TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Ghaderi, O., Pappalardo, G., Di Graziano, A., Dimauro, G. A., Wang, X., Farah, H., & Cafiso, S. (2026). Interaction of road, traffic and environment in the operational design domain of lane support systems: hybrid factorial–observational design and causal analysis. *Accident Analysis and Prevention*, *236*, Article 108691. https://doi.org/10.1016/j.aap.2026.108691

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

*This work is downloaded from Delft University of Technology.*

Accident Analysis and Prevention 236 (2026) 108691

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

# Interaction of road, traffic and environment in the operational design domain of lane support systems: hybrid factorial–observational design and causal analysis

Omid Ghaderi [a,1,*], Giuseppina Pappalardo [a,2], Alessandro Di Graziano [a,3], Giovanni Andrea Dimauro [a,4], Xuesong Wang [b,c,5], Haneen Farah [d,6], Salvatore Cafiso [a,7]

[a] Department of Civil Engineering and Architecture, University of Catania, Catania, Italy
[b] The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Shanghai 201804, China
[c] College of Transportation, Tongji University, Shanghai 201804, China
[d] Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN Delft, the Netherlands

ARTICLE INFO

*Keywords:*
Road safety
Lane marking
Advanced Driver Assistance Systems (ADAS)
Lane Support System (LSS)
Operational Design Domain (ODD)
Machine Learning (ML)

ABSTRACT

The Operational Design Domain (ODD) defines the conditions under which automated driving and driver-assistance systems are expected to operate. This study evaluates the ODD of a camera-based Lane Support System (LSS) using direct Mobileye 6.0 lane-detection quality outputs. A large-scale hybrid factorial–observational field design covered 6 different Light × Weather combinations across 9,351 road sections on two-lane rural roads with wide variability in lane-marking retroreflectivity (RL) and road horizontal alignment and cross section characteristics.

Statistical and machine-learning classification models were calibrated and compared to analyze the relationships between lane-marking quality and environmental, road, and traffic features. An AutoML-LightGBM pipeline with SMOTE-based class-imbalance treatment achieved the best accuracy of 0.81. SHAP analysis identified low RL, rain, night conditions, narrow lanes, and high curvature as contributors to critical detection conditions. Because standard ML is optimized for prediction rather than causal inference, Double Machine Learning was added to estimate adjusted effects from observational data. Higher RL, higher speed, and wider lanes were associated with better expected Mobileye quality scores, whereas rain, night conditions, and higher curvature were associated with lower detection quality. SHAP dependence and conditional SHAP analyses supported the identification of maintenance-mitigable infrastructure constraints and harder environmental/geometric ODD limits. One practical result is that RL transitions from low-quality detection mainly occur within 120–150 mcd/(m$^2$·lx) across the majority of environmental and physical conditions, although this transition is less evident under sharp curvature or rain.

## 1. Introduction

The development of automated Vehicles (AVs) offers significant potential for more efficient, eco-friendly, and safe mobility. Various sensors and advanced driver-assistance systems (ADAS) aid AVs in safe navigation, evaluating the reliability of these systems and their

* Corresponding author.
*E-mail addresses:* omid.ghaderi@unict.it (O. Ghaderi), giuseppina.pappalardo1@unict.it (G. Pappalardo), adigraziano@unict.it (A. Di Graziano), giovanni.dimauro@phd.unict.it (G.A. Dimauro), wangxs@tongji.edu.cn (X. Wang), H.Farah@tudelft.nl (H. Farah), dcafiso@unict.it (S. Cafiso).
[1] ORCID: 0000-0002-8859-6413.
[2] ORCID: 0000-0002-9793-1885.
[3] ORCID: 0000-0001-6484-2758.
[4] ORCID: 0009-0003-9262-6042.
[5] ORCID: 0000-0001-8046-3213.
[6] ORCID: 0000-0002-2919-0253.
[7] ORCID: 0000-0002-7247-0365.

https://doi.org/10.1016/j.aap.2026.108691
Received 2 April 2026; Received in revised form 17 July 2026; Accepted 18 July 2026
Available online 28 July 2026
0001-4575/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Operational Design Domain (ODD) is essential for manufacturers of AVs to define the ADAS' operating capability. It may also be of interest to insurers, regulators, service providers, and road agencies, enabling them to understand AVs' needs and capabilities. While rigorous assurance approaches such as formal verification can provide strong safety guarantees, their conclusions may become conservative or inconclusive under uncertainty and perturbations, which reinforces the importance of empirical ODD evaluation for ADAS functions (Liu et al., 2026).

SAE defines ODD as "*the operating conditions under which a given driving automation system or feature thereof is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics*" (Taxonomy, 2018). While Level 5 autonomous vehicles have an unlimited ODD, systems at Levels 1 through 4 have restricted ODDs that vary based on the original equipment manufacturer (OEM) and the system's characteristics. A reference standard (BSI) PAS 1883:2020 (The British Standards Institution, 2020). It establishes a hierarchical taxonomy for ODDs into three primary elements: scenery, environment, and dynamic elements. The "scenery" attribute shall consist of the non-movable elements of the operating environment, e.g. roads or traffic lights. The "environmental conditions" attribute shall consist of weather and atmospheric conditions. The "dynamic elements" attribute shall consist of the movable elements of the ODD, e.g. traffic or subject vehicle.

Advanced Driver Assistance Systems (ADAS) rely on cameras, radar, LiDAR, ultrasonic sensors, and AI-driven software to observe the surroundings, identify hazards, and deliver real-time driver support. ADAS encompasses various technologies, including adaptive cruise control (ACC), anti-lock braking systems (ABS), lane support system (LSS), automatic parking, etc. Among the different ADAs already in operation in L1+ vehicles, Lane Support Systems (LSS) represent an essential first step toward complete automation, offering significant and immediate safety enhancements on roadways. Field-effectiveness evidence has shown that lane-support technologies can reduce lane-departure crashes, with Lane Keep Assist combined with Lane Departure Warning associated with a 12% reduction (Leslie et al., 2021). As critical components for preventing unintentional lane departures through functionalities like lane keeping assistance and departure warnings, LSS is already a standard feature in modern vehicles (The European Parliament and the Council of the European Union, 2019). These systems rely heavily on camera-based sensors and computer vision algorithms to "read" lane markings. Research studies consistently highlight performance challenges for LSS under varying road and environmental conditions. This reinforces the need to characterize how real-world road, traffic, and environmental conditions constrain the operational capability of lane support functions.

Against this background, recent research has increasingly examined LSS performance not only as a vehicle-system issue, but also as a function of the surrounding operational context. In particular, the literature has investigated how physical infrastructure, environmental exposure, and traffic-related conditions contribute to defining the practical boundaries of the ODD for lane support functions.

Recent safety-oriented research has shown that the assessment of lane-keeping systems should also account for hazards arising from interactions between sensing, control logic, and the operating environment. Using a Systems-Theoretic Process Analysis of an Automated Lane Keeping System, Elizebeth et al. (2025) identified adverse weather, poor lighting, sharp curves, and limited environmental feedback as relevant conditions that constrain safe system operation within the ODD.

Farah et al. (2020) examined a Lane Keeping Assist (LKA) system's performance across various ODD classifications (inside the ODD, outside the ODD, and indeterminate) under diverse road types and conditions. The findings revealed inconsistencies between the ODD defined in the OEM manual and the driver's interpretation in all test scenarios.

Several studies have demonstrated that lane marking conditions significantly impact the accuracy and robustness of lane detection, underscoring the need for updated maintenance practices.

Lehoczký et al. (2022) conducted an in-depth study assessing the readiness of Slovak road infrastructure for the deployment of intelligent transportation systems, including Lane Keeping Assist (LKA) and Intelligent Speed Assist (ISA). The authors emphasized the critical role of well-maintained, visible road markings and signage in the effective operation of ADAS technologies. The visual assessment of the marking conditions revealed significant disparities in lane marking quality, especially in rural and secondary roads, where markings were often faint.

Cafiso and Pappalardo (2020) and Pappalardo et al. (2021) reported a 2.6% LSS fault occurrence rate, which was statistically correlated to the luminance coefficient of lane markings under diffuse lighting (Qd) and the inverse of the horizontal curvature radius (1/R). Speed did not result in a significant variable of the system fault.

Wang et al. (2021) showed that roadway geometry can directly constrain the ODD of automated driving systems. In their analysis of autonomous-vehicle operation at skewed intersections, geometric characteristics, such as the intersection angle and the associated required detection angle and distance, were found to define operational boundaries for safe system performance, indicating that infrastructure geometry should be considered an explicit component of ODD assessment. Qin et al. (2022) further showed that roadway geometry and operating conditions can constrain automated-driving ODD by increasing the sensing requirements for safe operation, indicating that demanding infrastructure configurations may exceed the capability of single-sensor systems. Ye and Wang (2022) further showed that roadway geometry and speed environment can constrain automated-driving ODD by increasing the sensing demands required for safe operation, particularly under more demanding alignment and grade conditions.

García and Camacho-Torregrosa (2020) conducted on-road pilot tests to evaluate lane-keeping systems, assessing their performance across lane widths ranging from 2.28 m to 3.80 m. The study found that the system could not operate at lane widths of 2.50 m or less, whereas it consistently operated on lanes 2.75 m or wider. Similarly, Marr et al. (2020) emphasized that lane widths below 2.8 m pose significant challenges for vehicle machine vision systems, particularly in the absence of edge markings. In another study, García et al. (2020) tested a Level 2 vehicle on various horizontal curves, revealing a clear relationship between vehicle speed and system disengagements.

Field tests conducted by Reddy et al. (2020) identified key factors influencing lane-keeping performance and examined the impact of vehicle speed and visibility conditions on system effectiveness. The highest lane-keeping accuracy was observed on straight sections and in lanes wider than 2.50 m, while speeds exceeding 90 km/h reduced performance. Another study by Cafiso et al. (2024) confirmed the relevance of horizontal curvature as a critical factor constraining the physical infrastructure, shaping the ODD of LSS.

In another study by Pappalardo et al. (2022) the authors examined Lane Departure Warning (LDW) performance on two-lane rural roads under different weather conditions. Using on-road testing and Generalized Estimating Equations (GEE), the study identified a significant rise in LSS faults during rainfall, whereas wet pavement alone had no substantial impact on system performance. The findings also confirmed the critical roles of lane marking quality (RLw) and the horizontal curvature radius (1/R) in defining the ODD for LDW.

Babić et al. (2021) examined differences in lane marking detection and field of view between daytime and nighttime conditions using the Mobileye 630 system. Their on-road tests conducted on rural roads found that nighttime conditions enhanced lane-marking detection, boosting confidence levels by 5% to 8% and reducing "nothing detected" incidents by 12%.

Chai et al. (2023) developed a surrogate evaluation model to assess the reliability of lane detection using images recorded from a front-mounted monocular camera and applying RANSAC software for lane detection over a limited road extension of 3.6 km. The study identified

environmental and infrastructure-related parameters such as road curvature, lighting conditions, and lane marking clearance as critical influences on perception performance. The model also highlighted the combined influence of curvature, lighting, and marking clearance.

Overall, the reviewed literature shows that Lane Support System operational capability is not determined by a single factor but by the combined influence of lane marking visibility, roadway geometry, environmental conditions, and driving context. Existing studies have progressively clarified the roles of individual variables, such as retroreflectivity, curvature, lane width, weather, and lighting, while also indicating that their effects may interact in nonlinear ways. This body of evidence confirms that the ODD of LSS should be interpreted as a multifactor domain shaped by infrastructure, environment, and traffic conditions, thus motivating a more integrated experimental and analytical framework.

### 1.1. Research gaps, objectives, and paper outline

Early studies on LSS performance focused mainly on lane-marking quality and road-geometry constraints, and were later extended to environmental conditions such as lighting, wet pavement, and rain. The reviewed literature shows that the operational capability of camera-based LSS is not determined by a single factor but by the combined influence of lane-marking quality, roadway geometry, environmental conditions, and driving context. Previous studies have already identified the relevance of individual variables such as retroreflectivity, curvature, lane width, weather, and lighting. However, many studies examined selected factors or limited combinations of conditions, often using short test sections, external video processing, dashboard warnings, or statistical models applied to only a few factors. Therefore, a need remains for large-scale field evidence based on direct LSS quality outputs, as well as for an integrated interpretation of how road, traffic, and environmental factors jointly define system-specific ODD boundaries. Given the nonlinear interactions among scenery, environmental, and dynamic ODD attributes, the combined contribution of multiple factors requires further exploration (Babić et al., 2021; Chai et al., 2023).

To address this gap, the present study uses a large-scale field campaign on two-lane rural roads, with planned factorial coverage of two lighting and three weather conditions, while observing infrastructure and traffic variables across a heterogeneous road network. The study does not treat traffic variables as experimentally controlled factors. Instead, it combines controlled coverage of Light × Weather scenarios with a selection of road sections purposefully selected to span variability in lane-marking retroreflectivity, curvature, lane width, shoulder width and section type. Direct Mobileye 6.0 lane-detection quality outputs are integrated with certified RL measurements, road geometry, traffic variables, and environmental conditions through GIS-based spatial data fusion.

Given the dataset's heterogeneity and the nonlinear relationships among ODD factors, several statistical and machine-learning models are compared to classify Mobileye lane-marking quality, with particular focus on the critical low-detection class. SHAP analysis is then used to interpret the fitted model and examine the conditional contribution of road, traffic, and environmental factors. Because SHAP attributions can be sensitive to correlated predictors, a Double Machine Learning (DML) causal-effect robustness analysis is added to estimate treatment effects from observational data. The causal interpretation of these estimates is conditional on exchangeability given the observed covariates, sufficient overlap across treatment values, and consistency of the treatment and outcome definitions. DML reduces bias from nuisance relationships by residualizing the outcome and treatment variables and using cross-fitting to improve the reliability of the adjusted effect estimates. Finally, a conditional SHAP analysis of RL under Light × Weather scenarios is used to support a practical ODD interpretation that distinguishes infrastructure-related constraints that can be partly mitigated through maintenance from harder physical and environmental constraints that may require vehicle-system or sensing improvements.

## 2. Field campaign design and data preparation

The field campaign was designed to capture the variability of LSS performance under different environmental, infrastructure, and traffic conditions. The planned factorial component of the campaign covered all combinations of two lighting conditions, day and night, and three weather conditions, dry, wet, and rainy. Road geometry, lane-marking retroreflectivity, lane width, shoulder width, and section type were not experimentally manipulated at the section level; rather, their variability was ensured by selecting an extended and heterogeneous two-lane rural road network, while traffic conditions, including speed and headway, were observed during the survey. Detailed data from multiple sources were aggregated into 50-m road sections, resulting in a final dataset of 9,351 samples.

### 2.1. Data acquisition, integration, and preprocessing

Data collection and treatment include data acquisition from multiple sources, data fusion into a common reference road section, and data validation and cleaning to address sensor errors and inconsistent or missing values.

In this research, to collect detailed and high-resolution data on LSS operation, the Mobileye 6.0 (Intel®) full protocol, a cutting-edge vision-based ADAS, was installed on the Automatic Road Analyzer (ARAN 9000) at the University of Catania. A data logger was specifically developed to record and synchronize data from Mobileye with other sensors, including a high-accuracy GNSS/INS system and cameras.

The Italian Road Agency (ANAS) provided the data on the retroreflectivity index (RL) (International Organization for Standardization, 2017; EN 1436, 2018) of the lane marking collected at 1 m resolution using the Delineation Photometric Certified Instrument (DELPHI), with measurements statistically aggregated (mean, max, min, standard deviation) every 50 m. Data on the road geometry features were derived from Open Street map (OSM) and ANAS inventory database. Fig. 1 illustrates the two mobile laboratories used for data acquisition.

Since all the data were accurately geo-localised, various spatial data fusion techniques in GIS (QGIS, Ver 3.28.0) were employed to associate the ADAS, OSM, DELPHI, and ANAS data with the same road section, which are located in different positions and collected at different frequencies. It is worth mentioning that each section of 50 m can be characterized by a variability in values of the different features, the dataset reports the main statistics of each parameter (e.g., min, max, mean, standard deviation).

ADAS data were collected across various weather and lighting conditions, creating a robust dataset to evaluate the ODD of the LSS. This dataset is essential for investigating correlations among road features (e.g., lane marking quality, cross-section width, horizontal alignment, section type), driving conditions (speed, vehicle headway), and environmental conditions (day, night, wet, rain, glare, lighting) with the LSS detection capability.

We excluded LSS data marked by the Mobileye system as failsafe due to blurred photos and glare, and inconsistent data associated with lane change signals, accounting for 6.43% of the total ADAS dataset. After data fusion and cleaning, sections with speeds lower than 40 km/h were also filtered out because the LSS is not fully operational at very low speeds. A comprehensive dataset was then created encompassing the three main fields of the ODD's taxonomy: the physical infrastructure attributes: road geometry (horizontal alignment, cross-section width, section type), lane marking quality (RL); the dynamic driving conditions: speed, headway; and the environmental factors: day/night, dry/wet/rain. The outcome variable is the lane marking detection quality rate (Q), which Mobileye classifies as Low (1), Medium (2), or High (3). Table 1 provides summary statistics for the dataset, which includes 9,351 sections covering 468 km of two-lane rural roads.

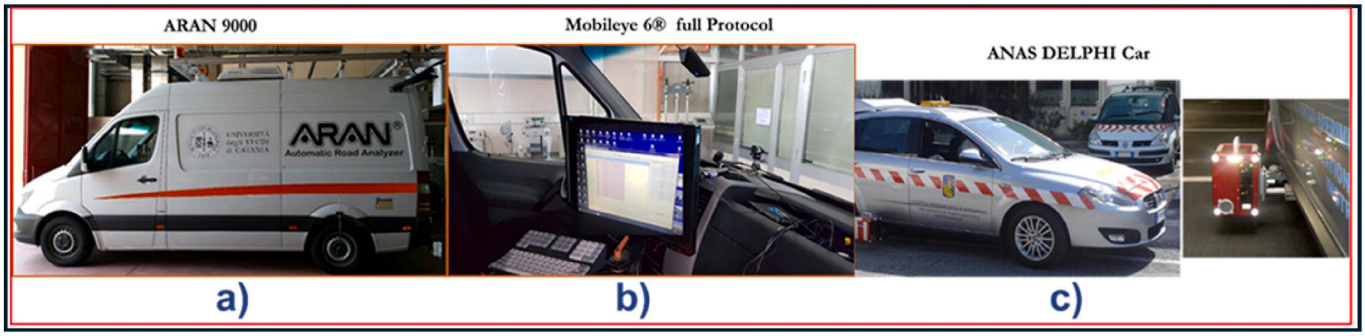



**Fig. 1.** Data acquisition: a) University of Catania ARAN 9000 vehicle; b) Mobileye 6 full protocol; c) ANAS DELPHI vehicle.

**Table 1**
Summary statistic parameters.

| ***Numerical*** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | ***RL* average** [mcd/m$^2$ /lx] | ***RL Min*** [mcd/m$^2$ /lx] | ***RL Max*** [mcd/m$^2$ /lx] | ***Speed* average** [km/h] | ***Headway Max*** [sec] | ***Headway Min*** [sec] | ***Headway* average *[sec]*** | ***Curvature Min*** [1/m] | ***Curvature Max*** [1/m] | ***Lane [m]*** | ***Shoulder [m]*** |
| **Count** | 9351 | 9351 | 9351 | 9351 | 9351 | 9351 | 9351 | 9351 | 9351 | 9351 | 9351 |
| **Mean** | 193.34 | 106.63 | 205.84 | 67.86 | 3.34 | 3.3 | 3.32 | 7.24E−04 | 1.63E−03 | 3.47 | 0.52 |
| **Standard Deviation** | 96.07 | 67.28 | 101.13 | 12.02 | 0.55 | 0.62 | 0.57 | 2.00E−03 | 2.92E−03 | 0.23 | 0.38 |
| **Coeff. of Variation** | 49.70% | 63.10% | 49.10% | 17.70% | 16.40% | 18.70% | 17.20% | 276.50% | 178.40% | 6.50% | 72.60% |
| **Minimum** | 42 | 40 | 42 | 40 | 0.6 | 0.1 | 0.56 | 1.00E−10 | 1.00E−10 | 3.15 | 0.1 |
| **Maximum** | 614 | 423 | 786 | 99.7 | 3.5 | 3.5 | 3.5 | 3.14E−02 | 3.58E−02 | 3.9 | 1.6 |
| ***Categorical*** | | | | | | | | | | | |
| **Light** | | | | **1 = Day** (64.61%) | | | | **0= Night** (35.39%) | | | |
| **Weather** | | | | **1 = Dry** (49.98%) | | | | **2 = Rain** (33.23%) | | **3 = Wet** (16.79%) | |
| **Section Type** | | | | **0 = Other** (94.54%) | | **1 = Tunnel** (0.12%) | | **2 = Bridge** (1.32%) | | **3 = Junction** (4.02%) | |
| ***Lane Marking Quality (Categorical)*** | | | | | | | | | | | |
| | | | **Class 1 Low** | | | **Class 2 Medium** | | | **Class 3 High** | | |
| **Qmax** | | | 6.90% | | | 13.50% | | | 79.60% | | |
| **Qmin** | | | 20.93% | | | 15.88% | | | 63.19% | | |
| **Qmode** | | | 11.56% | | | 15.98% | | | 72.46% | | |

## 3. Analytical framework

To develop a robust analysis of lane-marking detection quality, the workflow included feature selection, class-imbalance treatment using SMOTE, model comparison, model interpretation, and adjusted-effect analysis. Several statistical and machine-learning algorithms were evaluated, and AutoML-LightGBM achieved the best performance, especially for the lower-represented and most critical Class-1 detection condition, which is associated with an approximately 100% frequency of system faults. SHAP analysis was then applied to interpret global feature relevance and interaction patterns. Because SHAP values can be sensitive to correlated predictors, a complementary Double Machine Learning analysis was added to assess whether the main feature directions remained consistent after controlling for the remaining

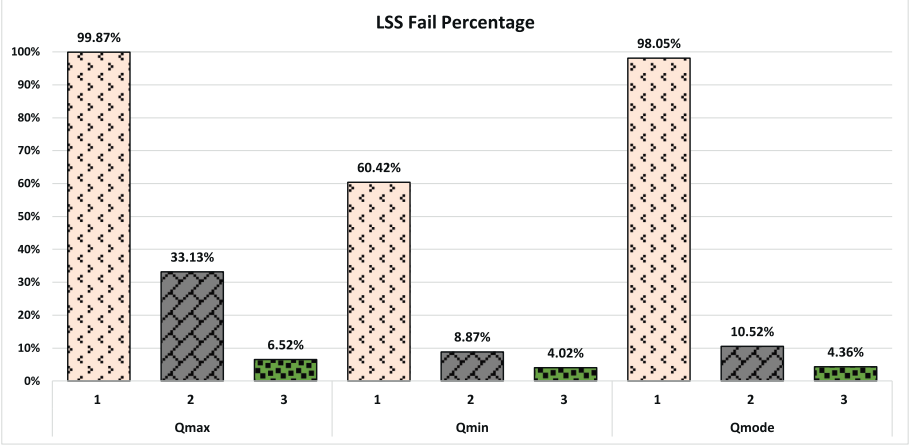



**Fig. 2.** Percentages of LSS faults for different marking detection quality scores.

covariates. Finally, a conditional SHAP analysis was used to examine the RL contribution under the six Light × Weather scenarios.

### 3.1. Target variable and feature selection

The LSS provides both classification of detection quality and an event indicating a system fault, making it impossible for the ADAS function, which has been recorded at 20 Hz. Lane marking quality represents how the LSS classifies lane marking detection with three quality scores (Q): Low (1), Medium (2), and High (3). Fig. 2 indicates that for class 1, the data shows a high frequency of LSS failures to detect lane markings, particularly when class 1 is referred to the max (Qmax) and mode (Qmode) values of the measures in the 50 m section. Therefore, among the Q statistics, we selected Qmode (i.e., the category with the highest frequency in the section) as the most representative of the lane marking quality classification for LSS performance. This choice is consistent with the section-level objective of the study, because Qmode represents the prevailing detection condition across the multiple LSS observations within each 50-m section and is therefore compatible with the section-level values of the other covariates. In contrast, Qmin may reflect an isolated local reduction over approximately 1 m. Nevertheless, Qmin or lower-percentile measures may be more appropriate for event-based or safety–critical analyses.

In order to identify the explanatory variables to be included in the analysis, we used feature importance analysis that is provided by different classifiers like Voting classifiers, which combine Random Forest (RF) and Support Vector Machines (SVM), Gradient Boosting Machines (GBM), and CatBoost. These algorithms consistently ranked feature relevance similarly. Ultimately, CatBoost was chosen for its reliable performance, superior accuracy, and ability to handle categorical data effectively (Dorogush et al., 2018; Prokhorenkova et al., 2018). CatBoost's ability to handle both numerical and categorical features, its updated permutation techniques, and its robust performance make it a powerful tool for fast, easy interpretation of feature scores through recursive feature elimination. Based on this preliminary analysis, the highest classification accuracy was achieved by including the nine variables in Fig. 3 in the model, selected as the 9 most important features from the full set of 14 variables in Table 1. Accordingly, to the use of $Q_{mode}$ as the target variable, $RL_{avg}$, $Speed_{avg}$, and Shoulder are identified as the most appropriate statistics of the parameter in the 50 m road section.

The CatBoost feature-importance analysis was used mainly as a descriptor-selection and dimensionality-reduction step, rather than as the sole basis for defining the conceptual ODD factors. Originally, the main ODD components, including RL, curvature, lane width, shoulder width, speed, headway, lighting, weather, and section type, were selected based on their relevance to LSS operation in previous literature. Since several features were represented by multiple statistical descriptors within each 50-m section, such as average, minimum, maximum, and variability measures, CatBoost was used to identify the most informative descriptor for each feature relative to the section's quality score mode (Qmode). Therefore, this step should not be interpreted as a model-specific procedure designed to favor CatBoost in the subsequent model comparison, but as a practical feature reduction of redundant descriptors before comparing alternative statistical and machine-learning models.

### 3.2. Class imbalance treatment

The overall accuracy of the CatBoost model was 0.78, while the F1 scores were 0.30, 0.60, and 0.89 for Classes 1, 2, and 3, respectively, indicating high variability, with a bias toward Class 3. The frequency distribution of Q-mode (Table 1) shows that classes 1 (11.56%) and 2 (15.98%) are underrepresented compared to class 3 (72.46%), which is predominant, resulting in an imbalanced dataset that biases calibration toward the most represented class 3. To prevent the model from being biased towards the highest frequency class and to ensure accurate classification of the most critical class 1, we applied the Synthetic Minority Over-sampling Technique (SMOTE) due to its proven performance (Fernández et al., 2018; Chawla et al., 2002; He and Garcia, 2009). SMOTE generates synthetic samples from the minority classes to balance the class distribution, while limiting excessive alteration of the original class distribution.

The resampling strategy aimed to reduce class imbalance by increasing the representation of Classes 1 and 2 relatives to the dominant Class 3. Three training-set resampling strategies were tested: no SMOTE, 50% SMOTE, and 100% SMOTE. The OD results obtained on the unchanged original test distribution constitute the primary reference for evaluating performance under the original class distribution. Fig. 4 reports the class-wise F1-scores for the primary OD evaluation and the supplementary BD diagnostic. In the OD evaluation, SMOTE was applied after the train-test split and only to the training subset, while the original test subset remained unchanged. The BD results provide a supplementary diagnostic of class-wise behavior under the corresponding resampled class distributions. Based on both evaluations, the 50% SMOTE configuration was retained for the subsequent model comparison and interpretability analyses. This configuration provided the optimal trade-off, improving minority-class performance while preserving the natural predominance of Class 3. Despite resampling, Class 2 remains the hardest class to classify because it represents a transitional detection condition between clearly low and clearly high lane-marking quality. This ambiguity cannot be fully resolved by resampling alone; however, it does not compromise the main objective of the study, which focuses on identifying the critical low-detection condition represented by Class 1.

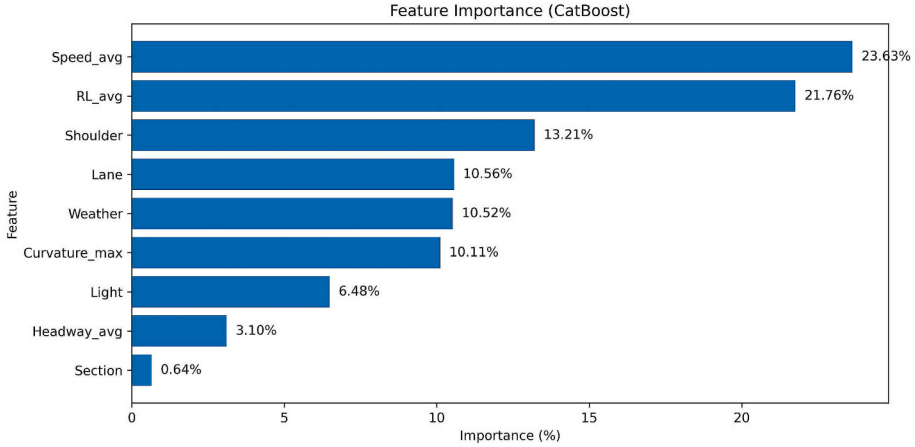



**Fig. 3.** Feature Importance Analysis for Qmode using Catboost.

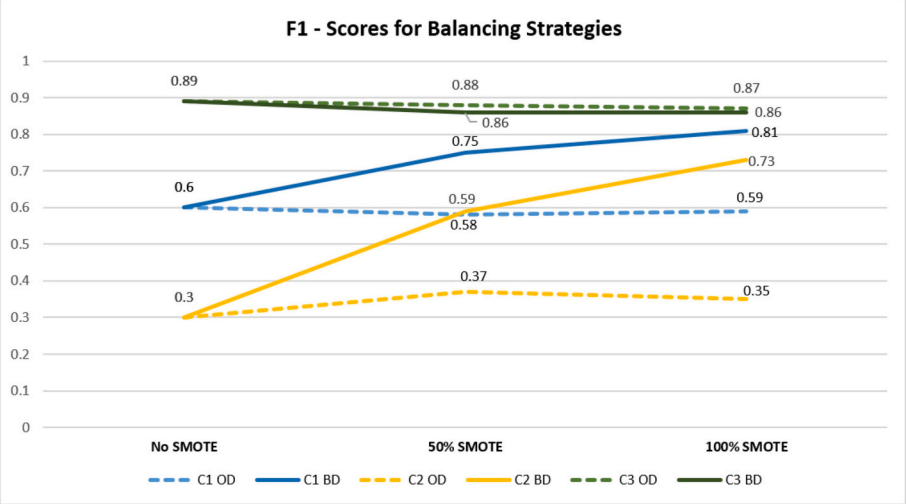



**Fig. 4.** F1-scores for different resampling strategies (C1, C2, and C3 are the three quality classes). OD: primary evaluation on the original imbalanced test data, with SMOTE applied only to the training set; BD: supplementary diagnostic on balanced test data.

## 4. Model development and comparison

As discussed previously, lane-marking quality is categorized into three ordered classes: Class 1 represents the lowest detection quality, Class 2 represents medium detection quality, and Class 3 represents the highest detection quality. These classes have an ordinal structure because they represent increasing levels of Mobileye lane-detection quality, although the numerical distance between adjacent classes should not be interpreted as strictly equal. For this reason, ordinal statistical models, including Ordinal Logistic Regression and Ordinal Gaussian Process Regression, were included as interpretable benchmark models. General classification models, such as Random Forest, Logistic Regression, LightGBM, and CatBoost, treat the classes as nominal categories unless an ordinal-specific loss function or decomposition strategy is explicitly implemented. The best-performing ensemble models, including LightGBM and CatBoost, were implemented using multiclass classification objectives rather than ordinal-specific loss functions. Accordingly, the ordinal meaning of the target was retained mainly in the interpretation of the results, particularly through the focus on Class 1 as the critical low-detection condition and Class 3 as the high-detection condition. The following sections describe the statistical benchmark models and the machine-learning classifiers evaluated in this study.

### 4.1. Statistical models

Statistical models are grounded in the theory of generalized linear models (GLMs) and are typically used for classification or prediction, relying on assumptions about the data's distribution. These models are highly interpretable, and the well-defined model form can provide insights into the relationships between variables. We selected Ordinal Logistic Regression and Ordinal Gaussian Process Regression as alternative statistical models for comparison with ML.

#### 4.1.1. Ordinal Logistic Regression

Ordinal Logistic Regression (Proportional Odds Model) is a well-known linear regression technique for modeling the relationship between an ordinal dependent variable and one or more independent variables. It assumes proportional odds ratios between all pairs of outcome categories, estimating a single set of coefficients for the log odds. This model is suitable when the response variable is ordinal, and the proportional odds assumption is valid (Kleinbaum and Klein, 2010).

#### 4.1.2. Ordinal Gaussian Process Regression

Ordinal Gaussian Process Regression extends Gaussian process regression to ordinal data, offering a nonparametric, flexible method for modeling underlying relationships within a probabilistic framework. This approach effectively handles complex, non-linear relationships, providing better accuracy and interpretability than traditional linear models (Neal, 1997; Chu et al., 2005).

Statistical models provide interpretable baselines for classification problems and conceptually explain the relationships among variables. However, their reliance on assumptions of linearity and proportional odds often limit performance when dealing with complex nonlinear relationships or high-dimensional data, requiring more flexible methods. Therefore, machine learning models are reassuring due to their adaptability in addressing such challenges.

### 4.2. Machine learning models

In machine learning, ordinal regression models have been developed by extending classification algorithms to leverage multiple binary classification tasks. In this study, three ensemble models, including a CatBoost fine-tuning model that was already used in the preliminary analysis were selected for ordinal classification. Machine learning models effectively address the limitations of statistical approaches through their flexibility, ensemble techniques, and ability to handle complex, nonlinear relationships between features.

#### 4.2.1. Ordinal Random Forest tuned by genetic algorithm

In the area of parallel ensemble machine learning, Ordinal Random Forest extends the traditional random forest algorithm to handle ordinal target variables, adjusting the impurity measure to better reflect the target variable's order, improving both accuracy and interpretability (Japkowicz and Shah, 2011; Agresti, 2010). This method modifies the splitting criteria in decision trees to respect the ordinal nature of the target variable, thereby enhancing predictive performance and interpretability compared to treating the ordinal variable as either continuous or nominal (Janitza et al., 2016; Hornung, 2020). This ORF model was further optimized using a genetic algorithm for hyperparameter tuning in the R environment.

#### 4.2.2. Automated Machine Learning-Light Gradient Boosting Machine (AutoML-LGBM)

LightGBM, or Light Gradient Boosting Machine, is a highly efficient gradient boosting ensemble model developed by Microsoft. It is

designed for speed and performance, especially with large datasets. LightGBM offers several advantages over traditional gradient boosting methods, including greater efficiency, higher accuracy, better scalability, and advanced features. Unlike other boosting algorithms, it uses a leaf-wise strategy for tree growth rather than the traditional depth-wise approach, making it effective for both regression and classification tasks (Ke et al., 2017). LightGBM settings were optimized using TPOT (Tree-based Pipeline Optimization Tool) through a custom configuration dictionary specifying ranges for hyperparameters such as the number of leaves, tree depth, learning rate, number of boosting rounds, minimum child samples, subsample fraction, fraction of features used to build each tree, regularization terms (alpha and lambda), and minimum gain required to make a split. Genetic programming in TPOT evolves machine learning pipelines by starting with a population of randomly generated pipelines and iteratively selecting, mutating, and recombining them based on performance. This evolutionary process is guided by a fitness function, in this case, model accuracy. TPOT runs multiple generations of LightGBM models with these hyperparameters, evolving and selecting the best configurations based on the fitness function (Hutter et al., 2019).

#### 4.2.3. Catboost

As discussed in the "Feature Selection" section, we chose the CatBoost classifier over other methods due to its strong performance with categorical data and robustness across varied data distributions. Compared to the preliminary analysis, we optimized CatBoost's hyperparameters using TPOT (Tree-based Pipeline Optimization Tool). TPOT uses a genetic algorithm to streamline the selection and setup of the best machine learning pipeline. It emphasizes fine-tuning hyperparameters like the number of iterations, learning rates, tree depths, and border counts. It also uses L2 regularization to penalize large coefficients; thus, it avoids overfitting and favors simpler models, which may generalize well across a wider range of data sets. The resulting technique not only accounts for the ordinal nature of target classes but also improves the predictive accuracy and robustness across several operational scenarios.

Machine learning techniques can achieve strong performance while maintaining computational efficiency by optimizing hyperparameters and leveraging ensemble methods.

## 5. Results and interpretation

### 5.1. Model performance comparison

This section evaluates the performance of the different classification models, focusing on their ability to accurately classify lane marking quality. We compared two different approaches based on Generalized Linear Models and Machine Learning, by testing five different models for ordinal classification, to select the best method to be used to analyze the relationships between the different features representing the ODD factors and the performance of the LSS represented by the target variable Qmode. Fig. 5 shows the overall accuracy of the five selected classification models. Among the statistical GLM and machine learning (ML) approaches, ML models, especially AutoML-LGBM and CatBoost, performed best. GLM models provided useful benchmarks but with lower performances due to the complex non-linear relationships. As we can see, the AutoML-LGBM method achieves the highest accuracy of 0.81 across 0.63–0.81.

LightGBM and CatBoost, both advanced implementations of the gradient boosting framework based on decision trees, demonstrated remarkable capabilities. They managed large datasets efficiently and handled categorical features without much preprocessing. The core idea of boosting, in which new models iteratively correct the errors of previous models, is central to their design. LightGBM uses a leaf-wise growth strategy, selecting the leaf that minimizes loss, which helps manage complex decision boundaries and reduces overfitting. On the other hand, CatBoost uses ordered boosting and symmetric tree structures to mitigate overfitting. It also includes innovative feature-encoding techniques for handling categorical data. These models are augmented with sophisticated strategies such as LightGBM's Gradient-based One-Side Sampling (GOSS), which prioritizes informative instances during training, and CatBoost's Exclusive Feature Bundling (EFB), which efficiently reduces feature dimensionality. This combination of strategies makes these models particularly adept at navigating the challenges posed by high dimensionality, class imbalance, and the diverse nature of feature types within the dataset, ensuring high predictive accuracy across various operational conditions.

Gaussian Process Regression and Ordinal Logistic Regression, which model the problem in a more rigid but bounded way of handling regression and classification tasks. GPR, on the one hand, offers an elegant probabilistic view that masterfully captures the uncertainties in ordinal data but does not scale to larger, noisier environments. Finally, Ordinal Logistic Regression is particularly well-suited to datasets with more linear feature relationships, but it is constrained in complex ones where interactions are nonlinear.

Therefore, further analysis was conducted using only the results from the AutoML-LGBM model. To assess the model's internal stability and potential overfitting, 5-fold cross-validation was employed together with learning and validation curve analyses. The model consistently achieved high training accuracy (~0.98) and stable validation accuracy (~0.81) across folds, indicating stable internal performance under

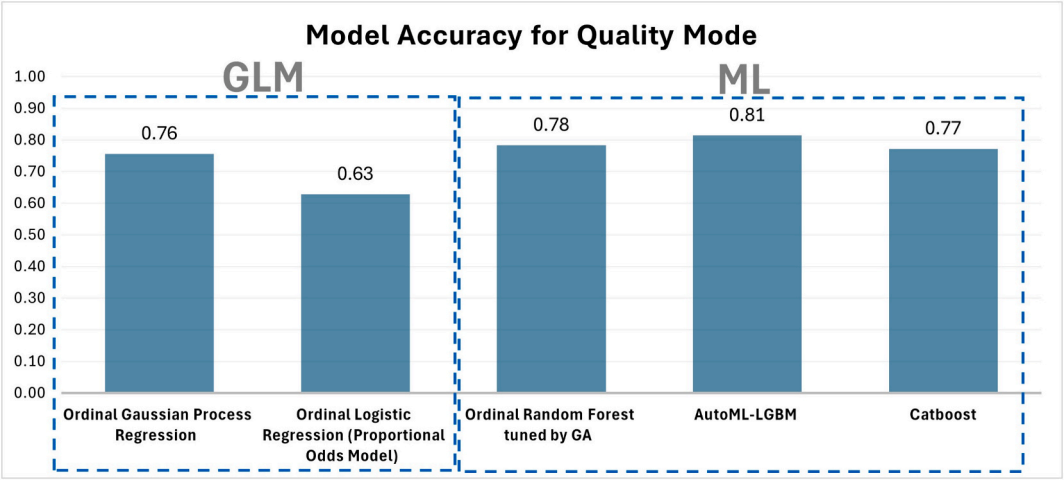



**Fig. 5.** Overall model accuracy.

standard 5-fold cross-validation. However, because adjacent 50-m sections may be spatially correlated, these estimates may be optimistic for independent road networks. Accordingly, the present validation does not establish transferability to independent road networks; grouped or spatial cross-validation is required for that assessment. Learning curves confirmed convergence between training and validation performance as the dataset size increased, while validation curves showed that optimal performance was achieved with low regularization values, confirming appropriate model complexity. These results support the internal stability of the model for subsequent interpretability analysis.

### 5.2. SHAP-based interpretation of class-1 prediction

In our research, understanding why the model makes a certain prediction is as crucial as the prediction's accuracy for its practical application. Statistical Generalized Linear Models (GLM), for example, can use their coefficients as a metric for the overall importance of each feature, but the coefficient cannot account for the local importance of the feature or their interactions. The same can be said for the feature importance of machine learning models, which is why SHapley Additive exPlanations (SHAP) is used for interpreting models. SHAP is grounded in the Shapley value framework from cooperative game theory (Shapley, 1953) and was later formalized as a unified approach for feature attribution in machine learning models (Lundberg and Lee, 2017). SHAP can be applied to any machine learning model to produce consistent explanations (Lundberg, 2020). It helps understand how different features affect model predictions, especially when they interact (Lundberg and Lee, 2017; Lundberg, 2020). More specifically, SHAP values quantify the contribution of each feature to the difference between the prediction for a given instance and a reference or baseline prediction, typically the expected model output over the dataset (Lundberg and Lee, 2017). Positive SHAP values indicate that the feature contributes to increasing the prediction, while negative SHAP values indicate a contribution to decreasing the prediction (Lundberg and Lee, 2017).

In this research, we performed SHAP analysis on the output of AutoML-LGBM, to explain the effect of features on the prediction of the lower class Q1. In the SHAP plots, the value indicates how much a particular feature has influenced the model: negative values indicate the feature decreases the model's output, while positive values indicate an increase. The red color indicates high feature values, and blue indicates low feature values compared to the average for the entire population. Each value is plotted as a dot on the chart. By aggregating local SHAP explanations across all samples, SHAP can also be used to characterize global model structure while preserving local faithfulness (Lundberg, 2020). Two types of SHAP plots are used: the **SHAP summary plot** uses violin shapes to represent the distribution of SHAP values for each feature, providing a visual summary of the feature's impact across the dataset; **SHAP Dependence Plot** shows the relationship between a feature and SHAP values, revealing potential interactions between features.

#### 5.2.1. Global feature effects

To display a summary of how each feature impacts the model's output, Fig. 6 shows the SHAP Beeswarm Chart for class 1. The features are sorted by importance, and the feature values are represented on a color scale where red indicates high values and blue indicates low values relative to the population average. The highest SHAP positive effect on class 1 prediction is observed for lower values of Speed, RL, and lane width, and for higher curvature. A clear separation of positive effects is observed for Daylight and Rain conditions. Meanwhile, shoulder width doesn't show a clear trend in its relationship with the outcome. Headway and section typology are the least relevant features in the model.

Fig. 6 presents the SHAP summary plot for the critical low-quality condition (Class 1). The global feature attribution indicates that Speed, RL, Lane, Shoulder, Weather, Light, and Curvature were the most influential variables in the fitted AutoML-LGBM model. Headway and Section showed negligible SHAP contribution, which is consistent with the capped Mobileye headway output and the sparse occurrence of specific section types in the real-world sample; therefore, they were not retained as focal variables in the subsequent DML and SHAP-dependence analyses. Low RL values, narrower lanes, night conditions, rainy weather, and higher curvature generally contributed toward Class 1 prediction, which is consistent with the expected degradation of vision-based lane-support performance under reduced marking visibility, constrained geometry, and adverse environmental conditions.

The high ranking of Speed requires a more cautious interpretation. The SHAP plot indicates that lower speeds were associated with increased Class 1 probability, which may appear counterintuitive because previous studies have suggested that LSS performance may decrease at speeds higher than 120 km/h (García et al., 2020) or remain

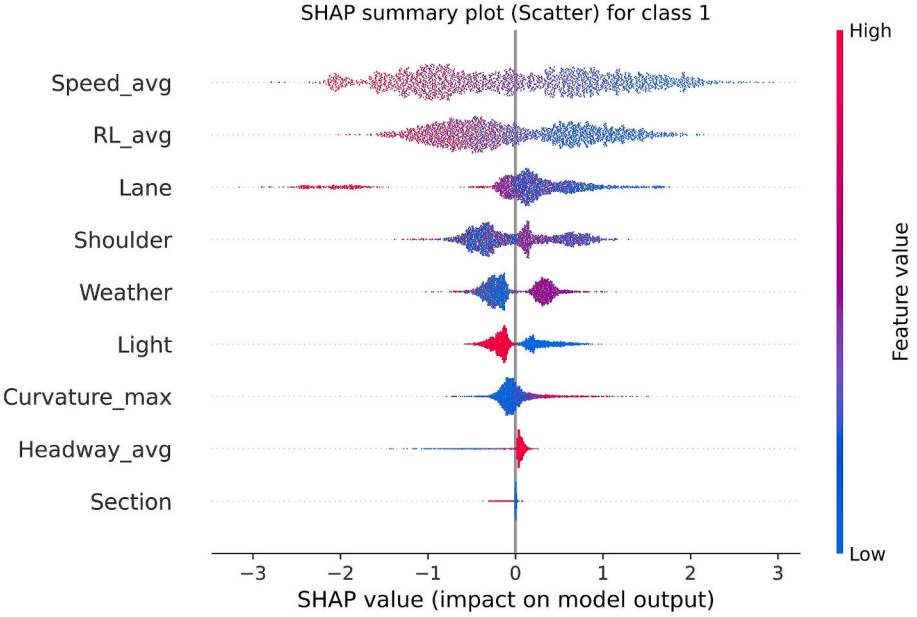



**Fig. 6.** SHAP summary plot (beeswarm) for the critical condition (Class 1) (Light: Day = 1, Night = 0; Env.: Dry = 1, Rain = 2, Wet = 3; Section: 1 = Tunnel, 2 = Bridge, 3 = Junction, 0 = Other).

relatively insensitive to speed within the tested range below 100 km/h (Cafiso and Pappalardo, 2020). Moreover, speed was not an experimentally controlled factor in the field campaign; it mainly reflected the operational driving context. Lower speed may therefore act as a proxy for more complex road and environmental conditions, including rainy weather, narrower lanes, sharper curves, and shorter headways. Consequently, the SHAP importance of speed should not be interpreted as evidence that lower speed causes poorer LSS performance; rather, it may partly reflect the more difficult road and environmental contexts in which lower speeds occurred.

Because SHAP values explain the behavior of the fitted model and can be sensitive to correlated predictors, an additional Double Machine Learning (DML) causal analysis was conducted before interpreting the detailed SHAP dependence and interaction plots.

SHAP feature importance and causal estimation in Double Machine Learning (DML) address fundamentally different questions: SHAP asks "What information did my model use to make a prediction?" while Double ML asks, "What is the covariate-adjusted causal effect of a specific feature?". The core difference is that SHAP measures the contribution of each feature to the model's output, whereas Double ML adjusts for the observed confounding relationships to estimate the causal mechanism of the feature (often called "the treatment").

#### 5.2.2. DML-based causal effect estimation

DML combines flexible machine-learning models for high-dimensional nuisance relationships with orthogonalized effect estimation, allowing adjusted treatment effects to be estimated while reducing sensitivity to nuisance-model error (Brathwaite and Walker, 2018; Chernozhukov et al., 2018; Wager and Athey, 2018). Conceptually, DML partials out the covariate contributions from both the outcome and the treatment and then estimates the residual-on-residual relationship using an orthogonal estimating equation. Cross-fitting is used to limit overfitting bias by estimating nuisance predictions for each observation from models trained on separate folds (Chernozhukov et al., 2018). Under regularity conditions and sufficiently accurate nuisance models, this orthogonalized construction is associated with favorable large-sample properties (Knaus, 2022; Farbmacher et al., 2022). In the present application, causal interpretation additionally relies on conditional exchangeability given the included road, traffic, and environmental covariates, adequate overlap in the focal variables, and consistency of the measured exposures and outcome.

In this study, the DML outcome variable was the lane-marking quality score (1–2-3), with higher values indicating better lane-marking quality. Because the numerical distances between the ordinal categories are not necessarily equal, treating this score as continuous is an approximation. Accordingly, the DML estimates are interpreted as directional average shifts in the expected quality score rather than effects on transitions between specific classes. Therefore, positive DML effects indicate an increase in the expected Mobileye quality score, while negative effects indicate a shift toward lower quality. Continuous variables, including RL, vehicle speed, curvature, lane width, and shoulder width, were robustly standardized using the interquartile range, so their effects are interpreted as the expected change in the quality score associated with a + 1 IQR (Inter Quartile Range) increase in the corresponding variable. Categorical factors were evaluated through binary contrasts, including Night versus Day and three weather contrasts: Wet versus Dry, Rain versus Wet, and Rain versus Dry. Given the previously identified better performance, LightGBM regressors were used as flexible nuisance models for both the outcome and treatment equations, and a linear final-stage model was used to estimate the adjusted marginal effects. Statistical uncertainty was estimated using bootstrap resampling. Headway and Section were not analyzed as focal treatment effects because they showed negligible SHAP contribution and limited variability in the data sample.

The results generally confirm the main directions observed in the SHAP analysis, while resolving concerns that the SHAP ranking could be affected by feature collinearity (Fig. 7). The strongest positive effect (0.25) was observed for vehicle speed, indicating that lower speeds, in the range of the experiment (i.e. 40–100 km/h) are associated with lower Mobileye quality score. The persistence of this association after adjustment for the observed covariates supports its relevance within the observed operating context; however, speed should be interpreted as an operational and contextual variable rather than as an isolated causal factor.

The DML results also confirmed the positive role of lane-marking retroreflectivity. A + 1 IQR increase in RL was associated with an increase in the expected Mobileye quality score, supporting the interpretation that higher marking visibility reduces the tendency toward the critical low-quality class. Among the other road factors, lane width showed a positive adjusted effect (+0.05) with higher quality scores expected for wider lane widths, and curvature showed a negative adjusted effect (−0.03), consistent with the SHAP interpretation and with the expected sensitivity of camera-based lane-support systems to road geometry. Shoulder width showed a small positive effect, but its confidence interval was close to zero, indicating no clear adjusted linear effect on detection quality.

The categorical contrasts further confirmed the role of environmental conditions. Rain showed the strongest negative effect when compared with Dry, and it also had a negative adjusted effect when compared with Wet. By contrast, the Wet versus Dry contrast was close to zero and its confidence interval crossed zero, indicating that wet pavement alone did not show a stable adjusted effect in the DML analysis. Night conditions also showed a negative adjusted effect compared with Day. These findings support the interpretation that active rain and nighttime conditions are important environmental ODD constraints for camera-based LSS.

Overall, the DML robustness analysis supports the main SHAP-derived conclusions: higher RL, higher speed, and wider lanes are associated with improved Mobileye quality scores, whereas rain, night conditions, and higher curvature are associated with lower expected quality. This strengthens the interpretation that the SHAP results reflect consistent operational patterns rather than only attribution instability under correlated predictors, supporting a further analysis of interaction patterns.

#### 5.2.3. Feature interaction patterns

The SHAP dependence plot is a visualization that relates the observed value of a feature to its SHAP contribution, thereby allowing both the main effect of that feature and possible interaction patterns with other variables to be examined (Lundberg, 2020). Unlike partial dependence plots, which only illustrate mean effects, SHAP dependence plots capture both mean effects and variations, indicating interactions. These plots visualize how changes in a feature's value affect the model's predictions. The x-axis represents the feature of interest, while the y-axis shows the SHAP values. Point colors represent the values of another feature, highlighting feature interactions. To limit the number of combinations, we selected the features most representative of the ODD factors in the road scene field that can be improved by both design (i.e., horizontal alignment, lane, and shoulder width) and maintenance (i.e., Marking RL). The mutual interaction of these features and their interaction with other features in the dynamic (i.e., speed, headway) or environmental (i.e., light, weather) fields of ODD are also analyzed because they could be considered for driving mitigation actions or the prediction of takeover driving control by the vehicle's systems in critical environmental conditions.

The SHAP dependence plots in Fig. 8 show the mean effect variation and interaction of the different features for class 1 using as reference RL-avg and Curvature (a − b); Lane Width (c − d); Shoulder Width (e − f); Speed (g − h); Light (i − j); Weather (k − l).

#### 5.2.4. ODD interpretation and practical Implications

The results observed in the SHAP dependence plot are discussed

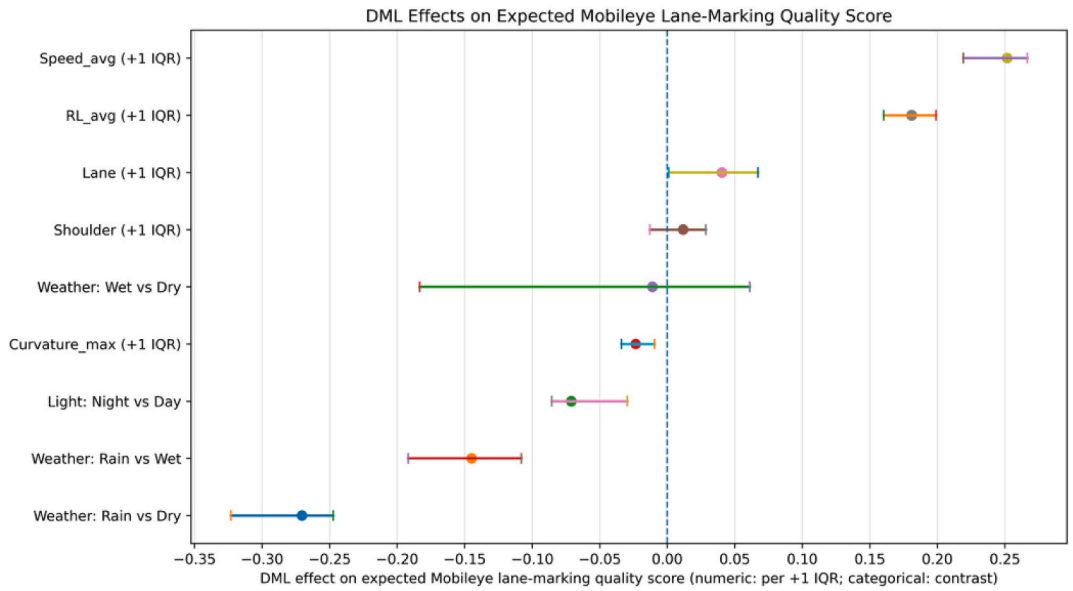



**Fig. 7.** DML-adjusted effects on expected Mobileye lane-marking quality score. For continuous variables, effects are reported per + 1 IQR increase. For categorical variables, effects are reported as binary contrasts. Positive values indicate an increase in the expected Mobileye quality score, while negative values indicate a shift toward lower expected quality.

below:

**RL-avg:** The role of RL was examined across the SHAP dependence plots in Fig. 8, particularly in plots a, c, e, g, i and k. These plots consistently show that low RL values contribute to the occurrence of low-quality lane-marking detection. The strongest positive SHAP contributions are generally observed when RL is below approximately 120 mcd/(m$^2$·lx), indicating that under this value retroreflectivity consistently increases the probability of Class-1 prediction. This value is already higher than the critical threshold of 80 mcd/(m$^2$·lx) commonly considered in maintenance standards for human visibility (Highways England, 2020), suggesting that camera-based LSS may require higher marking visibility than conventional human-oriented minimum requirements. The plots also indicate that the transition from positive to neutral or negative SHAP contribution occurs gradually, mainly within the 120–150 mcd/(m$^2$·lx) range. Therefore, RL should be interpreted here as a system-specific transition factor rather than through a single universal threshold. The following interaction analyses further examine how this RL contribution changes under different geometric, traffic, lighting, and weather conditions.

**Curvature & RL:** There is a clear effect of curvature to increase the Q1 probability for values higher than 0.006-0.007 (i.e., radius less than 150 m) (Fig. 8**b**). Increasing RL doesn't mitigate the effect of curvature. In curves with a sharp radius, the lane marking can fall out of the camera's field of view, which is limited to avoid optical distortion. Multi-camera and sensor fusion are adopted in the most advanced AVs to solve this issue.

Lane width & RL: Lane widths narrower than 3.5 m have a small increasing effect on class 1 prediction, which becomes particularly significant for widths lower than 3.2 m (Fig. 8d). Increasing RL reduces the probability of class 1 in narrow lanes (Fig. 8c).

**Shoulder width & RL:** For very narrow shoulders with a width of less than 0.25 m, the role of RL is revealed to be important in guaranteeing good system performance, as high class 1 predictions are always associated with minimum RL values (Fig. 8**f**). Shoulder widths in the range of 0.25–0.65 m showed the greatest impact, with limited effect from RL, whereas for wider shoulders, lane edge detection showed better performance across a wider range of RL. Because marking detection is strongly related to the pavement/marking contrast ratio (Cafiso and Pappalardo, 2020; Austroads, 2021; Pike et al., 2018; Davies, 2017). For shoulder widths less than 0.25 m, the pavement edge is very close to the marking border, which appears to improve LSS performance because detection is based on pixel-intensity variation beyond the pavement surface. Similarly, the most critical condition for widths between 0.25 and 0.62 m is the lack of sufficient paved shoulder width to provide uniform contrast between the marking and pavement on the external side.

**RL-avg & Speed:** Fig. 8g–h shows that lower speeds are associated with higher Class-1 SHAP values, especially when RL is also low. This unexpected result can be explained because Lane Departure Warning (LDW) and Lane Keep Assist (LKA) were fundamentally designed to prevent fatal, high-speed single-vehicle or head-on crashes. Therefore, they are optimized for the long, sweeping curves and constant speeds of highways with a minimum speed threshold of 50–60 km/h (e.g. Acura (45 mph), Honda (45 mph), Hyundai (40 mph), Chevrolet (37 mph), Audi 65 km/h), Kia (40 mph), Volvo (65 km/h)) to prevent erratic steering behaviors in dense, unpredictable environments. However, speed was not experimentally controlled and should be interpreted as an operational/contextual variable.

**RL & Light:** Night conditions made LSS performance more sensitive to RL quality (Fig. 8j), whereas in daylight conditions, LSS performance was significantly higher (Fig. 8i & j). Night driving remains the most challenging condition for both human and automated driving.

**RL & Weather:** Rain conditions are critical for marking detection, also with high values of RL (Fig. 8l). The thickness of the water film and water ponding can cover the marking, reducing the camera vision system's ability to detect pixel intensity. Compared to dry conditions, wet pavement improves marking/pavement contrast and, consequently, LSS performance (Fig. 8l).

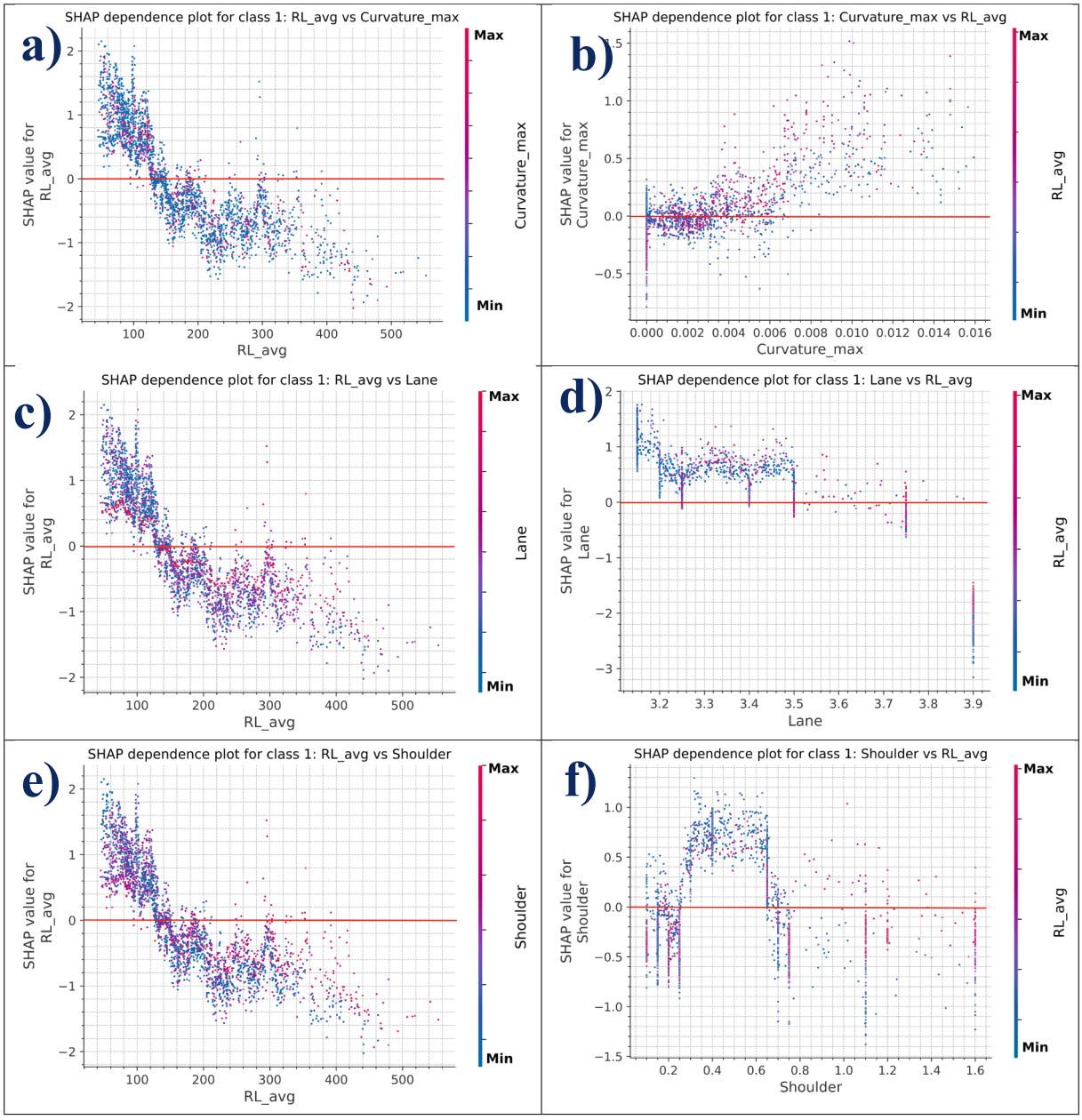



**Fig. 8.** SHAP Dependence Plots for critical system condition based on RL.

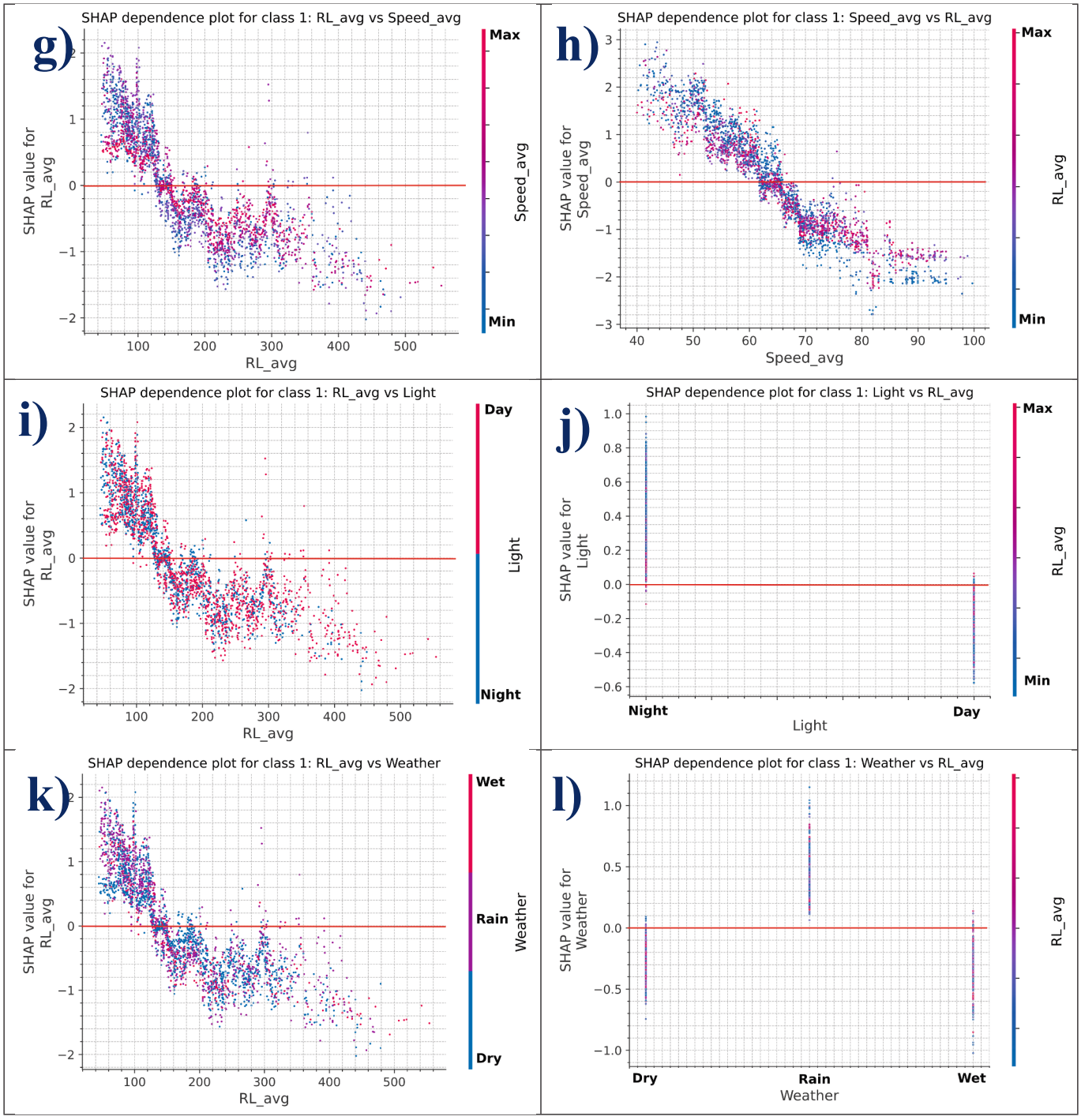



**Fig. 8.** (*continued*).

**Table 2**
SHAP effects on Class-1 risk under good vs poor retroreflectivity.

| Critical factor (range) | RL− (<120) | RL+ (≥150) |
|---|---|---|
| Curvature | | |
| R ≤ 150 m (~0.006–0.007 1/m) | (↑) | (↑) |
| 150 m < R < 250 m (~0.004–0.006) | (↑) | (=) |
| R ≥ 250 m (≤~0.004 1/m) | (=) | (=) |
| Lane width | | |
| <3.2 m | (↑) | (↑)/ (=) |
| 3.2–3.5 m | (=) / (↑) | (=) / (↑) |
| ≥3.6–3.9 m | (↓) | (↓) |
| Shoulder width | | |
| <0.25 m | (↑) | (↓) |
| 0.25–0.65 m (weak band) | (↑) | (↑) |
| >0.65 m | (↓)/ (=) | (↓) |
| Light | | |
| Night | (↑) | (↓) |
| Day | (=) / (↓) | (↓) |
| Weather | | |
| Rain | (↑) | (↑) |
| Wet (no rain) | (↓) | (↓) |
| Dry | (=) | (↓) |

Symbols: (↑) increase Q1 classification; (=) neutral effect; (↓) decrease Q1 classification

Overall, the SHAP analysis was very useful to demonstrate the combined effect of the different features with RL. From a practical standpoint, the potential benefits of improving the performance of lane markings to mitigate the adverse effects of features less suited for adaptation to AVs (e.g. road geometry, weather, and lighting) are particularly noteworthy.

To clarify these interactions for ODD interpretation, Table 2 summarizes the SHAP results for each factor, showing how it is affected by good (RL+) and poor (RL−) marking quality. The table uses (↑) (positive SHAP) to denote increased probability of low marking detection (Class-1), (↓) (negative SHAP) to denote reduced probability of Class-1, and (=) for negligible/unclear effect. Based on the SHAP dependence plots in Fig. 8 and the conditional RL analysis, RL values below approximately 120 mcd/(m$^2$·lx) are interpreted as poor retroreflectivity values (RL−), while values at or above approximately 150 mcd/(m$^2$·lx) are interpreted as good retroreflectivity values (RL+). The intermediate 120–150 mcd/(m$^2$·lx) interval is treated as a transition range rather than a sharp threshold. As highlighted in Table 2, sharp horizontal curvature with R ≤ 150 m remains a hard ODD boundary under both RL levels, whereas good retroreflectivity partly reduces the adverse contribution of several infrastructure and environmental conditions, including night driving, narrow lanes, and very narrow or wide shoulders.

**Light & Weather combined scenarios:** To further investigate the joint role of illumination and atmospheric conditions, SHAP dependence plots were generated crossing Light and Weather (Fig. 9a,b). This analysis shows that Night and Rain are the most adverse conditions, with consistent positive SHAP values that increase the probability of critical class-1 detections. In contrast, Day and Wet (no rain) consistently show negative SHAP values, indicating lower contribution toward Class 1. Mixed cases reveal nuanced patterns: Night–Dry still shows an adverse contribution due to the night effect, while Night–Wet shows moderate uplift (night positive, wet negative).

To further examine how retroreflectivity behaves under the combined Light–Weather scenarios, an additional conditional SHAP dependence analysis was conducted for RL. Unlike the previous Light–Weather plot, which focused on the environmental variables themselves, this analysis evaluates the total SHAP contribution of RL within each environmental scenario. The same global Class-1-vs-rest model was used, and the SHAP values were visualized separately for the six Light–Weather combinations. Therefore, Fig. 10 should be interpreted as a scenario-conditioned visualization of the RL contribution, not as six independent models or as a formal third-order SHAP interaction.

Fig. 10 shows that the role of RL is directionally consistent across all Light–Weather scenarios. At low RL values, the total SHAP contribution of RL is generally positive, indicating that poor retroreflectivity contributes toward the critical Class-1 prediction. As RL increases, the SHAP contribution decreases and becomes negative, meaning that higher RL shifts the model away from the critical low-quality class. The transition from positive to negative contribution occurs mainly within the approximate 120–150 mcd/(m$^2$·lx) range, although the exact shape and smoothness of the curve vary across scenarios.

This result supports the interpretation that RL acts as a mitigable

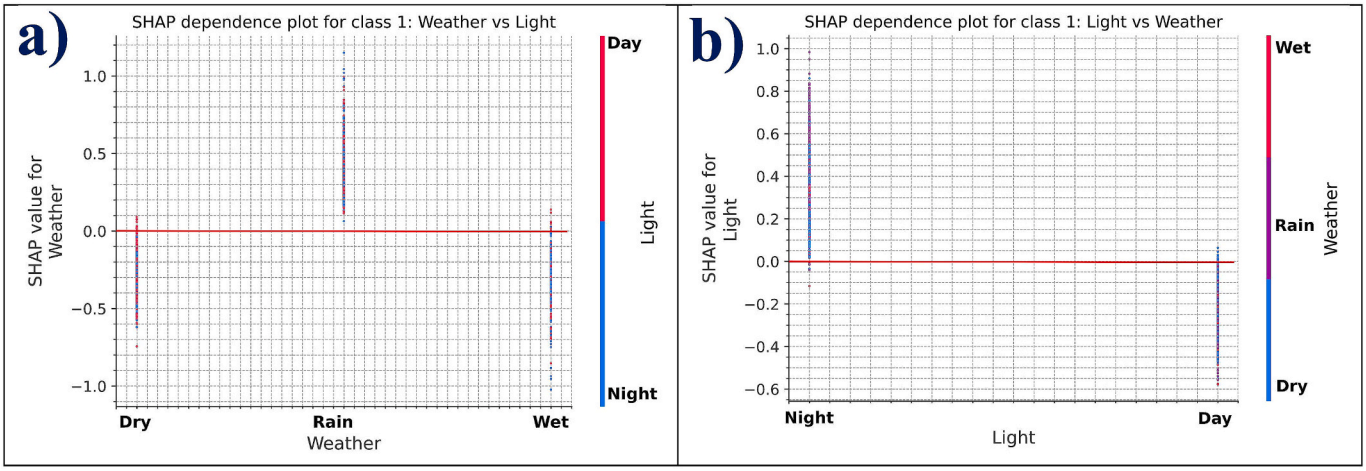



**Fig. 9.** SHAP Dependence Plots for critical environmental conditions.

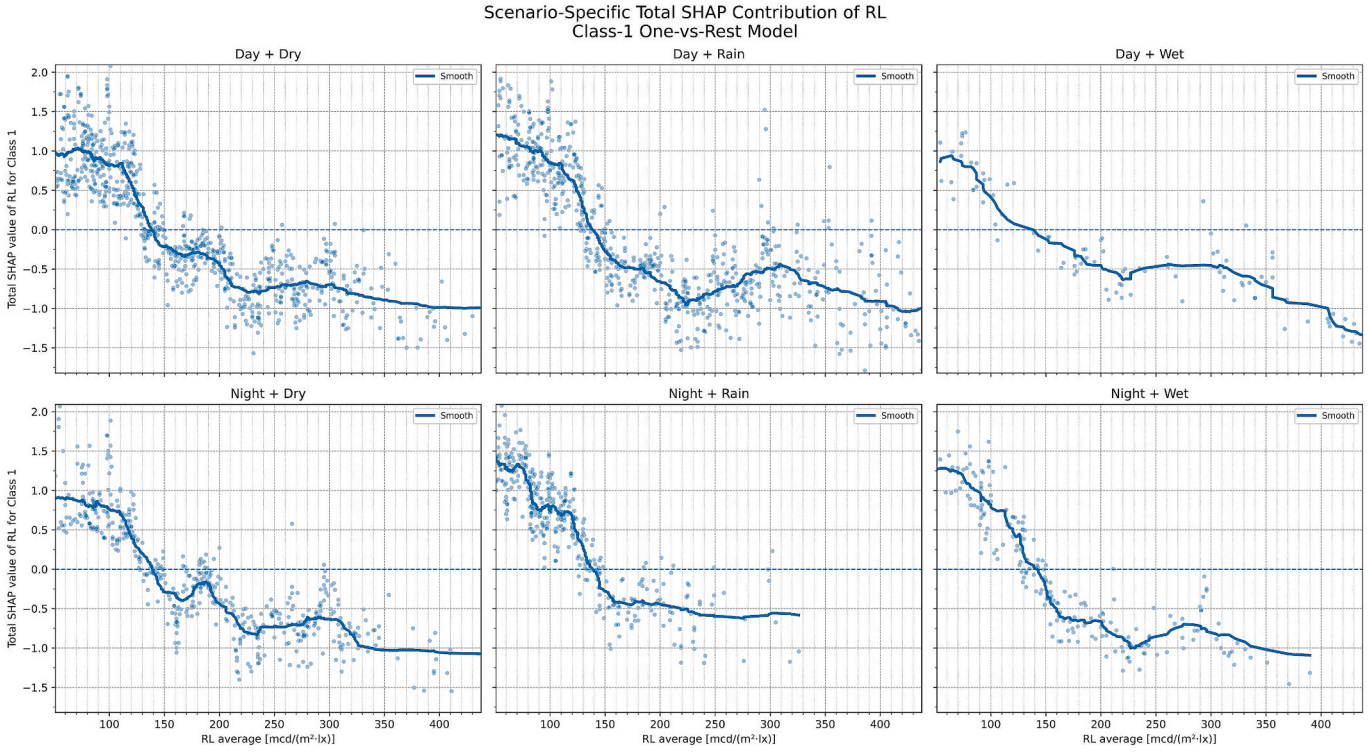



**Fig. 10.** Scenario-specific total SHAP contribution of RL for Class-1 prediction under Light–Weather combinations. Positive SHAP values indicate that RL contributes toward the critical Class-1 outcome, while negative values indicate contribution away from Class 1. The same global Class-1-vs-rest model was used, and SHAP values were conditionally visualized by environmental scenario.

**Table 3**
SHAP scenario analysis for Light–Weather combinations (class-1 risk).

| Condition | Overall effect | Light effect | Weather effect |
|---|---|---|---|
| Day + Dry (DD) | (=) | (=) | (=) |
| Day + Rain (DR) | (↑) | (=) | (↑) |
| Day + Wet (DW) | (↓) | (=) | (↓) |
| Night + Dry (ND) | (=) | (↑) | (↓) |
| Night + Rain (NR) | (↑) | (↑) | (↑) |
| Night + Wet (NW) | (=) | (↑) | (↓) |

Symbols: (↑) increase Q1 classification; (=) neutral effect; (↓) decrease Q1 classification.

infrastructure-related factor across different environmental conditions. Even under adverse conditions such as Night + Rain and Night + Wet, increasing RL reduces the contribution of retroreflectivity toward Class-1 prediction. However, this does not imply that high RL completely removes the adverse effect of rain or night conditions, since these conditions can still influence the final prediction through their own SHAP contributions and interactions with other road and traffic variables. Therefore, the conditional RL analysis reinforces the distinction between mitigable ODD constraints, such as insufficient marking retroreflectivity, and harder environmental constraints, such as active rain and poor illumination.

Table 3 summarizes these scenario effects, distinguishing the individual contributions of Light and Weather and their combined impact on marking detection.

## 6. Conclusions

This study evaluated the Operational Design Domain (ODD) of a camera-based Lane Support System (LSS) using direct Mobileye 6.0 lane-detection quality outputs integrated with lane-marking retroreflectivity (RL), road geometry, traffic variables, and environmental conditions. The field campaign covered all combinations of two lighting conditions and three weather conditions, while road infrastructure variables were observed across a heterogeneous sample of two-lane rural roads. The final dataset included 9351 road sections of 50 m, representing 468 km of roadway and covering the main ODD taxonomy fields of physical infrastructure, environmental conditions, and traffic-related dynamic conditions.

The modeling framework compared statistical and machine-learning approaches for classifying Mobileye lane-marking quality. CatBoost was used in the preliminary stage to support descriptor selection and identify the most informative variables among the available 50-m section statistics. The mode value of the Mobileye quality class (Qmode) was selected as the target variable because Class 1 in Qmode accounted for 98.05% of system faults in the dataset. In the second stage, several statistical and machine-learning models were compared. The best performance was obtained using an AutoML-LightGBM pipeline with SMOTE-based class-imbalance treatment, particularly for the critical low-detection Class 1. The modeling results were used primarily to support ODD interpretation rather than to claim deployment-level prediction on independent road networks. The intermediate Class 2 was more difficult to classify, reflecting its transitional nature between clearly low and clearly high lane-detection quality.

SHAP analysis was used to interpret the fitted model and identify the road, traffic, and environmental conditions contributing to Class-1 prediction. Because SHAP attributions can be sensitive to correlated predictors, a Double Machine Learning (DML) robustness analysis was added to estimate adjusted feature effects while controlling for the remaining covariates. The DML results supported the main SHAP-

derived directions: higher RL, higher speed, and wider lanes were associated with higher expected Mobileye quality scores, whereas rain, night conditions, and higher curvature were associated with lower expected quality. Speed should therefore be interpreted as an operational/contextual variable rather than as an isolated causal factor, since it was not experimentally controlled and reflects broader road, traffic, and environmental conditions.

The main findings from the SHAP dependence plots, DML analysis, and conditional RL interpretation can be summarized as follows:

- **RL-avg:** Low RL values consistently contributed toward critical Class-1 prediction. The strongest adverse contribution was observed below approximately 120 mcd/(m$^2$·lx), while the transition from positive to neutral or negative contribution generally occurred within the 120–150 mcd/(m$^2$·lx) range. This range is higher than the 80 mcd/(m$^2$·lx) value commonly considered in maintenance standards for human visibility (Highways England, 2020), suggesting that camera-based LSS may require higher marking visibility than conventional human-oriented minimum requirements. However, the 120–150 mcd/(m$^2$·lx) interval should be interpreted as a system-specific empirical transition zone for the tested Mobileye 6.0 LSS and road context, not as a universal maintenance threshold.
- **Curvature:** High curvature, particularly for radii lower than approximately 150 m, increased the tendency toward Class-1 prediction. Increasing RL did not fully mitigate the adverse contribution of sharp curvature, indicating that curvature represents a harder ODD constraint for camera-based lane detection. This limitation is likely related to camera field-of-view restrictions and the difficulty of maintaining lane-marking visibility in sharp curves. Vehicle-side improvements, such as multi-camera systems, extended field of view, or sensor fusion, may therefore be needed for such conditions.
- **Lane width:** Narrow lanes, especially below approximately 3.2 m, increased the contribution toward Class 1. In practice, lower RL, narrow lanes, and limited shoulder width often occur together on lower-category rural roads, increasing the probability of LSS detection difficulties. Unlike sharp curvature, however, the adverse contribution of narrow lanes was partly reduced when RL was higher, indicating that marking maintenance can partially mitigate this constraint.
- **Shoulder width:** SHAP dependence patterns suggested that shoulder widths in the range of approximately 0.25–0.65 m showed more critical behavior, especially when combined with low RL, although shoulder width did not show a clear independent adjusted effect in the DML analysis. Very narrow shoulders and intermediate shoulder widths can affect the pavement–marking contrast available to the camera system. In these conditions, maintaining adequate RL and ensuring a clear contrast between the marking and the surrounding pavement surface are particularly important.
- **Speed:** The SHAP summary plot showed that lower speeds contributed toward Class-1 prediction, which may appear counterintuitive compared with previous studies reporting performance degradation at higher speeds or limited speed relevance below 100 km/h. The DML analysis clarified this interpretation by showing that higher speed was associated with higher expected Mobileye quality after adjustment. Therefore, speed should not be interpreted as an independently controlled causal factor. Instead, it reflects the operational context in which lower speeds often occur, including adverse weather, sharper curvature, narrower lanes, or other driving constraints.
- **Weather:** Active rain was one of the most adverse environmental conditions for camera-based marking detection. Rain can reduce visibility through water film, water spray, and water ponding, which may cover the marking and reduce pixel-intensity contrast. Wet pavement without active rain showed a different behavior and, in some cases, improved marking/pavement contrast. However, both rain and wet pavement can remain critical for other sensing technologies, including LiDAR-based systems, making adverse pavement-surface and weather conditions relevant also for higher automation levels (Ye et al., 2025). Increasing RL can partly reduce the contribution toward Class 1, but it cannot fully compensate for active rain; pavement surface regularity and drainage remain essential for reducing water film thickness and ponding.
- **Light:** Night conditions increased the sensitivity of the system to RL quality, whereas daylight conditions generally reduced the contribution toward Class-1 prediction. The conditional Light × Weather analysis confirmed that Night + Rain is one of the most critical environmental combinations. At the same time, the scenario-specific RL analysis showed that higher RL reduces the RL-related contribution toward Class 1 even under Night + Rain and Night + Wet conditions. This indicates that insufficient retroreflectivity is a mitigable infrastructure-related constraint, while active rain and poor illumination remain broader environmental ODD constraints.

Overall, the results clarify the distinction between mitigable and hard ODD constraints for camera-based LSS. Insufficient lane-marking retroreflectivity, narrow lanes, and some shoulder-width conditions can be partly mitigated through infrastructure maintenance and design improvements. In contrast, sharp curvature and active rain remain harder operational constraints that cannot be fully compensated for by increasing RL alone. These conditions may require vehicle-system improvements, including wider camera field of view, multi-camera perception, sensor fusion, improved algorithms, and better integration between vehicle perception and infrastructure design.

From an infrastructure perspective, maintaining adequate retroreflectivity, improving pavement drainage, preserving marking–pavement contrast, and ensuring sufficient lane and shoulder conditions can improve the operational reliability of camera-based LSS. From a vehicle-system perspective, sharp curves, night conditions, and adverse environmental conditions highlight the need for broader sensing strategies and more robust perception systems. These findings emphasize the importance of coordinated adaptation between road agencies and vehicle manufacturers to improve ADAS readiness and define practical ODD boundaries.

The study is limited to single-carriageway, two-lane rural roads and to one camera-based LSS implementation. Multilane carriageways may pose additional challenges related to higher speeds, greater variability in marking types, and more complex lane-changing behavior (Ye et al., 2025). Urban roads may present narrow lanes, complex and variable marking patterns, and markings designed more for traffic control than lane keeping (García and Camacho-Torregrosa, 2020). In such environments, lane keeping may need to rely more strongly on high-accuracy localization, HD maps, and sensor fusion rather than on lane-marking detection alone.

The reported RL transition range and feature interactions are specific to the tested LSS, the investigated road network, and the observed pavement, marking, and environmental conditions. They may vary across vehicle models, camera systems, pavement materials, marking types, rain intensity, and road categories. In addition, adjacent 50-m sections may share similar geometry, marking quality, and environmental characteristics; therefore, future validation should include spatial analysis and grouped or spatial cross-validation by route, survey run, or independent road network, including approaches that adjust existing models or develop novel models for spatially dependent ordinal data (Xu et al., 2020). Future studies should also extend the analysis to multilane roads, urban roads, different sensor technologies, and larger datasets collected from multiple vehicle platforms. Since large-scale ADAS datasets are costly to collect, collaboration with vehicle manufacturers and road agencies will be essential for developing transferable ODD assessment frameworks.

## CRediT authorship contribution statement

**Omid Ghaderi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation. **Giuseppina Pappalardo:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Alessandro Di Graziano:** Writing – review & editing, Methodology. **Giovanni Andrea Dimauro:** Methodology, Data curation. **Xuesong Wang:** Writing – review & editing, Methodology. **Haneen Farah:** Writing – review & editing, Methodology. **Salvatore Cafiso:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

## Funding

This paper was partially supported by the European Union (Next-Generation EU), National Centre for HPC, Big Data and Quantum Computing, Mission 4 Component 2, Investment 1.4, CUP E63C22001000006, and by PRIN 2022 "Smart Road for Smart Drivers toward Smart Vehicles – Smart RoDriVe", CUP E53D23010410006.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The datasets generated and/or analyzed in this study are not publicly available due to a university non-disclosure agreement, but may be obtained from the project administrator on reasonable request.

## References

Agresti, A., 2010. Analysis of Ordinal Categorical Data, vol. 656. John Wiley & Sons.
Austroads, "ATS 4110 - Longitudinal Pavement Marking: Austroads Technical Specification," 2021.
Babić, Darko, Dario Babić, Mario Fiolić, Arno Eichberger, Zoltan Ferenc, Magosi, 2021. A Comparison of Lane Marking Detection Quality and View Range between Daytime and Night-Time Conditions by Machine Vision. Energies 14 (15), 4666. https://doi.org/10.3390/en14154666.
Brathwaite, T., Walker, J.L., 2018. Causal inference in travel demand modeling (and the lack thereof). Journal of Choice Modelling 26, 1–18.
S. Cafiso, H. Farah, O. Ghaderi, and G. Pappalardo, "Experimental analysis of road characteristics impact on the performance of lane support system," vol. 63, pp. 315–324, 2024, doi: 10.53136/979122181262621.
Cafiso, S., Pappalardo, G., 2020. Safety effectiveness and performance of lane support systems for driving assistance and automation – Experimental test and logistic regression for rare events. Accid. Anal. Prev. 148, 105791. https://doi.org/10.1016/j.aap.2020.105791.
Chai, C., Liu, T., Liu, J., Cao, X., Hu, Y., Li, Q., 2023. Surrogate evaluation model for assessing lane detection reliability of automated vehicles in complex road environments. Transp. Res. Rec. 2677 (7), 185–195.
Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.
V. Chernozhukov et al., "Double/debiased machine learning for treatment and structural parameters," 2018, Oxford University Press Oxford, UK.
W. Chu, Z. Ghahramani, and C. K. I. Williams, "Gaussian processes for ordinal regression.," Journal of machine learning research, vol. 6, no. 7, 2005.
C. Davies, "Effects of pavement marking characteristics on machine vision technology," 2017.
A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: gradient boosting with categorical features support," arXiv preprint arXiv:1810.11363, 2018.
Elizebeth, M.J., Khastgir, S., Jennings, P., 2025. Hazard analysis of an automated lane keeping system using systems-theoretic process analysis. Accid. Anal. Prev. 221, 108171.
EN 1436. Road Marking Materials Road Marking Performance for Road Users and Test Methods, Milano, Italy, 2018. Accessed: Feb. 12, 2023. [Online]. Available: https://store.uni.com/uni-en-1436-2018.
H. Farah et al., "An Empirical Analysis to Assess the Operational Design Domain of Lane Keeping System Equipped Vehicles Combining Objective and Subjective Risk Measures," IEEE Transactions on Intelligent Transportation Systems, vol. 22, pp. 2589–2598, 2020, [Online]. Available: https://api.semanticscholar.org/CorpusID:263876355.
Farbmacher, H., Huber, M., Laffers, L., Langen, H., Spindler, M., 2022. Causal mediation analysis with double machine learning. Econom. J. 25 (2), 277–300.
Fernández, A., Garcia, S., Herrera, F., Chawla, N.V., 2018. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. J. Artif. Intell. Res. 61, 863–905.
García, A., Camacho-Torregrosa, F.J., Baez, P.V.P., 2020. Examining the effect of road horizontal alignment on the speed of semi-automated vehicles. Accid. Anal. Prev. 146, 105732.
García, A., Camacho-Torregrosa, F.J., 2020. Influence of lane width on semi-autonomous vehicle performance. Transp. Res. Rec. 2674 (9), 279–286.
He, H., Garcia, E.A., 2009. Learning from Imbalanced Data. IEEE Trans. Knowl. Data Eng. 21 (9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239.
Highways England, "DMRB: Road Layout CS 126 Inspection and assessment of road markings and road studs (formerly TD 26/17)," Birmingham, 2020.
Hornung, R., 2020. Ordinal Forests. J. Classif. 37 (1), 4–17. https://doi.org/10.1007/s00357-018-9302-x.
Hutter, F., Kotthoff, L., Vanschoren, J., 2019. Automated machine learning: methods, systems, challenges. Springer Nature.
Janitza, S., Tutz, G., Boulesteix, A.-L., 2016. Random forest for ordinal responses: Prediction and variable selection. Comput. Stat. Data Anal. 96, 57–73. https://doi.org/10.1016/j.csda.2015.10.005.
Japkowicz, N., Shah, M., 2011. Evaluating learning algorithms: a classification perspective. Cambridge University Press.
G. Ke et al., "Lightgbm: A highly efficient gradient boosting decision tree," Adv. Neural Inf. Process. Syst., vol. 30, 2017.
D. G. Kleinbaum and M. Klein, "Ordinal Logistic Regression," in Logistic Regression: A Self-Learning Text, D. G. Kleinbaum and M. Klein, Eds., New York, NY: Springer New York, 2010, pp. 463–488. doi: 10.1007/978-1-4419-1742-3_13.
Knaus, M.C., 2022. Double machine learning-based programme evaluation under unconfoundedness. Econom. J. 25 (3), 602–627.
Lehoczký, P., Janeba, M., Galinski, M., Šoltés, L., Dec. 2022. Testing the Readiness of Slovak Road Infrastructure for the Deployment of Intelligent Transportation. Strojnícky Casopis - Journal of Mechanical Engineering 72, 103–112. https://doi.org/10.2478/scjme-2022-0020.
Leslie, A.J., Kiefer, R.J., Meitzner, M.R., Flannagan, C.A., 2021. Field effectiveness of general motors advanced driver assistance and headlighting systems. Accid. Anal. Prev. 159, 106275.
Liu, L., Sun, Q., Yang, L., Tian, Y.-C., Zhou, C., 2026. Enhanced verification of safety and security for advanced driver assistance systems. Reliab. Eng. Syst. Saf. 266, 111691. https://doi.org/10.1016/j.ress.2025.111691.
Lundberg, S.M., et al., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67.
S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Adv. Neural Inf. Process. Syst., vol. 30, 2017.
J. Marr, S. Benjamin, and A. Zhang, Implications of pavement markings for machine vision, no. AP-R633-20. 2020.
R. M. Neal, "Monte Carlo implementation of Gaussian process models for Bayesian regression and classification," arXiv preprint physics/9701026, 1997.
G. Pappalardo, S. Cafiso, A. Di Graziano, and A. Severino, "Decision Tree Method to Analyze the Performance of Lane Support Systems," Sustainability, vol. 13, no. 2, 2021, doi: 10.3390/su13020846.
G. Pappalardo, R. Caponetto, R. Varrica, and S. D. Cafiso, "Assessing the operational design domain of lane support system for automated vehicles in different weather and road conditions," Journal of Traffic and Transportation Engineering (English Edition), 2022, [Online]. Available: https://api.semanticscholar.org/CorpusID:248766235.
A. M. Pike, T. P. Barrette, and P. J. Carlson, "Evaluation of the Effects of Pavement Marking Characteristics on Detectability by ADAS Machine Vision," National Cooperative Highway Research Program (NCHRP): Washington, DC, USA, 2018.
L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," Adv. Neural Inf. Process. Syst., vol. 31, 2018.
Qin, D., Wang, X., Hassanin, O., Cafiso, S., Wu, X., 2022. Operational design domain of automated vehicles for crossing maneuvers at two-way stop-controlled intersections. Accid. Anal. Prev. 167, 106575.
N. Reddy, H. Farah, Y. Huang, T. Dekker, and B. van Arem, "Operational Design Domain Requirements for Improved Performance of Lane Assistance Systems: A Field Test Study in The Netherlands," IEEE Open Journal of Intelligent Transportation Systems, vol. 1, pp. 237–252, 2020, [Online]. Available: https://api.semanticscholar.org/CorpusID:228089891.
L. S. Shapley, "A value for n-person games," 1953.
S. A. E. Taxonomy, "Definitions for terms related to driving automation systems for on-road motor vehicles," SAE: Warrendale, PA, USA, vol. 3016, 2018.
The British Standards Institution (BSI), "PAS 1883:2020 Operational Design Domain (ODD) taxonomy for an automated driving system (ADS)–Specification," London, Aug. 2020.
The European Parliament and the Council of the European Union, "EC) No. 1224/2009. Regulation (EU) 2019/2144 of the European Parliament and of the Council," 2019. Accessed: May 13, 2023. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2019/2144/oj.
the International Organization for Standardization (ISO), "ISO 17361-2017 (E) Intelligent transport systems — Lane departure warning systems — Performance requirements and test procedures," Geneva, Switzerland, 2017. Accessed: Feb. 08, 2023. [Online]. Available: https://www.iso.org/standard/72349.html.
Wager, S., Athey, S., 2018. Estimation and inference of heterogeneous treatment effects using random forests. J. Am. Stat. Assoc. 113 (523), 1228–1242.

Wang, X., Qin, D., Cafiso, S., Liang, K.K., Zhu, X., 2021. Operational design domain of autonomous vehicles at skewed intersection. Accid. Anal. Prev. 159, 106241.
Z. Xu, S. B. Cannon, and W. D. Beavis, "Applications of spatial models to ordinal data," bioRxiv, pp. 2020–2029, 2020.
Ye, X., Wang, X., 2022. Operational design domain of automated vehicles at freeway entrance terminals. Accid. Anal. Prev. 174, 106776.
Ye, X., Wang, X., Cafiso, S.D., García, A., 2025. Readiness evaluation of freeways for lane-detection performance of Lidar-based automated vehicles: a field test analysis. IEEE Trans. Intell. Transp. Syst. 1–15. https://doi.org/10.1109/TITS.2025.3568591.