International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4732-4739
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0622

RoadVision AI: An Intelligent Vision-Based Road Infrastructure Monitoring System Using YOLOv8 and Depth Estimation

Kruthika H K[1], Dr. Veena M N[2], Yogitha C[3], Rahul K[4], Rahul Kumar S M[5]
[1,3,4,5]Master of Computer Applications Student, Department of Master of Computer Applications, PES College of Engineering, Mandya, 571401, Karnataka, India
[2]Professor & HOD, Department of MCA, PES College of Engineering, Mandya, 571401, Karnataka, India.
Email ID: *kruthikagowda208@gmail.com[1], veenadishal@pesce.ac.in[2], yogithayogitha2526@gmail.com[3], rahulrk0412@gmail.com[4], smrahulkumar31@gmail.com[5]*

Abstract
Road defects such as potholes, damaged drain holes, and missing or broken sewer covers pose a persistent challenge to road safety, ride comfort, and vehicle maintenance costs. Field-based inspection, which remains the conventional method of assessing pavement condition, depends on manual observation by maintenance personnel and is therefore slow, subjective, and difficult to repeat at scale across large road networks. This paper presents RoadVision AI, a vision-based road infrastructure monitoring system that automates the detection and severity assessment of common road defects directly from images. The system performs detection in a structured pipeline. First, a trained YOLOv8 object detection model localizes potholes, drain holes, and sewer covers within an uploaded road image. Once a defect region has been identified, OpenCV is used to extract visual descriptors such as contour geometry, edge density, texture, and darkness from the cropped region, while the MiDaS monocular depth estimation model generates a relative depth map to estimate pothole depth without requiring dedicated depth sensors. The extracted visual descriptors and depth values are then fused to compute a Visual Severity Score and a Road Damage Index, from which the system determines the risk category, evaluates overall road health, and generates an appropriate maintenance recommendation. A Streamlit-based web application allows users to upload images, review annotated detections, and monitor historical trends through an interactive Plotly dashboard. The YOLOv8 model was trained on 3,813 annotated road images collected from Roboflow and achieved a precision of 91.20%, recall of 87.79%, mAP@50 of 91.25%, and mAP@50–95 of 56.70%. This framework contributes an integrated road-monitoring pipeline that combines object detection, depth-aware severity estimation, and dashboard-based analytics within a single deployed system.

Keywords: *Road Infrastructure Monitoring, YOLOv8, Pothole Detection, Depth Estimation, MiDaS, OpenCV, Visual Severity Score, Road Damage Index, Streamlit, Computer Vision.*

1. Introduction

Timely identification of road surface defects is essential for maintaining driver safety, protecting vehicles from avoidable damage, and controlling the long-term cost of pavement upkeep. Road agencies traditionally rely on field personnel who walk or drive along a stretch of road to visually catalogue potholes, damaged drain holes, and broken sewer covers. Because such inspections must cover extensive road networks, they are labor-intensive, slow to complete, and inherently dependent on the judgement of the individual inspector. As a result, defects are often reported and repaired only after they have worsened, increasing both the safety risk to road users and the eventual cost of remediation. Recent progress in computer vision and deep learning has opened a practical alternative to manual surveying by enabling models to recognize pavement defects directly from images with a level of consistency that human inspection cannot easily match. The introduction of the YOLO object detection framework demonstrated that deep neural networks could achieve real-time detection with high computational efficiency [1], and the subsequent Ultralytics YOLOv8 release extended this efficiency with an anchor-free detection head and improved feature extraction layers, making it well suited to applications that demand both speed and accuracy [2]. A number of researchers have since applied YOLO-based architectures specifically to road damage detection and have reported strong localization performance under a range of environmental conditions [11]–[14]. These studies confirm that deep learning-based object detection is a reliable substitute for manual pavement inspection. However, a closer examination of the existing literature shows that most systems stop at object localization and classification; comparatively few frameworks extend detection into a fuller condition assessment that also considers visual severity, depth, historical trends, and maintenance decision support. In addition, many existing pipelines are built exclusively around pothole detection and do not simultaneously track other infrastructure elements such as drain holes and sewer covers, even though these components are just as relevant to road safety. To close this gap, this paper introduces RoadVision AI, an intelligent vision-based road infrastructure monitoring system capable of detecting and analyzing multiple categories of road defects from a single uploaded image. The framework begins with a YOLOv8 detector trained to recognize potholes, drain holes, and sewer covers. Detected regions are then passed through an OpenCV-based feature extraction stage that characterizes their visual properties, while a MiDaS monocular depth estimation module estimates pothole depth from the same RGB image without the need for stereo cameras or LiDAR hardware [3], [4]. The visual descriptors and depth estimate are combined into a Visual Severity Score, which in turn feeds into a Road Damage Index that quantifies overall pavement condition on a 0–100 scale. Based on this index, the system determines a risk category, recommends a maintenance action, and stores every inspection record for later review. A Streamlit web interface makes the entire pipeline accessible to non-technical users, while an integrated Plotly dashboard visualizes defect distribution, severity trends, and road health statistics over time [19], [20]. By unifying detection, severity assessment, and analytics within one application, RoadVision AI offers a more complete decision-support tool for road maintenance authorities than detection-only alternatives.

2. Literature Review

Automated road condition assessment has attracted sustained research interest as the number of accidents and vehicle damages attributable to potholes, cracks, and deteriorating drainage infrastructure continues to grow. Early object detection research, including the original YOLO formulation by Redmon et al. [1], established that a single neural network could perform real-time detection with high computational efficiency, laying the foundation for later domain-specific applications such as road damage recognition. Architectural refinements introduced in subsequent years, including deep residual learning [6], Feature Pyramid Networks [7], and focal loss for dense object detection [8], improved the ability of detection networks to localize small and visually ambiguous objects, a capability that is particularly relevant when identifying subtle pavement defects against a cluttered road background. Building on these foundations, several researchers have applied YOLO-based detectors directly to road defect identification. Yurdakul and Tasdemir proposed an enhanced YOLOv8 model that improved both the localization accuracy and the measurement precision of detected potholes [11], while Zuo et al. incorporated feature enhancement techniques into a YOLOv8 pipeline to strengthen crack detection performance under varying lighting and weather conditions [12]. Khare et al. extended this line of work to a broader set of infrastructure classes, demonstrating a YOLOv8-based detector capable of recognizing potholes, sewer covers, and manholes within a single model [13], and Doshi and Yilmaz showed that ensembling multiple deep networks reduces false detections and improves robustness across challenging road scenes [14]. Taken together, these studies confirm that YOLO-family detectors generalize well to infrastructure monitoring tasks, though most remain focused on the detection step itself rather than on a fuller assessment of defect severity. A separate but complementary line of research has addressed the estimation of physical defect characteristics beyond simple bounding-box localization. Ranftl et al. introduced a monocular depth estimation framework capable of producing dense, zero-shot depth maps from a single RGB image, removing the dependency on specialized depth sensors and offering a practical route to estimating pothole depth from ordinary photographs [3]. The OpenCV library, described by Bradski, has similarly become a standard tool for extracting contour geometry, edge density, texture, and other low-level visual descriptors that support detailed characterization of detected regions [4]. Deployment-oriented frameworks have also matured alongside these modeling advances: The Ultralytics implementation of YOLO streamlines the training and inference workflow for object detection models [2], Streamlit enables rapid development of interactive web interfaces for real-time image analysis [19], and Plotly supports the interactive charting needed to visualize detection statistics and historical inspection trends [20]. Despite this progress, the literature reveals a persistent gap between detection-focused systems and truly integrated road condition monitoring platforms. Relatively few frameworks combine object detection with visual feature analysis, monocular depth estimation, quantitative severity scoring, historical analytics, and maintenance recommendation within a single deployed pipeline, and fewer still monitor multiple infrastructure classes simultaneously. RoadVision AI is designed to address this limitation by unifying YOLOv8-based detection, OpenCV-based visual feature extraction, MiDaS-based depth estimation, Road Damage Index computation, and dashboard-based analytics into one coherent, deployable system for road infrastructure monitoring.

3. Methodology

The proposed RoadVision AI framework follows a structured, multi-stage pipeline that begins with image acquisition and proceeds through preprocessing, object detection, visual feature extraction, depth estimation, severity scoring, and dashboard visualization. Each stage is designed to operate on a single uploaded image and to return an interpretable result within seconds, allowing the system to be used directly by maintenance personnel without specialized training. The subsections that follow describe the dataset, the preprocessing pipeline, the detection and analysis modules, and the deployment architecture in turn.

3.1.Dataset Description

The framework is trained on a road infrastructure image dataset collected from Roboflow that spans three defect categories relevant to routine road maintenance: Pothole, Drain Hole, and Sewer Cover. The images were captured under diverse environmental conditions, including variation in illumination, road surface texture, camera viewing angle, and background clutter, so that the trained detector would generalize beyond a narrow set of visual conditions. Every image in the dataset was manually annotated with bounding boxes corresponding to its defect class prior to training. The complete dataset contains 3,813 annotated road images, divided into 3,076 training images, 491 validation images, and 246 testing images, as summarized in Table 1. This split was chosen to provide a sufficiently large training set for the detector to learn discriminative features while reserving separate validation and test partitions for unbiased performance evaluation.

Table 1 Dataset Composition Used for Training and Evaluation

Dataset Split	Number of Images
Training	3076
Validation	491
Testing	246
Total	3813

3.2.Image Acquisition

Road images are acquired through the Streamlit web interface, which accepts photographs captured from a smartphone, a vehicle-mounted camera, or any standard image source. Upon upload, the application verifies that the file conforms to a supported format such as JPG, JPEG, or PNG before it is passed further into the pipeline. This acquisition step is intentionally lightweight so that field personnel can capture and submit road images with minimal effort, supporting frequent and low-cost inspection cycles compared to scheduled manual surveys.

3.3.Image Preprocessing

Before being used for model training, the collected images During training, every input image was resized to 640 × 640 pixels, and pixel values were normalized automatically by the YOLOv8 preprocessing pipeline. Data augmentation techniques, including random horizontal flipping, scaling, translation, rotation, mosaic augmentation, and color-space transformation, were applied to expose the model to a wider range of visual conditions than the raw dataset alone could provide. The model was trained for 50 epochs with a batch size of 8.

3.4.Dataset Preparation

Prior to training, the annotated dataset was organized into the training, validation, and testing subsets described in Table 1, ensuring that no image appeared in more than one partition. This separation allows the model's learning progress to be tracked on the validation set during training while the testing subset remains reserved for a final, unbiased assessment of detection performance. Maintaining balanced coverage of all three defect classes across the subsets was an important consideration in preparing the dataset, since an uneven distribution could otherwise bias the detector toward the majority class.

3.5.YOLOv8-Based Road Defect Detection

The primary detection component of RoadVision AI is built on the YOLOv8 (You Only Look Once, version 8) object detection architecture. YOLOv8 was selected for this task because of its high detection accuracy, fast inference speed, and ability to perform localization and classification in a single forward pass. Relative to earlier YOLO variants, YOLOv8 employs an anchor-free detection mechanism together with refined feature extraction layers and an optimized prediction head, which together improve its sensitivity to small and irregularly shaped defects such as potholes. The trained model identifies three infrastructure classes: Pothole, Drain Hole, and Sewer Cover. Each predicted bounding box is represented as B = (x, y, w, h), where x and y denote the coordinates of the box center and w and h denote its width and height. Localization accuracy is evaluated using the Intersection over Union (IoU) metric:

$$\mathbf{IoU = Area\ (B_p \cap B^g) / Area\ (B_p \cup B^g)}$$

where B_p is the predicted bounding box and B^g is the corresponding ground-truth box. To convert the raw network outputs into class probabilities, YOLOv8 applies the Softmax function:

$$\mathbf{P_i = e\hat{}(z_i) / \Sigma_j\ e\hat{}(z_j)}$$

where P_i denotes the probability assigned to class i and the summation runs over all C candidate classes. In the proposed system, C = 3, corresponding to Pothole, Drain Hole, and Sewer Cover.

3.6.OpenCV-Based Visual Feature Analysis

Once a defect has been localized, the corresponding Region of Interest (ROI) is cropped from the original image and passed to an OpenCV-based analysis stage that extracts additional information about its physical appearance. Edge detection and contour extraction are applied to the cropped region to derive descriptors such as bounding box area, contour area, contour perimeter, edge density, texture score, aspect ratio, shape irregularity, darkness score, and water reflection characteristics. These descriptors are combined into a single Visual Severity Score (VSS):

$$\mathbf{VSS = \Sigma_i\ w_i\ F_i}$$

where F_i is the normalized value of the i-th visual feature extracted from the detected region, w_i is the corresponding feature weight, and n is the total number of visual features considered. The computed VSS is subsequently combined with the estimated pothole depth to determine the final severity level and risk category.

3.7.MiDaS-Based Monocular Depth Estimation

To assess pothole severity beyond what surface-level visual features alone can capture, RoadVision AI incorporates the MiDaS monocular depth estimation model to estimate the relative depth of detected potholes directly from a single RGB image. Unlike stereo cameras, LiDAR, or other dedicated depth-sensing hardware, MiDaS generates a dense depth map using a deep neural network trained for zero-shot cross-dataset transfer, making it practical to deploy on ordinary photographs captured in the field. The estimated depth values are normalized as:

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4732-4739
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0622

$$D_{norm} = (D - D_{min}) / (D_{max} - D_{min})$$

where D is the estimated depth value at a given pixel, and D_{min} and D_{max} denote the minimum and maximum depth values observed across the generated depth map. Based on the resulting normalized depth, detected potholes are grouped into Low, Medium, and High severity categories, providing a physically grounded complement to the surface-level Visual Severity Score computed in the previous stage.

3.8. Road Damage Index and Severity Assessment

The outputs of the detection, feature extraction, and depth estimation stages are ultimately combined to compute the Road Damage Index (RDI), a single quantitative measure of pavement condition expressed on a 0–100 scale, where higher values indicate greater deterioration and a correspondingly higher maintenance priority. The RDI is computed as a weighted combination of normalized severity parameters:

$$\mathbf{RDI = \Sigma_i\, w_i\, S_i}$$

where S_i is the normalized severity score of the i-th evaluation parameter, w_i is its corresponding weight, and n is the total number of parameters considered. Based on the computed RDI, the system classifies overall road condition into five health levels — Excellent, Good, Fair, Poor, and Critical — and uses this classification, together with the underlying severity level, to determine a risk category and generate an appropriate maintenance recommendation, ranging from routine monitoring to immediate pothole repair, drainage cleaning, or sewer cover replacement.

3.9. Performance Evaluation

Detection performance was evaluated using standard object detection metrics, namely precision, recall, mAP@50, and mAP@50–95, computed on the held-out testing subset described in Table 1. Training and validation losses were monitored across all 50 epochs to confirm stable convergence, and the resulting confusion matrix was examined to assess class-wise detection reliability across the Pothole, Drain Hole, and Sewer Cover categories. The complete quantitative results of this evaluation are reported in Section IV.

3.10. Streamlit-Based Web Deployment

RoadVision AI is deployed as a web-based application built on the Streamlit framework, which provides an accessible interface for uploading road images and reviewing analysis results with a single click. After an image has been processed by the detection, feature extraction, and depth estimation modules, the application displays the annotated image alongside the detected defect category, confidence score, estimated pothole depth, Visual Severity Score, Road Damage Index, risk level, road health assessment, and a corresponding maintenance recommendation. Each detection result is automatically recorded in a CSV file, allowing an inspection history to be maintained for future reference. These stored records feed an interactive Plotly-based dashboard that summarizes defect distribution, severity levels, Road Damage Index trends, and overall road health statistics over time, enabling maintenance authorities to monitor conditions across multiple inspections rather than viewing each image in isolation. An optional offline voice notification feature, implemented using the pyttsx3 text-to-speech library, further alerts users audibly whenever a high-risk defect is identified, supporting faster response in time-sensitive field conditions.

4. Results and Discussion

The performance of RoadVision AI was evaluated using the annotated dataset described in Section III, which was partitioned into training, validation, and testing subsets to provide a reliable assessment of detection performance across all three defect categories. The YOLOv8 model was trained for 50 epochs using an input resolution of 640 × 640 pixels and a batch size of 8. Throughout training, both training and validation losses decreased steadily, while precision, recall, and mean average precision improved gradually, indicating that the model was learning discriminative features for potholes, drain holes, and sewer covers without exhibiting signs of overfitting. The final evaluation on the testing subset produced a precision of 91.20%, a recall of 87.79%, an mAP@50 of 91.25%, and an mAP@50–95 of 56.70%, as summarized in Table 2. The relatively high precision indicates that most objects flagged by

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4732-4739
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0622

the detector correspond to genuine road defects, while the recall value shows that the majority of actual defects present in the test images were successfully identified. The gap between mAP@50 and the stricter mAP@50–95 metric reflects the added difficulty of achieving tight bounding-box overlap across a wider range of IoU thresholds, which is expected given the irregular shapes of many potholes and partially occluded sewer covers.

Table 2 Performance Metrics of the Trained YOLOv8 Detection Model

Performance Metric
Precision
Recall
mAP@50
mAP@50–95
Training Epochs

An examination of the confusion matrix generated during evaluation showed that potholes, drain holes, and sewer covers were correctly classified in the large majority of cases, with only a small number of misclassifications. Detection performance was particularly strong for potholes, which formed the largest portion of the dataset, while the remaining errors were concentrated in challenging conditions such as uneven lighting, visually similar surface textures, and partial occlusion of the defect region. These findings are consistent with results reported elsewhere for YOLOv8-based road damage detectors [11]–[14] and suggest that the current level of performance is well suited to practical deployment, while also highlighting environmental robustness as a natural direction for further improvement. Beyond detection accuracy, the value of RoadVision AI lies in the additional analysis performed once a defect has been located. The OpenCV-based feature extraction stage and the MiDaS-based depth estimation stage jointly determine the Visual Severity Score and the estimated pothole depth, which are then fused into the Road Damage Index to describe overall pavement condition on a continuous 0–100 scale. This severity-aware output allows the system to differentiate between a shallow, cosmetic pothole and a deep, hazardous one, a distinction that a detection-only pipeline cannot express. The Streamlit-based dashboard further extends this capability by aggregating individual inspection records into class-wise defect distributions, severity trends, and road health summaries, giving maintenance authorities a longitudinal view of road condition rather than a series of disconnected snapshots. Overall, the results confirm that combining YOLOv8-based detection with OpenCV feature analysis, MiDaS depth estimation, Road Damage Index computation, and dashboard analytics yields a system that is both quantitatively accurate and practically informative. The measured precision and recall demonstrate that the underlying detector is reliable across all three defect classes, while the severity and dashboard layers translate raw detections into the kind of prioritized, actionable information that road maintenance planning requires.

Conclusion

This paper presented RoadVision AI, a vision-based road infrastructure monitoring system that integrates YOLOv8 object detection, OpenCV visual feature extraction, and MiDaS monocular depth estimation to analyze road images and assess pavement condition. The system detects potholes, drain holes, and sewer covers; estimates pothole severity using a combined Visual Severity Score and depth-based analysis; computes a Road Damage Index; evaluates overall road health; and generates maintenance recommendations accordingly. A Streamlit-based dashboard further stores every detection record and presents interactive visualizations that support long-term monitoring rather than isolated, one-off inspections. Experimental evaluation showed that the trained YOLOv8 model achieved 91.20% precision, 87.79% recall, 91.25% mAP@50, and 56.70% mAP@50–95, demonstrating reliable detection performance across all three defect categories on the held-out test set. These results, together with the additional severity and analytics layers built on top of detection, indicate that RoadVision AI offers a more complete decision-support solution for road maintenance authorities than detection-only alternatives reported in prior work.

Future Enhancement

Future work will focus on improving detection robustness under challenging environmental conditions such as poor lighting, heavy shadowing, and wet road surfaces, which remain the primary source of residual detection errors observed in the current evaluation. Expanding the dataset with a larger and more diverse collection of road images, including additional defect types beyond potholes, drain holes, and sewer covers, would further improve the generalizability of the trained model. The framework could also be extended to support real-time video-based monitoring from vehicle-mounted cameras rather than single-image uploads, enabling continuous inspection during routine driving. Integrating GPS metadata with each detection would allow defects to be geo-tagged and mapped, supporting route-level prioritization of maintenance work. Finally, deploying the system on mobile platforms and validating it against inspection data from multiple road authorities would help assess its generalizability across different pavement types and climatic conditions.

Acknowledgement

The authors express their sincere gratitude to Dr. Veena M N, Professor & HOD, Department of Master of Computer Applications, PES College of Engineering, Mandya, for her guidance and continued support throughout the course of this research. The authors also thank the Department of Master of Computer Applications, PES College of Engineering, Mandya, for providing the infrastructure and resources necessary to carry out this work.

References

[1]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 779–788, 2016.
[2]. G. Jocher, A. Chaurasia, and J. Qiu, Ultralytics YOLO Documentation, Ultralytics, 2023.
[3]. R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun, "Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 3, pp. 1623–1637, 2022.
[4]. G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.
[5]. W. McKinney, "Data Structures for Statistical Computing in Python," Proc. 9th Python in Science Conf., pp. 56–61, 2010.
[6]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Proc. CVPR, pp. 770–778, 2016.
[7]. T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," Proc. CVPR, pp. 2117–2125, 2017.
[8]. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," Proc. ICCV, pp. 2980–2988, 2017.
[9]. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Int. Conf. Learning Representations (ICLR), 2015.
[10]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[11]. M. Yurdakul and Ş. Taşdemir, "An Enhanced YOLOv8 Model for Real-Time and Accurate Pothole Detection and Measurement," 2025.
[12]. H. Zuo, Z. Li, J. Gong, and Z. Tian, "Intelligent Road Crack Detection and Analysis Based on Improved YOLOv8," 2025.
[13]. O. M. Khare, S. Gandhi, A. M. Rahalkar, and S. Mane, "YOLOv8-Based Visual Detection of Road Hazards: Potholes, Sewer Covers, and Manholes," 2023.
[14]. K. Doshi and Y. Yilmaz, "Road Damage Detection using Deep Ensemble Learning," 2020.
[15]. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," Proc. CVPR, pp. 248–255, 2009.
[16]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, 2017.
[17]. R. Girshick, "Fast R-CNN," Proc. ICCV, pp. 1440–1448, 2015.
[18]. J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic

Segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 4, pp. 640–651, 2017.
[19]. Streamlit Inc., Streamlit Documentation, 2024.
[20]. Plotly Technologies Inc., Plotly Python Graphing Library Documentation, 2024.