obstacle on my track in open track section”).

A considerable challenge consists in the justification of equivalence classes used during perceptor evaluation: since the number of different environment conditions and – in the case of obstacle detection – the number of different object shapes to detect is unbounded. As a consequence, feasible validation test suites require the specification of finite collections of equivalence classes, such that a small number of representatives from each class suffices to ensure that *every* class member is detected. The equivalence class identification is problematic, because human perception frequently uses different classes as a trained neural network would use [20]. We have elaborated a new method for equivalence class identification, but this is beyond the scope of this paper (see Sect. 6). In any case, the stochastic independence between channels, achieved through different perception methods applied, reduces the probability that both perceptor channels will produce the same false negatives, to be accepted by the voter.

For the perceptor channels based on neural networks and machine learning, it has to be shown that the training and evaluation data sets are sufficiently diverse, and that the correct classification results have been obtained “for the correct reasons” [20]. In the case of camera sensors and image classifier perceptrors, this means that the image portion leading to a correct mapping into the ontology really represents the ontology element. Moreover, robustness, in particular, the absence of *brittleness* has to be shown for the trained neural network: small variations of images need to be mapped onto the same (or similar) ontology elements. Brittleness can occur as a result of overfitting during the training phase.

Evaluation of the “conventional” Sub-pipeline. We observe that the planning → prediction → control → actuation sub-pipeline does not depend on AI-techniques

and is fully specified by formal models at type certification time. Consequently, no discrepancies between the safety of the specified functionality and that of the intended functionality are to be expected. Therefore, the evaluation of the kernel and train interface unit is performed as any conventional automated train protection system. The ODD helps to identify the relevant system-level tests to be performed, such as transitions between track sections with different equipment, or different weather conditions influencing the train's braking capabilities. These tests, however, are no different from those needed to establish operational safety of non-autonomous trains. Moreover, the functional safety model induces tests covering equipment failures (e.g. failures of the sensor→perceptor sub-pipeline) and the resulting changes between the operational modes described above.

6 Conclusion

We have presented a new architecture for autonomous train controllers in open environments with the normal infrastructure to be expected in European railways today. It has been demonstrated how this could be evaluated and certified on the basis of the existing CENELEC standards, in combination with the novel ANSI/UL 4600 pre-standard dedicated to the assurance of autonomous, potentially AI-based, transportation systems. As a main result, it has been shown that such an evaluation is feasible already today, and, consequently, such systems are certifiable in the case of freight trains and metro trains, but not in the case of high speed trains. This restriction is necessary because no reliable solutions for obstacle detection in high speed trains seem to be available today.

For a “real-world” certification, the qualitative results of this paper need to be supported by concrete risk figures. This is currently investigated, with the application of stochastic model checking on a world model covering the operational design domain, as well as the trains and their ATP mechanisms discussed here. Moreover, the automated synthesis of safety supervisors from ATP-submodels of the world model will be explored with a novel methodological approach by Gleirscher et al. [9], complementing existing results [1]. For calculating the probabilities of residual perceptor errors and for verifying stochastic independence between channels, we have developed a new method based on statistical tests, algorithms for the explanation of image classification results, and the construction of equivalence classes; the effectiveness of this method will be evaluated.

Acknowledgements. We would like to thank Mario Gleirscher for stimulating discussions of earlier releases of this paper. We are also grateful to the anonymous reviewers who provided very helpful suggestions leading to the revised version of this paper.

References

1. Basile, D., ter Beek, M.H., Legay, A.: Strategy synthesis for autonomous driving in a moving block railway system with UPPAAL STRATEGO. In: Gotsman, A., Sokolova, A. (eds.) *Formal Techniques for Distributed Objects, Components, and Systems*. LNCS, pp. 3–21. Springer, Cham (2020)

2. Behm, P., Benoit, P., Faivre, A., Meynadier, J.-M.: Météor: a successful application of B in a large project. In: Wing, J.M., Woodcock, J., Davies, J. (eds.) FM 1999. LNCS, vol. 1708, pp. 369–387. Springer, Heidelberg (1999). <https://doi.org/10.1007/3-540-48119-2.22>
3. Bordini, R.H., Hübner, J.F., Wooldridge, M.: Programming Multi-agent Systems in AgentSpeak Using Jason. Wiley, West Sussex (2007)
4. CENELEC: EN 50128: 2011 Railway applications - Communication, signalling and processing systems - Software for railway control and protection systems (2011)
5. CENELEC: EN 50126 Railway Applications - The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS) - Part 1: Generic RAMS Process (2017)
6. CENELEC: Railway applications - Communication, signalling and processing systems - Safety related electronic systems for signalling (2018)
7. Fisher, M., Mascardi, V., Rozier, K.Y., Schlingloff, B.H., Winikoff, M., Yorke-Smith, N.: Towards a framework for certification of reliable autonomous systems. *Auton. Agent. Multi-agent Syst.* **35**(1), 8 (2020). <https://doi.org/10.1007/s10458-020-09487-2>
8. Flammini, F., Donato, L.D., Fantechi, A., Vittorini, V.: A vision of intelligent train control. In: Dutilleul, S.C., Haxthausen, A.E., Lecomte, T. (eds.) Reliability, Safety, and Security of Railway Systems. Modelling, Analysis, Verification, and Certification - 4th International Conference, RSSRail 2022, Paris, France, 1–2 June 2022, Proceedings. LNCS, vol. 13294, pp. 192–208. Springer, Cham (2022). <https://doi.org/10.1007/978-3-031-05814-1.14>
9. Gleirscher, M., Calinescu, R., Woodcock, J.: RISKSTRUCTURES: a design algebra for risk-aware machines. *Formal Aspects Comput.* **33**(4–5), 763–802 (2021). <https://doi.org/10.1007/s00165-021-00545-4>
10. Haxthausen, A.E., Peleska, J.: Formal development and verification of a distributed railway control system. *IEEE Trans. Softw. Eng.* **26**(8), 687–701 (2000)
11. Haxthausen, A.E., Lecomte, T., Peleska, J.: Standardisation considerations for autonomous train control - Technical Report. Technical report, Zenodo, February 2022. <https://zenodo.org/record/6185229>
12. ISO: ISO/DIS 21448: Road vehicles - Safety of the intended functionality. European Committee for Electronic Standardization (2021). iCS: 43.040.10, Draft International Standard
13. Kephart, J.O., Chess, D.M.: The vision of autonomic computing. *Computer* **36**(1), 41–50 (2003). <https://doi.org/10.1109/MC.2003.1160055>
14. Koopman, P., Kane, A., Black, J.: Credible autonomy safety argumentation. In: Proceedings of the 27th Safety-Critical Systems Symposium, February 2019. https://users.ece.cmu.edu/~koopman/pubs/Koopman19_SSS_CredibleSafetyArgumentation.pdf
15. Koopman, P., Wagner, M.: Toward a framework for highly automated vehicle safety validation. In: Proceedings of the 2018 SAE World Congress/SAE 2018-01-1071 (2018). https://users.ece.cmu.edu/~koopman/pubs/koopman18_av_safety_validation.pdf
16. Koopman, P., Wagner, M.D.: Autonomous vehicle safety: an interdisciplinary challenge. *IEEE Intell. Transp. Syst. Mag.* **9**(1), 90–96 (2017). <https://doi.org/10.1109/MITS.2016.2583491>
17. Marmo, R., Lombardi, L., Gagliardi, N.: Railway sign detection and classification. In: 2006 IEEE Intelligent Transportation Systems Conference, pp. 1358–1363 (2006)

18. Ristić-Durrant, D., Franke, M., Michels, K.: A review of vision-based on-board obstacle detection and distance estimation in railways. *Sensors* (Basel, Switzerland) **21**(10), 3452 (2021). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8156009/>
19. Siemens Mobility GmbH: World premiere: DB and Siemens present the first automatic train, October 2021. <https://press.siemens.com/global/en/pressrelease/world-premiere-db-and-siemens-present-first-self-driving-train>, Press release
20. Sun, Y., Chockler, H., Huang, X., Kroening, D.: Explaining image classifiers using statistical fault localization. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) *ECCV 2020*. LNCS, vol. 12373, pp. 391–406. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58604-1_24
21. The British Standards Institution (BSI), Centre for Connected & Autonomous Vehicles: PAS 1883:2020, Operational Design Domain (ODD) taxonomy for an automated driving system (ADS) - Specification, August 2022
22. Tonk, A., Boussif, A., Beugin, J., Collart-Dutilleul, S.: Towards a specified operational design domain for a safe remote driving of trains. In: *ESREL 2021, 31st European Safety And Reliability Conference*, p. 8p. Angers, France, September 2021. <https://hal.archives-ouvertes.fr/hal-03328878>, eSREL 2021, 31st European Safety And Reliability Conference, Angers, France, 19 September 2021–23 September 2021
23. Trentesaux, D., et al.: The autonomous train. In: 2018 13th Annual Conference on System of Systems Engineering (SoSE), pp. 514–520, June 2018
24. Underwriters Laboratories Inc.: ANSI/UL 4600-2020 Standard for Evaluation of Autonomous Products - First Edition. Underwriters Laboratories Inc., 333 Pfingsten Road, Northbrook, Illinois 60062-2096, 847.272.8800, April 2020
25. UNISIG: Basic System Description, Chapter 2, vol. Subset-026-2 of [26], February 2006. Issue 2.3.0
26. UNISIG (ed.): ERTMS/ETCS - Class 1 System Requirements Specification, vol. Subset-026, February 2006. Issue 2.3.0
27. UNISIG: ERTMS/ETCS System Requirements Specification, Chapter 3, Principles, Chapter 3, vol. Subset-026-3 of [26], February 2012. Issue 3.3.0
28. Withers, J., Stoehr, N.: Automated Train Operations (ATO) Safety and Sensor Development. Technical Report RR 20–21, U.S. Department of Transportation - Federal Railroad Administration, November 2020. <https://railroads.dot.gov/library/automated-train-operations-ato-safety-and-sensor-development>
29. Zhang, Z., Wang, Y., Brand, J., Dahnoun, N.: Real-time obstacle detection based on stereo vision for automotive applications. In: 2012 5th European DSP Education and Research Conference (EDERC), pp. 281–285 (2012)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

