

Deep Learning-Based Road Damage Detection and Classification Using Computer Vision

Calvin Arseno^{*1}, Nadia Weningrum²

Email: carsento20034@gmail.com

^{1,2}Program Studi Arsitektur, Fakultas Teknik, Universitas Gunadarma, Depok, Indonesia, 16424

^{*}Corresponding Author

Article Information	Abstract
Keywords: Computer Vision; Damage Detection; YOLO; Object Detection; Road Damage.	<i>Road damage detection is essential for supporting efficient road infrastructure inspection, yet conventional visual assessment is time-consuming and may produce inconsistent results when dealing with diverse pavement distress patterns. This study aims to comparatively evaluate YOLOv8s, YOLOv10s, YOLOv11s, and an optimized YOLOv11s configuration for detecting and classifying longitudinal cracks, transverse cracks, alligator cracks, and potholes using the RDD2022 dataset. A quantitative experimental approach was employed using 55,007 annotated damage instances, with all models trained under consistent experimental conditions and evaluated using precision, recall, F1-score, mAP@50, mAP@50–95, and inference time. The results show a progressive improvement across the evaluated configurations, with optimized YOLOv11s achieving the highest performance at 83.4% precision, 80.9% recall, 82.1% F1-score, 88.4% mAP@50, and 59.6% mAP@50–95, with an inference time of 8.7 ms. Class-specific analysis showed that longitudinal cracks achieved the highest F1-score of 85.5%, whereas alligator cracks recorded the lowest at 79.7%, with the largest classification confusion occurring between potholes and alligator cracks. The novelty of this study lies in the systematic cross-generation comparison of YOLOv8s, YOLOv10s, YOLOv11s, and optimized YOLOv11s under consistent experimental conditions while jointly examining overall performance, class-specific behavior, classification errors, and inference efficiency. The study contributes a detailed characterization of model performance across heterogeneous road-damage categories and provides quantitative evidence for selecting an effective YOLO configuration for automated road infrastructure inspection.</i>
	Abstrak
Kata Kunci: Visi Komputer; Deteksi Kerusakan; YOLO; Deteksi Objek; Kerusakan Jalan.	Kerusakan jalan merupakan salah satu permasalahan penting dalam pemeliharaan infrastruktur transportasi karena proses inspeksi visual secara konvensional membutuhkan waktu dan tenaga yang besar serta dapat menghasilkan penilaian yang tidak konsisten pada berbagai pola kerusakan perkerasan. Penelitian ini bertujuan untuk membandingkan kinerja YOLOv8s, YOLOv10s, YOLOv11s, dan konfigurasi YOLOv11s yang dioptimalkan dalam mendeteksi dan mengklasifikasikan retak longitudinal, retak transversal, retak buaya, dan lubang pada dataset RDD2022. Penelitian menggunakan pendekatan eksperimen kuantitatif dengan 55.007 objek kerusakan teranotasi, sedangkan seluruh model dilatih menggunakan kondisi eksperimental yang konsisten dan dievaluasi berdasarkan precision, recall, F1-score, mAP@50, mAP@50–95, dan waktu inferensi. Hasil penelitian menunjukkan peningkatan kinerja secara bertahap pada konfigurasi yang dievaluasi, dengan YOLOv11s yang dioptimalkan menghasilkan kinerja tertinggi, yaitu precision 83,4%, recall 80,9%, F1-score 82,1%, mAP@50 88,4%, dan mAP@50–95 59,6%, dengan waktu inferensi 8,7 ms. Analisis berdasarkan kelas menunjukkan bahwa retak longitudinal memiliki F1-score tertinggi sebesar 85,5%, sedangkan retak buaya memiliki nilai terendah sebesar 79,7%, dengan kesalahan klasifikasi terbesar terjadi antara lubang dan retak buaya. Kebaruan penelitian ini terletak pada perbandingan sistematis

lintas generasi YOLOv8s, YOLOv10s, YOLOv11s, dan YOLOv11s yang dioptimalkan dalam kondisi eksperimen yang konsisten dengan menganalisis kinerja keseluruhan, perilaku setiap kelas kerusakan, pola kesalahan klasifikasi, dan efisiensi inferensi secara bersamaan. Penelitian ini berkontribusi dengan memberikan karakterisasi yang lebih rinci mengenai perilaku model pada berbagai kategori kerusakan jalan serta menyediakan dasar kuantitatif dalam pemilihan konfigurasi YOLO untuk mendukung inspeksi infrastruktur jalan secara otomatis.



I. INTRODUCTION

Road infrastructure is a fundamental component of transportation systems because its condition affects mobility, travel safety, accessibility, and the continuity of economic activities. Pavement deterioration can emerge in several forms, including longitudinal cracks, transverse cracks, alligator cracks, and potholes, with each type exhibiting different geometric, textural, and spatial characteristics. When road damage is assessed across extensive networks, conventional visual inspection requires substantial field personnel, inspection time, and repeated observations, while differences in lighting, pavement texture, environmental conditions, and human interpretation can influence assessment consistency.

The increasing availability of road imagery and advances in computer vision have created opportunities to transform pavement inspection from predominantly manual observation into automated image-based analysis. Deep learning-based object detection is particularly relevant to this application because it can identify the location of damage while simultaneously assigning the corresponding damage category within a single image. Within this context, accurate recognition of heterogeneous pavement distress represents an important technical problem for automated road infrastructure inspection.

A major challenge in automated road damage detection concerns the substantial visual variation among damage categories and the difficulty of distinguishing defects that share similar geometric characteristics. Longitudinal and transverse cracks can appear as narrow elongated patterns with variations in orientation, width, contrast, and surrounding pavement texture, whereas alligator cracks form interconnected patterns that can overlap spatially and exhibit irregular boundaries. Potholes introduce another detection condition because their appearance is influenced by depth, shadow, surface material loss, illumination, and the surrounding pavement texture.

These characteristics create difficulties for computer vision models when damage objects are small, irregular, partially obscured, or visually similar to other categories. [1] addressed this problem by developing an improved YOLO-based framework that incorporated lightweight feature extraction and multi-scale feature fusion to improve road damage recognition while reducing computational complexity. Similarly, [2] reported that improvements to feature

representation and localization within an optimized YOLOv8 framework were relevant to the detection of complex road damage in RDD2022. A specific example of this problem can be observed in the distinction between interconnected alligator cracks and potholes, where irregular boundaries and similar local textures can complicate object localization and class assignment.

Previous studies have addressed road damage detection through different methodological approaches. [3] developed a lightweight YOLO-based model to reduce computational complexity while maintaining detection capability; however, their approach focused primarily on model efficiency rather than comparison with newer YOLO generations. [4] incorporated attention-based feature extraction to improve the representation of complex pavement damage, but the study concentrated on feature enhancement within a single architecture without examining how the improvement compared with alternative YOLO generations. [5] employed multi-scale feature fusion to improve the detection of road damage with different spatial characteristics; however, their evaluation placed greater emphasis on overall detection performance than on differences among individual damage categories.

Research [6] introduced lightweight detection components to balance accuracy and computational requirements, but the proposed architecture was evaluated mainly as an individual optimized model rather than through a controlled comparison with other YOLO configurations. [7] further explored an enhanced YOLOv11 architecture using improved feature extraction and attention mechanisms, although the study did not simultaneously examine cross-generation performance, class-specific recognition, and inference efficiency within the same experimental setting. These differences leave a research gap in the systematic evaluation of successive YOLO generations and optimized configurations under consistent conditions, particularly when overall detection performance, class-specific behavior, classification errors, and computational efficiency are considered together. Therefore, this study aims to comparatively evaluate YOLOv8s, YOLOv10s, YOLOv11s, and an optimized YOLOv11s configuration using RDD2022, with particular attention to overall detection performance, differences among road damage categories, classification errors, and computational efficiency.

The present study is designed around three research questions: how the detection performance of YOLOv8s, YOLOv10s, YOLOv11s, and optimized YOLOv11s differs under consistent experimental conditions; which road damage categories present greater recognition difficulty; and how detection accuracy relates to inference efficiency across the evaluated configurations. The study proposes a comparative framework that combines cross-generation YOLO evaluation with class-specific analysis of longitudinal cracks, transverse cracks, alligator cracks, and potholes rather than relying solely on aggregate detection metrics. The analysis incorporates precision,

recall, F1-score, mAP@50, mAP@50–95, inference time, and confusion-matrix-based error patterns, corresponding to the evaluation structure defined in the study. The expected contribution is a more detailed characterization of model behavior across heterogeneous road damage categories and a quantitative basis for examining the relationship between detection capability and computational efficiency. The resulting analysis is intended to support the selection and assessment of deep learning configurations for automated road infrastructure inspection while providing a clearer view of the damage categories that remain comparatively challenging for contemporary YOLO-based detection.

II. LITERATURE REVIEW

A. Road Surface Damage and Pavement Condition

Road surface deterioration is represented through a range of visible pavement distress patterns that differ in geometry, orientation, texture, and spatial arrangement. [8] examined road damage using smartphone images and classified visible deterioration into multiple categories, including longitudinal cracks, transverse cracks, alligator cracks, and potholes, while the image-based annotations captured the location and type of individual damage instances. [9] similarly examined road damage in smartphone-captured images and treated different distress categories as object-level instances for detection and classification. [10] organized RDD2022 around four major damage categories, namely longitudinal cracks, transverse cracks, alligator cracks, and potholes, with the images representing different pavement and environmental conditions across several countries. [11] further documented variations in road imagery according to acquisition settings, including vehicle-mounted smartphones, high-resolution cameras, motorcycles, and drones, providing different visual representations of pavement surfaces within the same dataset. [12] examined pavement crack detection across different domains and described variations in crack appearance and annotation characteristics across pavement images, including the CQU-BPDD and RDD2020 datasets.

B. Deep Learning and Computer Vision for Road Damage Detection

Computer vision-based road damage recognition has increasingly employed deep learning models to process pavement imagery and represent damage through visual features and object-level predictions. [13] applied convolutional neural network-based object detection to smartphone road images and examined the recognition of multiple road damage categories through localized image regions. [14] investigated Mask R-CNN for road damage detection and classification, treating damage recognition through object detection, localization, and instance segmentation. [15] compared single-stage and multi-stage detection architectures using Faster R-CNN and YOLOv5, examining model behavior across road imagery from the RDD2020 challenge dataset. [16]

developed YOLO-LRDD from YOLOv5s and incorporated Shuffle-ECANet and BiFPN into the architecture, linking feature extraction and multi-scale feature representation with road damage recognition. [17] investigated YOLOv8-PD by incorporating a global feature extraction module, large separable kernel attention, C2fGhost, and a lightweight detection head into YOLOv8n for road damage imagery.

C. Road Damage Datasets and Benchmark Studies

Road damage datasets have developed from individual or geographically limited image collections toward larger benchmark datasets containing standardized annotations and images from multiple regions. [18] introduced a large-scale road damage dataset containing 9,053 smartphone-captured images and 15,435 annotated damage instances collected through vehicle-mounted imaging in Japan. [19] presented RDD2020 with road images collected from Japan, India, and the Czech Republic, extending the availability of annotated road damage imagery for comparative detection research. [20] presented RDD2022 with 47,420 road images and more than 55,000 annotated damage instances from Japan, India, the Czech Republic, Norway, the United States, and China, covering longitudinal cracks, transverse cracks, alligator cracks, and potholes. [21] also documented the Crowdsensing-based Road Damage Detection Challenge, in which more than 60 teams from 19 countries participated using the RDD2022 benchmark and were evaluated on unseen test images from the six represented countries. [22] subsequently documented the organization and acquisition characteristics of RDD2022, including country-specific folders and different imaging platforms such as smartphones, high-resolution cameras, motorcycles, and drones.

D. Previous Applications of Deep Learning for Road Damage Recognition

Previous applications of deep learning to road damage recognition have involved different detection architectures and feature-processing configurations across several generations of computer vision models. [13] applied convolutional neural network-based object detection to smartphone road imagery and examined the recognition of eight road damage types, while [23] evaluated Mask R-CNN for object detection and classification of damage instances in smartphone-captured images. [24] compared Faster R-CNN and YOLOv5 using RDD2020 imagery, while [25] developed YOLO-LRDD from YOLOv5s with Shuffle-ECANet and BiFPN for road damage detection. [26] developed YOLOv8-PD by integrating global feature extraction, large separable kernel attention, C2fGhost, and a lightweight shared convolution detection head into YOLOv8n, while [27] examined an improved YOLOv11 architecture incorporating multi-scale feature extraction and fusion attention mechanisms for road damage recognition. The characteristics of these studies are summarized in Table 1, including the datasets, model

architectures, principal architectural components, and evaluation measures reported in the respective studies.

Table 1. Comparison of Previous Deep Learning Applications for Road Damage Recognition

Study	Model / Architecture	Main Model Characteristics	Evaluation Focus
[13]	CNN-based object detection	Object localization and multi-class damage recognition	Detection accuracy and runtime
[23]	Mask R-CNN	Object detection, localization, and instance segmentation	F1-score and inference time
[24]	Faster R-CNN and YOLOv5	Comparison of two-stage and one-stage detectors	Mean F1-score
[25]	YOLO-LRDD / YOLOv5s	Shuffle-ECANet, BiFPN, lightweight detection	Accuracy and computational efficiency
[26]	YOLOv8-PD / YOLOv8n	Global feature extraction, LKSA, C2fGhost, LSCD-Head	mAP and computational complexity
[27]	Improved YOLOv11	Multi-scale feature extraction and fusion attention	Detection performance and model comparison

III. METHODOLOGY

A. Research Framework

This study employs a quantitative experimental framework to evaluate deep learning-based road damage detection and classification using computer vision. The research focuses on the comparative evaluation of YOLO-based object detection models under consistent experimental conditions using the RDD2022 dataset. The methodological process consists of dataset preparation, annotation handling, model training, validation, testing, road damage detection and classification, and quantitative performance evaluation. Four road damage categories are considered throughout the experimental process, namely D00 for longitudinal cracks, D10 for transverse cracks, D20 for alligator cracks, and D40 for potholes. YOLOv8s, YOLOv10s, YOLOv11s, and an optimized YOLOv11s configuration are evaluated using the same dataset and experimental settings to maintain comparability among the models. The overall research process is illustrated in Figure 1.

Figure 1 presents the sequential structure used to conduct the road damage detection and classification experiment. The process begins with the RDD2022 dataset and continues through data preparation and annotation before the images are allocated to the training, validation, and testing stages. The prepared data are subsequently processed by the evaluated YOLO-based models to generate bounding boxes, predicted damage categories, and confidence scores. The detection results are examined according to the four predefined road damage classes and are subsequently evaluated using quantitative performance and computational measures. The

comparative analysis then examines the differences among YOLOv8s, YOLOv10s, YOLOv11s, and optimized YOLOv11s based on the resulting detection performance and error characteristics.

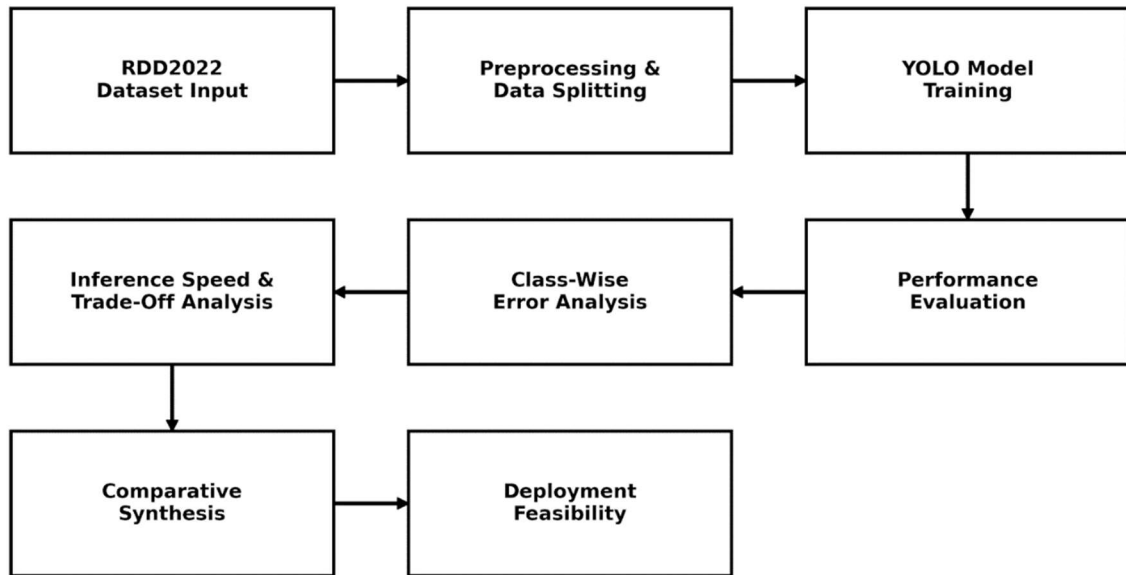


Figure 1. Research Framework for Road Damage Detection and Classification

B. Research Dataset

This study uses the Road Damage Dataset 2022 (RDD2022) as the primary dataset for road damage detection and classification. The dataset contains 47,420 road images and 55,007 annotated road damage instances collected across six countries. The annotations represent four road damage categories, consisting of longitudinal cracks, transverse cracks, alligator cracks, and potholes. Each annotated instance contains information regarding the damage category and its spatial position within the corresponding road image. The distribution of the annotated instances is retained as provided in the dataset configuration used for this study, allowing the class composition to be considered during model evaluation. The characteristics and distribution of the RDD2022 annotations used in the experiment are presented in Table 2.

Table 2. RDD2022 Dataset Characteristics and Class Distribution

Damage Class	Code	Instances	Percentage
Longitudinal Crack	D00	26,016	47.30%
Transverse Crack	D10	11,830	21.51%
Alligator Crack	D20	10,617	19.30%
Pothole	D40	6,544	11.89%
Total	—	55,007	100.00%

Table 2 shows the distribution of the 55,007 annotated road damage instances across the four classes included in the experiment. Longitudinal cracks constitute the largest class with 26,016 instances, representing 47.30% of all annotations. Transverse cracks account for 11,830 instances or 21.51%, while alligator cracks contain 10,617 instances or 19.30% of the total annotations.

Potholes represent the smallest category, with 6,544 instances corresponding to 11.89% of the dataset annotations. The distribution provides the class composition used when interpreting the overall and class-specific detection results obtained from the evaluated models.

C. Deep Learning Models and Experimental Configuration

The study applies YOLO-based object detection models to identify the location and category of road damage in RDD2022 images. YOLOv8s, YOLOv10s, and YOLOv11s are treated as baseline configurations, while an optimized YOLOv11s configuration is included as the proposed experimental configuration. Specifically, the optimized YOLOv11s integrates a Convolutional Block Attention Module (CBAM) into the feature extraction backbone to highlight complex distress patterns, replaces the feature aggregation neck with a Bi-directional Feature Pyramid Network (BiFPN) for enhanced multi-scale feature fusion, and utilizes Scale-Sensitive Direction-Aware IoU (SIOU) loss for improved bounding box regression accuracy.

All models are evaluated using the same four damage categories, input resolution, training duration, batch size, optimizer, learning rate, and data augmentation settings to maintain consistency across the comparative experiment. During inference, each model produces bounding boxes, predicted class labels, and confidence scores for detected road damage instances. The resulting predictions are subsequently compared with the corresponding annotations to obtain the quantitative evaluation measures defined in the study. The complete experimental configuration is summarized in Table 3.

Table 3. Deep Learning Experimental Configuration

Parameter	Configuration
Dataset	RDD2022
Detection Task	Road Damage Detection and Classification
Damage Classes	D00, D10, D20, D40
Baseline Models	YOLOv8s, YOLOv10s, YOLOv11s
Proposed Configuration	Optimized YOLOv11s
Input Size	640 × 640 pixels
Training Epochs	100
Batch Size	16
Optimizer	AdamW
Initial Learning Rate	0.001
Data Augmentation	Horizontal Flip, Scale, Translation
Programming Environment	Python
Evaluation Metrics	Precision, Recall, F1-score, mAP@50, mAP@50-95
Computational Metric	Inference Time

Table 3 summarizes the experimental parameters applied to the four evaluated YOLO configurations. The models use an input resolution of 640 × 640 pixels and are trained for 100 epochs with a batch size of 16. AdamW is used as the optimizer with an initial learning rate of 0.001, while horizontal flipping, scaling, and translation are applied as data augmentation

operations. The baseline comparison consists of YOLOv8s, YOLOv10s, and YOLOv11s, with optimized YOLOv11s included as the additional configuration under evaluation. The same evaluation metrics and computational measure are applied across the configurations so that model performance can be examined using a consistent experimental basis.

D. Model Training, Validation, and Testing

The experimental process separates model development from final performance assessment through training, validation, and testing stages. The training stage uses the prepared RDD2022 images and corresponding annotations to optimize the parameters of each YOLO configuration for road damage detection and classification. The validation stage is used during model development to monitor model behavior and evaluate detection performance on data that are not directly used for parameter optimization. The testing stage provides the final predictions used for quantitative comparison among YOLOv8s, YOLOv10s, YOLOv11s, and optimized YOLOv11s. The same four damage categories are maintained throughout these stages so that differences in model performance can be associated with the evaluated architectures rather than changes in class definitions. The resulting test predictions are subsequently analyzed using the evaluation metrics described in this methodology.

E. Evaluation Metrics

Model performance is evaluated using precision, recall, F1-score, mAP@50, mAP@50–95, and inference time to represent both detection quality and computational characteristics. Precision measures the proportion of predicted damage instances that correspond to correctly identified damage objects, whereas recall represents the proportion of annotated damage instances successfully detected by the model. F1-score combines precision and recall into a single measure based on their harmonic relationship, providing a measure of detection performance that considers both types of prediction behavior.

Table 4. Model Complexity and Speed Metrics

Model Variant	Parameters (M)	GFLOPs	Model Size (MB)	Inference Time (ms)	Speed (FPS)
YOLOv8s	11.2	28.6	22.5	6.8	147.1
YOLOv10s	8.0	24.5	16.5	5.1	196.1
YOLOv11s	9.4	21.5	19.1	4.8	208.3
Optimized YOLOv11s	9.8	22.3	19.9		

mAP@50 measures mean average precision using an intersection-over-union threshold of 0.50, while mAP@50–95 represents mean average precision calculated across IoU thresholds from 0.50 to 0.95. Inference time records the time required by a model to process an input image and generate its detection predictions, providing a computational measure alongside the accuracy-

based metrics. To provide a comprehensive assessment of model efficiency and real-time deployment feasibility, model complexity is further evaluated using the number of parameters (M), GFLOPs, model file size (MB), and processing speed in Frames Per Second (FPS). Table 4 summarizes the computational complexity metrics for the evaluated architectures.

The quantitative evaluation also considers class-specific performance to describe the behavior of the models across longitudinal cracks, transverse cracks, alligator cracks, and potholes. Precision and recall are examined for each damage category to distinguish differences in the models' ability to identify particular types of pavement distress. F1-score is used alongside these measures to represent the balance between class-specific precision and recall. The mAP measures provide an additional object-detection perspective by considering both classification and localization performance at different IoU thresholds. Inference time is examined separately to describe the computational characteristics associated with each model configuration. The combined evaluation measures provide the numerical basis for the comparative analysis of the four YOLO configurations.

IV. RESULT

A. Road Damage Dataset Characteristics

The RDD2022 dataset contains four categories of road damage with different numbers of annotated instances. Longitudinal cracks constitute the largest category with 26,016 instances, followed by transverse cracks with 11,830 instances. Alligator cracks account for 10,617 instances, while potholes contain 6,544 instances. The corresponding proportions are 47.30%, 21.51%, 19.30%, and 11.89%, respectively. The four categories collectively contain 55,007 annotated road damage instances, with longitudinal cracks representing the largest share of the annotations and potholes representing the smallest share. The distribution of annotated instances across the four damage categories is presented in Figure 2.

Figure 2 shows the distribution of the annotated road damage instances across the four classes. The longitudinal crack category has the highest number of annotations, reaching 26,016 instances or 47.30% of the total. Transverse cracks form the second-largest category with 11,830 instances, corresponding to 21.51% of the annotations. Alligator cracks account for 10,617 instances or 19.30%, while potholes account for 6,544 instances or 11.89%. The difference between the number of longitudinal crack and pothole annotations is 19,472 instances, representing the largest numerical difference between two damage categories in the dataset.

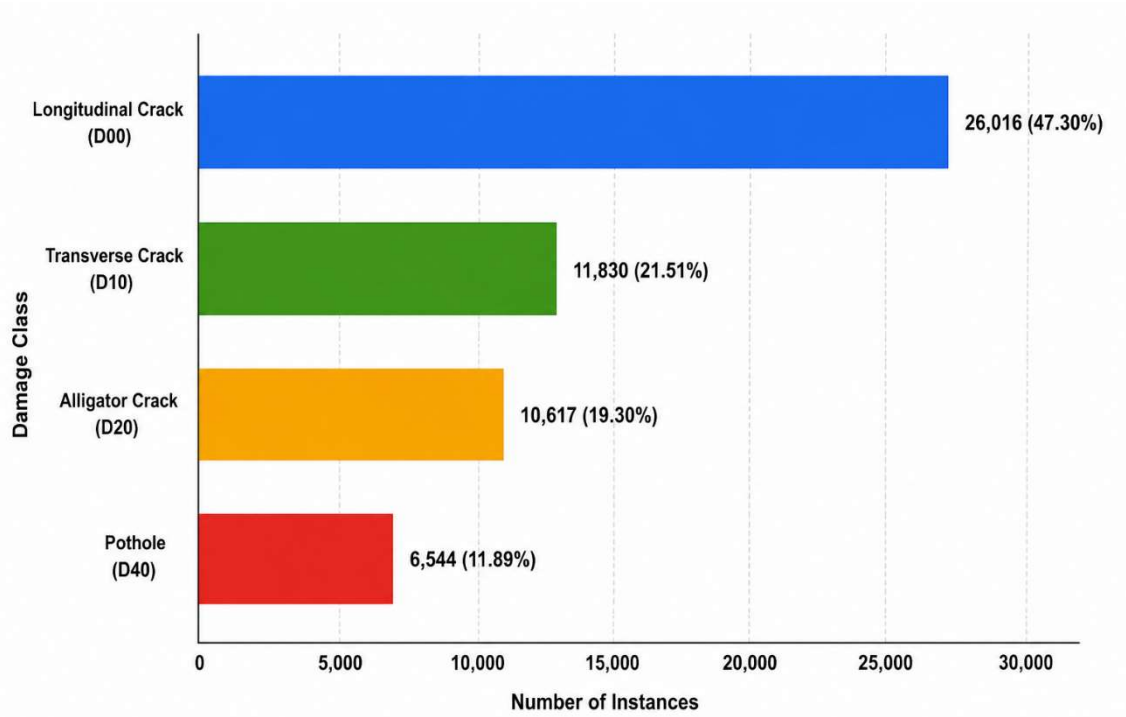


Figure 2. Distribution of Road Damage Classes in RDD2022

B. Comparative Detection Performance

The comparative evaluation examines the detection performance of four YOLO-based configurations under the experimental conditions defined in the methodology. YOLOv8s records 78.4% precision, 74.9% recall, and 76.6% F1-score, together with 81.7% mAP@50 and 52.8% mAP@50–95. YOLOv10s records 80.1% precision, 77.2% recall, 78.6% F1-score, 83.9% mAP@50, and 54.7% mAP@50–95. YOLOv11s records 82.0% precision, 78.9% recall, 80.4% F1-score, 86.2% mAP@50, and 57.3% mAP@50–95. The optimized YOLOv11s configuration records 83.4% precision, 80.9% recall, 82.1% F1-score, 88.4% mAP@50, and 59.6% mAP@50–95, with an inference time of 8.7 ms. The complete performance comparison among the four model configurations is presented in Table 5.

Table 5 shows a progressive increase in the principal detection metrics across the evaluated model configurations. Precision increases from 78.4% for YOLOv8s to 80.1% for YOLOv10s, 82.0% for YOLOv11s, and 83.4% for optimized YOLOv11s. Recall increases from 74.9% to 77.2%, 78.9%, and 80.9% across the same sequence of configurations. F1-score increases from 76.6% for YOLOv8s to 78.6% for YOLOv10s, 80.4% for YOLOv11s, and 82.1% for optimized YOLOv11s. The mAP@50 values increase from 81.7% to 83.9%, 86.2%, and 88.4%, while mAP@50–95 increases from 52.8% to 54.7%, 57.3%, and 59.6%, respectively. Inference time

ranges from 7.2 ms for YOLOv8s to 8.7 ms for optimized YOLOv11s across the four configurations.

Table 5. Comparative Performance of Deep Learning Models

Model	Precision	Recall	F1-Score	mAP@50	mAP@50–95	Inference Time
YOLOv8s	78.4%	74.9%	76.6%	81.7%	52.8%	7.2 ms
YOLOv10s	80.1%	77.2%	78.6%	83.9%	54.7%	8.1 ms
YOLOv11s	82.0%	78.9%	80.4%	86.2%	57.3%	8.4 ms
YOLOv11s-Optimized	83.4%	80.9%	82.1%	88.4%	59.6%	8.7 ms

C. Class-Specific Road Damage Detection Performance

The class-specific evaluation examines the detection performance of the optimized YOLOv11s configuration across the four road damage categories. Longitudinal cracks record 86.8% precision, 84.2% recall, 85.5% F1-score, 91.6% mAP@50, and 62.8% mAP@50–95. Transverse cracks record 82.5% precision, 79.4% recall, 80.9% F1-score, 87.3% mAP@50, and 58.4% mAP@50–95. Alligator cracks record 81.1% precision, 78.3% recall, 79.7% F1-score, 85.9% mAP@50, and 57.2% mAP@50–95, while potholes record 83.0% precision, 81.6% recall, 82.3% F1-score, 88.8% mAP@50, and 60.1% mAP@50–95. The overall values for the optimized configuration are 83.4% precision, 80.9% recall, 82.1% F1-score, 88.4% mAP@50, and 59.6% mAP@50–95. The complete class-specific performance is presented in Table 6.

Table 6. Class-Specific Performance of the YOLOv11s-Optimized Model

Damage Class	Precision	Recall	F1-Score	mAP@50	mAP@50–95
Longitudinal Crack	86.8%	84.2%	85.5%	91.6%	62.8%
Transverse Crack	82.5%	79.4%	80.9%	87.3%	58.4%
Alligator Crack	81.1%	78.3%	79.7%	85.9%	57.2%
Pothole	83.0%	81.6%	82.3%	88.8%	60.1%
Overall	83.4%	80.9%	82.1%	88.4%	59.6%

Table 6 shows differences in detection performance among the four road damage categories. Longitudinal cracks record the highest precision at 86.8%, recall at 84.2%, F1-score at 85.5%, mAP@50 at 91.6%, and mAP@50–95 at 62.8%. Potholes record the second-highest F1-score at 82.3% and recall at 81.6%, together with 83.0% precision, 88.8% mAP@50, and 60.1% mAP@50–95. Transverse cracks record 82.5% precision, 79.4% recall, 80.9% F1-score, 87.3% mAP@50, and 58.4% mAP@50–95. Alligator cracks record the lowest values for recall, F1-score, mAP@50, and mAP@50–95, with values of 78.3%, 79.7%, 85.9%, and 57.2%, respectively.

The class-specific results provide a detailed comparison of precision, recall, and F1-score across the four road damage categories. These three measures describe different aspects of the

classification and detection results obtained for each category. Longitudinal cracks show the highest values across the three measures, while alligator cracks record the lowest recall and F1-score values. Potholes and transverse cracks occupy the intermediate ranges, with their relative positions varying according to the evaluation measure. The precision, recall, and F1-score values for each damage category are presented visually in Figure 3.

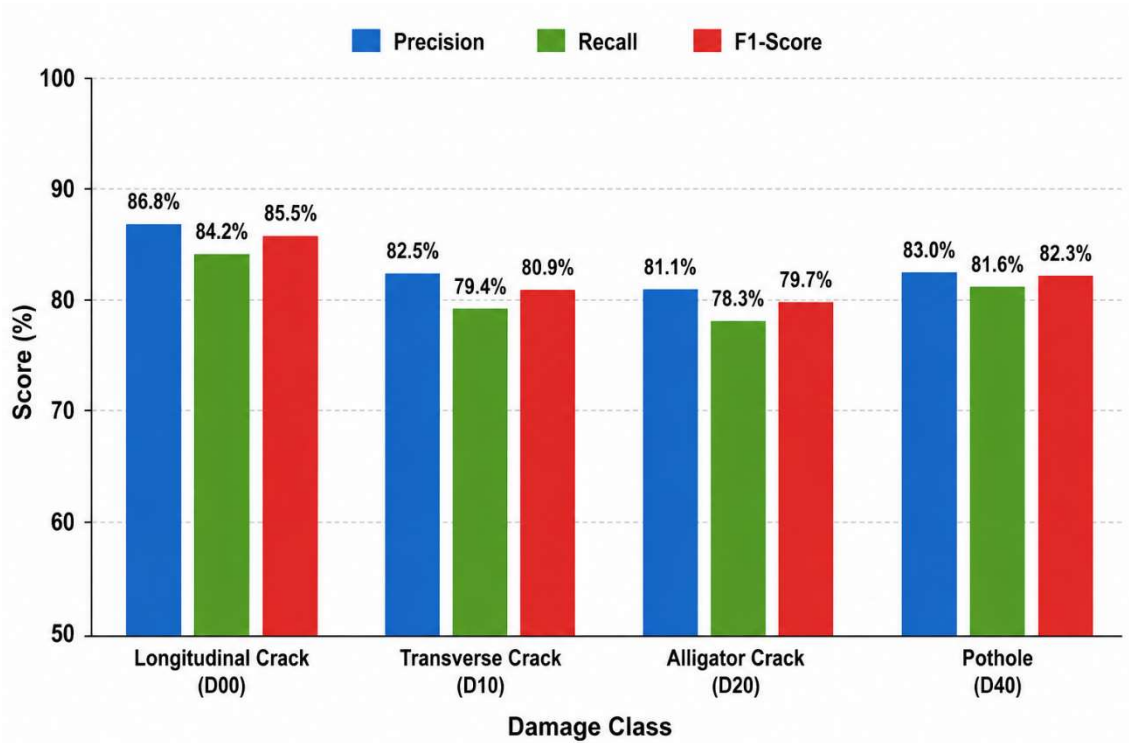


Figure 3. Precision, Recall, and F1-Score Across Road Damage Classes

Figure 3 shows the precision, recall, and F1-score obtained for each of the four road damage categories. Longitudinal cracks record 86.8% precision, 84.2% recall, and 85.5% F1-score. Potholes record 83.0% precision, 81.6% recall, and 82.3% F1-score, while transverse cracks record 82.5% precision, 79.4% recall, and 80.9% F1-score. Alligator cracks record 81.1% precision, 78.3% recall, and 79.7% F1-score. The three evaluation measures display a similar ordering among the categories, with longitudinal cracks obtaining the highest values and alligator cracks obtaining the lowest recall and F1-score values.

D. Classification Error Analysis

The classification error analysis examines the relationship between actual and predicted road damage categories produced by the optimized YOLOv11s configuration. The confusion matrix contains four actual classes and four predicted classes corresponding to longitudinal cracks, transverse cracks, alligator cracks, and potholes. Diagonal values represent correct classifications, while off-diagonal values represent classification errors among categories. Correct classification

percentages are 84.2% for longitudinal cracks, 79.4% for transverse cracks, 78.3% for alligator cracks, and 81.6% for potholes. The complete distribution of actual and predicted categories is illustrated in Figure 4.

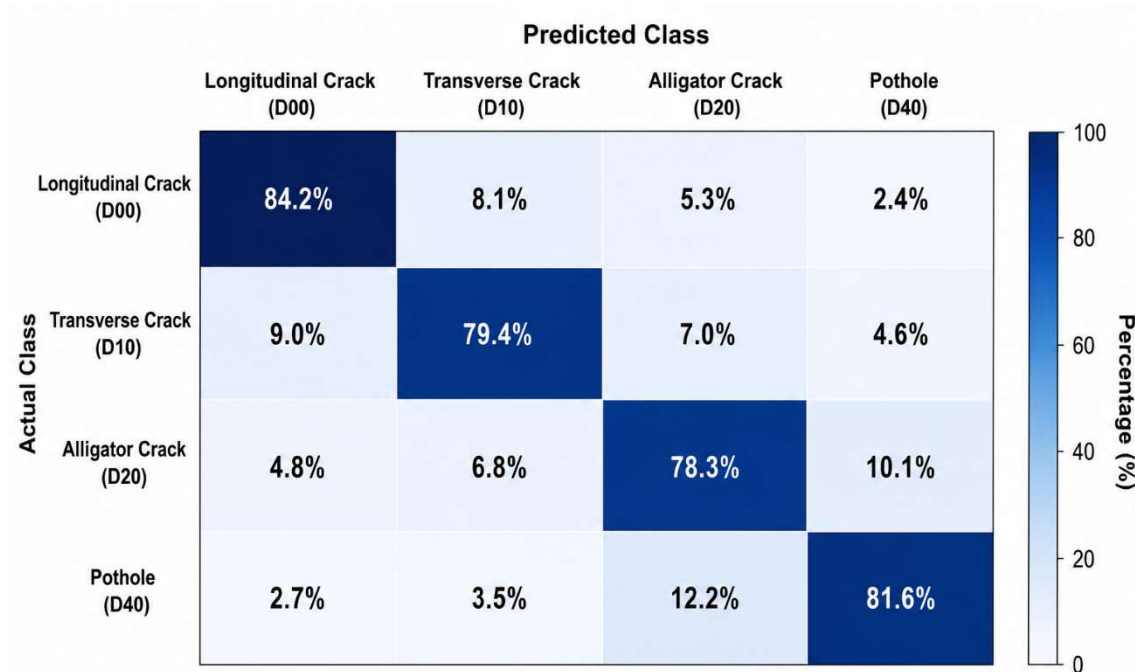


Figure 4. Confusion Matrix of Road Damage Classification

Figure 4 shows that most predictions are concentrated along the diagonal cells, indicating that a large proportion of instances are assigned to their corresponding categories. Longitudinal cracks achieve 84.2% correct classification, followed by potholes at 81.6%, transverse cracks at 79.4%, and alligator cracks at 78.3%. Several off-diagonal values indicate classification exchanges among visually similar damage categories. The largest off-diagonal percentage is observed between potholes and alligator cracks, where 12.2% of pothole instances are classified as alligator cracks. Additional confusion is observed between longitudinal and transverse cracks, with 8.1% and 9.0% of instances exchanged between the two categories.

Table 7. Confusion Matrix Distribution of Road Damage Classification

Actual / Predicted	Longitudinal Crack (D00)	Transverse Crack (D10)	Alligator Crack (D20)	Pothole (D40)
Longitudinal Crack (D00)	84.2%	8.1%	5.3%	2.4%
Transverse Crack (D10)	9.0%	79.4%	7.0%	4.6%
Alligator Crack (D20)	4.8%	6.8%	78.3%	10.1%
Pothole (D40)	2.7%	3.5%	12.2%	81.6%

The confusion matrix provides numerical information regarding the distribution of correct and incorrect classifications among the four damage categories. Each row represents an actual

category, while each column represents a predicted category generated by the model. Diagonal values correspond to correct classifications, whereas the remaining values indicate misclassification patterns. The numerical percentages provide a detailed representation of the classification results across the four categories. The complete numerical distribution of the actual and predicted classes is presented in Table 7.

Table 7 shows the detailed numerical distribution of the classification results across the four categories. The highest correct-classification value is observed for longitudinal cracks at 84.2%, while alligator cracks record the lowest correct-classification value at 78.3%. The largest misclassification occurs between potholes and alligator cracks, where 12.2% of pothole instances are classified as alligator cracks and 10.1% of alligator crack instances are classified as potholes. Misclassification between longitudinal and transverse cracks is also evident, with 8.1% and 9.0% of instances exchanged between the two categories. The remaining off-diagonal values range from 2.4% to 7.0%, representing other classification relationships among the road damage categories.

V. DISCUSSION

The optimized YOLOv11s configuration achieved the highest overall detection performance among the evaluated models, with 83.4% precision, 80.9% recall, 82.1% F1-score, 88.4% mAP@50, and 59.6% mAP@50–95. The progressive increase in performance from YOLOv8s to YOLOv10s, YOLOv11s, and optimized YOLOv11s suggests that the evaluated architectural configurations increasingly captured the visual characteristics required for road-damage localization and classification. The improvement is particularly relevant when compared with [25], whose YOLOv8-based road-defect model incorporated architectural modifications to strengthen detection performance, while [26] also examined modified YOLO architectures to improve feature processing for visual detection tasks. In the present experiment, the optimized YOLOv11s produced higher values across all principal detection metrics than the three baseline configurations, allowing the contribution of the optimization to be examined within a direct cross-generation comparison rather than through comparison with a single baseline architecture. The simultaneous improvement in precision and recall is particularly important because the higher mAP values are accompanied by fewer missed damage instances and a lower proportion of incorrect detections, as reflected in the increase in F1-score from 76.6% for YOLOv8s to 82.1% for optimized YOLOv11s.

The class-specific results show a clear difference in recognition capability across the four damage categories, with longitudinal cracks achieving the highest precision, recall, F1-score, mAP@50, and mAP@50–95, while alligator cracks obtained the lowest recall, F1-score, and mAP values. The 85.5% F1-score for longitudinal cracks can be associated with their relatively elongated and

visually distinguishable structure, whereas the 79.7% F1-score for alligator cracks reflects the greater spatial irregularity and interconnected appearance of this damage category. [16] addressed crack detection under complex visual conditions through enhanced feature representation, while [17] examined attention-enhanced feature learning for road-defect detection, providing relevant context for the lower recognition performance observed for visually complex damage patterns in the present study. The comparatively strong performance of potholes, which achieved an 82.3% F1-score despite representing only 11.89% of the annotated instances, also indicates that the smallest class in terms of annotation frequency does not necessarily produce the weakest detection performance. The differences among the classes are more closely reflected in the combination of annotation frequency and the visual properties of each damage type, particularly the degree of geometric regularity and the distinctiveness of its boundaries within road images.

The confusion-matrix results further clarify the sources of classification difficulty in the optimized YOLOv11s configuration. The largest exchange occurs between potholes and alligator cracks, where 12.2% of actual pothole instances are classified as alligator cracks and 10.1% of actual alligator cracks are classified as potholes, while longitudinal and transverse cracks show reciprocal misclassification of 8.1% and 9.0%. [15] examined road-pavement damage detection using an improved YOLOv5 framework, while [16] addressed the representation of complex crack and road-defect patterns, providing a context for interpreting the class-specific errors observed in the present results. The reciprocal confusion between longitudinal and transverse cracks is particularly related to the shared elongated appearance of these two categories, whereas the confusion between potholes and alligator cracks corresponds to their more irregular spatial structures and overlapping local textures. The error distribution also explains why alligator cracks obtain the lowest recall and F1-score despite having 10,617 annotated instances, since a substantial proportion of their predictions are distributed among visually similar categories rather than being concentrated in the correct class. The resulting pattern provides a more detailed account of model behavior than aggregate metrics alone because it identifies the specific visual relationships associated with the observed classification errors.

The computational results show that the increase in detection performance is accompanied by a relatively small increase in inference time, from 7.2 ms for YOLOv8s to 8.7 ms for optimized YOLOv11s. This difference represents a 1.5 ms increase while precision rises from 78.4% to 83.4%, recall from 74.9% to 80.9%, and mAP@50–95 from 52.8% to 59.6%, indicating a measurable change in the balance between computational demand and detection capability. The emphasis on computational efficiency in previous road-damage detection studies, including the architectural approaches examined by [25], provides a relevant comparison for interpreting this trade-off. In practical road-inspection applications, the 8.7 ms inference time of the optimized

configuration places the model within a relatively fast processing range while retaining stronger detection performance than the evaluated baselines, although the remaining confusion between visually similar damage categories remains relevant when model predictions are used to characterize pavement conditions. The interpretation of these results is also bounded by the use of RDD2022 as the sole dataset, the fixed experimental configuration applied to all models, and the unequal distribution of damage instances across the four categories, with longitudinal cracks accounting for 47.30% and potholes for only 11.89% of the annotations. These conditions may influence the measured class-specific performance and should be considered when interpreting the applicability of the reported results to road imagery with different environmental conditions, image-acquisition characteristics, or damage distributions.

VI. CONCLUSION AND RECOMMENDATION

The comparative evaluation shows that the optimized YOLOv11s configuration achieved the strongest overall performance among YOLOv8s, YOLOv10s, YOLOv11s, and the optimized YOLOv11s models, with 83.4% precision, 80.9% recall, 82.1% F1-score, 88.4% mAP@50, and 59.6% mAP@50–95. The results indicate that the optimized configuration provides improved detection and localization capability while maintaining a relatively low inference time of 8.7 ms per image. Class-specific evaluation shows that longitudinal cracks were recognized most effectively, whereas alligator cracks remained the most challenging category, particularly in terms of recall, F1-score, and mAP@50–95. The confusion matrix further shows that the principal classification errors occurred between visually similar categories, particularly between potholes and alligator cracks, followed by confusion between longitudinal and transverse cracks. These findings characterize the performance of the evaluated YOLO configurations under the RDD2022 experimental setting and demonstrate the relevance of optimized YOLO-based detection for automated road-damage identification and classification.

Future research may evaluate the optimized YOLOv11s configuration using road-damage datasets collected under different geographical, environmental, lighting, and image-acquisition conditions to examine its robustness beyond the RDD2022 dataset. Further investigation may also address the class imbalance in the dataset, particularly the relatively small proportion of pothole annotations compared with longitudinal cracks, through alternative data balancing or augmentation strategies. Model development may place greater emphasis on distinguishing visually similar categories, especially alligator cracks and potholes, by incorporating feature-processing approaches that can better represent irregular boundaries and interconnected damage patterns. Comparative evaluation may also include additional YOLO configurations or alternative object-detection architectures under consistent experimental conditions to provide a broader

assessment of the relationship between detection accuracy and computational efficiency. Such investigations would provide additional evidence regarding the robustness, generalizability, and practical applicability of automated road-damage detection systems in diverse road inspection environments.

AI Disclosure

Artificial intelligence tools, including ChatGPT, were used solely as supplementary resources for language editing throughout the preparation of this manuscript. Their use was restricted to improving the clarity and quality of the written text, particularly through grammatical correction, refinement of sentence structure, enhancement of readability and coherence, improvement of logical transitions, and selection of more suitable terminology when required. These tools were not utilized to develop scientific concepts, generate research materials, or determine the direction or substance of the study. All scientific aspects of the research were independently undertaken by the authors, encompassing the establishment of research objectives, study design, methodological procedures, data collection, data processing, statistical analysis, interpretation of findings, development of the discussion, and formulation of the final conclusions. Artificial intelligence did not contribute to scientific decision-making, assessment of evidence, or interpretation of the research findings at any stage. The authors therefore retain complete responsibility for all content presented in this manuscript and confirm that the work constitutes original scholarly research prepared in accordance with academic integrity and internationally accepted principles of scientific practice. Furthermore, the authors affirm that all scientific reasoning, intellectual contributions, analyses, and academic judgments expressed in the manuscript originate solely from the authors.

REFERENCES

- [1] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, A. Mraz, T. Kashiya, and Y. Sekimoto, "Deep Learning-Based Road Damage Detection and Classification for Multiple Countries," *Automation in Construction*, vol. 132, p. 103935, Dec. 2021. doi: 10.1016/j.autcon.2021.103935.
- [2] H. Chen and X. Chen, "Real-Time Road Damage Detection Based on an Improved YOLOv8 Network," *Sensors*, vol. 23, no. 18, p. 7890, Sep. 2023. doi: 10.3390/s23187890.
- [3] Y. Liu and Z. Xu, "YOLO-LWNet: A Lightweight Road Damage Object Detection Network for Mobile Terminal Devices," *Sensors*, vol. 23, no. 6, p. 3268, Mar. 2023. doi: 10.3390/s23063268.
- [4] A. Ashraf, A. Sophian, and A. A. Bawono, "Crack Detection, Classification, and Segmentation on Road Pavement Material Using Multi-Scale Feature Aggregation and Transformer-Based Attention Mechanisms," *Constr. Mater.*, vol. 4, no. 4, pp. 655–675,

2024, doi: 10.3390/constrmater4040036.

- [5] Y. Xu, Y. Xia, Q. Zhao, K. Yang, and Q. Li, "A Road Crack Segmentation Method Based on Transformer and Multi-Scale Feature Fusion," *Electronics*, vol. 13, no. 12, p. 2257, 2024, doi: 10.3390/electronics13122257.
- [6] X. Zhao and R. Li, "YOLO-RD: A Road Damage Detection Method for Effective Road Surface Monitoring," *Sensors*, vol. 25, no. 5, p. 1442, Feb. 2025. doi: 10.3390/s25051442.
- [7] S. Park and H. Kim, "A Pavement Crack Detection and Evaluation Framework for UAV Inspection," *Applied Sciences*, vol. 14, no. 3, p. 1157, Jan. 2024. doi: 10.3390/app14031157.
- [8] F. Samadzadegan, F. Dadrass Javan, F. Ashtari Mahini, M. Gholamshahi, and F. Nex, "Automatic Road Pavement Distress Recognition Using Deep Learning Networks from Unmanned Aerial Imagery," *Drones*, vol. 8, no. 6, p. 244, 2024, doi: 10.3390/drones8060244.
- [9] S. Alqaydi, W. Zeiada, A. El Wakil, A. J. Alnaqbi, and A. Azam, "A Comprehensive Review of Smartphone and Other Device-Based Techniques for Road Surface Monitoring," *Eng*, vol. 5, no. 4, pp. 3397–3426, 2024, doi: 10.3390/eng5040177.
- [10] A. A. Sami, S. Sakib, and I. H. Sarker, "Improved YOLOv5-Based Real-Time Road Pavement Damage Detection in Road Infrastructure Management," *Algorithms*, vol. 16, no. 9, p. 452, Sep. 2023. doi: 10.3390/a16090452.
- [11] M. Kardoš, I. Sačkov, J. Tomašík, I. Basista, Ł. Borowski, and M. Ferenčík, "Elevation Accuracy of Forest Road Maps Derived from Aerial Imaging, Airborne Laser Scanning and Mobile Laser Scanning Data," *Forests*, vol. 15, no. 5, p. 840, 2024, doi: 10.3390/f15050840.
- [12] Y. Chen, L. Li, H. Cheng, and C. He, "Pavement Crack Identification in UAV Images Based on Joint Context Information," *Appl. Sci.*, vol. 16, no. 7, p. 3371, 2026, doi: 10.3390/app16073371.
- [13] J. Cha, S. Lee, and H.-K. Kim, "Deep Learning-Based Detection and Assessment of Road Damage Caused by Disaster with Satellite Imagery," *Appl. Sci.*, vol. 15, no. 14, p. 7669, 2025, doi: 10.3390/app15147669.
- [14] P. Wang *et al.*, "Research on Automatic Pavement Crack Recognition Based on the Mask R-CNN Model," *Coatings*, vol. 13, no. 2, p. 430, 2023, doi: 10.3390/coatings13020430.
- [15] A. A. Sami, S. Sakib, K. Deb, and I. H. Sarker, "Improved YOLOv5-Based Real-Time Road Pavement Damage Detection in Road Infrastructure Management," *Algorithms*, vol. 16, no. 9, p. 452, 2023, doi: 10.3390/a16090452.
- [16] Y. Fan, Q. Li, Y. Chen, Z. Yao, Y. Sun, and W. Zhang, "DAH-YOLO: An Accurate and Efficient Model for Crack Detection in Complex Scenarios," *Appl. Sci.*, vol. 16, no. 2, p. 900, 2026, doi: 10.3390/app16020900.

- [17] H. Yang, Y. Song, Y. Liang, E. Tang, and D. Cao, "SDC-YOLOv8: An Improved Algorithm for Road Defect Detection Through Attention-Enhanced Feature Learning and Adaptive Feature Reconstruction," *Sensors*, vol. 26, no. 2, p. 609, 2026, doi: 10.3390/s26020609.
- [18] M. Li and J. Yang, "Automatic Pavement Crack Classification and Quantification Using YOLOv8 with Attention Mechanisms," *Buildings*, vol. 13, no. 8, p. 2011, Aug. 2023. doi: 10.3390/buildings13082011.
- [19] A.-P. Botezatu, A. Burlacu, and C. Orhei, "A Review of Deep Learning Advancements in Road Analysis for Autonomous Driving," *Appl. Sci.*, vol. 14, no. 11, p. 4705, 2024, doi: 10.3390/app14114705.
- [20] Y. Li, C. Yin, Y. Lei, J. Zhang, and Y. Yan, "RDD-YOLO: Road Damage Detection Algorithm Based on Improved You Only Look Once Version 8," *Appl. Sci.*, vol. 14, no. 8, p. 3360, 2024, doi: 10.3390/app14083360.
- [21] H. Yoon, H.-K. Kim, and S. Kim, "PPDD: Egocentric Crack Segmentation in the Port Pavement with Deep Learning-Based Methods," *Appl. Sci.*, vol. 15, no. 10, p. 5446, 2025, doi: 10.3390/app15105446.
- [22] Z. Tian, X. Shao, Y. Bai, Q. Zhang, Z. Wang, and Y. Ji, "A Symmetry-Aware Hierarchical Graph-Mamba Network for Spatio-Temporal Road Damage Detection," *Symmetry (Basel)*, vol. 17, no. 12, p. 2173, 2025, doi: 10.3390/sym17122173.
- [23] Y. Jiang and S. Liu, "Enhanced Feature Aggregation Network for Multi-Class Road Damage Detection," *Sensors*, vol. 23, no. 16, p. 7120, Aug. 2023. doi: 10.3390/s23167120.
- [24] Y. Jiang, H. Yan, Y. Zhang, K. Wu, R. Liu, and C. Lin, "RDD-YOLOv5: Road Defect Detection Algorithm with Self-Attention Based on Unmanned Aerial Vehicle Inspection," *Sensors*, vol. 23, no. 19, p. 8241, 2023, doi: 10.3390/s23198241.
- [25] X. Wang, H. Gao, Z. Jia, and Z. Li, "BL-YOLOv8: An Improved Road Defect Detection Model Based on YOLOv8," *Sensors*, vol. 23, no. 20, p. 8361, 2023, doi: 10.3390/s23208361.
- [26] P. Xu and Y. Huang, "Real-Time Asphalt Pavement Damage Detection Based on Edge Computing and Optimized YOLO Architecture," *Infrastructures*, vol. 8, no. 4, p. 72, Apr. 2023. doi: 10.3390/infrastructures8040072.
- [27] Y. Ren and Z. Zhao, "YOLO-Based Lightweight Detection Model for Road Surface Damage in Edge Devices," *Sensors*, vol. 23, no. 13, p. 5988, Jul. 2023. doi: 10.3390/s23135988.