

CAPITAL UNIVERSITY OF SCIENCE AND
TECHNOLOGY, ISLAMABAD



AI-Driven Pothole Detection with GPS Localization and Severity Assessment

by

Umar Zakria

A thesis submitted in partial fulfillment for the
degree of Master of Science

in the

Faculty of Engineering

Department of Civil Engineering

2026

Copyright © 2026 by Umar Zakria

All rights reserved. No portion of the material protected by this copyright notice may be replicated or utilized in any arrangement or by any means, electronic or mechanical including photocopy, recording or by any information storage and retrieval system without authorization from the author(s).



CERTIFICATE OF APPROVAL

AI-Driven Pothole Detection with GPS Localization and Severity Assessment

by

Umar Zakria

Registration No: (MCE243018)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Syed Bilal Ahmed Zaidi	UET, Taxila
(b)	Internal Examiner	Dr. Ishtiaq Hassan	CUST, Islamabad

Dr. M. Usman Farooqi

Thesis Supervisor

September, 2026

Dr. Majid Ali

Head

Dept. of Civil Engineering

September, 2026

Dr. Imtiaz Ahmad Taj

Dean

Faculty of Engineering

September, 2026

Author's Declaration

I, **Umar Zakria**, hereby state that my MS thesis titled “**AI-Driven Pothole Detection with GPS Localization and Severity Assessment**” is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.



(Umar Zakria)

Registration No: (MCE243018)

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled “**AI-Driven Pothole Detection with GPS Localization and Severity Assessment**” is exclusively my research work with no remarkable contribution from any other individual. Small contribution/help wherever taken has been acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the Higher Education Commission and CUST towards plagiarism. Therefore, I as an author of the above titled thesis declare that no part of my thesis has been plagiarized and any material used as reference is properly cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.



(Umar Zakria)

Registration No: (MCE243018)

Acknowledgement

First and above all, I praise **ALLAH**, the Almighty, for providing me the opportunity and granting me the capability to proceed successfully. In addition, to **Muhammad (PBUH)**, the Divine servant Leader, who has changed my life.

I would like to express my sincere gratitude to **Engr. Prof. Dr. Majid Ali** for his guidance, and supervision throughout this research. His valuable advice and constructive suggestions helped me at the initial stages of this project and provided me the right direction. I sincerely appreciate the time, effort, and knowledge he shared during the course of this study.

I am also thankful to my supervisor, **Engr. Dr. M. Usman Farooqi**, for his guidance, support, and valuable suggestions throughout this research. I would like to thank the Department of Civil Engineering, Capital University of Science and Technology, for providing the library, laboratory facilities, and academic environment that made this research possible.

Finally, I would like to thank my family for their constant encouragement, patience, and support throughout my studies. Their motivation has always been a source of strength and has helped me complete this thesis.

Umar Zakria

Abstract

With the expansion of road networks and aging infrastructure across the globe, timely detection of pavement defects has become increasingly crucial to ensure road safety, serviceability and sustainable asset management. Traditional pavement condition surveys are commonly carried out by manual visual inspection, which are labour intensive, time consuming, subjective and hard to implement over large road networks. In recent years, computer vision and artificial intelligence technologies have been developed to automatically detect pavement distresses. However, challenges exist for reliable pothole segmentation, cost reduction for deployment, and integration of detection results into practical pavement maintenance workflows.

This research proposes an unique automated pothole assessment framework using deep learning instance segmentation architectures and smartphone for reporting road condition results. The automated report generation of survey will aid authorities to take timely and informed decision about the maintenance of pavements. Manual measurements are compared with automated results verify the accuracy of proposed framework. Final application will provide prompt results on the camera screen with pothole severity, GPS localization, PCI assessment and map visualization. A custom pothole dataset was prepared using local road conditions, and manually annotated using Roboflow tool. Pothole dataset was trained, validated and tested using two instance segmentation architectures, YOLOv8s-seg and Mask R-CNN. The trained models were evaluated on standard segmentation metrics including precision, recall, and mean Average Precision (mAP) and computational performance metrics including inference speed, model complexity and resource requirements.

Experimental results showed that the overall performance of the detection and computational efficiency of YOLOv8s-seg were better than Mask R-CNN. YOLOv8s-seg obtained a mAP50 of 72.3% and recall of 68.2%. Moreover, YOLOv8s-seg was 7.5 times faster than Mask CNN in inference speed and required less computational resources, thus making it more applicable for real-time pavement

monitoring applications. Mask R-CNN provided smoother segmentation on large and small sized potholes however, it was computationally expensive to deploy practically.

The best performing YOLOv8s-seg model was then implemented in a field deployment framework, which consisted of smartphone video acquisition, GPS synchronization, pothole tracking, severity classification, and Pavement Condition Index (PCI) application. Segmentation pixel masks were converted to a square meter using a conversion factor to estimate the area of potholes. Results obtained through automated pipeline had a difference of 11.2% when compared to manual measurement. The location of each pothole detected in the recorded video was estimated using GPS interpolation with synchronized video timestamps. The automated deduct value (DV), corrected deduct value (CDV) and PCI calculations were performed in accordance with the ASTM D6433 procedures. Field validation indicated the feasibility of using a low-cost smartphone platform for automated pothole detection, geospatial reporting and pavement condition evaluation. The proposed framework offers a practical and scalable solution to support pavement maintenance planning and road asset management.

Keywords: Instance Segmentation, YOLO (You Only Look Once), Pavement defects, Pothole detection, Area estimate, PCI

Contents

Author's Declaration	iii
Plagiarism Undertaking	iv
Acknowledgement	v
Abstract	vi
List of Figures	xi
List of Tables	xiii
List of Abbreviations	xiv
1 Introduction	1
1.1 Background	1
1.2 Research Motivation and Problem Statement	3
1.2.1 Research Questions	4
1.3 Overall Goal and Specific Aim of this MS Study	4
1.4 Scope of Work and Study Limitations	4
1.4.1 Rationale Behind the Variable Selection	5
1.5 Brief Methodology	6
1.6 Research Impact on Industry	7
1.6.1 Research Novelty - Uniqueness	7
1.6.2 Research Significance and Benefit	7
1.6.3 Practical Implementation	8
1.6.4 National and Global Impact with Emphasis on SGDs Relevance	8
1.6.5 Research Challenges	8
1.6.6 Ethical and Management Considerations Including Risk Management	9
1.6.7 Research Deliverable, Sale and Marketing Potential	9
1.7 Thesis Layout	10

2	Literature Review	12
2.1	Background	12
2.2	Pavements	13
2.2.1	Distresses in Flexible Pavement	14
2.2.2	Potholes in Flexible Pavements	15
2.3	Distress Detection Techniques	16
2.3.1	Semi-Automated Pavement Assessment Techniques	17
2.3.1.1	Conventional Two-dimensional Image Processing	17
2.3.1.2	Three-dimensional Imaging and Sensors Based Methods	18
2.3.1.3	Unmanned Aerial Vehicle Based Inspection	20
2.3.2	Fully Automated AI-Based Assessment Techniques	22
2.3.2.1	Conventional Machine Learning Models	22
2.3.2.2	Deep Learning Object Detection Models	24
2.3.2.3	Pixel Segmentation Approaches for Pothole Detection	27
2.4	AI Model Evaluation and Application	30
2.4.1	Performance Review of YOLO models	30
2.4.2	Pavement Datasets and Limitations	34
2.4.3	Automated Pavement Condition Rating Frameworks	37
2.5	Summary	39
3	Methodology	41
3.1	Background	41
3.2	Study Area	42
3.3	Methodology	43
3.4	Dataset Preparation	45
3.4.1	Image Collection	45
3.4.2	Data Annotation and Image Pre-processing	47
3.5	Instance Segmentation Architectures and Training Strategy	49
3.5.1	Single-Stage YOLOv8 Instance Segmentation	49
3.5.2	Two-Stage Mask R-CNN Using Detectron2	51
3.6	Field Application Framework	52
3.7	Summary	54
4	Results and Analysis	55
4.1	Background	55
4.2	Evaluation Metrics	56
4.3	Instance Segmentation Model Results	57
4.3.1	Single-Stage YOLOv8 Instance Segmentation	57
4.3.2	Two-Stage Mask R-CNN Using Detectron2	61
4.4	Comparative Analysis of YOLO vs Mask R-CNN	64
4.4.1	Segmentation Performance	64
4.4.2	Computational Efficiency	67

4.5	Discussion	67
4.6	Summary	70
5	Practical Implementation of Automated Pothole Detection with PCI Evaluation	71
5.1	Background	71
5.2	Developed Framework	72
5.2.1	Pothole detection and GPS Integration	73
5.2.2	Pothole Area Estimate	74
5.2.3	Deduct Value	77
5.2.4	PCI Calculation	79
5.3	Field Assessment and Validation	80
5.4	Summary	87
6	Conclusion and Future Work	88
6.1	Conclusion	88
6.2	Future Recommendations	89
	Bibliography	91

List of Figures

2.1	Gradient-histogram analysis algorithm (a) Original Image (b) Pavement crack (c) Segmented image [23] (d) OTSU processing results [24] (e) Initial binary image (f) Hole filled image [25] (g) Median denoising of crack image (h) Diffused denoising [26]	19
2.2	PaveVision3D Ultra System equipment with 2D and 3D camera sensors [33]	20
2.3	Application of 3D and sensor-based systems (a) LRIS dual-camera lane wide coverage [34]. (b) Green line presents transverse profile of pavement surface [33]	21
2.4	Volumetric representation of pothole in Bentley Software [42]	22
2.5	Multicopter drone with camera sensor for data collection [43]	22
2.6	Traditional ML models used for pavement defect identification (a) Hu et. al [34] five output node model. (b) Hoang et. al [48] vector model	23
2.7	Typical network architecture of a YOLO-based object detector [57]	25
2.8	Illustration of Faster R-CNN [58]	27
2.9	Pothole detection approaches [70]	28
2.10	Outputs of the YOLO series model [89]	32
2.11	YOLO 8v-seg bounding box and pixel-level masking of pothole [90]	33
2.12	Performance comparison of YOLO models [91]	33
2.13	Mobile-Based Data Collection Setup for Pavement Distress Detection (a) Dash-mounted mobile device [99]. (b) Smartphone mounted on front windscreen [100]	35
2.14	Statistical overview of the RDD 2022 dataset [101] (a) Distribution of pavement distress classes. (b) Geographic distribution of data collection sites.	35
2.15	Rating scale of Pavement Condition Index (PCI) based on ASTM D6433-18 [110]	39
3.1	Data collection sites, Islamabad-Rawalpindi, Pakistan.	42
3.2	Methodological diagram for Pothole detection and PCI reporting . .	44
3.3	Mobile-based system for data collection (a) DVR camera mounted on car windshield. (b) Gooseneck Mobile holder assembly mounted on motorcycle handlebar.	46
3.4	Potholes images captured using different data acquisition platforms (a) Windshield view. (b) Handlebar view. (c) Dash view.	47
3.5	Instance segmentation architecture of YOLOv8-seg [111]	50

3.6	Instance segmentation of pothole using Mask R-CNN architecture [112]	51
4.1	YOLOv8s-seg a) Training and b) Validation Loss Curves	58
4.2	YOLOv8s-seg Training Performance Curve for mAP0.5 and mAP0.5-0.95	58
4.3	Precision-Recall curve	59
4.4	F1-confidence Curve	60
4.5	Mask RCNN training loss convergence profile	62
4.6	Mask R-CNN Segmentation AP@0.5 and AP@0.75 Convergence Curves	62
4.7	Mask R-CNN Internal mask head accuracy curve	63
4.8	Segmentation results of small to large sized potholes	64
4.9	Over detection by Mask RCNN	66
5.1	GPS Synchronization and Object Tracking for Pothole Detection	75
5.2	Camera FOV and Ground Coverage	76
5.3	Pixel polygon to Diameter conversion	77
5.4	Extraction of ASTM D6433 [110] DV coordinate points using Web-PlotDigitizer	79
5.5	ASTM D6433 [110] CDV curve implementation using numpy interpolation	80
5.6	Illustration of mobile application for Pothole detection	84
5.7	Flow diagram of end to end framework for mobile application	84
5.8	Survey Analysis report displayed on Mobile	85
5.9	Map view of detected potholes and PCI reporting	86

List of Tables

2.1	Comparison of DL and ML models on different aspects covered in literature review	24
2.2	Performance analysis of Object Detection models against various datasets	26
2.3	Comparative Analysis of Semantic and Instance Segmentation Architectures for Pavement Distress Detection	29
2.4	Inference speed and model size of YOLO seg models by Ultralytics [80]	34
2.5	Summary of Publicly Available Datasets for Pavement Distress Detection	36
3.1	Image distribution and pre-processing details	48
3.2	Comparison of training hyperparameters for Mask R-CNN and YOLOv8-Seg model	53
4.1	Size Based Segmentation Performance Comparison at mAP 0.95 . .	65
4.2	Comparison of segmentation performance on the validation dataset	67
4.3	Computational Comparison of YOLOv8s-seg and Mask R-CNN . .	68
4.4	Comparison with previous studies	68
4.5	Characteristics of YOLOv8s-seg and Mask R-CNN for Pothole Instance Segmentation	69
5.1	Severity Classification of Potholes using ASTM D6433-18 [110] . . .	78
5.2	Diameter equivalent (D_{eq}) vs. Actual (D)	82
5.3	Percentage error Automated PCI vs. Manual	83

List of Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
AASHTO	American Association of State Highway and Transportation Officials
ACM	Automatic Crack Monitor
AI	Artificial Intelligence
ANN	Artificial Neural Network
AP	Average Precision
ASTM	American Society for Testing and Materials
AUC	Area Under the Curve
CDV	Corrected Deduct Value
CNN	Convolutional Neural Network
COCO	Common Objects in Context
CRDDC	Crowd-based Road Damage Detection Challenge
CSV	Comma-Separated Values
DHDV	Digital Highway Data Vehicle
DL	Deep Learning
DV	Deduct Value
FN	False Negative
FOV	Field of View
FP	False Positive

FPN	Feature Pyramid Network
FPS	Frames Per Second
GPS	Global Positioning System
GPU	Graphics Processing Unit
HTC	Hybrid Task Cascade
Hz	Hertz
ID	Identification
IoU	Intersection over Union
IPM	Inverse Perspective Mapping
LiDAR	Light Detection and Ranging
LRIS	Laser Road Imaging System
LS-SVM	Least Squares Support Vector Machine
mAP	Mean Average Precision
ML	Machine Learning
PASER	Pavement Surface Evaluation and Rating
PCI	Pavement Condition Index
PMS	Pavement Management System
PSCI	Pavement Surface Condition Index
R-CNN	Region-based Convolutional Neural Network
RGB	Red Green Blue
RF	Random Forest
ROI	Region of Interest
RPN	Region Proposal Network
SSD	Single Shot Detector
SfM-MVS	Structure from MotionMulti View Stereo
SVM	Support Vector Machine
TDV	Total Deduct Value
TP	True Positive

UAV Unmanned Aerial Vehicle

YOLO You Only Look Once

Chapter 1

Introduction

1.1 Background

Roads are an essential mode for transport in a country in order to move people, goods and services, and is vital to the economic and social development of a country. The road network in Pakistan is about 264,175 km with nearly 70% of road network being paved and the National Highway Authority (NHA) controlling about 4.6-5% of the total road network. Efficient road infrastructure enhances regional connectivity, facilitates trade, and offers access to education, healthcare and jobs services [1]. World health Organization (WHO) has designated road safety a public issue. Road pavements continuously deteriorate due to traffic loading, environmental exposure, material aging, and inadequate maintenance. Pavement defects such as cracks, rutting, and potholes progressively reduce pavement performance and serviceability. Among these defects, potholes are considered one of the most common and hazardous forms of pavement distress. Potholes can increase vehicle operating costs, disrupt traffic flow, and create safety risks for road users. According to the Road Traffic Injury Research and Prevention Centre (RTIR&PC), more than 50% of road injuries are directly or indirectly associated with pavement defects [2]. Furthermore, pothole formation may increase traffic congestion by up to 15% during peak hours, while substantial rehabilitation expenditures are required annually to maintain deteriorated road infrastructure.

The accumulation of stagnant water within potholes can potentially contributing to the spread of diseases such as dengue fever and malaria [3].

To maintain pavement condition, transportation agencies traditionally rely on visual inspection methods performed by trained personnel using manual surveys and standardized reporting procedures. Although these methods provide direct observations of pavement condition, they are labor intensive, time consuming, subjective, and often require significant resources for large road networks [4]. Over the past two decades, pavement inspection methods have evolved from manual surveys to utilizing sensor based and computer vision technologies. However, these approaches often face challenges in accurately distinguishing different distress types and quantifying pavement damages [5]. With the advent of Artificial Intelligence (AI) and Deep Learning technologies, the automated pavement inspection systems have been rapidly emerged in the industry. Convolutional Neural Networks (CNNs) have also shown good results in pavement defect detection and classification in different road and lighting conditions [6]. Current studies use object detection models which detect potholes with rectangular bounding boxes. These techniques are useful for efficient localization, but do not adequately capture the true geometry and size of pothole damage.

While deep learning has demonstrated significant results for automated pothole detection, there are several gaps highlighted in this study. There is a small number of studies that have focused on pothole instance segmentation instead they are more focused on the traditional object detection methods [7]. In addition, the comparison of single-stage and two-stage instance segmentation methods for pothole assessment is still relatively limited. There are also limited studies available that have the capability to estimate the dimensions of the pothole from segmentation masks and incorporate geospatial data for automated reporting. Moreover, there is no publicly available pothole dataset, which represents the conditions of roads in Pakistan, making it difficult to create and test trained models under local conditions.

To address these gaps, this study develops a custom pothole dataset representing local road conditions and evaluates two instance segmentation architectures,

namely YOLOv8s-seg and Mask R-CNN, for automated pothole detection. The study also explores the severity classification of potholes based on segmentation masks using pixel to area conversion and incorporates GPS location mapping for geospatial reporting. Finally, an automated framework is developed to support the assessment of Pothole Severity, Deduct Value (DV), Corrected Deduct Value (CDV) and Pavement Condition Index (PCI) for pavement condition monitoring.

1.2 Research Motivation and Problem Statement

Traditional and manual techniques are widely used in road surveys in developing countries like Pakistan. The automated pavement distress detection approach is more efficient in terms of cost and time, because it helps minimize the resources used for surveys, manual documentation, traffic closures, and maintaining pavement condition database. The digitized assessment of road assets will assist engineers and consultants in making timely decisions regarding the rehabilitation of road networks. For government agencies such as the National Highway Authority, it will be much easier to geographically identify and assess the condition of road assets, whereas the current process is extremely complicated and time consuming. Hence, this research will not only assist engineers in pavement condition assessment but also reduce the burden on the national budget associated with maintaining large road and highway networks by enabling informed decision regarding maintenance, rehabilitation, and resource allocation.

“Many people risk injury or worse because of damaged roads. To keep everyone safe, it is important to spot and fix these problems as soon as possible. There is a real need for simple and reliable ways to check the condition of our roads, so that repairs can happen quickly and resources are used wisely, without needing special or expensive equipment. By harnessing intelligent image analysis, the goal is to automate reporting and maintenance processes, enhance safety, lower repair costs, and promote a more sustainable transportation infrastructure all while keeping resource requirements to a minimum.”

1.2.1 Research Questions

- i. Which instance segmentation architecture, single-stage (YOLO) or two-stage (Mask R-CNN), is most suitable for accurate and efficient pothole detection?
- ii. How to use the segmentation masks generated from the trained model to estimate the severity of the pothole?
- iii. How to combine GPS information and PCI calculation into an AI-based pothole detection framework for automated end to end pavement condition monitoring?

1.3 Overall Goal and Specific Aim of this MS Study

The overall aim of this research is to automate the pavement defect survey process for the rehabilitation and maintenance of road networks. The specific goal of this research is to leverage an AI based computer vision model with low-cost mobile camera technology to report potholes and their severity levels, as well as Pavement Condition Index (PCI) of the surveyed section, offering a scalable solution for urban and highway agencies.

1.4 Scope of Work and Study Limitations

This study focuses on the automated detection and assessment of potholes using AI based instance segmentation architecture. A custom pothole dataset is developed from smartphone mounted on car windshield and camera mounted on motorbike handle bar to provide more practicality to the dataset. Two instance segmentation architectures, YOLOv8s-seg and Mask R-CNN, are tested to evaluate their applicability in pothole detection under actual road scenes. Evaluation metrics used to evaluate both architectures are mAP, Precision, Recall and inference speed. The framework developed consists of the four components, pothole detection, severity

classification, GPS location mapping and PCI assessment within a single workflow. The severity of the pothole is estimated by area measurements obtained by converting segmentation mask pixels to area (m²) using conversion factor. The detected potholes are linked with GPS coordinates, which allows geo-spatial reporting of pavement defects.

The following limitations were identified during the development and field implementation of the proposed study:

- i. Available computing power and training period limited the training process.
- ii. The study focuses only on potholes and does not address other pavement distresses.
- iii. Fixed-height, forward-facing smartphone images may be affected by perspective distortion and lighting conditions.
- iv. Images were collected using both a car DVR and a motorbike-mounted camera; however, the effect of the two collection methods was not evaluated.
- v. Severity estimation uses an assumed depth range of 25-50 mm from ASTM D6433.
- vi. A fixed lane width of 3.5 m was assumed for pavement area calculations.
- vii. Pavement area was calculated using GPS start and end points, assuming straight road sections without curves, intersections, or roundabouts.

1.4.1 Rationale Behind the Variable Selection

Potholes are selected as the target pavement distress due to their significant impact on road safety, vehicle operating costs, and maintenance requirements [8]. To examine the detection accuracy and computational efficiency for practical pavement monitoring applications, the comparison of single-stage and two-stage instance segmentation architectures is carried out. YOLOv8s-seg and Mask R-CNN are chosen as instance segmentation architectures, as they offer pixel segmentation masking of the pothole, allowing the estimation of the area and classification of

the pothole severity [9]. Smartphone is used as main source to record survey as this approach is low cost, easily deployed and is appropriate for large-scale road surveys when compared to specialized pavement inspection equipment. Mobile GPS is used to report pothole location and managing asset dashboard. The segmentation masks are used for pixel-to-area conversion to estimate the size of potholes for severity assessment. ASTM D6433 guidelines are adopted for calculating deduct values, corrected deduct values, and pavement condition ratings due to standardized approach to pavement condition. Google Colab is used because it is accessible and has enough computational power for training and testing deep learning models, which is equipped with an NVIDIA T4 GPU. The road sections chosen are surveyed under normal traffic conditions to test the applicability of the proposed framework in real traffic conditions.

1.5 Brief Methodology

The research methodology consists of two stages. In the first stage, pothole data are collected and a custom model is trained. In the second stage, field deployment is conducted for pavement condition assessment. To train the model, road images and videos of potholes were first captured from roads of Rawalpindi and Islamabad using the vehicle mounted cameras under normal traffic conditions. A localized dataset created was then manually annotated with polygon-based masks to precisely define the boundaries of potholes. The prepared dataset was split into three parts: training, validation, and testing. Two instance segmentation methods were trained using the prepared dataset, one based on YOLOv8 and another based on Mask R-CNN. The models were evaluated and compared using standard performance measures such as precision, recall, and mean Average Precision (mAP). The best performing model was determined based on the comparative analysis and then deployed for field assessment.

The selected model was implemented through a mobile for automatic pothole detection in a real road survey. To estimate the dimensions and severity levels of the detected potholes, a pixel-to-area conversion factor was calibrated and used to estimate pothole area. The measurements were then used to estimate pavement

condition ratings using an automated assessment procedure. Results from the automated framework were evaluated and compared with manual calculations to report the percentage error. Lastly, the processed information, such as pothole location, severity, and pavement condition results, was displayed in an interactive map to support road maintenance and asset management.

1.6 Research Impact on Industry

1.6.1 Research Novelty - Uniqueness

This research develops a custom pothole dataset representing local road conditions for training and evaluating instance segmentation models. While there have been several studies on pavement defect detection using AI methods, the pothole detection using instance segmentation has received less attention than the traditional object detection. Moreover, few studies have been conducted to compare the single-stage and two-stage instance segmentation architectures for pothole assessment, such as YOLOv8s-seg and Mask R-CNN. In particular, there are few studies that combine the three aspects of automated pothole detection, instance segmentation, severity rating, GPS localization and Pavement Condition Index (PCI) evaluation in a unified low-cost system using smartphone camera. Consequently, the practical application of an end-to-end pavement condition monitoring system in actual road conditions has not been well studied and is still a gap in pavement management studies.

1.6.2 Research Significance and Benefit

This research is intended to enhance the condition assessment of pavements using AI and low-cost mobile imaging technology. The results will be used to develop more efficient and automated road survey techniques to use in pavement maintenance and rehabilitation. The proposed framework will help road agencies and engineers locate potholes, determine the condition of the pavement, and proactively decide when maintenance is needed.

1.6.3 Practical Implementation

The proposed framework has the potential to be a practical pavement inspection tool for road agencies, consultants and maintenance contractors. The use of the smartphone mounted on car windshield can enable pavement surveys to be done with standard low-cost equipment rather than specialized LIDAR inspection vehicles. A pothole inventory, severity data, geographical locations and Pavement Condition Index (PCI) evaluation can aid in planning maintenance, budgeting and prioritizing rehabilitation activities. The automated workflow also minimizes manual data processing and reporting, leading to quicker condition evaluation of roads. The framework can be modified for urban roads, highways and municipal road networks, and offers a practical solution for digital pavement asset management.

1.6.4 National and Global Impact with Emphasis on SGDs Relevance

The results of this research will contribute to the digital transformation of pavement management by making pavement condition surveys easier and more efficient. The proposed framework aligns with SDG 9 (Industry, Innovation, and Infrastructure) by advocating the application of AI in monitoring and maintaining infrastructure. It also contributes to SDG 11 (Sustainable Cities and Communities) by enhancing road safety and transportation systems. The study allows for timely maintenance interventions and decreases the need for rehabilitation activities, therefore supporting SDG 12 (Responsible Consumption and Production) with more efficient use of resources.

1.6.5 Research Challenges

The study faces challenges associated with acquiring pothole images under varying lighting conditions, pavement textures, and traffic environments. Creating a custom dataset demands a lot of manual annotation to produce accurate instance segmentation masks for model training and evaluation. Moreover, there are only a few public datasets for pothole available for local road conditions, which makes direct model benchmarking and comparison difficult. The proposed framework also

involves acquiring images through the use of smart phones, which can introduce distortion, motion blur and quality issues that can impact the accuracy of detection and area estimation. The GPS coordinates are recorded at a lower frequency than frames of video, and the interpolation methods must be used to estimate the locations of the potholes. Moreover, training deep learning models is also a costly process and ensuring the accuracy of pavement detection as well as efficient processing for practical pavement monitoring applications is a challenge. More pavement distress categories, better camera calibration and larger datasets could be used in future studies to further improve the performance of the framework.

1.6.6 Ethical and Management Considerations Including Risk Management

The study highlights transparency in reporting the findings, assumptions, and limitations, which will enable responsible AI use in pavement condition assessment. The images and GPS data taken during road surveys are only used for research purposes, with data confidentiality and privacy ensured. The framework is designed to facilitate engineering decision making and not to replace engineering judgment, acknowledging that there may be detection and classification errors. Throughout the study, compliance with the university research guidelines and safe field data collection procedures is ensured. In addition, the study encourages the adoption of low-cost, widely available technologies that can help to facilitate broader adoption by road agencies and infrastructure stakeholders. The study seeks to foster the responsible use of AI-enabled pavement monitoring systems by highlighting these factors.

1.6.7 Research Deliverable, Sale and Marketing Potential

The study provides an automated solution for detecting, measuring, reporting, and assessing the Pavement Condition Index (PCI) of potholes using low-cost smartphone imagery and GPS. The proposed system is designed for use by agencies, consultants, contractors and infrastructure management organizations who are looking for efficient pavement condition assessment methods. The framework

offers a practical solution instead of the traditional manual surveys, cutting down on inspection time, reporting and data processing. The custom pothole dataset and trained AI models also contribute to future research and development efforts in the field of pavement monitoring applications. The proposed solution is flexible and can be adapted to urban roads, highway, and municipal road networks, depending on the context in which they are used. In addition, the use of mobile device and ordinary vehicle increase accessibility and reduces deployment costs. The results of this research will encourage the use of digital technologies in pavement management and inform better planning and decision making for pavement maintenance.

1.7 Thesis Layout

Chapter 1 presents the introduction section, encompassing the background, research motivation, problem statement, research questions, objectives, scope of work, study limitations, research significance, research novelty, and thesis structure.

The literature review is presented in Chapter 2, which outlines conventional pavement inspection techniques, image processing methods, and machine learning and deep learning approaches. For pothole detection, single-stage and two-stage AI architectures are explored. Existing pavement datasets, model evaluation methods, and automated pavement condition assessment frameworks are also reviewed, and the identified research gaps are presented.

The research methodology is described in Chapter 3, which covers the study area, data preparation, image collection, annotation, single-stage YOLOv8 and two-stage Mask R-CNN model training parameters, and field deployment methodology.

Chapter 4 focuses on results and analysis, including model training performance, segmentation accuracy, computational efficiency, comparative evaluation of YOLOv8 and Mask R-CNN, discussion of findings, and summary of the experimental outcomes.

The field deployment section is presented in Chapter 5 and covers the implementation of the selected model for pothole detection, GPS integration, estimation

of the area of the pothole, classification of pothole severity, calculation of deduct value, assessment of PCI and field survey application.

Chapter 6 presents the conclusions and future recommendations based on the findings of the study. References are provided and annexures are provided at the end of the thesis.

Chapter 2

Literature Review

2.1 Background

Potholes are one of the most critical pavement defects because of their effect on road safety, vehicle operating costs and pavement deterioration. Therefore, timely pothole detection is vital for effective pavement rehabilitation. Pothole detection techniques have experienced significant transition over the past few decades because of traditional manual inspection systems, which were time-consuming, subjective and unsafe under live traffic conditions. Historically, these methods utilized manual tools (taps, rods, inspection sheets), requiring significant resources and time resulting in decentralized information. With the rapid technological shift in camera technology, these techniques transitioned to semi-automated methods where two-dimensional images were captured followed by simple image processing such as greyscale conversion to highlight cracks. But the performance of legacy methods was heavily dependent on two key factors, image quality and computational speed of the equipment. Subsequently, these techniques were unable to provide direct output for pavement health without human input. Three-dimensional point cloud imaging systems, with the use of portable devices such as UAVs (Unmanned Aerial Vehicles), have greatly increased the data acquisition capabilities

to provide detailed pavement surface representation, albeit with limitations including battery life, higher cost, and increased operational complexity. Fully automated techniques have emerged more recently with the incorporation of deep learning models, particularly convolutional neural networks (CNN) for object detection which have greatly enhanced the detection accuracy and efficiency through large annotated datasets of pavement distresses (e.g., rutting, alligator, potholes). Moreover, this has opened avenues for low-cost image capturing systems, with the use of a smartphone which can provide a practical and easy-to-implement solution. Although the image based detection techniques has played a paramount role in road condition surveillance, certain issues associated with environmental variability, dataset generalization and most importantly system integration remain major challenges. This chapter presents a detailed literature review of the suitable computer vision models, associated datasets and techniques using instance segmentation for measuring pothole severity which is required for the assessment of pavement condition. Section 2.2 highlights pavement defects in flexible pavement, focusing on potholes. Section 2.3 discusses potholes detection techniques as semi-automated approaches, including image acquisition and traditional image processing techniques and AI-based automated detection methods, providing a detailed analysis of available architectures such as object detection, instance segmentation, and semantic segmentation. Section 2.4 provides a detailed analysis of the YOLO series of models, on available parameters as performance, accuracy, and precision. It additionally examines available datasets and their limitations, followed by a review of existing studies on the integration of Pavement Condition Index (PCI) assessment. Finally, Section 2.5 summarizes the key findings and highlights the identified research gaps addressed in this study.

2.2 Pavements

Pavements are engineered surfaces designed to support the movement of vehicles and to isolate the vehicle loads from the natural soil. A pavement is a layered system that is designed to carry loads safely and distribute them to the underlying ground and to keep the stresses and deformations to acceptable levels over

the design life of the asphaltic roads. A pavement usually has four basic layers: subgrade, sub-base, base course and wearing course. These layers contribute to structural support, ride quality, skid resistance and protection from environmental influences. Pavements can be broadly divided into two types: flexible pavements and rigid pavements based on the material composition and the behavior of the pavement under loading. In flexible pavements, the traffic load is transferred to the subgrade gradually through the layers of the pavement, and the pavement consists of granular layers covered with a bituminous wearing course. Rigid pavements are mainly based on Portland Cement Concrete in which flexural strength of the concrete distributes the incoming wheel loads. Pavement type is selected and designed based on a variety of factors, such as the traffic loading, subgrade strength, material properties, environmental conditions, drainage, ride quality requirements, and desired level of reliability over their design life. Pavement are continually subjected to traffic loading and environmental conditions and their condition slowly degrades with time. Under repeated wheel loading, temperature changes, moisture intrusion and material aging, different types of pavement distress can occur, diminishing the structural capacity and rideability. These defects, if not treated in a timely fashion, can deteriorate to more serious and potentially unserviceable failures, and may pose hazards to the road users.

2.2.1 Distresses in Flexible Pavement

Road conditions impact overall economic level and development of a country, and it has been accepted as one of the key infrastructure indicators by the World Bank [10]. A well-maintained road network ensures seamless access to education, health, industry and goods transportation throughout the country. Roads play a vital role in regional connectivity and economic growth, providing significant geographical and economic advantages to a nation [11]. Road pavements are continuously under traffic loading and exposed to environmental effects throughout their service life. Distresses in the flexible pavement are caused mainly due to inadequate thickness, poor construction practices and surface drainage. And if not maintained properly pavements begin to crack and exhibit distresses over the time. These defects can

cause reduction in serviceability, pavement performance and overall deterioration which can lead to impact the safety of the road and higher maintenance costs.

In flexible pavements, failure is usually classified as either 2D (two-dimensional) or 3D (three-dimensional) defects. The 2D defects are mainly evaluated by surface area and appearance such as longitudinal cracks, transverse cracks, alligator cracking and block cracking. While, 3D defects refer to their surface area and depth, such as potholes, patches, rutting and corrugation [12]. These defects are typically classified into low, medium, and high severity levels depending on their dimensions and extent. The primary factors that contribute to pavement deterioration are traffic loading, climatic conditions, and inadequate surface drainage. These factors directly or indirectly accelerate the development of pavement distresses. These defects continue to worsen, eventually creating a safety hazard on road and the situation further leads to a costly reconstruction rather than routine rehabilitation.

The economic impact of delayed maintenance is substantial. In the United States alone, approximately USD 22.6 billion was spent on pavement maintenance activities in 2015, highlighting the financial burden associated with delayed and insufficient maintenance measures [13]. The same can be seen in Pakistan where motorways are regularly maintained and long-term rehabilitation costs are low. In contrast, the most serious faults are often ignored on major roads like the Grand Trunk (GT) Road, resulting in serious deterioration, expensive rehabilitation and further strain on the national budget.

2.2.2 Potholes in Flexible Pavements

Potholes are one of the most common and severe defects found in flexible pavements. They are the bowl-shaped areas of road surface texture that are formed due to the disintegration and removal of the road surface materials caused by traffic loading [14]. A pothole typically forms in an area where the pavement is cracked or has a surface defect that lets water into the layers of pavement. Over time, repeated wheel loads, moisture intrusion and deterioration of the substructure cause

the pavement to deteriorate and potholes to develop [15]. Their presence also reduces driving comfort and may cause damage to vehicle tires and suspension systems. Potholes grow larger and contribute to overall pavement deterioration and increased reconstruction costs.

Pothole defect has effects beyond just the deterioration of the pavement. They are a major contributor to road accidents, particularly during night-time and adverse weather conditions when visibility is reduced. Poor road surface conditions have been cited to cause about 1,000 traffic collisions and fatalities each year, many of which are preventable by adequate maintenance and repair [8]. In Pakistan, local welfare association stated that 100 people lost their lives and 1,048 people were injured in pothole accidents in the first 2 months of 2026 [16]. Potholes may also become stagnant water breeding grounds for mosquitos and can release dust into the atmosphere that has a harmful impact on environment and health in dry weather. Thus, early identification of potholes is crucial for increasing the safety of the road network, lowering maintenance expenses and sustainable transport systems.

2.3 Distress Detection Techniques

Pothole detection methods can generally be classified into traditional and modern approaches. Traditional methods are mostly based on manual road surveys, visual inspections, field measurements, and pavement condition reports produced by inspectors. While these techniques provide direct assessment of pavement conditions, they are labour-intensive, time consuming, costly, and require human judgement. To address these challenges, transportation agencies have increasingly turned to semi-automated and fully automated approaches for monitoring roads. Machine learning and deep learning techniques, specifically computer vision models have shown potential in the detection of pavement defects from images among these technologies. The modern methods provide quick, reliable and replicable solutions to the pavement condition assessment problem in comparison with the traditional methods.

2.3.1 Semi-Automated Pavement Assessment Techniques

2.3.1.1 Conventional Two-dimensional Image Processing

Imaging of Pavement distresses has evolved through 2D camera systems and basic image processing to high resolution 3D sensing and deep learning models (CNN, RCNN). Over time, the goals have been higher accuracy, lower cost, faster surveys, and better automation. Otsu [17], was a pioneer in development of an unsupervised, non-parametric threshold selection method to separate greyscale levels within an image for segmentation. This approach separated foreground and background features within an image to highlight cracks and has been widely adopted as a key technique in crack detection research. The method provided one of the earliest ways to convert greyscale images to binary cracks or no cracks masks to multilevel thresholds.

Following early segmentation approaches as the Otsus method of two-dimensional image processing began to emerge in the early 1990s when custom systems like the Automatic Crack Monitor (ACM) were built to collect pavement images containing linear cracks [18]. The system utilized gradient-histogram analysis algorithm that scanned the image from all sides to generate a binary image for crack extraction, is shown in Figure 2.1(a). The ACM used in this research converted binary images to separate crack pixels from background for further analysis Figure 2.1(b). From these binary images crack maps, the system extractive quantitative crack parameters for maintenance inputs. Despite using chain-code direction, detection was limited to specific types of cracking primarily longitudinal and temperature cracks. Other limitations included processing time, sensitivity to image gradient at edges and limited image sampling capacity (up to 90 images). Later, Zhu [19] also employed Otsus method (OTSU) on asphaltic pavement for crack identification on original image (Figure 2.1(c)), however the poor contrast between the grey regions and white background reduced the crack visibility (Figure 2.1(d)).

Researches have also reported classic image post-processing by converting base image into binary followed by denoising localized thresholding, edge detection

and simple feature extraction to measure crack parameters and classified various crack patterns [20]. Figure 2.1(e) and 2.1(f) illustrates conversion of binary image to a hole filled image. These methods were based primarily on two-dimensional grayscale or RGB images captured by standard digital cameras mounted on survey vehicles, but were still limited by their inability to capture irregular cracks, sensitivity to shadows, stains, low contrast and surface texture, and were moderately valid when compared to manual PCI ratings [21]. The setup cost is minimal, however core issue remained with two-dimensional intensity images that could not reliably distinguish true structural defects from visually similar objects such as oil, debris, or road markings [22]. This limitation can also be observed in Figure 2.1(g), where median filtering noising was converted to defused denoising in Figure 2.1(h) thereby reducing overall crack detection efficiency.

2.3.1.2 Three-dimensional Imaging and Sensors Based Methods

With advancements in sensor technologies and computational capabilities, pavement distress detection techniques have evolved towards three-dimensional imaging using specialized cameras in combination with LiDAR sensors and laser profiling systems. Laser-equipped survey vehicles are capable of generating detailed and precise 3D point cloud models to accurately spot defects such as potholes, rutting, longitudinal cracks with millimetre-level precision [27, 28]. Furthermore, 3D laser profiling technology can achieve submillimetre depth resolution, providing stable crack measurements across varying speeds and lighting conditions, with accuracy exceeding 98% [29]. However, there is higher capital and maintenance cost involved along with complex logistics which limit their adoption for smaller agencies.

A widely used industrial system, PaveVision3D Ultra, enabled network-level pavement evaluation at highway speeds of up to 100 km/h without dependency on external lighting conditions [30]. The PaveVision3D system is equipped with 2D and 3D line cameras mounted on a Digital Highway Data Vehicle (DHDV) capable of scanning lane-wide point clouds and texture images as illustrated in Figure 2.2. Li et al. [31], evaluated over 200 pavement images by employing pre-processing

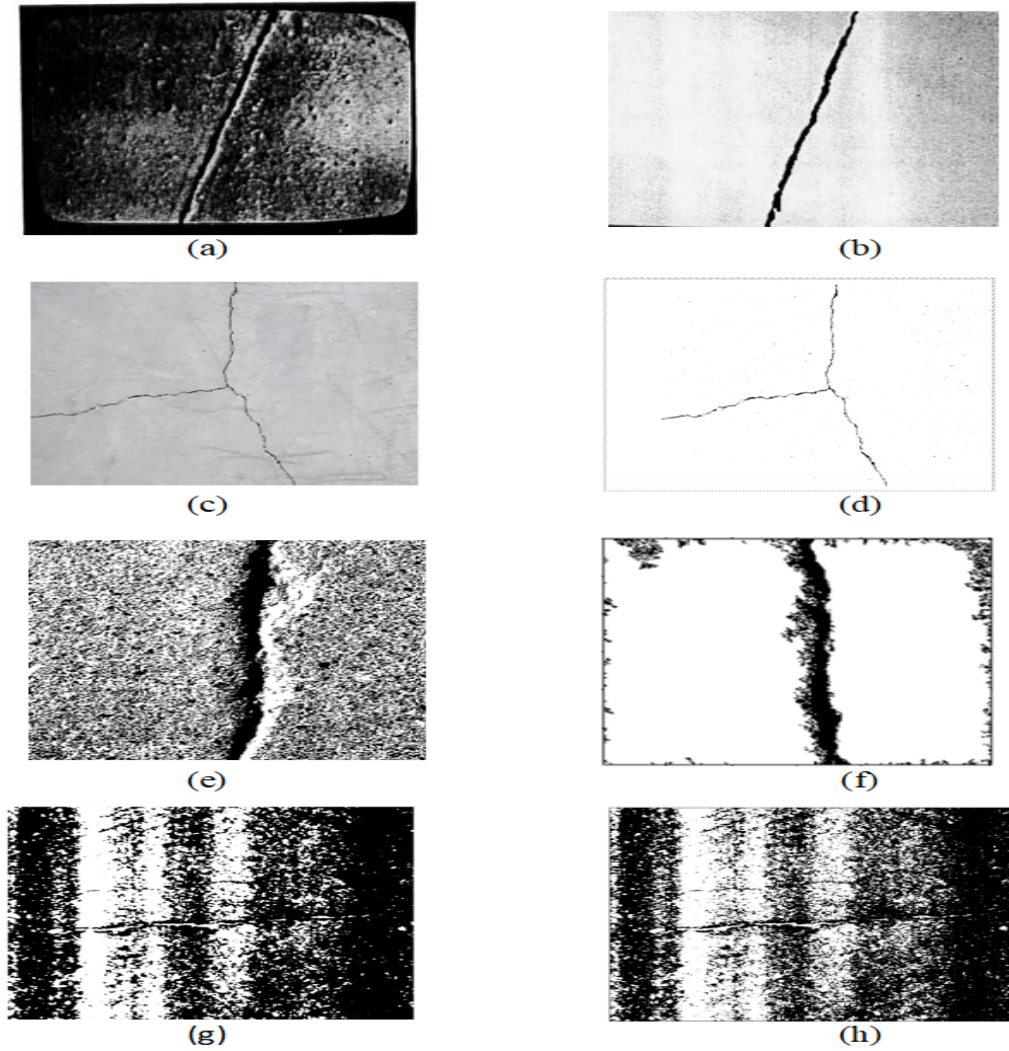


FIGURE 2.1: Gradient-histogram analysis algorithm (a) Original Image (b) Pavement crack (c) Segmented image [23] (d) OTSU processing results [24] (e) Initial binary image (f) Hole filled image [25] (g) Median denoising of crack image (h) Diffused denoising [26]

to reduce spurious noise followed by 2D tensor voting to achieve an average precision of 88.38%. Despite the higher accuracy, limitation resided in computation time of 10.3 s per image, largely due to tensor voting step, complex pre- and post-processing setup. Furthermore, the technique was more suited for analysing longitudinal crack depth, textures and rutting related defects but not the irregular crack as potholes. Figure 2.3 (b), illustrate a transverse profile of pavement crack scanned through a 3D camera.

Though three-dimensional camera-based systems offer a better detection of the

pavement defects, terrestrial LiDAR and laser scanning systems have more accurate geometric measurements through direct distance sensing. These systems allow detection of various types of distresses such as cracks, potholes, rutting, and defects in the joints with a vertical resolution of millimetres. Furthermore, they minimize crack geometry errors to a few centimetres compared to field measurements, providing overall accuracy ranging from 82% to 98% [27]. Canadas INO Corporation tested Laser Road Imaging System (LRIS) system shown in Figure 2.3 (a), a synchronized dual-camera setup able to cover 4m wide section of road in low lighting conditions [4]. Similar setup was used by Zhen Liu et al. [32], to develop PavDistress dataset based on a specialized pavement inspection vehicle mounted with a Basler raL2048-80km line-scan camera and an infrared laser-assisted lighting, which captured high-resolution grayscale surface images at 1 mm intervals on the S315 highway in China with a highway speed of 70 km/h. Each image covered a substantial road area (3.9 m x 2.1 m) while successfully labelling various types of distresses however, defects with a complex three-dimensional nature remain undetected.



FIGURE 2.2: PavVision3D Ultra System equipment with 2D and 3D camera sensors [33]

2.3.1.3 Unmanned Aerial Vehicle Based Inspection

UAVs, or drones, have emerged as flexible and economical tools for road inspections compared to traditional vehicle-based systems. They are not only cost effective but also time efficient, while maintaining stable height at centimetre-level 3D accuracy of crack geometry [35]. UAVs reduce the need of ground access which expedites the

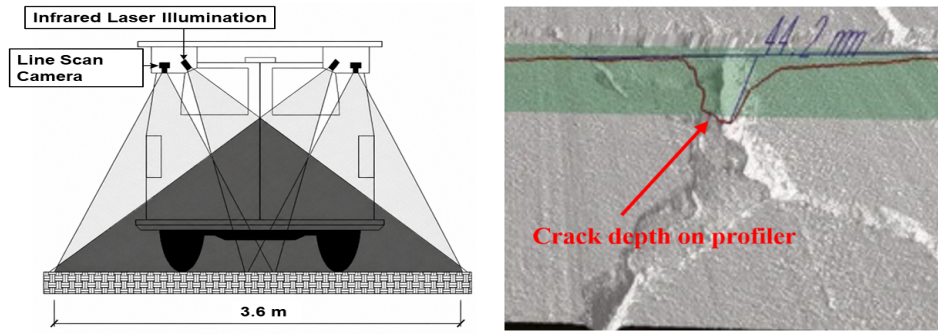


FIGURE 2.3: Application of 3D and sensor-based systems (a) LRIS dual-camera lane wide coverage [34]. (b) Green line presents transverse profile of pavement surface [33]

data collection and provide in-built computation ability to generate point cloud 3D models for potholes identification [36]. Romero-Chambi et al. [37] acquired potholes images through Phantom 4 Pro UAV, processed through Structure from Motion-MultiView Stereo (SfM-MVS) software and point clouds are compiled in Bentley software, shown in Figure 2.4. Their results recorded error of less than 1% at drone heights 10 m and 15 m. These optimal heights made UAVs more accurate, accessible, manoeuvrable, while offering safety benefits by reducing the need of surveyors in live traffic and hazardous locations. Samadzadegan et al. [38], captured 828 images of potholes using UAV, a multirotor drone attached with mapping camera ADTi 26MP, as illustrated in Figure 2.5. Images were processed using deep learning model to detect potholes more efficiently with higher accuracy of 81%. HighRPD, a dataset of 11,969 images collected at drone height of 50m, identifying potholes with a total of 1412 pit annotations, exhibiting a significant improvement in number of labels compared to previous studies [39]. Another study by Zhu et al. [40], utilized a UAV based dataset, namely UAPD, consisting of 3,151 images, and used deep learning models in post processing to report mean average precision (mAP) greater than 56%. Lee et al. [41] introduced an AI-based framework to reduce vehicle noise in images through street map generation, however this approach was not directly applied to pavement defect assessment.

Despite many advantages, UAVs are also limited by many factors such as battery life, airspace regulations and sensitivity to weather, lighting and wind which currently restrict their practicality for broader and long-distance surveys. UAVs

application is currently more efficient for local areas and assessments rather than full network surveys. Although, UAVs are suitable for detecting medium to large sized distresses, higher altitude for larger coverage reduces resolution making it difficult to identify smaller cracks and potholes. Images interrupted by the vehicles can also significantly diminish the effectiveness of drones during peak hours.

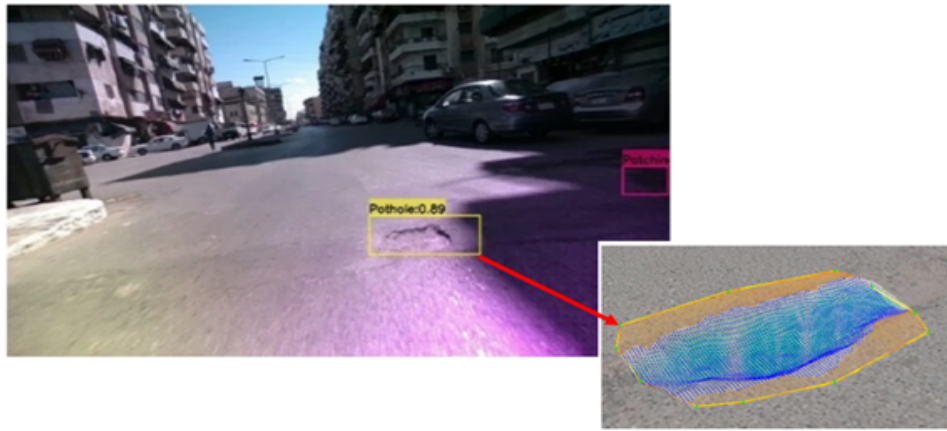


FIGURE 2.4: Volumetric representation of pothole in Bentley Software [42]

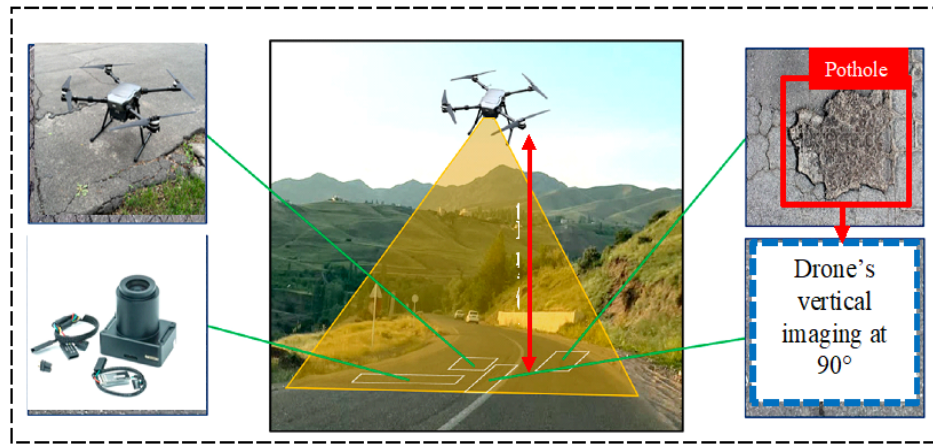


FIGURE 2.5: Multirotor drone with camera sensor for data collection [43]

2.3.2 Fully Automated AI-Based Assessment Techniques

2.3.2.1 Conventional Machine Learning Models

Artificial Neural Networks (ANNs) represent one of the foundational approaches in machine learning, inspired by biological neural networks. ANNs are composed of interconnected nodes linked through weighted connections, enabling the system

to learn relationships within given data. Hu [44], in 2013 developed a specialized ANN architecture consisting of six input nodes and five output nodes to classify five different types of pavement distresses, as illustrated in Figure 2.6 (a). Although the ANN based detection model was effective in defining pavement cracks, it required manually input date, which restricted the degree of automation.

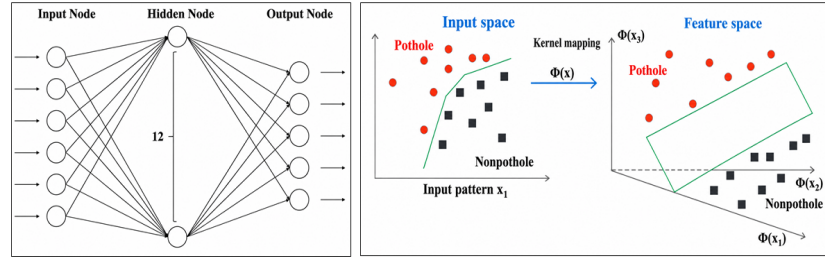


FIGURE 2.6: Traditional ML models used for pavement defect identification (a) Hu et. al [34] five output node model. (b) Hoang et. al [48] vector model

Besides ANNs, machine learning algorithms, including support vector machines (SVMs) and their variants, were also developed and tested for pavement distress using extracted image features such as textures and geometric properties. Hoang [45] in 2018 was one of the pioneers to implement the Least Squares Support Vector Machine (LS-SVM) model trained on texture-based features and classified pothole versus non-pothole regions (Figure 2.6 (b)). The approach demonstrated the effectiveness of LS-SVM in distinguishing surface defects and its potential applicability for automated pavement assessment by achieving an accuracy of approximately 89% and an area under the curve (AUC) of 0.96. However, the method was limited by the image dataset and its inability to estimate the size of potholes.

Machine learning models for pothole detection were proven to be low-cost, integrated solutions, especially with the use of hand-held devices such as smartphones, however, they faced limitations in generalization, threshold sensitivity, and varying road conditions [6, 46]. Sensor-based ML methods may further degrade performance when encountering irregular geometries such as potholes [47]. While deep learning (DL) approaches, particularly CNNs, deliver higher detection rates and improved performance across larger labelled datasets, they require more computational power. Overall, DL are regarded as the state-of-the-art models, while

TABLE 2.1: Comparison of DL and ML models on different aspects covered in literature review

S.No	Aspect	ML Models	DL Models	Reference
A.1	Models	SVM, LS-SVM, RF	CNNs (R-CNN, SSD, YOLO)	[44], [45], [49]
A.2	Extraction	Manual definitions	Automatic feature learning	[50], [46]
A.3	Dataset requirement	Small dataset but limited generalization	Requires large, annotated dataset	[50], [6]
A.4	Accuracy range	85 - 95%	95 - 97%, mAP up to 0.8	[6], [6, 46, 47, 51, 52]
A.5	Computational requirement	Low, suitable for limited hardware	Higher, can be optimized for light weight models	[6], [7]
A.6	Scalability	Limited to diverse environment conditions, large scale networks	Scalable with sufficient training data	[6], [50]
A.7	Strengths	Low cost, controlled dataset, applicable to sensor-based systems	High accuracy, self-learning, works in diverse conditions	[6, 46, 47, 51, 52]
A.8	Limitations	Lacks generalization, limited to varying road conditions, no deployment on mobile	Requires large dataset, annotation, high computation cost and deployment (Identified gap)	[6], [7]

classical ML approaches remain useful for resource-constrained or sensor-based systems. Table 2.1 presents limitations of ML in comparison with DL models discussed in section 2.3.2 on different aspects studied in literature review.

2.3.2.2 Deep Learning Object Detection Models

CNNs models are capable of learning relevant patterns directly from raw images. A typical CNN model consists of three main layers: a convolutional layer, a pooling layer, and a fully connected layer, The convolutional layer extracts image

features to differentiate crack patterns, the pooling layer reduces the spatial resolution of the feature maps, and the fully connected layer combines these features to generate the final classification of distress type [49]. CNN-based object detections are categorized into single-stage and two-stage detectors. Commonly used Single-stage based architectures are Single Shot Detector (SSD) and You Only Look Once (YOLO). Both skips the region proposal step by predicting bounding boxes and class labels (e.g., Pothole) in a single pass through the network (Figure 2.7). In contrast, two-stage object detectors architectures, such as Region-based Convolutional Neural Networks (Faster R-CNN), first generates region proposals and then classify the detected objects [53], as illustrated in Figure 2.8 Although R-CNN models typically achieve 4-10% higher mAP (mean Average Precision) than single-stage detectors (e.g., YOLO), this improved accuracy comes at the cost of significantly higher memory consumption, exceeding YOLO models by more than twenty times [7].

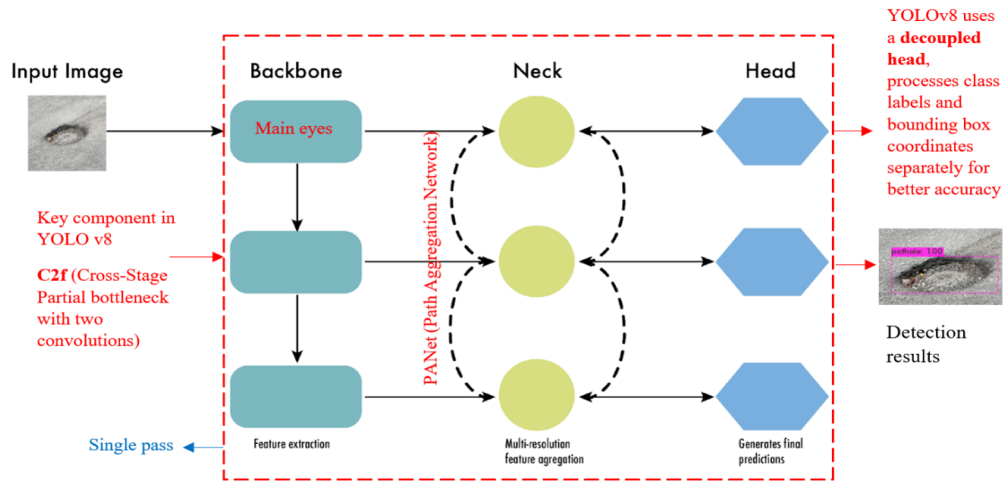


FIGURE 2.7: Typical network architecture of a YOLO-based object detector [57]

A comparative analysis of object detection models suggests that YOLO object detection models are widely adopted in pavement distress detection due to their fast inference speed and suitability for lightweight deployment, particularly for online and server-based inspection systems [54]. YOLOv8 used the RDD 2022 dataset and achieved mAP of 77% [6], the lightweight YOLOv8n (nano model) further improved performance to 84% [55], outperforming Faster R-CNN under certain conditions while maintaining efficiency for mobile deployment. Zhang et

TABLE 2.2: Performance analysis of Object Detection models against various datasets

Models	Datasets	Performance (mAP %)	Key Observations	References
CNN (ResNet)	MobileCrack	73.2	Limited suitability for vehicle-mounted applications	[55]
Faster R-CNN	UAV-based	79.3	Evaluated on 1,388 crack images; slower inference speed	[7]
Faster R-CNN	MIT-CHN-ORR	97	Supports multi-class detection (4 classes including potholes)	[54]
SSD (VGG-16)	Smart phone-based	77	9053 images – Alligator, ravelling cracks and patching	[59]
SSD MobileNet	RDD 2022	73	Lightweight model suitable for mobile deployment; includes pothole detection	[60]
YOLO v8	RDD 2022	77	Balanced accuracy with low error rate; supports 7 distress classes including potholes	[6]
YOLO 8s	UAV-based	74	Compact model with moderate accuracy; limited adoption in pavement studies	[7]
YOLO 8n	RDD 2022 - China	84	Nano model optimized for mobile deployment; includes pothole detection	[55]
OBC-YOLOv8n	RDD 2022	86	Custom model designed to capture complex and diverse road damage features	[55]
POT-YOLO	Smart phone-based	97	Hybrid model with precise pothole edge detection; lacks geometric measurement capability (Identified gap)	[56]

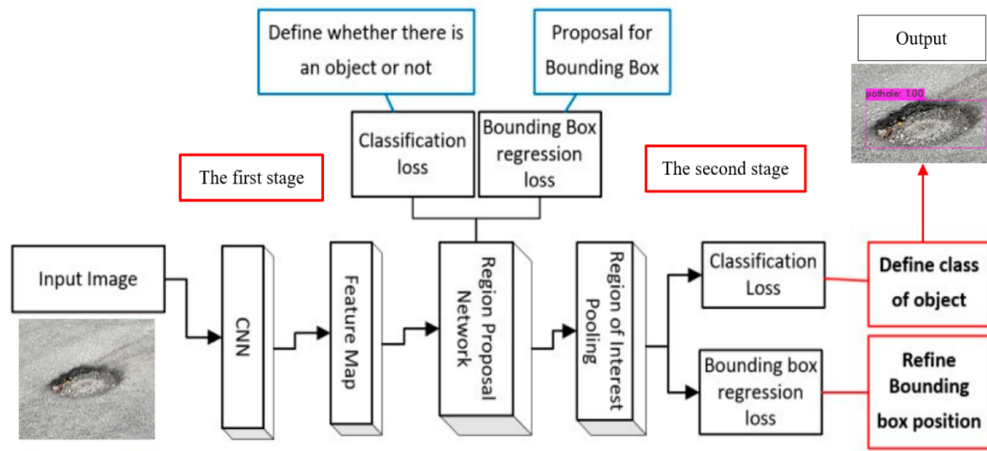


FIGURE 2.8: Illustration of Faster R-CNN [58]

al. [55] used OBC-YOLOv8n, a custom model, and demonstrated further gains in accuracy up to 86%. Bhavana et al. [56], recently developed a hybrid model, POT-YOLO, which showed excellent accuracy of 97% however, remained limited in providing geometric measurements. Table 2.2 presents a comparison of various object detection models used for pavement distress detection. Furthermore, majority of datasets represent countries such as China, India, and Indonesia, with minimal representation from Pakistan, which identifies a research gap in dataset generalization.

2.3.2.3 Pixel Segmentation Approaches for Pothole Detection

Despite the simplicity of object detection models as easy-to-implement and having low computational cost, they are limited to providing only bounding box identification which cannot provide the shape, measurement and severity of potholes, as illustrated in Figure 2.9 (a). Image segmentation also a pixel-level detection has two well-known approaches for potholes detection, Instance and Semantic segmentation [9]. Instance segmentation models, enable both object detection and pixel-level masking of each pothole with unique identification, providing higher level of granular information for severity assessment, maintenance cost and planning [61], as shown in Figure 2.9 (b). While major Semantic architectures such as U-Net, SegNet, and DeepLabv3+, classify all the pixels in potholes with a single class label, they do not differentiate between individual instances [62], as

indicated in Figure 2.9 (c). Therefore, when the two or more potholes overlap, semantic segmentation gives a combined masking area, which is not suitable to detect quantitative potholes defects or the Pavement Condition Index (PCI). Pereira et al. [63] utilized U-Net architecture for pothole image segmentation, reporting accuracy of 97% and a 0.86 IoU. Similarly, Li et al. [64] applied DeepLabv3+ segmentation method on Crack500 benchmark dataset achieving outperforming Intersection over Union (IoU) results for crack detection. Other researcher have developed specialized model, Han et al. [65] developed a novel semantic architecture, CrackW-Net and utilized widely used dataset Crack500 to demonstrate superior performance on pavement cracks. Similarly, Zhong et al. [66] employed a multi fusion model, Pavement Distress Detection Network (PDDNet), which outperformed the benchmark models (Mask R-CNN, U-Net) with mAP of 0.81 covering seven types of distresses.

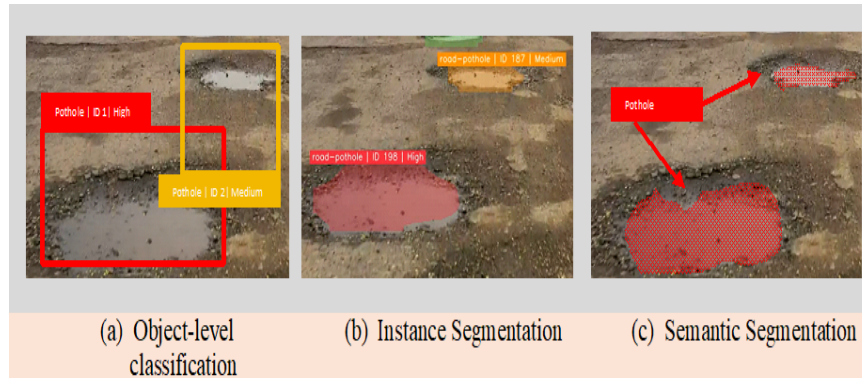


FIGURE 2.9: Pothole detection approaches [70]

Key instance segmentation architectures include Mask R-CNN (two-stage detector) and YOLO-seg (single-shot detectors). In comparison to semantic segmentation, it offers more finer details for complex geometric distress, such as potholes masking regions. Instance segmentation is also able to identify individual potholes for highly saturated defect zones. Recent studies have shifted toward instance segmentation, particularly for pothole detection, due to the aforementioned benefits that has facilitated its integration into end-to-end pavement assessment applications. An improved Hybrid Task Cascade (HTC) instance segmentation model utilizing ResNeXt101, BFP, and DCN was developed by Wang et al. [67] to accurately segment pavement distress with a recall rate of 96%. Li et al. [68] employed

TABLE 2.3: Comparative Analysis of Semantic and Instance Segmentation Architectures for Pavement Distress Detection

Method	Model	Performance Metrics	Key Limitations	References
Semantic	U-Net	Accuracy: 97%, IoU: 0.86	High accuracy but merges overlapping potholes.	[63]
Semantic	SegNet	Accuracy: 98%, IoU: 0.75	Covers crack, potholes, and patches but increased training time	[71]
Semantic	DeepLabv3+	Accuracy: 87%, IoU: 0.75	Strong feature extraction; limited to single-class (Potholes) segmentation	[72]
Semantic	Segformer	Accuracy: 86%, IoU: 85%	Quantify potholes, low depth error but lacks spatial data	[73]
Instance	Mask R-CNN	Precision: 70%, Recall: 41%	Covers Potholes and manholes, low recall rate for potholes	[74]
Instance	YOLOv8-seg	Precision: 82.2%, Recall: 74.4% mAP50: 80.5%	Inefficient on smaller potholes	[69]
Instance	YOLOv8-seg	Precision: 85%, Recall: 78% mAP50: 78.27%	Not suitable for varying weather and pavement textures	[75]

modified Mask R-CNN framework trained on a diverse multi-condition dataset of potholes, achieved a mAP of 92.1%. S. D et al. [69] trained the model using YOLOv8m instance segmentation architecture from the Ultralytics package, with a dataset of 3,074 annotated road images, the model achieved a precision of 82.2% and recall of 74.4%, demonstrating the model's efficiency while detecting and segmenting potholes across varying road conditions. YOLO based model in these studies reported slightly lower accuracy metrics compared to two-stage architectures (Mask R-CNN), but it provided a necessary balance between reduced computational demands and stable accuracy.

Literature review on pixel level segmentation approaches suggest an indicative trend toward semantic and instance segmentation architectures. While semantic

models were primarily benchmarked for general cracks until 2022, there has been a sharp rise in instance segmentation, particularly using the YOLO series of segmentation models, suited for real-time applications. Table 2.3 provides a comparison of both semantic and instance segmentation model on pavement distress detection. Semantic segmentation models such as U-Net, SegNet, and DeepLabv3+ demonstrate accuracy more than 85% with an average IoU of 75%. Mask R-CNN models are well known for their higher accuracy, but they are more resource-intensive, hence limiting their feasibility for mobile-based implementation. The YOLOv8 series of instance segmentation models show excellent performance with precision up to 99% and recall 85%, while maintaining inference speed suitable for real-time applications. This clearly indicates a growing trend towards lightweight instance segmentation models that balance accuracy and computational efficiency for deployment on mobile devices.

2.4 AI Model Evaluation and Application

2.4.1 Performance Review of YOLO models

The You Only Look Once (YOLO) family of models has emerged as the most popular architecture for high-speed object detection, primarily due to its ability to optimize the trade-off between predictive precision and hardware resource use. It is built on four main components, starting from input, passing through the backbone and neck, and giving final output [76]. Redmon et al. [77] first introduced YOLO in 2016, as a foundational algorithm for real-time object detection known as YOLO v1. YOLO v2 [78] and YOLOv3 [79] were further improved by the original author in 2018. Later the algorithm has undergone significant evolution (from v4 to v12) and maintained by Ultralytics [80]. In 2026, Ultralytics published YOLO version 26, designed specifically to operate in the low-power computational settings, establishing the future of mobile-based highway monitoring and maintenance [81]. Du et al. [82] were among the earliest to utilize the application of YOLO for pavement defect detection, developing a dataset of 45,788 images, achieving processing speeds up to nine times faster than traditional Faster R-CNN models, while also covering seven types of pavement distress including potholes. Asad et al. [6]

conducted an analysis of YOLO models (v1-v5) for pothole detection, reporting that YOLOv5 achieved the highest mAP of 95%. Sarange et al. [83] reported that YOLOv8 achieved a superior mAP of 82% and a precision of approximately 0.85 for pothole detection, compared to earlier versions. Although SSD demonstrated faster inference speeds in some cases, YOLOv8 provided a better balance between accuracy and efficiency. The most recent version of YOLO v8 architecture, released in 2023, introduced improvements in inference speed, optimized accuracy, and reduced model size through fewer parameters and FLOPs compared to earlier versions [84].

Recent studies showed, YOLOv8 has been more widely adopted for real-time object detection across various pavement distresses, including potholes. For example, autonomous vehicle systems built around YOLOv8 have been reported to achieve detection accuracies of 99% [85], and other researchers report mAP values of up to 85% when using mobile-based systems, which is sufficiently high for practical use [86]. Aquib et al. [87] compared YOLOv8 with YOLOv5 and YOLOv7, demonstrating that YOLOv8 outperformed earlier versions in both precision and accuracy. As illustrated in Figure 2.10, YOLOv8s (small-sized variant), shows performance closest to YOLOv5l (large-sized variant) with the bounding boxes confidence score exceeding 0.9, whereas YOLOv7l identifies one extra pothole hinting false-positives detection (Figure 2.10 (e)), for smaller sized potholes. Overall, recent studies indicate a clear trend toward the adoption of YOLOv8 due to its balanced performance, improved detection capability, and reduced computational complexity. As illustrated in Figure 2.12, YOLOv8 achieves higher detection performance with fewer parameters and lower latency compared to earlier YOLO versions, making it more efficient for mobile based applications.

YOLOv8 natively supports object detection, image classification, and instance segmentation tasks. Specifically, the YOLOv8-seg model combines the bounding box prediction (mAP_{box}) and pixel-level masking (mAP_{mask}), allowing more precise detection along with measurement of potholes (Figure 2.11). The model is available in multiple scaled variants (n, s, m, l, x), allowing a trade-off between computational efficiency and segmentation accuracy [88]. The fastest inference

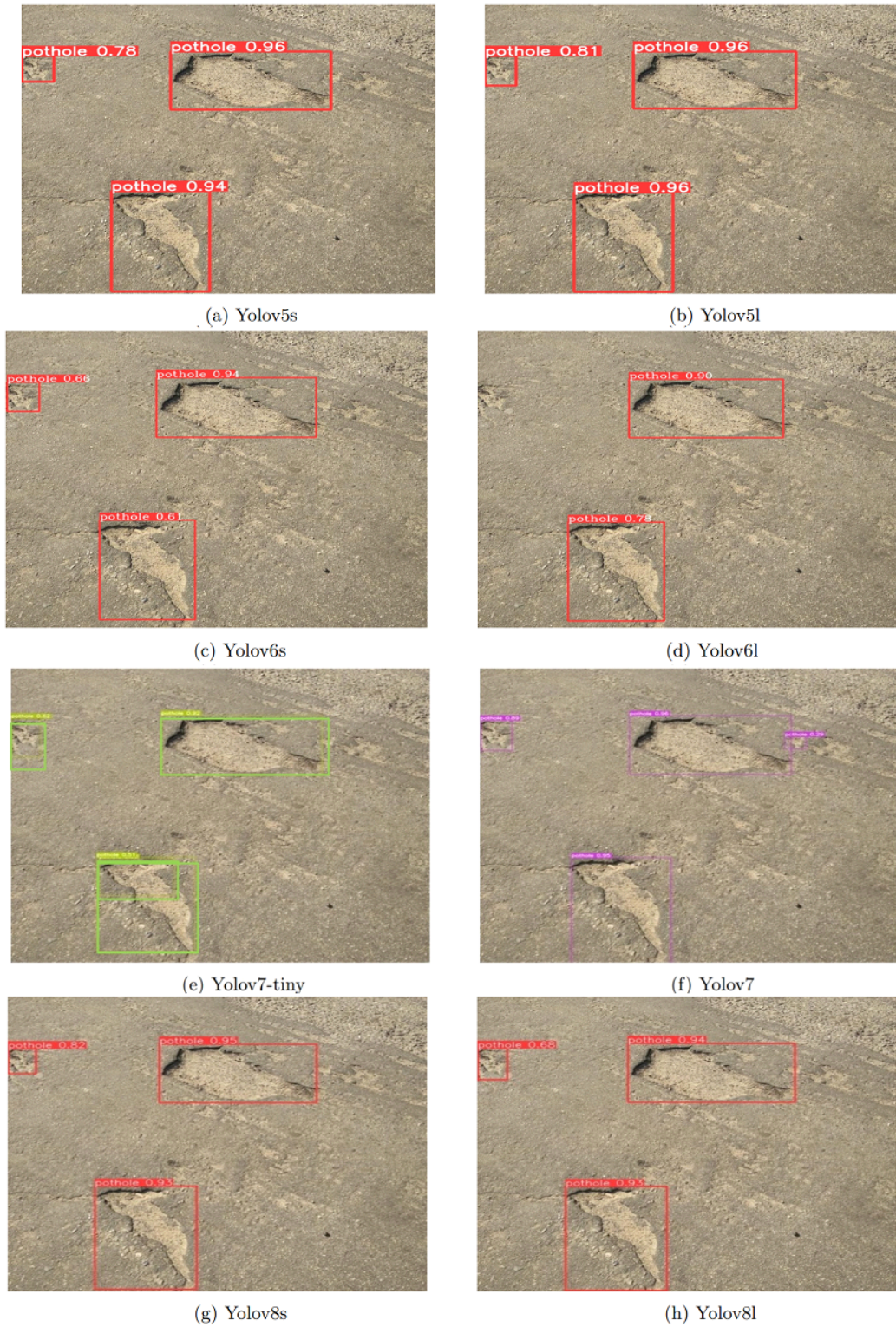


FIGURE 2.10: Outputs of the YOLO series model [89]

time is provided by YOLOv8n-seg, and the more accurate is YOLOv8l-seg, however it requires higher computational resources. As shown in Table 2.5, YOLOv8s-seg provides a balanced trade-off between accuracy and computational efficiency,



FIGURE 2.11: YOLO 8v-seg bounding box and pixel-level masking of pothole [90]

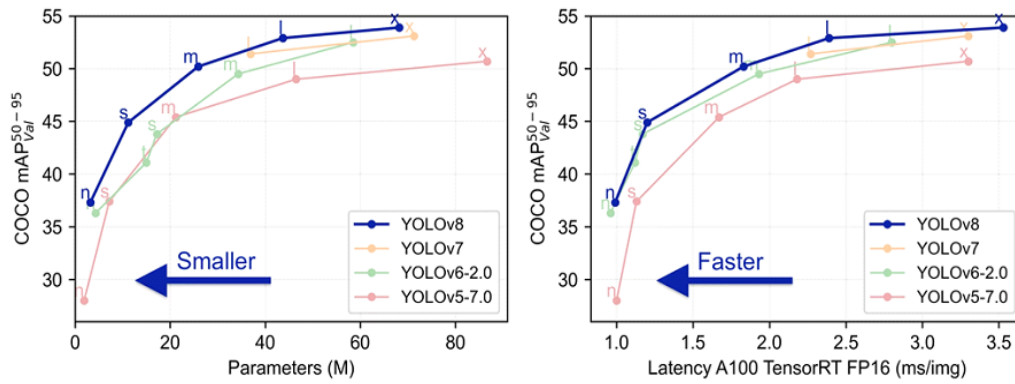


FIGURE 2.12: Performance comparison of YOLO models [91]

achieving higher segmentation accuracy than YOLOv8n-seg while maintaining significantly lower inference time compared to larger variants such as YOLOv8m-seg. Although newer models such as YOLOv26 demonstrate improved inference speed (approximately 24% faster than YOLOv8s-seg), their application in pavement distress segmentation remains limited.

Overall, the application of YOLOv8 has become more common, but the available literature suggests that current studies are more focused on potholes object detection while fewer studies reported the pixel-level segmentation which is required for pavement distress severity analysis. Despite the recent release of models such as YOLOv12 and YOLOv26-seg (2025-26), showing better performance in their speed and computational efficiency, there is still a lack of evidence on their stability and use case in pavement distress segmentation. The comparison shows

TABLE 2.4: Inference speed and model size of YOLO seg models by Ultralytics [80]

Model	mAP_box (50-95)	mAP_mask (50-95)	Speed (ms)	Params (M)	Use Case
YOLOv8n-seg	36.7	30.5	96.1	3.4	Fast and light weight, suitable for edge deployment
YOLOv8s-seg	44.6	36.8	155.7	11.8	Balanced between accuracy and speed, recommended for mobile deployment
YOLOv8m-seg	49.9	40.8	317.0	27.3	Higher accuracy and large model size
YOLOv8l-seg	52.3	42.6	572.4	46.0	Suitable for larger deployments
YOLO12s	48.0	–	–	9.3	Not trained on segmentation tasks
YOLOv26s-seg	47.3	40.0	118.4	10.4	Faster inference speed and accurate but lacks validation

Note: mAP = mean Average Precision; mAP_box = bounding box detection accuracy; mAP_mask = segmentation (mask) accuracy; ms = milliseconds (inference time per image); M = million parameters, model size

that YOLOv8 provides a suitable balance between detection performance, inference speed, and computational requirements among the evaluated YOLO variants. The segmentation capability of YOLOv8s-seg also makes it suitable for pixel-level pothole assessment required for area and severity estimation. Thus, YOLOv8s-seg is a consistent and practical option to use in mobile-based pothole detection systems.

2.4.2 Pavement Datasets and Limitations

The performance and generalization capability of deep learning models for pavement distress detection are highly dependent on the three key factors; image quality, dataset variability, and scale of the training dataset. To detect potholes, the dataset should be able to capture the different shapes, sizes, and surface texture under varied lighting and weather conditions to guarantee effective results in real-life conditions [92]. Several publicly available datasets have been developed to support automated pavement distress detection. For example, RDD series (2018-2022) has one of the largest collections of up to 47,000 images related to pavement

distress from many countries, in Central Area, Europe , and the United States, capturing various types of distresses. [93]. The RDD2022 dataset, introduced through the Crowd based Road Damage Detection Challenge (CRDDC 2022), further increased the data diversity by including those obtained with the help of dash-mounted mobile, setup presented in Figure 2.13 (a) [94].



FIGURE 2.13: Mobile-Based Data Collection Setup for Pavement Distress Detection (a) Dash-mounted mobile device [99]. (b) Smartphone mounted on front windscreen [100]

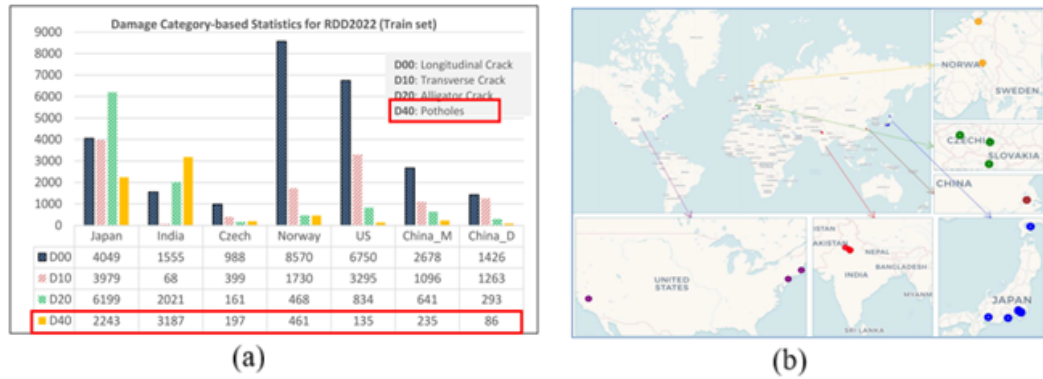


FIGURE 2.14: Statistical overview of the RDD 2022 dataset [101] (a) Distribution of pavement distress classes. (b) Geographic distribution of data collection sites.

In addition, UAV-based datasets such as UAV-PDD2023 provided an aerial pavement imagery with annotated labels, enabling alternative perspectives for distress detection [95]. Region-specific datasets such as GAPs (Germany) [96], EGY_PDD (Egypt) [12], and ISTD-PDS7 [97] contributed to geographical diversity, with dataset sizes ranging from 2,000 to over 14,000 annotated images. Furthermore, Rezaeimanesh et al. [98] developed a dataset (Attain) based in Iran, collected

TABLE 2.5: Summary of Publicly Available Datasets for Pavement Distress Detection

Dataset	No. of Images	Classes	Annotation Type	Region	Limitations
RDD 2022 [93]	47,000	7 (incl. potholes)	Bounding box	Japan, India, US, Czech Republic	No segmentation, few potholes
UAV-PDD2023 [95]	3,000	Multiple	Bounding box	China	UAV based, limited generalization
GAPs [96]	2,000	Cracks, potholes	Bounding box	Germany	Moderate size, no segmentation
EGY_PDD [12]	14,000	4 (incl. potholes)	Bounding box	Egypt	Region-specific, detection only
ISTD-PDS7 [97]	6,500	Multiple (incl. Potholes)	Bounding box	China	Limited potholes coverage
Attain [98]	2,293	10 (incl. potholes)	Pixel-level	Iran	Limited generalization and climatic factors

wide range of distresses for PCI calculation using a mobile device mounted on a car windscreen at a speed of 20-70 km/h, as shown in Figure 2.13 (b). Despite the availability of multiple datasets, they are still geographically concentrated, with limited evidence supporting pixel-level (segmentation) annotations and underrepresentation of potholes. Table 2.6 presents globally available datasets, highlighting their limitation with regards to their location and annotation type.

The primary limitation of publicly accessible datasets lies in their class imbalance, where potholes are not represented as frequently as other forms of distress. This is evident in Figure 2.14 (a), where RDD2022 dataset shows the lowest proportion of pothole images compared to other classes. Additionally, most datasets are region-specific, with images mainly collected from East Asia, Central Europe, and the Middle East, limiting model generalization across diverse road environments, as shown in Figure 2.14 (b). Another key limitation is the lack of pixel-level annotations, as many datasets rely on bounding-box labels, restricting their application for accurate pavement health assessment. Furthermore, there is no publicly available dataset representing potholes in Pakistan, highlighting a significant gap in region-specific data for model training and validation.

2.4.3 Automated Pavement Condition Rating Frameworks

Regional agencies use a range of rating indices to assess the health of the roads by incorporating certain physical characteristics of the pavement. These evaluation systems may vary regionally, typically based on factors such as local infrastructure, maintenance budgets and environmental conditions [102]. Traditional assessment procedure is consisted of three steps, first acquiring field data, either through 2D or 3D sensors mounted on a vehicle (Section 2.3). In second step, engineers inspect images manually based on their experience and decide level of severity of distresses. In final step defects are quantified and assigned specific rating index for a specific chainage [103]. This manual method of evaluating pavement health is tedious, time-consuming, and sometimes inconsistent. In a fully automated system, images are captured through a mobile-based setup, followed by deep learning techniques to identify defect types, estimate crack dimensions using advanced algorithms, and then evaluate the final condition of the pavement. The overall process is implemented as a single pipeline, from identification to pavement condition evaluation, involving minimal human intervention and making the maintenance process more cost-effective.

There are several methods for pavement assessment; the most commonly used include Surface Evaluation Rating (PASER) [104], Pavement Condition Index (PCI) [105], and Pavement Surface Condition Index (PSCI) [106]. The PCI method is the most widely used and is based on the ASTM D6433-18 standard [105], which evaluates pavement condition on a scale of 0 to 100, where 0-10 is regarded as failed and above 85 is considered Good (Figure 2.15). PCI evaluation follows a three-stage process: identification of distress type, evaluation of severity, and calculation of distress extent. Potholes are counted in numbers and categorized as per their severity levels namely Low, Medium and High as defined in ASTM D6433 (Table X1.1). The Deduct value (DV) for each category is determined using the Distress density graph (Figure X3.19) in ASTM D6433 [105], and final PCI is calculated using Equation 2.1 and 2.2.

$$\text{Density} = \frac{\text{Total distress quantity}}{\text{Section area}} \quad (2.1)$$

$$PCI = 100 - \max(CDV) \quad (2.2)$$

Where:

CDV = Corrected Deduct value

Maeda et al. [107] developed a mobile phone-based automated detection system that identified distress types (e.g., potholes) and reported their geospatial coordinates using the Google Maps API. The authors utilized the MobileNetV2 architecture to detect eight types of distresses with a precision greater than 77% however, the low-cost setup limits the vehicular speed at 23 km/h and lacks a detailed assessment report. Other studies have also reported the use of smartphones mounted on dashboards to capture images and identify distresses through deep learning techniques, limiting to the first stage of PCI calculation [12, 96–98]. Ibragimov et al. [76] proposed end-to-end solution automating PCI calculation by utilizing YOLOv8 in combination with OpenCV algorithms to identify and estimate crack widths for PCI calculations. This methodology achieved an accuracy of over 90% but limited to longitudinal cracks in asphaltic pavements, and the use of high-resolution images reduced its cost-effectiveness. Qureshi et al. [108] presented a detailed approach for automated pavement assessment using camera mounted on a vehicle by processing images using segmentation models. The approach employed data cleaning, image resizing and machine learning methods however it relied on road experts to train model by rating images on a scale of 0-10, limiting its use on variety of datasets. In another study, Rateke et al. [109] applied image classification method to report three levels of rating good, regular and bad on potholes with accuracy over 98% on all three types. The employed method was restricted to only three-level rating scale limiting its practical use in road network surveys.

Despite all advancements in compute vision models there no recent work that can effectively address the image quality, standardization and subject of use for automated pavement condition assessment. Available datasets lack the generalization, environmental variables and fall short in collecting complex potholes pattern and variations. Additionally, most studies on automated road assessment have focused object-based detection using older annotation tools as labellmg [109] for training,

validation and testing, which is not efficient and time consuming. Also, limited alignment with established standards (e.g., AASHTO and ASTM) for PCI calculation results in inconsistencies in output. Furthermore, there is no notable solution that provides end-to-end solution for road survey assessment from recording defects to store data on central PMS database for asset management.

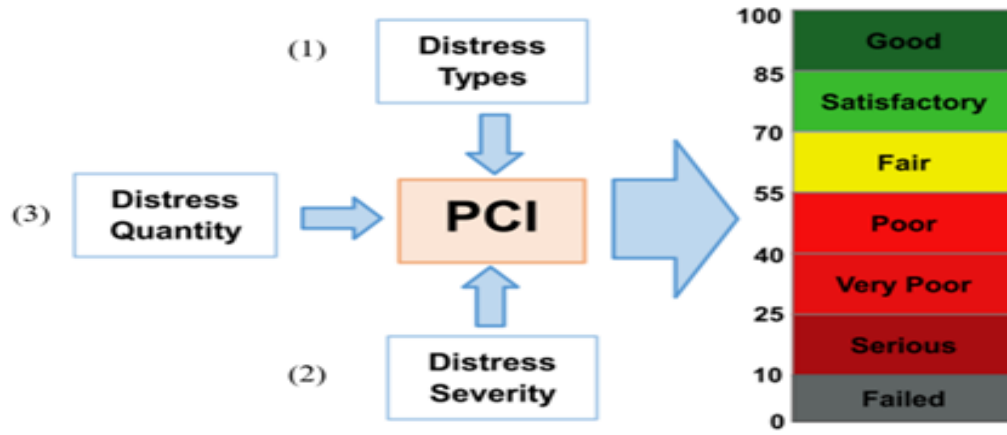


FIGURE 2.15: Rating scale of Pavement Condition Index (PCI) based on ASTM D6433-18 [110]

2.5 Summary

Comprehensive literature review has been undertaken to assess the development of pavement distress detection, from traditional inspection methods to fully automated artificial intelligence (AI) approaches. Traditional approaches were based on manual inspections and basic two-dimensional image processing, which were highly dependent on image quality, weather conditions, and relied more on human judgment. Three-dimensional sensing, LiDAR and the use of UAV to acquire data substantially enhanced the accuracy of detection and the representation of distresses. But they are still limited by the equipment cost, system complexity and scalability. Semi-automated methods reduced field workload but required high levels of human intervention in processing, leading to inefficiencies for network-wide monitoring. Lately the shift to deep learning techniques particularly CNN-based object detectors such as YOLO models have demonstrated strong performance in

real-time pothole detection. However, current methods are primarily limited to either bounding-box detection or semantic segmentation, which cannot represent the precise geometric characteristics needed for quantitative analysis of potholes.

Instance segmentation model have overcome challenges by allowing pixel-level representation of individual distresses however its implementation remains limited due to higher computational demands, lack of balance in the dataset, and limited testing under varying lighting conditions. Furthermore, there are high false positives reported in existing studies classifying debris and patches as potholes. Small sized potholes are also presented as challenge where AI models reported lower recall rates making them harder to detect. Publicly available datasets are also limited to specific countries which represent low number of potholes in comparison to other distress types, and absence of trained dataset for roads in Pakistan. Existing works also lack integration with the standardized PCI calculation, with a focus on the accuracy of detection rather than severity, GPS location, and system deployment. Therefore, literature review suggests a need for a scalable framework that combines light-weight instance segmentation model, localized dataset and computation of standard PCI for a road monitoring system. This research builds on these opportunities by developing a mobile-based framework for detection, measurement and geo-mapping for potholes.

Chapter 3

Methodology

3.1 Background

Accurate pavement condition assessment requires not only reliable detection of distresses but also their quantitative evaluation for a standardized pavement rating system. Recent research studies are limited to object-level identification of potholes without capturing their geometric properties needed to assess actual pavement conditions. Moreover, the unavailability of localized datasets and realistic deployment strategies limits their use in real-life road monitoring systems. This chapter presents the methodology adopted for automated pothole detection using a deep learning image segmentation approach that is cost-effective and suitable for Pakistan. To achieve this, baseline dataset was prepared by capturing local road conditions of Rawalpindi and Islamabad using camera mounted on a car dashboard and motorcycle. Two approaches are used to train the models, and the weights of the best-performing model are selected for field deployment. The final stage describes the application of the trained model in a practical framework to compute pavement condition rating in a real-world survey environment.

3.2 Study Area

The present study is conducted in twin cities of Islamabad and Rawalpindi of Pakistan, for its varied infrastructure, and high traffic volumes during peak times, thus causing undue stresses on the road infrastructure. The survey sites were chosen to represent different pavement distresses, traffic loads and maintenance practices within a given area.

For the survey, two different types of vehicles were chosen, a car with a camera mounted inside the dashboard and a 2-wheeler bike with a camera mounted on the handlebar. During the survey, two extreme weather conditions were selected, full sunny and cloudy, and images were tagged accordingly. The data collection was carried out between 10:00 AM and 4:00 PM to capture varying solar and contrast levels, over the 32 kilometres of asphaltic concrete pavement at five different locations (Figure 3.1).



FIGURE 3.1: Data collection sites, Islamabad-Rawalpindi, Pakistan.

Park Road and Kahuta Road were the major arterial roads, with 2-lane carriageways that experience high traffic loads and are poorly maintained, presenting higher severity pothole distress. While the PWD, Kuri road and NUST-PSO were minor arterial road with residential access. They are aged asphaltic roads with moderate traffic intensity and have low to medium severity potholes. The

NUST-PSO and Kahuta roads were surveyed using a motorcycle with a gooseneck-mounted camera attached to the handlebar, while Park Road and PWD Road were surveyed using a car with a DVR mounted on the windscreen.

Each selected road segment was surveyed at least once under the defined field conditions, while some sections, such as Kuri Road, were surveyed more than once using the car-mounted camera. The collected data were considered sufficient for evaluating the proposed framework under the surveyed conditions. These areas ensure that model is trained with potholes anomalies from the local area, such as varied pothole shapes and textures, not present in available open-source global data sets.

3.3 Methodology

Main purpose of this methodology is to improve the dataset quality by introducing local dataset conditions that compliment, poor lighting conditions, variability in different sizes of potholes with fixed camera perspective. To improve the performance for real world scenarios more complex and small sized potholes are used to evaluate the detection performance of trained models. The research methodology focuses on instance segmentation using custom dataset of potholes in two phases, first is to develop the model (training and evaluation) and then to deploy the model to the field to do the practical pothole detection and pavement status review as presented in Figure 3.2. The initial phase entails gathering of pothole images on the vehicle mounted cameras in various road and environmental conditions to ensure that the dataset captures diverse real-world scenarios from a forward-facing driver-perspective. The images obtained are then annotated in Roboflow with the help of polygon-based masking to accurately reflect the potholes edges. After annotation, image pre-processing and augmentation are applied, which includes resizing, filtering, and horizontal flipping, and the dataset is exported in COCO segmentation format for training. Two deep learning methods are selected to train and compare model performance. First, a single-stage instance segmentation approach using YOLOv8 and its lightweight variants, and second, a two-stage

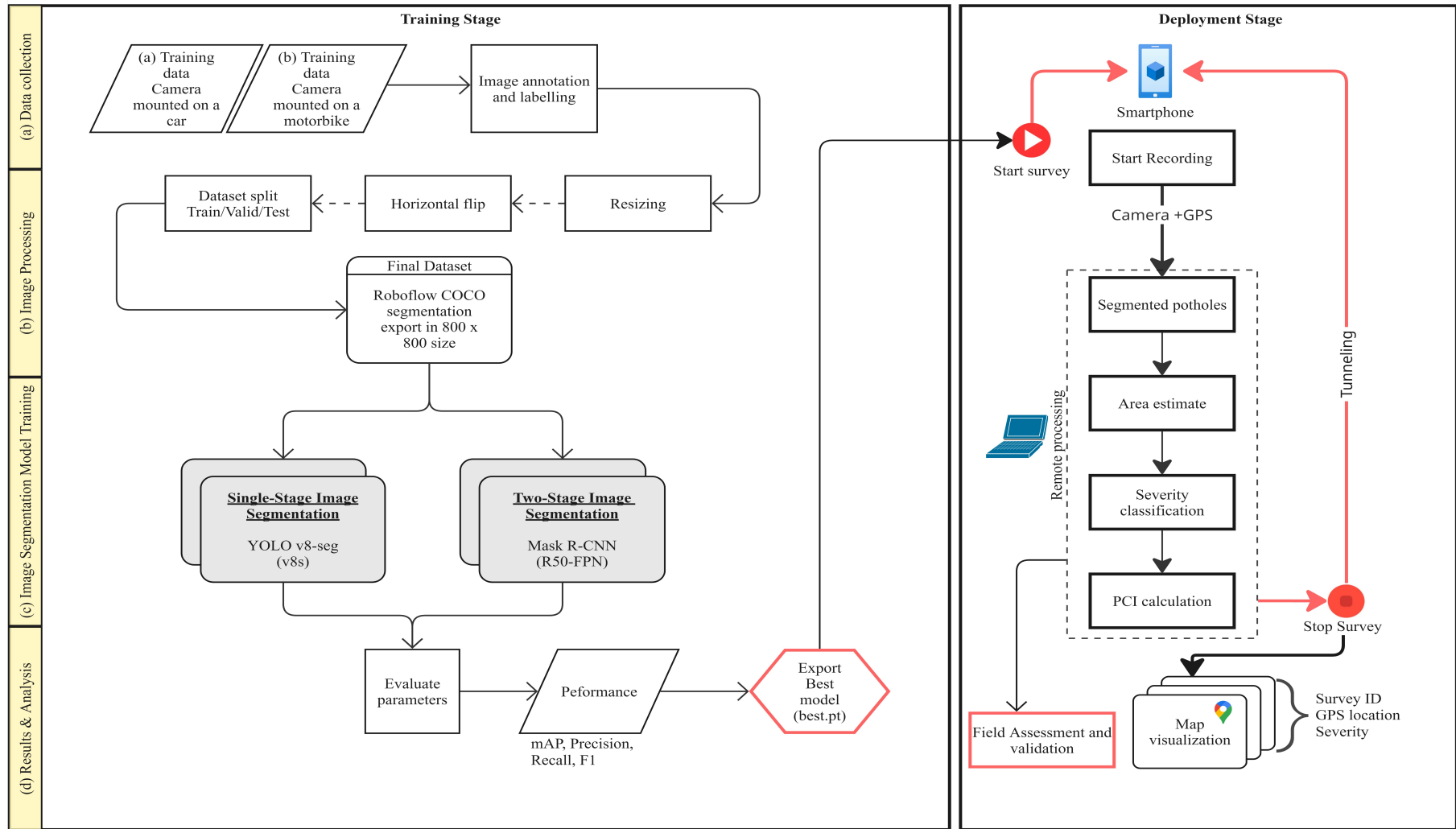


FIGURE 3.2: Methodological diagram for Pothole detection and PCI reporting

instance segmentation approach using Mask R-CNN implemented in the Detectron2 framework. Training hyperparameters are applied to control the learning process and computational efficiency. Models are then trained and validated using standard performance measures of mean Average Precision (mAP), precision, and recall. In the final step, a comparative study is conducted to determine the best model with regard to accuracy and efficiency in pothole detection and the chosen model is inputted in second phase of the study.

The best performing model is deployed to a real-time detection pipeline in the field deployment phase, with the help of a mobile camera system. A region of interest is applied to target the road surface, and image cropping to make sure that the detection is always consistent within the camera field of view. Geographic coordinates are recorded for each detected pothole using mobile GPS, and interpolation methods are applied between frames. The pothole regions are analysed in the final step to calculate Pavement Condition Index (PCI) values according to the pothole severity ranges complying with ASTM D6433 standard. The estimated results are directed back to mobile device for direct reviewing of PCI report and branched to OpenStreetMap for visualization of potholes with their GPS coordinates. This framework also enables storing of survey data under central database for asset-level reporting.

3.4 Dataset Preparation

3.4.1 Image Collection

This study presents a new, localised dataset, addressing gap in the literature by offering primary data relevant for the development of a pothole detection model. The data were collected from a range of urban and residential areas in Rawalpindi and Islamabad, which included a variety of road types, from main arterial roads to minor residential streets. These areas provided variation in pavement textures under different road widths ranging from 3.5m to 1.5m. The dataset is comprised of 350 pothole images captured using three different data collection platforms to

provide a high level of variability and diversity in camera angle and image quality. The main data was collected through a car DVR system recording at 30 FPS in Full HD (1920 x 1080 pixels) resolution, with vehicle speed ranging from 60 to 80 km/h (Figure 3.3 (a)). Images are then extracted and cropped from recorded videos to ensure consistent input size for the training stage. Additional data was collected using a smartphone fixed to the dashboard of the vehicle to obtain additional images with a resolution of 4032 x 3024 pixels. To build a cost-effective mobile system for collecting pothole images, a smartphone was mounted on motorcycle using a Gooseneck flexible mobile holder to capture images at the resolution of 1280 x 720 pixels (Figure 3.3 (b)). Third method of data collection was deliberately introduced to assess the performance of the model under the effects of vibration and lean angles, which would help to support the local surveying in Pakistan. All three setups maintained a fixed, forward-facing camera configuration at a consistent mounting height to ensure uniform perspective during data acquisition. A total of 350 images were initially collected, providing a localized dataset of pothole instances under different road and environmental conditions. Considering that the study focuses on a single distress class, the dataset was further expanded through augmentation to increase the variation available for model training.



FIGURE 3.3: Mobile-based system for data collection (a) DVR camera mounted on car windshield. (b) Gooseneck Mobile holder assembly mounted on motorcycle handlebar.

The images captured using the motorcycle and car-mounted camera setups show differences in viewing perspective, image quality, and pothole visibility due to

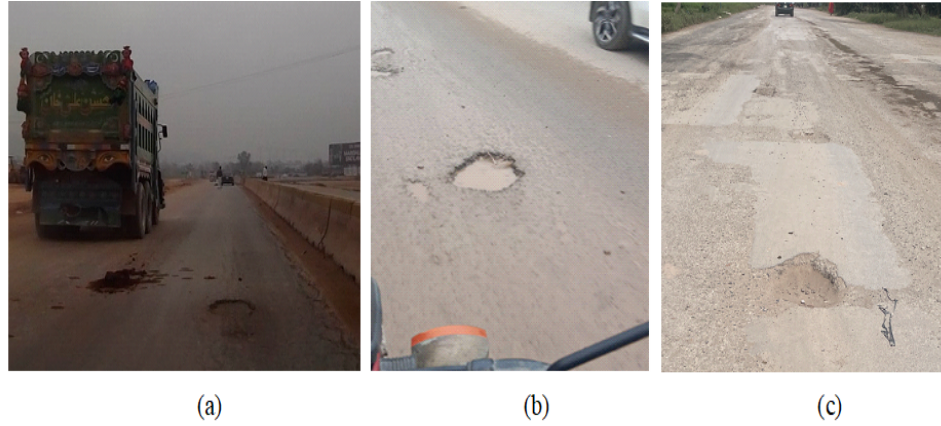


FIGURE 3.4: Potholes images captured using different data acquisition platforms (a) Windshield view. (b) Handlebar view. (c) Dash view.

the different camera mounting positions and recording systems. The motorcycle-mounted camera provides a closer and lower viewing perspective of the road surface, while the car-mounted DVR provides a relatively wider forward view from a higher mounting position. The collected data will ensure the model's ability to detect a wide variety of potholes, including small, early-stage potholes, medium-sized potholes with normal wear and tear, and large potholes with severe hazard levels, as shown in the Figure 3.4. The collected data includes varied condition of textures and lighting condition with inclusion of wet potholes as illustrated in Figure 3.4 (b). This approach ensures the trained model can identify wider range of pothole severities without compromising performance. In addition, the data was acquired at various times of the day to include varying shadow and light conditions. The final prepared dataset was named as PKPOT500 and uploaded to Roboflow public library assessed through link: <https://app.roboflow.com/cracktree/yolov8-seg-8iibs/10>.

3.4.2 Data Annotation and Image Pre-processing

A new project was created on Roboflow to prepare the data for supervised learning. This stage was about translating human knowledge into a format that can be used in the training the model by manually annotating each pothole with the instance segmentation polygonal masks. This approach was preferred over bounding boxes to allow the model to learn the irregular shapes associated with potholes.

All potholes present in the images were annotated and labelled as "pothole". Additionally, images were also tagged with metadata denoting a "sunny" or "rainy" environment so that different lighting and contrast levels could be tested during the validation phase. In order to adopt a Transfer Learning approach during the training process, the labels were saved in the COCO JSON format. Finally, the labelled images were divided into three distinct sets for training, validation and testing with a split ratio of 70%, 20%, and 10% respectively.

TABLE 3.1: Image distribution and pre-processing details

Camera setup	Tag	No. of images	Native Resolution	New Resolution	Image Augmentation	Split (%)
DVR 1080p 40FPS	Dvr	100	1920×1080	800×800	Flip and Brightness	70/20/10
Dashboard Mounted	Dash	100	4032×3024	800×800	Flip	70/20/10
Motor bike handle bar	Bike	150	1280×720	800×800	Brightness	70/20/10

Once the data was labelled, several pre-processing methods were applied to standardise the data set and enhance training. First, an auto-orientation filter was used to strip EXIF data and re-arrange the pixels to correct the horizon in all images. As the images in the dataset were in varied resolutions due to different quality of cameras being used in Car and Motorbike, they were resized to 800 x 800 pixels, a common input size for training the models. This will also offer a good balance between accuracy and fast inference times (FPS). To enhance the model's generalization over varying lighting and weather conditions, data augmentation techniques were applied on training subset. This comprised horizontal flips, the training data set by simulating alternative views of the lane, and brightness adjustments ($\pm 20\%$) to replicate varying illumination especially for darker conditions encountered in real-world scenarios. Image rotation of ± 90 and motion blur $\pm 2\%$ was also applied to simulate minor camera orientation variations and improve model learning for real-world scenarios. These techniques increased the training set by a factor of three, while the validation and test subsets were not augmented to prevent bias in the model testing and evaluation on the road images. To evaluate under a wider range of conditions a second set of images from a public

dataset was added to the test set and manually annotated. Roboflow annotation heat map suggests that maximum number of annotations lies on center third of an image size 800×800 . After augmentation and dataset expansion, the total number of images in the dataset was 920 from 350 images containing 1472 annotations. The overall details of the images collected in various resolutions and image pre-processing applied is outlined in Table 3.1

3.5 Instance Segmentation Architectures and Training Strategy

3.5.1 Single-Stage YOLOv8 Instance Segmentation

The YOLOv8-seg model uses a single-stage design with a backbone, neck, and decoupled prediction head to perform instance segmentation in a single pass. It is a CSP inspired architecture with C2f modules that enhance gradient flow and feature reuse through parallel connections to ensure efficient extraction of hierarchical features. The backbone, as illustrated in Figure 3.5, includes a Spatial Pyramid Pooling Fusion (SPPF) module to extract multi-scale features, while the neck and head are integrated in a unified architecture, which enables a single-stage framework. The neck performs multi-scale feature aggregation using up sampling and concatenation operations, conceptually similar to PANet, allowing the integration of low-level spatial details with high-level semantic features across different resolutions. This improves the identification of minor and irregular potholes. The head uses a decoupled structure with separate branches for classification, bounding box regression, and segmentation, which enhances the learning efficiency and prediction accuracy. Moreover, YOLOv8 implements an anchor-free detection approach, which simplifies bounding box prediction and makes the computation process more efficient. The segmentation branch creates pixel-wise masks of every detected instance, which allow delineating boundaries accurately. This unified architecture facilitates effective end-to-end learning, and therefore YOLOv8-seg can be effectively used to detect potholes efficiently at the instance-level. The Ultra-

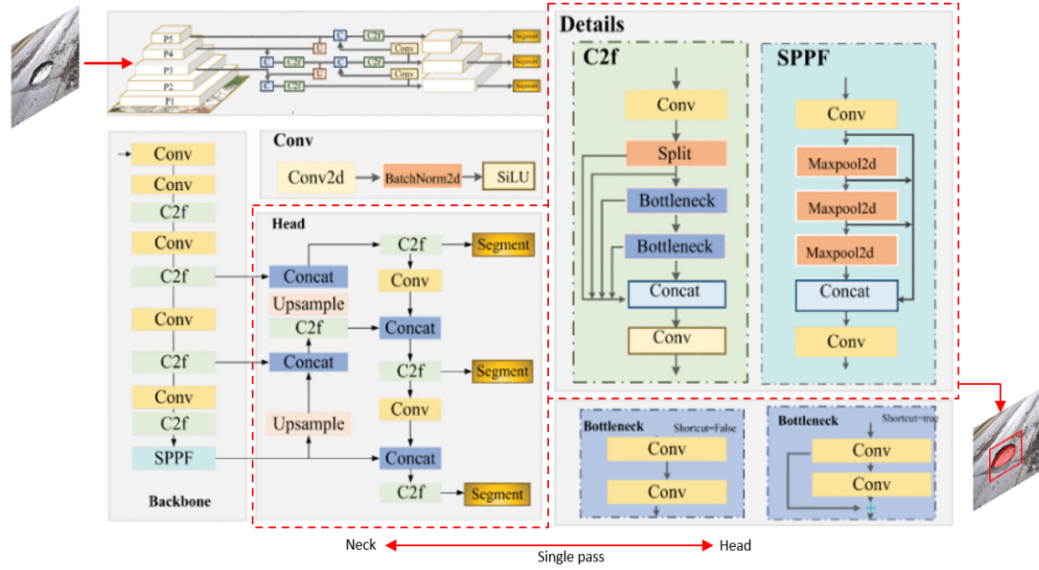


FIGURE 3.5: Instance segmentation architecture of YOLOv8-seg [111]

lytics framework was used to train the YOLOv8-seg model with the small sized YOLOv8s-seg as a starting point to find a balance between the accuracy and computational efficiency. The training was initialized with the custom pothole dataset annotated with polygon masks and exported to YOLOv8 TXT format through Roboflow. Training of the model was conducted on Google Colab using Python v3.12.12 and pytorch 2.9.0, T4 GPU acceleration with a fixed input image size of 800×800 , to ensure consistency with data pre-processing. The training process was defined over 100 epochs, and early stopping was used to end training when the validation performance did not increase any further. Other hyperparameters were kept as default with a learning rate of 0.001 and a weight decay of 0.0005. After training, the model produced optimal weights (best.pt), representing the best performance on validation.

The optimization was run with the default stochastic gradient descent (SGD) setup in Ultralytics, where the batch size varied automatically depending on the amount of memory available on the GPUs. To perform model evaluation, standard metrics such as mean Average Precision (mAP), precision, and recall were computed at each epoch. A confidence threshold of 0.5 was applied during inference to filter low-confidence detections. This approach ensured a balance between detection sensitivity and prediction reliability while maintaining consistency across experiments.

3.5.2 Two-Stage Mask R-CNN Using Detectron2

The instance segmentation model was trained using the Detectron2 framework built on PyTorch, with the Mask R-CNN model architecture with a ResNet-50 backbone. In particular, the `mask_rcnn_R_50_FPN_3x` model from the Detectron2 model zoo was used, which combines a ResNet-50 backbone with a Feature Pyramid Network (FPN) to incorporate multi-scale feature maps. The backbone network performs hierarchical feature extraction using residual bottleneck blocks and the FPN combines low-level and high-level features across the pyramid levels to enable multi-scale pothole detection, accommodating different sizes and shapes of potholes. As demonstrated in Figure 3.6, the Region Proposal Network (RPN) proposes potential object locations using shared convolutional features, followed by RoI Align for accurate spatial alignment between proposals and the feature maps. The RoI heads simultaneously conduct classification, box regression and pixel mask prediction to precisely define instances. Such a design is ideal for detecting and understanding pavement distresses, given the need for precise boundary definition and identification of defects that are closely located. The predefined anchor scales and multi-scale feature extraction strategy improve the detection of small-to-medium sized potholes commonly found in road defects, while preserving the spatial information. To accelerate the convergence and guarantee better

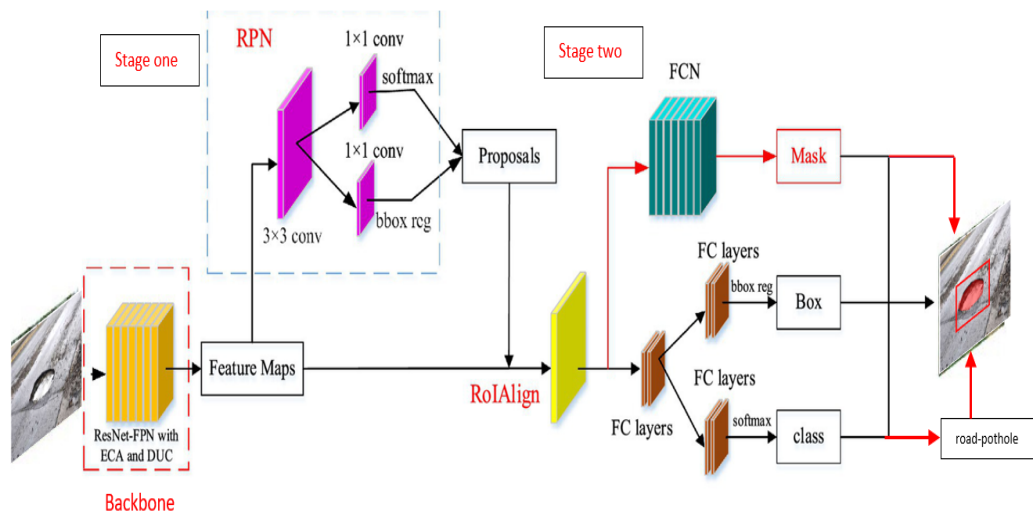


FIGURE 3.6: Instance segmentation of pothole using Mask R-CNN architecture [112]

performance on a custom dataset, the training was initialized with COCO pre-trained weights for transfer learning. The model was trained on Google Colab with NVIDIA T4 GPU. Roboflow was used to export the custom pothole dataset in COCO segmentation format (.JSON) in the Colab folder. The input image size was fixed to 800×800 to be consistent with the dataset and the model. Training was performed for a fixed number of iterations (10,000 iterations) which corresponds to about 20-25 epochs of the dataset size. Base learning rate of 0.00025 was selected for stable fine-tuning and a small batch size of 2 images per iteration due to hardware constraints. Additionally, momentum was set to 0.9 and weight decay to 0.0001 to stabilize training and reduce overfitting. The RoI head sampling rate was raised from 64 to 128 proposals per image to improve region-based learning, and anchor sizes were tweaked to effectively represent pothole features at a smaller scale. Horizontal flipping was also applied beyond Roboflow augmentation to further improve generalization. Training evaluation was done every 500 iterations, using COCO-style metrics, and inference was set at a confidence of 0.5 to achieve a balance between precision and recall. This approach focuses on stability, convergence and performance under small data constraints, to achieve consistent instance segmentation of potholes. Table 3.2 compares key hyperparameters that were set during the training of Mask R-CNN and YOLOv8-seg models, highlighting the simplified default settings in YOLOv8.

3.6 Field Application Framework

The trained model (best.pt) is used in mobile based pothole detection framework using Python (v3.10.11) pipeline under real pavement conditions. The mobile device (iPhone 12) is mounted at the vehicle dashboard with a fixed height of about 1.2 m above the road surface level to record the continuous video stream during the length of survey. The system can operate at a regular speed of 60-80 km/h that makes it effective in gathering data at a network level. To estimate real-world ground dimension, field of view (FOV) is derived from camera standard specifications and calibration. A region of interest (ROI) is applied by cropping

TABLE 3.2: Comparison of training hyperparameters for Mask R-CNN and YOLOv8-Seg model

Component	Hyperparameter	Mask R-CNN	YOLOv8-Seg
Framework	Version	Detectron2 v0.6	Ultralytics v8.3.26
Model	Architecture	Mask R-CNN (R50-FPN)	YOLOv8s-seg
Initialization	Pretrained Weights	COCO (Detectron2 zoo)	YOLOv8-seg
Training	Iterations	MAX_ITER = 8,500	Epochs = 100
	Batch Size	IMS_PER_BATCH = 4	Batch = 8
	Momentum	0.9	0.937 (default)
	Weight decay	0.0005	0.0005 (default)
Optimization	Learning Rate	BASE_LR = 0.0005	Lr0 = 0.01 (default)
	LR Scheduling	STEPS = 6000, 7500	default
	Warm-up	WARMUP_ITERS = 500	Internal
Image Processing	Input Size	MIN_SIZE_TRAIN = 800 MIN_SIZE_TEST = 800	imgsz = 800
Model Specific	RoI Head size	128	Not applicable
	Anchors	[16, 32, 64, 128]	Anchor-free
Augmentation	Training Augmentation	RANDOM_FLIP Horizontal	default
Early Stopping	Patience	Not used	20
Evaluation	Frequency	EVAL_PERIOD = 500	Every epoch
	Evaluation Method	COCO metrics (AP, AR, AP50)	YOLO metrics (mAP, Precision, Recall)
Inference	Confidence Threshold	0.5	0.5
Output	Weights	model_final.pth	best.pt

the upper and lower thirds of each video frame to remove irrelevant regions such as the sky and vehicle hood, retaining the road surface within the central portion of the image. During the survey session, geo-referencing of every pothole is assigned using a mobile GPS device, followed by interpolation to improve positional continuity along the survey path. After the session is stopped, video is transmitted to a remote machine via tunnelling service, where the trained model performs instance segmentation and report result back to mobile device. This cost-effective approach offsets the computation demand from the mobile device and still allows it to perform real time pothole detection.

Pavement condition assessment involves identifying surface distresses and their severity across a specified road section to derive standardized rating. Instead of merely performing pothole identification and geo-location, in this study each segmented pothole in the proposed pipeline is transformed into an equivalent diameter and used to compute a PCI rating (0-100) based on ASTM D6433 severity classification. The scaling factor based on camera projection is performed to convert pixel measurements of a masking region into real-world dimensions, including camera height, field of view, and image resolution. The calculated diameter is then used in accordance with the Table X1.1 of ASTM D6433 to establish the level of severity of each pothole. The existing processing pipeline does not estimate the depth of the potholes hence, a constant depth range of 25-50 mm is assumed to classify the severity. The identified levels of severity are then used to compute deduct values (DV) of each pothole using interpolation of ASTM semi-logarithmic deduct value curves. Finally, DVs for each severity level are then added up to get the final PCI rating of the conducted survey. The end product is a central database containing pothole counts, severity classification and PCI score, and can be visualized using OpenStreetMap.

3.7 Summary

The chapter introduces the methodology which is adopted to detect potholes and assess the condition of pavements using image-based instance segmentation. It also establishes a complete pipeline from pothole detection to quantitative pavement condition assessment by employing segmentation, geometric measurement, and PCI computation for a practical road monitoring system. The study focuses on training of models using custom dataset with the help of the single-stage and two-stage segmentation methods. The localized dataset of potholes was collected under different road, traffic, and environmental conditions, using camera mounted on a car and bike. Image augmentation and preprocessing was applied to prepare the dataset for the training. The use of two different approaches for pothole detection will provide a comparative analysis of model performance in the next chapter.

Chapter 4

Results and Analysis

4.1 Background

The experimental results and comparative analysis of the two deep learning frameworks, YOLOv8s-seg and Mask R-CNN, for automated pothole segmentation are comprehensively presented in this chapter. The evaluation was rigorously conducted using both qualitative and quantitative assessment techniques to systematically investigate segmentation accuracy, detection sensitivity, mask refinement capability, and overall computational efficiency under challenging real-world pavement conditions. The evaluation dataset comprised diverse test images featuring various pothole geometries, irregular surface textures, complex shadow effects, and confounding pavement defects. Furthermore, the quantitative assessment relied on critical performance metrics, including mean Average Precision (mAP), recall, inference speed, GPU memory utilization, total model size, and overall computational complexity. This detailed comparison specifically aims to elucidate the operational trade-offs between high-fidelity segmentation quality and edge-device real-time deployment efficiency across both model architectures. Ultimately, the insights derived from this chapter establish a thorough understanding of the structural strengths and practical limitations of each model, offering valuable guidance

for their implementation in automated pavement inspection and infrastructure condition assessment systems.

4.2 Evaluation Metrics

Segmentation accuracy metrics based on IoU and computational efficiency metrics were used to assess performance of the proposed instance segmentation frameworks. Precision (Equation-4.1) quantifies the number of correct classifications of pothole instances out of the total number of detections while Recall (Equation-4.2) indicates the competence of the model at detecting the real pothole regions in the dataset. True Positive (TP) is the number of pothole instances that are correctly detected, False Positive (FP) is the number of non-pothole regions that are incorrectly detected as a pothole, and False Negative (FN) are actual potholes that are not detected.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4.1 \text{ [73]})$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4.2 \text{ [73]})$$

The evaluation of the overlap between predicted pothole masks and the ground-truth pothole annotations was achieved by calculating the Intersection over Union (IoU) (Equation-4.3) to get a direct estimation of the accuracy of the pothole boundary segmentation. To test the segmentation performance at different IoU thresholds, Average Precision (AP) and mean Average Precision (mAP) (Equation-4.4) were used, with mAP 0.5 representing predictions at a minimum IoU threshold of 50%, and mAP 0.5-0.95 representing a stricter evaluation of predictions IoU ranging from 50% to 90%.

$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}} \quad (4.3 \text{ [73]})$$

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (4.4 \text{ [113]})$$

For the purpose of the confidence threshold analysis, the F1 score (Equation-4.5) was also employed, to evaluate the balance between Precision (PR) and Recall

(R). Additionally, Frames Per Second (FPS) and inference time were included to evaluate computational performance of trained model.

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.5 \text{ [73]})$$

4.3 Instance Segmentation Model Results

4.3.1 Single-Stage YOLOv8 Instance Segmentation

YOLOv8s-seg model showed stable convergence characteristics during training and validation, suggesting that the learning configuration and the composition of the dataset were appropriate to learn generalized pothole representations. As in Figure 4.1, the loss values dropped sharply in the initial epochs, then gradually stabilized in the middle of the training phase with no noticeable oscillations or loss of stability during the training and validation process. The overall training process was terminated at epoch 45, and early stopping was used at epoch 25, suggesting that further training may not yield significant improvements in validation performance values. The lightweight YOLOv8s-seg architecture is suitable for real-time pavement inspection applications, requiring only about 0.45 hours of training time on a Tesla T4 GPU and around 20.5 ms per image of inference time with low GPU memory usage. Finally, the trained model had around 11.8 million learnable parameters and 39.9 GFLOPs to balance segmentation accuracy and complexity. The results evaluated has demonstrated reliable pothole localization and segmentation accuracy at low Intersection over Union (IoU) values. The performance of the segmentation task was evaluated using the mAP metric, with the segmentation mAP of about 0.723 at an IoU of 0.5, (Figure 4.2) and the mAP with an IoU of 0.5-0.95 at about 0.417, showing that the model maintained an acceptable level of segmentation performance with moderate IoU thresholds. The performance of model decreased at higher IoU thresholds is mostly because of the pothole edges and texture were irregular, which made it more difficult to correctly align the pixel-level pothole mask at higher IoU thresholds. The same trend was noticed in the literature survey, where the mAP 0.5 was higher and mAP 0.5-0.95 was

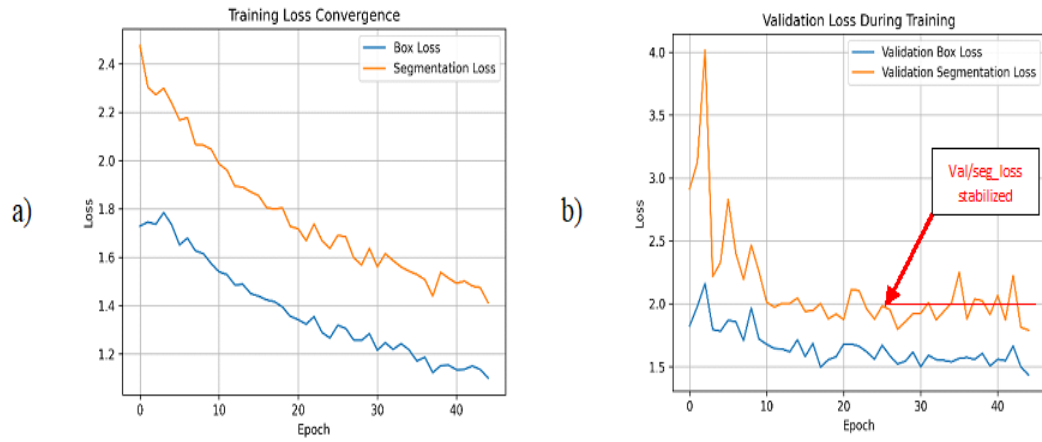


FIGURE 4.1: YOLOv8s-seg a) Training and b) Validation Loss Curves

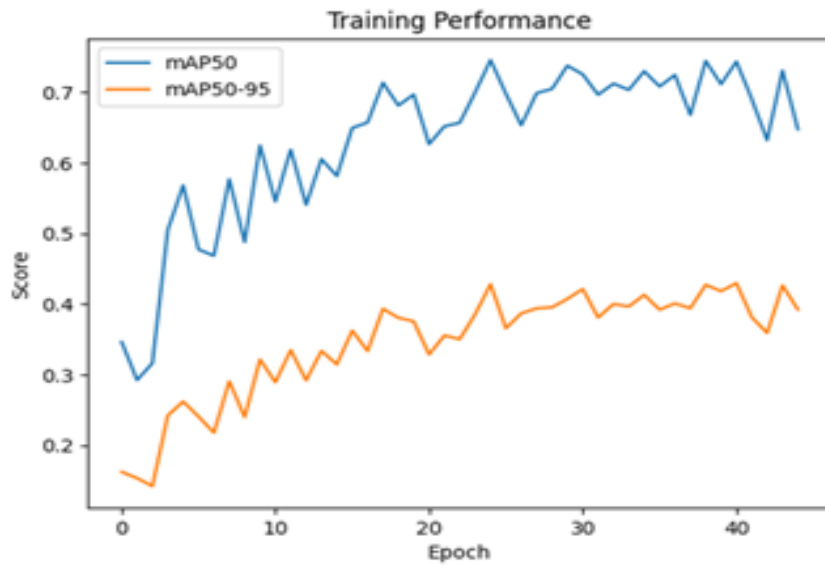


FIGURE 4.2: YOLOv8s-seg Training Performance Curve for mAP0.5 and mAP0.5-0.95

comparatively lower in the case of pothole segmentation. Additionally, the validation precision and recall values were relatively steady at around 68% for both bounding-box and segmentation outputs, thus providing a good balance between pothole detection and avoiding false detections. As shown in the Precision-Recall (PR) curve in Figure 4.3, the model achieved relatively high precision at moderate recall, demonstrating that the feature representations learned from the data were not only discriminative enough to differentiate the pothole regions from the rest of the road texture, but also prevent too many false positive detections. At point A, higher precision was achieved with lower recall, as model being able to detect

only the most confident potholes and fewer false positives. As recall gradually increased between point B and C, the model detected more potholes and increased true positives however, precision gradually decreased due to false detections and boundary inaccuracies. This is potentially caused by shadows, road patches, texture variations, and distant low-resolution potholes. The model started to detect more correct potholes as recall rate increased at point D, but the precision decreased gradually because of the presence of false pothole detections. The best balance between the precision and recall at mAP0.5 was achieved near at point C, 0.723 which is represented as Area Under the Curve (AUC).

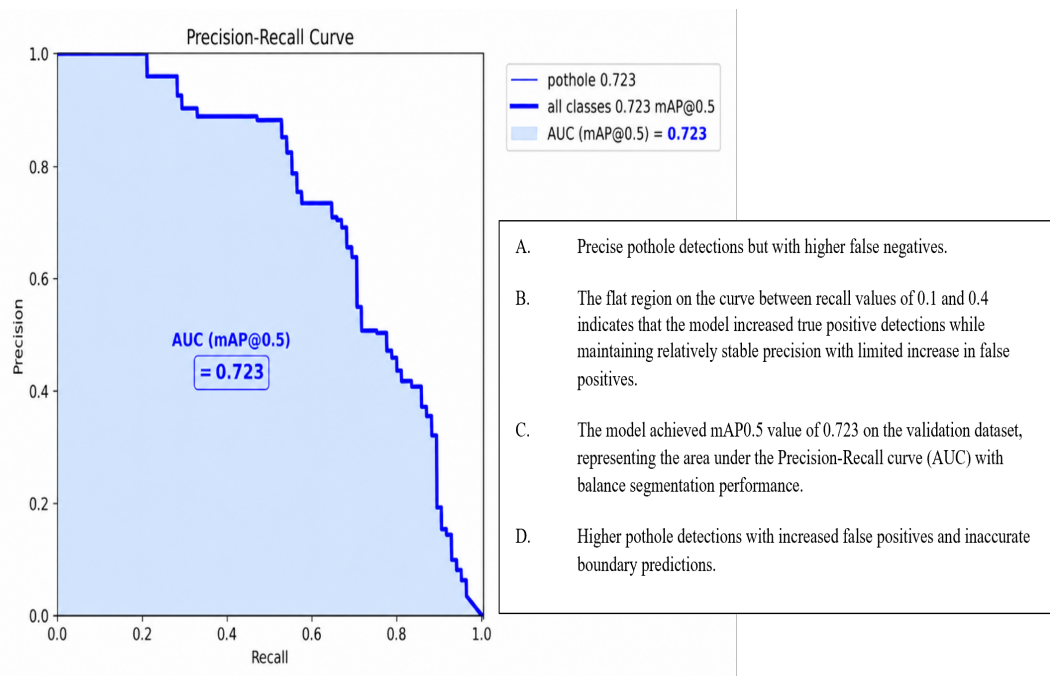


FIGURE 4.3: Precision-Recall curve

The F1 confidence curve in Figure 4.4 illustrates the best precision and recall balance at a confidence threshold of 0.514, with the best F1-score of around 0.68 in region B. This is the optimum point where the model achieves a balance between maximizing the true positive detections and minimizing the false positives. The left side of the graph (region A) represents the low end of the prediction threshold, where recall was still fairly high due to higher number of false positives. While the right region C shows very high precision (low false positives), the recall dropped substantially with the increase in confidence levels because many of the true potholes were rejected.

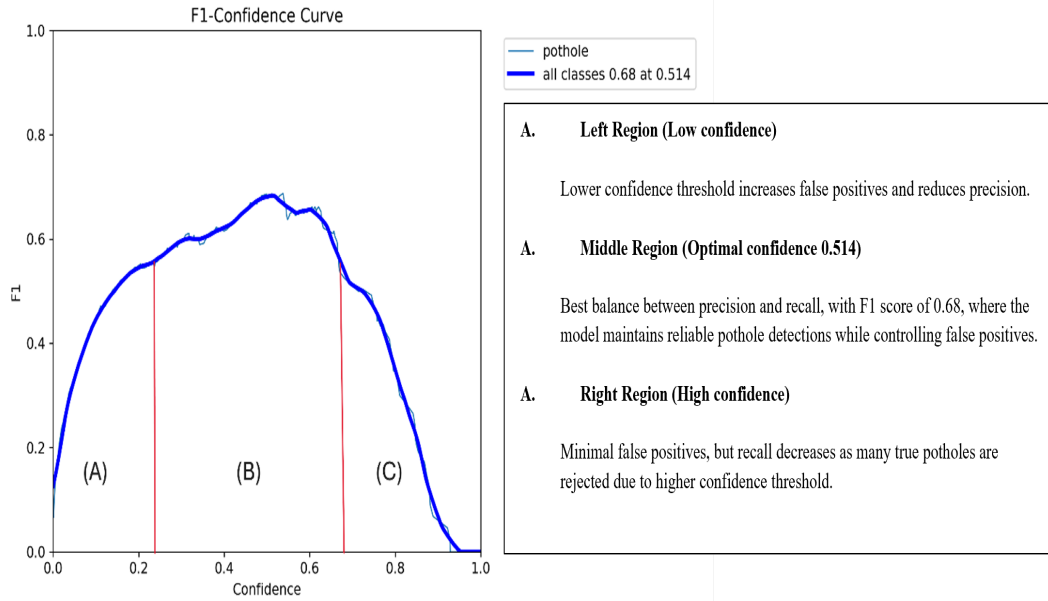


FIGURE 4.4: F1-confidence Curve

The segmentation results also demonstrated the limitation of the proposed framework under different pothole sizes and imaging conditions. The model was able to accurately detect approximately 93% of large potholes ($> 9,216 \text{ px}^2$), compared to 41.5% for medium potholes ($> 1,024 \text{ px}^2$) and 25.6% for small potholes ($< 1,024 \text{ px}^2$). This significant improvement on large potholes was mainly attributed to the visual boundaries being clearer, the ability of the model to represent the pothole surface better, and the presence of more distinguishable discontinuities within the pothole surfaces. In contrast, small potholes often filled small portions of the image, lacked spatial detail in the feature maps, and were mis-segmented, with pixel-wise masks that were not complete and lacked detections. The low-resolution DVR images, the angle of viewing from a distance, low-contrast pavement surfaces, and the irregularity of the roads were also important factors in these challenges, which were visually similar to the pothole regions. The segmentation consistency was influenced by factors like limited size of the dataset, irregular occurrence of potholes, and difficult camera angles however, the trained YOLOv8 segmentation model is still superior in masking potholes regions compared to conventional bounding boxes. The model also performed well with a moderate computational complexity, enabling real-time performance and making it suitable for field deployment.

4.3.2 Two-Stage Mask R-CNN Using Detectron2

Training behaviour of the proposed Mask R-CNN framework was stable throughout the training process, which showed that the hyperparameters and dataset composition were adequate to learn generalized pothole representations. As shown in Figure 4.5, the model not only shows a smooth decay in its error boundaries but also does not show the sign of gradient instability or overfitting. The convergence is smooth with few oscillations, which suggests that the network is converging to an optimized state. The mask head loss remained consistently lower than the total system loss, indicating that the second stage segmentation branch was able to learn the mask structures of potholes well when the region proposal stage was stabilized. As can be seen in the Figure 4.6, the AP50 has increased to 68.51% in the iteration 1900 and stayed at relatively constant value between 60% and 68% for the following iterations, whereas the AP75 has remained relatively lower at 41.33% throughout the training. The reduction between AP50 and AP75 shows that the model localised the pothole areas with a good degree of accuracy at moderate IoU thresholds, but it was still difficult to get the pixels to align correctly with the pothole boundaries because of the irregular geometry of the pothole, varying texture and disjointed edges of the pavement. The AP50 score for the final results on the test set was approximately 0.685 and the average recall was 54.2%, indicating the algorithm's pothole detection was relatively conservative, and that there were fewer false positives. The total training time took around 2.5 hours on an NVIDIA Tesla T4 GPU and the total inference time took around 155.85 ms per image (or 6.42 FPS) with GPU memory utilization reaching approximately 1,113 MB. In addition, the trained model contained approximately 43.92 million learnable parameters with 85.86 GFLOPs, reflecting the computational complexity associated with the two-stage instance segmentation architecture.

The segmentation results also revealed the benefits of the Mask R-CNN framework in different lighting conditions and pothole sizes. The model was able to obtain good AP results of around 74% for large potholes and significantly lower AP results of 32.6% for small potholes, indicating that the segmentation performance is

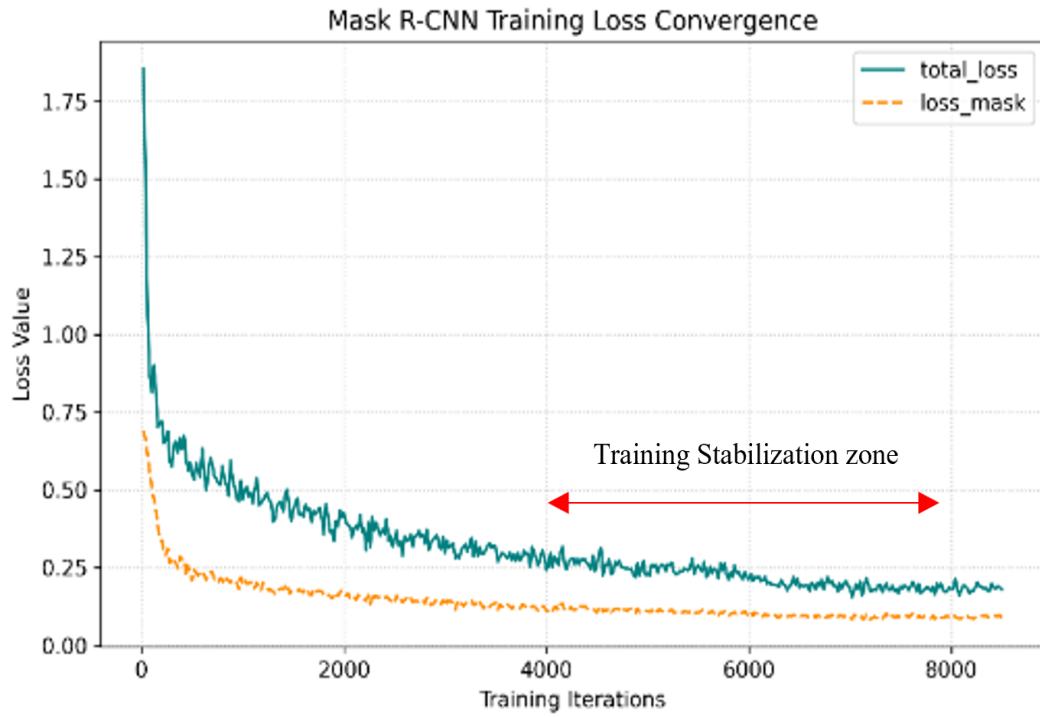


FIGURE 4.5: Mask RCNN training loss convergence profile

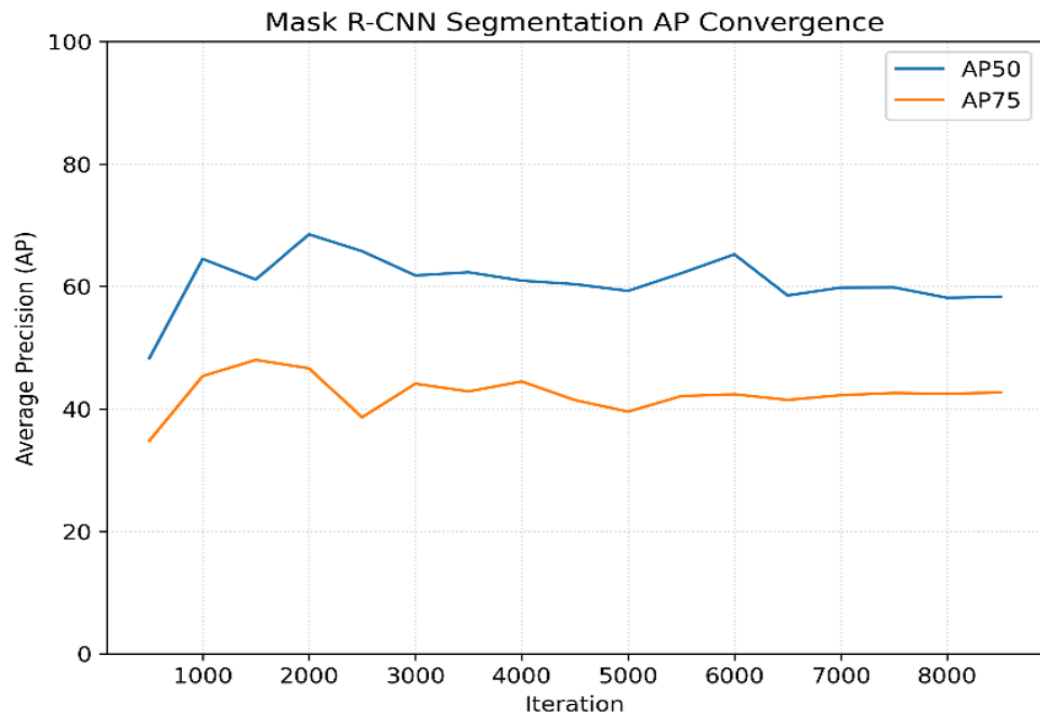


FIGURE 4.6: Mask R-CNN Segmentation AP@0.5 and AP@0.75 Convergence Curves

better with larger potholes. The large potholes had mostly more visual discontinuities and more pixel representation, which enabled the feature pyramid network

and mask refinement stage to keep the pothole boundaries smooth. Some small potholes were however missed partially because of lack of visual detail, poor region proposals, and poor edge visibility due to backbone down sampling operations.

As can be seen in Figure 4.7, the accuracy of the internal mask head improved rapidly with 1200 iterations and surpassed 0.9, reaching a maximum accuracy of 0.96 at the end of the training. This trend suggests that the second-stage segmentation branch could effectively learn pothole masks more effectively as stable region proposals were generated during training. The masks produced were mostly smoother and less fragmented than the single stage segmentation behaviour, especially in medium and large potholes under stable illumination. However, distant and small potholes in combination with low-resolution DVR images still remained challenging because small surface defects occupied limited pixel regions and visually resembled surrounding pavement defects such as patches, ravelling and striping. The two-stage architecture required higher computation cost and slower in inference speed but the overall results showed that Mask R-CNN could obtain reliable and smooth masking boundary, supporting its suitability for offline processing.

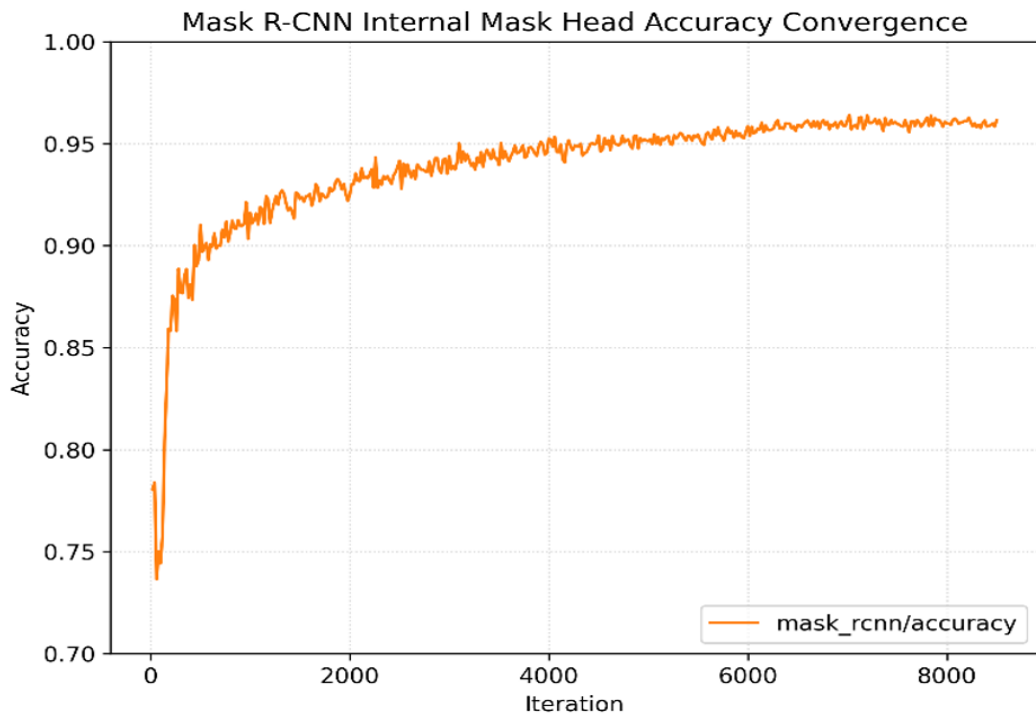


FIGURE 4.7: Mask R-CNN Internal mask head accuracy curve

4.4 Comparative Analysis of YOLO vs Mask R-CNN

4.4.1 Segmentation Performance

A set of test images were used with most difficult conditions including minimal colour variation, road patches, and neighbouring regions of cracks on both trained models. From the results shown in Figure 4.8, it is clear that both

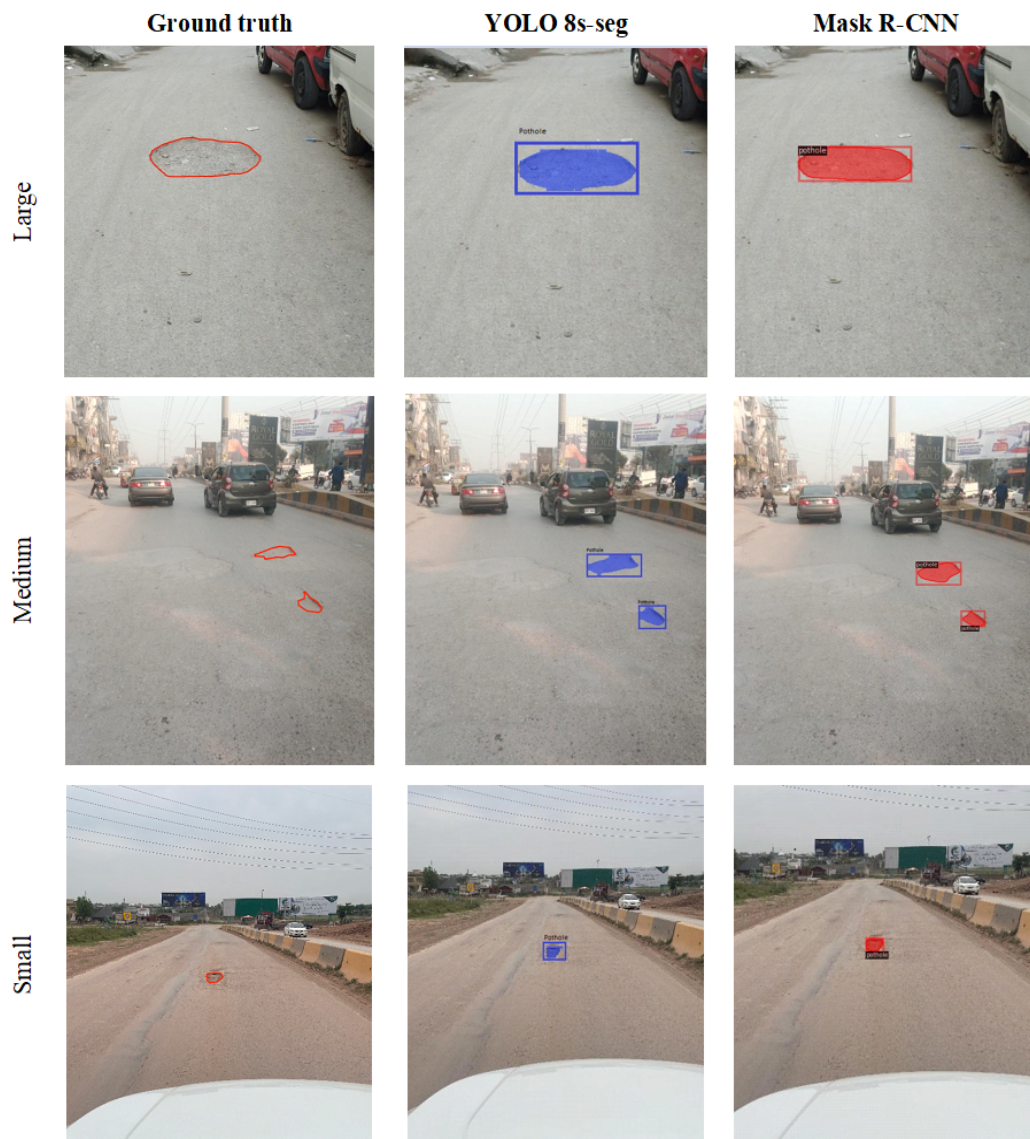


FIGURE 4.8: Segmentation results of small to large sized potholes

frameworks achieved good pothole detection performance for all the pothole sizes (small, medium and large), indicating that the both architectures could learn the

knowledge of pothole detection well. However, Mask R-CNN was able to generate smoother and more accurate mask boundaries that were closer to the ground-truth annotations especially in the case of irregular pothole edges and small surface defects. This action is primarily due to the second stage mask refinement of Mask R-CNN that refines region proposal and preserves their mask boundaries pixel by pixel, which contributes to more accurate mask boundaries. Compared to that, the YOLOv8s-seg model produced relatively rough edges for pothole masks and a certain boundary overflow, especially in cases of medium and large sized potholes. Overall, localization accuracy is similar between both models for clearly visible potholes. The difference in mask boundary refinement can affect the subsequent pothole area estimation, as boundary overflow or under-segmentation changes the number of pixels classified as pothole. This difference is therefore relevant to field evaluation, where the estimated pothole area is used for severity classification and subsequent PCI calculation.

As shown in Table 4.1, the segmentation performance as measured in mAP0.5-0.95 across small, medium and large potholes was consistently better for Mask R-CNN. This is because Mask R-CNN has two-stage mask refinement process which retained the pothole boundary with a higher accuracy under the strict IoU evaluation criterion. Therefore, it achieved 15% higher mAP0.5-0.95 compared to YOLOv8s-seg model on small and large sized potholes. YOLOv8s-seg, however, achieved similar performance for the medium-sized potholes and maintaining significantly better inference speed.

TABLE 4.1: Size Based Segmentation Performance Comparison at mAP 0.95

Object Size	YOLOv8s-seg mAP 0.95	Mask R-CNN mAP 0.95
Small	32.6%	41.9%
Medium	33.5%	34.4%
Large	44.4%	59.5%

Figure 4.9(a) further highlights that these two models are not equally sensitive to the detection of the visually weaker potholes. Mask R-CNN could successfully recognize several small and blurred potholes which were missed by YOLOv8s-seg. However, this higher recall also meant a rise in the number of false positive

detections in areas of the image where the texture is similar to that of a pothole. For example, in Figure 4.9(b), the dense region proposal mechanism and the mask refinement process used by Mask R-CNN led to detect fewer false positives where certain areas of road patches looked similar to potholes. YOLOv8s-seg, on the other hand, exhibited more cautious prediction behaviour with less false positive, but some small potholes were not detected under challenging imaging conditions.

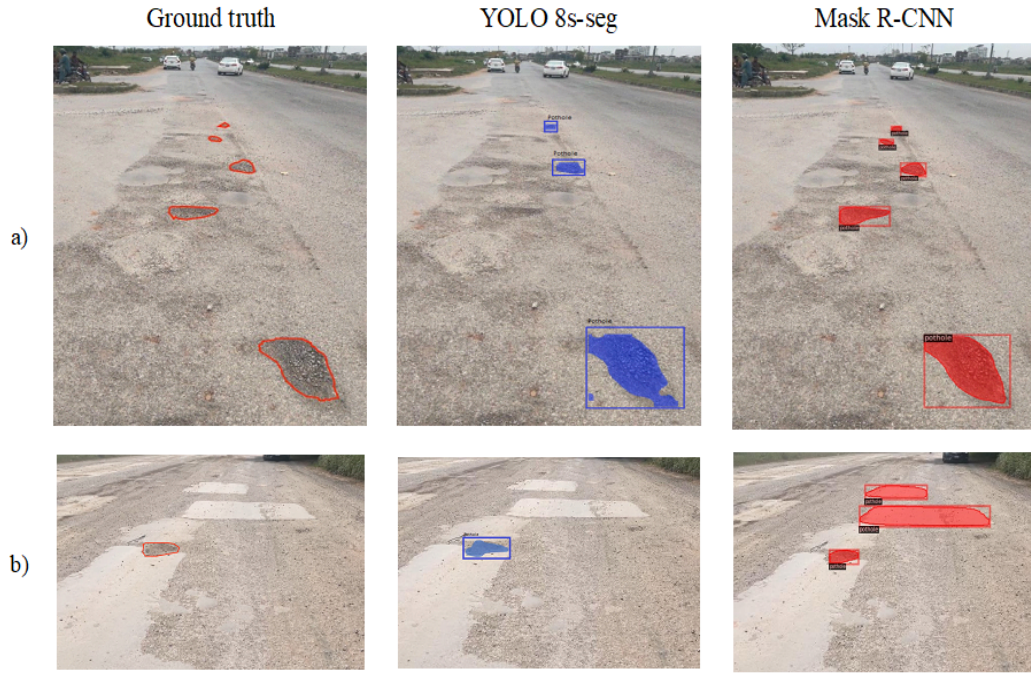


FIGURE 4.9: Over detection by Mask RCNN

Furthermore, the internal mask head accuracy for Mask R-CNN during training was around 96%, but the final segmentation AP was lower because of the conservative region proposal behaviour of the model, which often missed small or ambiguous potholes before the mask refinement stage. However, Mask R-CNN can always generate more refined and clear boundaries for the potholes which shows that it has a better pixel-level learning ability of masks. Table 4.2 demonstrates that YOLOv8s-seg has achieved better results in segmentation at $\text{IoU} = 0.5$ while both frameworks performed relatively similar results under the $\text{mAP}_{0.5-0.95}$ evaluation metric. YOLOv8s-seg, however, yielded the significantly higher recall, showing that Mask R-CNN had a lot of false negatives in the detection process but generated precise segmentation results for the regions where it was able to detect a pothole.

TABLE 4.2: Comparison of segmentation performance on the validation dataset

Model	mAP 0.5	mAP 0.95	Recall
YOLOv8s-seg	72.3%	41.7%	68.2%
Mask R-CNN	68.5%	41.3%	54.2%

4.4.2 Computational Efficiency

Table 4.3 shows a computational comparison of YOLOv8s-seg and Mask R-CNN during the inference ran on the NVIDIA Tesla T4 GPU. Results revealed that YOLOv8s-seg is much more efficient due to its single-stage model that can perform object localization and segmentation at the same time in the same network. This also means YOLOv8s-seg is much faster and lower on GPU memory than Mask R-CNN, saving nearly three times the memory and 7.5 times lesser inference time. Whereas, Mask R-CNN used an additional three stages of sequential region proposal generation, feature refinement, and mask prediction to create the pipeline thus consuming more memory on the GPU. Likewise, YOLOv8s-seg had significantly fewer learnable parameters (11.8M) and fewer GFLOPs (39.9) than Mask R-CNN, suggesting that the network is less complex and requires less computational resources during inference. Mask R-CNN has significantly higher numbers of parameters and GFLOPs, largely stemming from the more complex feature backbone, region proposal network, and second-stage mask refinement operations needed for accurate pixel-level segmentation. Moreover, the much smaller size of the model YOLOv8s-seg shows that it requires lesser storage and deployment demands, thus, fitting the use case of real-time online pavement monitoring applications in resource constrained environments.

4.5 Discussion

This study compared the segmentation performance and computational efficiency of single-stage YOLOv8s-seg framework with the two-stage Mask R-CNN using wide range of pothole images. It can be concluded that the two-stage method is less efficient and more complex in implementation than the single-stage YOLO

TABLE 4.3: Computational Comparison of YOLOv8s-seg and Mask R-CNN

Model	Inference Time (ms)	FPS	Inference GPU Memory (MB)	Parameters (M)	GFLOPs	Model Size
YOLOv8s-seg	20.50	48.80	349.79	11.80	39.90	23.8 MB
Mask R-CNN	155.85	6.42	1,113.16	43.92	85.86	334.8 MB

ms=milli seconds; FPS=frame per seconds; M=Million; GFLOP=Giga Floating Point Operations (Computational workload)

architecture. In summary, YOLOv8s-seg outperformed the Mask R-CNN in terms of recall score and mAP0.5, demonstrating its superior ability to detect potholes in different scenarios such as varying pothole sizes, distant viewing conditions, and surface defects partially visible in potholes. Higher IoU evaluation thresholds were also used for large and irregular pothole segmentation to create a more challenging test set, which revealed that Mask R-CNN exhibited better pixel-level segmentation refinement with smoother and more precise pothole boundaries that were closer to the ground-truth masks. The conservative region proposal behaviour, however, was sometimes insufficient to detect visually weak potholes, causing the overall recall to be reduced while obtaining good internal mask head accuracy. Table 4.4 presents previous work on pothole instance segmentation

TABLE 4.4: Comparison with previous studies

Study	Model	Variant	Images	mAP0.5
S. D et al. [69]	YOLOv8m-seg	Medium	3,074 (mix)	80.5%
Mirajkar et al. [75]	YOLOv8n	Nano	N/A	78.27%
M et al. [114]	YOLOv8	N/A	130	99.2%
Present	YOLOv8s-seg	Small	350	72.3%

based on YOLOv8 precision is reported from 78.27% to 80.5%, and the recall is in the range of 74.4% to 78%. [69, 75]. In addition, the YOLOv8 LiDAR framework reported in [67] also achieved mAP0.5 of 78.5% however the majority of studies mainly emphasized pothole detection and didn't report the segmentation mask or the capability for area estimation. When compared to the previous studies, the

performance of the YOLOv8s-seg model developed in this research was comparatively lower due to the smaller size of the dataset (350 pothole images) collected from different camera mounting positions, distant viewing angles, and small potholes covering limited image pixels. Moreover, the YOLOv8 light weight variant as used in this study has enhanced computation efficiency, while compromising the accuracy.

TABLE 4.5: Characteristics of YOLOv8s-seg and Mask R-CNN for Pothole Instance Segmentation

Feature	YOLOv8s-seg	Mask R-CNN
Detection Framework	Single-stage segmentation framework performing pothole localization and mask prediction simultaneously in one inference stage.	Two-stage segmentation framework that first generates candidate pothole regions before performing mask refinement and segmentation.
Segmentation Behaviour	Achieved higher recall, segmentation accuracy with comparatively coarser mask boundaries.	Produced smoother and more accurate pothole boundaries.
Computational Efficiency	Computationally lightweight with lower GPU memory consumption.	Computationally intensive due to sequential proposal generation and mask refinement stages.
Deployment Suitability	More suitable for real-time pavement monitoring and vehicle-mounted inspection systems.	More suitable for offline pavement analysis requiring refined pothole boundary extraction.

As for computational performance, the single-stage YOLOv8 yielded faster (48.8 FPS) and more precise localization of potholes with relatively stable segmentation performance under moderate IoU levels. This is an especially helpful behaviour for real-life pavement inspection cases where network length surveys could greatly influence computational capabilities of the application. Furthermore, the YOLO model remained significantly faster in terms of inference speed with significantly lower GPU memory usage and fewer learnable parameters, as well as a reduction in GFLOPs. The results are in line with the recent pavement distress studies that showed YOLO-based architecture is more robust and computationally efficient than the two-stage segmentation frameworks. In terms of deployment, the lighter

YOLOv8 architecture is better suited for the inspection survey where camera is mounted in a vehicle and results are computed at the run time.

The experiments also identified some shortcomings in both of the frameworks. Low resolution DVR imagery, shadows, camera angle, and similar pavement textures contributed to segmentation inconsistencies. The performance was further impacted by small potholes that occupied smaller portions as pixels in the image. Furthermore, the dataset size was limited to specific pavement conditions, including wet and dry potholes and crack leading to pothole formation, which may have affected the generalization under unseen road environments. The segmentation and computation properties observed during the experimental evaluation of YOLOv8s-seg and Mask R-CNN are presented in Table 4.5.

4.6 Summary

The experimental results showed a trade-off between segmentation performance and detection efficiency between the two instance segmentation frameworks. For large and irregular pothole and under higher IoU, Mask R-CNN always produced smoother and more refined pothole boundaries which were closer to the ground-truth masks. For the pothole detection coverage, YOLOv8s-seg had the best generalization on different pothole sizes and partially visible pothole defects with mAP_{0.5} of 72.3% and a higher recall value of 68.2%. The size-based evaluation at mAP_{0.5-0.95} has revealed that Mask R-CNN outperforms in segmentation performance for both small and large potholes, whereas YOLOv8s-seg has relatively strong segmentation for medium sized potholes. Moreover, computational analysis revealed that there are significant differences in the efficiency of both frameworks, with YOLOv8s-seg having about 7.5 times less inference time, lower memory needs on the GPU, and smaller model weight than Mask R-CNN. Overall, YOLOv8s-seg appears to offer a more practical use in detection coverage, and computational efficiency for real-time pavement monitoring tasks, while Mask R-CNN is more suitable for offline analysis applications.

Chapter 5

Practical Implementation of Automated Pothole Detection with PCI Evaluation

5.1 Background

This chapter outlines the deployment framework designed for the automated pothole assessment based on the smartphone camera video and GPS integration. The framework utilizes the trained YOLOv8s-seg model to detect potholes through instance segmentation and estimate pothole area, equivalent diameter, and severity classification. The proposed workflow further incorporates geographic localization, distress density evaluation and automated Pavement Condition Index (PCI) evaluation based on the guidelines of ASTM D6433. This automated calculation of Deduct Value, Total Deduct Value, Corrected Deduct Value and PCI rating is achieved by implementing ASTM assessment curves in the processing pipeline. Finally, field validation is carried out to validate the proposed framework by comparing it to manual measurements and traditional calculations to report the percentage error in the calculation.

5.2 Developed Framework

The deployment phase of the proposed framework highlights the automated pavement condition assessment by incorporating the trained YOLOv8s-seg instance segmentation model, GPS recording, and PCI final evaluation. The deployment framework is made up of two key stages: first is to detect the potholes and second downstream processing, which involves area estimation, Deduct Value (DV), and calculation of the Pavement Condition Index (PCI). In the first step, the trained YOLOv8s-seg model from Chapter 4 is loaded in the code block to detect the potholes in real time. The roadway video is continuously recorded from the survey vehicle while GPS coordinates are also logged for every pothole detected. The field survey is performed at a speed of about 60 km/h with video recording at 30 frames per second and GPS data are recorded for each pothole detected.

The post-processing framework is composed of three computational blocks to quantify pavement distress, DV and assess PCI. The segmented pothole regions obtained from instance segmentation are processed by pixel to area conversion method in the first block. The pothole area estimated is then used to estimate the pothole diameter which is needed to calculate the severity level of a pothole. Based on the ASTM severity classification table, detected potholes are categorized into low (L), medium (M), and high (H) severity classes according to their estimated diameter. After the severity classification, the second block determines the deduct value (DV) using the distress density and severity classification of the detected potholes. Lastly, in third block the computed deduct values are fed into the PCI assessment stage where pavement condition rating is assigned for the surveyed area.

The deployment framework has been designed with Python 3.10.11 based on a PyTorch environment and several supporting libraries for real-time pothole detection. The video processing is performed using OpenCV, numerical and geometric computations are performed using NumPy. Tracking and ID assignment of potholes between consecutive frames are performed using ByteTrack. Furthermore, communication between the mobile and remote processing machine is established

using Ngrok, and GPS data of pothole detections are recorded on a map using the OpenStreetMap API.

5.2.1 Pothole detection and GPS Integration

The pothole detection framework is implemented using a sequential processing pipeline integrating video frames, GPS integration, instance segmentation, object tracking, and timestamp synchronization. While inspecting the pavement, video frames and GPS coordinates are logged at the same time using smartphone, where GPS measurements are only assigned together with timestamp information whenever a pothole is detected. Prior to segmentation, region-of-interest (ROI) is applied to limit false positives caused by perspective distortion, roadside objects, sky regions, and large pavement textures. The ROI is defined by excluding the upper and lower one third portions of the frame, thereby retaining the central road region (33% to 66% frame height) where potholes are predominantly visible (Equation-5.1). Pothole detection is then performed using the trained YOLOv8s-seg instance segmentation model, which generates pixel-level masks for individual potholes within each frame. Since the vehicle continuously moves during data acquisition, the same pothole can appear across multiple consecutive frames. Therefore, ByteTrack-based object tracking is integrated into the processing workflow to maintain temporal consistency and suppress duplicate pothole detections by assigning persistent tracking IDs to detected potholes across adjacent frames.

$$\begin{aligned}
 \text{top_crop} &= \lfloor h \times 0.33 \rfloor \\
 \text{bottom_crop} &= \lfloor h \times 0.66 \rfloor \\
 \text{roi_frame} &= \text{frame}[\text{top_crop} : \text{bottom_crop}, :]
 \end{aligned} \tag{5.1}$$

GPS coordinates are synchronized with detected potholes using timestamp-based interpolation between consecutive GPS samples. Video frames are processed at approximately 30 FPS, whereas GPS measurements simultaneously acquired through the smartphone are recorded at approximately 1-second intervals (1 Hz), primarily due to geolocation API (watchPosition) limitation and mobile battery optimization. Therefore, to avoid registering the same GPS coordinate between 0 to 30

frames, intermediate GPS coordinates are estimated to align pothole detections with their corresponding geographic positions using Equation 5.2.

$$\begin{aligned} \text{lat}(t) &= \text{lat}_1 + \frac{t - t_1}{t_2 - t_1}(\text{lat}_2 - \text{lat}_1) \\ \text{lon}(t) &= \text{lon}_1 + \frac{t - t_1}{t_2 - t_1}(\text{lon}_2 - \text{lon}_1) \end{aligned} \quad (5.2)$$

Where lat_1 , lon_1 , and lat_2 , lon_2 represent consecutive GPS coordinates recorded at times t_1 and t_2 respectively. $\text{Lat}(t)$ and $\text{Lag}(t)$ are the interpolated latitude and longitude coordinates at the intermediate frame time t .

If a vehicle travels at a speed of 60 km/h (16.7 m/s) while recording video at 30 FPS, and the GPS sampling rate is approximately 1 Hz, a pothole detected at frame 15 will have a timestamp value of $t = 0.5$ s and an approximate displacement of 8.35 m between GPS samples recorded at start $t_1 = 0$ s and at end $t_2 = 1$ s. As illustrated in Figure 5.1, two GPS coordinates are recorded at the start and end positions, while the pothole is detected between them at frame 15. Therefore, using Equation 5.2, the intermediate GPS coordinate is interpolated and assigned to the detected pothole location. At lower survey speeds, more video frames are captured over the same road distance, which can improve the spatial resolution of pothole detection however, GPS localization accuracy is primarily influenced by GPS sampling and timestamp synchronization rather than vehicle speed alone. For closely spaced potholes, the timestamp-based interpolation assigns a separate GPS coordinate to each pothole according to its respective detection frame, allowing their locations to be distinguished along the vehicle trajectory. However, the positional accuracy remains dependent on the smartphone GPS accuracy and the distance travelled between consecutive GPS samples.

5.2.2 Pothole Area Estimate

Images captured from a smartphone mounted on the vehicle windshield are recorded in a perspective view, where potholes closer to the camera appear larger than those farther away. This means that the width and length of a pothole in image space is not directly proportional to the true width and length of the pothole. Khamkaew

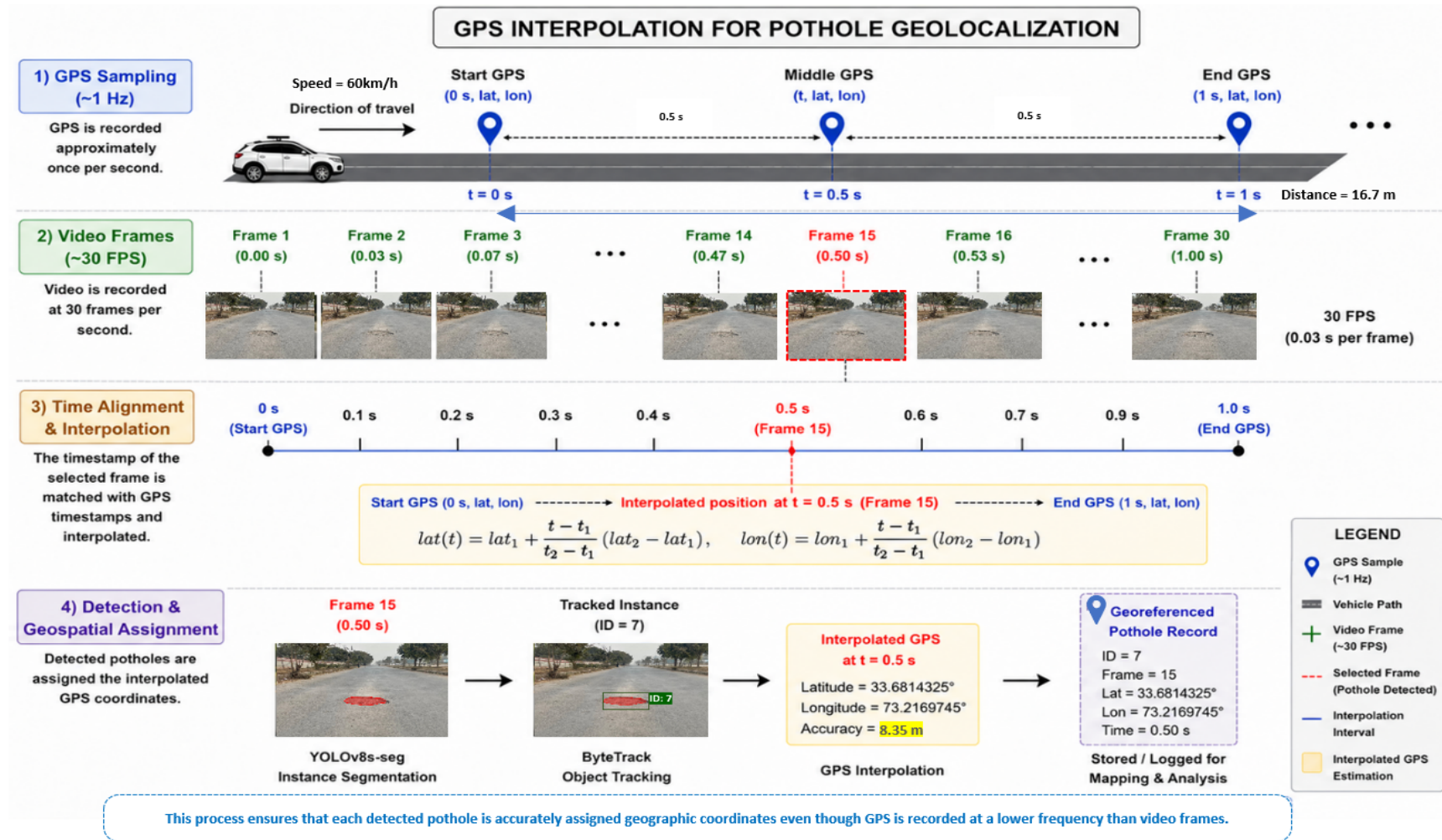


FIGURE 5.1: GPS Synchronization and Object Tracking for Pothole Detection

et al. [5], has overcome this by incorporating the Inverse Perspective Mapping (IPM) and homography transformation techniques to estimate the dimensions of potholes from the transformed bird's eye view (BEV) image. In this work, an empirical scaling calibration approach is used, assuming a fixed camera geometry. Field of view (FOV) and pavement coverage from the recorded images are kept similar between images, while the mounting height of the smartphone camera is kept at an approximately fixed height and viewing angle during the survey. Based on simplified camera geometry assumptions, a vertical FOV of 50° and a height of 1.41 m of a smartphone are used to estimate a vertical pavement coverage of 1.31 m. The relationship between camera geometry, field of view and pavement coverage is conceptually shown in figure 5.2.

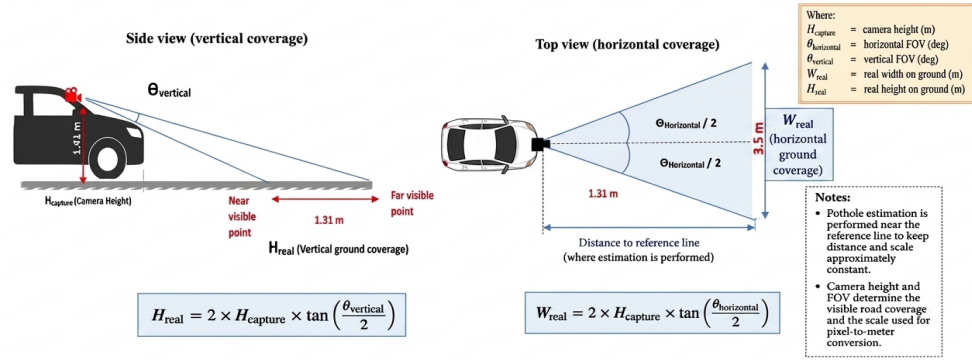


FIGURE 5.2: Camera FOV and Ground Coverage

The size of detected potholes is estimated only when the detected potholes crossed a predefined reference line that is within the central ROI (region of interest) of the frame to minimize perspective distortion and maintain approximately constant scaling conditions. The ROI was designed to exclude regions of upper and lower background, and to keep central pavement region where most pothole defects exist in the dataset. This way, all potholes were estimated to be at roughly the same distance from the camera and in a relatively fixed FOV area.

For calibration, a pothole with known physical dimensions is manually measured in the field and recorded using the same camera setup. The corresponding pothole

region in the image is segmented using the YOLOv8s-seg model, and the total number of pixels area (N) inside the segmentation mask is computed. The measured physical pothole area and segmented pixel area are then used to establish an empirical pixel-to-area conversion factor (S). The segmented pothole area in pixels is converted into physical area measurements using the calibrated scaling factor (A_{eq}). Finally, the calculated physical area is transformed into an equivalent pothole diameter (D_{eq}), as illustrated in Figure 5.3. The estimated diameter will be utilized in next section to evaluate the severity classification of pothole.

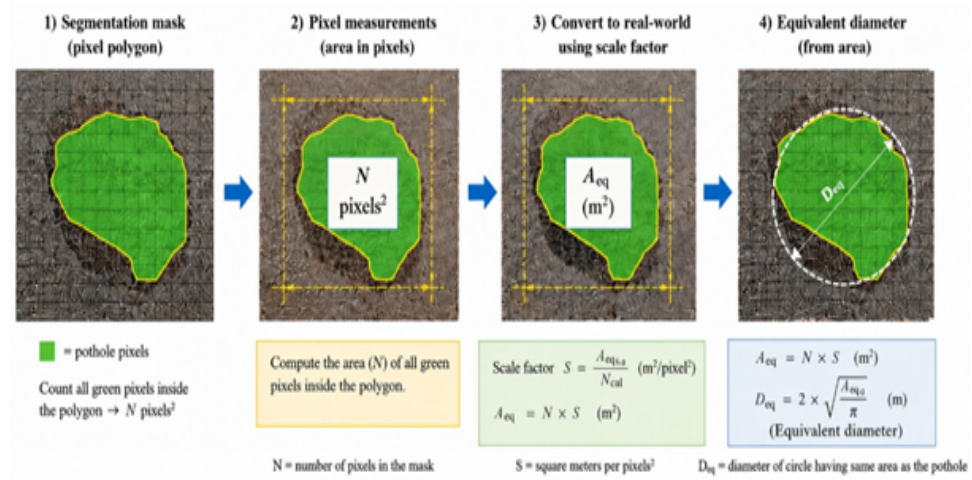


FIGURE 5.3: Pixel polygon to Diameter conversion

5.2.3 Deduct Value

Pothole diameter (D_{eq}), estimated in Section 5.3.2 is used to determine pothole severity according to ASTM D6433-18. Since pothole depth estimation is outside the scope of this study, all potholes are assigned to the ASTM depth range of 25-50 mm. Based on this assumption, each pothole is classified to Low (L), Medium (M), or High (H) severity using the ASTM severity classification criteria presented in Table 5.1.

After severity classification is determined, the distress density is calculated using the distress area divided by total surveyed area. The surveyed road area is estimated using the distance between the start and end GPS coordinates recorded during the survey using `geopy.distance` multiplied by a standard lane width of 3.5

TABLE 5.1: Severity Classification of Potholes using ASTM D6433-18 [110]

Maximum Depth	Equivalent pothole diameter (D_{eq})		
	100–200 mm	200–450 mm	450–750 mm
13 mm to 25 mm	Low	Low	Medium
> 25 mm to 50 mm	Low	Medium	High
> 50 mm	Medium	Medium	High

Adopted from ASTM D6433-18: Table X1.1

m. The total area of potholes belonging to each severity level is divided by the surveyed road area to obtain the distress density percentage, using Equation 5.3.

$$\text{Distress Density}_i = \frac{A_i}{A_{\text{section}}} \times 100 \quad (5.3)$$

Where:

Density_i = Distress density (%) for severity level (L, M, H)

A_i = Total pothole area for Low (L), Medium (M), and High (H) A_{section} = Surveyed Road Area

ASTM D6433 provides separate DV curves L, M, and H severity potholes, where DVs are expressed as a function of distress density. The curves exhibit a non-linear relationship between distress density and deduct value. To enable automated DV computation, 21 Density to DV coordinates (x,y) are extracted from each ASTM curve using WebPlotDigitizer and stored as comma-separated value (CSV) files for each category, while intermediate values are computed using np.interp function. Figure 5.4 illustrates the extraction of density-DV coordinates from the ASTM D6433 and low severity pothole DV curve are exported as a CSV lookup table for automated deduct value computation. The final output will provide the DVL for low, DVM for medium and DVH for high severity potholes.

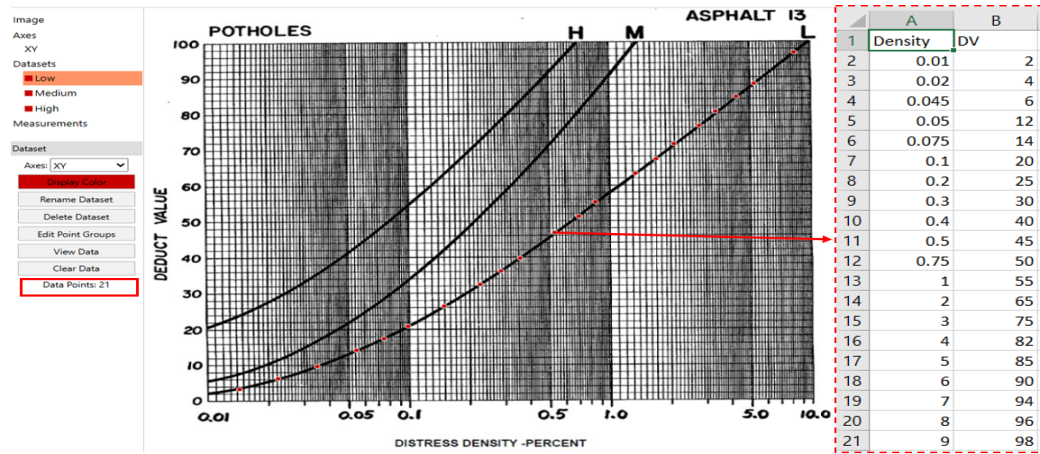


FIGURE 5.4: Extraction of ASTM D6433 [110] DV coordinate points using WebPlotDigitizer

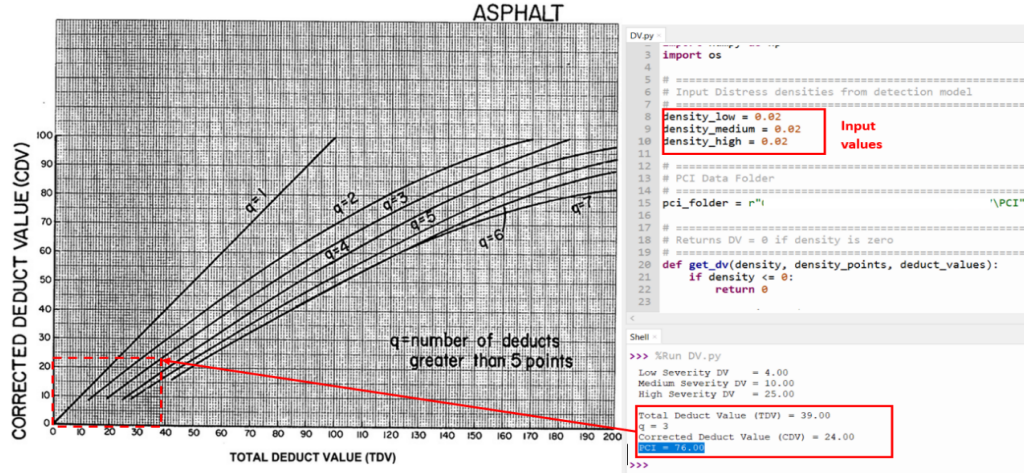
5.2.4 PCI Calculation

The Total Deduct Value (TDV) is obtained by summing all deduct values obtained from each severity classification in Section 5.2.3. The TDV represents the combined impact of all detected potholes within the surveyed pavement section and is calculated using Equation 5.4.

$$TDV = DV_L + DV_M + DV_H \quad (5.4)$$

The Corrected Deduct Value (CDV) is determined using the ASTM D6433-18 correction curves represented by the q1-q7 non-linear curves. Values from the corresponding curves are extracted using *WebPlotDigitizer* at intervals of 10 TDV units and stored in separate lookup tables for q1, q2, and q3. The parameter q is defined as the number of deduct values greater than 2, as specified in ASTM D6433-18. For example, if deduct values are obtained for all three severity levels and each deduct value exceeds 2, then q=3. Similarly, if only two severity levels produce deduct values greater than 2, then q=2. Based on the calculated TDV and the corresponding q value, the CDV is obtained through linear interpolation between adjacent coordinate sets (x,y) in the selected lookup table. Figure 5.5 presents a sample PCI calculation for a distress density of 0.02 in each severity category. The resulting DV produce a TDV of 39.0. As all three deduct values are

greater than 2, the q3 correction curve is selected and calculated a CDV of 19.50 through linear interpolation.



a stretch of road about 1.2 km long and 3.5 m wide. The mobile device was installed on the dashboard of the vehicle at a fixed height in a calibrated camera configuration, shown in Figure 5.6. In the field, vehicle speed was maintained at approximately 60 km/h or below. At this speed, the vehicle travels approximately 0.56 m per video frame, allowing timestamp-based interpolation to provide separate estimated positions for closely spaced detections. During the survey, five potholes (P1-P5) were identified and ground-truth data for comparison was manually collected on site measuring the maximum dimensions of each pothole. All the potholes were correctly identified and the proposed system detected 6 potholes with single false positive detection. The one incorrect false positive was a due to soil stain on the pavement surface that had visual properties like a pothole, indicating more work needs to be done in training data. The GPS coordinates recorded by the application were found to be consistent with the observed pothole locations, and the calculated survey length (1.2 km) closely matched the actual road length with negligible error, suggesting the reliability of the GPS acquisition and interpolation procedure.

Table 5.2 shows the comparison between the field diameters of potholes (D) and equivalent diameters (Deq) estimated by converting from pixel to area conversion factor (N). The results present that there is a small error between the actual measurement and the two potholes (P1, P2), whereas for the remaining potholes (P3-P5), there is a moderate error. One pothole (P5) has an error above 30%, which is believed to be either an error in the segmentation of the pothole boundary or a field measurement error. The larger error of P5 suggests that pothole segmentation is not able to properly detect some large, irregularly shaped potholes, which can lead to under or over estimation of the equivalent diameter (Deq). The average absolute percentage error obtained in this study was 11.2%, which is comparable to the 11.3% error reported by Khamkaew et al. [5] using IPM method for potholes area estimation. Similarly, Faisal & Gargoum, [115] reported an average area estimation error of 7.89% using a low-cost LiDAR-based detection system for potholes when compared with manual measurements. Although the LiDAR approach achieved slightly higher measurement accuracy, it relied on three-dimensional point cloud data, whereas in this study pothole dimensions are

estimated through smartphone video imagery. These comparisons with the previous studies suggest that the proposed method achieves a level of measurement accuracy consistent with previously reported approaches. However, limited studies have reported pothole area estimation using pixel-level segmentation, and therefore, limited direct comparison is available with the proposed smartphone-based method.

TABLE 5.2: Diameter equivalent (D_{eq}) vs. Actual (D)

ID	D_{eq} (mm)	D (mm)	% Error
P1	541	530	+2.2
P2	694	650	+6.8
P3	716	750	-4.5
P4	783	850	-7.9
P5	679	1038	-34.6

These measurement errors were also studied on how they affect the pavement condition assessment using PCI analysis. Table 5.3 compares the manually computed PCI values with those obtained through proposed framework. The automated system yielded distress density of 0.047% compared to manual measurements 0.056%, with a difference of 16.1%. The calculated Deduct Value (DV) using the proposed framework was 37.04, whereas the manual evaluation provided a value of 45.00, highlighting a difference 17.7% difference. This relatively small increase in error (1.6%) between distress density and DV indicates that the automated calculation method can provide accurate results with minor deviation from the manual ASTM D6433 evaluation. Finally, the automated PCI value computed was 62.96 with a difference of 14.5% from manual calculation however, both PCI ratings falling within the same condition category of Fair. This indicates that the proposed framework can report reliable PCI rating despite minor errors in pothole area estimation. Pietersen et al. [116], utilized drone based imagery on airfield and reported a difference of 4.5% between two methods. Although, the PCI difference obtained in the present study is higher, both approaches demonstrate that

automated distress detection and measurement can provide pavement condition ratings consistent with manual assessment methods.

TABLE 5.3: Percentage error Automated PCI vs. Manual

Parameter	Automated Assessment	Manual Assessment	% Difference
Road Area (m ²)	3941	3941	0
Distress Area (m ²)	1.854	2.210	−16.1
Distress Density (%)	0.047	0.056	−16.1
DV	37.04	45.0	−17.7
CDV	37	45	−
PCI	63	55	+14.5
Pavement Rating	Fair	Fair	−

From the user interface, the application home screen presents the camera view with a start recording option along with the current GPS coordinates. As presented in Figure 5.6, once recording starts, the mobile device begins recording video and GPS points simultaneously. After survey completes, the user stops the recording, which initiates the upload process to the remote machine. Ngrok creates a tunnel between the remote machine and the mobile device to facilitate the upload and download process. Figure 5.7 shows the complete end-to-end process displayed on the user's mobile device and back-end processing from a remote machine. Following Steps 1 and 2, once processing of the recorded video reaches 100%, the results are presented in a dashboard reporting pothole severity level, pothole IDs, and GPS coordinates for each detected pothole in tabular form (Figure 5.8). In the final step, the user can navigate to the map view of the survey, where the location of each pothole is displayed with the ability to remove incorrect pothole detections. Figure 5.9 presents the map view of potholes detected during the survey together with a tabular view of pothole locations and the final PCI report within the same interface.

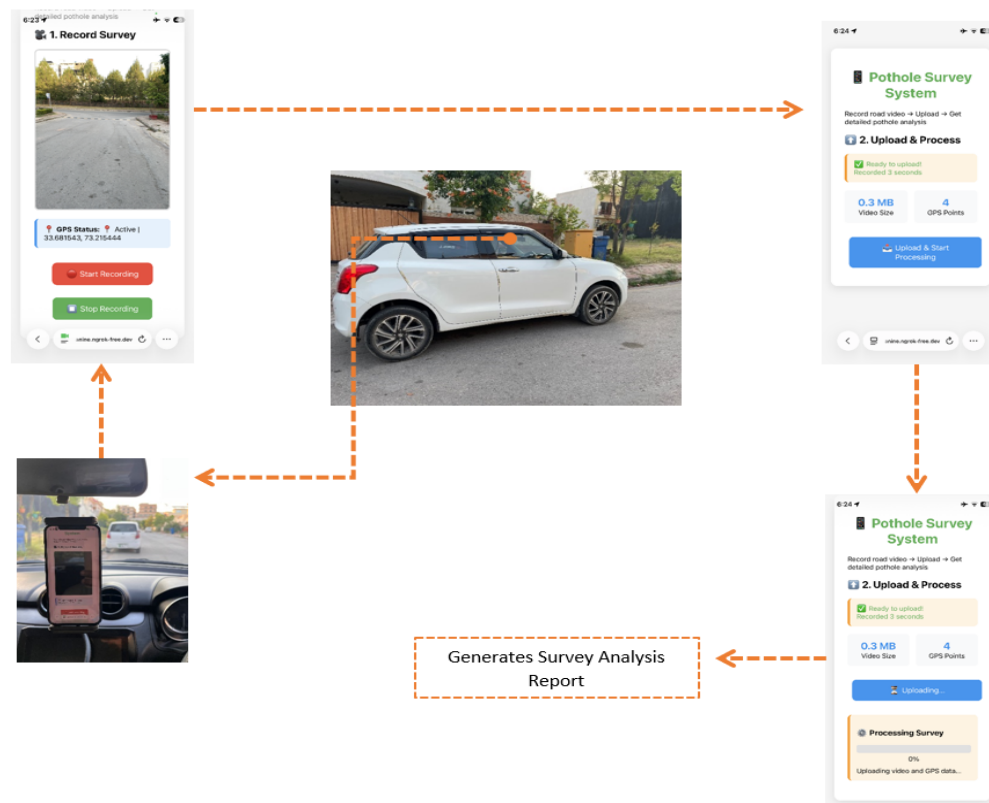


FIGURE 5.6: Illustration of mobile application for Pothole detection

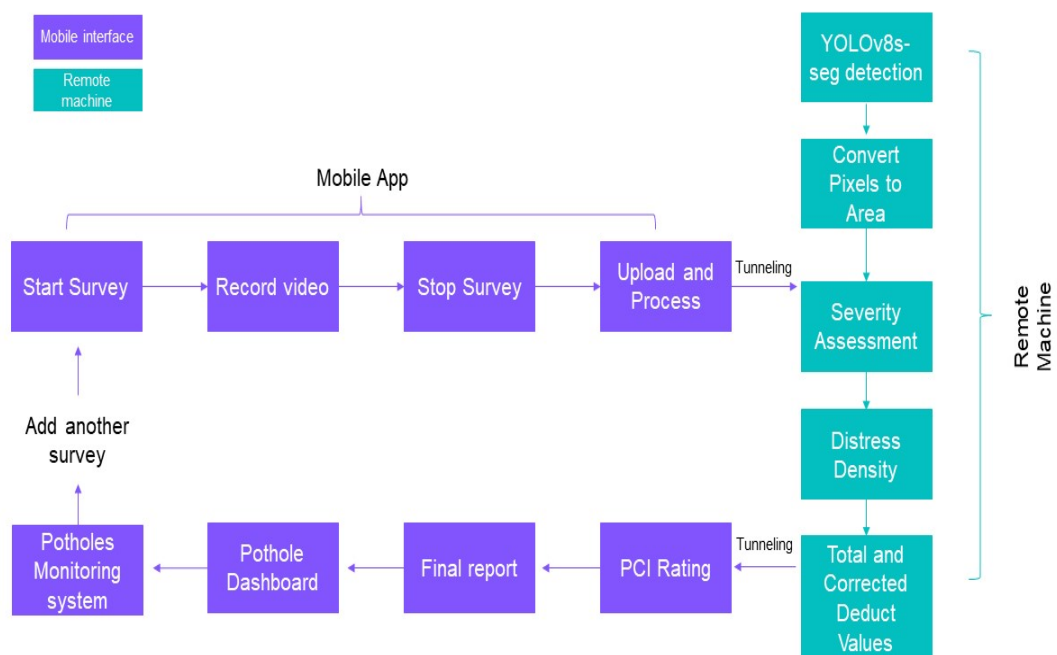


FIGURE 5.7: Flow diagram of end to end framework for mobile application

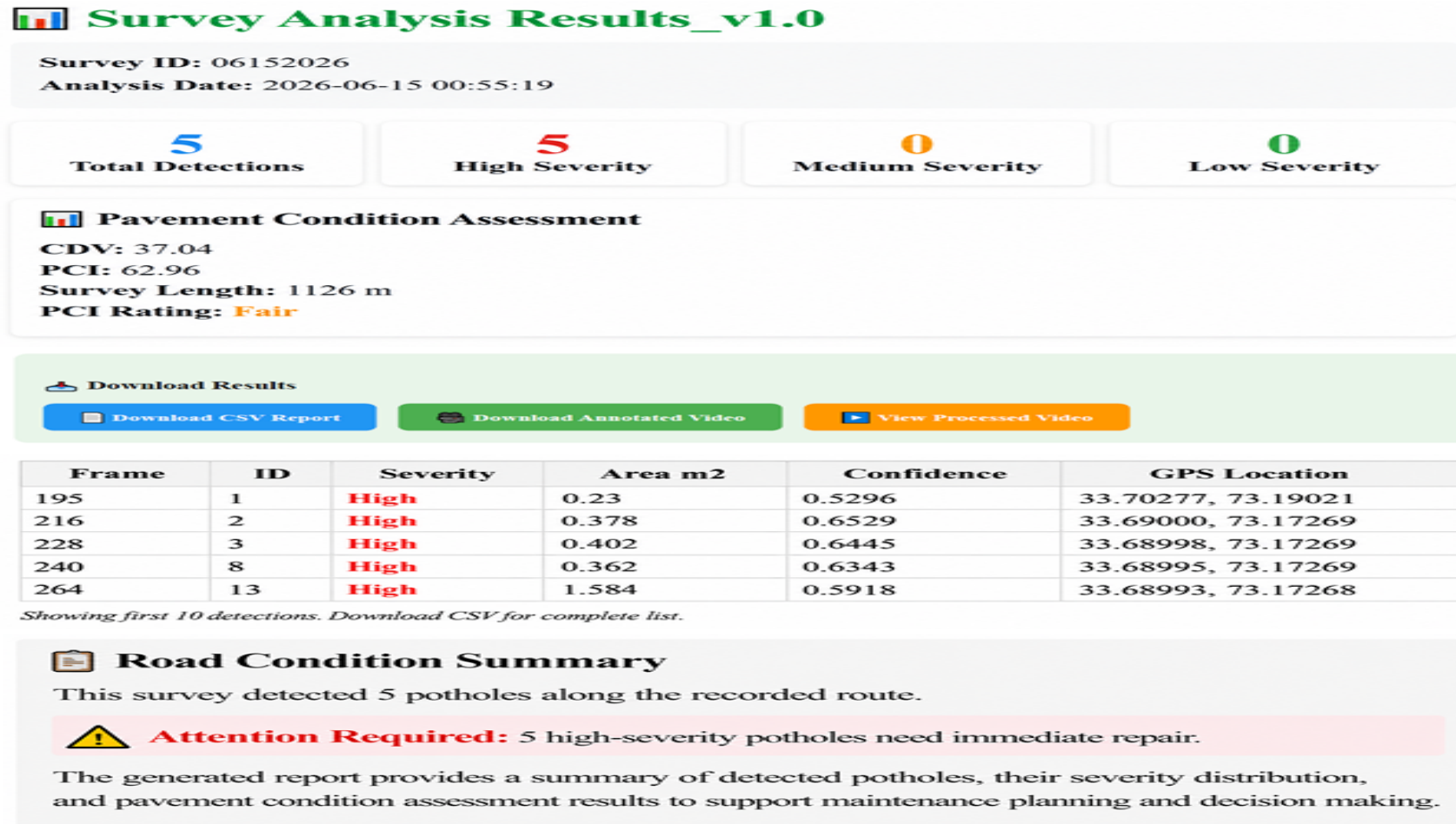


FIGURE 5.8: Survey Analysis report displayed on Mobile

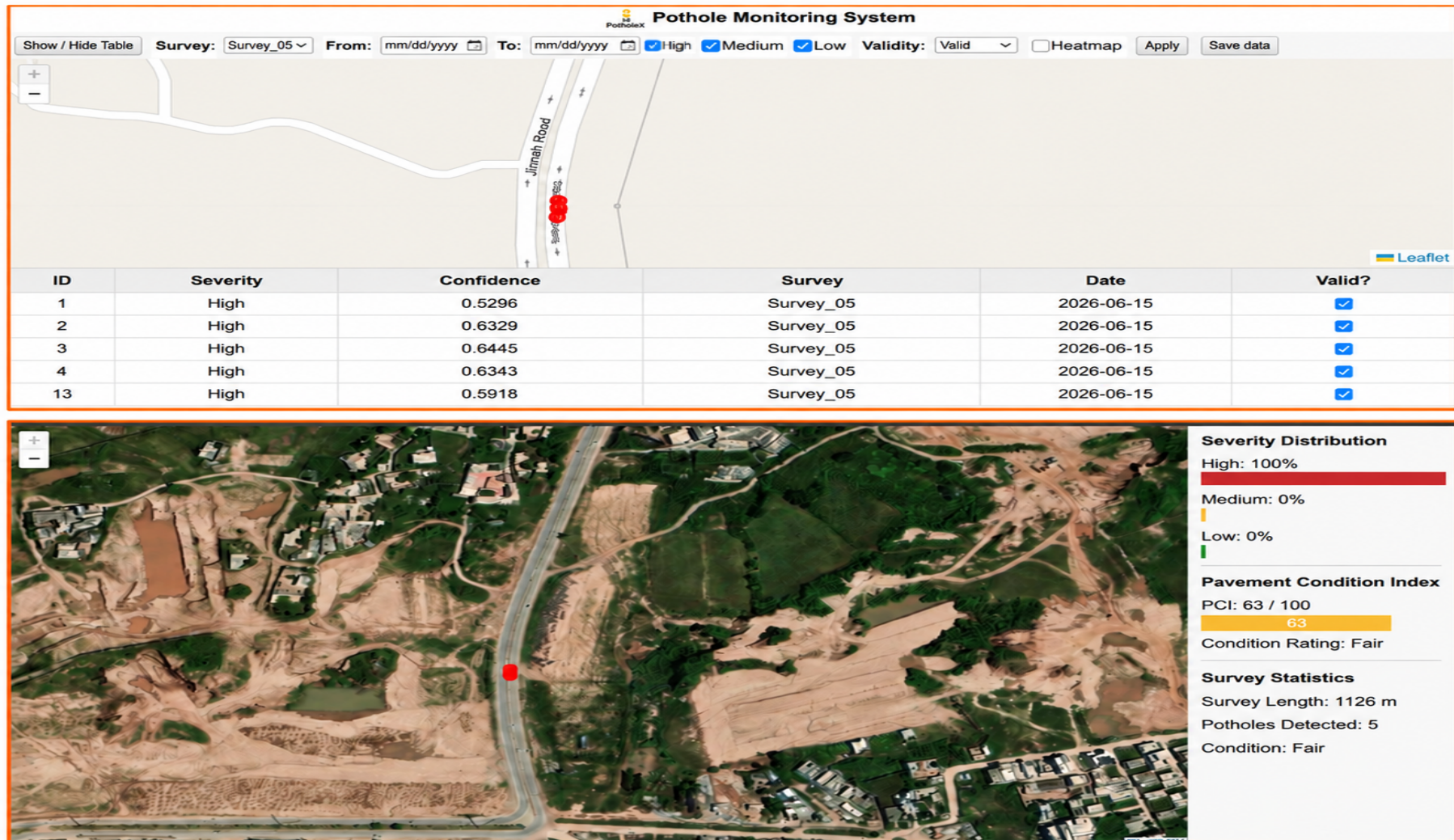


FIGURE 5.9: Map view of detected potholes and PCI reporting

5.4 Summary

This chapter demonstrated the practical deployment of the proposed pavement assessment framework by integrating pothole instance segmentation, GPS integration, and automated PCI evaluation into a single workflow. The developed system has the capability to detect potholes from a smartphone camera, estimate the size of the potholes from the segmentation masks, assign severity level and calculate the pavement condition rating without manual interpretation of the ASTM graph. The difference between the estimated diameter of the pothole and the reported actual diameter was 11.2% as per the field validation. This measurement deviation has not impacted automated distress density, deduct value, and PCI calculations, and reported the same pavement condition rating as manual evaluation. The findings show the proposed system is capable of reliably and practically detecting, quantifying, assigning a severity level to potholes, and determining the pavement condition rating by implementing a standard mobile camera device.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This study used instance segmentation to automatically identify and measure potholes for the creation of a low-cost monitoring system for potholes in mobile applications. To represent different pavement conditions, a customized pothole dataset was developed with three camera mounting configurations, each under different viewing angles, based on the pothole detection performance from the various viewing angles. Two instance segmentation models, YOLOv8s-seg and Mask R-CNN were trained and evaluated, and the accuracy of segmentation and computing cost of the two models were compared. Furthermore, the best performing model was coupled to mobile GPS for mapping the location of the potholes detected. A framework was developed based on ASTM D6433 to estimate the Deduct Value (DV) and Corrected Deduct Value (CDV) in order to facilitate the pavement condition assessment, and to make automatic pavement severity classification based on PCI and report pothole severity for the surveyed pavement sections. From this study, following conclusions are drawn:

- i. Among 350 pothole images, YOLOv8s-seg achieved superior overall segmentation performance with mAP 0.5 of 72.3% compared to 68.5% achieved by Mask R-CNN, indicating stronger pothole detection coverage across varying pavement conditions.

- ii. In terms of computational efficiency, the lightweight architecture of YOLOv8 showed significantly better performance with an inference speed of about 7.5 times faster and 69% lesser memory usage compared to Mask R-CNN, making it suitable for mobile application systems for pavement monitoring.
- iii. In the quality of the pixel-level segmentation, Mask R-CNN achieved higher mAP 0.5-0.95 scores for small pothole (41.9%) and large pothole (59.5%) compared to YOLOv8s-seg, showing superior performance for higher IoU thresholds.
- iv. Both frameworks performed best on large potholes, with lower performance on medium and small potholes, indicating that size and visible surface area of the pothole are important factors for the segmentation accuracy.
- v. The average deviation of the proposed approach of pixel to area conversion from the field measurement was 11.2%, which shows the suitability of this method for field use in the classification of severity and PCI.

6.2 Future Recommendations

Following recommendations may be considered for further enhancing the capabilities of AI based pothole detection application.

- i. Expanding the pothole dataset to include more diverse pavement conditions and improving annotation boundary precision to enhance model generalization and segmentation accuracy across varying road and environmental conditions.
- ii. More in-depth analysis of the effects of different viewing angles from the car DVR and motorcycle-mounted camera and low lighting condition.
- iii. Incorporating pothole depth estimation from Machine learning RGB imagery to eliminate the current dependence on assumed depth ranges during severity assessment.

-
- iv. Implementing advanced perceptive transformation techniques by flattening the pavement image to estimate pothole area.
 - v. Improve current PCI classification block by applying polynomial regression model to provide more accurate DV and CDV.
 - vi. Validate the framework on larger road networks and under diverse lighting, traffic, and pavement conditions to improve generalization capability.

Bibliography

- [1] C. H. Yang, J. Kim, and S. Shin, “Road hazard assessment using pothole and traffic data in south korea,” *Journal of Advanced Transportation*, vol. 2021, pp. 1–10, 2021.
- [2] K. Sheeraz, N. A. Memon, A. H. Memon, N. A. Memon, and S. F. Jafri, “Discovering blackspots caused by pavement distresses using lagrange polynomial and safety inspections of traffic accidents: A case study for karachi, pakistan,” *Mehran University Research Journal Of Engineering and Technology*, vol. 42, no. 4, pp. 123–139, 2023.
- [3] G. Kombe, “A sustainable approach for road pothole repair using waste plastic bottles and aggregates,” *Frontiers in Engineering and Built Environment*, vol. 5, 2024.
- [4] X. Zhigang, C. Yanli, and L. Jinlong, “Research progress on automatic processing technology for road surface damage images,” *Journal of Transportation Engineering*, vol. 19, no. 01, pp. 172–190, 2019.
- [5] N. Khamkaew, K. Jitpaired, D. Rinchumphu, Y. Nonjui, T. Me-ngoen, N. Khansila, and B. Thueakchaikham, “Pothole detection and dimension estimation via image transformation and scaling with thai road data integration,” *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. X-4/W7-2025, pp. 75–81, 2025.

- [6] M. H. Asad, S. Khaliq, M. Yousaf, M. Ullah, and A. Ahmad, "Pothole detection using deep learning: A real-time and ai-on-the-edge perspective," *Advances in Civil Engineering*, 2022.
- [7] X. Chen, C. Liu, L. Chen, X. Zhu, Y. Zhang, and C. Wang, "A pavement crack detection and evaluation framework for a uav inspection system based on deep learning," *Applied Sciences*, vol. 14, no. 3, p. 1157, 2024.
- [8] S. S. Reddy, M. Janarthanan, I. U. Khan, and K. Amrutha, "A deep learning framework for real-time pothole detection from combined drone imagery and custom dataset using enhanced yolov8 and custom feature extraction," *Mathematics*, vol. 14, no. 5, p. 898, 2026.
- [9] F. Deng, J. Jin, X. Chen, Z. An, and Y. Du, "Enhancing automated health monitoring of road infrastructure through a hierarchical robust deep learning approach," *Structural Health Monitoring*, vol. 25, no. 1, pp. 37–59, 2026.
- [10] Q. Cesar and G. Surhid, "Road infrastructure and economic development : some diagnostic indicators," Policy Research working papers ; no. WPS 921. Transport Washington, D.C. : World Bank Group., Report, 2010.
- [11] C. P. Ng, T. H. Law, F. M. Jakarni, and S. Kulanthayan, "Road infrastructure development and economic growth," *IOP Conference Series: Materials Science and Engineering*, vol. 512, no. 1, p. 012045, 2019.
- [12] M. F. Abdelkader, M. A. Hedeya, E. Samir, A. A. El-Sharkawy, R. F. Abdel-Kader, A. Moussa, and E. El-Sayed, "Egy-pdd: a comprehensive multi-sensor benchmark dataset for accurate pavement distress detection and classification," *Multimedia Tools and Applications*, vol. 84, no. 31, pp. 38 509–38 544, 2025.
- [13] F. Administration and Highway, "Policy and governmental affairs," Report, 2015.

- [14] A. Abed, M. Rahman, N. Thom, D. Hargreaves, L. Li, and G. Airey, “Analysis and prediction of pothole formation rate using spatial density measurements and pavement condition indicators,” *Transportation Research Record*, vol. 2677, pp. 651–664, 2023.
- [15] M. A. Afridi, S. Erlingsson, and L. Sjgren, “Municipal street maintenance challenges and management practices in sweden,” vol. 9, 2023.
- [16] News Desk, “Pothole-hit roads leave 100 dead, over 1,000 injured in karachi in two months,” *Pakistan Today*, April 14 2026, accessed: June 23, 2026.
- [17] N. Otsu, “A threshold selection method from gray-level histograms,” *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.
- [18] D. Mahler, Z. Kharoufa, E. Wong, and L. Shaw, “Pavement distress analysis using image processing techniques,” *Computer-aided Civil and Infrastructure Engineering*, vol. 6, pp. 1–14, 1991.
- [19] Y. Zhu, “Automatic pavement crack identification system research based on digital image processing,” *Xi. an: Chang. an University*, 2008.
- [20] A. B. Alayat and H. A. Omar, “Pavement surface distress detection using digital image processing techniques,” *Jurnal Kejuruteraan*, 2023.
- [21] I. Abbas and M. Ismael, “Automated pavement distress detection using image processing techniques,” *Engineering, Technology & Applied Science Research*, 2021.
- [22] W. Qureshi, S. I. Hassan, S. McKeever, D. Power, B. Mulry, K. Feighan, and D. O’Sullivan, “An exploration of recent intelligent image analysis techniques for visual pavement surface condition assessment,” *Sensors (Basel, Switzerland)*, vol. 22, 2022.

-
- [23] D. Mahler, Z. Kharoufa, E. Wong, and L. Shaw, "Pavement distress analysis using image processing techniques," *Computer-aided Civil and Infrastructure Engineering*, vol. 6, pp. 1–14, 1991.
- [24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.
- [25] A. B. Alayat and H. A. Omar, "Pavement surface distress detection using digital image processing techniques," *Jurnal Kejuruteraan*, 2023.
- [26] Z. Lv, Z. B. Hao, Y. Zhu, and C. Lu, "A review on automated detection and identification algorithms for highway pavement distress," *Applied Sciences*, vol. 15, no. 11, p. 6112, 2025.
- [27] B. Al-Mistarehi, A. Shtayat, R. Imam, and A. Abdallah, "An automated assessment technique for pavement defects using a laser scanner and deep machine learning," *Civil Engineering Journal*, 2025.
- [28] S. Ding, G. Yang, K. Wang, E. Yang, and A. Zhan, "3d imaging based pavement texture evaluation," *International Journal of Pavement Engineering*, vol. 24, 2023.
- [29] D. Zhang, Q. Zou, H. Lin, X. Xu, L. He, R. Gui, and Q. Li, "Automatic pavement defect detection using 3d laser profiling technology," *Automation in Construction*, vol. 96, pp. 350–365, 2018.
- [30] K. C. P. Wang, L. Li, W. Luo, and A. Larkin, "Potential measurement of pavement surface texture based on three-dimensional image data," *Conference Proceedings*.

- [31] B. Li, K. Wang, A. Zhang, Y. Fei, and G. Sollazzo, “Automatic segmentation and enhancement of pavement cracks based on 3d pavement images,” *Journal of Advanced Transportation*, 2019.
- [32] Z. Liu, W. Wu, X. Gu, and B. Cui, “Pavedistress: A comprehensive dataset of pavement distresses detection,” *Data in Brief*, vol. 57, 2024.
- [33] B. Li, K. Wang, A. Zhang, Y. Fei, and G. Sollazzo, “Automatic segmentation and enhancement of pavement cracks based on 3d pavement images,” *Journal of Advanced Transportation*, 2019.
- [34] X. Zhigang, C. Yanli, and L. Jinlong, “Research progress on automatic processing technology for road surface damage images,” *Journal of Transportation Engineering*, vol. 19, no. 01, pp. 172–190, 2019.
- [35] I. Feitosa, B. Santos, and P. Almeida, “Pavement inspection in transport infrastructures using unmanned aerial vehicles (uavs),” *Sustainability*, 2024.
- [36] J. Jing, X. Yang, H. Cheng, X. Feng, H. Zheng, I. Brilakis, and J. Liu, “A critical review on pavement distress detection using images and point clouds from visual features to geometric modeling,” *Journal of Road Engineering*, vol. 5, no. 4, pp. 583–606, 2025.
- [37] E. Romero-Chambi, S. Villarroel-Quezada, E. Atencio, and F. Muoz-La Rivera, “Analysis of optimal flight parameters of unmanned aerial vehicles (uavs) for detecting potholes in pavements,” *Applied Sciences*, vol. 10, no. 12, p. 4157, 2020.
- [38] F. Samadzadegan, D. Javan, A. Mahini, M. Gholamshahi, and F. Nex, “Automatic road pavement distress recognition using deep learning networks from unmanned aerial imagery,” *Drones*, 2024.

-
- [39] J. He, L. Gong, C. Xu, P. Wang, Y. Zhang, O. Zheng, G. Su, Y. Yang, J. Hu, and Y. Sun, "Highrpd: A high-altitude drone dataset of road pavement distress," *Data in Brief*, vol. 59, p. 111377, 2025.
- [40] J. Zhu, J. Zhong, T. Ma, X. Huang, W. Zhang, and Y. Zhou, "Pavement distress detection using convolutional neural networks with images captured via uav," *Automation in Construction*, vol. 133, p. 103991, 2022.
- [41] S. Lee, S. Kim, and S. Moon, "Development of a car-free street mapping model using an integrated system with unmanned aerial vehicles, aerial mapping cameras, and a deep learning algorithm," *Journal of computing in civil engineering*, vol. 36, no. 3, p. 04022003, 2022.
- [42] E. Romero-Chambi, S. Villarroel-Quezada, E. Atencio, and F. Muoz-La Rivera, "Analysis of optimal flight parameters of unmanned aerial vehicles (uavs) for detecting potholes in pavements," *Applied Sciences*, vol. 10, no. 12, p. 4157, 2020.
- [43] F. Samadzadegan, D. Javan, A. Mahini, M. Gholamshahi, and F. Nex, "Automatic road pavement distress recognition using deep learning networks from unmanned aerial imagery," *Drones*, 2024.
- [44] P. Hu, "The digital image identification system of highway pavement disease application research," *Chongqing Jiao tong University Chongqing, China*, 2013.
- [45] N.-D. Hoang, "An artificial intelligence method for asphalt pavement pothole detection using least squares support vector machine and neural network with steerable filter-based feature extraction," *Advances in Civil Engineering*, 2018.

- [46] C. Wu, Z. Wang, S. Hu, J. Lpine, X. Na, D. Ainalis, and M. Stettler, “An automated machine-learning approach for road pothole detection using smart-phone sensor data,” *Sensors (Basel, Switzerland)*, vol. 20, 2020.
- [47] D. Li, Z. Duan, X. Hu, D. Zhang, and Y. Zhang, “Automated classification and detection of multiply pavement distress images based on deep learning,” *Journal of Traffic and Transportation Engineering (English Edition)*, 2023.
- [48] N.-D. Hoang, “An artificial intelligence method for asphalt pavement pot-hole detection using least squares support vector machine and neural network with steerable filter-based feature extraction,” *Advances in Civil Engineering*, 2018.
- [49] H. S. Munawar, A. W. A. Hammad, A. Haddad, C. A. P. Soares, and S. T. Waller, “Image-based crack detection methods: A review,” *Infrastructures*, vol. 6, no. 8, p. 115, 2021.
- [50] Y. Safyari, M. MahdianPari, and H. Shiri, “A review of vision-based pothole detection methods using computer vision and machine learning,” *Sensors (Basel, Switzerland)*, vol. 24, 2024.
- [51] M. Gao, X. Wang, S. Zhu, and P. Guan, “Detection and segmentation of cement concrete pavement pothole based on image processing technology,” *Mathematical Problems in Engineering*, vol. 2020, pp. 1–13, 2020.
- [52] K. Ragab, “Smart pothole detection using deep learning based on dilated convolution,” *Sensors (Basel, Switzerland)*, vol. 21, 2021.
- [53] M. Asad, S. Khaliq, M. H. Yousaf, M. Ullah, and A. Ahmad, “Pothole detection using deep learning: A real-time and ai-on-the-edge perspective,” *Advances in Civil Engineering*, vol. 2022, 2022.

- [54] A. Sekar and V. Perumal, “Automatic road crack detection and classification using multi-tasking faster rcnn,” *Journal of Intelligent & Fuzzy Systems*, vol. 41, no. 6, pp. 6615–6628, 2021.
- [55] S. Zhang, Z. Liu, K. Wang, W. Huang, and P. Li, “Obc-yolov8: an improved road damage detection model based on yolov8,” *PeerJ Computer Science*, vol. 11, 2025.
- [56] N. Bhavana, M. M. Kodabagi, B. M. Kumar, P. Ajay, N. Muthukumaran, and A. Ahilan, “Pot-yolo: Real-time road potholes detection using edge segmentation-based yolo v8 network,” *IEEE Sensors Journal*, vol. 24, no. 15, pp. 24 802–24 809, 2024.
- [57] J. Terven, D.-M. Crdova-Esparza, and J.-A. Romero-Gonzalez, “A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas,” *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, 2023.
- [58] R. Roberts, G. Giancontieri, L. Inzerillo, and G. Di Mino, “Towards low-cost pavement condition health monitoring and analysis using deep learning,” *Applied Sciences*, vol. 10, no. 1, p. 319, 2020.
- [59] C. P. Yabi, G. F. Gbehoun, B. C. K. Tamou, E. Alamou, M. Gibigaye, and E. N. Farsangi, “Analysis and classification of distress on flexible pavements using convolutional neural networks: A case study in benin republic,” *Infrastructures*, vol. 10, no. 5, p. 111, 2025.
- [60] F. Andika and Y. Bandung, “Road damage classification using ssd mobilenet with image enhancement,” in *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, Conference Proceedings, pp. 540–545.

- [61] A. He, A. A. Zhang, X. Xu, Y. Ding, H. Zhang, and Z. Dong, “Automated instance segmentation of asphalt pavement patches based on deep learning,” *Structural Health Monitoring*, vol. 24, no. 1, pp. 602–621, 2025.
- [62] H. Wang, J. Kang, S. Zhang, Y. Xiang, and J. Li, “A customized multi-class pavement distress segmentation method for routine repair monitoring,” *Engineering Applications of Artificial Intelligence*, vol. 154, p. 111002, 2025.
- [63] V. Pereira, S. Tamura, S. Hayamizu, and H. Fukai, “Semantic segmentation of paved road and pothole image using u-net architecture,” in *2019 International Conference of Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, Conference Proceedings, pp. 1–4.
- [64] F. Li, Y. Mou, Z. Zhang, Q. Liu, and S. Jeschke, “A novel model for the pavement distress segmentation based on multi-level attention deeplabv3+,” *Engineering Applications of Artificial Intelligence*, vol. 137, p. 109175, 2024.
- [65] C. Han, T. Ma, J. Huyan, X. Huang, and Y. Zhang, “Crackw-net: A novel pavement crack image segmentation convolutional neural network,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 22 135–22 144, 2021.
- [66] J. Zhong, M. Zhang, Y. Ma, R. Xiao, G. Cheng, and B. Huang, “A multitask fusion network for region-level and pixel-level pavement distress detection,” *Journal of Transportation Engineering, Part B: Pavements*, vol. 150, no. 1, p. 04024002, 2024.
- [67] D. Wang, Z. Liu, X. Gu, and W. Wu, “Feature extraction and segmentation of pavement distress using an improved hybrid task cascade network,” *International Journal of Pavement Engineering*, vol. 24, no. 1, p. 2266098, 2023.

- [68] L. Li, J. Liu, J. Xing, Z. Liu, K. Lin, and B. Du, "Road pothole detection based on crowdsourced data and extended mask r-cnn," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 9, pp. 12 504–12 516, 2024.
- [69] R. S. S. D, A. J. S, S. S, and T. V. S. K, "Pothole detection and instance segmentation using yolo v8," in *2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*, Conference Proceedings, pp. 1185–1190.
- [70] F. Deng, J. Jin, X. Chen, Z. An, and Y. Du, "Enhancing automated health monitoring of road infrastructure through a hierarchical robust deep learning approach," *Structural Health Monitoring*, vol. 25, no. 1, pp. 37–59, 2026.
- [71] J. Zhong, J. Zhu, J. Huyan, T. Ma, and W. Zhang, "Multi-scale feature fusion network for pixel-level pavement distress detection," *Automation in Construction*, vol. 141, p. 104436, 2022.
- [72] R. Fan, H. Wang, Y. Wang, M. Liu, and I. Pitas, "Graph attention layer evolves semantic segmentation for road pothole detection: A benchmark and algorithms," *IEEE Transactions on Image Processing*, vol. 30, pp. 8144–8154, 2021.
- [73] M. Roman-Garay, H. Rodriguez-Rangel, C. B. Hernandez-Beltran, P. Lepej, J. E. Arreygue-Rocha, and L. A. Morales-Rosales, "Architecture for pavement pothole evaluation using deep learning, machine vision, and fuzzy logic," *Case Studies in Construction Materials*, vol. 22, p. e04440, 2025.
- [74] B. Kulambayev, M. Nurlybek, G. Astaubayeva, G. Tleuberdieva, S. Zholdasbayev, and A. Tolep, "Real-time road surface damage detection framework based on mask r-cnn model," *International Journal of Advanced Computer Science and Applications*, 2023.

- [75] R. Mirajkar, A. Yenikar, S. Nawalkar, R. Kaul, A. Rokade, and K. Rothe, “Enhanced pothole detection in road condition assessment using yolov8,” in *2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE)*, Conference Proceedings, pp. 429–433.
- [76] E. Ibragimov, Y. Kim, J. H. Lee, J. Cho, and J.-J. Lee, “Automated pavement condition index assessment with deep learning and image analysis: An end-to-end approach,” *Sensors*, vol. 24, no. 7, p. 2333, 2024.
- [77] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Conference Proceedings, pp. 779–788.
- [78] J. Redmon and A. Farhadi, “Yolo9000: better, faster, stronger,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Conference Proceedings, pp. 7263–7271.
- [79] F. Redmon, Joseph and Ali, “Yolov3: An incremental improvement,” *arXiv preprint arXiv:1804.02767*, 2018.
- [80] G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics YOLOv8,” version 8.0.0, 2023.
- [81] 2026.
- [82] Y. Du, N. Pan, Z. Xu, F. Deng, Y. Shen, and H. Kang, “Pavement distress detection and classification based on yolo network,” *International Journal of Pavement Engineering*, vol. 22, pp. 1659–1672, 2020.
- [83] S. Sarange, P. Manmat, P. Ghate, and S. Kale, “Evaluation of yolov8 and ssd object detection models for pothole detection systems,” *Journal of Information Systems Engineering and Management*, 2025.

- [84] G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics YOLOv8,” version 8.0.0, 2023.
- [85] M. Khan, M. A. Raza, G. Abbas, S. Othmen, A. Yousef, and T. Jumani, “Pothole detection for autonomous vehicles using deep learning: a robust and efficient solution,” *Frontiers in Built Environment*, 2024.
- [86] P. Tang, M. Lv, Z. Ding, W. Xu, and M. Jiang, “Pothole detection-you only look once: Deformable convolution based road pothole detection,” *IET Image Process.*, vol. 19, 2024.
- [87] M. Aquib, M. Faisal Hussain, E. RamaLakshmi, and D. T. Prathima, “Yolo models for accurate pothole detection: A comparative study,” in *Journal of Physics: Conference Series*, vol. 3191. IOP Publishing, Conference Proceedings, p. 012002.
- [88] R. Bai, M. Wang, Z. Zhang, J. Lu, and F. Shen, “Automated construction site monitoring based on improved yolov8-seg instance segmentation algorithm,” *IEEE Access*, vol. 11, pp. 139 082–139 096, 2023.
- [89] M. Aquib, M. Faisal Hussain, E. RamaLakshmi, and D. T. Prathima, “Yolo models for accurate pothole detection: A comparative study,” in *Journal of Physics: Conference Series*, vol. 3191. IOP Publishing, 2024, Conference Proceedings, p. 012002.
- [90] K. M, G. G, S. B. T, S. Sk, J. Y, and M. Agoramoorthy, “A novel deep learning approach for enhanced roadway pothole detection using yolov8 instance segmentation algorithms,” *International Journal of Computational and Experimental Science and Engineering*, 2025.
- [91] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2961–2969, 2017.

- [92] A. K. Prajapati and I. Guler, “A generative approach to generalize deep learning models for pavement distress segmentation,” *Data Science for Transportation*, vol. 7, 2025.
- [93] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, “Rdd2022: A multinational image dataset for automatic road damage detection,” *Geoscience Data Journal*, vol. 11, no. 4, pp. 846–862, 2024.
- [94] D. Arya, H. Maeda, Y. Sekimoto, H. Omata, S. K. Ghosh, D. Toshniwal, M. Sharma, V. V. Pham, J. Zhong, and M. Al-Hammadi, “Rdd2022-the multi-national road damage dataset released through crddc’2022,” (*No Title*), 2022.
- [95] H. Yan and J. Zhang, “Uav-pdd2023: A benchmark dataset for pavement distress detection based on uav images,” *Data in Brief*, vol. 51, p. 109692, 2023.
- [96] Z. Xu, Z. Sun, J. Huyan, W. Li, and F. Wang, “Pixel-level pavement crack detection using enhanced high-resolution semantic network,” *International Journal of Pavement Engineering*, vol. 23, no. 14, pp. 4943–4957, 2022.
- [97] W. Song, Z. Zhang, B. Zhang, G. Jia, H. Zhu, and J. Zhang, “Istd-pds7: A benchmark dataset for multi-type pavement distress segmentation from ccd images in complex scenarios,” *Remote. Sens.*, vol. 15, p. 1750, 2023.
- [98] M. Rezaeimanesh, A. Golroo, M. S. Fahmani, M. J. Amani, F. Hasanitabaar, M. S. Entezari, and S. Karimi, “Attain: Inclusive annotated pavement distress types and severity dataset,” *Data in Brief*, vol. 61, p. 111715, 2025.
- [99] D. Arya, H. Maeda, Y. Sekimoto, H. Omata, S. K. Ghosh, D. Toshniwal, M. Sharma, V. V. Pham, J. Zhong, and M. Al-Hammadi, “Rdd2022-the multi-national road damage dataset released through crddc’2022,” (*No Title*), 2022.

- [100] M. Rezaeimanesh, A. Golroo, M. S. Fahmani, M. J. Amani, F. Hasanitabaar, M. S. Entezari, and S. Karimi, “Attain: Inclusive annotated pavement distress types and severity dataset,” *Data in Brief*, vol. 61, p. 111715, 2025.
- [101] D. Arya, H. Maeda, S. K. g. Ghosh, D. Toshniwal, and Y. Sekimoto, “Rdd2022: A multinational image dataset for automatic road damage detection,” *Geoscience Data Journal*, vol. 11, no. 4, pp. 846–862, 2024.
- [102] W. S. Qureshi, S. I. Hassan, S. McKeever, D. Power, B. Mulry, K. Feighan, and D. OSullivan, “An exploration of recent intelligent image analysis techniques for visual pavement surface condition assessment,” *Sensors*, vol. 22, no. 22, p. 9019, 2022.
- [103] N. S. P. Peraka and K. P. Biligiri, “Pavement asset management systems and technologies: A review,” *Automation in Construction*, vol. 119, p. 103336, 2020.
- [104] D. Walker, L. Entine, and S. Kumer, “Paser-manual asphalt roads,” *Wisconsin Transportation Information Center, College of Engineering, University of Wisconsin, Madison, WI*, 2002.
- [105] A. International, *Standard practice for roads and parking lots pavement condition index surveys*. ASTM International, 2018.
- [106] J. McCarthy, L. Fitzgerald, J. McLaughlin, B. Mulry, D. Obrien, and K. Dowling, “Rural flexible roads manual pavement surface condition index,” *Department of Transport, Tourism and Sports: Dublin, Ireland*, vol. 4, 2014.
- [107] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyaama, and H. Omata, “Road damage detection and classification using deep neural networks with smart-phone images,” *ComputerAided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1127–1141, 2018.

- [108] W. S. Qureshi, D. Power, I. Ullah, B. Mulry, K. Feighan, S. McKeever, and D. O’Sullivan, “Deep learning framework for intelligent pavement condition rating: A direct classification approach for regional and local roads,” *Automation in Construction*, vol. 153, p. 104945, 2023.
- [109] T. Rateke and A. Von Wangenheim, “Road surface detection and differentiation considering surface damages,” *Autonomous Robots*, vol. 45, no. 2, pp. 299–312, 2021.
- [110] A. International, *Standard practice for roads and parking lots pavement condition index surveys*. ASTM International, 2018.
- [111] R. Bai, M. Wang, Z. Zhang, J. Lu, and F. Shen, “Automated construction site monitoring based on improved yolov8-seg instance segmentation algorithm,” *IEEE Access*, vol. 11, pp. 139 082–139 096, 2023.
- [112] L. Shen, J. Su, R. Huang, W. Quan, Y. Song, Y. Fang, and B. Su, “Fusing attention mechanism with mask r-cnn for instance segmentation of grape cluster in the field,” *Frontiers in Plant Science*, vol. Volume 13 - 2022, 2022.
- [113] r. etinkaya, E. D. atmabacak, and E. ztrk, “Detection of fractured endodontic instruments in periapical radiographs: A comparative study of yolov8 and mask r-cnn,” *Diagnostics*, vol. 15, no. 6, p. 653, 2025.
- [114] K. M, G. G, S. B. T, S. Sk, J. Y, and M. Agoramoorthy, “A novel deep learning approach for enhanced roadway pothole detection using yolov8 instance segmentation algorithms,” *International Journal of Computational and Experimental Science and Engineering*, 2025.
- [115] A. Faisal and S. Gargoum, “Cost-effective lidar for pothole detection and quantification using a low-point-density approach,” *Automation in Construction*, vol. 172, p. 106006, 2025.

-
- [116] R. A. Pietersen, M. S. Beauregard, and H. H. Einstein, “Automated method for airfield pavement condition index evaluations,” *Automation in Construction*, vol. 141, p. 104408, 2022.