Incorporating ML models in a New Use-case Setting:Pothole Detection and Severity Classification using Machine Learning

ZIYAD SHAHIN, University at Buffalo, USA
JYOTHISMARIA JOSEPH, University at Buffalo, USA

Abstract

Potholes are a significant hazard to roadway safety, contributing to accidents, vehicle damage, and increased repair costs. Traditional pothole detection methods, such as manual surveys, are costly, error-prone, and time-consuming. To address these limitations, this study proposes an automated system for pothole detection and severity classification using machine learning techniques. While numerous studies have focused on detecting potholes, they often overlook severity assessment. This research leverages the YOLOv8 and YOLOv11 models for pothole detection and a custom classification pipeline for severity estimation. Initial attempts at depth estimation through fine-tuning pre-trained models were unsuccessful for smaller potholes. Consequently, we shifted to a visual-based approach, training a YOLOv8 model on a custom annotated dataset for severity classification. The proposed system combines real-time pothole detection with a two-stage pipeline for severity classification, providing authorities with an efficient and scalable solution for pothole management. Our results indicate a promising path for improving pothole detection and prioritization based on severity, ultimately enhancing road safety and minimizing maintenance costs.

ACM Reference Format:
Ziyad Shahin and Jyothismaria Joseph. 2025. Incorporating ML models in a New Use-case Setting:Pothole Detection and Severity Classification using Machine Learning. 1, 1 (April 2025), 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Potholes pose a serious threat to roadway safety and infrastructure, contributing to accidents, vehicle damage, and substantial repair costs. Pothole-related crashes can lead to severe injuries and fatalities, underscoring the urgent need for an efficient automated detection system. Timely identification and repair are critical for minimizing both risks and costs.

Traditional pothole detection methods rely on manual field surveys and visual inspections to estimate pothole size and repair costs. However, these methods are costly, error-prone, subjective, unsafe, and time-consuming. Potholes also inflict significant damage on vehicles, increasing out-of-pocket repair expenses for drivers. According to a 2022 AAA survey, pothole-related vehicle repairs rose by 57% compared to the previous year, affecting 44 million U.S. drivers at an average cost of $406 per repair often requiring multiple repairs annually.

Authors' Contact Information: Ziyad Shahin, University at Buffalo, Buffalo, USA; Jyothismaria Joseph, University at Buffalo, Buffalo, USA.

ACM XXXX-XXXX/2025/4-ART
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Numerous previous studies have attempted to automate pothole detection using various machine learning techniques. Several studies examined different versions of YOLO "You Only Look Once" model, Faster R-CNN, CNN, and SSD object detection algorithm for pothole classification. While these models perform well in detecting potholes, they often lack the capability to assess severity—an essential aspect for prioritization of higher severity potholes. To address this gap, other studies have proposed methods such as using the triangle similarity property (geometric optics) to estimate pothole dimensions, clustering techniques to identify dense pothole regions and rank them.

Initially, we experimented with fine-tuning pre-trained depth estimation models to estimate the severity of potholes. However, we encountered limitations with these models, as they struggled to generate accurate depth maps for smaller potholes, particularly those under 50 cm in size. This challenge led us to shift our focus toward image classification for severity estimation, allowing us to infer pothole severity implicitly based on visual indicators such as depth and size. By leveraging image classification, authorities can more effectively identify and prioritize potholes that require immediate attention, thereby optimizing maintenance efforts and resources.

2 Literature Review

Numerous studies have explored pothole detection using a variety of machine learning techniques. This section reviews recent research to highlight key advancements and limitations.

The research [11] introduces a transformer-based model, particularly the SegFormer framework, for pothole segmentation using high-resolution images captured from a road inspection vehicle. The model takes advantage of the high-resolution RGB images captured by inspection vehicles or unmanned ground robotic vehicles, allowing for the detection of smaller potholes that may have been missed by traditional inspection methods. The SegFormer (b5) model was employed with optimal hyperparameters from prior experiments to achieve the best performance. The results showed that the network achieved an accuracy of 0.9621, precision of 0.8307, recall of 0.7217, F1-score of 0.7724, and IoU of 0.6291. A major limitation of this paper is that the study had a small training dataset of only 240 images. To address these limitations, we collected a dataset that incorporated over 1500 images, ensuring diverse conditions and a more robust model.

The research [19] aims to develop an automated system for tracking and detecting potholes using the YOLOv8 model. The study also evaluates the Faster R-CNN and YOLOv4 models. The results indicate that YOLOv8 achieved perfect precision (1.0) at a confidence level of 0.89, meaning that all detections above this threshold were potholes with no errors. However, the study lacks precision data for confidence levels below 0.89, which is crucial for making informed threshold decisions. A limitation of this approach is that pothole severity classification is based on the percentage of road damage, which varies due to inconsistent image dimensions.

This study [18] examined different variants of YOLO, namely, YOLOv1, YOLOv2, and YOLOv3. Combining two open-source datasets summing up to 445 images, yielding mAP@50 of 0.797, 0.822, and 0.833 for YOLOv1, YOLOv2, and YOLOv3 respectively, showing that newer versions of YOLO provide better results.

The research conducted in [12] evaluated YOLOv5, YOLOv7, and YOLOv8 models for pothole detection, deployed on Edge AI devices for real-time inference. YOLOv8 achieved the highest accuracy at 78%, followed by YOLOv5 at 70%, and YOLOv7 at 67%. The comparison highlights the performance differences of these models in real-time applications.

The authors of [16] utilizes the YOLOv5 model on edge devices for detecting potholes on highways. The model was evaluated using a dataset of images captured under varying illumination, weather, and road conditions. However, the size of the dataset was not explicitly mentioned, which makes it difficult to assess the generalizability of the model.

In [17], the authors achieved a mean average precision (mAP) of 0.92 and recall of 0.89 using YOLOv8 for pothole detection. However, the manually collected dataset using dashcam may result in the model's performance decline under challenging conditions, such as poor lighting or bad weather, which could impact detection accuracy. The study found that fine-tuning the pre-trained model on the pothole dataset enhanced results compared to training from scratch. The work in [3] also utilized the YOLOv8 model; however, the performance was much lower with a 0.790 mAP@0.5.

The study [9] implemented a UAV-based road defect monitoring system utilizing the YOLOv5 object detection architecture. Their framework aimed to identify and classify road surface defects (such as cracks and potholes) using high-resolution imagery collected from UAVs.

The research conducted in [4] introduced an improved YOLOv8-GD algorithm for real-time road pothole detection, enhancing detection accuracy and speed. By replacing the YOLOv8 model's neck feature fusion network with an aggregation and distribution mechanism (GD), they improved feature fusion and model generalization. The system showed significant improvements in precision, recall, and overall performance, offering a more efficient solution than traditional road inspection methods.

Authors of [13] developed a multi-label deep learning model capable of detecting various roadway disruptions (including road defects, animals, debris, fire, fog, flood, and humans), supporting automated infrastructure monitoring using image-based data. Study [10] proposed a deep learning framework based on the ResUNet-a architecture for pixel-wise semantic segmentation of cracks and potholes. Their model integrates residual connections, atrous convolutions, pyramid scene parsing pooling, and multi-task inference, achieving superior performance compared to traditional U-Net architectures. The model was evaluated using multiple crack datasets and a high-resolution pothole dataset collected in Greece, demonstrating strong accuracy in segmenting road surface defects.

This study [1] employed the YOLOX model for pothole detection, utilizing bounding box annotations to identify potholes in images. The annotations were initially created in the Pascal VOC format and later converted to the COCO format to ensure compatibility with the YOLOX model.

All the studies [1, 3, 4, 9–13, 16–19] mentioned above have successfully detected potholes; however, an important aspect of the potholes, how severe they are was not addressed. Assessing severity is essential, as it enables the prioritization of more dangerous potholes that pose a greater threat to road users.The following section presents how severity assessment has been conducted in the literature using various approaches, ranging from simple image-based classification to advanced 3D depth estimation techniques.

The research [8] proposes an automated pothole detection and reporting system that utilizes an edge computing device to send images of detected potholes, along with GPS latitudes and longitudes, to a centralized server. The server then counts the number of potholes and prioritizes them based on severity. The methodology involves Selective Search for object detection, TV-VGG for classification, and Spatial Proximity Amalgamation for severity estimation. One key limitation of the system is that severity classification is solely based on the number of potholes in an image. However, there can be cases where a single deep pothole may pose a greater hazard than multiple smaller ones.

In the paper [15] they employed deep learning to detect potholes and for estimating the dimensions they used built-in vehicle technologies. They proposed the usage of the YOLOv5 model for pothole detection. They also highlighted the usage of data augmentation methodologies to improve the model. They achieved high precision, Recall and F1-score with the implementation of YOLOv5 However, a key limitation of this method is the cost associated with built-in vehicle technology, making it less accessible for widespread deployment.

The study [20]proposes a framework that integrates YOLOv5n-p6 for pothole detection and ZoeDepth for metric depth estimation, ensuring lightweight deployment and robust accuracy. To enhance model performance, data augmentation techniques are employed during the training phase to increase data diversity. The depth estimation process relies on key reference points within the pothole, and a pinhole camera is used to pothole the area with some particular key assumptions that facilitate the calculation of the estimation. This study does not provide exact numerical results for area estimation. Additionally there is no clear metric outlined for Zoe Depth's accuracy.

The study [14]introduces a computationally efficient and cost-effective pavement pothole detection framework. They proposed to enhance the single-stage CNN architecture and introduce a preventive co-adaptive to detect potholes in pavements which is done by modifying the ResNet50 backbone framework used in integration with the RetinaNet architecture. The RetinaNet model was trained for 250 epochs in batches of 4 over 72 hours. It also incorporates an additional method to conduct depth estimation using 3D point clouds using the photogrammetric technique of SfM by employing photogrammetric software. It had really good results in training and testing where recall is around 0.96 and F1-score is 0.98. The depth estimation mean percentage error for the observed pothole samples was 0.409. While the system is lightweight, the 3D point cloud generation process still involves computational overhead.

This study[2] provides methods to approach the problem of assessing road damage and details a number of public datasets and models that can be used to tackle it. It focuses on using publicly available datasets and dashcam imagery for automated pothole detection and size estimation. It tests the current popular choices of methods to accomplish the tasks like FasterRCNN and YOLOv5 for detecting potholes and AdaBins for monocular depth estimation. They state that detection models perform well, depth estimation remains unreliable due to scale ambiguity, and the study lacks ground truth data for accurate validation.

To address this limitation, the authors of [7] compared YOLOv7 and CNN, showing YOLO's superiority in detecting potholes accurately with fast processing. The addition of a depth estimation module, using an image-processing-based triangular similarity measure, was used for pothole dimension estimation. Similarly, in[6] the authors address the limitations of previous studies by employing the YOLOv3 and YOLOv4 models and adding a depth estimation module. An image-processing-based triangular similarity measure is used for pothole dimension estimation.

The research in [5] employs the YOLOv5 model with data captured from a D455 depth camera and LiDAR sensor to acquire 3D point cloud data. The RGB images are segmented using a self-trained YOLOv5-seg network, and corresponding virtual point clouds are generated via coordinate transformation. The 3D point clouds are then registered with the virtual point clouds to enrich the semantic information of the pothole. Finally, Euclidean clustering is used to segment the pothole point clouds, providing the necessary information. This process may be computationally expensive and slow due to the intensive tasks of semantic segmentation, coordination, point cloud registration, and clustering, especially when dealing with large datasets. Additionally, implementing such a system can be costly in terms of both hardware and software resources.

3 Methodology

The primary objective of our project is to detect potholes and classify their severity. To achieve this, various methods were implemented and evaluated to identify the most effective approach in terms of accuracy, cost, and overall efficiency.

3.1 Pothole Detection:

For the pothole detection component of the project, the proposed approach was to utilize a YOLO (You Only Look Once) model to enable real-time pothole identification. YOLOv5m was selected as our baseline model, and YOLOv11 as our proposed model.

YOLOv11 was proposed as an alternative to YOLOv5 to explore potential improvements in detection accuracy and model efficiency, leveraging advancements in model architecture and training methodology. Given that YOLOv11 represents a more recent development, this investigation aimed to assess whether its innovations could offer measurable benefits for real-time pothole detection compared to YOLOv5 the baseline model.

Model training for pothole detection involves using a labeled dataset to train the model to detect potholes in new images or videos. Both models were trained for 50 epochs, with their performance evaluated using testing and validation sets. Evaluation metrics included precision, recall, F1 score, and mean average precision.

The results of this initial analysis indicated slightly better performance for YOLOv5m. However, this marginal improvement is not sufficient to determine the preferred model. Furthermore, the size of the model (for example, YOLOv11s, YOLOv11m, YOLOv11l) can significantly affect the detection accuracy.

Therefore, YOLOv5m, YOLOv5l, YOLOv11m, and YOLOv11l were each trained for 450 epochs to allow a more comprehensive comparison. The full results of all models are presented in the results section.

3.2 Severity Classification:

Our initial plan was to fine-tune a depth estimation model, such as Midas, to generate depth maps of potholes, which could then be used to classify the severity of each pothole. The motivation behind this approach was to develop a geometry-aware classification pipeline where deeper or larger potholes could be quantitatively identified based on predicted depth values. However, while generating the dataset in Blender and testing it with the depth model, we encountered problems. The model failed to produce clear depth maps for potholes with a depth of less than 50 cm when using a perspective camera intended to mimic real-world conditions. The resulting depth maps failed to capture significant depth variation for moderate potholes. While switching to an orthographic camera provided clearer depth maps, this setup does not reflect the imaging conditions typically encountered in real-world scenarios, rendering it impractical for our intended use case.

Since our primary goal was to estimate the severity of potholes, we pivoted our approach. We created a custom annotated dataset in which potholes were labeled as either major or medium based on severity. Instead of relying on explicit depth estimation, we fine-tuned a classification model to learn visual cues related to pothole size and severity implicitly. Currently we are using YOLO-v8cls to facilitate our classification task.

3.3 Overall Pipeline:

We experimented with two distinct processing pipelines to handle the task of both detection and severity classification robustly. First, we experimented with end-to-end YOLO models, which are capable of both object detection and classification. We tested different YOLO model variants to evaluate performance trade-offs across model size and inference time.

Second, we designed a two-stage pipeline to recognize the availability of a larger dataset containing only detection-level annotations (i.e., pothole locations without severity labels). In this setup:

- Stage 1: A YOLOv11l model is used to detect potholes. The bounding boxes from this model are used to crop individual pothole regions from the original images.
- Stage 2: These cropped regions are fed into a separate YOLOv8 classification model explicitly trained on the severity-labeled dataset.

This modular pipeline architecture gives us flexibility: We can utilize the full detection dataset for locating potholes while leveraging the smaller, custom annotated dataset for learning severity classification.

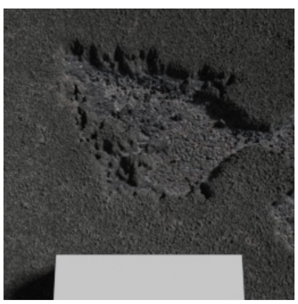
4 Data Sets

4.1 Pothole Detection Dataset

For the pothole detection part of the problem, we will be combining Roboflow and Kaggle Pothole Detection Dataset and Pothole Detection for YOLOv4 Dataset. The first dataset has 665 images, and the YOLO annotation for each image is in an XML file. The second dataset has 1,562 training images and 421 testing images, and each image has YOLO annotation in a text file. We merged these datasets together in Roboflow and used it as a database for training and testing. After removing the duplicated images, the final dataset that we used to train YOLOv11 has 1,982 images with a total of 5,021 potholes altogether. The dataset was then split into 70% for the training set, 20% for the validation set, and 10% for the test set. This combined dataset captures diverse weather conditions, lighting variations, and camera angles, enhancing the model's robustness.

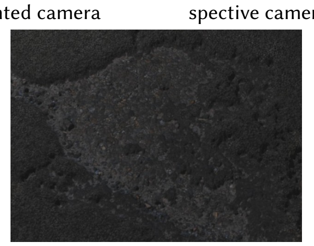
4.2 Severity Classification Dataset

4.2.1 Blender Synthetic Data. As the initial approach was to fine-tune a depth estimation model data with the ground truth depth was needed to test and validate the model. Therefore, blender was used to generate potholes with varying sizes. Examples of the dataset is presented in figures.1a,1b, and 1c



(a) Blender Synthetic Pothole mimicing car mounted camera

(b) Severe Synthetic Pothole perspective camera



(c) Synthetic Pothole perspective camera

Fig. 1. Examples of pothole Synthetic Data

4.2.2 Classification Dataset. A custom annotated dataset of pothole images was developed for training the classification model. Each image was labeled based on the perceived severity of the pothole, categorized into two classes: major and medium.

- Major potholes refer to those that are either very deep, or moderately deep but large in area, representing a high level of severity.
- Medium potholes include those that are moderately deep but small in area, or shallow but large in area, indicating a comparatively lower severity.

This dataset was developed by aggregating images from various existing pothole datasets in which potholes were originally annotated without severity labels. Examples from the dataset are presented in figures 2a, and 2b

(a) Major Pothole

(b) Medium Potholes

Fig. 2. Examples of Severity Annotated Potholes

5 Results

The evaluation used a standard test split alongside a custom-designed test subset to assess the model's robustness and generalization. The standard test set included images randomly selected from the overall dataset. Additionally, we created a secondary test set featuring images that reflect various real-world conditions, such as potholes photographed on clear days and water-filled potholes. This approach helped us conduct a comprehensive analysis of the model's performance beyond the usual comprehension.

We used a combination of object detection and classification metrics to do quantitative analysis. We computed mean Average Precision at IoU threshold 0.5 (mAP@50) for object detection to evaluate how accurately the model identified and localized potholes. We also used precision, recall, and the F1-score to measure the correctness of predictions per class. We also analyzed the confusion matrix to identify trends in misclassification and understand where the model struggles, especially in distinguishing between visually similar classes such as medium and minor potholes.

5.1 Object Detection Performance:

To evaluate the detection part of the project, we experimented with multiple model variants across 2 proposed procedures:

- Severity-classified detection: where models learned to distinguish between major, medium potholes.
- Binary detection: where the model was trained on a much larger dataset with pothole vs background only (no severity differentiation).

In the severity-classified setup, we tested four models: YOLOv5m, YOLOv5l, YOLOv11m, and YOLOv11l. Among these, YOLOv11m outperforms all others, achieving the highest precision (0.737) and mAP@50 (0.670) while maintaining a moderate inference time of 23.9ms per image. In the severity-classified setup, we tested four

Model	Precision	Recall	mAP@50	Inference Time
YOLOv5m	0.725	0.695	0.655	18.2 ms
YOLOv5l	0.714	0.711	0.660	33.6 ms
YOLOv11m	0.737	0.697	0.670	23.9 ms
YOLOv11l	0.730	0.680	0.664	31.1 ms
YOLOv11l (Large dataset w/o severity)	0.941	0.8942	0.972	31.6 ms

Table 1. Detection model comparison under severity-aware and binary-only settings.

models: YOLOv5m, YOLOv5l, YOLOv11m, and YOLOv11l. Among these, YOLOv11m outperforms all others, achieving the highest precision (0.737) and mAP@50 (0.670) while maintaining a moderate inference time of 23.9ms per image.

In contrast, we also trained YOLOv11l on a significantly larger dataset that did not include severity labels. This model was evaluated on a binary detection task — classifying whether a pothole is present. It achieved high precision (0.941) and recall (0.8942), demonstrating the model's ability to generalize when the task is simplified. These results suggest that removing the severity classification task allows the model to focus purely on object localization, resulting in better overall accuracy.

5.2 Severity Classification Performance:

To evaluate how well each model distinguished between pothole severity levels, we computed class-wise precision and recall for the two labeled classes used during training: major and medium potholes. These metrics were calculated using the output of the final prediction by each pipeline configuration, enabling us to assess both one-stage detection and classification models and a pipeline two-step setup.

Model	Class	Precision	Recall
YOLOv5m	major	0.867	0.800
	medium	0.584	0.591
YOLOv5l	major	0.854	0.802
	medium	0.575	0.62
YOLOv11m	major	0.879	0.800
	medium	0.575	0.620
YOLOv11l	major	0.879	0.800
	medium	0.600	0.594
YOLOv11l (detection) + YOLOv8 (classification)	major	0.9048	0.7125
	medium	0.7273	0.780

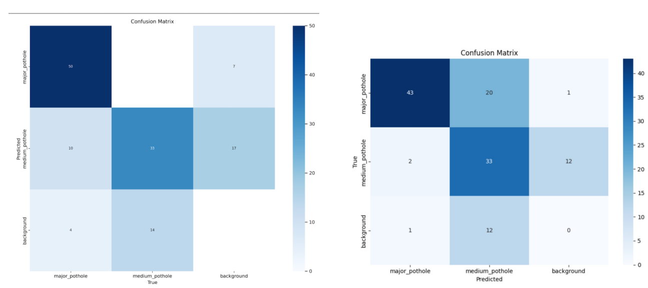
Table 2. Class-wise precision and recall for major and medium pothole severity using different model architectures and pipeline approaches.

Across single-step YOLO-based models, YOLOv11m and YOLOv11l consistently outperformed earlier YOLOv5 variants in identifying major potholes. For instance, YOLOv11m achieved a precision of 0.879 and recall of 0.800 for the major class, compared to 0.867/0.800 with YOLOv5m and 0.854/0.802 with YOLOv5l. However, all models that were detecting and classifying using the same model in a single step showed limited performance for the medium class, with precision ranging from 0.575 to 0.600 and recall from 0.591 to 0.620, indicating a higher likelihood to confuse medium potholes with other classes or backgrounds. This is likely due to the smaller dataset on which it was trained.

To address this challenge, we implemented a two-step pipeline combining YOLOv11l for initial pothole detection with a separate YOLOv8 image classifier for severity classification. This pipeline approach led to a noticeable improvement in severity differentiation. The precision and recall for major potholes increased to 0.9048 and 0.7125, respectively, while for medium potholes, both metrics improved significantly to 0.7273 and 0.780. Allowing this two-step separate pipeline helped us utilize the annotated large dataset to detect potholes and get much better object detection. Then, a secondary model was trained on the same data as a single-step YOLO-based model to classify the potholes into severity.

5.3 Performance Under Challenging Conditions:

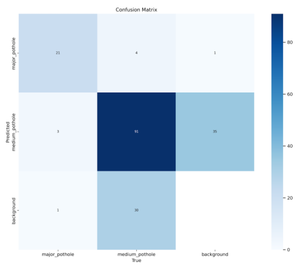
To assess the robustness of our models in real-world environments, we evaluated the two best-performing systems: the YOLOv11m model trained for severity-based detection and the pipeline approach combining YOLOv11l for detection with a YOLOv8 classifier for severity prediction. These models were tested on two curated subsets of the test data: one consisting of potholes captured on clear, sunny days and another featuring water-filled potholes typically due to rainfall. These subsets were chosen to assess the models on environments they can encounter in real life.



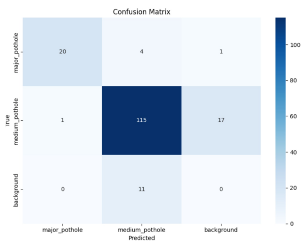
(a) YOLOv11m (rainy) (b) Pipeline (rainy)

Fig. 3. Confusion matrices comparing YOLOv11m and the pipeline method under rainy conditions.

When comparing the results across both conditions, the pipeline method demonstrated more stable performance and better class separation than YOLOv11m. Under clear conditions, the pipeline achieved strong severity differentiation, correctly identifying 115 medium potholes with minimal confusion, while YOLOv11m struggled slightly more with medium-to-background misclassifications. The performance gap widened in the water-filled subset: the pipeline method maintained relatively consistent predictions. However, it performed relatively poorly in classifying major potholes, whereas YOLOv11m misclassified a significant number of medium potholes as background and showed reduced reliability. This comparison



(a) YOLOv11m (clear day)

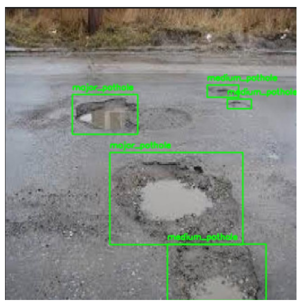


(b) Pipeline (clear day)

Fig. 4. Confusion matrices comparing YOLOv11m and the pipeline method under clear conditions

suggests that the two-stage pipeline architecture offers greater robustness, especially in visually ambiguous scenarios.

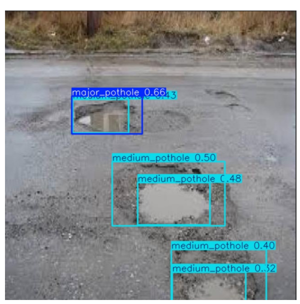
(a) Test image 1



(b) Test image 2

Fig. 5. Testing performance of the proposed pipeline

(a) Test image 1



(b) Test image 2

Fig. 6. Testing performance of YOLOv11m

Figures. 5, and 6 shows the performance on testing dataset of the proposed pipeline, and YOLOv11m, respectively. Notably, figures 5b, and 6b highlight the superiority of the proposed pipeline. This superiority is expected as this approach leverages a YOLOv11l trained on a large dataset acheiving high precision (0.941) to detect potholes. Detected potholes are then individually cropped from the original images and passed into a YOLOv8 classification model, which is specifically trained on a dedicated severity dataset—allowing for more accurate severity assessment.

6 Conclusion and Future Works

Following a comprehensive evaluation of multiple object detection architectures and severity classification strategies, we selected the pipeline-based approach as the final model configuration. This design, which integrates a YOLOv11l detector with a separate YOLOv8-based classifier, demonstrated a more robust performance in both standard and challenging visual conditions. This structure allowed the detection stage to benefit from training on a significantly larger dataset without the constraints of severity annotations, thereby improving generalizability. Simultaneously, the dedicated classification stage provided more accurate severity estimation, focusing exclusively on cropped pothole regions. This approach also offered improved scalability and adaptability for real-world deployment with high precision and recall of how potholes are differentiated from the background.

In future work, we would like to explore alternative classification models, expand the training dataset, and incorporate more detailed annotations to further improve performance in classification. Overall, the pipeline configuration presents a more reliable and effective solution for pothole detection and severity assessment in diverse operational environments.

7 Contributions

- Jyothismaria Joseph
  - Report Writing
  - Model Training (YOLO)
  - Synthetic Data creation in Blender
  - Literature Review
- Ziyad Shahin
  - Report Writing
  - Model Training (YOLO)
  - MiDas Calibration Experimentation
  - Literature Review

References

[1] M.P. B and S. K.C. 2022. Enhanced pothole detection system using YOLOX algorithm. *Auton. Intell. Syst.* 2, 22 (2022). doi:10.1007/s43684-022-00037-z

[2] Adithya Badidey, Ryan Dalby, Zhongyi Jiang, David Sacharny, and Thomas C. Henderson. 2022. Monocular Road Damage Size Estimation using Publicly Available Datasets and Dashcam Imagery. In *2022 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. 1–8. doi:10.1109/MFI55806.2022.9913878

[3] Rounak Kumar Bala, Ankit Bhattacharyya, Ankan Roy, Adrija Roychowdhury, and Debraj Chatterjee. 2023. Pothole Detection using YOLO v8 Algorithm. In *2023 9th International Conference on Smart Computing and Communications (ICSCC)*. 1–5. doi:10.1109/ICSCC59169.2023.10335077

[4] Jie Chen, Qianli Ma, Yunxiang Fei, Jun Yin, and Xinzhe Chen. 2025. Design and Implementation of real-time road surface detection System Based on YOLOv8-GD Algorithm. In *Proceedings of the 2024 8th International Conference on Electronic Information Technology and Computer Engineering (EITCE '24)*. Association for Computing Machinery, New York, NY, USA, 529–533. doi:10.1145/3711129.3711220

[5] Xiaokang Chen, Nan Ma, and Mohan Wang. 2023. YOLOv5-Pothole: An Improved Pothole Perception Method Based on YOLOv5-Seg. In *2023 2nd International Conference on Artificial Intelligence, Human-Computer Interaction and Robotics (AIHCIR)*. 75–79. doi:10.1109/AIHCIR61661.2023.00018

[6] Pranjal A. Chitale, Kaustubh Y. Kekre, Hrishikesh R. Shenai, Ruhina Karani, and Jay P. Gala. 2020. Pothole Detection and Dimension Estimation System using Deep Learning (YOLO) and Image Processing. In *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*. 1–6. doi:10.1109/IVCNZ51579.2020.9290547

[7] Saranya E, Nivetha R, Abirami S, Mohaideen Arsath M, and Dharaneesh S. 2024. Revolutionizing Road Maintenance: YOLO Based Pothole Detection System. In

2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS), Vol. 1. 1991–1997. doi:10.1109/ICACCS60874.2024.10716937

[8] Zachary Jeffreys, Kshama Kumar, Zhuojing Xie, Wan D Bae, Shayma Alkobaisi, and Sada Narayanappa. 2024. PotholeVision: An Automated Pothole Detection and Reporting System using Computer Vision. In *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing* (Avila, Spain) *(SAC '24)*. Association for Computing Machinery, New York, NY, USA, 695–697. doi:10.1145/3605098.3636118

[9] Iason Katsamenis, Nikolaos Bakalos, Eftychios Protopapadakis, Eleni Eirini Karolou, Georgios Kopsiaftis, and Athanasios Voulodimos. 2023. Real time road defect monitoring from UAV visual data sources. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Greece) *(PETRA '23)*. Association for Computing Machinery, New York, NY, USA, 603–609. doi:10.1145/3594806.3596561

[10] Iason Katsamenis, George Kopsiaftis, Athanasios Voulodimos, Ioannis Rallis, Ioannis Georgoulas, Charalampos Zafeiropoulos, and Anastasios Doulamis. 2024. A Deep Learning Framework for Segmentation of Road Defects Using ResUNet-a. In *Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments* (Crete, Greece) *(PETRA '24)*. Association for Computing Machinery, New York, NY, USA, 449–455. doi:10.1145/3652037.3663935

[11] Iason Katsamenis, Athanasios Sakelliou, Nikolaos Bakalos, Eftychios Protopapadakis, Christos Klaridopoulos, Nikolaos Frangakis, Matthaios Bimpas, and Dimitris Kalogeras. 2023. Deep transformer networks for precise pothole segmentation tasks. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Greece) *(PETRA '23)*. Association for Computing Machinery, New York, NY, USA, 596–602. doi:10.1145/3594806.3596560

[12] Aniruddhsinh Parmar, Ruchi Gaiiar, and Nagendra Gajjar. 2023. Drone based Potholes detection using Machine Learning on various Edge AI devices in Real-Time. In *2023 IEEE International Symposium on Smart Electronic Systems (iSES)*. 22–26. doi:10.1109/iSES58672.2023.00016

[13] Eftychios Protopapadakis, Iason Katsamenis, and Anastasios Doulamis. 2020. Multi-label deep learning models for continuous monitoring of road infrastructures. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Greece) *(PETRA '20)*. Association for Computing Machinery, New York, NY, USA, Article 66, 7 pages. doi:10.1145/3389189.3397997

[14] Eshta Ranyal, Ayan Sadhu, and Kamal Jain. 2023. AI assisted pothole detection and depth estimation. In *2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS)*, Vol. 1. 1–4. doi:10.1109/MIGARS57353.2023.10064547

[15] Cuthbert Ruseruka, Judith Mwakalonge, Gurcan Comert, Saidi Siuhi, Frank Ngeni, and Quincy Anderson. 2024. Augmenting roadway safety with machine learning and deep learning: Pothole detection and dimension estimation using in-vehicle technologies. *Machine Learning with Applications* 16 (2024), 100547. doi:10.1016/j.mlwa.2024.100547

[16] Bhavan Kumar S B, Guhan S, Manyam Kishore, Santhosh R, and Alfred Daniel J. 2023. Real-time Pothole Detection using YOLOv5 Algorithm: A Feasible Approach for Intelligent Transportation Systems. In *2023 Second International Conference on Electronics and Renewable Systems (ICEARS)*. 1678–1683. doi:10.1109/ICEARS56392.2023.10085336

[17] Saluky, Yoni Marine, Ahmad Zaeni, Ari Yuliati, Onwardono Rit Riyanto, and Nurul Bahiyah. 2023. Pothole Detection on Urban Roads Using YOLOv8. In *2023 10th International Conference on ICT for Smart Society (ICISS)*. 1–6. doi:10.1109/ICISS59129.2023.10291192

[18] R. Sathya, B. Saleena, and B. Prakash. 2023. Pothole Detection Using YOLOv3 Model. In *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. 1–7. doi:10.1109/SCEECS57921.2023.10063116

[19] R. Swathika, N. Radha, Parthiban D, and Sriram J. 2024. Pothole Detection and Road Damage Assessment System Utilizing Image Processing Techniques. In *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*. 269–278. doi:10.1109/ICICI62254.2024.00052

[20] Dehao Wang, Yiwen Xu, Haohang Zhu, and Kaiqi Liu. 2024. A Novel Framework for Pothole Area Estimation Based on Object Detection and Monocular Metric Depth Estimation. In *2024 IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*. 1–6. doi:10.1109/ICSIDP62679.2024.10869122