Towards Safer Roads: Intelligent Pothole Detection with YOLOv11

M. Lakshmi Chetana[1], P. Hemanth Sai Vikas[2] and Sivadi Balakrishna[3]
{chetanamanne7165@gmail.com[1], hemanthvikas.pingali@gmail.com[2], drsivadibalakrishna@gmail.com[3]}

Department of Advanced Computer Science and Engineering, Vignan's Foundation for Science, Technology and Research, Vadlamudi, Guntur – 522213, Andhra Pradesh, India[1, 2, 3]

Abstract. The preservation of road infrastructure and the safety of vehicular traffic are critical challenges in modern transportation systems. Potholes, in particular, pose significant risks to road users and contribute to accidents, making their timely detection and repair essential. This study proposes the concept design and realization of an intelligent automated pot- hole detection system utilizing YOLOv11, A cutting-edge object detection model powered by deep learning algorithms. In this approach, deep learning algorithms are trained to detect potholes in image data collected from road surfaces under varying environmental conditions, including different lighting and weather scenarios. The real-time capabilities of YOLOv11 enable accurate and rapid detection, making the system suitable for integration into autonomous vehicles, traffic monitoring systems, and road maintenance operations. The system is developed using Python, OpenCV, and deep learning frameworks such as TensorFlow and PyTorch. Its performance is evaluated against Important measurements like processing speed, intersection over union (IoU), and mean average precision (mAP) are used in order to measure both accuracy as well as computing speed. The framework for detecting potholes, which was created, can help facilitate the improvement of intelligent transportation systems through the incorporation of automation, resulting in enhanced road safety and simpler maintenance procedures.

Keywords: Pothole detection, YOLOv11, Object detection, Computer vision.

1 Introduction

Roads are the most popular means of travel in India, with millions reliant on them every day. Still, the increasing figure of road accidents over the last several years has become a very important issue of concern for safety. Among the key factors behind these accidents is the existence of potholes, which can initiate sudden stops, vehicle instability, and life-threatening injuries, especially for two-wheeler drivers and pedestrians [1].

Potholes usually occur as a result of a combination of heavy traffic, poor road maintenance, harsh weather conditions such as heavy rainfall, and ground movements. When roads are not properly maintained, tiny cracks appear and gradually develop into big potholes due to the continuous pressure of vehicles and water infiltration. These dangers not only lead to serious damage to vehicles but also pose a threat to human lives. To counter this, recent research has been aimed at creating automated and precise pothole detection systems [2].

ICITSM-Part I 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357949

Deep learning object detection models, particularly real-time models such as YOLO (You Only Look Once), have been highly promising in this area. The progressive development of YOLO models from initial versions to more light-weight and precision-oriented variants such as YOLOv8 has made practical deployment possible on a range of platforms ranging from drones to handheld devices. Research has also examined the possibilities for extending models like YOLOv11 to more general traffic safety use cases, including the detection of accidents, highlighting their potential for smart infrastructure [3].

In addition, recent developments have incorporated edge computing and digital twin technologies under the Internet of Vehicles (IoV) model for improving pothole detection in real time. Through processing information at the edge, such systems minimize latency while facilitating instantaneous identification of hazards. The addition of digital twins ensures constant monitoring and predictive maintenance and thus the response and dependability of vehicle networks in solving road anomalies [4]. A sample image from dataset is illustrated in Fig. 1.

Fig. 1. Sample image from dataset.

A comprehensive comparison of different YOLO architectures highlights the trade-offs between speed, accuracy, and computational complexity, making it easier to select the most appropriate model for specific real-world use cases [5].

Furthermore, recent surveys on multi-object detection models underline the increasing relevance of such technologies in intelligent transportation systems, offering scalable and efficient solutions for urban mobility challenges [6].

Building upon these advancements, the present work aims to develop a pothole detection system using the latest YOLOv11 framework, focusing on real-time performance, model efficiency, and suitability for Indian road conditions.

This paper's remaining sections are organized as follows: The literature on object and pothole detection methods is reviewed in Section II. The suggested approach, dataset characteristics, model structure, and implementation plan are all defined in Section III. The experimental setup, findings, and performance evaluation of the YOLOv11-based system are described in Section IV. Section V wraps up the research and suggests possible directions for further investigation.

2 Related works

The identification of potholes is extremely crucial for safe travel on roads. Detection of potholes has now become much easier compared to traditional methods like manual inspections and vibration-based techniques, which are time-consuming and expensive. With the help of advanced technologies like computer vision and deep learning, Automated pothole detection has emerged as a promising area in research.

N.Bhavana et al. [7] developed a POT-YOLO, a real-time pothole detection system built upon an edge segmentation- enhanced YOLOv8 architecture. In their methodology, they first converted pothole videos into image frames. To reduce distractions in the converted images, they applied Contrast Stretching Adaptive Gaussian Star Filter (CAGF). For detecting pothole regions, the Sobel detector was used. In performance evaluations, the YOLOv8-based model demonstrated superior results compared to models similar to Faster R-CNN and Mask R-CNN.

Li et al. [8] proposed a pothole detection system that utilizes crowdsourced information integrated with an enhanced Mask R-CNN model. The integration of multiple data sources improves model generalizability across different road conditions. The enhanced Mask R-CNN model enhances segmentation precision for accurate pothole area identification. The paper shows the potential of integrating community- sourced data with state-of-the-art deep learning methods for infrastructure inspection. This approach provides a scalable solution for real-time road maintenance applications.

Wang et al. [9] proposed an intelligent pipeline drainage defect detection system based on an enhanced YOLO- DeepSort framework. The system is a combination of the fast object detection feature of YOLO and DeepSort's tracking algorithm to track defects over time. This combination makes it possible to continuously evaluate the condition of the pipelines, making it possible for timely maintenance. The method illustrates the potential in using the fusion of detection and tracking algorithms for infrastructure health monitoring. The research indicates advancements in detection precision and working efficiency in pipeline inspections.

Rout et al. [10] has proposed a hybrid system by combining Enhanced super Resolution GAN(ESRGAN) with YOLOV7 variants(standard,tiny,X) for detecting potholes in low-resolution dash cam footage. In their methodology they processed 1784 Images (1265 training) through ESRGANS Residual-in-Residual Dense blocks to upscale 640X360 pixels inputs to 1100X800 pixels before YOLOv7 analysis. The Yolov7X model achieved best performance on the other hand the smaller YOLO v7 tiny variant maintained real-time capability at 0.0093 per frame. The study demonstrated ESRGAN ability to enable high accuracy detection from low cost cameras.

Sai et al. [11] introduced a YOLOv8 based system to detect potholes automatically using a dataset of 631 road images. Ambiguous pothole images were eliminated to reduce overfitting on irrelevant data. Following this, three YOLOv8 models—Nano, Small, and Medium—were trained. Among them, the YOLOv8-Nano model demonstrated the most optimal performance compared to the other two models.

Aparna et al. [12] developed a pothole detection system where they have used thermal imaging and Convolutional Neural Networks (CNN'S) In their proposed methodology at first they collected images of roads under various conditions (day/night, water-filled, dry etc.) using the FLIR ONE camera. The collected Images were resized into (240X295 pixels format) by cropping and resizing to standard dimensions. To enhance the size of the dataset some data augmentation techniques including zooming, rotation and noise injection were applied . Both custom CNN models and fine-tuned pre-trained ResNet models (ranging from ResNet-18 to ResNet-152) were implemented in the study. The ResNet-101 model specifically utilized input images resized to 224×224 pixels for training and evaluation.

Dhiman and Klette et al. [13] introduce a systematic work on pothole detection using automation, comparing classical stereo vision methods with deep learning methods. They test four techniques: two stereo-vision methods (SV1 and SV2) and two deep learning models (LM1 and LM2). The work identifies an important trade-off between fine segmentation for infrastructure inspection and real-time detection for application within vehicles. The work also considers issues of irregular pothole boundaries and environmental noise. Their research has established critical milestones and has greatly impacted developments in intelligent transportation systems.

Au Yang Her et al. [14] introduced a real-time pothole detection approach by enhancing the YOLOv5 object detection framework, specifically designed for motorized use on Malaysian roads. Their system utilized different YOLOv5 variants (including m6, s6, and n6) to analyze live video streams captured using an Intel RealSense D435i camera mounted on a moving vehicle. For onboard computing and fast inference, the configuration used a Nvidia Jetson Xavier X module. The approach combines deep learning-based detection with embedded hardware to enable real-time analysis. The study demonstrated YOLOv5's potential for real-time and on-vehicle pothole detection in Malaysian road conditions.

Dharneeshkar et al. [15] presented a deep learning framework aimed at detecting potholes with the help of YOLO variants to address road surface damages on Indian roads. Their methodology involved capturing road images and training object detection models such as YOLOv2, YOLOv3, and YOLOv3-tiny to identify potholes from 2D visual data. The framework makes use of the speed of convolutional neural networks and specializes in vision-based detection and presents a low-cost substitute to existing vibration or 3D reconstruction methods. The study emphasizes the adaptability of YOLO models for region-specific road monitoring applications in developing countries.

While various YOLO-based pothole detection systems show promising results, many rely on limited or region-specific datasets, which reduces their ability to generalize to different environments and road conditions. High-accuracy models like YOLOv8 and ResNet-101 require powerful GPUs, limiting their use on low-resource or edge devices. In contrast, lightweight models offer real-time processing but often compromise detection precision. Some methods also involve complex preprocessing steps or require specialized sensors like thermal cameras, increasing the cost and implementation com- plexity. Additionally, environmental factors such as lighting, shadows, and road texture can still affect detection accuracy. The absence of standardized benchmarks further complicates model comparisons and hinders the development of universally applicable solutions.

3 Methodology

The proposed workflow of Pothole Detection using YOLOv11-n is shown in Fig 2. The collected dataset, which already includes augmented images. This dataset is categorized in to training and validation sets to make it ready for model development. The YOLOv11-n variant is then initialized by setting the required parameters and declaring important hyperparameters. In the following stage, the model is trained by specifying the number of epochs and applying pre-trained weights to enhance its accuracy. The trained model is then tested using the validation data to determine its overall performance and reliability. Finally, key evaluation metrics including precision, recall, and mean Average Precision (mAP) are computed, and the model is utilized to perform inference on unseen images to effectively detect potholes.

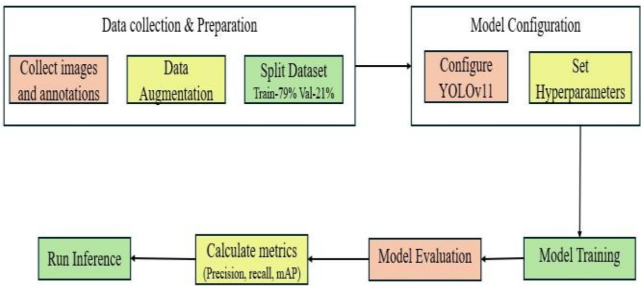

Fig. 2. Methodology.

3.1 Dataset

The data used in this research was obtained from Kaggle and augmented with real-time pothole images gathered from other online sources to provide diversity and realism. The image set contains a total of 3,226 images, including not just images with a single pothole but also images with multiple potholes in the same frame. Each image in the dataset is annotated using the YOLO format, where annotations follow the structure:

<class_id> <x_center> <y_center> <width>

<height>

, with all values representing normalized coordinates. The dataset already contains augmented images, so there is no need for further augmentation in preprocessing. The mean original image resolution was about 410×410 pixels; however, to improve model performance and detection quality, the input image size was increased to 624×624 pixels during training. The data was classified into training and validation sets of 2,602 images and 624 images, respectively.

3.2 YOLOv11

YOLO (You Only Look Once) is a popular real-time object detection framework that uses a single convolutional neural network to identify object classes and predict their bounding boxes in one unified step. YOLOv11 is the latest version in this series, developed by Ultralytics, and it sets a new benchmark with respect to accuracy, inference speed, and computational performance. Being an extension of its predecessors, YOLOv11 offers several architectural enhancements and optimization techniques to improve results achieved across various computer vision tasks

Fig 3 [3] illustrates the architecture of the YOLOv11 model, The architecture is structured into three core modules: Backbone, Neck, and Head.

The YOLOv11 architecture is organized into three fundamental components:the Backbone, Neck, and Head, each having a different purpose in the object detection pipeline. The Backbone takes charge of obtaining rich visual information from the input image by going through a cascade of convolutional layers and Cross Stage Partial (C2f) modules that enhance computational efficiency and gradient flow with shortcut connections. It is concluded with an SPPF (Spatial Pyramid Pooling - Fast) layer, which is responsible for capturing multi-scale features. The Neck serves as a feature aggregator, taking features from various levels of the backbone through upsampling and concatenation operations. Such integration enables the model to maintain fine-grained spatial information while including high-level semantic features, which improves its capability to detect objects of different sizes.

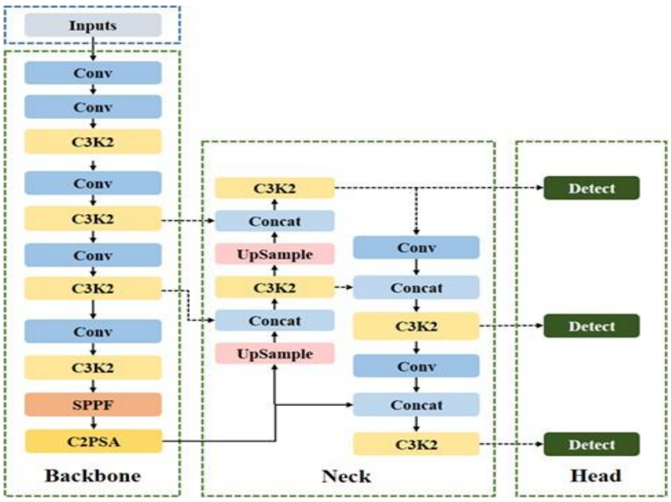

Fig. 3. Network architecture of YOLOv11.

Lastly, the Head predicts by processing the combined features from the Neck and generating bounding boxes and class probabilities at three scales (80×80, 40×40, and 20×20) to accurately detect small, medium, and large objects in one forward pass. Fig. 4 shows the Key architectural modules in YOLOv11 [16].

Besides its standard architecture, YOLOv11 integrates three powerful modules that play important roles in its performance:

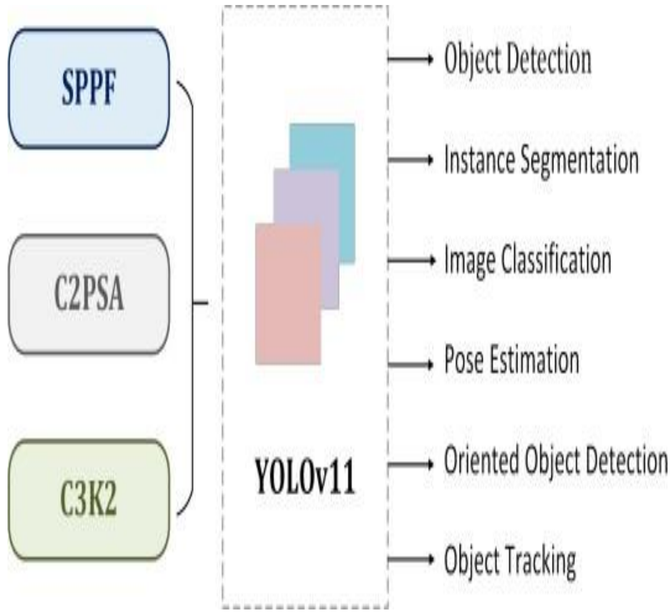

Fig. 4. Key architectural modules in YOLOv11 [16].

- SPPF (Spatial Pyramid Pooling-Fast): Enhances multi- scale feature representation by pooling features at different scales without increasing inference time.

- C2PSA (Cross-Stage Partial Self-Attention): Improves attention across spatial dimensions, this allows the model to more accurately focus on the important regions within the image.C3K2 (Efficient Convolutional Block): Increases the efficiency and speed of the model through optimized convolutional operations.

3.3 Training

To train the YOLOv11 model for pothole detection, we used a .yaml configuration file that contains essential details about the training dataset. This file defines the paths to the training and validation data, the number of object classes, and their corresponding names. The model was trained for 115 epochs to ensure it effectively learns the features. All input images were resized to 640×640 pixels, and a batch size of 64 was employed to expedite training process. The device was set to 0 to enable GPU usage, which further accelerates training. Additionally, rect=True was included to preserve the original aspect ratio of images during resizing.

3.4 Inference

Once the YOLOv11 model is trained, it is used for detecting potholes in new images through inference. The model takes an input image and returns bounding boxes along with confidence scores for each detected pothole, indicating the location and certainty of the detection. When source=0 is specified during inference, the system accesses the webcam and processes real- time images, allowing the model to perform pothole detection on live image feeds.

4 Results and Analysis

This section provides a detailed assessment of the proposed Pothole Detection using YOLOv11. The results are demonstrated using both quantitative metrics and visual outputs. The evaluation is carried out on a custom pothole dataset, which was preprocessed, annotated, and classified into train, validation and test sets. The YOLOv11 model was trained using the Ultralytics framework and tested on images from the validation set.

4.1 Input and Detection Output

Fig. 5. Sample Input image (before detection).

Fig. 6. Output image (after detection).

A test image is provided, illustrating a real-world road scenario with clearly observable surface deterioration. Fig. 5 shows the Sample Input image (before detection). Fig. 6 shows the Output image (after detection).

After processing the input image through the trained YOLOv11 model, the detected potholes are highlighted using bounding boxes. The model performs real-time inference and outputs the detected objects with corresponding confidence scores.

4.2 Model Training Analysis

The confusion matrix shows how many accurate and inaccurate predictions the model produced in comparison to the real annotations.

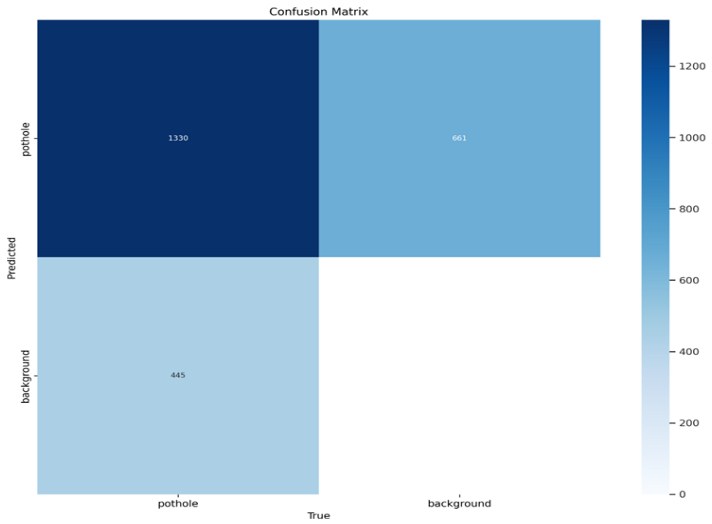

Fig. 7. Confusion matrix.

The confusion matrix shown in Fig 7. indicates that the YOLOv11 model successfully detects the majority of potholes with 1330 true positive predictions. However, there are 661 false positives and 445 false negatives, which suggest the model at times confuses potholes with similar background features.

When assessing object detection models, the Precision- Recall curve is essential, particularly when working with unbalanced data. It depicts how precision and recall fluctuate when the confidence threshold changes, helping students comprehend the trade-off between these two measurements.

We employed three main performance metrics precision, recall, and mean Average Precision (mAP) to assess the efficacy of the YOLOv11 model.

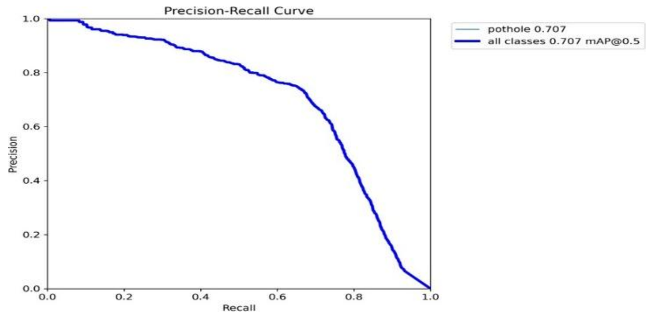

Fig. 8. Precision Recall Curve.

- **Precision:** Precision indicates the correctness of positive predictions. It informs us about how many of the pre- dicted potholes were indeed correct. Fig. 8 shows the Precision Recall Curve.

Formula:

$$precision = \frac{\textbf{Correct Detections}}{\textit{Correct Detections} + \textit{Incorrect Detections}} \quad (1)$$

- **Recall:** Recall indicates how well the model identifies all true potholes. It reflects the capability of the model to identify all objects of interest in the image.

Formula:

$$Recall = \frac{\textit{Correct Detections}}{\textit{Correct Detections} + \textit{Missed Detections}} \quad (2)$$

- **Mean Average Precision (mAP):** By combining precision and recall for every object class and different confidence thresholds, mAP generates a total performance score.

  To calculate mAP, we first determine the Average Precision (AP) for each class, which is the area of the Precision-Recall curve. The mAP is the average of all APs.

Formula:

$$mAP = \frac{1}{C}\sum_{i=1}^{c} AP_i \quad (3)$$

where C is the total number of object classes.

The training and validation loss curves for the YOLOv11- n model over 120 epochs are shown in Fig 9. The box, classification, and DFL loss graphs, which represent the losses, decline gradually, reflecting successful learning. Meanwhile, performance metrics such as precision, recall, mAP50, and mAP50–95 increase steadily, reflecting that the model is becoming better at recognizing and classifying potholes. In general, the trends reflect fine model convergence and accuracy.

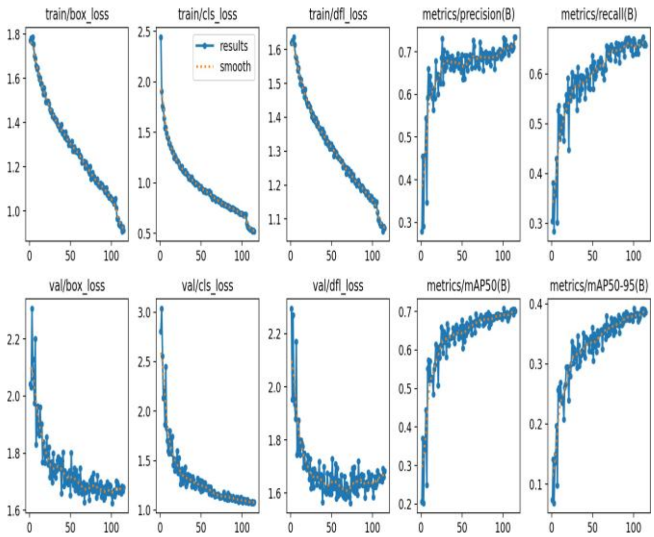

Fig. 9. Training and validation loss curves along with evaluation metrics.

4.3 Evaluation results

The performance metrics—precision, recall, mAP@0.5, and mAP@0.5:0.95—offer a summary of how well the model performs on unseen validation data.

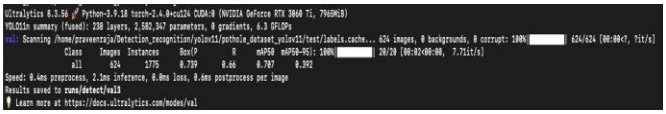

Fig. 10. YOLOv11n Model Performance Summary After Training.

With a precision of 0.739 and a recall of 0.66 at an IoU threshold of 0.5, the model demonstrated strong performance. Good localization and classification capabilities were shown by the mAP of 0.392 for IoU thresholds between 0.5 and 0.95. With an average of 0.4 ms for preprocessing, 2.1 ms for inference, and 0.6 ms for post-processing per image, the system showed effective processing speeds. Because of this, the model can be used in systems like surveillance cameras, drones, or vehicle-mounted systems that need to monitor road conditions quickly and accurately. Fig. 10 shows the YOLOv11n Model Performance Summary After Training.

5 Comparison with Yolo Models

We assess our model's pothole detection efficacy by contrasting its recall performance with earlier iterations of YOLO. One important parameter is recall, which shows how well the model can detect every real pothole.

Table 1. Recall Comparison Between Yolo Models.

Reference	Model	Recall (%)
[11]	YOLOv8-nano	60.0
[11]	YOLOv8-small	35.0
[11]	YOLOv8-medium	50.0
	YOLOv11-nano (Ours)	66.0

The proposed YOLOv11-n model achieves a higher recall of 66.0%. This improvement reflects the effectiveness of YOLOv11-n in minimizing false negatives and enhancing detection reliability, especially in real-world scenarios. Table 1 shows the Recall Comparison Between Yolo Models.

6 Conclusion and Future Work

A very efficient pothole detection system based on deep learning was implemented with the YOLOv11 model. Due to increasing road accidents caused by poor road condition, especially potholes, the suggested approach offers real- time precise implementation that can be used on various platforms from mobile phones to drones and surveillance systems. The YOLOv11-n model performed higher recall than other YOLO models including variants like YOLOv8-nano, YOLOv8-small, YOLOv8-medium, indicating its higher capacity to identify potholes with fewer false negatives. The model's stability and dependability were further validated by other assessment criteria like the precision-recall curves and confusion matrix. Incorporating this model into smart city infrastructure allows authorities to ease road monitoring, streamline repair priorities, and improve public safety.

In the future, additional enhancements could include expanding the dataset to encompass different road types, weather conditions, and geographical locations to improve generalization. Incorporating real-time video stream processing would enable continuous pothole tracking for application in autonomous vehicles and real-time traffic management. A severity evaluation mechanism could also be added to assist municipalities in prioritizing repairs by pothole depth or size. Edge device deployment for real-time low-latency inference in resource-scarce settings is also a future prospect of interest.

References

[1] Z. A. Y. Saluky, Yoni Marine *et al.*, "Pothole detection on urban roads using yolov8," 2024. Available at https://ieeexplore.ieee.org/document/ 10291192.

[2] D. Shuaiyu Fan, Hao Cai, "Lightweight road pothole detection network based on fcos," 2024. Available at https://ieeexplore.ieee.org/ document/10505391.

[3] Z. Zhou, "Traffic accident detection based on yolov11," 2024. Available at https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber= 10898397.

[4] J. Sana Saleh *et al.*, “Real-time pothole detection with edge intel- ligence and digital twin in internet of vehicles,” 2024. Available at https://ieeexplore.ieee.org/document/10798477.

[5] R. C. Baranidharan Balakrishnan *et al.*, “Comparative study on various architectures of yolo models used in object recognition,” 2022. Available at https://ieeexplore.ieee.org/document/10037635.

[6] A. M. S Balakrishna, “Progress in multi-object detection models: a comprehensive survey,” 2022. Available at https://doi.org/10.1007/ s11042-022-14131-0.

[7] M. M. K. N. Bhavana *et al.*, “Pot-yolo: Real-time road potholes detection using edge segmentation-based yolo v8 network,” 2024. Available at https://ieeexplore.ieee.org/document/10542661.

[8] J. L. Linchao Li *et al.*, “Road pothole detection based on crowdsourced data and extended mask r-cnn,” 2024. Available at https://ieeexplore. ieee.org/document/10474576.

[9] Y. D. Niannian Wang *et al.*, “Advanced yolo-deepsort-based system for drainage pipeline defects intelligent detection,” 2024. Available at https:

[10] //ieeexplore.ieee.org/document/10681414.

[11] G. D. Nirmal Kumar Rout *et al.*, “Improved pothole detection using yolov7 and esrgan,” 2023. Available at https://arxiv.org/abs/2401.08588.

[12] D. D. V. K. Kurra Kaushik Sai *et al.*, “Pothole detection using deep learning,” 2024. Available at https://ieeexplore.ieee.org/document/ 10537377.

[13] Y. B. Aparnaa *et al.*, “Convolutional neural networks-based potholes detection using thermal imaging,” 2020. Available at https://www. sciencedirect.com/science/article/pii/S1319157818312837.

[14] Dhiman and R. Klette, “Pothole detection using computer vision and learning,” 2020. Available at https://ieeexplore.ieee.org/document/ 8788687.

[15] W. K. Y. Au Yang Her *et al.*, “Real-time pothole detection system on vehicle using improved yolov5 in malaysia,” 2022. Available at https:

[16] //ieeexplore.ieee.org/document/9968423.

[17] S. D. V. Dharneeshkar J *et al.*, “Deep learning-based detection of potholes in indian roads using yolo,” 2020. Available at https:// ieeexplore.ieee.org/document/9112424.

[18] R. Khanam* and MuhammadHussain, “Yolov11: An overview of the key architectural enhancements,” 2024. Available at https://arxiv.org/ abs/2410.17725.