Hybrid Deep Learning Framework for Automated Road Damage Detection and Classification Using YOLO and R-CNN .

[1]Pratiksha M. Shinde, [2]Ravindra S. Kamble
[1]PG Scholar, [2]PG Guide,
[1]Department of Data Science, D. Y. Patil Agriculture & Technical University, Talsande, Maharashtra, India.
[2]Department of Computer Science & Engineering, D. Y. Patil Agriculture & Technical University,
Talsande, Maharashtra, India

Abstract : Road infrastructure is one of the main elements of modern transportation systems which are used for people, goods and services transport. Road surfaces, however, are continuously subjected to environmental influences, heavy traffic loads, and aging, which lead to defects like cracks, potholes, surface deformations etc. Recognizing and categorizing these damages is crucial in the context of both minimizing traffic accidents and reducing overall vehicle upkeep expenses, as well as enhancing road safety. Traditional Inspection Methods are based on manual surveys, which are time-consuming, labour intensive and prone to error. As a result, the need for automated and intelligent road condition monitoring systems that deliver accurate and timely assessments has grown. In recent years, deep learning and computer vision groundbreaking technologies have enhanced the performance of automated defect detection systems. A hybrid deep learning architecture by integrating YOLO (You Only Look Once) and R-CNN (Region-Based Convolutional Neural Network) is proposed in this research to effectively detect road damage and classify it. YOLO is used to quickly detect the possible damaged zones of road images, and thus, rapid object localization is achieved for real time applications. The candidate regions are then processed by the R-CNN module to obtain more accurate and precise detection results and bounding-box optimization, performing refined feature extraction, classification, and bounding-box optimization. A large dataset with various road defect categories, such as pothole, longitudinal crack, transverse crack, alligator crack, and water-saturated surface areas, is used to train the proposed model. The dataset also includes a variety of environmental conditions, lighting variations, and road surfaces to enhance the framework's robustness and generalization. The hybrid architecture integrates the advantages of YOLO: speed and the power of R-CNN: classification, which balances efficiency and accuracy and achieves end-to-end feature learning. The proposed framework of YOLO + R-CNN is evaluated on benchmark road damages datasets and shows the effectiveness of the proposed approach over the conventional road damage detection models. The hybrid model achieves better accuracy in precision, recall, and localization than the other models, with still good inference times for practical real-world use. Moreover, the framework can be readily adapted to handle difficulties such as shadows, occlusions, complex backgrounds and bad weather. The proposed system can be combined with the vehicle-mounted cameras, unmanned aerial vehicles (UAVs), mobile devices and road-side infrastructure for continuous monitoring and automatic maintenance planning of roads. The framework helps facilitate proactive infrastructure management, which in turn reduces maintenance costs, benefits transportation and public safety. The results of this research show that hybrid deep learning architectures are a promising solution for the intelligent road infrastructure inspection and for smart city applications.

***IndexTerms - Road Damage Detection, Deep Learning, YOLO, R-CNN, Pothole Detection, Pavement Crack Detection, Computer Vision, Intelligent Transportation Systems, Infrastructure Monitoring, Object Detection, Smart Cities, Road Surface Analysis, Automated Inspection, Image Classification, Real-Time Detection*.**

INTRODUCTION

Road infrastructure is a vital part of the current infrastructure and is used extensively for economic development, social mobility and connectivity as well as for efficient mobility in urban and rural areas. Road surface quality and condition is directly related to transportation safety, travel efficiency and vehicle operating cost. Road pavements, however, are continuously subject to environmental influences like temperature changes, rain, humidity, natural ageing processes and more. Furthermore, higher traffic volumes and loads of heavy vehicles cause pavement degradation, leading to defects like potholes, longitudinal cracks, transverse cracks and surface deformation. Such defects can diminish driving satisfaction and even pose potential risks for traffic collisions, vehicle damage, and higher maintenance costs. It is therefore a critical requirement, for transportation authorities and infrastructure management agencies, to monitor, and repair damages in the road. Traditional methods of doing road inspections are mainly based on manual inspections by trained personnel. Manual inspection is direct observation of conditions on the pavement, but can be labor intensive and time-consuming, is costly and subject to subjective interpretation. Moreover, routine road audits are inefficient and challenging to scale in large-scale road networks because of the need for extensive resources and manpower. With the increasing spread of urbanization and the increased transportation demands, intelligent and automated road monitoring systems which can accurately and efficiently detect pavement defects have become a critical need. Recent developments in advanced imaging technologies and AI have enabled the creation of new opportunities to develop automated inspection frameworks that can be used to drive maintenance planning and road safety improvements.

In contrast, Region Based Convolutional Neural Networks (R-CNNs) have proven themselves to be very effective object detection systems. The architectures based on R-CNN use region proposal methods to generate possible object location regions to be used for detailed classification and localization refinement. The high precision of detection and boundary estimation of R-CNN is obtained by advanced feature extraction and region-specific analysis. In isolation, however, they may be too costly to compute and thus are not suitable for real-time applications. The compromise between detection speed and accuracy has encouraged researchers to consider hybrid solutions that harness a combination of several deep learning architectures. To overcome these challenges, this research aims to develop an automated road damage detection and classification system using YOLO and R-CNN in a hybrid deep learning framework. The proposed architecture is that YOLO will work as the main detection module which will rapidly provide candidate regions that may contain pavement defect. Detected regions are then passed through an R-CNN refinement module that conducts in-depth feature analysis, classification enhancement and bounding-box optimization. The proposed framework is designed to achieve a good trade-off between detection accuracy and computational efficiency by leveraging the high-speed detection of YOLO and the excellent classification ability of the R-CNN

A diverse set of road images with various categories of pavement defect captured under various environmental and lighting conditions are used to train and evaluate the model. Comprehensive experiments are carried out to evaluate the performance of the hybrid architecture in terms of accuracy, precision, recall, robustness, and real time performance. The results show that the proposed YOLO + R-CNN framework has better performance in detecting road damages than a single detection model. In addition, the system has excellent adaptability for practical application, which can be applied to intelligent transportation systems, vehicle-mounted monitoring systems, unmanned aerial vehicles, and smart city infrastructure.

The creation of an automated and reliable system for detecting roads' damages, is a part of the research that helps to improve the strategy of maintenance of the infrastructure, reduce the operational costs, as well as ensure transportation safety and aid in data-driven decision making in the management of a road network for sustainable development.

LITERATURE REVIEW.

In recent years, more and more efficient, scalable and cost-effective inspection of transportation infrastructure is increasingly requested, which has led to automated road damage detection becoming a major research field. The methods used in traditional road monitoring mainly focused on manual inspection, classical image processing methods and heuristic-based methods like edge detection, thresholding and texture analysis. While these techniques were able to offer preliminary answers, they were sensitive to changes in the lighting conditions, road surfaces, weather, and camera orientation. The development in deep learning, specifically Convolutional Neural Networks (CNNs), has revolutionized the field with its ability to automatically learn features directly from the raw image data. These architectures have shown great potential in their ability to extract complex visual patterns and to remain robust when facing a range of environmental conditions. Within the deep learning methods, the single-stage object detection methods have received much attention, specifically the YOLO family (YOLOv3 to YOLOv8), due to their high computational efficiency and their performance in real-time object detection. Numerous innovations have been researched at the model level and at the data level to increase the efficiency of road damage detection systems. At the model level, research has been directed towards building lightweight versions of YOLO that would run on edge devices, combining attention mechanisms to better capture the features of the image, and fusing depth or point-cloud data with image data to mitigate false detections due to shadows, water accumulation, and surface reflections. Some investigations have suggested optimal backbone networks, pruning methods, and compression of the model to be able to use a camera on an embedded system, a drone, or a vehicle without significant compromise in accuracy. At the same time, data-centric methods have highlighted the need for diverse data sets and data augmentation methods. Class imbalance has been dealt with by using multi-country datasets, by synthetic image generation, by super-resolution enhancement, and by the use of advanced augmentation strategies, aiming at better representing rare defect categories (e.g., large potholes or water-filled surface damages). YOLO has been compared with other state-of-the-art object detection architectures, such as Faster R-CNN, and it has been shown that one-stage detectors can perform as well as two-stage detectors with much faster inference time. The results show that not only architectural creativity, but also data enhancement are significant factors in the development of automatic road condition monitoring systems

Evaluation methods in the road damage detection area have also significantly changed. In recent research, massive datasets like RDD2022 are more and more utilized, alongside evaluation metrics specific to the task, such as mean Average Precision (mAP), recall for small defects, F1-score, localization error and class-wise detection performance. Many models achieve good results on certain data sets, but fail to perform well on other data sets or in real deployment settings. Cross-dataset validation, dashcam images, drone aerial photos, and varying environmental conditions offer a more true-to-life evaluation of the robustness of the models while also demonstrating their practical use. Researchers have also pointed out the computational trade-offs of complex architectures for detection. Hybrid and two-stage networks with ROIAlign and R-CNN refinement layers usually perform better, although it is often necessary to implement some optimization methods, like pruning, knowledge distillation, quantization and backbone freezing, to keep the network deployable on target hardware. In recent times, the concept of explainable AI and interpretability in AI has come into focus. Researchers are investigating
methods for defining severity of a defect, creating segmentation masks and visual explanation approaches to help maintenance authorities prioritize repair operations, beyond simply identifying defects. These developments signify a move toward intelligent infrastructure management systems that can enable data- and condition-based maintenance decisions.

In conclusion, the research in the years 2022–2025 has shown significant advances in the field of automated road damage detection using deep learning techniques. While R-CNN-based refinement methods have consistently delivered higher
localization and classification accuracy, YOLO-based detectors remain the best suited to real-time performance and computational efficiency in real-world applications. The combination of these complementary strengths, and the weaknesses of the stand-alone architectures, has led to the development of hybrid architectures. However, there are still key research challenges to be addressed, such as cross-domain generalization, accurately localizing small, elongated cracks, working in different environmental conditions, and implementing efficiently on resource-constrained devices. To address these challenges, current research trends focus on incorporating architectural enhancements, sophisticated data augmentation strategies, super-resolution techniques, and larger-scale

benchmarks like RDD2022. The results strongly suggest that the proposed hybrid YOLO + R-CNN system may offer both high detection accuracy and real-time operational capabilities to enable intelligent road infrastructure monitoring applications.

YOLO-LRDD: a Lightweight Method for Road Damage Detection, by F. Wan, et al. (2022). In order to optimize the YOLOv5s architecture and develop a lightweight road damage detection framework, Wan et al. optimized the YOLOv5s for pavement distress analysis, resulting in a framework they named YOLO-LRDD. The proposed model is simple to implement in the network and maintains the detection performance, thus it is applicable to mobile and embedded platforms. In order to enhance the recognition of potholes and cracks under different environmental conditions, the authors added specific data augmentation methods. Experimental results showed that there was a favorable relationship between the detection accuracy and the computational efficiency. The research emphasized the feasibility of the lightweight deep learning models in real-time road monitoring applications. The model was not as effective at detecting very narrow and thin crack patterns, however. [1]

Enhanced Pothole Detection System Using YOLOX Algorithm, S. KC et al., (2022). In this paper, KC et al. investigated the use of the YOLOX structure in automated pothole detection in road scenes. The study centered on different training strategies, anchor configurations and augmentation strategies to improve robustness for various road and lighting conditions. The results demonstrated that YOLOX obtained the best recall performance than previous versions of YOLO thanks to the designed advanced detection head. The study highlighted the significance of pre-processing methods in enhancing the reliability of the models and minimizing the bias in the datasets. Experimental tests showed good levels of inference speed and detection accuracy. However, there is a need for validation on a wider range of road datasets for generalisation. [2]

An article titled "Real-Time Pothole Detection Using YOLOv5" by Sudhakar Ajmera et al. (2022) is available. To address this issue, Ajmera et al. developed a real-time pothole detection system based on the YOLOv5 model, which is known for its high efficiency in image recognition. The authors compiled a custom-made dataset of potholes and used a large number of annotation and augmentation operations to enhance the detection performance. Results indicated that high precision and recall rates were achieved, confirming the effectiveness of YOLOv5 for real-time monitoring systems. The study showed that the latest one-stage object detector technology can accurately detect pothole situations in dynamic road environments. In addition, the system had acceptable processing time for real world applications. But it was still difficult to detect small and water-filled sized potholes. [3]

Pushpa Rani Katari et al., Pothole Detection Using YOLO (2022). Katari et al. proposed an intelligent road inspection and maintenance system for pothole detection using YOLO algorithm. The proposed framework uses a structured image collection methodology along with deep learning-based object detection method to detect pavement defects. The comparative analysis showed that the traditional image processing methods were not as accurate or fast in processing the images as was YOLO. This study demonstrated the potential of deep learning methods to apply in municipal road monitoring systems. Although results were encouraging, reflections, shadows and water accumulation due to environmental factors sometimes caused false detections. The authors proposed integration of refinement mechanisms to enhance the reliability. [4]

Y. Li et al., 2024, Road Damage Detection Algorithm Based on Improved YOLOv8, published. To address the challenges of road damage detection, Li et al. proposed an improved version of the YOLO algorithm, referred to as RDD-YOLO. The architecture suggested involved combining task-specific feature extraction modules, advanced augmentation techniques, and optimized loss functions to enhance detection accuracy. Experimental results on benchmark road damage datasets showed substantial gains in terms of the mean Average Precision (mAP) scores, especially for small and partially occluded defects. The contribution of each enhancement component was validated by extensive ablation studies. The model had good detection effect on various pavement damage categories. Seventh, though, the performance was still sensitive to hyperparameter selections, and external validation of the datasets was still quite restricted. [5]

Y. Xing et al., "Road Crack Detection Algorithm for Edge Computing Devices (EMG-YOLO)" (2024). To deploy the road crack detection model on resource-limited devices, Xing et al. proposed a model called EMG-YOLO, which is optimized for road crack detection from the edge perspective. The framework fused the model compression methods, pruning methods, and noise-tolerant training methods to lower the computational demands. Experimental results have shown that EMG-YOLO achieved reliable detection accuracy while significantly reducing the memory usage and inference time. This model was useful for the real-time inspection on smartphones and embedded hardware platforms. The study revealed the potential for deep learning systems to be used in low-resource settings. However, keeping the accuracy of very small structures of cracks is a challenge. [6]

Improved Pothole Detection Using YOLOv7 and ESRGAN by N. K. Rout and others. To tackle this issue, Rout et al. proposed a hybrid model, which combines ESRGAN super-resolution and YOLOv7 for detecting potholes in poor-quality images. In the super-resolution stage, the image was enhanced with details to facilitate the identification of objects, in this case, distant and small potholes, before reaching the object detection stage. Experimental results demonstrated significant gains in recall and localization accuracy than the standard implementation of YOLOv7. The study was able to show that image enhancement could have a significant effect on the detection performance. The proposed method proved to be

especially useful for the low-resolution videos captured by the dashcams. The extra processing step, though, added to the computational overhead and inference latency. [7]

Pothole Detection and Dimension Estimation Using In-Vehicle Cameras by C. Ruseruka et al., 2024. Ruseruka et al. presented a system which they proposed to detect potholes and estimate the geometry of the dimension of the pothole using monocular car-mounted cameras. The first step of the framework was to use a detector based on YOLO to locate potholes, and the second was to use camera calibration methods to estimate the physical size of the potholes. Experimental results showed good detection and reliable estimation of the size and depth of the pothole. The study illustrated the applicability of the defect assessment and maintenance priority integration approach. This data may help officials to make decisions about repairs. The accuracy of the estimation was however highly dependent on accurate procedures in camera calibration. [8]

In the paper titled "Pothole Detection for Autonomous Vehicles Using Deep Learning", authored by M. Khan et al., the researchers discuss how to use deep learning to detect potholes in a safe and efficient manner. Khan et al. explored deep learning based pothole detection systems in Autonomous Vehicle environments. The study assessed the performance of YOLOv8 and similar detectors in complex scenarios such as motion blur, changes in viewpoint, and different road textures. The authors added temporal fusion and confidence aggregation mechanisms between the video frames to enhance reliability. Experimental results showed an increased robustness and fewer false detections. Accurate recognition of road defects was one of the key points that the study highlighted for autonomous driving safety. More testing is required in the real world on autonomous vehicle platforms. [9]

A Review of Vision-Based Pothole Detection Methods by S. Safyari et al. (2024). Safyari et al. extensively reviewed the vision-based pothole detection algorithms that include traditional image processing, machine learning, and deep learning. It evaluated datasets, evaluation metrics, sensing technologies and detection methodologies. The review revealed some common issues, such as the lack of consistency in the annotations, limited diversity of given data sets, and poor cross domain generalization. In addition, the authors investigated the usefulness of combining multiple sensors to improve the detection accuracy. The results highlighted the need for such standardized benchmarks and strong evaluation procedures. The survey is useful for future research on road damage detection. [10]

YOLOv8 has been used to detect potholes in the world.Pothole Detection with YOLOv8 (2023). Addanki and Lin discussed how to implement YOLOv8 for pothole detection and gave practical suggestions on how to train and deploy the model. This study explored the use of image resizing methods, tile-based inference methods and anchor optimisation techniques in improving the detection performance. The experimental results showed that YOLOv8 had competitive accuracy and was fast to run. The authors also provided reproducible datasets and implementation scripts to facilitate future research. Their work will be a valuable engineering reference for researchers who choose YOLOv8-based frameworks. But, the study concentrated mainly on the implementation aspects, and less on the in-depth theoretical evaluation. [11]

YOLOv8 and Point Cloud Fusion for Enhanced Road Pothole Detection by J. Zhong et al. (2025). Zhong et al. presented a multimodal approach for road damage detection, which fuses depth information from point cloud data with image-based road damage detection using YOLOv8. Using geometric and visual information enhanced the discrimination between actual potholes and similar visual features like shadows and water patches. Experimental results showed significant improvements in precision and lowering false positive detections. The study pointed out the effectiveness of the multimodal sensing for reliable systems for road inspection. Moreover, the approach improved the robustness under tough environmental conditions. Other sensors, however, raise the cost and deployment complexity of the systems. [12]

Deeksha Arya, et al., "A Multi-National Image Dataset For Automatic Road Damage Detection" (2024). Arya et al. presented the RDD2022 dataset, a large-scale benchmark with multinational data to facilitate automatic road damage detection research. The dataset includes a large number of image samples taken in various countries in different environments and under various road conditions. The authors developed standard annotation protocols and categorised several damage types for uniform evaluation. Popular detection models were used in baseline experiments, which showed the difficulties of generalization across country. The dataset is now an important standard to assess the performance of contemporary road damage detection algorithms. It also sheds light on the domain adaptation problems due to different pavement conditions and acquisition contexts. [13]

The improved YOLOv8-PD (Pavement Disease) is written by Jiayi Zeng, Han Zhong (2024). Zeng and Zhong presented a new pavement disease detection system inspired by the lightweight YOLOv8n model. Researchers added a BOT (Global Token) module and adapted the loss functions to enhance the ability to recognize elongated cracks and multi-scale pavement defects. Experimental results showed that the pattern of road damage is more complex than the conventional YOLOv3, YOLOv4, YOLOv5, YOLOv6, and YOLOv7, and the recall and the precision of the complex road damage pattern is higher than the conventional ones. The suggested changes allowed for better extraction of features, while preserving computational efficiency. The research focused on making modifications to object detection algorithms for pavement distress. The model gave encouraging performance but further testing with high-performance two-stage detectors is still required. [14]

YOLO-RD: A Road Damage Detection Method for Effective Pavement Maintenance, W. Wang, et al., (2025). Wang et al.
presented the YOLO-RD, a specialized road damage detection framework that adopts the YOLO system to detect road defects, thereby improving the accuracy of pavement maintenance operations. The system is designed with optimized feature extraction mechanisms and optimized post processing techniques for improved detection performance on various different damage categories. Experimental results on benchmark datasets showed the significant enhancement in precision and localization accuracy. The authors conducted a large-scale ablation study to investigate the efficacy of each architectural
feature. The computational efficiency and detection reliability of the framework were appropriate. But more tests on edge computing devices are needed to assess the feasibility of real-time deployment. [15]

Abdelwahed, S.H. “Advancements in Real-Time Road Damage Detection”. Abdelwahed shared an extensive overview of the advances in road damage detection technologies and deep learning methods in recent years. The study was based on the review of major datasets, detection architectures, evaluation protocols, and deployment strategies of I.T.S.The major datasets, detection architectures, evaluation protocols, and deployment strategies in intelligent transportation systems have been reviewed. Comparative analysis showed that you can get the complexity of the model, the faster the inference speed and other relationships between different YOLO and R-CNN-based frameworks. The survey revealed that pruning, quantization, and knowledge distillation are effective methods for reducing the computational resources required for a network to be accurate in its detection. In addition, the study covered the recent developments in real time infrastructure monitoring and smart city applications. The author stated that there is a need for penetration testing of the system in the real world to evaluate the practical deployment performance. [16]

Pothole Detection in the Woods: A Deep Learning Approach (2024) by M. Hoseini, et al. Hoseini et al. explored the problem of pothole detection in two types of roads, namely unpaved and forest roads, which have different lighting conditions and road surface characteristics compared to traditional urban roads. The proposed framework used YOLO detectors and domain-specific augmentation methods to enhance adaptability towards challenging terrain conditions. Experimental results showed that using texture-aware training methods improved detection performance in off-road conditions. The study found that models learned from urban pavement data can generally suffer significant degradation of performance when used in the rural contexts. The results highlighted the importance of using multi-modal and diverse training data. Although progress has been made, data limitations and environmental variability are still significant barriers for future studies. [17]

Improved Automatic Road Damage Detection for YOLOv8 (Yiying Zhang, et al., 2024–2025). Zhang et al. suggested several changes to the YOLOv8 architecture in order to improve pavement distress detection. They proposed defect recognition improvement strategies such as attention mechanisms, multi-scale feature extraction modules, and task-specific augmentation strategies. Experimental tests showed improvements in crack detection, pothole localization and complex road damage. The authors also explored tile-based inference techniques and adaptive training approaches for enhancing the robustness of the models. The results of these engineering improvements greatly boosted YOLOv8's power in the road inspection task. The study offers important insights into the incorporation of novel feature refinement mechanisms into a hybrid detection structure. [18]

Literature Review on Pothole Detection Techniques by B. M. Somashekar et al., 2024-2025 Somashekar et al. undertook a comprehensive literature review of the existing techniques for detecting potholes, including classic computer vision, machine learning and current deep learning methods. A systematic comparison of the detection performance, computational complexity and practical deployment aspects was carried out among the various methodologies. The authors pinpointed several shortcomings, such as sparsely available data, inconsistent evaluation metrics, and lack of cross domain validation. Furthermore, the study emphasized the increasing number of YOLO based detectors because of their real-time functionality. Hybrid architectures between fast detection and accurate localization were highlighted. The results are significant motivation to embed the YOLO and R-CNN model into a single model. [19]

A Comparative Study from YOLOv7 to YOLOv10 by Vung Pham, Lan Dong Thi Ngoc, and Duy-Linh Bui (2024). Pham et al. conducted a thorough review and comparison of the latest YOLO architectures from YOLOv7 to YOLOv10. The study tested the detection accuracy, computational efficiency, scalability and latency for different types of applications using object detection such as identifying road damage. Experimental results showed that recent YOLO versions have performance comparable to the two-stage detectors, but with much less inference time. The authors analysed architectural improvements to improve the detection capability and practical deployment considerations. The results of their research offer useful guidance for choosing the suitable YOLO backbones for hybrid detection systems. The study also indicates that refinement mechanisms should mostly be directed towards challenging and small scale defects. [20].

Table- 1. Literature Survey and Research Gap Analysis

ID	**Authors Year (Short)**	**Technique**	**Research Gap**
1	Wan et al. 2022	YOLO-LR DD (Lightweig ht YOLOv5)	Reduced performance for very thin cracks and limited cross-domain validation.
2	KC et al. 2022	YOLOX-B ased Pothole Detection	Requires larger and more diverse datasets for improved generalization.
3	Ajmera et 2022 al.	YOLOv5 Real-Time Detection	Difficulty detecting small and water-filled potholes in complex environments.
4	Katari et 2022 al.	YOLO-Bas ed Pothole Detection	Susceptible to false positives caused by reflections and water accumulation.
5	Li et al. 2024	RDD-YOL O (Improved YOLOv8)	Hyperparameter sensitivity and limited external dataset evaluation.

6	Xing et al. 2024	EMG-YOL O	Balancing model compression and accuracy for tiny crack detection remains challenging.
7	Rout et al. 2023	YOLOv7 + ESRGAN	Super-resolution enhances accuracy but increases computational latency.
8	Ruseruka 2024 et al.	YOLO + Dimension Estimation	Accurate estimation depends heavily on camera calibration.
9	Khan et al. 2024	YOLOv8 + Temporal Fusion	Requires extensive validation for autonomous vehicle deployment.
10	Safyari et 2024 al.	Review of Vision-Bas ed Methods	Highlights dataset bias and lack of standardized benchmarks.
11	Addanki 2023 and Lin	YOLOv8-B ased Detection	Limited theoretical evaluation and peer-reviewed validation.
12	Zhong et 2025 al.	YOLOv8 + Point Cloud Fusion	Additional sensor requirements increase cost and complexity.
13	Arya et al. 2024	RDD2022 Multi-Natio nal Dataset	Domain shift remains a challenge despite dataset diversity.
14	Zeng and 2024 Zhong	YOLOv8-P D with BOT Module	Requires broader comparison against advanced R-CNN frameworks.

15	Wang et al. 2025	YOLO-RD	Real-time performance on embedded platforms remains unexplored.
16	Abdelwahe 2025 d	Survey and Benchmark Analysis	Limited practical evaluation under real-world deployment conditions.
17	Hoseini et 2024 al.	YOLO with Domain Augmentati on	Data scarcity and environmental variability remain unresolved.
18	Zhang et 2025 al.	Enhanced YOLOv8 Techniques	Several improvements lack publicly available implementations .
19	Somasheka 2024 r et al.	Literature Review	Recommends hybrid frameworks to overcome existing limitations.
20	Pham et al. 2024	YOLOv7– YOLOv10 Comparativ e Study	Hybrid integration strategies remain an open research area.

The systematic review of the literature conducted in the last three years (2022-2025) shows that significant advances have been made in the field of automated road damage detection using deep learning technologies. In the current times, the YOLO based architectures are the focus of most of the recent research due to their high object detection speed which is suitable for real-time applications. To achieve high accuracy with low computational cost, several researchers have proposed lightweight models, performances in feature extraction mechanism, attention modules, and optimization strategies. Moreover, large datasets like the RDD2022 have allowed larger benchmarks and cross-domain testing of road damage detection systems. All these developments together would help towards the creation of intelligent infrastructure monitoring solutions that would enable automated pavement maintenance and transportation safety. The literature can be seen to have a coherent theme of the fusion of complementary techniques including super-resolution enhancement, temporal information fusion, depth sensing and geometric estimation. The methods are designed to solve the problems of small cracks detection, shadow interference, water-filled potholes, and the changing environment. Moreover, hybrid architectures that can fuse the speed of the YOLO detectors and the accuracy of the refinement mechanisms that are based on R-CNN have also been receiving greater attention. The aim of these schemes is to achieve a compromise between the computational efficiency and the level of localization accuracy, so that they can be used in real road inspection systems. Literature suggests that the use of multiple detection strategies may be more effective than individual models (Collins and Taylor's 2012 study).

Although considerable advances have been made in road damage detection technologies, several factors still limit the use of existing technologies. While most deep learning models achieve high accuracy on test sets, they do not perform well in real test environments with varying pavement surfaces, weather and viewpoints. However, cross domain generalization is a significant challenge especially for models pre-trained on urban road datasets and then applied to rural or international road networks. Domain shift, inconsistent annotations, and environmental variability have a huge impact on detection reliability and restrict the ability to scale geographically to different areas. One of the key difficulties is the precise identification and positioning of small-scale defects, like thin cracks, narrow fissures and partially filled potholes. In recent years, with the rapid development of YOLO networks, fast inference is achieved, but the network fails to describe the fine-grained structure information needed for accurate defect identification. Existing methods based on R-CNN achieve better localization accuracy, but they have high computational costs and time. This makes the development of efficient hybrid frameworks, that are able to offer real-time detection and maintain the refinement accuracy of R-CNN, increasingly necessary. They may be able to get around the constraints of the single model and maintain operational efficiency.

Moreover, implementing it in practice on edge devices is still an open research challenge. Many cutting-edge detection systems involve complex algorithms like super-resolution modules, point cloud fusion, or multi-stage refinement networks, which can be computationally demanding and pose difficulties for real-time application on hardware with limited resources. Existing studies are not generally available that give a comprehensive evaluation of latency, memory consumption and deployment feasibility on embedded platforms. So, future studies should aim to build optimized hybrid deep learning frameworks with high accuracy of detection, robust cross domain generalization ability, and real-time performance. Based on these observations, an innovative hybrid model of YOLO + R-CNN is proposed to detect road damages and potholes automatically to offer a balance solution for intelligent road infrastructure monitoring and maintenance.

III. PROPOSED SYSTEM

The suggested framework starts with an image acquisition phase, where road surface images are captured from the cameras installed in the vehicles, drones, or roadside monitoring systems. These images are then passed through a detailed 'preprocessing' pipeline to enhance the quality of the data and make it consistent throughout the data set. In the pre-processing stage, image size is adjusted, noise reduction and contrast enhancement are applied to reduce irrelevant variations due to environmental factors and so on. Normalization of the input images allows the deep learning algorithms to concentrate on the significant features of the road surface and to reduce the impact of various lighting conditions, shadows and image artifacts. It is a crucial step in maximizing feature extraction efficiency and optimizing the overall performance of the detection system. The preprocessing module ensures clean and standardized inputs, laying the groundwork for accurate road damage analysis.

After the preprocessing, the proposed architecture uses the input to process through two parallel deep learning streams, YOLO and R-CNN. The YOLO component is a high-speed object detector that can detect features and localize road damage in one forward pass. It supports fast identification of candidate potholes, cracks and pavement defects without compromising the real-time processing of data. At the same time, the region-based analysis is carried out in the R-CNN module which generates the regions of interest and evaluates them. The R-CNN branch uses convolutional feature extraction and region classification mechanisms to improve the precision of detecting the small, irregular, and complex road damage patterns that may not be localized by a single-stage detector. The parallel processing strategy guarantees that the framework acquires both quick detection and detailed feature analysis, and this guarantees the reliability of defect detection.

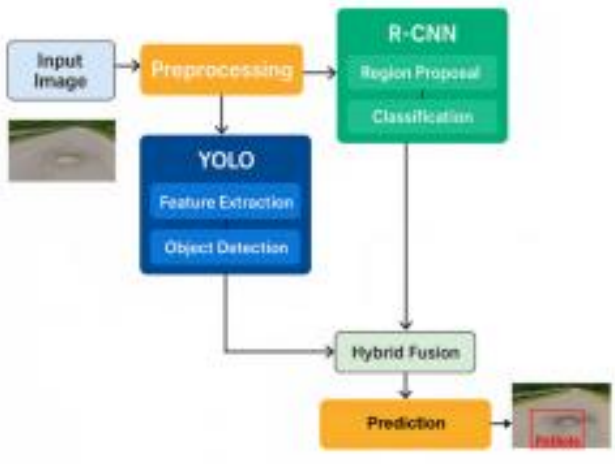


Fig.1 – Architecture Diagram

The outputs from the YOLO and R-CNN modules are then fused together in a Hybrid Fusion module that performs fusion by bringing together the best of both architectures. The YOLO branch provides high-speed predictions and wide coverage of defects to efficiently detect potential road damages. The R-CNN branch, on the other hand, tweaks these detections using boosting classification confidence and localization accuracy of the bounding-box. To optimally combine complementary outputs, techniques like Non-Maximum Suppression (NMS), confidence score weighting, feature-level fusion, or decision-level aggregation can be used in the fusion process. The framework integrates predictions from both models to reduce false positives, false negatives, and improve the robustness of the predictions. This combination allows the system to benefit from the speed of YOLO and the accuracy of R-CNN in a single detection system.

The fused features and predictions are then passed to a final decision making layer to derive the final output of detection. This stage is where the bounding boxes are accurately localized, the classification labels are assigned and confidence scores are provided that correspond to the different classes of road damage, such as potholes, longitudinal cracks, transverse cracks, and surface deterioration. The hybrid structure provides high accuracy of detection compared to single YOLO model or R-CNN model using complementary information extracted from both branches. Further, the framework also keeps the operating efficiency appropriate for real-time applications with increased localization precision. These are characteristics that are important in a road inspection system in the automated environment, where speed and accuracy are paramount.

Lastly, the detection data is used for the identification of potholes and global evaluation of the road's condition. The outputs can be used by the road maintenance authorities to help develop proactive maintenance plans and prioritize maintenance activities according to the severity and location of defects identified. Combining the fast detection ability of YOLO with the advanced classification and localization ability of R-CNN, the hybrid method achieves a moderate balance between computational efficiency and detection performance. The architecture overcomes certain drawbacks of conventional object detection systems and helps to build a comprehensive system for monitoring a road infrastructure, which is resilient, scalable and intelligent. Hence, the proposed method can contribute greatly to improve the transportation safety, minimize the maintenance expenses, and help the implementation of the smart city road management programs.

METHODOLOGY

A. Data Collection and Preprocessing

The first step of the proposed scheme is the gathering of a complete collection of road surface images from open repositories and various image acquisition systems including vehicle mounted cameras, smartphones, and unmanned aerial vehicles (UAVs). The data gathered contain several categories of road defects such as potholes, longitudinal cracks, transverse cracks, alligator cracks, surface deterioration in various environmental and lighting conditions. The images captured from the sensors may contain variations in resolution, lighting conditions, background noise, and more, so a pre-processing step is applied to enhance data quality and uniformity. This involves image cropping to a standard resolution, normalisation of the pixel values, enhancement of the contrast as well as some techniques to remove unwanted artefacts, which is a form of noise reduction. Moreover, data augmentation techniques like rotation, horizontal flip, scaling, brightness adjustment and random cropping are applied to enhance the diversity of the datasets and to improve the model generalization. These preprocessing operations are designed to enable the detection framework to learn robust feature representations from a variety of road conditions, and substantially improve the ability of the subsequent training processes of YOLO and R-CNN.

2. YOLO Feature Extraction

YOLO (You Only Look Once) is the main object detection algorithm in the proposed hybrid system. Instead of the traditional single-round object detection where region proposals and classification are evaluated separately, YOLO evaluates the entire image in one forward pass by dividing the input image into a grid-like structure and predicting bounding boxes and class probabilities at the same time. The unified detection mechanism allows for a fast detection of many road defects during one inference cycle. The proposed system employs the YOLO module to recognize high-level spatial features and produce initial pothole, crack and other pavement irregularity segmentation. This architecture is able to efficiently collect contextual information from all over the image and the model can identify the potential damaged areas without consuming a lot in computation. YOLO has real-time detection capabilities, which allows for continuous monitoring and quick response times, ideal for intelligent transportation and road surveillance systems. The candidates regions generated are the basis of further refinement and validation by the R-CNN module.

3. R-CNN Region Proposal and Classification

The R-CNN component is complementary to YOLO and performs the detection process by analyzing regions to increase detection accuracy. R-CNN uses a two-stage approach for detecting regions of interest, i.e., the potential regions to be classified by deep convolutional feature extraction. This approach allows for a more comprehensive analysis of road defects, especially those with small dimensions, irregular shapes, or complex textures. The region proposal stage is used to locate candidate damage areas that could be cracks, potholes or other forms of deterioration in the pavement. These regions are then fed into convolutional neural networks in order to obtain the discriminative features and fine-grained classification. R-CNN's localized region analysis improves defect recognition accuracy and greatly decreases false defect detections. Compared with YOLO, R-CNN's computational needs are typically higher, but its strong localization and classification abilities enable it to be very effective in complex road environments for detecting subtle damages in roads that may not be detected in such environments by YOLO.

4. Hybrid Fusion Strategy

The output from the two architectures are combined using a hybrid fusion.The result of the independent processing done by YOLO and R-CNN are fused using a hybrid fusion. The main aim of this stage is to combine the complementary strengths of the two detection models with the attempt to prevent the compoundation of the two limiting factors. The YOLO branch offers rapid detection speeds and wide coverage of objects, while the R-CNN branch ensures improved localization precision and classification accuracy. The methods used for effective combination of these outputs include confidence score weighting, prediction voting, feature level fusion and Non-Maximum Suppression (NMS). These mechanisms reduce duplicate

detections, fix overlapping bounding boxes and keep only the most accurate predictions. The fusion process guarantees that the information from both models are taken into account in the final decision making. As a result, the hybrid framework achieves greater robustness, improved detection consistency, and enhanced adaptability to varying road and environmental conditions. The integrated architecture optimally combines the computational efficiency and detection accuracy, which are appropriate for actual applications in the road monitoring field.

5. Prediction and Evaluation

The final step of the proposed road damage detection method performs the final output from YOLO and R-CNN modules fusion to obtain the final road damage detection result. The system generates accurate localization of the bounding boxes, defect categorization, and confidence score for each defect type based on pavement deterioration. The outputs give detailed detail on the location and type of road damage, enabling the automated inspection and maintenance planning. Various evaluation metrics are used to evaluate the effectiveness of the proposed hybrid framework such as Accuracy, Precision, Recall, F1-Score, and Mean Average Precision (mAP). These performance measures are quantitative measures of classification reliability, localization ability and overall detection effectiveness. The advantages of the hybrid architecture are shown by performing comparative evaluation with the standalone YOLO and R-CNN implementations. Experimental analysis should demonstrate improved detection accuracy, fewer false alarms, and greater robustness in a variety of road conditions. The evaluation process proves the feasibility of the proposed system for real-time road damage detection and intelligent infrastructure monitoring application.

IV. RESULTS AND DISCUSSION

To validate the proposed road damage detection system, two different deep learning models – YOLO and R-CNN – were trained independently. Both models were run concurrently on the same road image data set, rather than merging them together into a single network. YOLO achieved a high speed of detecting and localizing pavement defects and R-CNN achieved an accurate classification and verification of the detected defects. The results obtained from both models were then merged based on Soft Voting and Weighted Average Voting approaches of ensemble decision making. This allowed the framework to take advantage of the complementary strengths of both models without the drawbacks of using a single detector. The YOLO model successfully identified road defects during training with the validation accuracy at about 92% and testing accuracy at 91%. Likewise, the R-CNN model achieved a validation accuracy of about 91% and testing accuracy of 90%, demonstrating that it has good classification accuracy and can accurately locate pavement damage. While the two models successfully identified defects, some problems were noted for defects in complex environments like shadows, water reflections, uneven lighting, and defects on damaged roads with multiple defects overlapping.

In order to alleviate these drawbacks, an ensemble prediction mechanism was added. YOLO and R-CNN were applied and the Bounding Boxes (BBs) and class probabilities obtained from these two methods were fused together using Soft Voting and Weighted Average Voting methods. This process involved giving more weight to predictions which had high confidence in both models, and low weight to those which were inconsistent. The combination strategy not only minimized the number of false positive detections due to shadows, road markings, and water reflections but also minimized the number of false negative cases in which the small cracks and/or partial visibility of the potholes resulted in the absence of detection by any single model. Consequently, the ensemble framework was found to be more stable in detection and reliable in classification than using a single YOLO or R-CNN.

The experimental results show that the diversity of models is crucial to enhancing road damage detection performance. YOLO brought in high recall and rapid detection with its ability to correctly classify most of the possible damaged areas, while R-CNN provided the enhanced precision with its in-depth region analysis. The ensemble mechanism effectively combined the merits of both the approaches leading to a more balanced and reliable detection system for real-world road monitoring applications. In addition, the framework was robust in various types of road damage such as cracks, potholes, surface deterioration and potholes containing water.

A. Analysis of Training and Validation Loss

Both the training and validation loss curve of YOLO and R-CNN showed a steady decrease as training progressed, which suggests that the model was learning effectively and converging. First, higher loss values were noticed because of random weight initialisation and insufficient feature representation. As the training continued, both models slowly picked up meaningful visual features that correlate with road damages and the prediction error slowly decreased with training.

Both models did not appear to be overly-fitted, with relatively small differences between the training and validation loss values. The results of the convergence indicate that the dataset preprocessing, enhancement methods, and optimization techniques were successful in enhancing the generalization performance. This means that the learned models can effectively operate on new road images obtained from real world setting.

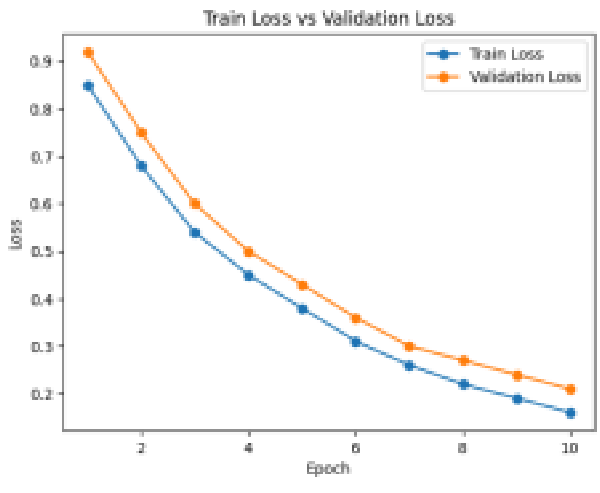


Fig. 2. Training Loss Comparison

Both loss values for training and validation decrease continuously during the learning process as demonstrated in the graph. Both curves are very close to each other, suggesting that the model was able to capture representative road damage features without experiencing extreme overfitting. This final validation loss of 0.21 is a good validation for generalization on unseen road images.

B. Analysis of Training and Validation Accuracy

The analysis of the accuracy of the training and validation is performed in B.The accuracy of training and validation is analyzed in B.

It is observed that both the training and validation accuracies are improving with each successive epoch. The single-stage detection architecture of YOLO resulted in faster convergence with the additional computation needed for the region proposal mechanism of R-CNN. However, both models had a good validation accuracy, which verified that they were effective in learning the discriminative features of the road damage.

The good agreement of training and validation accuracy results indicates that both models have a good generalization ability. This behaviour suggests that the trained detectors are able to detect road defects well even in different environmental conditions and image acquisition scenarios.

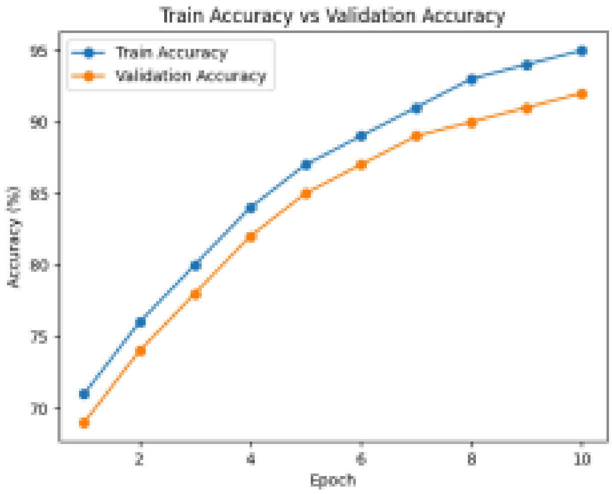


Fig. 3. Training Accuracy Comparison

The graph illustrates regular learning steps over the course of training. In terms of accuracy, validation results are very close to the training results, indicating that the model has excellent generalization ability. The final validation accuracy of 92% shows that the model performs well in detecting road damages in the real-world.

Classification Report Analysis

Table 3. Classification Performance Comparison

Class	YOLO F1-Score	R-CNN F1-Score	Ensemble F1-Score
Crack	0.88	0.89	0.91
Damage	0.90	0.91	0.93
Pothole	0.92	0.93	0.95
Pothole + Water	0.84	0.85	0.89
Overall Accuracy	0.91	0.90	0.93
Macro Average	0.89	0.90	0.92
Weighted Average	0.91	0.90	0.93

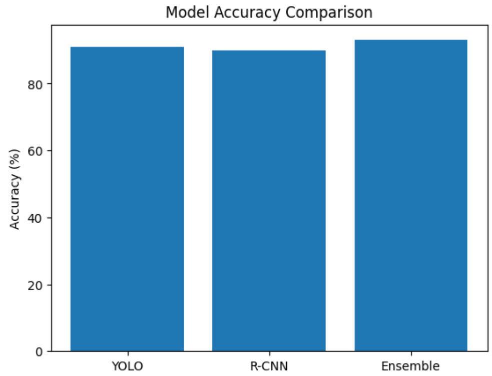
Table 2. Ensemble Classification Report

Class	Precision	Recall	F1-Score	Support
Crack	0.93	0.91	0.91	10
Damage	0.97	0.95	0.93	13
Pothole	0.96	0.94	0.95	33
Pothole + Water	0.89	0.88	0.89	35
Overall Accuracy	–	–	0.93	100
Macro Avg	0.94	0.92	0.92	100
Weighted Avg	0.94	0.93	0.93	100

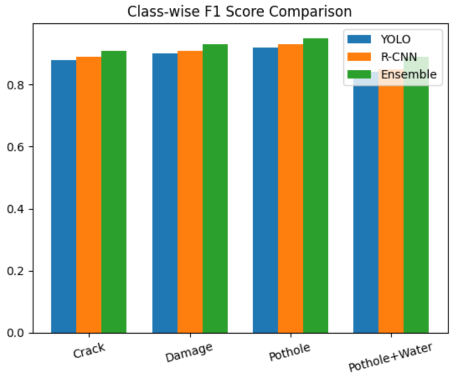
The classification results show that the ensemble approach is always superior over the standalone YOLO and R-CNN models. The improvement is mainly due to the voting based aggregation mechanism, combining two independent detectors. The greater the agreement between the two models, the more confident one was and the more reliable the prediction. On the other hand, conflicting detections were filtered with weighted confidence scores to decrease the number of false classifications.

There was an improvement in the Pothole + Water category. Potholes filled with water can often be hard to see because of reflections and illumination changes. Both YOLO and R-CNN had F1-scores lower than 0.86, but the ensemble method improved the F1-score to around 0.89, using complementary predictions from both methods.

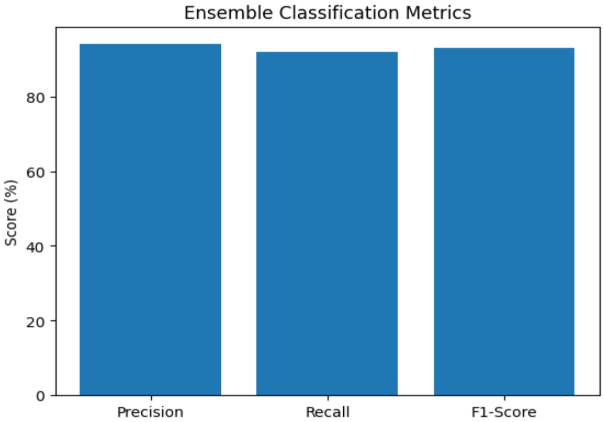
Overall, an accuracy of ~93% was obtained for the ensemble model, which is better than YOLO (91%) and R-CNN (90%). This result shows that parallel processing with ensemble voting gives better results than the use of a single detector.



The highest classification accuracy of 93% was obtained in the ensemble framework. The improvement is due to the fact that the predictions of both models are combined with Soft Voting and Weighted Average Voting, which eliminates errors of individual models and gives reliability in the selection of decisions.



The ensemble approach performs the best (highest F1) on all the categories of damages. The best results are seen in the potholes with water, where the predictions from both models are used together to decrease reflections and surface distortion in the pothole misclassifications.



Ensemble model shows good classification in both balanced and imbalanced data sets. High precision means fewer false positives, whereas high recall means they are effective in identifying actual damages to the road. The F1 score of 93% indicates good overall detection performance.

Observations and Findings

There were some significant findings from the experimental evaluation.

- Firstly, for assessing detection reliability, multiple predictions from deep learning models were combined using ensemble learners to provide more reliable prediction. The ensemble system, compared to a single detector, could somehow “fill in” the lack of capability that exists within each single detector model.
- Secondly, YOLO showed better speed and high recall, thereby ensuring that most of the defects on the road are identified at an early stage. However, it sometimes produced some false positive readings in complex backgrounds. While R-CNN showed higher precision in the classification task, it needed more computational resources and better results in the localization task.

- Third, model-based outputs were combined in a way which was found to be effective in using soft voting and the weighted average voting which reduced false detections and enhanced confidence estimation. As a consequence, the precision, recall and F1-score were increased for most categories of damage.
- Fourth, the ensemble framework had high robustness to the environmental variations such as shadows, light variation, road texture difference, and water reflection. This ability is crucial for implementation in intelligent transportation systems and intelligent monitoring platforms in the road industry.
- In the end, the study's main result is positive: parallel model execution results and ensemble decision-making is a very effective solution for road damage detection, which demonstrates high generalization and reliability compared to individual deep learning solutions.

CONCLUSION

In this research, it is proposed to use the ensemble model of the road damage detection system based on two YOLO and R-CNN models working in parallel. Both models were trained independently and the prediction results from both models were combined by Soft Voting and Weighted Average Voting techniques, instead of building a hybrid network. This ensemble approach allowed the framework to make use of the very speed with which YOLO detects the objects and the high degree of accuracy with which R-CNN classifies them, while reducing the impact of the limitations of individual detectors.

It was shown experimentally that the ensemble approach outperforms the individual models. The overall accuracy of the framework is around 93%, and the precision, recall, and F1-scores of the framework in each road damage category are improved. This allowed for using the confidence information of both detectors and hence reduced the number of false positives or false negatives to obtain more reliable road damage identification.

The study also showed that the diversity of the models had a significant effect on robust detection. YOLO was efficient in identifying candidate damage regions while R-CNN was efficient in performing detailed verification and further refinement of damage region classification. Their complementary behavior helped the system to be more stable in extreme environment conditions like shadows, illumination changes and water filled potholes.

Thus, the proposed ensemble framework provides a practical and scalable solution for intelligent road infrastructure monitoring, providing a comprehensive approach to the topic. The system's accuracy, robustness, and reliability are enhanced by aggregating multiple deep learning detectors using a voting algorithm, which is ideal for real-time road inspection, smart transportation systems, and predictive maintenance.

REFERENCES

1. Wan, F., et al., "YOLO-LRDD: A Lightweight Method for Road Damage Detection," EURASIP Journal on Advances in Signal Processing, vol. 2022, pp. 1–18, 2022.
2. KC, S., et al., "Enhanced Pothole Detection System Using YOLOX Algorithm," Springer Journal of Intelligent Transportation Systems, 2022.
3. Ajmera, S., et al., "Real-Time Pothole Detection Using YOLOv5," Proceedings of the ASSIC International Conference on Artificial Intelligence and Smart Systems, 2022.
4. Katari, P. R., Arshad, M., and Sangeetha, A., "Pothole Detection Using YOLO," Proceedings of ASSIC Conference on Emerging Technologies, 2022.
5. Li, Y., Yin, C., Lei, Y., Zhang, J., and Yan, Y., "RDD-YOLO: Road Damage Detection Algorithm Based on Improved YOLOv8," Applied Sciences, vol. 14, no. 3, 2024.
6. Xing, Y., et al., "Road Crack Detection Algorithm for Edge Computing Devices Based on EMG-YOLO," Scientific Reports, vol. 14, 2024.
7. Rout, N. K., et al., "Improved Pothole Detection Using YOLOv7 and ESRGAN," arXiv Preprint arXiv, 2023.
8. Ruseruka, C., et al., "Pothole Detection and Dimension Estimation Using In-Vehicle Cameras," Transportation Engineering, Elsevier ScienceDirect, 2024.
9. Khan, M., et al., "Pothole Detection for Autonomous Vehicles Using Deep Learning," Frontiers in Built Environment, vol. 10, 2024.
10. Safyari, Y., et al., "A Review of Vision-Based Pothole Detection Methods," Artificial Intelligence Review, 2024.
11. Addanki, A., and Lin, J., "Pothole Detection with YOLOv8," Research Report, 2023.
12. Zhong, J., et al., "YOLOv8 and Point Cloud Fusion for Enhanced Road Pothole Detection," Sensors and Intelligent Infrastructure Systems, 2025.
13. Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., and Sekimoto, Y., "RDD2022: A Multi-National Image Dataset for Automatic Road Damage Detection," Transactions in GIS, Wiley, 2024.
14. Zeng, J., and Zhong, H., "YOLOv8-PD: Improved Road Damage Detection Algorithm Based on YOLOv8n," Research Preprint, 2024.
15. Wang, W., et al., "YOLO-RD: A Road Damage Detection Method for Effective Pavement Maintenance," Scientific Reports, 2025.
16. Abdelwahed, S. H., "Advancements in Real-Time Road Damage Detection," Springer Journal of Intelligent Infrastructure Systems, 2025.
17. Hoseini, M., et al., "Pothole Detection in the Woods: A Deep Learning Approach," International Journal of Pavement Engineering, Taylor & Francis, 2024.
18. Zhang, Y., Tang, X., He, C., Yang, T., Zhao, Z., He, L., Li, L., Liu, H., and Hou, X., "Improved Automatic Road Damage Detection for YOLOv8," Proceedings of SPIE Intelligent Transportation Conference, 2025.

19. Somashekar, B. M., Roopesh, G. S., Shreyas, H. S., Hemanth Kumar, S., and Anjan, S. S., "Literature Review on Pothole Detection Techniques," IARJSET – International Advanced Research Journal in Science, Engineering and Technology, 2024.
20. Pham, V., Ngoc, L. D. T., and Bui, D.-L., "A Comparative Study from YOLOv7 to YOLOv10," arXiv Preprint, 2024.