VERIFICATION AND VALIDATION OF MACHINE LEARNING APPLICATIONS IN ADVANCED DRIVING ASSISTANCE SYSTEMS AND AUTOMATED DRIVING SYSTEMS

Chung-Jen Hsu
Bowhead Mission Solutions
USA

Paper Number 23-0048

ABSTRACT

The verification and validation processes of machine learning applications in advanced driving assistance systems or automatic driving systems are presented, and the processes are implemented by using the forward collision warning of pedestrian automatic emergency braking. Supervised learning is one of the machine learning branches using image datasets to train the deep neural network for detecting or identifying the target object or scenario in a vision-based application. The verification process consists of specifying the requirements of a safety functionality, identifying the target objects in the Operation Design Domain (ODD) and pre-crash scenarios, and evaluating the quality and quantity of images based on safety requirements, also the coverage of ODD and pre-crash scenarios. The validation process consists of designing test procedures based on the specified ODD and pre-crash scenarios, conducting a sufficient number of tests, recording the test results, and evaluating the test results based on specified metrics. Eight published pedestrian datasets from 2010 to 2020 are reviewed. Three datasets contain the raining condition, but no dataset had images collected during snowing days. Fog or smoke images are not available in all datasets, and the headlight condition is not addressed in all datasets. The 3 datasets containing pedestrians in the nighttime did not label the vehicle's headlight status as low or high beam. All reviewed datasets had no annotations of pre-crash scenarios that the subject vehicle is maneuvering or not. The validation of pedestrian detection uses the activation of forward collision warning as the evaluation metric. Eleven vehicles were tested in 4 pre-crash scenarios with different pedestrian orientations and speeds: the test pedestrian crossing from the nearside, crossing from the offside, stationary facing away, and walking away in front of the vehicle. The vehicle speed under test is 40 kph and the test pedestrian's speed is 5 or 8 kph. The light conditions are daytime, nighttime with low beam, and nighttime with high beam without streetlighting in a test track. The statistical test results show that some vehicles under test behave inconsistently when the test pedestrian is crossing or not crossing. Test results in the nighttime with high beam are similar to that of the daytime; however, the test results in the nighttime show significant variations compared with that of daytime. No trend or similarity can be found among all vehicles under test, the same vehicle may behave inconsistently under different light conditions and pedestrian orientations. Also, the pedestrian detection time is longer when the test pedestrian is not crossing for some vehicles. The vision-based machine learning application for the vehicle safety functionality reveals the underlying uncertainty of a deep neural network, and it results in the inconsistent performance in differentiated ODD conditions and pre-scenarios.

INTRODUCTION

Machine learning (ML) techniques have been widely used in the safety functions and vehicle control of Automated Driving Systems (ADS). ADS perform object and event detection and response consisting of monitoring the driving environment and execute appropriate responses to objects and events. The driving environment can also be referred to as Operational Design Domain (ODD) specifying the operating domains or conditions in which Advanced Driving Assistance Systems (ADAS) or ADS are designed to function safely. Object and event detections can be achieved by using cameras, radars or lidars to retrieve images for further processing. Supervised ML models can be used to identify vehicles, pedestrians, and other objects such as traffic signs, obstacles, and lane markings. Supervised learning is one of the ML paradigms being extensively applied for detecting objects through the training of a Deep Neural Network (DNN) with sufficient images of the target objects [1]. The detection accuracy depends on the quality and quantity of the training images and DNN modeling. Collecting and labeling images containing target objects are time consuming, and this effort is proportional to the numbers of object categories. Vehicles, pedestrians, cyclist, traffic signals, signs, lane markings, etc. are some objects to be identified in ADAS/ADS applications. If training images are not labeled correctly or they are unable to cover most of the target object categories, then the detection accuracy will not be sufficient for the safety requirements of ADAS/ADS even though the DNN modeling is impeccable. Scene semantic segmentation is to identify multiple objects and segment them as a relational group revealing a specific scenario in an image [2]. Pre-crash scenarios are crucial to safety applications to recognize the scene semantics in different ODD conditions. If an ADAS/ADS application can recognize driving scene semantics using a DNN, then the categories of scenes and related characteristics should also be examined for verifying the safety limitations of ADAS/ADS in the same manner as verifying object detections. The Verification and Validation (V&V) of ML applications in ADAS/ADS are not addressed comprehensively by using conventional engineering approaches as specified in automotive standards including ISO 26262 and 21448 [3]. One of the V&V challenges of ADAS/ADS safety applications is lack of transparency in ML development processes including the DNN modeling and training data. Due to the complex and proprietary characteristics of DNN modeling, it is difficult to verify its robustness by reviewing DNN's structures and algorithms; however, the validation can be achieved by measuring the level of detection accuracy. This study intends to tackle the V&V of ML applications in ADAS/ADS safety functionalities by verifying the training datasets and validating the performance from the safety perspective.

RELATED WORKS

Borg et al. [4] conducted a review of V&V for ML in the automotive industry. This study found a gap between current safety standards and contemporary ML-based safety-critical systems from the V&V perspective. Potential methodologies of V&V in ML applications can be categorized as: formal methods, control theory, probabilistic methods, process guidelines, and simulated test cases. The challenges are no clear certification processes of safety-critical systems with DNNs, a lack of transparency in ML processes, and concerns of the robustness and state-space explosion. The challenges essentially originate from the workflow of supervised ML that training data are fed into a DNN, and test data are used to validate whether the design requirements are fulfilled [5]. The insufficient robustness and out of scope state-spaces are caused by the lack of comprehensive coverage in training data or the design defect in a DNN. Depending on the design purposes of ML applications in ADAS/ADS, this drawback poses safety risks on scene identification, motion planning, decision making, vehicle control, or communication [6]. The first step of scene identification is to perceive the objects of interest that might result in a safety risk. The perception tasks of ADAS/ADS are implemented by using cameras, lidars, or other sensors. Sensors provide images of vehicles, pedestrians, cyclist, lane markings, etc. that had been collected and labeled as training datasets for developing ADAS/ADS applications. Yurtsever et al. [7] surveyed 18 driving datasets being used for ADS developments; however, only 6 of them covers various weather conditions in the daytime and nighttime. The insufficiency of ODD coverage for training datasets emerges as a safety risk. Burton [8] proposed to set criteria of selecting training data based on the semantic analysis of triggering conditions (pre-crash scenarios) or other causes of errors for safety assurance. Schwalbe and Schels [9] conducted a survey on methods for the safety assurance of ML-based systems and summarized that data representativity requirements including the scenario coverage, input space ontology, and experience collection can be used to validate ADAS/ADS functionalities. Willers et al. [10] proposed mitigation approaches to safety concerns including the data distribution's approximation of real world, data shifting over time, inadequate separation of tests and training data, and dependence on the labeling quality. Other safety concerns related to the DNN modeling are the brittleness of DNNs, unreliable output confidence information, unknown behavior in rare critical situations, and incomprehensible behavior. The mitigation approaches to addressing data concerns are the sensible data acquisition strategy, iterative analysis of test results, data labeling guidelines, continuous learning and updating, and data partitioning guidelines. Cheng et al. [11] measured the robustness, interpretability, completeness, and correctness of DNNs by metrics including the scenario coverage, neuron activations, neuron activation pattern, adversarial confidence loss, scenario-based performance degradation, interpretation precision, occlusion sensitivity covering, and weighted accuracy/confusion. Calculations of these metrics require analyses of DNNs attributes and corresponding images. Amershi et al. [12] also stated that ML components are more difficult to handle as distinct modules than traditional software components. ML models may be entangled with data in complex ways and experience non-monotonic erroneous behaviors. The validation of ADAS/ADS applications can be achieved by various testing methodologies to address issues of residual risks, including the pre-deployment road tests, closed course testing, full/simplified vehicle environment simulations, and subsystem simulations [13]. Residual risks are unexpected scenarios/environment, unexpected human driver behavior, degraded infrastructure, and road hazards. A direct measurement of the failure rate remains a viable approach to validate the ML applications in ADAS/ADS with the consideration of residual risks [8].

METHODS

Supervised learning is a paradigm of utilizing a large and representative set of labeled data to train a ML model. The training dataset is the crucial factor of determining the accuracy of object detection in ADAS/ADS applications. A vision-based system requires objects of interest in the training dataset. The rationale is without the objects of interest in the training dataset the probability of detecting the objects of interest can be close to zero, but not zero for false positives may exist in DNNs. The best practice of improving the detection accuracy is to provide high quality images in the training dataset and develop a decent DNN that can achieve a high detection rate. To ensure the detection accuracy in a vision-based ML application, the first step is to verify the training dataset that should have the required quality and quantity in the desired ODD conditions and pre-crash scenarios. The verification is the process of evaluating whether the training dataset meets the safety requirements.

Verification

The process of verifying a training dataset is shown in Figure 1. The first step is specifying the requirements of an ADAS/ADS safety functionality. This is comparable with the specification of software safety requirements in the design phase of software development as defined in ISO 26262 [14]. The objects in potential risks of collision need to be specified based on the safety requirements. Most common objects in the driving ODD are vehicles, pedestrians, motorcycle, cyclists, and other objects that may be struck by the subject vehicle. Table 1 lists the top-level categories of ODD classifications [15]. The images of a training dataset can be categorized according to physical infrastructure, operational constraints, objects, and environmental conditions. In addition, the semantics of images can be categorized based on the pre-crash scenario groups including control loss, road departure, animal, pedestrian, cyclist, lane change, opposite direction, rear-end, and crossing paths as listed in Table 2 [16]. Each image can be labeled according to the categories of ODD and pre-crash scenarios. Image labeling is a labor-intensive task that most existing ML datasets are labeled in the object level. A semantic level labeling task demands more human endeavors, so the automation is desirable for mitigating the cost and time of labor. Recent research of ADS started to work on the semantic scene identification [2, 17, 18, 19], these techniques can be applied to the labeling of ODD and pre-crash scenarios [20, 21]. After categorizing and calculating the quantity of images in categories of ODD and pre-crash scenarios, the distribution of images in each category can be reviewed and a reasonable inference can be made. For example, if no nighttime illumination of pedestrian images is available in the training dataset, then the detection rate of nighttime pedestrians will be low most likely. The number of images in a specified category can be an evaluation metric. Also, the target object's distribution in ODD and pre-crash categories indicates a sensible expectation of detection rates in those categories. High density categories may have a better detection rate; on the contrary, low density or no coverage categories may have a low or even zero detection rate. This information can also be the context of testing ODD conditions in the validation process. For example, the edge testing cases can be designed based on the rare conditions in the training dataset. The distribution of images in pre-crash scenarios is useful for providing the potential risk assessment across the coverage map. Different pre-crash scenarios represent differentiated viewing angles of the target object profiles. The vehicle profiles are different in the lane change, opposite direction, rear end, and crossing path scenarios. A pedestrian profile is different when they are crossing or not crossing (facing) a roadway. Lastly, the age of a training dataset may be a safety concern. When a dataset only contains outdated vehicles that may result in some new vehicles undetected.

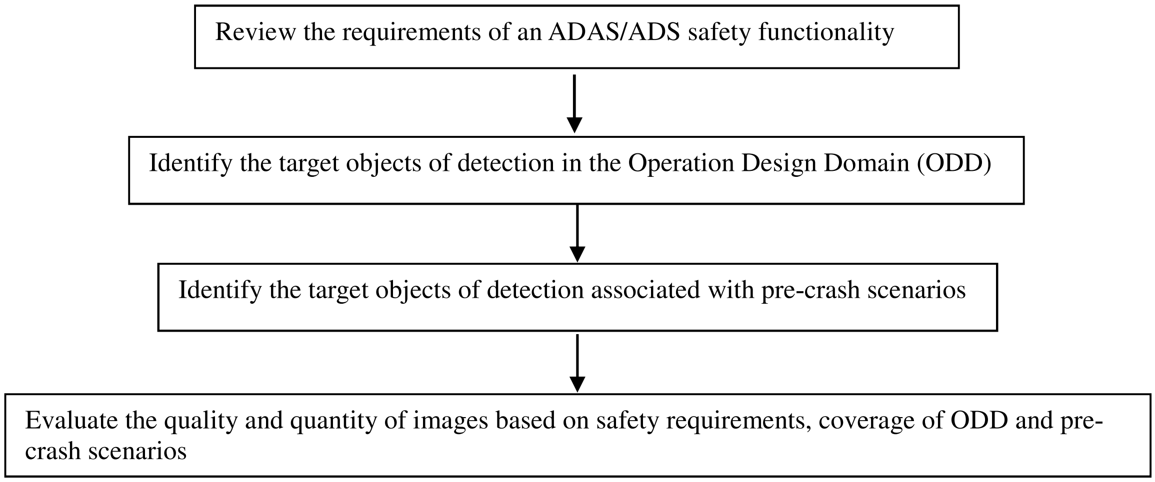

Figure 1. The process flow of verifying a training dataset.

Table 1. ODD classification with top-level categories

ODD Element	**Conditions**	**Categories**
Physical Infrastructure	• Roadway Types • Roadway Surfaces • Roadway Edges • Roadway Geometry	• Divided/undivided highway, arterial, urban, rural, parking, bridges, multi-lane/single lane, managed lanes (HOV, reversible lanes), on-off ramps, one-way, private roads, intersections • Asphalt, concrete, unpaved • Lane markers, temporarily lane markers, shoulder, barriers, curb • Horizontal/vertical alignment (curves, hills), superelevation, lane width
Operational Constraints	• Speed Limit • Traffic Conditions	• High/low • Traffic density, others (emergency vehicles, construction, closed road, special event)
Objects	• Signage • Roadway Users • Obstacles/Objects	• Signs (stop, yield, pedestrian, railroad, school zone, etc.), traffic signals, crosswalks, railroad crossing, stopped buses, construction signage • Vehicle types (cars, light trucks, large trucks, buses, motorcycles,), stopped vehicles, pedestrians, cyclists • Animals, debris
Environmental Conditions	• Weather • Weather-induced Roadway Conditions • Particulate Matter • Illumination	• Precipitation, wind, snow, temperature • Standing water, flooded, icy, snow • Fog, smoke, smog, dust/dirt • Dark, streetlights, dawn/dusk, low sun angle, day light, headlights (low/high beam), oncoming vehicle lights

Table 2. Pre-crash scenarios and groups

Scenario Group	Pre-Crash Scenarios	Subject Vehicle Maneuver
Control Loss	• Control loss/maneuver • Control loss/no maneuver	• Maneuver: performing a maneuver (e.g., passing, turning, changing lanes) • No maneuver: driving straight or negotiating a curve
Road Departure	• Road edge departure/maneuver • Road edge departure/no maneuver	
Animal	• Animal/maneuver • Animal/no maneuver	
Pedestrian	• Pedestrian/maneuver • Pedestrian/no maneuver	
Cyclist	• Cyclist/maneuver • Cyclist/no maneuver	
Lane Change	• Turning/same direction • Parking/same direction • Changing lanes/same direction • Drifting/same direction	• Turn and cut across the path of another vehicle initially traveling in the same direction • Enter or leave a parked position and collide with another vehicle • Change and encroach into another lane other vehicle traveling in the same direction • Drift into an adjacent lane other vehicle traveling in the same direction
Opposite Direction	• Opposite direction/maneuver • Opposite direction/no maneuver	• Make a maneuver (e.g., passing) and encroach into another vehicle traveling in the opposite direction • Drift and encroach into another vehicle traveling in the opposite direction
Rear-End	• Rear-end/striking maneuver • Rear-end/Lead Vehicle Accelerating • Rear-end/Lead Vehicle Moving • Rear-end/Lead Vehicle Decelerating • Rear-end/Lead Vehicle Stopped	• Change lanes or pass another vehicle and closes in on a vehicle ahead in the same lane • Close in on an accelerating lead vehicle ahead in the same lane • Close in on a moving vehicle ahead in the same lane • Close in on a decelerating lead vehicle ahead in the same lane • Close in on a stopped lead vehicle ahead in the same lane
Crossing Paths	• Right turn into path • Right turn across path • Straight crossing paths • Left turn across path, lateral direction • Left turn into path • Left turn across path, opposite direction	• Turn right and into the same direction of another vehicle crossing from a lateral direction • Turn right and into the opposite direction of another vehicle crossing from a lateral direction • Go straight and collide with another straight crossing vehicle from a lateral direction • Turn left and cross the path of another vehicle traveling in the opposite direction from a lateral direction (left) • Turn left into the path of another vehicle traveling in the same direction from a lateral direction (right) • Turn left and cross the path of another vehicle traveling in the opposite direction

Validation
Validation is the process of evaluating the degree to which a ML model/application and its data can provide an accurate result of the intended uses. Essentially, the validation of an ADAS/ADS application can be implemented after the verification of its training dataset that reveals the coverage and distribution across the spectrum of ODD and pre-crash scenarios. Depending on the design specification of a safety function, the validation tests can be conducted focusing on the selected ODD and pre-crash scenarios. Also, validation test procedures can be designed based on the historical crash data. High crash frequency scenarios may be tested with a higher priority and number of test runs. Well-design test procedures should be able to address safety concerns including rare critical situations, unreliable confidence information of DNN output, and brittleness of DNNs [10]. An ADAS/ADS safety functionality consists of software and hardware working together to achieve the goal of crash avoidance. The DNN is a part of software processing images from sensors (cameras) and generates the detection results. The uncertainty of a DNN output and the risk of hardware glitches result in the safety performance of a vehicle under test. Broken sensors or alarming devices result in no alarm that can be easily distinguish from software malfunctions in a few test runs. A vehicle-level test is feasible to validate the DNN of a safety application excluding the hardware failure. Safety metrics for evaluating the DNN performance also need to include the response time in addition to the pass/fail metric. Sufficient test runs are needed to collect data for calculating the reaction time and figuring out the boundary or capability of the DNN under test. The flowchart of validating a ML safety application is shown in Figure 2.

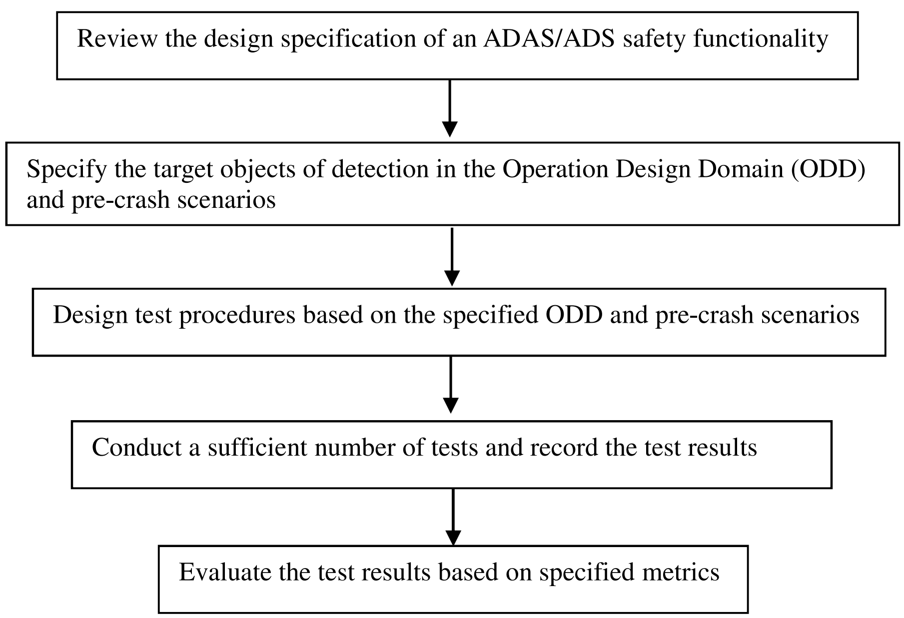

Figure 2. The process flow of validating an ADAS/ADS functionality with ML applications.

VERIFICATION OF TRAINING DATA

In consideration of selecting an ADAS/ADS safety application for the proposed methodology, the pedestrian crash avoidance is appropriate for a sensible reason. Pedestrian detection approaches are feature-based or ML-based or hybrid using different image processing algorithms [22, 23, 24]. Vehicle detections may use sensors such as lidar and radar along with cameras to provide inputs to the detection software, but the pedestrian detection mostly relies on cameras. The rationale is when the training dataset containing pedestrians in the specified ODD conditions and scenarios, the DNN may be able to detect the pedestrian at risk successfully. By inspecting the training dataset in line with the proposed verification process, the boundaries or limitations of pedestrian detection can be identified.

Safety Functionality Requirements
Pedestrian Automatic Emergency Braking (PAEB) is a safety function of ADAS. A PAEB system consists of cameras, Forward Collision Warning (FCW), and last moment automatic braking to prevent a collision with

pedestrians. FCW can be used as an indicator of whether a pedestrian at risk is detected or not. FCW is designed to warn the driver to maneuver or brake as early as possible.

Detection Object in Operation Design Domain

The pedestrian detection is to identify humans in the environment where the subject vehicle is traveling with potential collision risks. After reviewing the ODD conditions as listed in Table 1, roadway types, weather, particulate matter, and illumination are directly related to the performance of pedestrian detections. Most pedestrians appear at intersections, arterials, and parking lots in urban areas, less may be seen on all types of roadways in rural areas. Verifying the roadway distribution of pedestrian images can be beneficial to understating the background of the training dataset. Although the pedestrian detection capability and performance may not be correlated to the background of road types, such information can be used for the design of edge test scenarios. For example, a pedestrian is walking on the freeway shoulder.

Detection Objects in Pre-crash Scenarios

Two pre-crash scenarios of pedestrians are considered in Table 2 that the subject vehicle is maneuvering or not. When a vehicle is making a turn at an intersection where pedestrians are crossing, the viewing angle from the subject vehicle to a crossing pedestrian is changing in the process of turning. Ideally, the training dataset is expected to contain a variety of pedestrian profiles from various viewing angles.

Verification of Training Datasets

Eight pedestrian datasets are reviewed as listed in Table 3. The publication year, number of images, number of pedestrians, image resolution, pedestrian annotation (labeling), camera setup, and data collection areas reveal the background information of each dataset. The recent published datasets are reviewed since 2010 for aged pedestrian datasets may not have the sufficient quantity, resolution, and annotation to be used for training recent pedestrian detectors. One dataset labeled pedestrian images as the full, part, or just head of a pedestrian depending on the level of occlusion. Most datasets had a camera installed on the vehicle recording videos of pedestrians, two datasets collected pedestrians or human images from internet sources. The data collection area provides the information of where the pedestrian images were obtained.

Table 3. Dataset characteristics

Dataset	Year	# Image	#Pedestrian	Resolution	Annotation	Setup	Area
Caltech [25]	2010	250k	289k	640*480	Full, body	Vehicle	LA metropolitan
KITTI [26]	2012	15k	9k	1240*376	Full	Vehicle	Mid-size city, rural areas
CityPersons [27]	2017	5k	35k	2048*1024	Full, body	Vehicle	27 cities, Germany
CrowdHuman [28]	2018	24k	552k	-	Full, body, head	Internet images	40 cities worldwide
NightOwls [29]	2018	281k	56k	1024*640	Full	Vehicle	7 cities, 3 countries in Europe
EuroCity [30]	2019	47k	219k	1920*1024	Full	Vehicle	31 cities, 12 countries in Europe
TJU-DHD [31]	2020	75k	373k	1624*1200 2560*1440	Full, body	Vehicle, phone	Road, off road (campus)
WiderPerson [32]	2020	13k	39k	1400*800	Full	Internet images	

Table 4 lists the verification result of pedestrian datasets based on the ODD conditions and pre-crash scenarios. Ideally, the dataset verification process should have a labeling tool that is able to identify the ODD's condition/category and pre-crash scenario. However, such a tool is not available currently and the manual labeling of thousands of images is a huge task. The developers of ML pedestrian detection algorithms are expected to verify their training datasets with the consideration of ODD conditions and pre-crash scenarios. The review of datasets in this paper is based on the revealed information from the dataset publications. Three datasets contain raining weather conditions, but no dataset mentioned images collected during snowing days. Snow, fog, or smoke images are not available in all datasets, for such conditions may be rare during the data collection period and were not mentioned in the publications. The headlight condition is not addressed in all datasets, the 3 datasets containing pedestrians at night did not annotate the vehicle's headlights as low or high beam. When the data collection vehicle was traveling at night, it would be a reasonable inference that both low and high beams had been used. As for the pre-crash scenarios as the subject vehicle is maneuvering or not, it is most likely the vehicle had been making turns, changing lanes, and keeping straight with one or more pedestrians ahead in the period of data collection.

Table 4. Review of datasets based on ODD and pre-crash scenarios

Condition	**Category**	**Dataset**							
		Caltech	KITTI	CityPersons	Crowd Human	Night Owls	Euro City	TJU-DHD	Wider Person
Roadway	Intersection	✓	✓	✓	✓	✓	✓	✓	✓
	Arterial	✓	✓	✓	✓	✓	✓	✓	✓
Weather	Sunny/cloudy	✓	✓	✓	✓	✓	✓	✓	✓
	Rain					✓	✓	✓	
	Snow								
Particulate Matter	Fog/Smoke/Dust								
Illumination	Day	✓	✓	✓	✓		✓	✓	✓
	Night streetlight					✓	✓	✓	
	Headlight low beam					+	+	+	
	Headlight high beam					+	+	+	
Maneuver	Turning/changing lane	✪	✪	✪		✪	✪	✪	
Not Maneuver	Straight	✪	✪	✪		✪	✪	✪	

+: The dataset may contain low or high beam headlight images, but they are not identified specifically.

✪: The data collection vehicle should have both pre-crash scenarios in the data collection process.

VALIDATION OF SAFETY APPLICATIONS

The validation process as elucidated in Figure 2 is implemented in the following. The ML safety application is PAEB, and FCW uses the pedestrian detection to trigger an alarm. Vehicle level tests of PAEB were conducted in 2020 under the supervision of National Highway Traffic Safety Administration (NHTSA). Eleven vehicles equipped with FCW in PAEB of model year 2020 from 10 manufacturers including 4 sedans, 5 SUVs, 1 minivan, and 1 pickup truck were tested. The validation of pedestrian detection is to test whether FCW is activated in the selected ODD conditions and pre-crash scenarios.

Design Specifications of FCW in PAEB

FCW's operational speed ranges and limitations of 11 tested vehicles are collected from owner manuals and listed in Table 5. Eight vehicles' operational speeds are not higher than 50 mph, 3 of them are higher than 50 mph. The lower end of operational speeds is from 3 to 7 mph. The negative factors of vision-based pedestrian detection algorithms are summarized from the operational limitations listed in the owner manuals. Just like the limitations of human eyes, extreme light, air, and inclement weather conditions hinder the capability of pedestrian detections. In addition, the shape, movement speed, color, posture, and clothing of pedestrians are sensitive to detection results. Occlusions or carrying objects are negative factors to the detection accuracy, and the detectable pedestrian height is from 1 to 2 meter for some vehicles. Although it is not mentioned in all owner manuals, the ideal detection can only be achieved in the straight and flat road alignments.

Table 5. FCW operation speed range and limitations of tested vehicles

Vehicle	Operation Speed (mph) Low	High	Limitations
1	6	50	Curves, heavy fog rain, snow, dark, occlusion, glare, light variation/reflections
2	6	50	Curves, heavy fog rain, snow, dark, occlusion, glare, light variation/reflections
3	3	75	Not available
4	3	62	Height:1-2 m, groups, occlusion, unusual shape, movement (running)
5	5	45	Not walking upright, sudden appearance, small, clothes blend into background, too bright or dark, inclement weather
6	3	37	Less than 1m, carrying luggage, severe weather
7	4	43	Up to 50 mph for moving pedestrians, 43 mph for stationary pedestrians, snow, rain, fog, glare, sudden appearance, occlusion, blend into background, special clothing or object, tight curve
8	6	37	Small children, pedestrians on wheelchair/skateboard, not upright, darkness, strong light caused pedestrian in shadow, sudden change in brightness, occlusion, carrying luggage
9	7	100	1-2 m, in a group, next to obstacle, using umbrella, similar clothing color to background, carrying luggage, not upright, dark, sudden appearance, inclement weather, strong light from the front, dust, smoke, steam, steep up/down hill, darkness
10	7	50	1-2 m, abrupt appearance, not directly in front, near obstacle, occlusion, strong light, same color in the surrounding, oversize clothing, moving fast, not upright, pushing an object, inclement weather, steam, smoke, darkness, abrupt changing brightness, curve
11	3	50	Shorter than 0.8m, clothing covering body contour, poor background contrast, carrying a large object

ODD and Pre-crash Scenarios of Pedestrian Detection Tests

The test ODD conditions and pre-crash scenarios can be specified after reviewing the limitations of vehicles under test. A test process may start from easy or most common ODD conditions and pre-crash scenarios, then increase the difficulty level gradually depending on the testing requirements. Light conditions and pedestrian movement orientations are two major test variables including day and night under low and high beams, crossing from the nearside or offside, or walking toward/backward in front of the vehicle. The selected NHTSA test scenarios [33] are:

- S1b: the vehicle encounters a crossing adult from the nearside (closest to the curb)
- S1e: the vehicle encounters a crossing adult running from the offside (closest to the center of the road)
- S4a: the vehicle encounters a stationary adult on the nearside of the road facing away
- S4c: the vehicle encounters an adult on the nearside of the road walking in the same direction

These 4 test scenarios have both day and night test results for the validation.

Test Procedures

A test pedestrian mannequin with swinging arms is used to simulate an adult pedestrian whose speed and direction can be controlled. The vehicle under test is driven at the specified speed approaching the test mannequin moving in the orientation as defined in the test scenarios. The test site has no overhead signs or other significant structures to cause occlusions. Each trial was conducted without other vehicles, obstructions, or stationary objects within one lane width on either side of the driving lane. All tests are conducted without inclement weather conditions such as fog, smoke, or ash. Also, the daytime tests were conducted with good visibility without direct sunlight or glare. The nighttime tests were conducted without streetlighting. The speeds and orientations of the test vehicle and pedestrian are listed in Table 6.

Table 6. Vehicle and pedestrian speeds and orientations under tests

Test Scenario	Vehicle Speed (kph)	Pedestrian Speed (kph)	Pedestrian Orientation
S1b	40	5	Crossing nearside
S1e	40	8	Crossing offside
S4a	40	0	Stationary facing away
S4c	40	5	Walking away

Evaluation of Test Results

The accuracy of pedestrian detections depends on various limitation factors as elucidated in Table 5. The test scenarios are designed to exclude most unfavorable factors and focus on the most common ODD conditions and pre-crash scenarios. The pedestrian profile or shape can be different from the camera's viewing angle in the daytime or nighttime. The low or high beam light is the major source of light on the test pedestrian. The performance of pedestrian detections can be validated by comparing the detection results. The statistical test — Fisher's exact test [34] is used to evaluate whether the pedestrian detection algorithm is independent of a pedestrian's crossing or not under 3 light conditions. In other words, the test is to find out whether the detection algorithm's performance is the same under various profile conditions. The null hypothesis is that the pedestrian detection algorithm is independent of the test pedestrian's profile. A two-tale P value is used to determine whether the null hypothesis can be rejected or not. When the P value is greater than the significance level (α) 0.05, there is no evidence to reject the null hypothesis. The independence means the pedestrian detection algorithm behaves similarly when the pedestrian's profile varies. Alternatively, the detection algorithm behaves differently when the pedestrian's profile varies. Lastly, the detection time in terms of time to collision is summarized for comparing the performance of all vehicles under test.

Crossing versus Not Crossing Test scenarios S1b and S1e are crossing pedestrians from the nearside or offside, and test scenarios S4a and S4c are not crossing pedestrians but walking or standing in front of the vehicle under test. The test results of FCW are either a warning activated or not. The totals of warnings and no warnings for crossing and not crossing scenarios in the daytime are listed for each vehicle under test in Table 7. Fisher's exact tests are also conducted to evaluate the pedestrian detection of each vehicle. Most vehicles are able to detect the test pedestrian consistently whether it is crossing or not. Vehicles 4, 5, and 7 under test behaved inconsistently when they are detecting the test pedestrian in different orientations.

Table 7. Validation of pedestrian detection for crossing and not crossing scenarios in daytime

Vehicle	Crossing		Not Crossing		Fisher's Test	
	Warning	No Warning	Warning	No Warning	P Value	Null Hypothesis
V1	11	0	10	0	1	Not reject
V2	7	1	8	0	1	Not reject
V3	8	0	5	0	1	Not reject
V4	0	6	6	0	0.002	Reject
V5	10	0	4	4	0.023	Reject
V6	8	1	10	0	0.474	Not reject
V7	11	0	1	4	0.003	Reject
V8	10	0	8	0	1	Not reject
V9	11	0	10	0	1	Not reject
V10	10	0	10	0	1	Not reject
V11	10	0	6	0	1	Not reject

The totals of warnings and no warnings for crossing and not crossing scenarios in the nighttime (low beam) are listed for each vehicle under test in Table 8. Fisher's exact test result shows only vehicle 5 did not behave consistently when detecting the test pedestrian crossing or not. However, the variations of warnings and no warnings increase as compared to the daytime test results.

Table 8. Validation of pedestrian detection for crossing and not crossing scenarios in nighttime (low beam)

Vehicle	Crossing		Not Crossing		Fisher's Test	
	Warning	No Warning	Warning	No Warning	P Value	Null Hypothesis
V1	8	0	10	0	1	Not reject
V2	3	3	5	1	0.546	Not reject
V3	3	3	6	0	0.182	Not reject
V4	0	7	4	3	0.07	Not reject
V5	8	0	0	4	0.002	Reject
V6	0	6	0	5	1	Not reject
V7	11	0	3	1	0.267	Not reject
V8	0	5	0	6	1	Not reject
V9	8	0	6	0	1	Not reject
V10	10	0	10	0	1	Not reject
V11	3	3	4	3	1	Not reject

The totals of warnings and no warnings for crossing and not crossing scenarios in the nighttime (high beam) are listed for each vehicle under test in Table 9. Fisher's exact test result shows 3 vehicles (4, 5, and 8) did not behave consistently when detecting the test pedestrian crossing or not. The test results are similar to the daytime test results.

Table 9. Validation of pedestrian detection for crossing and not crossing scenarios in nighttime (high beam)

Vehicle	Crossing		Not Crossing		Fisher's Test	
	Warning	No Warning	Warning	No Warning	P Value	Null Hypothesis
V1	7	0	8	0	1	Not reject
V2	9	0	6	0	1	Not reject
V3	11	0	10	0	1	Not reject
V4	0	6	5	1	0.015	Reject
V5	10	0	3	4	0.015	Reject
V6	0	5	5	3	0.075	Not reject
V7	10	0	11	0	1	Not reject
V8	5	5	0	7	0.044	Reject
V9	10	0	10	0	1	Not reject
V10	9	1	10	0	1	Not reject
V11	8	0	12	0	1	Not reject

Only vehicle 5 is not able to detect the test pedestrian consistently under crossing or not crossing conditions in daytime, nighttime with low beam, and nighttime with high beam.

Detection Time Table 10 shows the detection times of all vehicles under test in 4 test scenarios and 3 light conditions. In the daytime, the average detection times of vehicles 1, 2, 6, 8, 9, 10, and 11 are longer when the test pedestrian is not crossing. In the nighttime (low beam), the average detection times of vehicles 2 and 10 are longer when the test pedestrian is not crossing. In the nighttime (high beam), the detection times of vehicles 3, 7, 9, 10 and 11 are longer when the test pedestrian is not crossing. On average, the vehicles under test take a longer time to detect the test pedestrian when it is not crossing.

Table 10. Averages of FCW detection times of vehicles under test

Vehicle	Day				Night (low beam)				Night (high beam)			
	S1b	S1e	S4a	S4c	S1b	S1e	S4a	S4c	S1b	S1e	S4a	S4c
V1	1.16	0.66	1.27	1.66	1.07	0.63	1.07	1.37	1.22	0.56	1.07	1.64
V2	0.40	1.14	2.04	2.11	0.61	0.30	0.78	0.73	1.58	1.09	0.79	0.33
V3	1.53	1.07	0.66	1.64	*	1.06	0.66	1.21	1.42	1.26	1.49	1.47
V4	*	*	1.95	2.26	*	*	1.95	0.63	*	*	1.54	2.48
V5	1.10	0.92	0.91	0.35	0.97	0.77	*	*	1.02	0.76	*	1.57
V6	1.09	1.15	1.42	1.68	*	*	*	*	*	*	*	1.70
V7	0.76	0.9	1.12	*	0.84	0.93	*	0.34	0.78	0.89	2.08	1.64
V8	1.53	1.43	1.07	1.75	*	*	*	*	1.64	*	*	*
V9	1.48	1.47	2.66	2.57	0.97	0.75	0.85	1.19	1.30	0.69	1.63	2.10
V10	1.88	1.10	2.23	2.27	1.41	1.08	2.23	2.28	1.71	0.93	2.25	2.28
V11	1.65	1.29	2.04	1.84	0.18	0.06	*	0.38	1.54	1.25	2.01	2.09
Average	1.26	1.11	1.64	1.81	0.86	0.70	1.27	1.02	1.36	0.93	1.61	1.73

*: no data. Unit: second.

CONCLUSIONS

The verification and validation methodology of ML applications for safety functionalities in ADAS/ADS is presented, and an example of FCW in PAEB is demonstrated. The verification of training data provides insights and potential weaknesses of a ML application from the safety perspective in terms of ODD conditions and pre-crash scenarios. Most current pedestrian datasets are lack of inclement weather, weak illumination, air particulate matter, and vehicle/pedestrian maneuvering annotations. The validation process follows the lead of the verification result that the vehicle's headlight and background illumination conditions are needed to be tested under different pedestrian pre-crash scenarios. The test results show that some vehicles under test behave inconsistently when the test pedestrian is crossing as compared to not crossing. Test results in the nighttime with high beam headlight is similar to that of the daytime; however, the test results in the nighttime show significant variations compared with that of daytime. No trend or similarity can be found among all vehicles under test, the same vehicle may behave inconsistently under different light conditions and pedestrian orientations. Also, the pedestrian detection time is longer when the test pedestrian is not crossing on average. The vision-based ML application for the vehicle safety functionality reveals the uncertainty of a DNN, and it results in the inconsistent performance under differentiated ODD conditions and pre-scenarios.

REFERENCES

[1] Kuutti S., Bowden R., Jin Y., Barber F., Fallah S. "A Survey of Deep Learning Applications to Autonomous Vehicle Control." arXiv:1912.10773v1; 2019.
[2] Feng D., Haase-Schutz C., Rosenbaum L. et al. "Deep Multi-modal Object Detection and Semantic Segmentation for Autonomous Driving: Datasets, Methods, and Challenges." arXiv:1902.07830v4; 2020.
[3] Mohseni S., Pitale M., Singh V., Wang Z. "Practical Solutions for Machine Learning Safety in Autonomous Vehicles." arXiv:1912.09630; 2019.
[4] Borg M., Englund C., Wnuk K., et al. 'Safely Entering the Deep: A Review of Verification and Validation for Machine Learning and a Challenge Elicitation in the Automotive Industry." Journal of Automotive Software Engineering. 2019; 1(1), 1–19.
[5] Pullum L. "Verification and Validation of Systems in Which AI is a Key Element". SEBoK Editorial Board. 2021. Accessed 5/11/2022. www.sebokwiki.org.
[6] Elallid B. B., Benamar N., Hafid A. S., Rachidi T., Mrani N. "A Comprehensive Survey on the Application of Deep and Reinforcement Learning Approaches in Autonomous Driving." Journal of King Saud University – Computer and Information Sciences; 2022. https://doi.org/10.1016/j.jksuci.2022.03.013.
[7] Yurtsever E., Lambert J., Carballo A., Takeda K. "A Survey of Autonomous Driving: Common Practices and Emerging Technologies." IEEE Access, 2020; 8: 58443–58469.
[8] Burton S. "A causal model of safety assurance for machine learning." 2022; arXiv:2201.05451v1.
[9] Schwalbe G., Schels M. "A Survey on Methods for the Safety Assurance of Machine Learning Based Systems."

10th European Congress on Embedded Real Time Software and Systems; 2020, Toulouse, France.
[10] Willers O., Sudholt S., Raafatnia S., Abrecht S. “Safety Concerns and Mitigation Approaches Regarding the Use of Deep Learning in Safety-Critical Perception Tasks.” arXiv:2001.08001v1; 2020.
[11] Cheng C-H, et al. “Towards Dependability Metrics for Neural Networks.” arXiv: 1806.02338v2; 2018.
[12] Amershi S. et al. “Software Engineering for Machine Learning: A Case Study.” IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice; 2019, pp. 291-300, doi: 10.1109/ICSE-SEIP.2019.00042.
[13] Koopman P., Wagner M. “Toward a Framework for Highly Automated Vehicle Safety Validation.” 2018 SAE World Congress, SAE 2018-01-1071.
[14] Salay R., Queiroz R. and Czarnecki K. “An Analysis of ISO 26262: Using Machine Learning Safely in Automotive Software.” arXiv :1709.02435v1; 2017.
[15] Staplin, L., Mastromatto, T., Lococo, K. H., Kenneth W. Gish, K. W., & Brooks, J. O. “The effects of medical conditions on driving performance (Report No. DOT HS 812 623).” Washington, DC: National Highway Traffic Safety Administration, 2018.
[16] Swanson, E., Foderaro, F., Yanagisawa, M., Najm, W. G., & Azeredo, P. “Statistics of light-vehicle pre-crash scenarios based on 2011-2015 national crash data (Report No. DOT HS 812 745).” Washington, DC: National Highway Traffic Safety Administration, 2019.
[17] Dvornik N., Mairal J., Schmid C. “On the Importance of Visual Context for Data Augmentation in Scene Understanding.” arXiv:1809.02492v3; 2019.
[18] Rahman Q. M., Sunderhauf N., Corke P., Dayoub F. “FSNet: A Failure Detection Framework for Semantic Segmentation.” arXiv:2108.08748v2; 2021.
[19] Cheng C.-H., Knoll A., Liao H.-C. “Safety Metrics for Semantic Segmentation in Autonomous Driving.” arXiv:2105.10142v2; 2021.
[20] Gyllenhammar M. et al. “Towards an Operational Design Domain That Supports the Safety Argumentation of an Automated Driving System.” Proceeding of the 10th European Congress on Embedded Real Time Software and Systems, Toulouse, 2020.
[21] Riedmaier S., Ponn T., Ludwig D., Schick B., Diermeyer F. “Survey on Scenario-Based Safety Assessment of Automated Vehicles.” IEEE Access 2020; 8: 84756–84777.
[22] Ragesh N. K., Rajesh R. “Pedestrian Detection in Automotive Safety: Understanding State-of-the-Art.” IEEE Access. 2019; 7: 47864–47890.
[23] Cao J. et al. “From Handcrafted to Deep Features for Pedestrian Detection: A Survey.” IEEE Transactions on Pattern Analysis and Machine Intelligence. 44(9):4913-4934; 2022.
[24] Hasan I. et al. “Pedestrian Detection: Domain Generalization, CNNs, Transformers and Beyond.” arXiv:2201.03176v2, 2022.
[25] Dollar P., Wojek C., Schiele B., and P. Perona. “Pedestrian detection: An evaluation of the state of the art.” IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(4):743–761, 2010.
[26] Geiger A., Lenz P., and R. Urtasun. “Are we ready for autonomous driving? the kitti vision benchmark suite.” Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2012.
[27] Zhang S., Benenson R., and B. Schiele. “Citypersons: A diverse dataset for pedestrian detection.” Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[28] Shao S. et al. “Crowdhuman: A benchmark for detecting human in a crowd.” arXiv:1805.00123, 2018.
[29] Neumann L. et al. “Nightowls: A pedestrians at night dataset.” Proc. Asian Conference on Computer Vision, 2018.
[30] Braun M., Krebs S., Flohr F., and D. M. Gavrila. “Eurocity persons: A novel benchmark for person detection in traffic scenes.” IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(8):1844–1861, 2019.
[31] Pang Y., et al. “Tju-dhd: A diverse high-resolution dataset for object detection.” IEEE Transactions on Image Processing, 2020.
[32] Zhang S. et al. “Widerperson: A diverse dataset for dense pedestrian detection in the wild.” IEEE Transactions on Multimedia, 22(2):380–393, 2020.
[33] Pedestrian automatic emergency brake system confirmation test (working draft). Washington, DC: National Highway Traffic Safety Administration, 2019.
[34] Dowdy S., Wearden S., and D. Chilko. “Statistics for Research” 3rd Edition, 2004.