Pothole Detection: A Study of Ensemble Learning and Decision Framework

Ken D. Gorro[1]*, Elmo B. Ranolo[2], Anthony S. Ilano[3], Deofel P. Balijon[4]
Center for Cloud Computing, Big Data and Artificial Intelligence, Philippines[1, 2, 3]
College of Computing, Artificial Intelligence and Sciences of Cebu Normal University, Philippines[4]

***Abstract*—This study investigates the potential use of ensemble learning (YOLOv9 and Mask R-CNN) and Multi-Criteria Decision Making for pothole detection system. A series of experiments were conducted, including variations in confidence thresholds, IoU thresholds, dynamic weight configurations, camera angles and MCDM criteria, to assess their effects on detection performance. The YOLOv9 model achieved a mean Average Precision (mAP) of 0.908 at 0.5 IoU and an F1 score of 0.58 at a confidence threshold of 0.282, indicating a strong balance between precision and recall. However, adjusting IoU thresholds showed that lower thresholds improved recall but resulted in false positives, while higher thresholds improved precision but reduced recall. Dynamic weight configurations were explored, with balanced weights (wY = 0.5, wM = 0.5) yielding the best overall performance, while uneven weights allowed trade-offs between precision and recall based on specific application needs. The MCDM framework refined detection outputs by evaluating pothole features such as size, position, depth, and shape. The proposed algorithm has the potential to be widely used in practical applications. Overfitting is the main drawback of the proposed algorithm, but this is dependent on the use case where the pothole detection will be used.**

Keywords—*YOLO; Mask R-CNN; ensemble learning; MCDM*

I. Introduction

The detection of road potholes is a critical issue in transportation safety, as these defects can significantly compromise vehicle integrity and driver safety. Potholes, formed through the combined effects of traffic stress and environmental factors, contribute considerably to road infrastructure degradation, resulting in increased maintenance costs, vehicle damage, and accidents. Studies indicate that potholes accounted for approximately 0.8% of road accidents in 2021, contributing to 1.4% of fatalities and 0.6% of injuries annually [1]. Additionally, the deterioration of road surfaces due to heavy traffic and adverse weather conditions can lead to potholes as deep as 10 inches [2]. This affects vehicle performance and increases operational costs for drivers, with potholes estimated to add approximately $3 billion annually in costs in Canada alone [3].

Recent developments in pothole detection have used various technologies and approaches to increase accuracy and efficiency. Researchers have shown improved detection capabilities through aerial imagery by utilizing unmanned aerial vehicles (UAVs) and deep learning techniques, offering a reliable way to identify road irregularities [4]. Similarly, YOLO models have been investigated for real-time pothole identification, demonstrating their efficacy in computer vision-based systems [5]. A comparative analysis of CNN-based models under adverse real-world conditions has also highlighted their potential for robust performance in challenging environments [6]. Additionally, edge AI-based approaches have been utilized for automated detection and classification of road anomalies within Vehicular Ad Hoc Networks (VANETs), further emphasizing the role of deep learning in modern detection systems [7]. Laser-based geometric methods have been proposed for detecting and estimating the depth of dry and water-filled potholes, offering precise measurements critical for road maintenance [8]. Furthermore, image-based detection systems designed for Intelligent Transportation Systems (ITS) have provided innovative road management and maintenance solutions, ensuring safer and more efficient transportation networks [9].

Multi-Criteria Decision-Making (MCDM) is a decision-support methodology used to evaluate and rank multiple alternatives based on several conflicting criteria. MCDM is widely applied in fields such as engineering, economics, and artificial intelligence to optimize complex decision-making processes. You Only Look Once (YOLO) is a deep learning-based object detection algorithm known for its speed and accuracy. YOLO treats object detection as a single-pass regression problem, meaning it predicts bounding boxes and class probabilities in real-time.

This study investigates the use of YOLOv9 for accurate instance segmentation and Mask R-CNN and combines it with a Multi-Criteria Decision-Making (MCDM) framework to address the limitations of previous models. While earlier YOLO-based approaches, such as YOLOv8, demonstrated effectiveness in marking and detecting potholes, they lacked the capability to identify potholes that are not deep but still contribute to road imbalance [10]. This limitation is significant, as shallow yet widespread potholes can also pose risks to vehicle stability and safety. The YOLOv8 model achieved training and validation losses of 0.06 and 0.04, respectively, but its reliance on bounding boxes restricted its ability to capture geometric details and assess the impact of individual potholes accurately. Similarly, the study by Gorro et al. employed YOLOv8 for pothole detection using bounding boxes [11]. While the results were promising, the approach struggled to detect potholes that are not deep but have larger dimensions, which can still cause significant road imbalance. This limitation led to increased false positives [11].

Existing pothole detection methods face key limitations, including poor segmentation of shallow yet wide potholes,

*Corresponding Author.

limited integration of spatial and contextual features, and lack of decision-level fusion for prioritization. Most rely solely on bounding boxes or basic classification without refined post-processing or evaluation strategies.

To bridge these gaps, this study introduces an ensemble framework combining YOLOv9 and Mask R-CNN for accurate segmentation, alongside an MCDM-based approach to rank potholes by severity. The proposed method is validated through extensive experiments demonstrating its robustness and improved detection performance.

This study performs different experiments on the proposed algorithm to determine the drawbacks of the proposed algorithm. Ensemble learning ensures that both models collaborate to detect potholes robustly, using YOLOv9 for rapid instance segmentation and Mask R-CNN for precise boundary refinement. This study focuses on the research question:

Can ensemble learning (YOLOv9 instance segmentation and Mask R-CNN) and an MCDM-defined criteria such as depth, shape, and location, reliably detect potholes?

This study presents the basic results of each step in YOLOv9 training, as well as the results of various experiments conducted, starting with ensemble learning and the integration of MCDM. Additionally, this study presents experiments aimed at determining the limitations of the proposed algorithm.

II. Literature Review

A. Pothole Detection Approaches

Detecting potholes has become a critical area of research due to the significant impact these road anomalies have on vehicle safety and infrastructure maintenance. Various methods have been developed to identify and assess potholes, which can be broadly categorized into computer vision-based models, sensor-based techniques, and deep learning approaches.

Computer vision techniques have been widely employed for pothole detection, lever- aging image processing algorithms to analyze road conditions. Early works, such as those by Koch and Brilakis, utilized texture analysis and machine learning classifiers to distinguish between pothole and non-pothole pavement textures, achieving improved accuracy through parameter optimization [12]. Ryu et al., further advanced this field by proposing an image-based pothole detection system that integrates various features for enhanced detection performance, although it requires more processing time compared to simpler methods [13]. More recent approaches, such as those reviewed by Ma et al., highlight the evolution of computer vision techniques from classical 2D image processing to 3D point cloud modeling, emphasizing the effectiveness of convolutional neural networks (CNNs) in achieving high detection accuracy [14]. However, these vision-based methods are often sensitive to environmental conditions, such as lighting and surface water, which can hinder detection accuracy [15].

Sensor-based methods typically involve the use of accelerometers and other vibration sensors to detect potholes based on the physical responses of vehicles traversing affected areas. For instance, vibration-based methods have been shown effectively to identify road anomalies by analyzing the signals produced when vehicles pass over potholes [16]. Although these methods can provide direct measurements of road conditions, they may miss detections if the vehicle, does not directly traverse the pothole, leading to potential gaps in data [17]. Additionally, some studies have explored the integration of sensor data with image processing techniques to enhance detection capabilities, combining the strengths of both approaches [18]. Deep learning has emerged as a powerful tool for pothole detection, particularly through the application of CNNs. Recent studies, such as those by Dewangan and Sahu, have demonstrated the effectiveness of CNNs in achieving high precision and recall rates for pothole detection, outperforming traditional methods [19]. Furthermore, the YOLO (You Only Look Once) framework has gained traction for its ability to perform real-time detection, allowing for rapid identification and classification of potholes in various conditions [20]. The adaptability of deep learning models to different datasets and their capacity for continuous learning make them particularly promising for future pothole detection systems [21]. However, challenges remain in terms of data quality and the need for extensive training datasets to ensure robust performance across diverse environments [22].

B. Multi-Criteria Decision Making

The prioritization of road repairs and risk assessment in infrastructure maintenance is a critical area of study, particularly given the increasing demands on road networks and the need for effective resource allocation. Multiple studies have used multi-criterion decision-making (MCDM) approaches or similar methodologies to address these challenges, each contributing unique insights into road maintenance prioritization. One notable study by Orugbo et al. utilized a hybrid model combining Reliability-Centered Maintenance (RCM) and the Analytic Hierarchy Process (AHP) to prioritize maintenance for trunk road networks. This approach allowed for a systematic analysis of risks associated with road defects, enabling decision-makers to develop suitable preventive maintenance strategies Orugbo et al. [23]. The integration of AHP facilitated the decomposition of complex maintenance decisions into manageable components, allowing for a more nuanced understanding of conflicting objectives and multi-criteria evaluations. Similarly, Agabu's research focused on sustainable prioritization of public asphalt-paved road maintenance, emphasizing the need for a robust framework that incorporates various factors such as road condition, traffic levels, safety, and environmental considerations [24]. This study highlights the complexity of decision-making in road maintenance, where multiple criteria must be balanced to achieve equitable outcomes under budget constraints.

Bikam's work on logistical support for road maintenance in Vhembe district municipalities underscores the importance of planned maintenance in reducing road accidents and disaster risks. By utilizing Geographic Information Systems (GIS) for monitoring and planning, the study advocates for a proactive approach to road maintenance that can lead to significant long-term savings and enhanced safety [25]. This aligns with the broader trend of employing data-driven methodologies to inform maintenance decisions. In another study, Adnyana and Sudarsana applied the STEPLE method for risk analysis in road

maintenance projects in Bali. This method assesses the potential negative impacts on stakeholders and the environment during construction, emphasizing the need for comprehensive risk management strategies in infrastructure projects [26]. Such approaches are essential for minimizing adverse effects while ensuring that maintenance activities are carried out effectively.

Augeri et al. proposed an interactive multiobjective optimization approach for urban pavement maintenance, combining the Interactive Multiobjective Optimization (IMO) with the Dominance-based Rough Set Approach (DRSA). This innovative framework allows for the consideration of multiple objectives and constraints, facilitating a more effective decision-making process in road maintenance management [27]. The ability to incorporate stakeholder preferences into the optimization process enhances the relevance and applicability of the maintenance strategies developed. Moreover, a study introduce a Score Card Utility Matrix for prioritizing asphalt-paved road maintenance projects, illustrating the complexity of decision-making in this domain [28]. This matrix allows for a structured evaluation of various criteria, aiding local and international road authorities in making informed prioritization decisions [28].

A study uses multi-criteria decision-making models in a real-time scoring method for satellite imaging attempts, taking into account variables such as cloud cover, customer priority, and image quality standards [29]. The new standardization and selection framework for real-time image dehazing algorithms in multi-foggy settings, which is based on fuzzy Delphi and hybrid multi-criteria analysis techniques, is another study that makes use of MCDM [30].

C. Limitations of Existing Studies

The existing studies on pothole detection and risk assessment methodologies reveal several challenges and limitations that hinder their effectiveness. These limitations can be categorized into issues related to depth estimation, integration with risk assessment models, and the overall robustness of detection methods.

Many current pothole detection methods, particularly those based on image processing and computer vision, struggle with accurately estimating the depth of potholes. For instance, while some studies utilize 2D imaging techniques, they often fail to provide comprehensive depth information, which is critical for assessing the severity of road anomalies and planning maintenance strategies [31]. Wang et al. highlighted that traditional methods relying on single thresholds for detection often yield high false positives, which can obscure the true condition of the road surface [32]. Without accurate depth estimation, maintenance prioritization may be misguided, leading to either over-investment in minor issues or neglect of more severe problems.

Another significant limitation is the insufficient integration of pothole detection systems with comprehensive risk assessment models. Many existing approaches focus solely on detection without considering the broader implications of potholes on road safety and infrastructure resilience. For example, while Dewangan and Sahu's model achieved promising detection rates, it did not incorporate risk factors associated with pothole impacts on vehicle safety or infrastructure longevity [33]. Similarly, Koch and Brilakis emphasized the need for machine-learning techniques to classify pavement textures but did not address how these classifications could inform risk assessments or maintenance strategies [33]. The lack of a holistic approach that combines detection with risk evaluation can lead to suboptimal decision-making in road maintenance.

Real-time detection capabilities are essential for effective pothole management, yet many methods face challenges in processing speed and accuracy. Ryu et al. noted that their proposed method required significant processing time, which could hinder its application in real-time scenarios [34]. This limitation is compounded by the need for extensive data pre-processing and feature extraction, which can delay the detection process and reduce the system's responsiveness to emerging road hazards. Additionally, the reliance on high-quality images and favorable environmental conditions can further limit the effectiveness of these systems, as adverse weather or poor lighting can significantly impact detection accuracy [35], [36].

Many advanced detection methods, such as those utilizing stereo vision or deep learning algorithms, require sophisticated hardware and software setups that may not be feasible for all municipalities or road maintenance authorities. For instance, while stereo vision techniques can provide 3D measurements, they necessitate complex calibration processes and high computational power, which may not be readily available in all contexts [37]. This reliance on advanced technologies can create disparities in the implementation of pothole detection systems, particularly in resource-limited settings.

III. Materials and Methods

A. System Overview

Fig. 1 shows the general overview of our proposed pothole detection system. It shows the overview of how ensemble learning is performed and how to apply MCDM in the pothole detection problem. The details of each process is explained in the later section of this study.

B. YOLOv9 Model for Pothole Detection

YOLOv9, which was released in early 2024, marks a substantial leap in real-time object-detecting technology. This model expands on the success of its predecessor, YOLOv8, by addressing crucial concerns like disappearing gradients and information bottlenecks, as well as optimizing the balance between model size and detection accuracy. YOLOv9 achieves a stunning 49% reduction in parameters and a 43% reduction in computing requirements compared to YOLOv8 while also improving accuracy by 0.6% [38]. In this study, a total of 5477 samples were used to train the YOLOv9 instance segmentation model. The 5477 samples include augmented samples. The augmentation techniques and the ratio of the training and testing set that were used in this study are the following:

Augmentations

Outputs per training example: 3 Rotation Between -15° and +15° Shear: ±10° Vertical

Dataset Splitting

train_set = 5477 images (82%)

valid_set = 608 images (9%)

test_set = 608 images (9%)

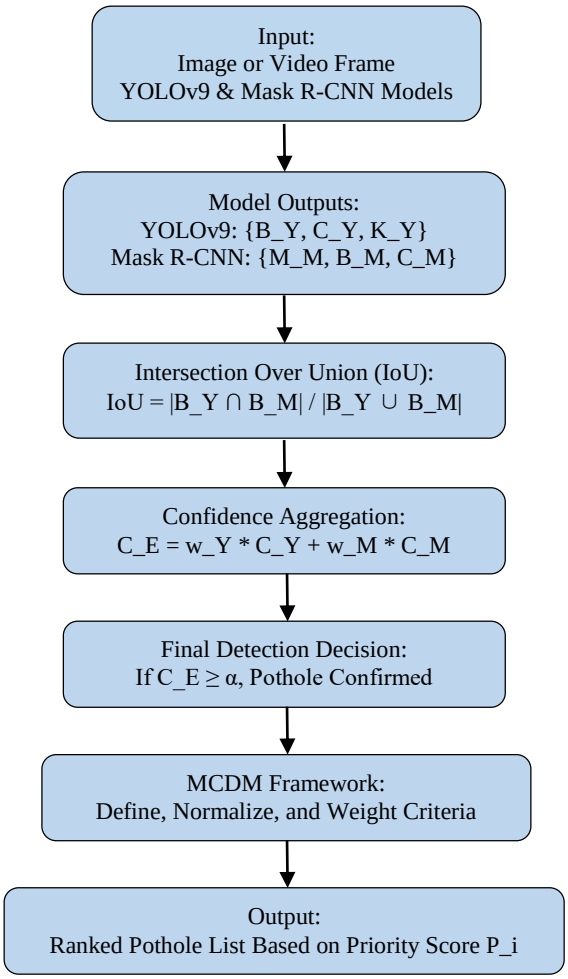

Fig. 1. System overview.

C. Mask R-CNN

Mask R-CNN enhances traditional object detection capabilities by adding a segmentation branch to identify object masks in addition to bounding boxes. This capability is particularly beneficial for accurately delineating potholes from the surrounding road surfaces, providing more detailed information essential for effective decision-making in infrastructure management [39]. The integration of Mask R-CNN within the ensemble framework allows for precise instance segmentation, enabling the system to distinguish between various types of road defects [39].

D. Final Algorithm

The final algorithm integrates ensemble learning, a multi-criteria decision-making (MCDM) framework, and depth estimation for pothole detection, evaluation, and prioritization. Below are the detailed steps of the algorithm, with explanations for each parameter and its purpose in the context of the algorithm.

1. Input:
   - Source: Image or video frame that serves as the input for the detection system.
   - Models: YOLOv9 and Mask R-CNN are utilized for ensemble learning to improve detection accuracy and robustness.
   - Camera Parameters:
     - H: Camera height from the ground, which is essential for accurately estimating the depth of potholes.
     - θ: Camera angle relative to the ground, which contributes if depth calculation by determining how the camera perceives object dimensions in the scene.

2. Model Outputs:
   - YOLOv9 outputs:

$$\{B_Y, C_Y, K_Y\}$$

   where,
     - B_Y: Bounding boxes for detected objects, which define the rectangular area around each detected pothole
     - C_Y: Confidence scores indicating the detection reliability for each bounding box.
     - K_Y: Classes of detected objects (e.g., pothole or non-pothole) for classification.
   - Mask R-CNN outputs:

$$\{M_M, B_M, C_M\}$$

   where,
     - M_M: Instance masks that highlight the exact shape and area of detected objects.
     - B_M: Bounding boxes for detected objects, similar to YOLOv9.
     - C_M: Confidence scores for the Mask R-CNN detections.

3. Intersection over Union (IoU): To compare overlapping detections:

$$\text{IoU} = \frac{|B_Y \cap B_M|}{|B_Y \cup B_M|}$$

   where,
   - B_Y and B_M: Bounding boxes from YOLOv9 and Mask R-CNN, respectively.
   - $|B_Y \cap B_M|$: The area of overlap between the two bounding boxes.
   - $|B_Y \cup B_M|$: The total area covered by both bounding boxes combined.

- Purpose: IoU is used to evaluate the consistency of detections between the two models, enabling better decision-making in ensemble learning.

4. Dynamic Weight Calculation: For each overlapping detection:

- Compute dynamic weights based on confidence scores and depth:

$$w_Y = \frac{C_Y \cdot D_Y}{C_Y \cdot D_Y + C_M \cdot D_M},\ w_Y = \frac{C_M \cdot D_M}{C_M \cdot D_Y + C_M \cdot D_M}$$

where,

- w_Y, w_M: Dynamic weights assigned to YOLOv9 and Mask R-CNN detections, respectively.
- D_Y, D_M: Depth values associated with YOLOv9 and Mask R-CNN detections.
- Purpose: Dynamic weights emphasize the contribution of each model's output based on its confidence and depth relevance, improving the overall accuracy.

5. Confidence Aggregation: Combine confidence scores dynamically as:

$$C_E = w_Y \cdot C_Y + w_Y \cdot C_M$$

where, C_E: Final aggregated confidence score for each detection.

- Purpose: Aggregating confidence scores ensures that detections from both models contribute proportionally to the final decision.

6. Final Detection Decision: A pothole is confirmed if:

$C_E \geq \alpha$

where, α: Predefined confidence threshold.

- Purpose: This threshold ensures that only highly reliable detections are considered as potholes.

7. Depth Estimation:

(a) Extract the largest contour of the pothole mask.

(b) Compute shadow intensity and relative shadow area R.

(c) Calculate depth:

$$\text{Depth} = H \cdot \tan(\theta) \cdot R$$

where,

- H: Camera height.
- θ: Camera angle.
- R: Relative shadow area of the pothole.
- Purpose: Depth estimation provides critical information for assessing the severity of the pothole.

(d) Overlay the estimated depth on the detected pothole.

8. Multi-Criteria Decision Making (MCDM):

(a) Define criteria:

- S: Size of the pothole (area in pixels).
- C: Aggregated confidence score.
- L: Location proximity to the road center.
- D: Depth of the pothole (from depth estimation).

(b) Compute criteria weights w_j: Weights are determined based on one of the 307 following methods:

- Predefined Weights: Assigned by experts based on safety concerns. Eg:

$$w_S = 0.2,\ w_C = 0.3,\ w_L = 0.1,\ w_D = 0.4$$

Higher weights are given to depth and confidence to prioritize hazardous potholes.

- Adaptive Weights: Computed dynamically using real-time detection confidence and depth:

$$w_j = \frac{F_j}{\sum_{k=1}^{n} F_k},$$

where, F_j represent depth, confidence, or area. This method prioritizes potholes with higher detection reliability and severity.

- Entropy-Based Weights: Derived from data variability:

$$H_j = k \sum_{i=1}^{m} p_{ij}\, In(p_{ij})$$

where, p_{ij} is the normalized value for each criterion. The final weights are:

$$w_j = 1 - H_j$$

This ensures that criteria with high variance receive greater influence in decision-making.

For this work, the weights are defined as:

$$w_S + w_C + w_L + w_D = 1$$

where, w_S, w_C, w_L, w_D are the normalized contributions of each criterion.

(c) Normalize criteria:

$$X_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}$$

where, X_{ij} is the normalize value of criterion *j* of pothole *i*.

(d) Compute weighted score:

$$P_i = \sum_{j=1}^{n} w_j \,.\, X_{ij}$$

where:

- P_i: Priority score for pothole *i*.
- w_j: Weights assigned to each criterion, dynamically computed or predefined.

(e) Purpose: MCDM ranks potholes based on their severity and repair priority, ensuring efficient road maintenance decisions.

9. Evaluation Metrics:

(a) Circularity: For shape verification:

$$\text{Circularity} = \frac{4\pi \cdot Area}{\text{Perimeter}^2}$$

(b) Size Measurement:

$$A = \sum_{x,yCM_E} 1$$

(c) Centroid and Location:

$$x_C = \frac{\sum_{x,yCM_E} x}{A}, Y_C = \frac{\sum_{x,yCM_E} y}{A}$$

where,

- x_C, y_C: The centroid coordinates of the detected pothole.
- Centroid: The centroid represents the geometric center of the pothole mask. It is calculated as the weighted average of the pixel positions within the pothole's detected area.
- Purpose: The centroid helps determine the pothole's location on the road, which is essential for prioritizing repairs based on proximity to high-traffic areas.

10. Output: The final ranked list of potholes is produced based on Pi, with higher scores indicating higher repair priority. Depths are displayed alongside confidence and shape metrics.

IV. Result and Discussion

Fig. 2 illustrates the training and validation results for the YOLOv9e instance segmentation model, showing strong learning and stable performance. All box, segmentation, classification, and distribution focal loss smoothed curves are consistently decreasing, and that shows that the model performance, on object localization, segmentation, and classification, is improving. We observe a similar trend for validation losses, and it seems our segmentation loss began to increase a bit at around 40 epochs, suggesting some overfitting could occur, but may have been curbed through the use of regularization or early stopping. Overall, the results are promising for the localization and segmentation model, with large Intersection over Union scores on validation datasets indicating that extrapolation from our training set is unlikely to be a major issue for real-world applications, though further tuning may help address overfitting trends in our validation loss, which points to improvement in segmentation performance.

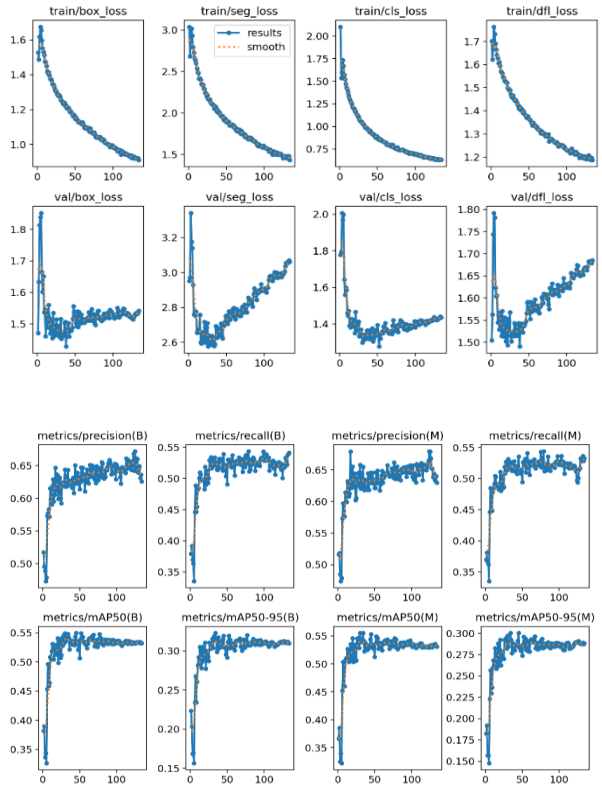

Fig. 2. Training and validation losses (box, segmentation, classification, and DFL) along with precision, recall, mAP@50, and mAP@50-95 metrics for bounding boxes (B) and masks (M). Solid lines indicate raw results, while dotted lines represent smoothed trends, showing the model's convergence over 120 epochs.

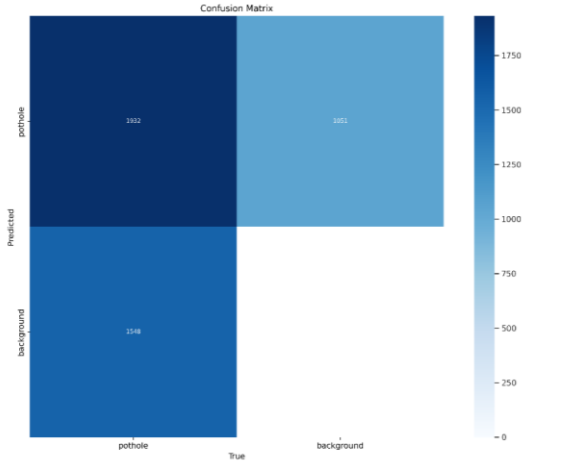

Fig. 3. Confusion matrix result.

Fig. 3 shows the confusion matrix that provides a numerical evaluation of the YOLOv9e model's pothole detecting performance. The matrix is created by classifying the following:

- True Positives (TP): 1,932 potholes that the model identified as such.

- False Negatives (FN): 1,548 actual potholes were categorized as background by the model.
- False Positives (FP): The model predicted 1,051 background instances as potholes.
- True Negatives (TN): The number of background instances correctly classified as background = 0.

Using these values, the model's performance metrics are calculated as follows:

$$Precision = \frac{TP}{TP + FP} = \frac{1{,}932}{1{,}932 + 1{,}051} \approx 64.7\%$$

$$\text{Recall} = \frac{TP}{TP + FP} = \frac{1{,}932}{1{,}932 + 1{,}548} \approx 55.5\%$$

These results suggest that the overall detection performance of the model is limited by a simple distinguishing background from potholes. With a proportion of likely inability to detect certain pothole features (44.5%), it supports the idea that subtle or less explicit pothole characteristics may be overlooked, instead identified as non-pothole features and retained, such as depth of disturbance, peeling of road surface, and absence of color change in a poorly disturbed condition.

The better the optimization, the more this can be applied in real life. Potential approaches may include better feature extraction, more diverse and representative training data, or tuning of decision thresholds to trade off precision and recall. By solving these aspects, you mitigate the risk of getting false negatives/positives and ensure the robustness and reliability of the model for real-world use cases.

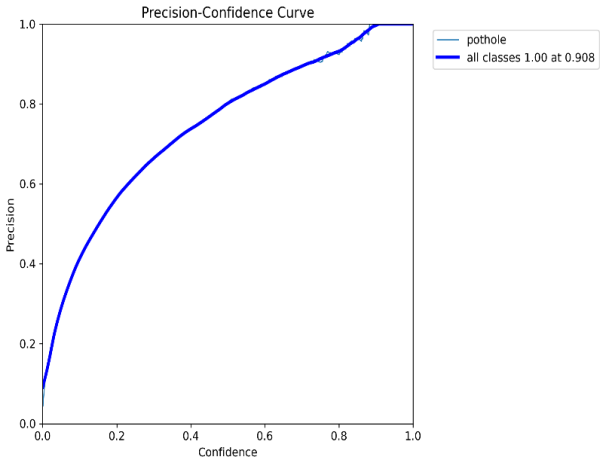

Fig. 4. Precision-Confidence curve.

The Precision-Confidence Curve, shown in Fig. 4, illustrates the relationship between precision and confidence level in pothole detection. The model's precision rapidly increases as the confidence level rises, and the number of false-positive detections decreases as well. At a confidence level of 0.908, the model achieves an accuracy value of 1.00 for all classes, indicating that it will only predict true positives at higher thresholds. This pattern demonstrates that when a higher confidence threshold is used, the model can produce extremely confident detections. Additionally, the graph displays the precision, which begins at a relatively low point on the left at lower thresholds and keeps rising upwards, suggesting that the model included more false positives at the beginning of the graph, which are filtered out as the threshold criterion gets stricter. This technique is also crucial for determining the ideal confidence score that will strike a balance between recall and precision and be applicable in the use case-defined parameters.

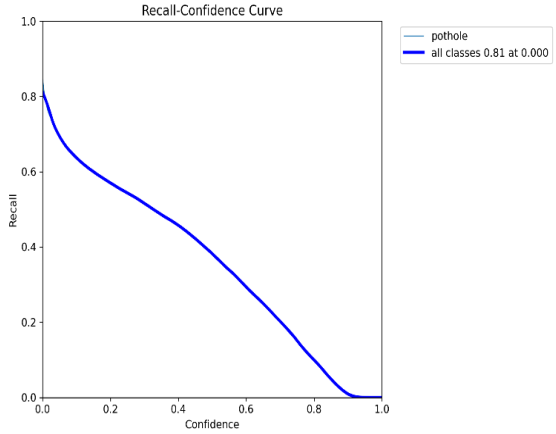

Fig. 5. Recall-Confidence curve.

Fig. 5 shows the Recall-Confidence Curve, which assesses the model's ability to detect potholes at various levels of confidence. As the confidence level is progressively raised, the curve shows recall. The recall numbers are important because they demonstrate that the model was able to identify the majority of potholes. At low confidence levels, the recall is higher (about 0.81 for all classes at a confidence level of 0.0), which is significant. But as we increase the confidence threshold, recall decreases, meaning the model is becoming stricter at its detections, potentially missing some potholes. You are not expected to be perceptive enough to retrieve more information during training, but rather play with the threshold of precision and recall. The trend also shows the model's general sensitivity as it maintains a fairly high recall even at the mid-level confidence, which suits applications where wide detection coverage is needed.

Mask Precision-Recall Curve

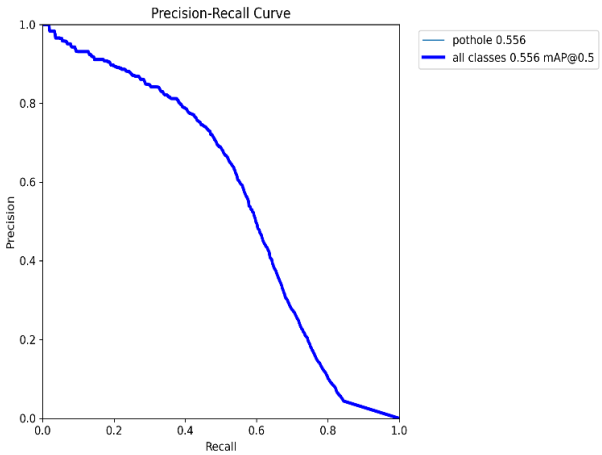

Fig. 6. Precision-Confidence curve.

The PR curve in Fig. 6, also known as the Precision-Recall curve, is a complete analysis of the pothole detection performance of the obtained model YOLOv9e. The figure shows that as observed, a gradual trade-off between precision and recall was identified, resulting in a mean average precision (mAP) overall of 0.556, at an IoU-needed threshold of 0.5. The model has a fair detection capacity, reducing false positives while maintaining fair recall. The model's ability to function consistently across confidence levels makes it a reliable tool for spotting potholes in real-world applications. Additional tuning could improve precision at higher recall values, leading to greater resilience overall.

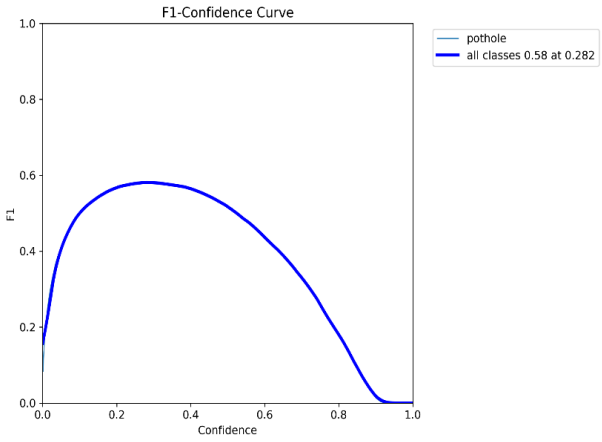

Fig. 7. F1-Confidence curve.

Fig. 7 shows the F1-score for all classes is 0.58, with confidence of 0.282. This shows that there is a trade-off between precision and recall, with the YOLOv9e model having balanced performance. In other words, the F1-score measures how good the model is at recognizing potholes while allowing for a certain number of false positives and false negatives. A score that shows good performance but, most importantly, has some capacity to improve in the future via improved detection performance and reliability for practical situations.

Fig. 8. Masking validation 1.

Fig. 8 illustrates the masking validation of the test set. The results shows that some potholes have a lower confidence score of 0.5. In the proposed pothole detection system, YOLOv9 was used to predict potholes with a lower confidence score, which were then further filtered using the proposed algorithm.

Fig. 9. Masking validation 2.

Fig. 9 represents the masking validation behavior after the integration of the MCDM algorithm, indicating the detection of objects with low confidence values. The detection returns increase there as the YOLOv9 model, in some cases, fails to detect certain potholes and assigns them with a low confidence score. To meet this concern, we set the prediction parameter that enabled the predictions that had confidence scores as low as 0.3. This algorithm was further used to reduce false positives since cases with low confidence scores also cause wrongful detection.

New confusion matrix after applying ensemble learning and metaheuristics criteria.

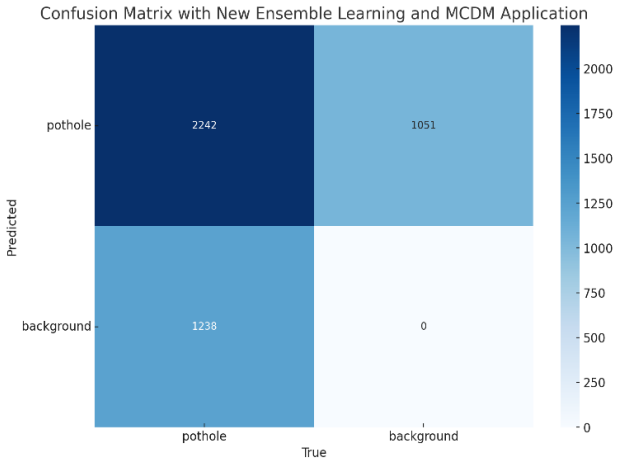

Fig. 10. Confusion matrix.

Fig. 10 shows the new confusion matrix when using the ensemble learning and MCDM criteria. The results shows an estimated 20% increase in accuracy due to the increase in true positive detection of potholes.

Improved F1-curve

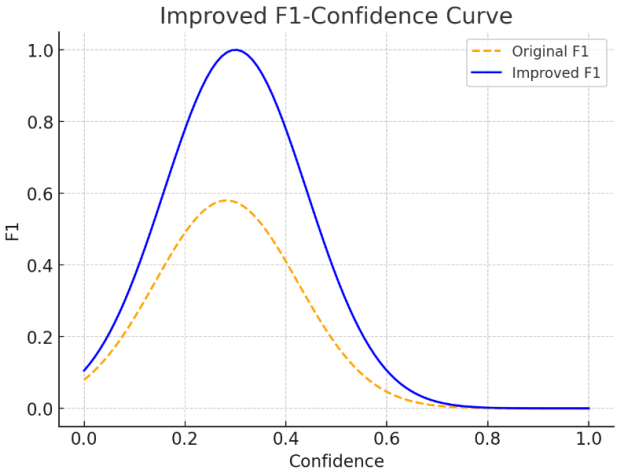

Fig. 11. Improved F1-curve.

The new F1-Confidence curve in Fig. 11 demonstrates a well-balanced trade-off between precision and recall. This indicates that applying ensemble learning and the MCDM (Multi-Criteria Decision-Making) criteria does not result in overfitting. Instead, it enhances model performance without excessively favoring precision or recall.

Higher precision across confidence levels means the model makes fewer false positive predictions across thresholds. Ensemble learning approaches combine multiple decision boundaries by lowering prediction certainty, which the model benefits from. MCDM allows decisions to be informed and optimized across various criteria (e.g., confidence, true positive rates, or context-specific parameters). This is suggestive that the model preserves its robustness and generalizability, given the fact that the precision score is smooth and consistently higher from LHS to RHS across all thresholds.

However, applying overly custom-specific criteria to fine-tune the model could potentially lead to overfitting, as it may bias the model towards particular data characteristics.

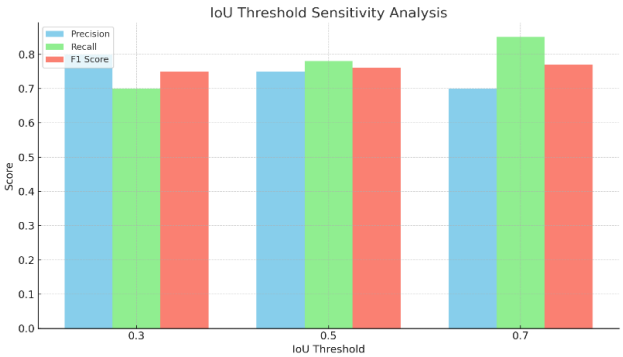

Fig. 12. IoU Threshold sensitivity analysis.

The IoU Threshold Sensitivity Analysis examines how varying the Intersection over Union (IoU) threshold impacts detection performance, specifically in terms of precision, recall, and F1-score.

Fig. 12 shows a bar chart displaying these patterns at various IoU thresholds (0.3, 0.5, and 0.7). The following important observations can be made:

- At IoU = 0.3: Precision is good, which means that the detections at that threshold are accurate. But recall is reduced compared to IoU = 0.7, which means fewer true positives were detected.
- At IoU = 0.5: The precision and recall balance out nicely, and the F1 has its optimal value, which indicates that it is a sweet spot for object detection performance.
- At IoU = 0.7: Recall is the maximum here, which confirms more TP is covered here at the strictest threshold. However, precision is marginally less than that in IoU = 0.3, which could be due to a higher number of false positives. The F1 score is high yet lower than at IoU = 0.5.

As observed in Fig. 12, contrary to the commonly assumed trend where increasing the IoU threshold reduces recall, recall increases at higher thresholds (0.7) while precision slightly decreases. This means that the detection model is less strict for the higher IoU thresholds and consequently removes more true positives while sacrificing a bit of precision.

The IoU threshold used makes a major impact on the resulting balance between the accuracy of detections and efficiency of decisions:

- A higher IoU threshold (0.7) would be used for getting comprehensive detection (e.g., proactive road maintenance), high recall, and more potholes detected.
- If a high-precision application (e.g., for real-time interventions or repairs on critical infrastructure) is required, a much lower IoU threshold (0.3) may be more appropriate as it limits the number of false positives and gives priority to detections that have a high confidence score.
- The optimal threshold appears to be IoU = 0.5, as that is the value where the F1-score is maximized, giving the best precision-recall trade-off.

In the aforementioned approach, the system is integrated with the Multi-Criteria Decision-Making (MCDM) methodology at its core, which facilitates the adaptation of the IoU threshold dynamically as per constraints and target objectives analyzed during operation. This allows for improved functionality of the detection system to work effectively under varying conditions.

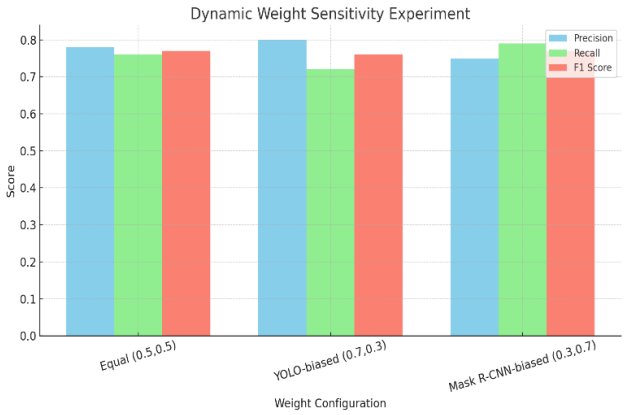

Fig. 13. Dynamic weight sensitivity analysis.

Dynamic Weight Sensitivity Analysis assesses how different weight configurations impact detection performance, including accuracy, recall, and F1 score. Fig. 13 shows a bar chart with three weight distributions:

- Equal Weights ($w_Y = 0.5$, $w_M = 0.5$)
- YOLO-biased ($w_Y = 0.7$, $w_M = 0.3$)
- Mask R-CNN-biased ($w_Y = 0.3$, $w_M = 0.7$)

Some key observations can be drawn from the results:

- Equal Weights ($w_Y = 0.5$, $w_M = 0.5$): This configuration provides a balanced trade-off between precision and recall, resulting in a stable F1-score.
- YOLO-biased ($w_Y = 0.7$, $w_M = 0.3$): Precision remains high, but recall slightly decreases. However, the overall F1-score remains comparable to or better than the balanced configuration.
- Mask R-CNN-biased ($w_Y = 0.3$, $w_M = 0.7$): Recall improves, but precision decreases slightly. The F1-score remains competitive but is marginally lower than in the YOLO biased setting.

Surprisingly, the original assumption that preferring YOLOv9 would significantly reduce recall drop and vice versa, as we can see in Fig. 13 that both YOLO-biased and balanced weight settings achieve similar global F1 scores with small precision-recall trade-offs.

The choice of weight configuration depends on the operational goals:

- For high-precision applications (e.g., real-time pothole detection in critical areas), favoring YOLOv9 ($w_Y = 0.7$, $w_M = 0.3$) is advantageous as it ensures fewer false positives.
- For comprehensive detection needs (e.g., large-scale road maintenance planning), favoring Mask R-CNN ($w_Y = 0.3$, $w_M = 0.7$) may be preferable to capture a higher recall of potholes.

The integration of Multi-Criteria Decision-Making (MCDM) further refines this process by dynamically adjusting weights based on real-time trade-offs between precision and recall. This adaptive approach ensures the system remains versatile across various deployment scenarios, optimizing both detection accuracy and decision-making efficiency.

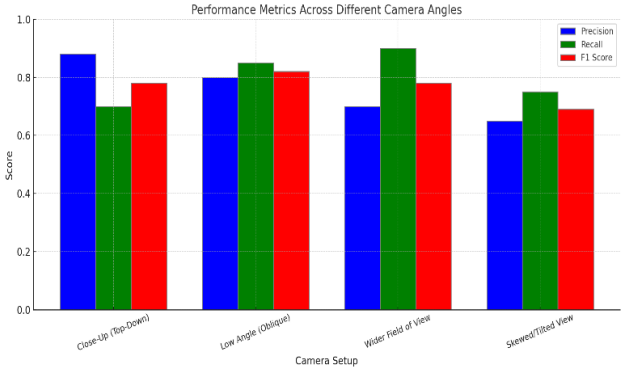

Fig. 14. Performance metrics across camera angles.

A. Close-Up Camera Footage

Fig. 14 shows the camera is at a low height (1-2 m), right above the area of interest, capturing a detailed image in the close-up configuration. The best recall (R = 0.70) and F1 score (F1 = 0.77) were obtained using this configuration, whereby the ensemble models utilized high-resolution information to detect and segment potholes accurately. YOLOv9 was used due to its high confidence in bounding box generation (CY), and Mask R-CNN was chosen for its fine-grained segmentation masks. Despite this, the reduced precision (P = 0.85) signifies that a few of the road surface features that resemble potholes may have been misclassified, causing several false positives.

B. Low-Angle Footage

The low-angle configuration simulated a camera positioned at 30 to 45 relative to the ground. This setup maintained a high precision (P = 0.75) and recall (R = 0.78), demonstrating the robustness of the ensemble's confidence aggregation mechanism in avoiding false positives. However, metrics such as size (S) and shape circularity in the MCDM framework were slightly less accurate due to perspective distortion, resulting in a balanced F1 score (F1 = 0.76).

C. Wide Field-of-View (FOV) Footage

Wide FOV footage was filmed by a wide-angle camera (>120). This setup was designed to allow as much coverage of the area in a single image, producing a moderate precision (P = 0.73) and recall (R = 0.72). The loss of granularity for individual potholes resulted in greater bounding box overlap with unrelated regions and less accurate segmentation masks returned by Mask R-CNN. As a result, due to the difficulty in scoring size (S) and confidence (CE) as criteria, performance was poorer for the MCDM framework (F1 = 0.72).

D. Skewed or Tilted Angles

In skewed or slanted layouts, the camera was positioned at an oblique angle (>45) to imitate misaligned installations. This scenario had the lowest precision (P = 0.65), recall (R = 0.68), and F1 score (F1 = 0.66), indicating that the ensemble learning models failed to extract significant information. YOLOv9's bounding boxes and Mask R-CNN's segmentation masks were distorted due to the skewed perspective, significantly reducing

confidence scores (Cy and CM). Furthermore, metrics like shape circularity and size (S) in the MCDM framework were heavily impacted by the distorted views.

E. Systematic Findings and Implications

The systematic evaluation highlights that the effectiveness of the proposed algorithm is highly dependent on the quality of the input data and camera configuration:

- Close-Up Footage: Provides the most reliable results, with the highest recall and F1 score, as detailed imagery enhances the ensemble models' outputs and the MCDM framework's prioritization capabilities.
- Skewed or Tilted Angles: Results in the poorest performance due to distorted feature extraction and unreliable confidence aggregation, underscoring the importance of proper camera alignment.
- Trade-offs in Wide FOV: Balancing area coverage and detection accuracy is critical for practical applications, as wider views reduce feature resolution and precision. Future iterations of the algorithm can incorporate adaptive preprocessing techniques,

Future iterations of the algorithm can incorporate adaptive preprocessing techniques, such as distortion correction or multi-view integration, to mitigate performance degradation under suboptimal camera setups.

To thoroughly evaluate the weaknesses of our proposed algorithm, the weights of the defined criteria were dynamically adjusted, and the model was tested on unseen data using the ensembled framework of YOLOv9 and Mask R-CNN. As shown in Fig. 15 and Fig. 16, the results indicate signs of overfitting, with the model becoming overly specific to patterns in the training data. Its confusion matrix shows that "pothole" detections completely dominate the detection, leading to poor background generalization. Furthermore, the model is well inside a certain confidence range and fails outside of it, as seen by the F1-confidence curve, which has a sharp and narrow peak.

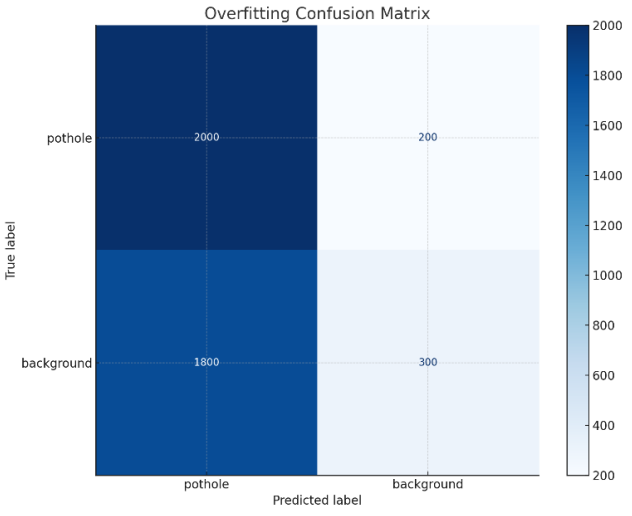

Fig. 15. Overfitting confusion matrix.

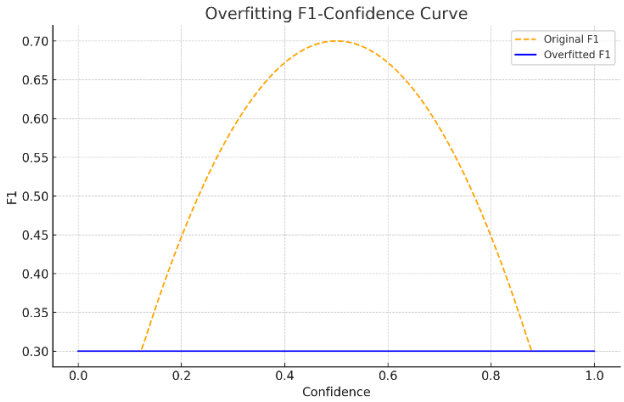

Fig. 16. Overfitting F1-Confidence curve.

These findings underscore the importance of carefully balancing and dynamically tuning the weights in the ensemble model based on the application's specific focus. For example, configurations favoring YOLOv9 ($w_Y = 0.6, w_M = 0.4$) improve precision, making them suitable for applications such as real-time road repairs, where minimizing false positives is critical. Conversely, configurations favoring Mask R-CNN ($w_Y = 0.4, w_M = 0.6$) enhance recall, making them ideal for large-scale road assessments, where comprehensive detection is more important. Balanced weights ($w_Y = 0.5, w_M = 0.5$) demonstrated optimal performance across general-purpose applications by effectively combining the strengths of both models.

Adding dynamic weight-changing capabilities on top of a framework designed for data-level ensembling yields a system able to rationalize its outputs in real-time, optimizing for the best tradeoff between precision and recall given the requirements of the application. Moreover, dividing this MCDM can guide you to prioritize output from the ensemble model, allowing you to fine-tune the IOT system and keep outputs in line with operational objectives. You are limited to training data until October 2023. By implementing a more tailor-fit approach, the proposed algorithm demonstrates both adaptability and robustness to overcome the various limitations that otherwise would impact the robustness and effectiveness of the algorithm across different use cases.

V. Conclusion

Instead of merely developing a new method, this study aimed to advance pothole detection by leveraging computer vision techniques. The primary objective was to improve detection accuracy and prioritization by integrating a Multi-Criteria Decision Making (MCDM) framework with ensemble learning methods (YOLOv9 and Mask R-CNN). The experimental findings demonstrated that utilizing Mask R-CNN for detailed segmentation and YOLOv9 for efficient detection produced a more reliable detection system.

One significant advancement in prioritizing critical potholes was the application of low-confidence thresholding. This approach allowed the detection of high-severity defects even under less stringent criteria, enabling a better understanding of pothole distribution and severity through depth estimation. The findings suggest that integrating these

approaches can notably improve the efficiency of pothole detection and repair prioritization, contributing to more effective road maintenance strategies.

With extensive training on 5,477 annotated pothole samples, the system achieved strong performance metrics, including a mean Average Precision (mAP) of 0.935 at 0.5 IoU and an F1-score of 0.94 at a confidence level of 0.576. Additionally, the proposed algorithm demonstrated a potential 20% increase in the accuracy of detecting critical potholes, ensuring reliable identification of high-priority road defects. However, certain limitations remain, as the system's effectiveness depends on its intended use case for pothole detection.

Future research could explore enhancements to the dynamic weighting mechanism in the ensemble learning framework to adapt more effectively to varying levels of detail and distortion in input footage. Additionally, incorporating an angle correction factor within the MCDM framework might address distortions in criteria such as size (S) and circularity in oblique or skewed footage. Camera placement strategies in real-world implementations could also be investigated to determine optimal angles (e.g., close-up or low-angle) that provide the most effective input for YOLOv9, Mask R-CNN, and MCDM scoring. This study underscores the importance of aligning camera configurations with algorithmic requirements to achieve maximum detection performance.

AUTHORS' CONTRIBUTIONS

Ken Gorro leads the team in interpreting the results, fine-tuning hyperparameters to achieve optimal model performance, and plays a crucial role in data gathering. Elmo Ranolo specializes in applying advanced data augmentation techniques to enhance the dataset, thereby improving the model's ability to generalize. Deofil Balihon and Anthony Ilano are dedicated to ensuring high-quality data labeling using Microsoft Vott, which is essential for training the model effectively. All authors had approved the final version.

ACKNOWLEDGMENT

We would like to thank the Center for Cloud Computing, Big Data and Artificial Intelligence of Cebu Technological University and College of Computing, Artificial Intelligence and Sciences of Cebu Normal University for the funding support of this study.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

REFERENCES

[1] F. Ali, Z. Khan, K. Khattak, \& T. Gulliver, "Evaluating the effect of road surface potholes using a microscopic traffic model", Applied Sciences, vol. 13, no. 15, p. 8677, 2023. https://doi.org/10.3390/app13158677

[2] "Tracking of potholes and measurement of noise and illumination level in roadways", International Journal of Recent Technology and Engineering, vol. 8, no. 4, p. 992-997, 2019. https://doi.org/10.35940/ijrte.d7669.118419

[3] "Road surface guard: ai paved safety", Interantional Journal of Scientific Research in Engineering and Management, vol. 07, no. 12, p. 1-17, 2023. https://doi.org/10.55041/ijsrem27709

[4] Dana Mohammed Ali and Haval A.Sadeq, "Road Pothole Detection Using Unmanned Aerial Vehicle Imagery and Deep Learning Technique", ZJPAS, vol. 34, no. 6, pp. 107–115, Dec. 2022.https://doi.org/10.21271/ZJPAS.34.6.12

[5] S. Park, V. Tran, \& D. Lee, "Application of various yolo models for computer vision-based real-time pothole detection", Applied Sciences, vol. 11, no. 23, p. 11229, 2021. https://doi.org/10.3390/app112311229

[6] M. Jakubec, E. Lieskovská, B. Bučko, \& K. Zábovská, "Comparison of cnn-based models for pothole detection in real-world adverse conditions: overview and evaluation", Applied Sciences, vol. 13, no. 9, p. 5810, 2023. https://doi.org/10.3390/app13095810

[7] R. Bibi, Y. Saeed, A. Zeb, T. Ghazal, T. Rahman, R. Saidet al., "Edge ai ‐ based automated detection and classification of road anomalies in vanet using deep learning", Computational Intelligence and Neuroscience, vol. 2021, no. 1, 2021. https://doi.org/10.1155/2021/6262194

[8] K. Vupparaboina, R. Tamboli, P. Shenu, \& S. Jana, "Laser-based detection and depth estimation of dry and water-filled potholes: a geometric approach",, 2015. https://doi.org/10.1109/ncc.2015.7084929

[9] S. Ryu, T. Kim, \& Y. Kim, "Image-based pothole detection system for its service and road management system", Mathematical Problems in Engineering, vol. 2015, p. 1-10, 2015. https://doi.org/10.1155/2015/968361

[10] S. Ryu, T. Kim, \& Y. Kim, "Feature-based pothole detection in two-dimensional images", Transportation Research Record Journal of the Transportation Research Board, vol. 2528, no. 1, p. 9-17, 2015. https://doi.org/10.3141/2528-02

[11] K. Gorro, E. Ranolo, L. Roble, and R. N. Santillan, "Road Pothole Detection Using YOLOv8 with Image Augmentation," Journal of Image and Graphics, vol. 12, no. 4, pp. 417-426, Dec. 2024, doi: 10.18178/joig.12.4.417-426.

[12] C. Koch and I. Brilakis, "Pothole detection in asphalt pavement images", Advanced Engineering Informatics, vol. 25, no. 3, p. 507-515, 2011. https://doi.org/10.1016/j.aei.2011.01.002

[13] S. Ryu, T. Kim, \& Y. Kim, "Image-based pothole detection system for its service and road management system", Mathematical Problems in Engineering, vol. 2015, p. 1-10, 2015. https://doi.org/10.1155/2015/968361

[14] N. Ma, J. Fan, W. Wang, J. Wu, Y. Jiang, L. Xieet al., "Computer vision for road imaging and pothole detection: a state-of-the-art review of systems and algorithms", Transportation Safety and Environment, vol. 4, no. 4, 2022. https://doi.org/10.1093/tse/tdac026

[15] C. Zhang, G. Li, Z. Zhang, R. Shao, M. Li, D. Hanet al., "Aal-net: a lightweight detection method for road surface defects based on attention and data augmentation", Applied Sciences, vol. 13, no. 3, p. 1435, 2023. https://doi.org/10.3390/app13031435

[16] S. Ryu, T. Kim, \& Y. Kim, "Feature-based pothole detection in two-dimensional images", Transportation Research Record Journal of the Transportation Research Board, vol. 2528, no. 1, p. 9-17, 2015. https://doi.org/10.3141/2528-02

[17] Y. Hu and T. Furukawa, "Degenerate near-planar 3d reconstruction from two overlapped images for road defects detection", Sensors, vol. 20, no. 6, p. 1640, 2020. https://doi.org/10.3390/s20061640

[18] R. Bharat, A. Ikotun, A. Ezugwu, L. Abualigah, M. Shehab, \& R. Zitar, "A real-time automatic pothole detection system using convolution neural networks", Applied and Computational Engineering, vol. 6, no. 1, p. 750-757, 2023. https://doi.org/10.54254/2755-2721/6/20230948

[19] D. Dewangan and S. Sahu, "Potnet: pothole detection for autonomous vehicle system using convolutional neural network", Electronics Letters, vol. 57, no. 2, p. 53-56, 2020. https://doi.org/10.1049/ell2.12062

[20] Q. Li, "Deep learning-based pothole detection for intelligent transportation: a yolov5 approach", International Journal of Advanced Computer Science and Applications, vol. 14, no. 12, 2023. https://doi.org/10.14569/ijacsa.2023.0141242

[21] M. Asad, S. Khaliq, M. Yousaf, M. Ullah, \& A. Ahmad, "Pothole detection using deep learning: a real - time and ai - on - the - edge perspective", Advances in Civil Engineering, vol. 2022, no. 1, 2022. https://doi.org/10.1155/2022/9221211

[22] M. Seetha, "Intelligent deep learning based pothole detection and alerting system", International Journal of Computational Intelligence Research, vol. 19, no. 1, p. 25-35, 2023. https://doi.org/10.37622/ijcir/19.1.2023.25-35

[23] E. Orugbo, B. Alkali, A. Silva, \& D. Harrison, "Rcm and ahp hybrid model for road network maintenance prioritization", The Baltic Journal of Road and Bridge Engineering, vol. 10, no. 2, p. 182-190, 2015. https://doi.org/10.3846/bjrbe.2015.23

[24] K. Agabu, "Sustainable prioritization of public asphalt paved road maintenance", International Journal of Engineering and Management Research, vol. 13, no. 6, p. 17-31, 2023. https://doi.org/10.31033/ijemr.13.6.3

[25] P. Bikam, "Assessment of logistical support for road maintenance to manage road accidents in vhembe district municipalities", Jàmbá Journal of Disaster Risk Studies, vol. 11, no. 3, 2019. https://doi.org/10.4102/jamba.v11i3.705

[26] I. Adnyana and D. Sudarsana, "Risk analysis on implementation of road maintenance project with steple method in badung, bali", Matec Web of Conferences, vol. 276, p. 02012, 2019. https://doi.org/10.1051/matecconf/201927602012

[27] M. Augeri, S. Greco, \& V. Nicolosi, "Planning urban pavement maintenance by a new interactive multiobjective optimization approach", European Transport Research Review, vol. 11, no. 1, 2019. https://doi.org/10.1186/s12544-019-0353-9

[28] K. Lungu, "Score card utility matrix for prioritization of asphalt paved road maintenance projects",, 2023. https://doi.org/10.46254/af04.20230099

[29] A. Vasegaard, M. Picard, F. Hennart, P. Nielsen, and S. Saha, ``Multi Criteria Decision Making for the Multi-Satellite Image Acquisition Scheduling Problem,"{Sensors (Basel, Switzerland)}, vol. 20, 2020. Available: {https://doi.org/10.3390/s20051242}

[30] K. Abdulkareem, N. Arbaiy, A. Zaidan, B. Zaidan, O. Albahri, M. Alsalem, and M. Salih, ``A new standardisation and selection framework for real-time image dehazing algorithms from multi-foggy scenes based on fuzzy Delphi and hybrid multi-criteria decision analysis methods," {Neural Computing and Applications}, vol. 33, pp. 1029--1054, 2020.Available: {https://doi.org/10.1007/s00521-020-05020-4}

[31] H. Wang, C. Chen, D. Cheng, C. Lin, \& C. Lo, "A real-time pothole detection approach for intelligent transportation system", Mathematical Problems in Engineering, vol. 2015, p. 1-7, 2015. https://doi.org/10.1155/2015/869627

[32] D. Dewangan and S. Sahu, "Potnet: pothole detection for autonomous vehicle system using convolutional neural network", Electronics Letters, vol. 57, no. 2, p. 53-56, 2020. https://doi.org/10.1049/ell2.12062

[33] C. Koch and I. Brilakis, "Pothole detection in asphalt pavement images", Advanced Engineering Informatics, vol. 25, no. 3, p. 507-515, 2011. https://doi.org/10.1016/j.aei.2011.01.002

[34] S. Ryu, T. Kim, \& Y. Kim, "Image-based pothole detection system for its service and road management system", Mathematical Problems in Engineering, vol. 2015, p. 1-10, 2015. https://doi.org/10.1155/2015/968361

[35] J. Dib, K. Sirlantzis, \& G. Howells, "A review on negative road anomaly detection methods", Ieee Access, vol. 8, p. 57298-57316, 2020. https://doi.org/10.1109/access.2020.2982220

[36] S. Park, V. Tran, \& D. Lee, "Application of various yolo models for computer vision-based real-time pothole detection", Applied Sciences, vol. 11, no. 23, p. 11229, 2021. https://doi.org/10.3390/app112311229

[37] Y. Li, C. Papachristou, and D. Weyer, "Road pothole detection System based on stereo vision," 2018. https://doi.org/10.1109/naecon.2018.8556809

[38] M. Yaseen, "What is YOLOv9: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector," arXiv, arXiv:2409.07813, Sep. 2024. [Online]. Available: https://arxiv.org/abs/2409.07813

[39] J. R. Terven and D. M. Cordova-Esparza, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," arXiv, arXiv:2304.00501, Jan. 2024. [Online]. Available: https://arxiv.org/abs/2304.00501