Testing Predictive Automated Driving Systems: Lessons Learned and Future Recommendations

Rubén Izquierdo Gonzalo, Carlota Salinas Maldonado, Javier Alonso Ruiz, and Ignacio Parra Alonso
Are with the Computer Engineering Department, University of Alcalá, Alcalá de Henares, Madrid, 28801, Spain.
E-mail: ruben.izquierdo@uah.es; carlota.salinas@uah.es; javier.alonso@uah.es; ignacio.parra@uah.es

Digital Object Identifier 10.1109/MITS.2022.3170649
Date of current version: 20 May 2022

*Corresponding author.

David Fernández Llorca*
Is with the Joint Research Centre, European Commission, Sevilla, 41092, Spain and the Computer Engineering Department, University of Alcalá, Alcalá de Henares, Madrid, 28801, Spain.
E-mail: david.fernandez-llorca@ec.europa.eu

Miguel Ángel Sotelo
Is with the Computer Engineering Department, University of Alcalá, Alcalá de Henares, Madrid, 28801, Spain.
E-mail: miguel.sotelo@uah.es

Abstract—Conventional vehicles are certified through classical approaches, where different physical certification tests are set up on test tracks to assess the required safety levels. These approaches are well suited for vehicles with limited complexity and limited interactions with other entities as last-second resources. However, these approaches do not allow the evaluation of safety with real behaviors for critical and edge cases nor the evaluation of the ability to anticipate them in the mid or long term. This is particularly relevant for automated and autonomous driving functions that make use of advanced predictive systems to anticipate future actions and motions to be considered in the path planning layer. In this article, we present and analyze the results of physical tests on the proving grounds of several predictive systems in automated driving functions developed within the framework of the BRidging Gaps for the Adoption of Automated VEhicles (BRAVE) project. Based on our experience in testing predictive automated driving functions, we identify the main limitations of current physical testing approaches when dealing with predictive systems, analyze the main challenges ahead, and provide a set of practical actions and recommendations to consider in future physical testing procedures for automated and autonomous driving functions.

To place a vehicle on the market, car manufacturers need prior authorization, granted by the competent authority, after proving that the vehicle complies with all applicable regulatory standards and safety certification requirements. Whether through vehicle type approval or self-certification approaches, original equipment manufacturers (OEMs) must pass stringent certification processes to validate a component, a system, or the entire vehicle [1].

Background on Certification Tests

Conventional vehicles are certified through classical approaches, where different physical certification tests are set up on test tracks or test benches to assess the required safety level using various performance criteria. These approaches are well suited for components, systems, and vehicles with limited complexity and limited interactions with other entities (e.g., braking tests). However, as the complexity of systems increases (e.g., electronic stability control), classical approaches cannot address all relevant safety areas due to two main reasons. The first is the large number of safety-related systems (including multiple electrical and electronic systems), which increases risks from systematic failures and random hardware failures. This is reasonably well addressed by the existing functional safety and Automotive Safety Integrity Levels requirements in the automotive industry (e.g., ISO 26262). The second is the enormous variability of possible multiagent scenarios, which, on the one hand, implies the need for a formal safety model [2], and on other hand, has led to the introduction of simulation-based safety-oriented audits as a way to complement physical vehicle testing [3].

With the introduction of assisted [Society of Automotive Engineers (SAE) Levels 1 and 2], automated (SAE Level 3), and autonomous (SAE Levels 4 and 5) driving systems [4], [5], the overall complexity increases in terms of the number of software functions, variants of multiagent scenarios and interactions, and potentially affected safety areas [6]. The complexity of these systems, and therefore the difficulty to test them, increases with the level of automation, being particularly important in the step from SAE Level 3 to 4 since the automated driving system must be able to reach a minimal risk condition within its Operational Design Domain without user/passenger intervention [7].

New innovative testing approaches, including procedures of different natures, are needed for future vehicle safety regulatory frameworks and for assessments under current exemption procedures [8]. New online/in-service safety monitoring and verification mechanisms [9] that act after the market deployment of automated driving systems [6] are also needed as a way of reducing the need to test all possible combinations at the time of type approval. Several national and international regulatory and standardization initiatives and projects are already underway to tackle all these problems [10].

One of the most solid regulatory proposals is being developed by the Working Party on Automated/Autonomous and Connected Vehicles (GRVA) of the United Nations Economic Commission for Europe (UNECE) World Forum for Harmonization of Vehicle Regulations (WP.29). It is based on three pillars that must be assessed together [11].

1) *Audit and Assessment*: This includes the use of simulation to cover all types of scenarios but especially edge case scenarios that are unlikely to occur in real-world traffic
2) *Physical Certification Tests:* These assess critical scenarios, performed in controlled environments on test tracks (closed roads) and involving sophisticated equipment, such as lightweight global vehicle [12], articulated pedestrian [13], and bicyclist [14] targets
3) *Real-World Test Drives*: These are devised as a "driving license test" for automated driving systems to assess the overall capabilities and behavior of the vehicle in nonsimulated traffic on public or open roads.

This approach has been the one adopted by United Nations to regulate the approval of advanced emergency braking systems (AEBS) [15] and, more recently, automated lane keeping systems (ALKS) [16]. These regulations have recently been integrated in countries such as Japan and Germany, enabling the commercialization of the first SAE Level 3 automated driving systems by two different car manufacturers [17], [18].

A similar approach was provided by the PEGASUS project [19], including laboratory, simulation, testing site tests, and field tests, with particular emphasis on the definition of use cases and test scenarios. In another project, ENABLE-S3 [20], the goal was to reduce the testing efforts of traditional road testing by focusing on virtualization. One of the main contributions was the use of semivirtual systems, such as the DrivingCube [21], which combines both simulation and ready-to-drive vehicles on a chassis dynamometer and on a power-train testbed. This approach can be considered as an intermediate step between pure simulation-based verification and physical certification testing or as a subfield of simulation using vehicle in the loop (VIL).

The three approaches mentioned (i.e., simulation, including VIL simulation, and physical and real-world testing) have strengths and weaknesses [22], which is why it is important to implement them holistically [11]. In Table 1, we illustrate the advantages and disadvantages of all testing approaches. As can be observed, the methods are somehow complementary. For example, although simulation-based testing allows full controllability, repeatability, and variability in a very efficient way, these methods exhibit very low fidelity and lack real-world behaviors. We can increase fidelity at the cost of increasing complexity and thus decreasing efficiency, from VIL simulation to physical testing on closed tracks. But the absence of real behaviors remains a problem, which can only be partially compensated by testing in real traffic conditions (open roads).

Another relevant variable refers to the degree to which the driving functions can be optimized on the specific scenarios, which can be seen as a shortcut by OEMs to overfit the performance of their systems to the test scenarios. This has a negative impact on the possible fidelity of the tests, while the performance of the systems in real traffic remains unknown. In general, if the simulation conditions are known a priori, if the physical test conditions are of closed tracks, or if the proving grounds or the test area are in real traffic, all test methods are potentially subject to overfitting. Still, there are some differences. For example, on the one hand, the uncontrolled conditions of open road testing make this method less prone to overfitting. On the other hand, the low variability and the strict control and repeatability conditions of the scenarios in the physical certification on closed roads are favorable conditions for the optimization of the systems to the proposed scenarios.

A closer look at the different methods reveals that complementarity is limited as the scenarios addressed by each approach are of a different nature. In Figure 1, we illustrate three different types of scenarios (typical, critical, and edge), referring to their probability of occurrence with

Table 1. The main features of the different testing approaches.

Approaches	Control./Repeat.	Scalability	Variability	Fidelity	Efficiency	Test Overfitting*	Real Behaviors†
Simulation	★★★★★	★★★★★	★★★★★	★☆☆☆☆	★★★★★	★★★☆☆	☆☆☆☆☆
VIL simulation	★★★★★	★★★★☆	★★★★☆	★★☆☆☆	★★★☆☆	★★★★☆	☆☆☆☆☆
Physical track	★★★★☆	★★★☆☆	★☆☆☆☆	★★★☆☆	★★☆☆☆	★☆☆☆☆	☆☆☆☆☆
Open roads	★☆☆☆☆	★☆☆☆☆	★★☆☆☆	★★★★★	★★★☆☆	★★★★☆	★★★★★

Note: * *Test overfitting* refers to the degree to which the systems can be optimized on specific test scenarios. A high score means a low probability of overfitting. † *Real behaviors* refers to the degree to which the test method can include actual behaviors of other agents (pedestrians, cyclists, other drivers, and so on). Control.: controllability; Repeat.: repeatability

Incorporating real behaviors in critical and edge scenarios, both in simulation environments and in physical tests on tracks, is not straightforward.

respect to the degree of complexity and potential risks. As can be observed, the distribution of scenarios follows a long-tail distribution, which requires significant scale [2] to discover and properly handle the long tail of rare events [23]. In Table 2, the types of scenarios that can be addressed for each testing method are depicted. As can be inferred, high fidelity is achievable only for typical scenarios, with higher uncertainty for critical and edge scenarios.

In addition, current testing approaches do not allow the assessment of safety with real behaviors for critical and edge cases. This is particularly relevant for automated and autonomous driving functions that make use of predictive perception, i.e., systems that learn and model the behaviors and interactions of traffic agents to anticipate future actions and motions to be considered in the path planning layer. In fact, higher levels of automation (autonomous) are expected to be achieved thanks to predictive capabilities [5]. These predictive systems are expected to enable autonomous driving to become more like manual driving, increasing safety margins, reducing risks, and providing smoother and more acceptable motion trajectories. All these factors have a direct effect on the perceived safety, risk, and trust of users, which are directly linked to user acceptance [24].

Incorporating real behaviors in critical and edge scenarios, both in simulation environments and in physical tests on tracks, is not straightforward. However, the advantages in safety and comfort [25] that predictive autonomous driving brings require new efforts to improve test methods for future certification processes. In an attempt to take the first steps, in this article, we present and analyze the results of physical tests on the proving grounds of several predictive systems in automated driving functions developed within the framework of the BRAVE project [26]. A number of use cases involving vehicles and vulnerable road users (VRUs) on different scenarios were defined, some of them directly equivalent to the European New Car Assessment Programme (Euro NCAP) test protocols for automatic emergency braking (AEB) systems [27], [28].

We focus our work on physical tests because they have a higher level of maturity (e.g., Euro NCAP test protocols) and offer more room for improvement in terms of variability and overfitting as well as a better relationship between fidelity and controllability/repeatability [22]. These tests are also essential to validate the fidelity (reality gap) of simulation-based methods [11]. Based on our experience in testing predictive automated driving functions, we identify the main limitations of current physical testing approaches when dealing with predictive systems, analyze the main challenges ahead, and provide a set of practical actions and recommendations to consider in future physical testing procedures for automated and autonomous driving functions. From this practical bottom-up approach (from direct empirical evidence in testing to conceptual identification of needs for future certification processes), we aim to contribute to and complement the high-level approaches carried out worldwide to adapt regulatory standards and safety certification requirements for increasingly advanced autonomous driving systems.

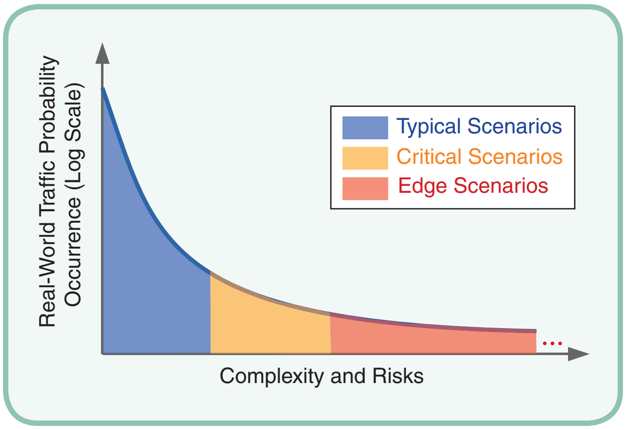

FIG 1 The types of scenarios, probability of occurrence in real-world traffic (log-scale), complexity, and risk. The distribution follows a long-tail distribution. Refer to [11] for more details on the scenarios.

Table 2. The distribution of scenarios by the testing approach.

Approaches	Typical	Critical	Edge
Simulation	✓	✓	✓
VIL simulation	✓	✓	✓
Physical track		✓	
Open roads	✓		

An Overview of Predictive Perception Systems

The experiments were performed using the University of Alcalá's (UAH's) DRIVERTIVE vehicle [29]. DRIVERTIVE is a mechanically automated vehicle that carries multiple sensors for environment perception, such as an HDL-32E lidar, three radars, and multiple cameras. The objective of the experiments was to validate the goodness of predictive systems to outperform conventional last-second reaction systems. These experiments involved different traffic agents, such as vehicles, pedestrians, and cyclists in critical, controlled situations. For self-containment reasons, in this section, we briefly introduce the predictive systems used in the experiments.

> The objective of the experiments was to validate the goodness of predictive systems to outperform conventional last-second reaction systems.

Regarding the interaction with vehicles, two deep learning approaches were used for the inference of both the intention and trajectories of vehicles. The intention prediction system uses a classification approach with a Resnet50 [30] backbone to classify enhanced images encoded as a single red, green, blue image, including context, interactions, and vehicle motion patterns [31]. This model estimates the probability of keeping or changing lanes. The trajectory prediction model [32] uses radar targets to create a top-view time-evolving map. The most likely future representation of this input map is inferred by a U-net model [33], [34] trained for this purpose. The output of this model is used as a triggering mechanism to assess critical cut-in maneuvers.

For pedestrians, body and facial key point detectors [35] act as the core for prediction systems. Deep learning approaches use body key points to anticipate changes in pedestrian motion patterns [36] and also to predict the intention of crossing from the sidewalk or at a crosswalk [37], [38]. Face key points are of paramount importance for the detection of crossing intention as eye contact is a powerful nonverbal channel of communication often used to express intention to drivers. This eye-contact detection was very useful in improving response time in tests with moving head dummies using the eye-contact signal as a trigger for the braking response.

Finally, interaction experiments with bicyclist dummies were conducted using an instance segmentation approach. This method provides object-level detections that allow for radar data fusion. Cyclist path prediction was implemented using a standard approach based on a physical model [39]. However, in the tests, no actions were expected from the cyclist (e.g., switching dynamics [40]), and nothing could be predicted. Interaction with the dummy was limited to maintaining the safety distance, being detected at 100 m, which provided enough space to ensure a smooth and safe maneuver.

The Experimental Validation of Predictive Automated Systems

This section provides experimental results derived from physical tests developed on the proving ground at Union Technique de l'Automobile et du Cycle (UTAC) facilities [41]. These tests are intended to evaluate autonomous capabilities at standardized Euro NCAP protocol tests as well as custom-defined autonomous tests. The goal of these experiments is to validate and measure the safety and goodness of predictive systems in critical and controlled circumstances. This section begins with a description of the experiments and their correspondence with Euro NCAP tests. For the sake of space, the most relevant configurations are analyzed in detail in two sections: "The Assessment of Vehicle–VRU Interaction," related to VRUs, i.e., pedestrian and bicyclist interactions, and "The Assessment of Vehicle–Vehicle Interaction," related to vehicle–vehicle interactions.

A Description of the Experiments

Several rounds of experimentation were conducted on the proving grounds at UTAC (depicted in Figure 2), a technology center located at Linas-Monthléry, France. The experiments were carried out in the framework of the BRAVE project [26] using the UAH's DRIVERTIVE vehicle. During the experiments with the vehicle under test (VUT), the following use cases were tested: pedestrian (VRU-1 and VRU-2), cyclist (VRU-3), cut-in vehicle (VEH-1), and intersection (VEH-2). A detailed description of each use case is provided in the next section. All use cases were designed and tested to efficiently assess the performance and the added value of the predictive systems developed in BRAVE.

Some of the tests conducted in this work have a direct equivalence with Euro NCAP use cases. Other tests have been specifically devised and designed to be tested in the BRAVE project as a means to provide further recommendations to the Euro NCAP. Table 3 provides the equivalence between tests conducted in BRAVE and use cases defined by the Euro NCAP.

As can be observed, the VEH-2 intersection use case is totally new, having no equivalence in Euro NCAP tests. It

FIG 2 The proving grounds at UTAC's premises.

has been completely devised and designed in the framework of the BRAVE project in an attempt to provide recommendations to the Euro NCAP on how to properly test automated vehicles when dealing with interactions with other vehicles at intersections.

Table 3. The equivalence between BRAVE and Euro NCAP use cases.

BRAVE Test	Euro NCAP Test Equivalence
VRU-1 GRAIL	
Configuration 1: Crossing on green LED signal	
Configuration 2: CPLA-50 deceleration then overtake	—
Configuration 3: Longitudinal then crossing on green LED	CPLA-50 day
VRU-2 PEDESTRIAN	—
Configuration 1: CPNA-50. full stop	
Configuration 2: CPNA-O-50, full stop	CPFA-50
Configuration 3: CPNA-25, AES	CPNC-50*
VRU-3 CYCLIST	CPNA-25 day
Configuration 1: CBLA-25, overtake	
Configuration 2: CBLA-25, deceleration	—
Configuration 3: CBLA-25, deceleration then overtake	CBLA-50 day
VEH-1 CUT-IN	—
Configuration 1: Deceleration within central lane	
Configuration 2: AES onto the left lane	Cut-in 50/10
VEH-2 INTERSECTION	—
Configuration 1: Straight line with crossing target	
Configuration 2: Straight line with turning target	—
	—

*The position of the car that causes the occlusion is lateral rather than longitudinal. The dummy corresponds to an adult instead of a child. CPLA: car-to-pedestrian longitudinal adult; CPNA: car-to-pedestrian nearside adult; CPFA: car-to-pedestrian farside adult; CPNC: car-to-pedestrian nearside child.

Table 4. The descriptions of configurations in VRUs' use cases.

VRU Use Case	Configuration
VRU-1	Configuration 1: Pedestrian crossing at 50% impact (stop)
	Configuration 2: Pedestrian walking in parallel (not crossing)
	Configuration 3: Pedestrian turns head and crosses
VRU-2	Configuration 1: Pedestrian crossing at 50% impact (stop)
	Configuration 2: Occluded pedestrian crossing at 50% impact (stop)
	Configuration 3: Pedestrian crossing at 25% impact (avoid)
VRU-3	Configuration 1: Overtake
	Configuration 2: Reduce and follow
	Configuration 3: Reduce, follow, and overtake

The Assessment of Vehicle–VRU Interaction

Several tests were conducted to assess vehicle–VRU interactions. For each use case, several configurations were exhaustively tested, as described in Table 4. The schematic description of all the use cases tested is provided as follows.

- *VRU-1*: An external human–machine interface (HMI) [GReen Assistant Interfacing Light (GRAIL) system [42]] onboard the vehicle is used to interact with pedestrians willing to cross the street.
- *VRU-2*: The emergency reactions of the VUT are tested in different situations when interacting with pedestrians whose predicted trajectory intersects the VUT estimated trajectory.
- *VRU-3*: The emergency reactions of the VUT are tested when interacting with cyclists.

For the sake of space, the description of experiments is provided only for the most challenging configurations, namely VRU-1 Configuration 3, VRU-2 Configuration 2, and VRU-3 Configuration 3. Data showing a summary of the results obtained for all the use cases in their different configurations are provided at the end of the section.

VRU-1 Configuration 3: Pedestrian Walks in Parallel, Turns Head Toward the Car, and Crosses the Street

These tests were conducted using a dummy with an articulated head that can be turned to emulate that the pedestrian is looking at the driver before starting to cross the street. Initially, the pedestrian walks in parallel with the road. After a while, the pedestrian stops and turns his head toward the car, emulating eye contact with the driver. The VUT is then expected to detect the pedestrian's intention to cross the street and, consequently, to decrease speed, gradually and smoothly, until coming to a full stop. At the same time, the VUT automatically activates the GRAIL system as a means to signal to the pedestrian that the VUT has detected his intention to cross, and it is going to brake. The graphical representation of this configuration is depicted in Figure 3. Figure 4 depicts an example of how the computer vision system detects the dummy face. The green dots overlaid on the face represent the face features recognized by the algorithm. These features are used to find out whether the dummy is looking at the VUT.

Figure 5 shows the data logged during the experiments conducted at 40 km/h. In the figure, the longitudinal distance between the VUT bumper and the pedestrian (bumper to pedestrian) and the VUT speed are depicted in blue and orange, respectively. The dashed orange curve shows the VUT reference speed, while the solid orange curves depict the VUT real speed. The figures also show the moments of pedestrian detection (pedestrian considered) and face detection (pedestrian looking) by means of green triangles. The first triangle determines the moment when the GRAIL system is preactivated (even though the pedestrian is first detected well in advance by the onboard camera). The second triangle determines the moment when the pedestrian face is looking at the driver.

It can be clearly observed that the VUT speed starts to diminish as soon as the GRAIL system is switched on. The VUT comes to a full stop in front of the pedestrian at a distance ranging between 6 and 7 m, providing safety and a comfortable margin for the pedestrian to cross. The quick decisions taken by the system lead to anticipative maneuvers that allow the VUT to perform smooth and safe actions. Once more, this way of operation is expected to contribute to increasing the feeling of comfort in the VUT passengers and the feeling of safety and respect in pedestrians.

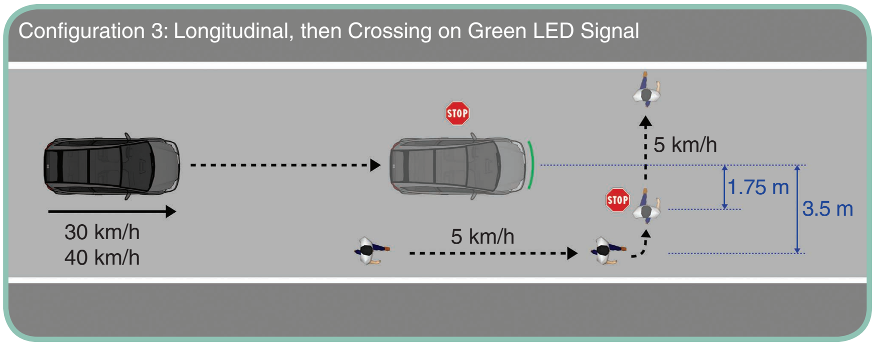

FIG 3 Use case VRU-1 Configuration 3: Walk, stop, and cross.

VRU-2 Configuration 2: Occluded Pedestrian Crossing at 50% Impact (Stop)

This test was conducted with a dummy emulating a pedestrian who emerges from behind a parked car and starts to cross the street unexpectedly. As in the previous configuration, the trajectory of the pedestrian is estimated to intersect the VUT trajectory (if both the VUT and the pedestrian keep a constant velocity) at the central point of the VUT. Thus, the overlap between the VUT and the pedestrian at the time of impact is estimated to be 50%. This test was conducted at two different VUT speeds: 30 and 40 km/h. The pedestrian walks at a constant velocity of 5 km/h with an initial lateral offset of 6 m with respect to the main axis of the VUT. At the beginning of the test, the pedestrian is not visible from the VUT given that it is occluded by a parked car. Figure 6(a) shows a graphical representation of this configuration, while Figure 6(b) shows a snapshot of the dummy being detected by the computer vision algorithm during the execution of the test. As can be observed, the detected pedestrian (highlighted by a blue bounding box) is only partially visible from the VUT given that it is occluded by the parked car.

The early detection of the upper body of the pedestrian makes it possible to start the reaction maneuver in time. Figure 7 provides a graphical representation of the main variables measured during the execution of the test with the VUT speed of 40 km/h. Figure 7(a) shows the VUT-VRU longitudinal distance (on the left) in blue and the VUT speed (on the right) in orange. Both the reference VUT speed (solid) and the current speed (dashed) are provided. As observed, the VUT reference speed is set to zero as soon as the pedestrian is detected. This is intended to execute the braking maneuver as abruptly as possible, given that the situation is critical, and safety is the only variable to consider.

The VUT is capable of coming to a full stop at a distance of just a few centimeters (3 cm) in front of the pedestrian. The VUT cannot start the braking maneuver until the pedestrian starts to be perceived by the onboard sensors. This happens

FIG 4 The detection of a dummy face using computer vision.

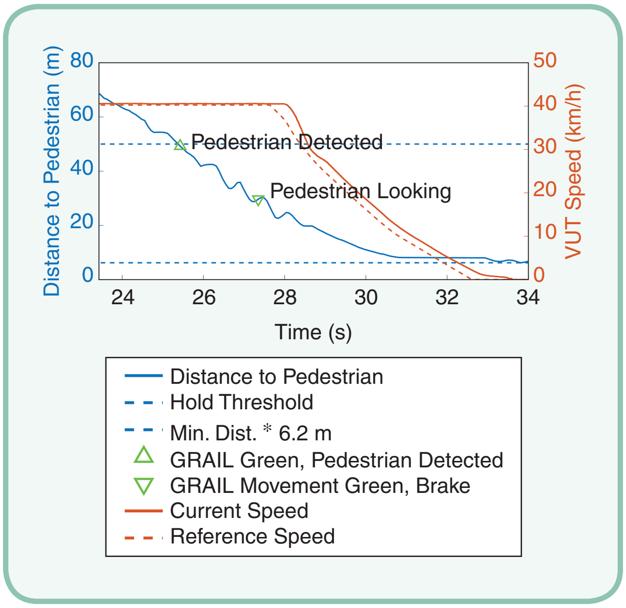

FIG 5 Data logged in VRU-1 Configuration 3 at 40 km/h. Min. Dist.: minimum distribution.

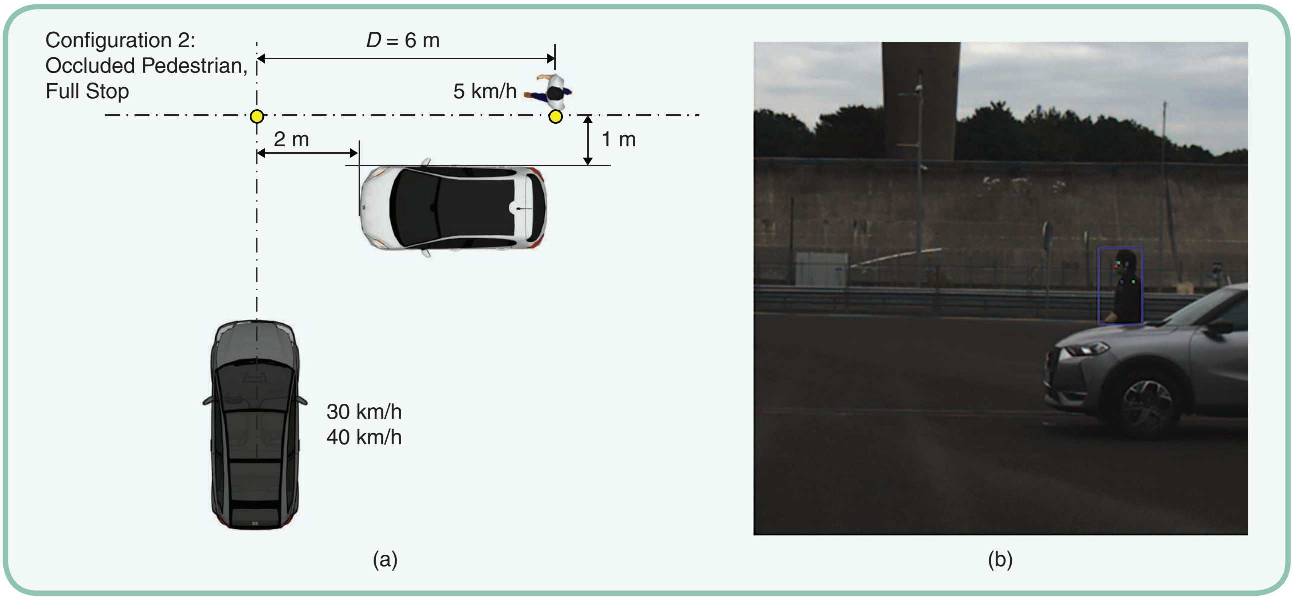

FIG 6 Use case VRU-2 Configuration 2: occluded pedestrian crossing at 50% impact (stop). (a) Graphical representation. (b) VRU detection example.

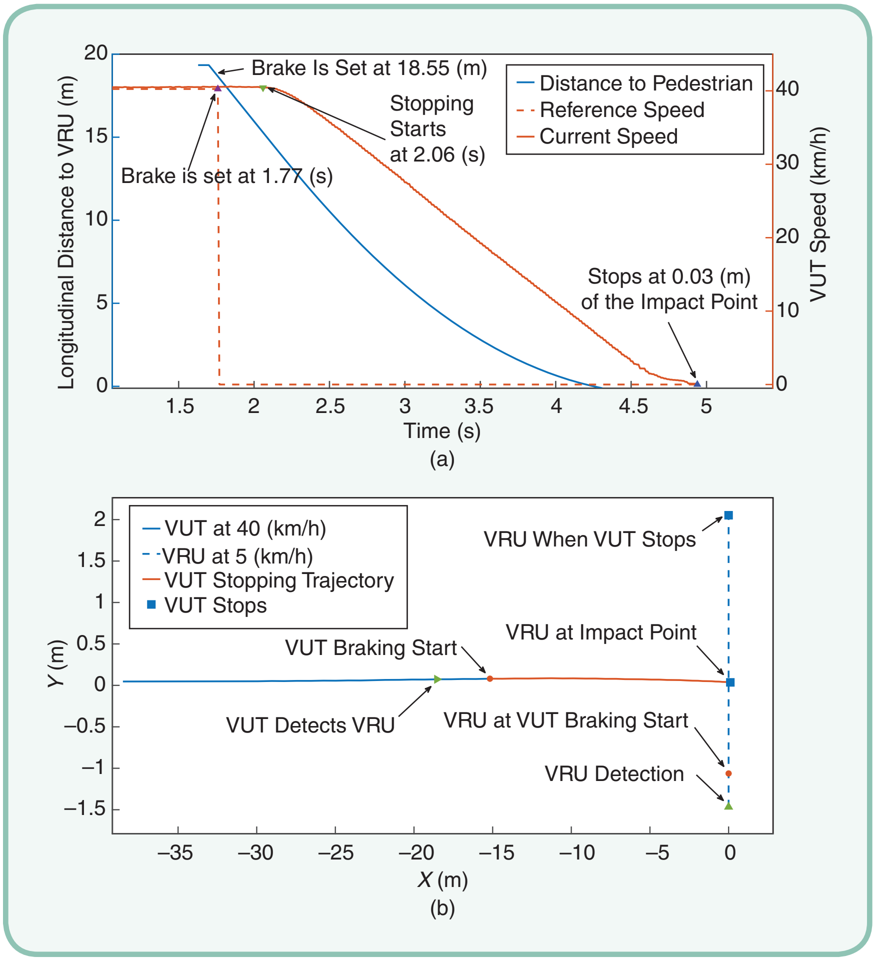

FIG 7 VRU-2 Configuration 2: results at VUT = 40 km/h. (a) VUT speed and VRU distance. (b) VUT and VRU trajectories.

with certain delay given the occlusion conditions [see details in Figure 7(b)]. In these circumstances, it can be stated that a VUT velocity of 40 km/h is the limit to guarantee safety in this typical urban situation. As a consequence, the recommendation to set a speed limit of 30 km/h in urban areas is fully supported by the results obtained in this test.

VRU-3 Configuration 3: Reduce, Follow, and Overtake

In this configuration, the VUT approaches the cyclist, observes that there is oncoming traffic on the adjoining lane, and waits until the lane is free. After that, the VUT starts to execute a smooth overtaking maneuver with sufficient lateral distance with respect to the cyclist to ensure comfort and safety simultaneously. The graphical description of this configuration is depicted in Figure 8. The VUT moves at a velocity ranging 30/40 km/h. The cyclist moves at a constant speed of 15 km/h. The initial lateral offset between the VUT main axis and the cyclist trajectory is 0.45 m. Figure 9 shows the cyclist dummy used in the experiment.

Figure 10 shows the VUT reference (blue dashed) and current (blue solid) speed and the VUT lateral offset (in orange), respectively. As can be observed, the VUT performs a smooth maneuver divided in three steps. In the first step, the VUT keeps a constant speed of 40 km/h. In a second step, the VUT decreases velocity until reaching a value of 20 km/h that is kept constant for a while until the oncoming traffic disappears. In the final step, the VUT performs the overtaking maneuver while keeping a constant speed of 20 km/h and leaving a lateral safety distance of 4 m with respect to the cyclist. As in the previous configurations, the VUT is not allowed to initiate the overtaking maneuver until the distance with the cyclist reaches a predetermined value, and no predictive system was implemented, given that it was considered to provide no added value in the conditions specified in the test. The passengers' comfort and the cyclist's safety are once more guaranteed due to the early detection and smooth action.

The conclusion of these experiments with cyclists is that although comfort and safety are guaranteed, the artificial intelligence (AI)-based predictive systems can offer much more for the sake of safety and anticipation in critical situations. However, it would be necessary to further develop the dummy technology to enable additional dummy movements (turning head, raising arm, inclining the body to the left or to the right, changing the pedaling pace, and so on). The deployment of such technology would allow the testing of much more challenging cyclist-based use cases.

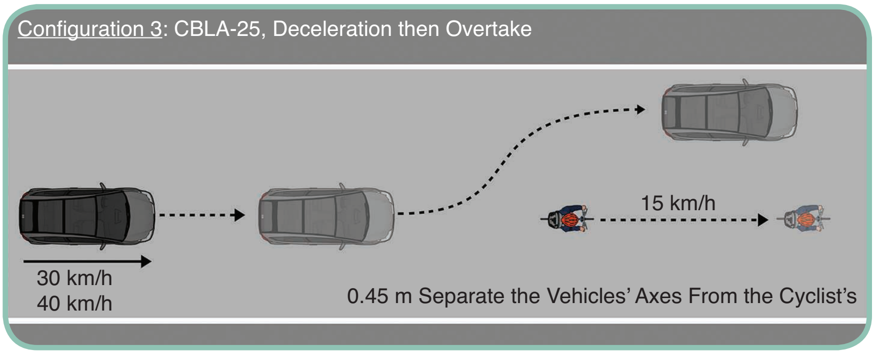

FIG 8 Use case VRU-3 Configuration 3: reduce, follow, and overtake.


FIG 9 The cyclist dummy used in VRU-3 experiments.

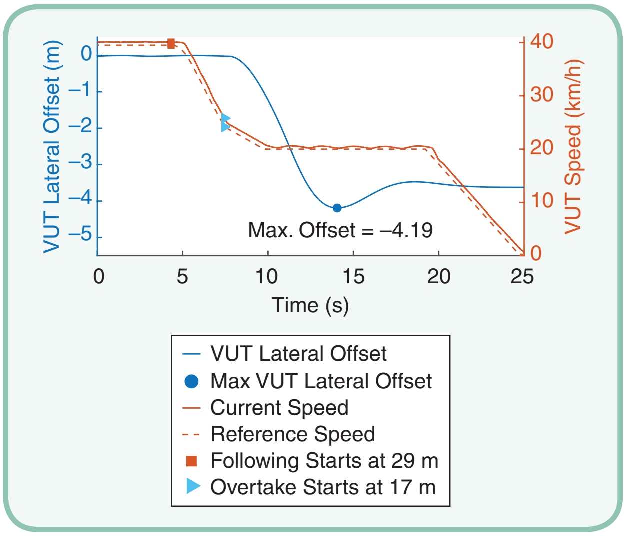

FIG 10 VRU-3 Configuration 3: VUT speed and lateral offset. Max.: maximum.

The Assessment of Vehicle–Vehicle Interaction

Several tests were conducted to assess vehicle–vehicle interactions following different configurations, as described in Table 5. The schematic description of all the use cases tested is provided as follows.

- *VEH-1 Cut-in*: The VUT interacts with other vehicles willing to merge into the mainstream lane from an entry ramp.
- *VEH-2 Intersections*: The VUT interacts with other vehicles entering an intersection.

As in the previous section, the description of experiments is provided only for the most challenging configurations, namely VEH-1 Configuration 2 and VEH-2 Configuration 2. Data showing a summary of the results obtained for all the use cases in their different configurations are provided at the end of the section.

Table 5. Descriptions of configurations in use cases VEH-1 and VEH-2.

VEH Use Case	Configuration
VEH-1	Configuration 1: Reduce and follow
	Configuration 2: Overtake
VEH-2	Configuration 1: Crossing at intersection
	Configuration 2: Turning at intersection

VEH-1 Configuration 2: Change Lane

Two different configurations were tested in this use case. Both configurations simulate a vehicle entering a highway

> As can be observed, the VEH-2 intersection use case is totally new, having no equivalence in the Euro NCAP tests.

from an entry ramp. In both cases, the trajectories of the VUT and the merging vehicle must be synchronized to generate a smooth maneuver. The VUT drives at a constant speed of 50 km/h. The global vehicle target (GVT) drives at a constant speed of 10 km/h. These relative speeds simulate a common highway scenario in which the main traffic flow drives at 120 km/h and the side traffic flow merges at 80 km/h. The first configuration (Configuration 1) evaluates the adaptive cruise control functionality, assuming that the GVT will merge in front of the VUT and that there is no chance to use the adjoining lane. The VUT should start to brake as soon as it considers that the GVT will merge into the VUT's trajectory.

The second configuration (Configuration 2) evaluates the automatic emergency steering (AES) functionality, assuming that the GVT will merge in front of the VUT and that the adjoining lane is available to be used in a lane change maneuver. The VUT should start to change the lane as soon as it considers that the GVT will merge into the VUT's trajectory. The VUT's performance relies on the prediction or detection of an oncoming lane change in both configurations. The earlier the lane change is detected, the longer the time gap and the relative distance to the GVT will be. This anticipation will allow the implementation of smooth actions, leading to higher safety and comfort levels. The graphical description of this configuration is depicted in Figure 11, while Figure 12 shows the vehicle dummy used in the experiments.

In addition, two different use cases were applied for each configuration, triggering the merging maneuver by the GVT as a function of the time to collision (TTC) between the VUT and the GVT once the merging maneuver by the GVT is completed (the time that the GVT needs to perform the lane change maneuver is fixed and known). The TTC means the remaining time before the VUT would strike the GVT, assuming that both the VUT and GVT would continue to travel with the speed they are traveling. For the first case, a TTC = 0 s was used. That is, the GVT initiates the merging maneuver to supposedly end with its rear bumper in contact with the front bumper of the VUT if none of the vehicles modify their speed and assuming in this case a relative speed of 40 km/h.

The second case was defined to be more challenging, with a TTC = –1.5 s (in this case, negative TTC values correspond to a post-collision situation). That is, the GVT starts the merging maneuver so as to supposedly end it with a relative distance from the front bumper of the VUT to the rear bumper of the GVT of –16.67 m for a relative speed of 40 km/h. In other words, the time that the VUT would have to react in this second case is 1.5 s less than in the first case. Both cases would result in a collision if the VUT took no action. By using the TTC between the VUT and the GVT at the end of the merging maneuver of the GVT, these scenarios can be easily adapted to different distances and speeds.

As already mentioned, the relative velocity between the VUT and the GVT is 40 km/h, emulating two vehicles driving on a highway at 120 km/h (VUT) and 80 km/h (GVT), respectively. On this occasion, the VUT checks the adjoining lane and determines that there is no oncoming traffic. After that, the VUT proceeds to execute a lane

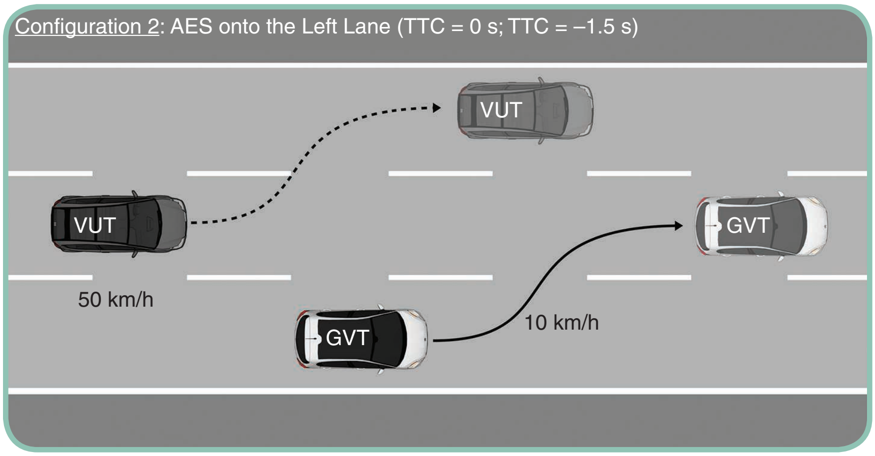

FIG 11 Use case VEH-1 Configuration 2: overtake.


FIG 12 The vehicle dummy used in VEH use cases.

change overtaking maneuver to leave plenty of room for the GVT to merge safely. The further the anticipation, the smoother the maneuver, and the larger the safety margin. This maneuver is denoted as *AES*.

Figure 13(a) shows the longitudinal distance between the VUT and GVT, in blue; the estimated distance (dashed); and the lateral offset, in orange, for TTC = −1.5 s (the most challenging conditions) using the baseline predictive system based on a Kalman filter in use case VEH-1 Configuration 2. Similarly, Figure 13(b) shows the same variables using the predictive system developed in BRAVE. Note the different scale in the time axis.

In this case, the reaction time of the baseline system is 1.2 s, while the reaction time of the BRAVE predictive system is 0.5 s. Under these demanding conditions, the difference of anticipation is 0.7 s in favor of BRAVE predictive system. This difference is to endow the system with additional reaction time. As a consequence, the predictive system allows us to achieve further comfort and safety. The BRAVE predictor leads to higher anticipation times systematically. The average lane change detection time achieved by the BRAVE predictor is 0.77 s (very similar to 0.65 s, the result achieved on video sequences after exhaustive testing). This is a bit more than 300 ms faster than the average reaction time of humans [43], which is around 1.08 s, and 800 ms faster than the average reaction time of the baseline predictive system based on Kalman filtering (1.56 s). Another relevant remark is the fact that the BRAVE vehicle scored full points on all use cases similar to the Euro NCAP tests.

VEH-2 Configuration 2: Turning at Intersections

Two different configurations were tested in this use case. Both configurations simulate a couple of vehicles, the VUT and the GVT, entering an intersection coming from perpendicular directions. In the first case, the GVT continues forward at the intersection, cutting the VUT trajectory. In such circumstances, the VUT must decelerate and accommodate its speed to give way to the GVT. In the second case, the GVT turns right at the intersection without interrupting the VUT trajectory at all. The VUT initially decreases its speed until it predicts the turning maneuver of the GVT. In that moment, the VUT reference speed is resumed to the cruise value that it had before entering the intersection. This can be seen as a case of false positive detection, which, in this case, increases safety in an uncertain situation. Furthermore, the ability to anticipate the turn of the GVT allows optimizing the speed of the VUT at the intersection. False positive testing involves use cases where the GVT finally aborts the expected maneuver. This use case is relevant given that it looks at evaluating systems that can improve the efficiency of traffic at intersections. The graphical description of this scenario is depicted in Figure 14.

The onboard sensors were used to detect the location of the GVT and the relative distance and velocity with respect to the VUT. Figure 15(a) shows the positions of the VUT and GVT during the execution of a test with an initial VUT speed of 40 km/h. The closest distance between the VUT and the GVT during the test was again around 4 m. Figure 15(b) shows the VUT reference and current speeds. As can be observed, the VUT reduces speed following a smooth pattern. The reduction of the VUT speed is not as big as it was in Configuration 1, given that the VUT eventually predicts the intention of the GVT to turn at the intersection. As soon as the intention is detected (around 2 s after starting to decrease speed, marked by an orange square), the VUT resumes its reference speed until

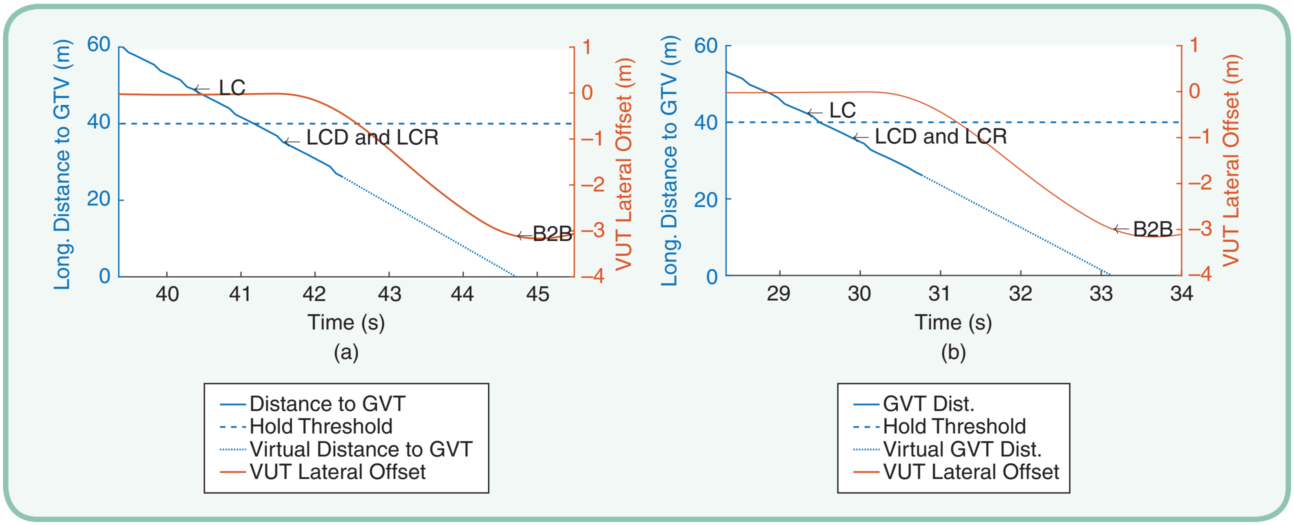

FIG 13 Use case VEH1 Configuration 2: TTC = −1.5 s; longitudinal distance to GVT and VUT lateral offset. (a) Baseline prediction: lane change detection after 1.2 s. (b) BRAVE prediction: lane change detection after 0.5 s. Dist.: distribution; LC: lane change; LCD: lane change detection; LCR: lane change reaction; B2B: bumper-to-bumper.

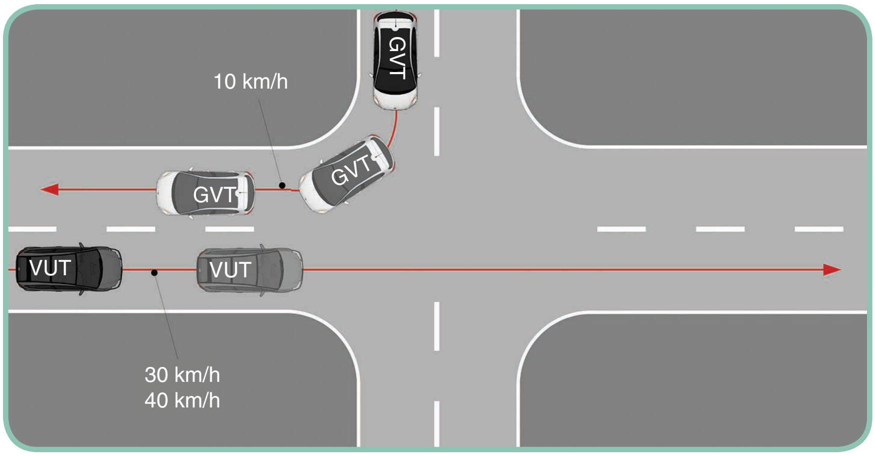

FIG 14 Use case VEH-2 Configuration 2: turning at intersection.

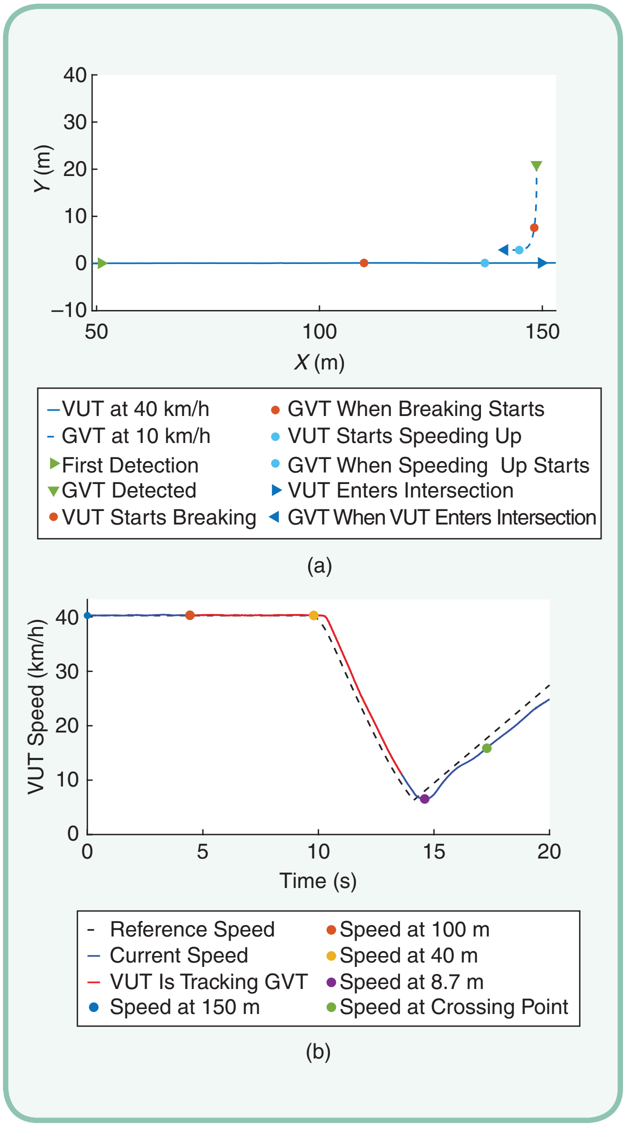

FIG 15 Use case VEH-2 Configuration 2. (a) VUT and GVT positions. (b) VUT speed (real and reference).

reaching the cruise speed it had before entering the intersection. Safety and comfort are once more the desired variables to optimize.

This use case has been completely devised in this project, given that it corresponds to none of the use cases considered by the Euro NCAP. The deployment of this use case allows us to issue some recommendations on how to define and execute this type of validation test in the Euro NCAP. The results achieved in these tests, in terms of safety gap and reaction time (time needed to estimate the intention of the GVT since it enters the intersection), set a baseline for further testing of more complex intersection use cases in the Euro NCAP.

Discussion

After addressing the technical descriptions of the tests performed and the results obtained, this section discusses the main findings and the limitations encountered in testing the predictive systems. It also offers potential actions to be implemented in future test procedures. The proposed recommendations are classified according to their difficulty of implementation or time horizon.

Main Findings

After the thorough experimentation conducted at UTAC's premises to test the predictive systems developed in the BRAVE project, the following findings can be remarked upon.

- In general, it can be safely stated that predictive systems provide added value in terms of safety (greater anticipation time and larger safety gap), efficiency, and comfort (smoother maneuvers with minimum jerk and minimal reduction of speed).
- After extensive experimentation in a proving ground under standardized conditions, BRAVE's predictive systems have been proven to outperform the state-of-the-art vehicles tested at UTAC in the same use cases. As a matter of fact, UTAC, as an independent Euro NCAP tester, issued the following qualitative report comparing the performance of the BRAVE vehicle and other vehicles with similar capabilities tested on the same use cases [44]:
  - "The BRAVE vehicle tested scored full points on the accomplished tests. EuroNCAP protocols used to obtain scores were written to assess AEB systems.
  - Comparison between the BRAVE vehicle and other manufacturers' vehicles with Automated Driving (AD) functions:

- BRAVE vehicle is the best performing vehicle as far as relative distance with GVT is concerned for cut-in use cases.
- BRAVE vehicle was the only vehicle able to avoid any collision fully autonomously for the cut-in –1.5s TTC use case.
- BRAVE vehicle behaves more smoothly than other vehicles."

- BRAVE's predictive system overcomes humans' prediction ability on lane changes by 300 ms as well as the prediction ability of the baseline predictive system (based on Kalman filtering) by 800 ms. Thus, it demonstrates the added value of AI-based prediction systems as an asset to achieve higher comfort and safety in automated driving, especially in critical scenarios, such as lane changing and merging maneuvers on highways.
- BRAVE has proved the added value of prediction of intentions in pedestrian use cases by detecting in advance the pedestrian (dummy) face looking for eye contact with the driver. The combined use of anticipated gaze detection and the activation of the GRAIL interface leads to smooth behavior, thus contributing to an enhanced feeling of comfort and respect both for pedestrians and passengers of automated cars.
- Last-second reactions have been improved (shortened) to increase the ability to deal with the most challenging conditions in pedestrian use cases, both in braking and avoidance scenarios. Predictive systems are useful in scenarios with high visibility (a pedestrian approaching the curbstone in an open area while being fully visible from the car), but they provide no added value in extreme cases where the pedestrian is occluded (a pedestrian crossing the street while suddenly emerging from behind parked cars). In any case, safety was guaranteed, even in those challenging conditions.
- The current predictive systems provide no added value in the cyclist use case, given that the cyclist is detected from a far distance and no abrupt cyclist's maneuver or reaction is carried out. Although safety and comfort were guaranteed during the tests, more technologically advanced cyclist dummies that more accurately mimic real cyclists' behavior would be needed to make the most out of the potential that predictive systems can offer. These are some of the advanced dummy features that would be needed for such purpose: the cyclist raising his arm to signal a change of direction; the cyclist turning his head (as a clear indication of changing direction); the cyclist leaning to the left or right; and variable pedaling pace.
- Intersection use cases involving two vehicles have been deployed successfully, proving the added value of prediction in false positive cases. However, further research would be needed to deal with more complex use cases at intersections involving multiple vehicles.

Main Limitations

The main shortcomings identified in the current procedures for testing prediction systems are described next. As illustrated in Figure 16, it is important to note that these limitations are interrelated and not mutually exclusive.

Limited Fidelity

Standardized tests are easy to carry out, but they do not reflect the behavior of road agents in a realistic manner since the movements of the robotized dummies are too rigid and linear. Consequently, they are easy to detect and predict. The aspect of the dummies and the GVT is also standardized, and thus, it is repetitive. Similarly, the testing scenarios are totally open, without any objects in the background, making the test much easier from the perception point of view, whether based on vision, radar, or lidar. In addition, only one dummy (human or vehicle) is considered at a time during the tests, making the scenarios too simplistic compared to real driving situations (see Figure 17) and eliminating the possibility of including interactions between multiple agents.

Limited Variability

The tests are conducted under strict repetitive conditions. Thus, there is little variability in the testing conditions, which are limited to changes of velocity, TTC, and percentage of overlapping area at the point of impact. The potential variability of the appearance of human dummies (e.g., clothing, hair color, skin color, height, and so on) as well as of the dummy vehicle (e.g., color, size, and so on), or even of background conditions (e.g., cluttered background) is not sufficiently exploited. These strict repeatability requirements can lead to tailored solutions that AI systems can easily learn.

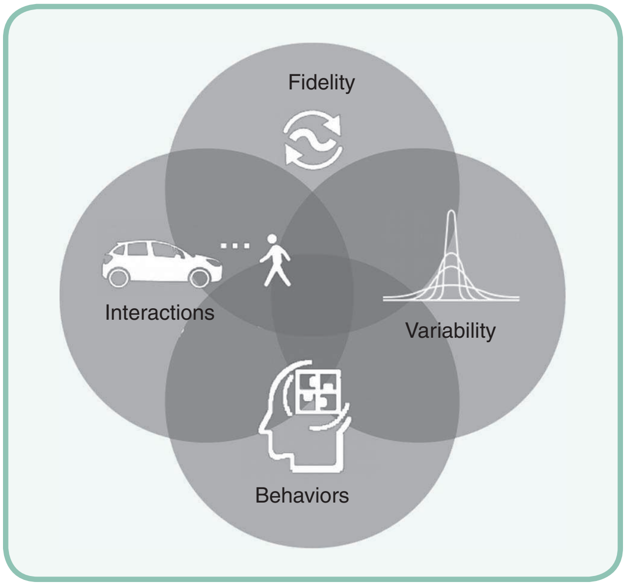

FIG 16 A Venn diagram illustrating the main identified weaknesses when testing predictive systems.

Limited Behaviors

The dummies execute standard, preprogrammed behaviors that are limited to very basic actions, such as crossing, not crossing, and changing lane. There are no scenarios representing multiagent interactions. A larger variation of behaviors, including interactions, is needed for testing advanced predictive systems.

The Lack of Real Behaviors

Most of the systems that have been tested in this project are "last resort" solutions (i.e., electronic stability program, electronic braking system, antilock braking system), meaning that the capability to provide mid- or long-term anticipation is not tested at all. In addition, the dummies do not perform any reaction as a consequence of the actions of the autonomous vehicle during the tests (e.g., the effect of external HMI systems cannot be assessed). Thus, it is not possible to test the effect of the evaluated systems on other agents' behavior. Advanced predictive systems should be tested on more realistic circumstances where real interactions between agents take place as a means to assess the real value that predictive capabilities can bring in this field.

The previously mentioned limitations impose constraints on the validation methods to assess the real potential of predictive systems in the context of automated driving. However, the results obtained suggest that, even with the current certification context, predictive systems provide substantial benefits in terms of anticipation, safety, and comfort. In the following text, we propose some potential actions intended to pave the way forward in the near future.

Recommendations for Future Actions

Improving Fidelity

Regarding the dummies, we have identified several possible actions. A first proposal is to have robotized dummies performing more realistic trajectories by adding probabilistic noise to the linear trajectories. Another possibility is to modify the baseline trajectories by making them a bit more erratic and, consequently, much less linear and more realistic. A second proposal is to develop dummies that can perform more realistic movements, such as pedestrians bending their upper bodies forward before starting to walk, pedestrians turning their torsos, cyclists turning their heads toward oncoming traffic, or vehicles with robotized turn signals. A third proposal is to include multiple dummies in the testing scenarios so that autonomous vehicles have to reason about several road users and their potential future movements and interactions. In terms of environments, any efforts to

FIG 17 An illustration of the fidelity gap between the test and real scenarios. (a) The test environment. (b) The view from the vehicle. (c) Examples of real environments.

provide more realistic environments (e.g., realistic urban scenarios, such as those used in the DARPA Urban Challenge [45]) would be beneficial to improve fidelity.

Improving Variability

The variability of the testing conditions can be largely increased by having the dummies performing different behaviors, apart from the standard crossing, not crossing, and so on. Examples of such behaviors are: partially changing trajectory (direction and orientation), stop and go, variable velocity, acceleration profiles, and so on. The use of dummies with different conditions of clothing, hair, and skin color can also be interesting. It is important to maintain a tradeoff between repeatability and variability, but the trend in new certification processes includes testing under real traffic conditions where repeatability cannot be guaranteed in any case.

Finally, false positive testing is also necessary for each use case, i.e., an aborted cut-in or an emerging pedestrian who ultimately does not enter the crossing zone. Predictive systems must be able to predict both positive and negative cases and act accordingly.

Adding Real Behaviors and Interactions

The only way to test real interactions between road users and automated vehicles in a safe manner is to have real road users in a simulated environment where they can safely interact with automated vehicles that will be equipped with adaptive motion planning and HMI strategies based on predictive features. The benefits of using simulated environments with real road users in the loop are twofold. On the one hand, all variables and conditions can be fully controlled during the execution of the tests; on the other hand, the reactions of road users can be accurately measured, providing the means to assess the effect that the autonomous vehicle's actions cause on other agents. This will definitely open the gate to the development of autonomous vehicles with real interacting capabilities that will mimic or even surpass human driving abilities.

Conclusion

Physical tests conducted in controlled environments on test tracks are mainly developed to evaluate last-second reactions and do not allow changes in the trajectory or speed of the VUT until the last moment, even if the vehicle has anticipated the situation correctly, which drastically reduces the ability of advanced predictive systems to improve safety and comfort. The conducted tests have proven that the developed advanced predictive systems have accomplished them increasing safety and comfort compared with basic systems, even with these limitations.

Current testing procedures present some limitations to assess the real potential of advanced predictive systems in real driving situations. These limitations are totally opened scenarios, interactions limited to a single element, the fixed and known appearance of the interacting element, strict repetitive conditions, the execution of preprogrammed trajectories, rigid body movements, and the impossibility of anticipating. Vision-based systems may be more undervalued than radar- and lidar-based systems because they perceive more limitations.

After analyzing the aforementioned limitations, the ability to measure the actual performance of predictive systems can be potentially increased by randomizing experiments and adding probabilistic noise to the dummies' linear trajectories. The development of more realistic dummies can increase reality in actions, such as pedestrians bending or turning their torsos, cyclists turning their heads, or vehicles with robotized turn signals. Several interacting dummies need to be included in the tests to create more complex but essentially similar scenarios.

Tests need to be improved in terms of variability by increasing the number of possible final situations, including different behaviors, such as fully and partially developed or even aborted maneuvers, to correctly evaluate the goodness of the advanced predictive systems. Along the same line, it is necessary to implement non-last-second tests to improve safety and comfort by anticipating oncoming events. Our future work is mainly focused on addressing the implementation of the proposals discussed in this article within the regulation and working groups of various institutions, such as the Euro NCAP (e.g., AEBS and automated driving, occupant status monitoring, HMI), European Commission General Safety Regulation 2, and UNECE Validation Method for Automated Driving).

Acknowledgments

This work was mainly funded by research grant BRAVE Project, Horizon 2020 (Contract 723021) and partially funded by Research Grants S2018/EMT-4362 SEGVAUTO 4.0-CM (Community of Madrid) and DPI2017-90035-R and PID2020-114924RB-I00 (Spanish Ministry of Science and Innovation). D. Fernández Llorca acknowledges funding from the HUMAINT project by the Directorate-General Joint Research Centre of the European Commission. We want to express our utmost gratitude to UTAC for providing the proving grounds and dummies and for the support in conducting the experiments.

About the Authors

Rubén Izquierdo Gonzalo earned his Ph.D. degree in information and communication technologies 2020 from the Universidad de Alcalá (UAH). He is currently an assistant professor at the Computer Engineering Department, UAH, Alcalá de Henares, Madrid, 28801, Spain. His research interests are focused on the prediction of vehicle behaviors but also in control algorithms for highly automated and cooperative vehicles. His work has developed a predictive adaptive cruise control and automatic emergency steering system for cut-in collision avoidance.

Carlota Salinas Maldonado earned her Ph.D. degree in engineering and automatics from the Universidad Complutense de Madrid in 2015. She is currently an assistant professor at the Computer Engineering Department, University of Alcalá, Alcalá de Henares, Madrid, 28801, Spain. Her research interests include autonomous vehicle navigation, data fusion systems, lidar, computer vision, and machine learning algorithms.

Javier Alonso Ruiz earned his Ph.D. degree in computer science from the Polytechnic University of Madrid in 2009. He is currently an assistant professor at the Computer Engineering Department, University of Alcalá (UAH), Alcalá de Henares, Madrid, 28801, Spain. His research interests are focused on autonomous vehicle navigation and control. He was awarded with the finalist prize for the Best Ph.D. Thesis in Intelligent Control 2010 by the Spanish Committee of Automatics. He succeeded in obtaining a Marie Curie Intra-European Fellowship to work on cognitive vehicles at the Karlsruher Institut für Technologie, Germany, and he received the Prize of the Social Council of the UAH to the University-Society Knowledge Transfer in 2018.

Ignacio Parra Alonso earned his Ph.D. degree in telecommunications engineering from the University of Alcalá (UAH), in 2005 and 2010, respectively. He is currently an associate professor with the Computer Engineering Department, UAH, Alcalá de Henares, 28801, Madrid, Spain. His research interests include intelligent transportation systems and computer vision. He received the Master Thesis Award in eSafety from the Ayuda al Automovilista. Lectureship at the Technical University of Madrid, Spain, in 2006.

David Fernández Llorca earned his Ph.D. degree in telecommunication engineering from the University of Alcalá (UAH) in 2008. He is currently a scientific officer at the Joint Research Centre, European Commission, Sevilla, 41092, Spain. He is also a full professor with the UAH, University of Alcalá (UAH), Alcalá de Henares, Madrid, 28801, Spain. He has authored more than 130 publications and more than 10 patents. He received the IEEE Intelligent Transportation Systems Society (ITSS) Young Researcher Award in 2018 and the IEEE ITSS Outstanding Application Award in 2013. He is the editor-in-chief of *IET Intelligent Transport Systems*. His current research interest includes trustworthy artificial intelligence for transportation, predictive perception for autonomous vehicles, human–vehicle interaction, end-user oriented autonomous vehicles, and assistive intelligent transportation systems. He is a Senior Member of IEEE.

Miguel Ángel Sotelo earned his Ph.D. degree in electrical engineering in 2001 from the University of Alcalá in 2001and his MBA degree from the European Business School in 2008. He is currently a full professor at the Department of Computer Engineering of the University of Alcalá, Alcalá de Henares, Madrid, 28801, Spain. His research interests include self-driving cars and predictive systems. He is author of more than 300 publications in journals, conferences, and book chapters. He has been recipient of the Best Research Award in the domain of automotive and vehicle applications in Spain in 2002 and 2009 and the 3M Foundation Awards in the category of eSafety in 2004 and 2009. He has served as project evaluator, rapporteur, and reviewer for the European Commission in the field of Information and Communication Technologies for Intelligent Vehicles and Cooperative Systems in the Sixth Framework Programme (EU) and the Seventh Framework Programme (EU). He is a member of the IEEE Intelligent Transportation Systems Society (ITSS) Board of Governors and Executive Committee. He served as the editor-in-chief of the *Intelligent Transportation Systems Society Newsletter* (2013), the editor-in-chief of *IEEE Intelligent Transportation Systems Magazine* (2014–2016), an associate editor of *IEEE Transactions on Intelligent Transportation Systems* (2008–2014), a member of the Steering Committee of the IEEE Transactions on Intelligent Vehicles (since 2015), and a member of the Editorial Board of *The Open Transportation Journal* (2006–2015). He served as the general chair of the 2012 IEEE Intelligent Vehicles Symposium that was held in Alcalá de Henares in June 2012. He was the recipient of the 2010 Outstanding Editorial Service Award for *IEEE Transactions on Intelligent Transportation Systems*, the IEEE ITSS Outstanding Application Award in 2013, and the Prize to the Best Team with Full Automation in Grand Cooperative Driving Challenge 2016. He was president of the IEEE ITSS (2018–2019).

References

[1] H. R. Martins, "Type approval homologation and self-certification," Ford Company, Detroit, MI, USA, Tech. Rep., 2010. [Online]. Available: https://www.researchgate.net/publication/311949142_Overview_of_Type_Approval_Homologation_and_Self-Certification

[2] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "On a formal model of safe and scalable self-driving cars," *CoRR*, 2017. [Online]. Available: http://arxiv.org/abs/1708.06374

[3] A. Lutz *et al.*, "Simulation methods supporting homologation of electronic stability control in vehicle variants," *Veh. Syst. Dynam.*, vol. 55, no. 10, pp. 1432–1497, 2017, doi: 10.1080/00423114.2017.1322705.

[4] R. Schram, "Euro NCAP's first step to assess automated driving systems," European New Car Assessment Programme, Euro NCAP, Leuven, Belgium, Paper Number 19-0292, Tech. Rep., 2019.

[5] D. Fernández Llorca and E. Gómez, "Trustworthy autonomous vehicles," Office of the European Union, Luxembourg, EUR 30942 EN, JRC127051, Tech. Rep., 2021.

[6] M. Edwards, M. Seidl, M. Tress, A. Pressley, and S. Mohan, "Study on the assessment and certification of automated vehicles," European Commission, Brussels, Belgium, Tech. Rep., 2016. [Online]. Available: https://circabc.europa.eu/sd/a/b6f6de76-184e-4967-93dd-9d7f1e1e3984/item%204-2017-01%20Commission%20study%20on%20vehicle%20certification.pdf

[7] "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," SAE International, Warrendale, PA, USA, J3016 202104, 2021.

[8] M. C. Galassi and A. Lagrange, "New approaches for automated vehicles certification: Part I—Current and upcoming methods for safety assessment," European Commission, Brussels, Belgium, EUR 30087 EN, JRC119345, Tech. Rep., 2020.

[9] M. Althoff and J. M. Dolan, "Online verification of automated road vehicles using reachability analysis," *IEEE Trans. Robot.*, vol. 30, no. 4, pp. 903–918, 2014, doi: 10.1109/TRO.2014.2312453.

[10] G. Baldini, "Testing and certification of automated vehicles (AV) including cybersecurity and artificial intelligence aspects," European Commission, Brussels, Belgium, EUR 30472 EN, JRC121631, Tech. Rep., 2020.

[11] "Future certification of automated/autonomous driving systems," UNECE WP29 GRVA, Geneva, Switzerland, Informal document GRVA-02-09, Tech. Rep., 2019.

[12] "Global vehicle target specification," Euro NCAP, Leuven, Belgium, Tech. Bull., 2018. [Online]. Available: https://cdn.euroncap.com/media/39159/tb-025-global-vehicle-target-specification-for-euro-ncap-v10.pdf

[13] "Articulated pedestrian target specifications," European Automobile Manufacturers Association, Tech. Rep., 2015.

[14] "Bicyclist target ACEA specifications," European Automobile Manufacturers Association, Brussels, Belgium, ACEA, Tech. Rep., 2018. [Onlilne]. Available: https://www.acea.auto/files/Bicyclist_target-ACEA_specifications.pdf

[15] "Uniform provisions concerning the approval of motor vehicles with regard to the Advanced Emergency Braking System (AEBS) for M1 and N1 vehicles," United Nations, Geneva, Switzerland, Addendum 151—UN Regulation No. 152, Tech. Rep., 2020.

[16] "Uniform provisions concerning the approval of vehicles with regard to Automated Lane Keeping Systems," United Nations, Geneva, Switzerland, Addendum 156—UN Regulation No. 157, Tech. Rep., 2021.

[17] "Honda launches next generation Honda sensing elite safety system with Level 3 automated driving features in Japan," Honda, 2021. [Online]. Available: https://hondanews.com/en-US/honda-corporate/releases/release-e86048ba0d6e80b260e72d443f0e4d47-honda-launches-next-generation-honda-sensing-elite-safety-system-with-level-3-automated-driving-features-in-japan

[18] "First internationally valid system approval for conditionally automated driving," Mercedes-Benz, Stuttgart, Germany, 2021. [Online]. Available: https://group.mercedes-benz.com/innovation/product-innovation/autonomous-driving/system-approval-for-conditionally-automated-driving.html

[19] "Project for the establishment of generally accepted quality criteria, tools and methods as well as scenarios and situations," PEGASUS Project, 2019. [Online]. Available: https://www.pegasusprojekt.de/en/about-PEGASUS

[20] "European initiative to enable validation for highly automated safe and secure systems." ENABLE-S3 Project, 2019. [Online]. Available: https://enable-s3.eu/

[21] "DRIVINGCUBE—A new way to speed up the validation and approval process of ADAS/AD systems," AVL, Graz, Austria, 2021. [Online]. Available: https://www.avl.com/-/avl-drivingcube

[22] E. Thorn, S. C. Kimmel, and M. Chaka, "A framework for automated driving system testable cases and scenarios," U. S. Dept. of Transportation, National Highway Traffic Safety, Tech. Rep., 2018.

[23] A. Jain, L. D. Pero, H. Grimmett, and P. Ondruska, "Autonomy 2.0: Why is self-driving always 5 years away?" *CoRR*, vol. abs/2107.08142, 2021. [Online]. Available: https://arxiv.org/abs/2107.08142

[24] S.-A. Kaye, K. Somoray, D. Rodwell, and I. Lewis, "Users' acceptance of private automated vehicles: A systematic review and meta-analysis," *J. Safety Res.*, vol. 79, pp. 352–367, Dec. 2021, doi: 10.1016/j.jsr.2021.10.002.

[25] M. Elbanhawi, M. Simic, and R. Jazar, "In the passenger seat: Investigating ride comfort measures in autonomous cars," *IEEE Intell. Transp. Syst. Mag.*, vol. 7, no. 3, pp. 4–17, 2015, doi: 10.1109/MITS.2015.2405571.

[26] "BRidging gaps for the adoption of Automated Vehicles." Accessed: May 11, 2022. [Online]. Available: http://www.brave-project.eu/

[27] "Test protocol—AEB car-to-car systems," Euro NCAP, Leuven, Belgium, 2019.

[28] "Test protocol—AEB VRU systems," European New Car Assessment Programme, Tech. Rep., 2020.

[29] I. Parra Alonso, R. Izquierdo Gonzalo, J. Alonso, Á. García-Morcillo, D. Fernández-Llorca, and M. Á. Sotelo, "The experience of drivertive-driverless cooperative vehicle-team in the 2016 GCDC," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 4, pp. 1322–1334, 2018, doi: 10.1109/TITS.2017.2749963.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[31] R. Izquierdo, A. Quintanar, I. Parra, D. Fernández-Llorca, and M. A. Sotelo, "Experimental validation of lane-change intention prediction methodologies based on CNN and LSTM," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, 2019, pp. 3657–3662, doi: 10.1109/ITSC.2019.8917331.

[32] R. Izquierdo, A. Quintanar, I. Parra, D. Fernández-Llorca, and M. A. Sotelo, "Vehicle trajectory prediction in crowded highway scenarios using bird eye view representations and CNNs," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, 2020, pp. 1–6, doi: 10.1109/ITSC45102.2020.9294732.

[33] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[34] R. Izquierdo *et al.*, "Vehicle lane change prediction on highways using efficient environment representation and deep learning," *IEEE Access*, vol. 9, pp. 119,454–119,465, Aug. 2021, doi: 10.1109/ACCESS.2021.3106692.

[35] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "Openpose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, pp. 172–186, Jan. 2021.

[36] R. Quintero Mínguez, I. Parra Alonso, D. Fernández-Llorca, and M. A. Sotelo, "Pedestrian path, pose, and intention prediction through gaussian process dynamical models and pedestrian activity recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1803–1814, 2019, doi: 10.1109/TITS.2018.2836305.

[37] J. Lorenzo, I. Parra, F. Wirth, C. Stiller, D. F. Llorca, and M. A. Sotelo, "RNN-based pedestrian crossing prediction using activity and pose-related features," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2020, pp. 1801–1806, doi: 10.1109/IV47402.2020.9304652.

[38] J. Lorenzo *et al.*, "Capformer: Pedestrian crossing action prediction using transformer," *Sensors*, vol. 21, no. 17, p. 5694, 2021, doi: 10.3390/s21175694.

[39] S. Zernetsch, S. Kohnen, M. Goldhammer, K. Doll, and B. Sick, "Trajectory prediction of cyclists using a physical model and an artificial neural network," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2016, pp. 833–838, doi: 10.1109/IVS.2016.7535484.

[40] J. F. P. Kooij, E. A. I. P. F. Flohr, and D. M. Gavrila, "Context-based path prediction for targets with switching dynamics," *Int. J. Comput. Vis.*, vol. 127, no. 3, pp. 239–262, 2019, doi: 10.1007/s11263-018-1104-4.

[41] "Proving grounds," Union Technique de l'Automobile et du Cycle (UTAC), 2022. [Online]. Available: https://www.utacceram.com/proving-grounds

[42] M. A. Sotelo, "GRAIL—An intelligent interface between vehicles and vulnerable road users," Apr. 24, 2018. [Online]. Available: http://www.brave-project.eu/index.php/2018/04/24/grail-an-intelligent-interface-between-vehicles-and-vulnerable-road-users/

[43] A. Quintanar, R. Izquierdo, I. Parra, D. Fernández-Llorca, and M. A. Sotelo, "The prevention challenge: How good are humans predicting lane changes?" in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2020, pp. 45–50, doi: 10.1109/IV47402.2020.9304640.

[44] "D4.6. Assessment of the effectiveness of vehicles and VRUs prediction of intentions," BRAVE, 2021. [Online]. Available: http://www.brave-project.eu/wp-content/uploads/2021/01/BRAVE-D4.6-v1.0.pdf

[45] M. Buehler, K. Iagnemma, and S. Singh, Eds., *The DARPA Urban Challenge*. Berlin: Springer-Verlag, 2009, vol. 56.

ITS