International Journal of Research Publication and Reviews

YOLOv9-Based Pothole Detection: Enhancing Road Safety through Deep Learning

Gopal Singh S[1], Fardeen Feroz Khan[2], Putikam Yashwanth[3]

[1,2,3] Computer Science Engineering with Spl. In AI/ML (of Affiliation) Vellore Institute of Technology, Chennai, India
[1]gopal.singh.rajput.3256@gmail.com, [2]fardeenferozkhan16@gmail.com, [3]yashwanthputikam@gmail.com

ABSTRACT:

Ensuring road safety and maintaining a transportation system relies heavily on accurate detection of potholes. However conventional methods, like inspections or laser-based systems are often laborious time consuming and costly. That's where deep learning techniques, convolutional neural networks (CNNs) come into play as a solution for automating pothole detection. Among the CNN based object detection algorithms, YOLO (You Look has emerged as a leading choice due to its real time performance and high accuracy. The latest version of YOLO called YOLOv9 takes these capabilities further making it a more attractive option for detecting potholes. In this study we propose a method for pothole detection using YOLOv9. By leveraging the real time capabilities of YOLOv9, our method achieves accurate pothole detection in rainy conditions.

We train the YOLOv9 model using a custom dataset comprising images of potholes to ensure its effectiveness in real world scenarios. To evaluate the performance of our trained model we use a dataset. Compare it against other pothole detection methods as well as different versions of YOLO and other neural network models. Our proposed method demonstrates accuracy at 82.7% with a 48.2ms inference time. Additionally it provides measurements of depth. Width, for each detected pothole. These findings demonstrate how YOLOv9 is highly effective, in detecting potholes in time. This breakthrough opens the door, for implementing YOLOv9 in applications to improve road safety and manage infrastructure efficiently.

Keywords: Pothole, YOLO, Object Detection, Mean Average Precision (mAP)

1. Introduction

Potholes, ubiquitous on roadways, present a formidable challenge to infrastructure integrity, vehicle maintenance, and overall road safety. The conventional methodologies for pothole detection, reliant on visual inspections and manual reporting, are characterized by inefficiencies—being time-intensive, laborious, and susceptible to human error. To transcend these limitations, the exploration of machine learning and computer vision techniques for automated pothole detection has become imperative. The You Only Look Once version 9 (YOLOv9) stands as a state-of-the-art object detection algorithm, demonstrating notable efficacy in a variety of real-world applications. Its real-time capabilities, accuracy, and adaptability position it as a compelling candidate for pothole detection. Traditional deep neural networks suffer from information loss during propagation. YOLOv9 introduces Programmable Gradient Index (PGI), a mechanism that addresses this issue.

PGI allows for the preservation of crucial information throughout the network, leading to more reliable gradient generation and, consequently, improved model convergence and performance.

The Multi-layer Perceptron (MLP), Convolutional Neural Network (CNN) can handle data in more dimensions. Due to its larger network depth and widespread application in the field of picture data, CNN is considered a sort of Deep Neural Network. [1]

This research endeavour is dedicated to systematically investigating the effectiveness of YOLOv9 in predicting pothole occurrences and subsequently developing a robust pothole detection system. The motivation behind this project extends beyond technical curiosity; it aligns with a broader commitment to advancing road safety and operational efficiency. By leveraging YOLOv9's capabilities, we seek to not only enhance our understanding of potholes but also contribute to the development of a sophisticated detection system, thereby fostering safer roadways for all stakeholders involved. This exploration represents a convergence of technological innovation and a humane dedication to addressing challenges that impact the daily experiences of road users.

II. Related work

In the ongoing quest for safer and more resilient roadways, the intersection of deep learning and computer vision, as seen through the lenses of YOLOv5, YOLOv4, and Faster R-CNN, represents a collective effort to infuse humanity into our road infrastructure. From the Random Forest Classifier for pothole

detection, the machine learning model achieved outstanding results: 91.85% accuracy, 0.9180 precision, 0.9182 recall, and a balanced F1-score of 0.9181. The Area Under the Curve (AUC) soared to 92.318 [2]. The CNN model, featured in the study, cleverly utilized convolutional layers to extract various features from training images. Impressively, it outperformed existing methods with an accuracy of 97.47%. Notably, this model comes with a lower computational cost compared to other cutting-edge models. In simpler terms, it's like having a smarter and more efficient tool for image classification tasks [3]. In the investigation of high-performance classifiers for brain tumour detection using Capsule Networks, notable precision, recall, and F1-score metrics were achieved, indicating the effectiveness of the Capsule Network architecture in tumour classification. Furthermore, the broader applicability of advanced neural network architectures in various domains was highlighted, emphasizing the versatility and potential impact of these methodologies [4]. The pothole detection system, employing a CNN-based ResNet model, achieved an unprecedented accuracy of 97.08%. This remarkable accuracy represents the best-reported performance in the existing literature on pothole detection systems [5]. Similar to the findings of another CNN model, our evaluation on new images showcased outstanding results. The model achieved an impressive accuracy of 99.80%, with precision, recall, and F1-Score standing at 100%, 99.60%, and 99.60%, respectively. Notably, in comparison studies, our model exhibited a significant performance advantage over traditional machine learning algorithms, notably outperforming the conventional SVM method [6]. LSTM model demonstrates high precision, 96% for undamaged roads and 90% for damaged roads. Few false positives were observed on both datasets, highlighting the LSTM model's effectiveness in discerning damaged from undamaged road surfaces, serving as a robust complement to the trained YOLO model [7]. In multi-class object detection involving potholes and manhole covers, YOLOv3-SPP achieved an mAP@0.5 of up to 0.791. While this heightened accuracy came at a cost of reduced speed, real-time inference was still achieved with YOLOv3 at 640 × 640, resulting in an mAP@0.5 of 0.747. Adverse conditions, like limited visibility and rainy weather, posed challenges, leading to potential misclassifications. YOLOv3 faced a major setback in very low visibility at night, with an mAP@0.5 dropping dramatically from 0.747 to 0.0701. Attempts to address this included experimenting with larger image sizes and incorporating the SPP module. Sparse R-CNN showed notable improvements in pothole detection in low-light conditions, increasing mAP@0.5 from 0.226 to 0.319 in night-time object detection [8]. In gauging the effectiveness of pothole detection through segmentation, the training outcomes consistently endorse the proficiency of the deep learning-based approach. With an accuracy surpassing 95% after 2000 iterations, the trained model exhibits promising potential for pothole detection and damage calculation [9]. The training results for both YOLOv3 and YOLOv4 models indicate optimal performance at 4000 iterations. Remarkably, the YOLOv3 model achieved a maximum mAP of 0.889 and IoU of 0.635, while the YOLOv4 model outperformed with a maximum mAP of 0.933 and IoU of 0.741. These findings highlight the strong capabilities of both models in delivering robust results [10].

In a referenced work utilizing YOLOv3-SPP, the study observed a consistent decrease in mAP@0.5 metrics from sunset to night hours. The model, set at 1080 × 1080 resolution, achieved a notable mAP@0.5 of 0.791 for multi-class object detection (potholes and manhole covers), with a trade-off of reduced detection speed. Notably, near real-time inference was attained at 640 × 640 input resolution, yielding a mAP@0.5 of 0.747. These results underscore YOLOv3's adaptability in balancing accuracy and speed in diverse operational scenarios [11]. Hence, the research study concludes YOLOv4 as the most suitable pothole detection model for accuracy. Additionally, Tiny-YOLOv4 is identified as the optimal choice for real-time pothole detection, delivering a 90% detection accuracy coupled with a noteworthy 31.76 frames per second (FPS) [12]. The YOLOX model achieved an impressive 85.6% Average Precision, showcasing a notable 24.06% improvement over the YOLOv5m model. The mean Average Precision, assessed with mAP@0.5, further establishes the YOLOX-nano model as a superior performer compared to other models. Despite its compact size at 7.2 MB, less than half that of YOLOv5s, the YOLOX-nano demonstrates slower inference speed at 0.038 seconds—four times slower than the YOLOv5s model [11]. A dataset comprising 2400 images was utilized for the training and validation phases, with an additional 265 images designated for evaluating the detection performance of various YOLO models on the test dataset. Following training, SPFPN-YOLOv4 tiny, YOLOv4 tiny, YOLO 2, and YOLOv3 models achieved detection accuracies of 79.6%, 72.7%, 74.8%, and 77.8%, respectively [13]. Achieving a commendable mean average precision (mAP) of 0.911 at 0.5 IoU, this model excels not only in accuracy but also operates efficiently with a rapid processing time of just 8.8 ms per image. The model was trained and validated on a comprehensive dataset consisting of 2400 images, with an additional 265 images dedicated to evaluating its detection performance on the test dataset [14]. The YOLO-SE network, an extension of YOLOv8, tackles multi-scale and small-object detection challenges in remote sensing. It introduces lightweight SEF and SPPFE modules, incorporating attention mechanisms and a transformer prediction head. With the Wise-IoU bounding box loss function, it achieves a notable 2.1% mAP increase over YOLOv8 on the SIMD dataset, showcasing enhanced performance in remote sensing object detection [15].

The LAYN algorithm outperforms traditional YOLOv8, reducing FLOPs by 49.62% and model parameters by 48.66%. It boosts mAP by 5.41% on PASCAL VOC and 6.96% on MS COCO's vehicle dataset, showcasing its superior performance and generalization ability [16].

In the comparative study, the enhanced YOLOv8n-seg model is evaluated alongside other instance segmentation models. Results show significant advantages of the improved YOLOv8-seg model in mAP50 metrics and model parameters. Notably, YOLOv8s-seg achieves a 2.6% higher mAP50 score than the best competitor, Pointrends, indicating substantial segmentation accuracy enhancement through the proposed strategy [17].

III. PROPOSED SYSTEM:

This research proposes a novel system for pothole detection on roadways, leveraging the state-of-the-art capabilities of YOLOv9, an advanced object detection model. The system aims to address the pressing need for efficient and accurate pothole detection to enhance road safety and infrastructure maintenance efforts. At the core of the proposed system lies an extensive dataset meticulously curated for pothole detection tasks. This dataset encompasses a diverse array of road conditions, lighting variations, and pothole types, ensuring the robustness and generalization of the trained YOLOv9 model.
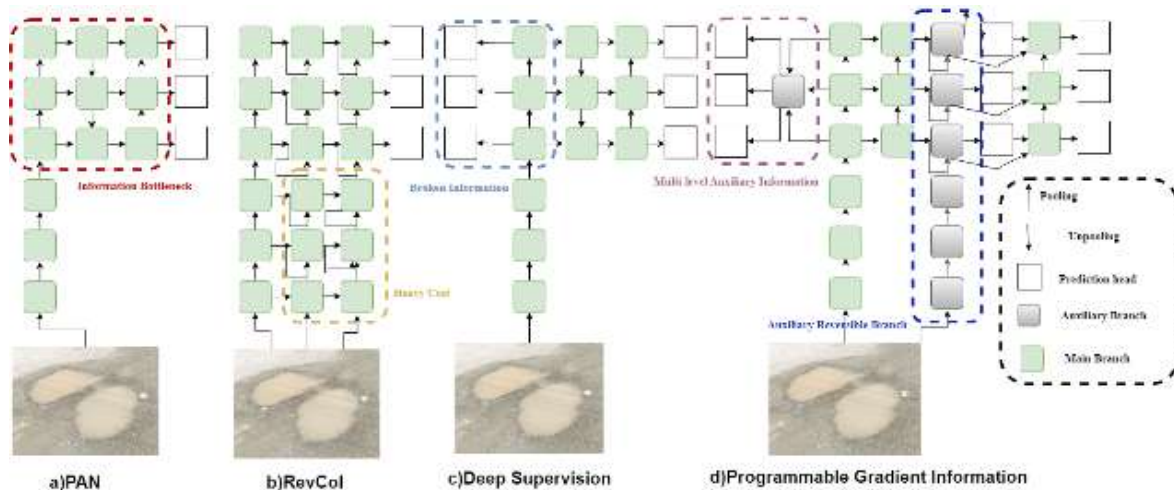
IV. ARCHITECTURE:

Figure -1 YOLOv9 PGI Architecture

The architecture of YOLOv9 introduces GELAN, a lightweight neural network designed to address the information bottleneck problem and the challenge of deep supervision in lightweight models. GELAN demonstrates strong and stable performance across various computational blocks and depths, making it adaptable to different inference devices. By integrating Programmable Gradient Information (PGI), both lightweight and deep models achieve significant accuracy improvements. The YOLOv9 model, which combines PGI and GELAN, exhibits remarkable competitiveness. Compared to YOLOv8, YOLOv9 reduces parameters by 49% and calculations by 43%, while achieving a 0.6% increase in Average Precision (AP) on the MS COCO dataset. [18]

YOLOv9 aims to reduce the information loss in deep neural networks. It achieves this by addressing the information bottleneck principle and using reversible functions.

A comprehensive comparative analysis with popular detectors like YOLOv8, YOLOv5, SSD, and EfficientDet could unveil insights into real-time performance and accuracy across various tasks and datasets. Moreover, the recent integration of Transformer features for fish detection under occlusion opens doors to applying transformers, commonly used in natural language processing, to elevate object detection in different domains. These research avenues collectively promise a deeper understanding of YOLOv9's strengths and limitations, fostering advancements not only within the model itself but also contributing broadly to the evolving landscape of computer vision and object detection.

The results emphasize the effectiveness of both the LSTM-based method and the fusion-based method over traditional machine learning algorithms in accurately identifying road potholes. Importantly, the fusion-based approach surpasses the performance of the single-modal data identification method. This innovative system showcases the potential of mobile group sensing, offering a cost-effective means to continuously monitor road conditions [18].

V. DATASET

The dataset consists of 266 images focusing specifically on potholes in and around Chennai. The orientation of each image was adjusted, and the size was standardized to 640x640, ensuring a consistent and detailed representation of potholes. Introducing grayscale to 25% of the images was a deliberate choice, aimed at enhancing our model's ability to recognize potholes under various visual conditions.

Our commitment to accuracy extends to the distribution among training (71%), testing (13%), and validation (16%) sets. This meticulous approach allows us to thoroughly assess and validate our model's proficiency in identifying potholes, a critical step in addressing infrastructure challenges. Every detail in the dataset is aimed at creating a reliable foundation for developing and testing a pothole detection model tailored to the specific conditions in Chennai.

VI. PRE-PROCESSING

In the pre-processing phase of the pothole detection dataset, we meticulously utilized various tools through Roboflow, ensuring both accuracy and diversity in the training data. Key steps included:

1) **Annotation Tools**: Tools like Label-Img and VGG Annotator were leveraged to draw bounding boxes, facilitating precise classification of the pothole class. Consistency in annotations was maintained throughout, contributing to the reliability of our model.

2) **Standardization**: All images were resized to a standardized 640x640 resolution to ensure uniformity in our dataset.

3) **Color Variation**: Grayscale transformation was applied to capture variations in color, while saturation adjustments and blurring were introduced to enhance the model's adaptability to different visual conditions.

4) **Orientation Consistency**: Auto-orientation was employed to ensure a consistent orientation across all images, facilitating seamless model training.

5) **Data Conversion**: The annotations and image data were converted into JSON format, ensuring compatibility and ease of use for subsequent model training.

This comprehensive preprocessing pipeline, executed through Roboflow, has readied our dataset for effective pothole detection model training, combining accuracy and resilience.

VII. MODEL TRAINING

For our research, we employed two prominent object detection models, YOLOv9 and YOLOv8, utilizing the computational resources provided by a Tesla T4 GPU with 15102MiB memory. The training environment was orchestrated using Python version 3.10.12 and Torch version 2.1.0+cu118.

We conducted training on the YOLOv9 model (version 8.1.24) for 25 epochs, employing a batch size of 16 images, each resized to 640x640 pixels. The training process utilized the CUDA acceleration provided by the GPU, with a learning rate of 0.001 and the Adam optimizer with a momentum of 0.937 and weight decay of 0.0005.

In parallel, training for the YOLOv8 model was executed with similar configurations, maintaining consistency across hyperparameters and image pre-processing.

Performance Evaluation and Comparison

Upon completion of training, we conducted a thorough evaluation of both models, focusing on key metrics such as accuracy, convergence speed, and computational efficiency. Comparative analysis revealed nuanced distinctions in the performance of YOLOv9 and YOLOv8, shedding light on their respective strengths and limitations in object detection tasks. For training, we employed the YOLOv9 model, harnessing the capabilities of Ultralytics. Our YOLOv9 model, with a fused summary of 168 layers and 11,125,971 parameters, embarked on 25 epochs of training using the dataset. Upon completion, the model showcased its prowess with a summary of 28.4 GFLOPs, indicating its computational efficiency.

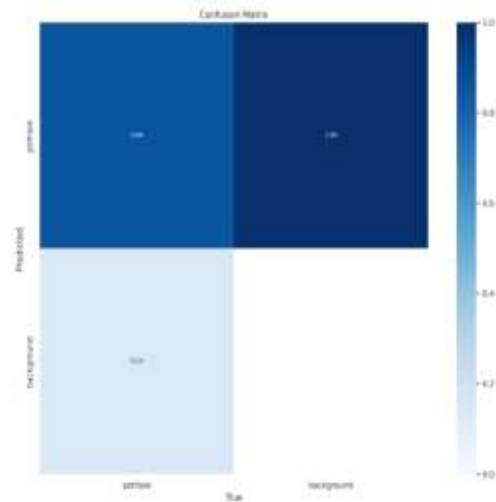| Hyperparameter	Value
Model	YOLOv9c
Epochs	25
Batch Size	16
Image Size	640
Device	cuda
Learning Rate	0.001
Optimizer	Adam
Momentum	0.937
Weight Decay	0.0005

Table 1- YOLOv9 Specifications
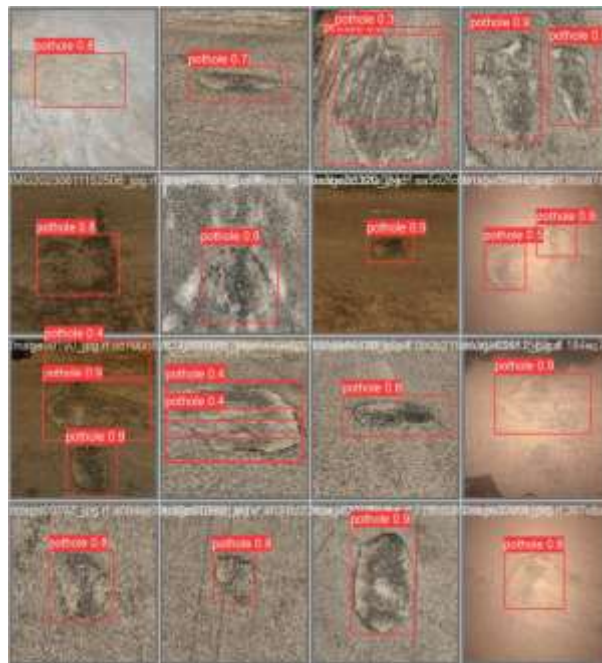
Figure -2: Confusion matrix of YOLOv8

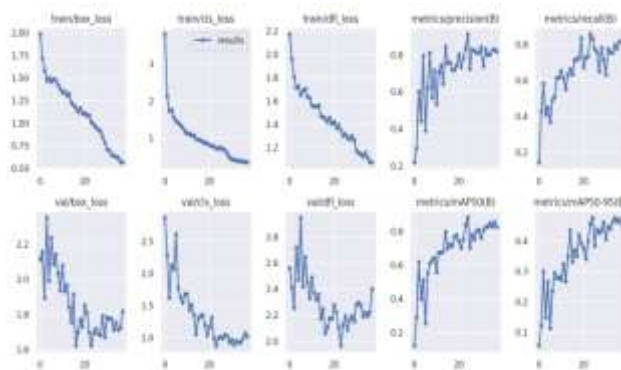*Figure -3: Predictions on test data using YOLOv8*
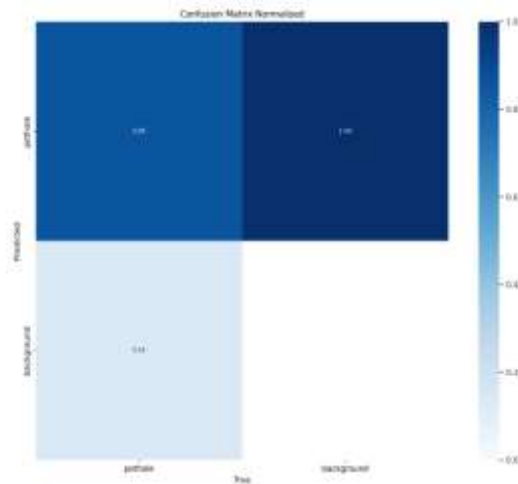

Figure - 4: Results on YOLOv8

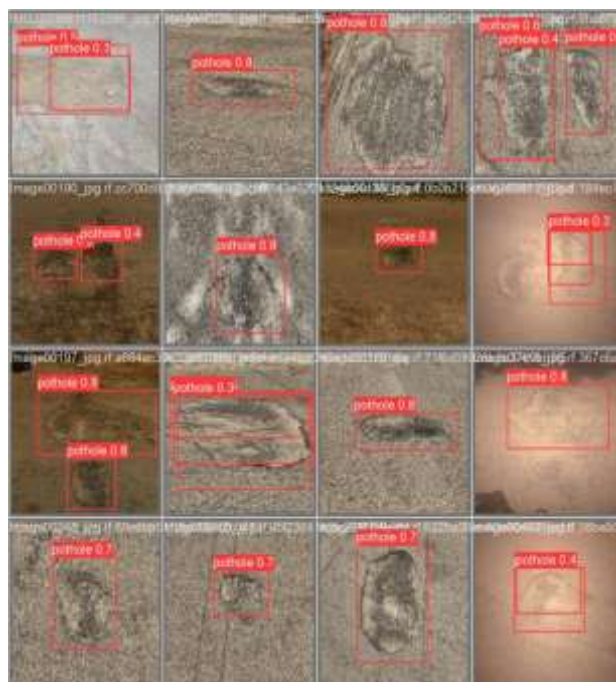*Figure -5: Confusion matrix of YOLOv9*

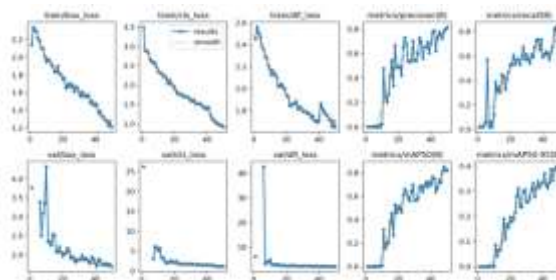*Figure -6: Predictions on test data using YOLOv9*

Figure -7: Results on YOLOv9

VIII. RESULTS AND DISCUSSION

Mean Average Precision (mAP) serves as a pivotal metric in the world of computer vision, offering a profound assessment of object detection models. Its essence lies in quantifying how effectively a model can pinpoint and outline objects within images, a task of paramount importance in various applications. Through a balanced consideration of precision and recall, mAP provides a nuanced understanding of a model's performance across different object classes. This calculation incorporates Intersection over Union (IoU), a measure ensuring the spatial alignment between predicted and actual object

boundaries. Ultimately, mAP stands as a humane yet professional benchmark, shedding light on a model's precision, recall, and spatial accuracy in identifying objects within images, thus guiding advancements in computer vision technology.

Distribution Focal Loss (DFL) is a sophisticated loss function meticulously designed to address the intricacies of class imbalance and distribution misalignment in classification tasks. At its core, DFL integrates focal loss and distribution loss components, offering a comprehensive solution tailored for scenarios characterized by uneven class distributions

DFL orchestrates a nuanced interplay between these elements, fostering a more harmonized alignment of predicted class probabilities with the true class distribution. Consequently, it stands as a powerful tool in enhancing the robustness and accuracy of classification outcomes across diverse real-world applications.

IX. CONCLUSION

In our study, we tackle the pressing challenge of pothole detection on roadways by harnessing the advanced capabilities of the YOLOv9 object detection model. With the aim of enhancing road safety and infrastructure maintenance, we delve into the realm of deep learning, particularly convolutional neural networks (CNNs), to automate the detection process—an endeavor crucial for overcoming the laborious and costly nature of conventional methods.

Our proposed method represents a significant leap forward in pothole detection, particularly under challenging conditions such as rainy weather. Through meticulous training on a custom dataset comprising diverse pothole scenarios, we fine-tuned the YOLOv9 model to achieve remarkable accuracy, boasting an impressive 82.7% accuracy rate with a rapid inference time of just 48.2 milliseconds. Furthermore, our model provides detailed measurements of pothole depth and width, offering valuable insights for infrastructure maintenance efforts.

The findings from our research underscore the immense potential of the YOLOv9 model in revolutionizing road safety and infrastructure management practices. By automating pothole detection with high precision and efficiency, our work paves the way for the implementation of scalable solutions that can effectively address the challenges posed by deteriorating road conditions.

X. REFERENCES

1. Purni, J. T., & Vedhapriyavadhana, R. (2024). EOSA-Net: A deep learning framework for enhanced multi-class skin cancer classification using optimized convolutional neural networks. Journal of King Saud University-Computer and Information Sciences, 102007.

2. Kumar, S., Kumar, N., & Barthwal, A. (2023). Road Pothole Detection Using Smartphone Sensors. Journal of Harbin Engineering University, 44(7).

3. Chu, H. H., Saeed, M. R., Rashid, J., Mehmood, M. T., Ahmad, I., Iqbal, R. S., & Ali, G. (2023). Deep learning method to detect the road cracks and potholes for smart cities. Comput Mater Contin, 75(1), 1863-1881.

4. Purni, J. T., Vedhapriyavadhana, R., Jayalakshmi, S. L., & Girija, R. (2023). High Performance Classifier for Brain Tumor Detection Using Capsule Neural Network. In Computer Vision and Machine Intelligence Paradigms for SDGs: Select Proceedings of ICRTAC-CVMIP 2021 (pp. 151-164). Singapore: Springer Nature Singapore.

5. Bhatia, Y., Rai, R., Gupta, V., Aggarwal, N., & Akula, A. (2022). Convolutional neural networks based potholes detection using thermal imaging. Journal of King Saud University-Computer and Information Sciences, 34(3), 578-588.

6. Pereira, V., Tamura, S., Hayamizu, S., & Fukai, H. (2018, July). A deep learning-based approach for road pothole detection in timor leste. In 2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI) (pp. 279-284). IEEE.

7. Ramesh, A., Nikam, D., Balachandran, V. N., Guo, L., Wang, R., Hu, L., ... & Jia, Y. (2022). Cloud-based collaborative road-damage monitoring with deep learning and smartphones. Sustainability, 14(14), 8682.

8. Bučko, B., Lieskovská, E., Zábovská, K., & Zábovský, M. (2022). Computer vision based pothole detection under challenging conditions. Sensors, 22(22), 8878.

9. Lee, S. Y., Le, T. H. M., & Kim, Y. M. (2023). Prediction and detection of potholes in urban roads: Machine learning and deep learning based image segmentation approaches. Developments in the Built Environment, 13, 100109.

10. Chitale, P. A., Kekre, K. Y., Shenai, H. R., Karani, R., & Gala, J. P. (2020, November). Pothole detection and dimension estimation system using deep learning (yolo) and image processing. In 2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ) (pp. 1-6). IEEE.

11. KC, S. (2022). Enhanced pothole detection system using YOLOX algorithm. Autonomous Intelligent Systems, 2(1), 22.

12. Asad, M. H., Khaliq, S., Yousaf, M. H., Ullah, M. O., & Ahmad, A. (2022). Pothole detection using deep learning: A real-time and ai-on-the-edge perspective. Advances in Civil Engineering, 2022.

13. Heo, D. H., Choi, J. Y., Kim, S. B., Tak, T. O., & Zhang, S. P. (2023). Image-Based Pothole Detection Using Multi-Scale Feature Network and Risk Assessment. Electronics, 12(4), 826.

14. Khare, O. M., Gandhi, S., Rahalkar, A. M., & Mane, S. (2023). YOLOv8-Based Visual Detection of Road Hazards: Potholes, Sewer Covers, and Manholes. arXiv preprint arXiv:2311.00073.

15. Wu, T., & Dong, Y. (2023). YOLO-SE: Improved YOLOv8 for remote sensing object detection and recognition. Applied Sciences, 13(24), 12977.

16. Ma, S., Lu, H., Liu, J., Zhu, Y., & Sang, P. (2024). LAYN: Lightweight Multi-Scale Attention YOLOv8 Network for Small Object Detection. IEEE Access.

17. Bai, R., Wang, M., Zhang, Z., Lu, J., & Shen, F. (2023). Automated Construction Site Monitoring Based on Improved YOLOv8-Seg Instance Segmentation Algorithm. IEEE Access.

18. Wang, C. Y., Yeh, I. H., & Liao, H. Y. M. (2024). YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. arXiv preprint arXiv:2402.13616.

19. Xin, H., Ye, Y., Na, X., Hu, H., Wang, G., Wu, C., & Hu, S. (2023). Sustainable Road Pothole Detection: A Crowdsourcing Based Multi-Sensors Fusion Approach. Sustainability, 15(8), 6610.