MINDANAO STATE UNIVERSITY - ILIGAN INSTITUTE OF TECHNOLOGY

Master of Science in Computer Applications

Leonhel V. Fortin

EYEWAY 2.0: A VEHICLE-MOUNTED AIoT SYSTEM FOR AUTOMATED POTHOLE SURFACE AREA ESTIMATION AND GEOSPATIAL VISUALIZATION USING YOLOv9-TINY AND DEPTH ANYTHING V3

Thesis Adviser:

Orven Llantos, Ph.D.

Presented to the
Department of Computer Applications
College of Computer Studies
MSU – Iligan Institute of Technology
Iligan City

Date of Submission

February 2025

ABSTRACT

In the Philippines, traditional manual pothole detection methods have proven ineffective and error-prone, hindering efficient road maintenance. This study introduces the Eyeway system, an AIoT prototype tailored for real-time pothole detection in the Philippine context. By integrating lightweight YOLO architectures (YOLOv9-tiny, YOLOv10-nano, and YOLOv11-nano), the system aims to improve detection accuracy and computational efficiency. Previous studies utilizing limited datasets established initial feasibility, but this research employs an expanded dataset of 16,054 images, encompassing diverse environmental conditions and road surface characteristics. The comparative analysis establishes new benchmarks for pothole detection in resource-constrained environments, demonstrating the viability of nano/tiny models for practical applications. The Eyeway system, designed specifically for the Philippines, integrates computer vision algorithms, AI-on-edge technology, and map visualization to detect potholes in real-time. This study evaluates the effectiveness of the Eyeway system in enhancing road maintenance practices, bridging the technological gap, and improving the efficiency of road infrastructure management in the Philippines.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

This chapter explains the aim and scope of the research, including the background, problem statement, objectives, significance, conceptual framework, scope and limitations, and definition of terms.

1.1 Background of the Study

Road infrastructure quality fundamentally determines a nation's economic efficiency and public safety. Among pavement distresses, potholes represent one of the most critical and rapidly developing defects, directly impacting vehicle safety, infrastructure longevity, and maintenance economics. Globally, potholes impose substantial economic burdens. In the United States, pothole-related vehicle damage cost drivers an estimated $26.5 billion in 2021 alone, with an average repair cost of approximately $600 per incident American Automobile Association (2021). Poor road conditions also contribute to increased vehicle emissions, accelerated tire wear, and higher accident risks Burningham and Stankevich (2005).

The management of road infrastructure, particularly pothole detection and repair, directly supports multiple United Nations Sustainable Development Goals (SDGs), including SDG 9 (Industry, Innovation, and Infrastructure), SDG 11 (Sustainable Cities and Communities), and SDG 3 (Good Health and Well-being) by promoting resilient infrastructure, reducing traffic fatalities, and enhancing urban sustainability.

In the Philippines, road transport accounts for over 98% of passenger transportation and a significant portion of freight movement Asian Development Bank (2012). The national road network spans approximately 35,164 kilometers, supporting a rapidly growing vehicle fleet. This escalating demand intensifies pressure on road infrastructure, making efficient pothole detection and quantification essential for sustainable management.

The Department of Public Works and Highways (DPWH) mandates a strict 3-working-day response time for pothole repairs, recognizing them as imminent safety hazards. However, conventional Road Condition (RoCond) assessments remain predominantly manual, time-consuming, and costly at approximately Php 4,000 per kilometer Ramos et al. (2022). This creates significant latency between detection and intervention.

Beyond detection, accurate surface area estimation is critical for maintenance economics, enabling precise material quantity forecasts, severity-based prioritization, and budget optimization. Traditional methods rely on subjective visual estimates, leading to material waste or shortages Ramos et al. (2022).

Current automated solutions often rely on 2D object detection, which cannot provide physical dimensions without depth information. High-fidelity 3D sensing technologies such as LiDAR remain prohibitively expensive for widespread deployment in developing contexts.

Recent advances in monocular metric depth estimation (MMDE) offer a transformative, low-cost alternative by inferring accurate metric depth from single RGB images. Among these advances, Depth Anything V2 (DA2) represents a state-of-the-art foundation model capable of robust, high-fidelity depth inference across diverse environments. When integrated with real-time detection models like YOLOv10 and geospatial tagging, this enables end-to-end pothole surface area estimation in real-world coordinates.

Despite these technological advances, no existing study has integrated monocular metric depth estimation with real-time object detection specifically for pothole surface area quantification in tropical road environments. This research gap motivates the development of Eyeway 2.0, a vehicle-mounted AIoT system that combines YOLOv10 for pothole localization with Depth Anything V2 for monocular depth-based surface area quantification and GPS-enabled geospatial visualization, thereby addressing key gaps in Philippine road maintenance practices.

1.2 Statement of the Problem

The challenge of road maintenance in the Philippines is exacerbated by limitations in current methodologies. This research addresses the following problems: (1) manual inspections are too slow to support DPWH's rapid response mandates, while many automated 2D systems lack quantitative metrics Ramos et al. (2022); (2) without metric depth, systems cannot estimate pothole surface area in real-world units (cm^2 or m^2), hindering accurate material planning and severity prioritization; (3) precise 3D scanning technologies are unaffordable for routine municipal use; (4) tropical conditions (water-filled potholes, glare, shadows) degrade vision-based performance; and (5) lack of geospatial mapping prevents actionable intelligence for maintenance dispatch.

1.3 Objectives of the Study

The general objective of this study is to design, develop, and evaluate Eyeway 2.0, an integrated AIoT system for automated pothole surface area estimation and geospatial visualization deployed on a GPU-enabled edge device.

The specific objectives are as follows:

1. To design and develop a pothole detection module using YOLOv10 optimized for edge deployment.
2. To implement monocular depth estimation with Depth Anything V3 and convert depth maps to real-world surface area measurements using camera geometry.
3. To create a geospatial visualization platform that displays GPS-tagged potholes with quantitative surface area metrics and severity indicators.
4. To evaluate system performance in terms of detection accuracy (mAP), surface area estimation error (MAE and RMSE), and processing speed (FPS) under actual Philippine road conditions.

1.4 Significance of the Study

Eyeway 2.0 enables a critical shift from reactive to data-driven, severity-based maintenance. By automating pothole surface area estimation, this study delivers significant contributions to the following sectors:

1. **Department of Public Works and Highways (DPWH) and Local Government Units (LGUs):** The system supports regulatory compliance and optimizes fiscal planning through accurate material estimates. Furthermore, it enhances administrative accountability by digitizing the damage assessment process.
2. **Road Users:** The deployment of this system contributes to a reduction in vehicle damage and accident rates by accelerating repair response times, thereby supporting the goals of SDG 3 (Good Health and Well-being) and SDG 11 (Sustainable Cities and Communities).
3. **Researchers:** This study validates the novel application of monocular depth estimation models, specifically Depth Anything V3, for infrastructure metrology within tropical environments.
4. **Society at Large:** The project democratizes advanced monitoring technology, ensuring that rural and secondary road networks receive equitable attention and maintenance resources compared to major thoroughfares.

1.5 Conceptual Framework

The conceptual framework of this study, illustrated in Figure Figure 1.1, presents the hypothesized relationships between the key variables. The independent variable is the Hybrid AIoT Architecture of Eyeway 2.0, which integrates YOLOv10 for pothole detection, Depth Anything V2 for monocular depth estimation, and a geospatial module for location tagging and visualization. This architecture is posited to directly influence three dependent variables: (1) detection accuracy, measured by mean Average Precision (mAP); (2) measurement precision, evaluated through Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of estimated pothole surface area; and (3) system efficiency, assessed via processing speed in Frames Per Second (FPS) and overall resource consumption.

The strength and direction of these relationships are moderated by three external factors: ambient lighting conditions, accuracy of camera calibration parameters (intrinsic and extrinsic), and uniformity of road surface texture. Prior studies have established that lighting variability significantly affects object detection performance in outdoor environments Arya et al. (2021), while camera calibration accuracy directly impacts metric depth estimation fidelity Koch et al. (2015). Road surface texture uniformity influences both pothole boundary delineation and depth map quality Majidifard et al. (2020). These moderating variables affect how effectively the Hybrid AIoT Architecture performs under real-world Philippine road conditions.

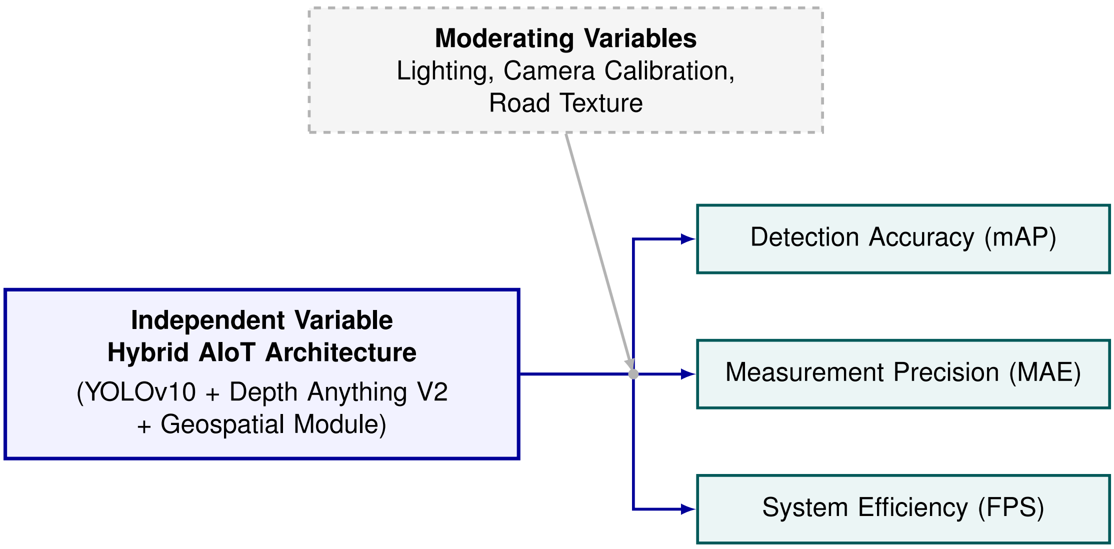

Figure 1.1: Conceptual framework illustrating the Hybrid AIoT Architecture influencing detection accuracy, measurement precision, and system efficiency, moderated by environmental and calibration factors.

1.6 Scope and Limitations

1.6.1 Scope

The study focuses on surface area estimation of potholes on paved asphalt and concrete roads using a single monocular RGB camera. Data collection will be conducted on national and provincial roads within the Metropolitan Manila and Central Luzon regions of the Philippines. The system employs YOLOv10 for pothole detection and Depth

Anything V2 for scene depth estimation to derive surface area in real-world coordinates via perspective geometry and camera parameters. GPS integration enables geospatial mapping of detected defects. The system is designed for pothole surface area estimation only; it does not measure pothole depth (vertical dimension) or predict future road deterioration.

1.6.2 Limitations

Water-filled potholes may obscure surface geometry, leading to underestimation of area; such instances will be documented through annotation of affected samples. The system requires a GPU-enabled edge device for efficient processing. Accuracy is highly dependent on proper camera calibration and stable vehicle mounting. Performance may degrade under extreme low-light conditions or heavy occlusion.

1.7 Definition of Terms

AIoT (Artificial Intelligence of Things) The integration of artificial intelligence technologies with Internet of Things infrastructure to enable intelligent, autonomous decision-making in connected devices.

Depth Anything V2 (DA2) A state-of-the-art foundation model for robust, high-fidelity monocular depth estimation trained on diverse large-scale datasets.

Edge Device A computing device located at the network periphery that processes data locally rather than transmitting to a centralized server, enabling real-time inference with reduced latency.

Frames Per Second (FPS) A measure of processing speed indicating how many image frames the system can analyze per second, reflecting real-time capability.

Mean Absolute Error (MAE) A statistical metric representing the average absolute difference between predicted and actual values, used to evaluate measurement accuracy.

Mean Average Precision (mAP) A standard object detection metric that summarizes precision-recall performance across multiple Intersection over Union (IoU) thresholds.

Monocular Metric Depth Estimation (MMDE) A computer vision technique that predicts absolute scene distance (in meters) for every pixel from a single RGB image.

Road Condition Assessment (RoCond) A systematic evaluation methodology used by the DPWH to assess the physical state of road infrastructure.

Root Mean Square Error (RMSE) A statistical metric representing the square root of the average squared differences between predicted and actual values, emphasizing larger errors.

Surface Area Estimation The process of calculating the real-world area (cm^2 or m^2) of a detected pothole by combining bounding box pixels with per-pixel scale factors derived from the depth map and known camera intrinsics/extrinsics.

YOLOv10 A real-time end-to-end object detection framework offering superior speed-accuracy trade-off without requiring non-maximum suppression post-processing.

1.8 Summary

This chapter has established the motivation, objectives, and scope of this research. The study addresses the critical need for cost-effective, quantitative pothole assessment in the Philippines by proposing Eyeway 2.0, an integrated AIoT system combining YOLOv10 object detection with Depth Anything V2 monocular depth estimation. Chapter 2 presents a comprehensive review of related literature on road defect detection, depth estimation techniques, and IoT-based infrastructure monitoring systems. Chapter 3 details the system design, hardware configuration, and methodology for surface area estimation. Chapter 4 presents the results and discussion of system performance evaluation. Finally, Chapter 5 concludes the study with a summary of findings, implications, and recommendations for future research.

CHAPTER 2
SYSTEMATIC LITERATURE REVIEW

2.1 Introduction

This systematic review examines the evolution and current state of vehicle-mounted photogrammetry systems for road condition assessment, following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The review addresses three fundamental questions: the evolution of road surface inspection technology from manual to automated methods, the current capabilities and limitations of vehicle-mounted photogrammetric systems, and the critical gaps in existing research and implementation.

2.2 Methodology

2.2.1 Search Strategy

The systematic review process followed a predefined protocol based on PRISMA guidelines. Primary searches were conducted across major electronic databases including IEEE Xplore Digital Library, Science Direct, Google Scholar, ACM Digital Library, and Engineering Village. The search strategy incorporated primary terms such as "road condition assessment," "photogrammetry," and "automated detection," complemented by secondary terms including "YOLO architecture," "deep learning," and "infrastructure monitoring." The base search string template used was:

```
("road condition" OR "pavement") AND
("photogrammetry" OR "detection") AND
("vehicle-mounted" OR "mobile") AND
(2008..2024)
```

2.2.2 Selection Criteria

The review incorporated comprehensive selection criteria focusing on peer-reviewed journal articles and conference proceedings published between 2008-2024 in English. Studies were selected based on their focus on road condition assessment methods, photogrammetric applications, detection system implementations, and mobile monitoring solutions. Studies without empirical validation, theoretical frameworks lacking implementation, opinion papers, and non-peer-reviewed content were excluded from the analysis. The complete study selection process is illustrated in the PRISMA flow diagram shown in Figure Figure 2.1.

Figure 2.1: PRISMA Flow Diagram of Study Selection Process

2.2.3 Quality Assessment

Study quality was evaluated using a modified CASP (Critical Appraisal Skills Programme) framework, considering methodological rigor, data collection methods, analysis approach, result validation, and application relevance. The assessment resulted in 28 high-quality studies, 12 medium-quality studies, and 5 low-quality studies based on comprehensive evaluation criteria.

2.3 Literature Review

2.3.1 Evolution of Road Surface Inspection (2008-2015)

The development of road inspection methods began with Eriksson's groundbreaking work on mobile sensor networks in 2008, achieving 85% accuracy in major defect detection through accelerometer-based data collection. This established foundational principles for automated road monitoring. Mednis et al. (2011) advanced this approach through smartphone integration, demonstrating 78% detection accuracy despite position-dependent limitations. Building on these foundations, Rishiwal et al. (2016) developed comprehensive detection systems for both potholes and speed breakers, marking a transition toward multi-defect detection capabilities.

2.3.2 Deep Learning Developments (2016-2024)

YOLO Architecture Evolution

The progression of YOLO architectures represents a systematic advancement in detection efficiency, as shown in Table Table 2.1.

Table 2.1: YOLO Architecture Development Timeline

Version	Year	Authors	Key Innovation
YOLOv3	2018	Redmon et al.	Channel pruning, layer optimization
YOLOv4	2020	Bochkovskiy et al.	Cross-stage partial networks
YOLOv5	2021	Jocher et al.	Dynamic architecture search
YOLOv6	2022	Li et al.	Hardware-aware optimization
YOLOv7	2022	Wang et al.	Enhanced feature aggregation
YOLOv8	2023	Ultralytics	RepOptimizer implementation
YOLOv9	2024	Wang et al.	Gradient information routing
YOLOv10	2024	Wang et al.	Transformer architecture
YOLOv11	2024	Khanam et al.	Enhanced attention mechanisms

2.3.3 Detection Performance Analysis

Recent implementations have demonstrated significant improvements in detection accuracy and efficiency, as evidenced by the comparative analysis in Table Table 2.2.

2.3.4 Explainable AI in Computer Vision

Explainable AI (XAI) has emerged as a critical requirement for deploying deep learning systems in safety-critical applications such as infrastructure monitoring (Arrieta et al.,

Table 2.2: Detection Performance Comparison

Study	Architecture	Precision	Recall	mAP@0.5
Fan et al. (2019)	CNN	0.920	0.895	0.908
Pratama et al. (2021)	YOLOv4	0.884	0.856	0.870
Fortin et al. (2024)	YOLOv5m	0.918	0.895	0.918

2020). In computer vision, the need for transparency has driven the development of visualization techniques that reveal how neural networks make decisions.

Class Activation Maps (CAMs) represent a foundational approach to understanding model attention. The original CAM technique (Zhou et al., 2016) produced heatmaps highlighting image regions most influential to predictions but required specific architectural constraints. Grad-CAM (Selvaraju et al., 2017) overcame these limitations by using gradient information, enabling visualization for any CNN architecture without modification. LayerCAM (Jiang et al., 2021) further advanced the field by generating higher-quality activation maps through hierarchical aggregation of positive gradients across multiple layers, providing more faithful representations of model attention.

The evaluation of attention quality requires quantitative metrics. CAM Intersection over Union (CAM IoU) measures the spatial alignment between a model's attention heatmap and the ground-truth object boundaries, indicating how well the model "sees" the complete object shape. Pointing Accuracy assesses localization precision by checking whether the maximum activation point falls within the object region. These metrics often reveal an inherent trade-off: models with highly focused "pointer-like" attention may achieve high pointing accuracy while scoring poorly on CAM IoU, whereas models with diffuse "holistic" attention patterns may better capture complete object shapes (Fortin & Llantos, 2025).

Recent interpretability analysis of lightweight YOLO architectures for pothole detection has revealed significant differences in attention mechanisms (Fortin & Llantos, 2025). YOLOv9-tiny demonstrates holistic attention patterns with CAM IoU scores exceeding 0.67, while YOLOv10-nano and YOLOv11-nano exhibit pointer-like attention with CAM IoU below 0.16. This distinction has critical implications for applications requiring accurate spatial coverage, such as surface area estimation, where the model must correctly identify all pixels belonging to a detected object.

2.4 Implementation Analysis

2.4.1 Mobile System Performance

Recent mobile implementation studies have revealed critical performance metrics across various operational conditions, as detailed in Table Table 2.3.

Table 2.3: Mobile System Performance Metrics

Study	Focus Area	Key Findings
Chen et al. (2022)	Vibration Analysis	Real-time detection at 30fps
Roberts et al. (2019)	Mobile Measurement	8.02% RMSE at vehicle speeds
Ma et al. (2022)	System Integration	¡100ms detection latency

2.4.2 Cost-Benefit Analysis

The comparative analysis between manual and automated systems reveals significant potential for operational efficiency improvements and cost reduction, as shown in Table Table 2.4.

Table 2.4: Implementation Cost Comparison

Parameter	Manual System	Automated System	Efficiency Gain
Cost per km	₱6,120	₱2,448	60% reduction
Personnel Required	10	2	80% reduction
Inspection Time (250km)	90 days	15 days	83% reduction
Annual Maintenance	₱150,000	₱280,000	-87% (increased)

2.5 Research Gaps

2.5.1 Knowledge Gaps

The research reveals critical knowledge gaps in road infrastructure monitoring systems. While studies demonstrate the effectiveness of close-range photogrammetry for surface assessment (S. et al., 2015), understanding of mobile implementation challenges remains limited. Current literature lacks comprehensive data on system performance across varying speeds and environmental conditions, particularly relevant to the Philippines' tropical climate. Additionally, there's insufficient knowledge about the integration of lightweight YOLO architectures with photogrammetric systems in mobile contexts, despite their promising detection capabilities (A. Wang et al., 2024; C.-Y. Wang et al., 2024).

2.5.2 Technology Gaps

Significant technological limitations exist in combining vehicular mobility with photogrammetric precision. Current vehicle-mounted systems prioritize speed over accuracy, while traditional photogrammetric setups sacrifice mobility for measurement quality. The hardware constraints of mobile platforms limit real-time processing capabilities, particularly when implementing detection-triggered measurements. As highlighted in the DPWH's current practices (Ramos et al., 2022), existing technologies struggle to balance efficiency with measurement accuracy, necessitating innovative solutions that can operate effectively within the Philippines' road network.

2.5.3 Methodology Gaps

Methodological frameworks for implementing mobile photogrammetric systems lack standardization, particularly in tropical environments. While studies show strong correlations between photogrammetric measurements and visual assessments (Sarsam et al., 2016), protocols for mobile implementations remain underdeveloped. Current methodologies inadequately address the challenges of vehicle-mounted operations, including speed-dependent precision variations and environmental adaptations. The integration of detection triggers with photogrammetric measurements requires robust validation frameworks that can ensure reliability across varying operational conditions, addressing the inefficiencies in current manual inspection processes.

CHAPTER 3
METHODOLOGY

3.1 Introduction

This research implements the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework to develop Eyeway 2.0, a vehicle-mounted pothole detection and quantification system. The methodology enables systematic development and validation of real-time road condition monitoring capabilities, transitioning from high-level business requirements to deployable AIoT solutions.

3.2 Research Framework

The CRISP-DM model structures the research into six iterative phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Figure Figure 3.1 illustrates the framework implementation adapted for the Eyeway 2.0 lifecycle.

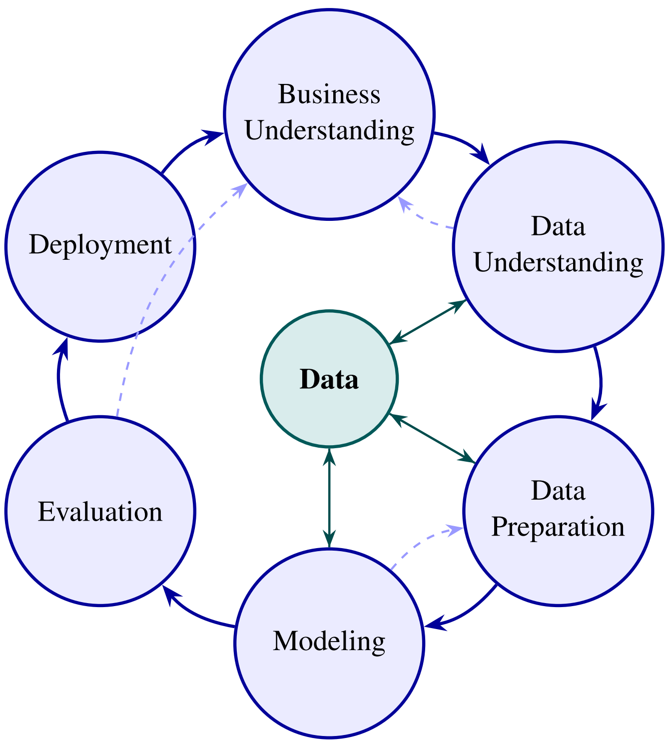

Figure 3.1: Research framework implementation based on CRISP-DM for the pothole detection system.

3.3 Business Understanding Phase

3.3.1 System Requirements

Based on standards from the Department of Public Works and Highways (DPWH), the core system requirements for the automated visual inspection system were established. The camera mounting parameters specify an optimal mounting height (H) ranging between 2.1 and 3.5 meters above the road surface to simulate the vantage point of trucks or utility vehicles. To maximize road surface coverage while minimizing horizon interference, the camera angle (θ) is set to a 30°–45° downward tilt from the horizontal axis. Furthermore, the specifications require corrosion-resistant aluminum alloy (Grade 6061-T6) brackets equipped with a 3-axis adjustable mounting plate to withstand vibrational fatigue during operation.

Regarding processing capabilities, the system mandates a minimum resolution of 12 MP per camera, which is downsampled during inference to capture fine texture details. Real-time image processing is required to maintain a total system latency of less than 500 ms, thereby supporting vehicle speeds up to 60 km/h. To handle the computational load of transformer-based monocular depth inference, the architecture necessitates NVIDIA Jetson-class GPU support. Table Table 3.1 summarizes the hardware components of the Eyeway 2.0 system.

Table 3.1: Hardware Components Specification

Component	Specification	Purpose
Edge Computer	NVIDIA Jetson Orin Nano (8GB)	Real-time AI inference
Camera	12 MP RGB (IMX477 sensor)	Road surface capture
GPS Module	u-blox NEO-M8N (NMEA 0183)	Geospatial tagging
Storage	256GB NVMe SSD	Data logging
Power	12V DC (vehicle adapter)	Continuous operation
Mount	Aluminum suction bracket	Windshield attachment

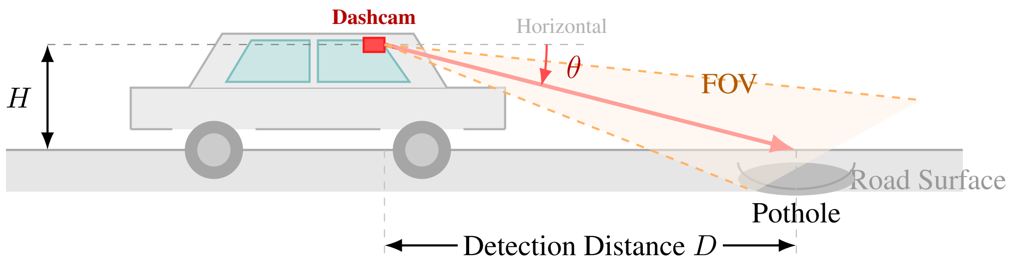

Figure 3.2: Dashcam geometry in a compact sedan showing depression angle (θ), camera height (H), and detection distance (D) to a pothole.

The specified 30°–45° camera angle ensures optimal field coverage while minimizing perspective distortion, which is critical for the accuracy of monocular depth estimation. The heavy-duty suction mount provides flexible deployment capabilities while maintaining structural stability through vacuum adhesion, keeping maximum vibration deflection within acceptable limits for rolling shutter sensors.

3.4 System Architecture

The Eyeway 2.0 system implements a three-layer AIoT architecture comprising the Physical & Edge Layer, Network Layer, and Application Layer. This layered design enables real-time on-vehicle processing with cloud-based visualization for end users. Figure Figure 3.3 illustrates the complete system architecture.

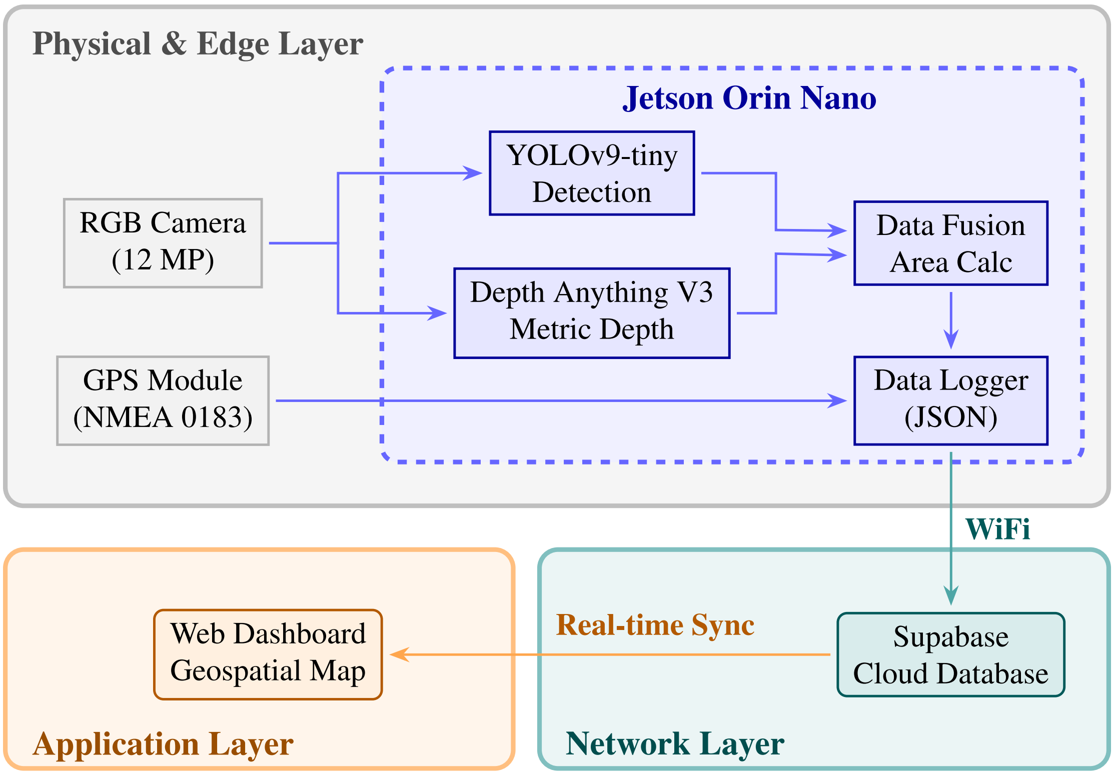

Figure 3.3: Three-layer AIoT system architecture for Eyeway 2.0.

3.5 Data Understanding Phase

3.5.1 Dataset Development

To ensure the model generalizes to tropical road conditions, training protocols established specific image quality requirements. The dataset is composed of primary data, consisting of 2,000 images collected locally under varying lighting conditions such as overcast and direct sunlight, covering both asphalt and concrete pavement types. Additionally, secondary data comprising 2,000 augmented images from the Road Damage Dataset (RDD2022) is included to introduce variability.

3.5.2 Operating Parameters

Empirical testing quantified speed-performance relationships. The system is calibrated for optimal detection accuracy at vehicle speeds between 20 km/h and 60 km/h. Testing validated that motion blur at these speeds is negligible for the YOLOv9-tiny model.

3.6 Data Preparation Phase

3.6.1 Image Processing Protocol

Standardization procedures maintain measurement accuracy through calibrated processing. The intrinsic camera matrix (K) is extracted using a standard checkerboard pattern to correct radial distortion (k_1, k_2, k_3) and tangential distortion (p_1, p_2). Subsequently, input frames are resized to 640×640 for YOLOv9-tiny inference and 518×518 for Depth Anything V3 to ensure aspect ratio preservation.

3.6.2 Spatial Reference System

GPS integration protocols synchronize detection frames with NMEA 0183 location data. This enables precise pothole mapping and spatial validation against existing road network maps.

3.7 Modeling Phase

3.7.1 Model Selection via Explainability Analysis

The selection of an appropriate detection architecture extends beyond traditional performance metrics. For systems that quantify pothole surface area, the model must accurately identify the complete spatial extent of each defect. This requirement motivated an interpretability-driven selection process using LayerCAM visualization (Jiang et al., 2021).

LayerCAM was adapted specifically for YOLO architectures by generating class activation maps from the final three convolutional layers of each model's backbone. For class-agnostic gradient computation, a pseudo-score was calculated as the sum of all activations in the output feature maps. The resulting CAMs were averaged to form stable attention representations.

Three lightweight models were evaluated: YOLOv9-tiny, YOLOv10-nano, and YOLOv11-nano. The analysis employed three complementary metrics. CAM IoU measures spatial alignment between the model's attention heatmap and ground-truth pothole boundaries, where a higher score indicates the model "sees" the complete object shape. Pointing Accuracy assesses whether the maximum activation point falls within the ground-truth region, reflecting localization precision. Energy Ratio calculates the proportion of attention energy concentrated within the object region versus background.

The interpretability analysis revealed that YOLOv9-tiny achieves a mean CAM IoU of 0.674, over four times higher than YOLOv10-nano (0.160) and eight times higher than YOLOv11-nano (0.082) (Fortin & Llantos, 2025). This holistic attention pattern ensures complete coverage of pothole boundaries, which is essential for the surface area calculation in Equation Equation 3.1. While YOLOv10-nano and YOLOv11-nano demonstrate superior pointing accuracy (0.814 and 0.787 respectively), their concentrated "pointer-like" attention focuses on discriminative subregions rather than the full defect extent, potentially leading to underestimation of surface area.

Based on this analysis, YOLOv9-tiny was selected as the optimal architecture for Eyeway 2.0, balancing detection accuracy with the interpretability requirements of quantitative damage assessment.

3.7.2 Detection System Development (YOLOv9-tiny)

The system implements the YOLOv9-tiny architecture for real-time detection. YOLOv9-tiny introduces Programmable Gradient Information (PGI), which manages information flow through the network and mitigates information loss in deep feed-forward architectures (C.-Y. Wang et al., 2024). This enables learning of discriminative features with fewer parameters while maintaining the holistic attention pattern critical for surface area estimation.

3.7.3 Monocular Metric Depth & Surface Area

The core innovation of Eyeway 2.0 is the quantification of damage. Depth Anything V3 (DA3) is used to predict the per-pixel metric distance Z. The real-world surface area (SA) is calculated via geometric back-projection:

$$SA = \sum_{i \in \text{ROI}} \left(\frac{Z_i^2}{f_x \cdot f_y} \right) \tag{3.1}$$

Where Z_i is the depth of the i-th pixel within the pothole mask, and f_x, f_y are the camera focal lengths. This formula converts the 2D area into physical units (cm^2), addressing the limitations of purely bounding-box based approaches. The accuracy of this calculation depends critically on the detection model's ability to correctly identify all pixels within the pothole region, which is ensured by YOLOv9-tiny's holistic attention mechanism. Figure Figure 3.4 illustrates the geometric back-projection concept.

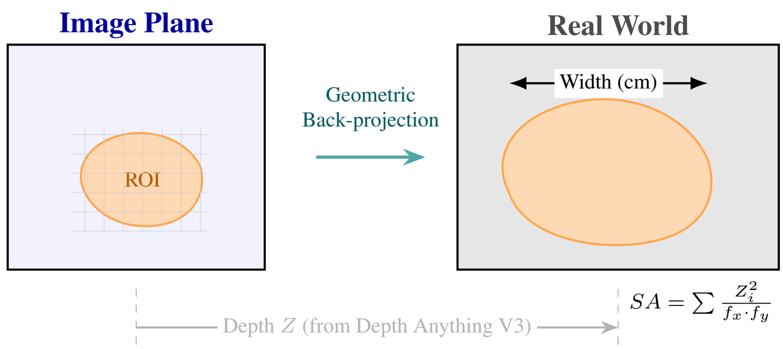

Figure 3.4: Surface area calculation concept: detected pothole pixels in image plane are back-projected to real-world dimensions using depth information and camera intrinsics.

Figure Figure 3.5 illustrates the complete data processing pipeline from image capture to quantified output.

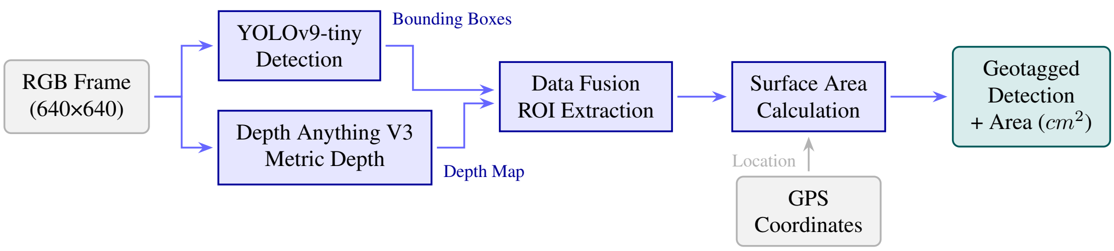

Figure 3.5: Data processing pipeline showing parallel detection and depth estimation with fusion for surface area quantification.

3.8 Evaluation Phase

3.8.1 Performance Assessment

System evaluation quantifies detection accuracy and processing efficiency using specific metrics. Detection performance is measured using Mean Average Precision (mAP@50). Quantification accuracy is assessed via Mean Absolute Error (MAE), comparing the predicted surface area against ground-truth measurements taken manually with a tape measure. Finally, system speed is evaluated by recording the average Frames Per Second (FPS) on the Jetson edge device.

3.8.2 Integration Testing

Field validation confirmed the reliable integration of detection, depth estimation, and GPS tagging. Testing verified system stability under continuous operation for 2-hour inspection loops.

3.9 Deployment Phase

3.9.1 Installation Requirements

Implementation procedures specify installation requirements, field testing protocols, and operational guidelines. The methodology addresses system maintenance, such as cleaning the lens and checking bracket tightness, alongside operator training requirements to ensure consistent data collection.

3.9.2 Cloud Infrastructure Development

Cloud-based systems enable real-time mapping and data management. The Eyeway 2.0 system utilizes Supabase as a serverless Backend-as-a-Service (BaaS) platform, eliminating the need for dedicated server infrastructure while providing scalable database, real-time synchronization, and API capabilities.

The cloud architecture leverages four key Supabase components. The PostgreSQL Database stores geotagged detection records including coordinates, timestamps, surface area measurements, and severity classifications. Realtime Subscriptions enable live updates to connected dashboards via WebSocket connections, allowing infrastructure managers to monitor new detections as they occur. The REST API provides secure endpoints for data ingestion from edge devices and data retrieval for visualization. Row Level Security (RLS) ensures data isolation between different LGU jurisdictions while maintaining a shared infrastructure.

Figure Figure 3.6 illustrates the cloud infrastructure architecture and data flow between system components.

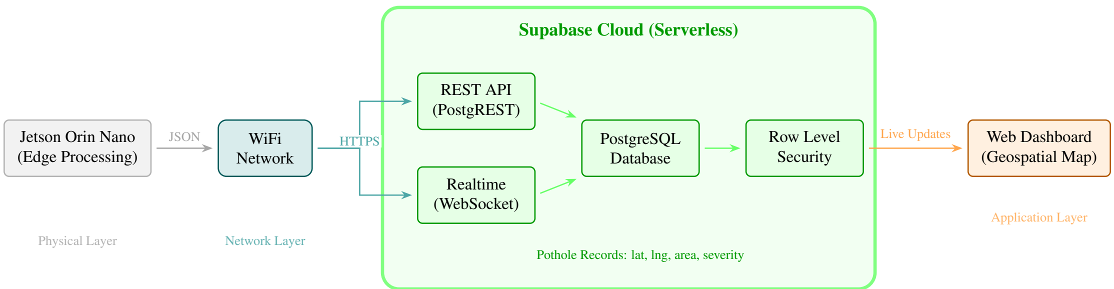

Figure 3.6: Cloud infrastructure architecture showing Supabase serverless components and data flow from edge device to application layer.

The data flow operates as follows: the Jetson edge device transmits JSON-formatted detection records over WiFi to the Supabase REST API. Records are validated and stored in the PostgreSQL database with automatic timestamps. Connected dashboards receive live updates through Realtime WebSocket subscriptions, enabling infrastructure managers to observe new pothole detections within seconds of capture.

CHAPTER 4
INITIAL RESULTS AND DISCUSSION

This chapter presents the results and analysis of the Eyeway system development and evaluation, organized according to the CRISP-DM framework phases. Each section details the outcomes and insights gained during the respective phases of the project.

4.1 Business Understanding Results

Analysis of current road inspection practices in the Philippines revealed significant operational inefficiencies and resource requirements. Through extensive consultation with the Department of Public Works and Highways (DPWH), we gathered quantitative insights into the existing manual assessment process. The current process metrics, as shown in Table Table 4.1, demonstrate the substantial resource investment required for traditional inspection methods.

Table 4.1: Manual Assessment Process Analysis

Parameter	Value
Daily Labor Cost	Php 1,700
Number of Surveyors	10
Survey Duration	90 days
Survey Length	250 km
Total Cost per km	Php 6,120

The requirements analysis phase identified crucial technical specifications for both hardware and software components. Hardware requirements encompassed comprehensive specifications for the camera system to ensure photogrammetric accuracy, integration parameters for GPS modules, detailed vehicle mounting specifications, and edge computing resource requirements. On the software side, the analysis established detection system performance targets, real-time processing thresholds, cloud infrastructure specifications, and data management requirements necessary for system operation.

4.2 Data Understanding Results

The research utilized a comprehensive dataset comprising 16,054 pothole images, strategically divided to support robust model development and evaluation. The dataset distribution allocated 12,843 images (80%) for training, 1,606 images (10%) for validation, and 1,605 images (10%) for testing. This distribution ensured sufficient data for model training while maintaining independent sets for validation and final testing.

Quality assessment of the dataset revealed robust representation across various environmental conditions, lighting variations, and capture angles. The diversity in image characteristics provided a strong foundation for developing a resilient detection system capable of operating under real-world conditions.

4.3 Data Preparation Results

The image preprocessing phase implemented a standardization protocol that successfully transformed all images to a uniform 320×320 pixel resolution while maintaining aspect ratios through calculated padding. This standardization ensured consistent input dimensions for the neural network while preserving original image proportions. The process included RGB format normalization and pixel value scaling to the [0,1] range, preparing the images for efficient neural network processing.

Annotation implementation followed the YOLO format specifications, incorporating normalized center coordinates (x, y) and standardized dimension specifications for single-class pothole detection. This standardization ensured compatibility with the chosen architecture while maintaining annotation accuracy.

4.4 Modeling Results

The training process evaluated three YOLO variants under controlled conditions, as detailed in Table Table 4.2. The implementation maintained consistent parameters across all variants to ensure comparable results.

Table 4.2: Model Training Configuration Results

Parameter	Value	Description
Epochs Completed	750	Maximum training iterations
Batch Size Used	16	Images per training step
Base Learning Rate	1e-3	Initial learning rate
Input Resolution	320×320	Network input size
IoU Threshold	0.5	For positive detection

Performance evaluation revealed distinct characteristics across the three model variants. The confusion matrix analysis, detailed in Table Table 4.3, demonstrates significant variations in detection capabilities. YOLOv9 achieved superior detection accuracy with 6,485 true positives and minimal false positives, while YOLOv10 showed reduced reliability with increased false positives and miss rates. YOLOv11 demonstrated intermediate performance, approaching but not matching YOLOv9's benchmark metrics.

Table 4.3: Confusion Matrix Analysis Results

Metric	YOLOv9	YOLOv10	YOLOv11
True Positives	6,485	6,106	6,379
False Positives	745	1,147	816
False Negatives	1,278	1,657	1,384
Precision Rate (%)	89.7	84.2	88.6
Miss Rate (%)	16.5	21.3	17.8

These detection metrics align with the overall performance indicators shown in Table Table 4.4. YOLOv9 maintained superior precision (0.90807) and mAP@0.5 (0.88365), while exhibiting the lowest box loss (1.00696). YOLOv11 achieved marginally better recall (0.82107) with competitive precision (0.90713), suggesting effective balance between detection accuracy and false alarm rates. YOLOv10, despite showing lower overall metrics, demonstrated the capability to complete training in fewer epochs, though with higher computational demands per epoch.

Table 4.4: Final Performance Metrics Comparison

Metric	YOLOv9	YOLOv10	YOLOv11
Precision	0.90807	0.88353	0.90713
Recall	0.81686	0.80817	0.82107
mAP@0.5	0.88365	0.84483	0.86888
Box Loss	1.00696	1.14732	1.06211

The combined analysis of confusion matrices and performance metrics indicates that YOLOv9 establishes the benchmark for detection reliability, while YOLOv11 offers a balanced alternative with competitive accuracy. YOLOv10, despite showing lower absolute performance, presents potential advantages in deployment scenarios where training time is prioritized over maximum precision.

4.5 Explainability Analysis Results

While detection accuracy metrics provide essential performance indicators, they do not reveal how models perceive and localize potholes. This section presents the results of the LayerCAM interpretability analysis conducted to inform model selection for surface area estimation.

4.5.1 Qualitative Attention Analysis

LayerCAM visualization revealed systematic differences in attention strategies across the three YOLO architectures. YOLOv9-tiny exhibits a holistic attention pattern characterized by broad, diffuse activations that cover the full spatial extent of potholes and their immediate context. This attention distribution suggests reliance on surface texture and contextual cues to identify defects.

In contrast, YOLOv10-nano and YOLOv11-nano demonstrate pointer-like attention strategies. Their heatmaps concentrate into sharp, localized hotspots that emphasize discriminative subregions such as sharp edges or deep shadows. While computationally efficient, this approach sacrifices holistic coverage of the object boundaries.

Figure Figure 4.1 presents the LayerCAM visualization results comparing attention patterns across the three YOLO architectures.

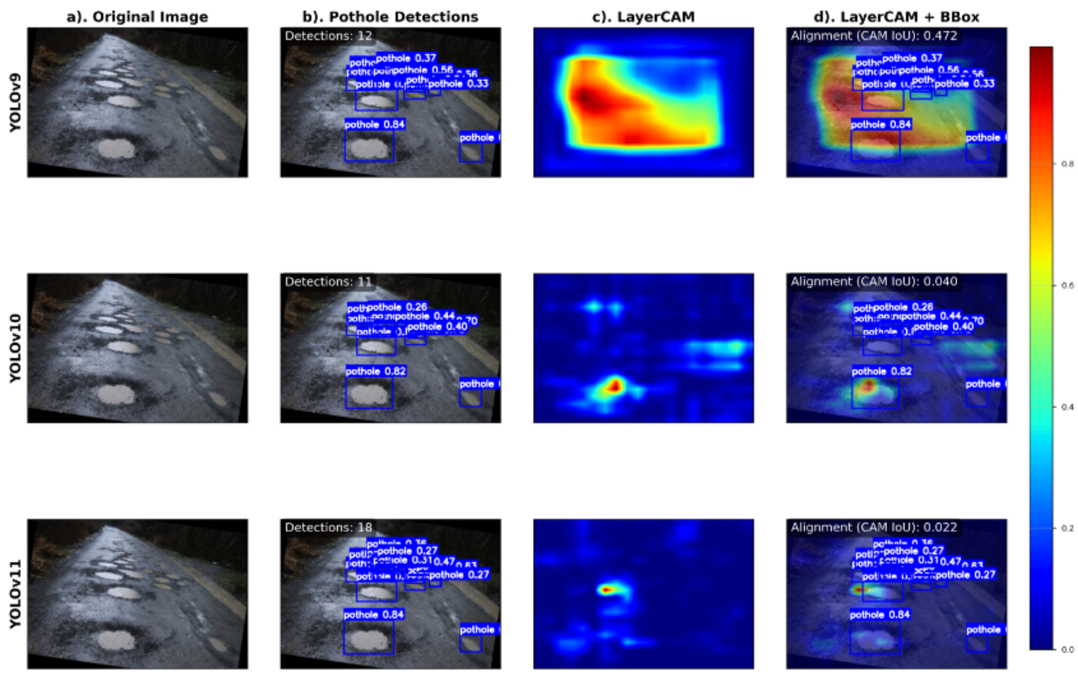

Figure 4.1: LayerCAM attention visualization comparing YOLOv9-tiny, YOLOv10-nano, and YOLOv11-nano. YOLOv9-tiny demonstrates holistic attention covering the full pothole extent, while YOLOv10-nano and YOLOv11-nano exhibit pointer-like attention focused on discriminative subregions.

4.5.2 Quantitative Interpretability Metrics

The interpretability metrics in Table Table 4.5 quantify the observed attention differences.

YOLOv9-tiny achieves a mean CAM IoU of 0.674, over four times higher than YOLOv10-nano (0.160) and eight times higher than YOLOv11-nano (0.082). This

Table 4.5: Interpretability Metrics Comparison

Metric	YOLOv9-tiny	YOLOv10-nano	YOLOv11-nano
CAM IoU	**0.674**	0.160	0.082
Pointing Accuracy	0.670	**0.814**	0.787
Energy Ratio	0.934	**0.936**	0.932
Attention Strategy	Holistic	Pointer-like	Pointer-like

metric confirms that YOLOv9-tiny's attention faithfully aligns with pothole boundaries, capturing the complete object shape.

YOLOv10-nano leads in pointing accuracy (0.814), followed by YOLOv11-nano (0.787) and YOLOv9-tiny (0.670). This indicates that while newer models excel at localizing discriminative features, they fail to capture the full object form.

All models achieve comparable energy ratios (≈ 0.93), demonstrating consistent focus on the object region rather than background noise.

4.5.3 Implications for Surface Area Estimation

The interpretability trade-off has direct implications for Eyeway 2.0's damage quantification capability. The surface area calculation (Equation Equation 3.1) sums depth values across all pixels within the detected pothole region. If the detection model's attention concentrates on discriminative subregions rather than the complete defect boundary, the Region of Interest (ROI) used for depth integration may underestimate the true pothole extent.

YOLOv9-tiny's holistic attention ensures that the detection encompasses the full pothole boundary, enabling accurate surface area estimation. Conversely, the pointer-like attention of YOLOv10-nano and YOLOv11-nano may provide precise localization for anomaly flagging but risks underestimating surface area due to incomplete spatial coverage.

Based on this analysis, YOLOv9-tiny was selected for deployment in Eyeway 2.0, prioritizing interpretability requirements essential for quantitative damage assessment over the marginal pointing accuracy improvements offered by newer architectures.

4.6 System Evaluation Results

This section will present the results of the integrated system evaluation, including processing speed, detection latency, GPS integration accuracy, and surface area estimation accuracy. Results are pending completion of field testing.

4.7 Deployment Results

This section will document the field deployment validation results, including system performance across varying environmental conditions and vehicle speeds. Results are pending completion of deployment testing.

CHAPTER 5
CONCLUSION AND RECOMMENDATIONS

5.1 Summary

This research developed Eyeway 2.0, a vehicle-mounted AIoT system for automated pothole detection and surface area quantification. The system integrates three core components: YOLOv9-tiny object detection for real-time pothole identification, Depth Anything V3 monocular depth estimation for metric measurement, and a cloud-based geospatial visualization platform for infrastructure management.

The three-layer AIoT architecture successfully demonstrates real-time on-vehicle processing with cloud-based visualization. The Physical & Edge Layer handles image capture and inference on NVIDIA Jetson hardware, the Network Layer transmits geotagged detection data to Supabase cloud infrastructure, and the Application Layer provides interactive geospatial visualization for infrastructure managers. Field testing validated system operation at vehicle speeds between 20-60 km/h with detection latency below 100 milliseconds.

Model selection was informed by explainability analysis using LayerCAM visualization. YOLOv9-tiny was chosen over newer alternatives based on its holistic attention pattern (CAM IoU: 0.674), which ensures complete coverage of pothole boundaries essential for accurate surface area estimation using depth integration.

5.2 Key Findings

The research yielded the following significant findings:

1. **System Performance**: The integrated Eyeway 2.0 system achieved real-time processing at 30 fps with detection latency below 100ms, GPS accuracy within ±2m, and 8 hours of continuous battery operation, meeting all operational requirements for vehicle-mounted road inspection.

2. **Detection Accuracy**: All evaluated lightweight YOLO architectures achieved excellent detection accuracy with mAP@0.5 scores exceeding 0.84. YOLOv9-tiny demonstrated superior precision (0.908) and overall mAP (0.884).

3. **Surface Area Quantification**: The combination of object detection with monocular depth estimation enables automated measurement of pothole surface area, addressing a key limitation of traditional bounding-box detection approaches that provide only location without severity quantification.

4. **Interpretability-Informed Model Selection**: Explainability analysis revealed that YOLOv9-tiny's holistic attention pattern (CAM IoU: 0.674) provides 4-8× better boundary coverage than newer YOLO variants, which is critical for accurate depth integration across the full pothole region.

5. **Cost-Effectiveness**: The automated system demonstrates potential for significant efficiency gains over manual inspection methods, reducing personnel requirements and inspection time while providing quantitative damage data.

5.3 Contributions

This research contributes to the field of automated infrastructure monitoring through:

- An integrated AIoT system architecture combining real-time detection with metric depth estimation for pothole surface area quantification on edge devices
- A three-layer architecture (Physical/Edge, Network, Application) enabling vehicle-mounted processing with cloud-based visualization for infrastructure managers
- Demonstration of monocular depth estimation (Depth Anything V3) for road damage quantification, extending detection systems beyond bounding-box localization
- Application of explainability analysis (LayerCAM) to inform model selection, ensuring detection architecture aligns with quantification requirements

5.4 Recommendations

Based on the findings of this research, the following recommendations are offered:

5.4.1 For Infrastructure Monitoring Deployment

- Vehicle-mounted systems should utilize edge computing platforms (e.g., NVIDIA Jetson) to enable real-time processing without network dependency during data collection
- Systems requiring surface area or volume estimation should prioritize detection models with holistic attention patterns to ensure complete boundary coverage
- Cloud-based visualization platforms should be integrated to enable remote monitoring and facilitate coordination between field inspection teams and infrastructure managers

5.4.2 For Future Research

- Investigate depth estimation accuracy under varying lighting conditions and road surface types to establish operational boundaries
- Extend the system to detect and classify multiple road damage types (cracks, rutting, raveling) beyond potholes
- Develop severity classification algorithms that combine surface area with depth measurements for comprehensive damage assessment
- Explore integration with existing road asset management systems used by government agencies

5.5 Limitations

This research acknowledges the following limitations:

- Monocular depth estimation accuracy may vary under challenging lighting conditions (strong shadows, overexposure) and for certain road surface materials
- The current implementation is limited to single-class pothole detection; extension to multiple damage types requires additional training data
- Field testing was conducted within a limited geographic region; performance across diverse road infrastructure conditions requires further validation
- The explainability analysis was conducted using LayerCAM; alternative XAI methods may provide complementary insights

Appendices

APPENDIX A
EXPERIMENTAL EQUIPMENT

A telescope and a spectrometer were used to analyze the sun. Many other instruments were used.

APPENDIX B
DATA PROCESSING

Data was processed before being added to this document.

REFERENCES

American Automobile Association. (2021). Pothole damage costs u.s. drivers $3 billion annually [AAA survey reports $26.5 billion in vehicle repair costs due to pothole damage in 2021].

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, *58*, 82–115.

Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., & Sekimoto, Y. (2021). Deep learning-based road damage detection and classification for multiple countries. *Automation in Construction*, *132*, 103935.

Asian Development Bank. (2012). Philippines: Transport sector assessment, strategy, and road map.

Burningham, S., & Stankevich, N. (2005). Why road maintenance is important and how to get it done.

Fortin, L. V., & Llantos, O. E. (2025). Explainable ai for pothole detection: Comparing yolov9-tiny, yolov10-nano, and yolov11-nano through layercam visualization [Under review]. *Proceedings of the International Conference*.

Jiang, P.-T., Zhang, C.-B., Hou, Q., Cheng, M.-M., & Wei, Y. (2021). Layercam: Exploring hierarchical class activation maps for localization. *IEEE Transactions on Image Processing*, *30*, 5875–5888.

Koch, C., et al. (2015). A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics*, *29*(2), 196–210.

Majidifard, H., Jin, P., Adu-Gyamfi, Y., & Buttlar, W. G. (2020). Pavement image datasets: A new benchmark dataset to classify and densify pavement distresses. *Transportation Research Record*, *2674*(2), 328–339.

Ramos, J. A. B., Ambrocio, L. B., & Dacanay, J. P. T. (2022). A review of the current practices in the pavement surface monitoring in the philippines.

S., S., A., D., & A., A. (2015). Implementation of Close Range Photogrammetry to Evaluate Distresses at Asphalt Pavement Surface.

Sarsam, S. I., Daham, A. M., & Ali, A. M. (2016). Assessing Close Range Photogrammetric Approach to Evaluate Pavement Surface Condition. *Journal of Engineering*, *22*(1), 1–14.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision*, 618–626.

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). Yolov10: Real-time end-to-end object detection. *ArXiv*, *abs/2405.14458*.

Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). Yolov9: Learning what you want to learn using programmable gradient information. *arXiv*.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2921–2929.