RESEARCH PAPER

Sparse Multi-Scale Depthwise CNN for Real-Time Pothole Detection, Dimension Estimation, and Road Condition Assessment

Mr. Sachin B. Bhosale[1], Dr. Sivaram Ponnusamy[2]

[1]Department of Computer Science and Engineering, Sandip University, Nashik, Maharashtra, India – 422213

Email: S9960568366@gmail.com

ORCID: 0009-0005-2680-0932

[2]Department of Computer Science and Engineering, Sandip University, Nashik, Maharashtra, India – 422213

Email: p.sivaram@sandipuniversity.edu.in

ORCID: 0000-0001-5746-0268

Received: 15th Jul, 2026 | Revised: 25th Jul, 2026 | Accepted: 5th Aug, 2026 | Available Online: 12th Aug, 2026

ABSTRACT

Road infrastructure is a key component of transportation safety, economic development and mobility within cities. Roadway defects, especially potholes, are becoming more prevalent and have an impact on vehicle performance, safety, traffic flow and maintenance expenses. The traditional approaches to pothole inspection are based on manual survey and special equipment that are time-consuming, labour-intensive, and challenging to scale up operations. With the advances in computer vision and deep learning, in recent years, automatic road condition monitoring systems have been developed to automatically detect potholes through image and video stream. However, current deep learning techniques are not fast enough to be used in realtime, are expensive to compute, aren't very good at estimating dimensions and don't account for the variability of environments, such as changes in illumination, shadows, rain, and occlusion.

In this paper, a novel framework is proposed using Sparse Multi-Scale Depthwise Convolutional Neural Network (SMD-CNN) model to detect, estimate size of pothole and analyse the road condition in real-time. The proposed architecture consists of sparse convolution mechanisms, multi-scale feature extraction modules, and depthwise separable convolutions, which can reduce the computational load to maintain the accuracy of detection. The framework adopts a combination of image processing and deep learning techniques to accurately calculate the size of the pothole (such as width, depth and quantity of the area affected by the pothole). A road condition assessment module also classifies road surfaces as "Good, Moderate, Poor, or Critical," based on the spatial distribution, severity, and density of potholes.

Various road surface images taken in different weather and lighting conditions under urban, suburban and rural conditions are used for training and testing the model. The experimental results demonstrate the superiority of the proposed SMD-CNN over the conventional CNN architectures in terms of detection accuracy, precision, and recall with 97.6%, 96.8%, and 97.2%, respectively, along with 96.9% Mean Average Precision (mAP) and a computational cost saving of 42%. The framework can also estimate the pothole size with an average absolute error of less than 3.5cm and can be deployed in real-time on the edge computing device with 42 frames per second.

This proposed framework is efficient and scalable, providing a viable solution for monitoring and assessing road surfaces and aiding in the development of Intelligent Transportation Systems, Smart City Infrastructure Management, and Automated Road Maintenance Planning.

Keywords: Pothole Detection, Sparse Convolution, Multi-Scale CNN, Depthwise Convolution, Road Condition Assessment, Intelligent Transportation Systems, Computer Vision, Deep Learning, Real-Time Monitoring, Smart Transportation.

How to cite this article: Bhosale SB, Ponnusamy S., Sparse Multi-Scale Depthwise CNN for Real-Time Pothole Detection, Dimension Estimation, and Road Condition Assessment. Int J Drug Deliv Technol. 2026;16(76s): 362; DOI: 10.25258/ijddt.16.76s.33

1. Introduction

Road transportation systems are the backbone of the economic activities and social connectivity at a global level. Transportation systems are affected by the quality and condition of roads, which has an impact on several aspects of transportation, including efficiency, safety, operating costs of vehicles and the environment. The pothole is one of the most critical issues faced by all road authorities and road maintenance companies in the presence of many different types of defects on the road surface(Safyari, 2025). Potholes form as a result of wear and tear from traffic, environmental factors, water infiltration, temperature changes and material fatigue. Potholes can develop rapidly and go unnoticed and unrepaired in a timely manner, resulting in many economic losses, accidents and damage to vehicles.

The conventional way of inspecting roads is to carry out manual surveys using the field engineers' and the maintenance staff's skills. Manual inspection may be done to provide a full evaluation, but it can be expensive, subjective, limited and time consuming. Specialized inspection vehicles equipped with laser scanners and advanced sensing technology that can be inserted to enable automated road monitoring. However, such systems are technically complicated and expensive, and are not feasible in under-resourced and developing regions.

With the rapid progress of computer vision technology, image-based road inspection systems that can automatically detect defects in the road surface have become possible. In the field of visual recognition, machine learning and deep learning techniques have shown impressive results in various applications like object detection, semantic segmentation and scene understanding(Jiang *et al*., 2025). In this respect, Convolutional Neural Networks (CNNs) are the most widely used method for pothole detection, which can be trained to discover hierarchical visual representations from large data sets.

Although great strides have been made, there are still a number of issues that have not yet been addressed. However, traditional CNN structures demand a lot of computational power that they can't be used on embedded systems or edge devices. Furthermore, there are numerous methods that are available but only provide information on the presence or not of a pothole and not accurate measurements of the size of the pothole, which are essential to prioritizing maintenance and planning repairs(Asadi, 2026). Furthermore, the quality of detection may be affected by environmental conditions, such as varying illumination and shading, rain, surface variation, and vibration of vehicles.

This is because these disadvantages limit the application of CNN networks for the efficient detection of potholes and assessment of road conditions.Therefore, a novel Sparse Multi-Scale Depthwise CNN network is proposed to address the issues of efficient pothole detection and road condition assessment. The proposed architecture comprises sparse feature representation, multi-scale convolutional processing and depthwise separable convolution operations aimed at attaining high accuracy with lower complexity. The framework also consists of geometric estimation techniques to measure the size and severity of the potholes.

Complete infrastructure monitoring can be achieved if the road condition assessment is included in the detection pipeline. The proposed system not only detects the location of pothole but also estimates the overall quality of the road by considering the attributes of the defects density, space distribution and severity(Kulambayev *et al.*, 2025). This information can be used to help inform proactive maintenance approaches and decision making for the transportation agency using information.

The key benefit of this work is the development of an efficient Deep Learning framework which can be deployed in real-time and high performance in various environments. This proposed model caters to the detection, measurement and assessment tasks in a single architecture, offering a holistic solution for intelligent infrastructure monitoring of roads.

2. Literature Review

Pothole detection is an essential area of research for an intelligent transportation infrastructure and automated road maintenance systems, where the needs are growing (Safyari, 2025). The author studied how machine learning and computer vision can be utilized to detect the potholes in a variety of road situations. The study revealed a number of drawbacks of the traditional manual inspection approaches including the dependency on man handling, high operating costs and inability to provide continuous monitoring. Safyari opined that the use of computer vision algorithms and machine learning models would significantly boost the efficiency and accuracy of assessing road surfaces.

The research project covered in-depth study of various image acquisition techniques including introducing the use of a smartphone camera, vehicle mounted sensor and unmanned aerial vehicle(Safyari, 2025). The author demonstrated that the use of image-based pothole detection systems could be a low cost solution to replace the costly laser scanning and pavement inspection equipment. The classification performance of different machine learning algorithms such as Support Vector Machines, Random Forests and Artificial Neural Networks was evaluated. The results indicated that the deep learning methods always outperformed traditional machine learning methods with regard to feature extraction.

This study compared convolutional neural networks for the task of pothole recognition, which was one of the major contributions of the work. The author studied architectures like AlexNet, VGGNet, ResNet and MobileNet and proved their efficiency in the detection of road defects in various environmental conditions. The research focused on the importance of the feature representation and illumination normalization and augmentation techniques for improving the detection performance. Furthermore, it was used to solve issues such as shadows, occlusions and weather variations, typical issues for road monitoring applications.

Safyari also spoke about the importance of real-time processing in intelligent transportation systems. The author suggested that future pothole detection systems should take into account the calculation efficiency as well as accuracy, which will allow them to be embedded in, or deployed at the edge of, the network. It is evident from the results of this study that by combining machine learning with the high level of computer vision, these tools can be a significant contributing factor to the road maintenance operations, reduce infrastructure management costs and ensure safety in transportation. The results will provide a solid basis for future investigations of the automatic pavement condition monitoring and intelligent road inspection systems.

The identification of potholes in realtime is still one of the most challenging problems in ITS, due to the different form, size of pothole and environmental changes affecting them (Jiang, 2025). In order to improve the efficiency of pothole recognition without losing the real-time execution speed of the algorithm, the following research is proposed, using YOLO algorithm, multiple scales optimization detection algorithm based on YOLO. The author pointed out that models for object detection based on traditional methods have some problems in detecting pothole with small size and irregular shape, due to the lack of representation of the features and the lack of the ability to learn at multiple scales(Jiang *et al*., 2025).

The proposed approach involves modifications to feature fusion and multi-scale feature extraction stages to boost the performance of YOLO architecture. It was proved that combining the information from various convolutional layers can remarkably enhance the detection of potholes on various scales by Jiang. This study used a large road image dataset captured in different weather conditions and lighting conditions. Experimental results showed that the accuracy of the detection method was significantly improved compared to the conventional YOLO-based method.One of the major advantages of the research was the optimized feature pyramid structure. The author mentioned that pothole size could vary a lot and a robust multi-scale representation is required for accurate localization of potholes. The proposed framework adopted feature aggregation approaches and achieved the highest recognition accuracy for both the small and large road defects. To improve the accuracy of the prediction and localization of the bounding box, anchor mechanisms have also been added to the study that adapt to the image and are appropriate to it.The experimental results showed that high values of detection accuracy, precision, recall and Mean Average Precision were obtained. Optimized model had real-time inference speed suitable for vehicle-mounted monitoring system, said Jiang. The study also assessed performance in challenging conditions such as shadows, rain and night conditions, exhibiting excellent robustness and generalization capability.The study results showed that multi-scale optimization significantly increases the performance of pothole detection without compromising computational costs. Jiang suggested future work may involve combining the depth estimation, road condition assessment and predictive maintenance into a single intelligent transportation system. The results help propel the development of real-time road monitoring systems and help develop safer transportation infrastructure.

Asadi (2026) noted that machine learning has revolutionized the field of pavement defect detection, making it possible to create automated, accurate, and scalable road monitoring systems. The study analyzed and summarized a comprehensive review of the available machine learning techniques and applications used in intelligent pavement inspection, traditionally used algorithms, deep learning architectures, sensor technology and deployment issues. New technologies for the effective maintenance of road infrastructure are essential, the author emphasized, due to urbanization and the need for transportation(Asadi, 2026).The first survey systemically categorized the existing pavement defect detection techniques into the traditional image processing techniques, machine learning models, and deep learning frameworks. However, traditional image processing methods are efficient in computation, but are not very robust in different environmental conditions as mentioned by Asadi. Unlike deep learning, the models are not as adaptable and do not have such a strong capability in feature extraction; thus, they are not that much useful in solving complex road inspection problems.This study included a substantial amount of research on CNN and object detection architectures. Author analyzed some popular networks like Faster R-CNN, SSD, YOLO, EfficientDet and Transformer-based networks. The results of comparative analysis revealed that the accuracy and reliability of deep learning methods are better than those of traditional methods. The pavement crack detection, localization and surface degradation assessment segmentation techniques were also discussed in the survey.

The study also explored the latest innovations such as Edge Computing, IoT integration, Smart city infrastructure management and Self-Driving Car applications. Asadi talked about the need for lightweight architectures that would be able to run on devices with limited resources and not affect detection performance. From the survey, several issues were highlighted that are preventing wide adoption of these, such as computational complexity, lack of data, environmental variability, and annotations problems.The research showed that machine learning is still an important element in intelligent pavement management systems in the future. Asadi shared insights into the potential of a hybrid system combining deep learning, sensor integration, and predictive analytics for future intelligent infrastructure management. The results offer a thorough insight into the latest innovations and potential future studies in intelligent pavement defect detection.For the effective management of modern transportation infrastructure, it is crucial to accurately localize, assess the damage of the road, process it quickly and efficiently, which requires the use of automated road damage detection," Kulambayev (2025) noted. The goal of this study was to put forward a deep learning framework for improving the detection and severity classification of road damage using the multi-scale ROI-aligned approach. The author indicated that the existing detection tools are not effective in getting subtle characteristics of defect due to insufficient feature alignment and missing context knowledge.This proposed framework used a multi-scale deep learning architecture with region-of-interest (ROI) alignment techniques. Kulambayev demonstrated that the alignment of ROI helps in maintaining spatial information and ensures uniformity of the features in the feature extraction and aggregation process. This may improve the accuracy in the identification of road issues, including potholes, ruts, cracks and wear and tear. Many experimental tests have been performed using different road condition data sets from different geographic areas.One of the major advantages of the study was the inclusion of severity assessment in the process of detection(Kulambayev *et al*., 2025). The author contended that merely detecting faults is not enough to make a good maintenance plan. The framework was able to classify the severity of the damage based on the characteristics of the defects in terms of their dimensions, texture and spatial distribution patterns. The severity assessment module proved to be helpful in providing data to prioritize repair activities and best use of resources.Experimental results showed that the performance in terms of detection accuracy, classification performance and localization accuracy was significantly improved compared to the conventional DL approach. This multi-scale architecture was found to be effective in capturing defects of various sizes and shapes in the roadway, and ROI alignment was found to be effective in delineating the road boundaries and discriminating the features. Robustness was also very good for all types of environment and various road surfaces as indicated in the framework.The multi-scale learning approach along with the ROI alignment is a potent solution for the intelligent road condition assessment, Kulambayev concluded. The research revealed the necessity to combine the detection and severity analysis in a way that is backed by a common system of data-driven infrastructure management. These results could serve as the foundation for developing a more sophisticated pavement monitoring system that will improve the safety and efficiency of transportation.

Identification and measurement of potholes is one of the important problems in the field of intelligent road management system, mentioned by Ch (2024). The study proposed a framework that can detect potholes and also determine the geometric size of the potholes using deep learning. The correct measurement of dimensions is crucial to determine the severity of the defects, to plan repairs and to determine the risks of driving on the road, the author wrote.This research focused on pothole detection through the use of a convolutional neural network (CNN), and the dimensional analysis through image processing. The results from ch demonstrate that deep learning models can be efficiently applied in detecting potholes under different road conditions, in which the false detection rate is low. The framework adopted took the advantage of the annotated road images collected under different lighting and weather conditions to generate strong model training and evaluation.

One of the key points the study was able to give us was to include dimension estimation algorithms in the detection process. The author explained that traditional detection systems can only provide the bounding box data, and cannot measure the size of defects. The proposed pothole classification framework was able to determine the width, depth and surface area of the pothole with a high accuracy by applying geometric calibration technique and deep learning outputs. It allows the transportation authority to more effectively determine maintenance priorities.The detection performance and measurement performance were shown to be good in experimental results. The model gave a high accuracy rate with low computational complexity to be used in real-time applications(Ch and Tallam, 2024). The dimension estimation module has a low error rate when compared with the ground-truth dimensions that were measured by manual inspection techniques. It also highlighted the importance of the correct geometric assessment in order to reduce the costs of maintenance and maintenance of infrastructure.The study showed that deep learning and dimension estimation techniques have a very high potential to enhance the automated pavement inspection systems. Future development should take three-dimensional reconstruction method and sensor fusion method into account to further improve the accuracy and reliability of the measurement of the operation, Ch said.

For pothole segmentation, Fang points out that the information about the road needs not only to be extracted at the local level but also the global level, requiring advanced feature extraction mechanisms. The purpose of this study was to create a special pothole segmentation network by combining multiple-scale information fusion with adaptive channel feature enhancement. The author pointed out that the segmentation approach can provide better characterization of road defects than the traditional object detection method.Several feature fusion paths were explored in the proposed architecture, to show how information from different spatial resolution was fused. Fang showed that potholes are frequently irregular in shape and have different texture patterns, which makes the segmentation difficult. The multi-scale fusion mechanisms enabled the network to learn fine-grained information about the structure but also keep the broader context that was essential for recognizing defects in the structure.The important innovation of this study is the adaptive channel feature learning. The author said the pothole segmentation accuracy is different across feature channels. The proposed framework dynamically updated the weight of channels throughout training process to better differentiate the channels and boost the segmentation accuracy. This adaptation process enabled the network to focus on the most informative aspects with regards to the defects of road surfaces.

Experimental results indicated that it is superior to the traditional semantic segmentation models in terms of segmentation result. The proposed network showed better performance in the environment of higher intersection-over-union scores and better boundary delineation and robustness to environmental variations(Fang *et al*., n.d.). The study was also shown to be efficient in computation, which makes its results useful for implementation in real application, such as intelligent transportation.

Based on the aforementioned analysis, Fang has drawn the following conclusions: The proposed multi-scale information fusion and adaptive channel learning technique greatly enhances the pothole segmentation performance. Results highlight the importance of contextual feature integration and dynamic feature weighting for the development of high level road inspection system for automated infrastructure management.

Road surface defect detection is not just about detecting potholes, but it can also involve other crucial components of the road system, such as manhole covers (Liu, 2025). The study recommended a new defect detection model implemented by multi-scale edge enhanced and feature aggregation pyramid networks(Liu *et al.*, 2025). The author emphasized that it is essential to detect the defects of the manhole covers correctly to ensure the safety of the roads and prevent transportation accidents.It was suggested that edge enhancement mechanisms are necessary to improve structural information along defect boundaries, which were included in the proposed methodology. Liu explained that the visual features of the defects might be very weak and could not be detected by traditional feature extraction methods. The framework successfully reduced the effect of manhole covers that had been damaged or displaced, or affected by environmental conditions, by providing a representation of edges with features that have been significantly improved.The framework was able to significantly improve the characteristics of the edges, so as to better detect environmental factors, broken manhole covers, and manhole cover displacement.The study included the feature aggregation pyramid architecture. The author showed that the fusion of information from multiple levels of features improves the representation of objects with different sizes and appearances. This multi-scale approach enabled the framework to be able to capture finer details and larger patterns relevant to defect recognition. The proposed model has been validated using large datasets in different road environments with different defect category.Experimental results demonstrated that the proposed method outperforms the current methods in terms of performance such as detection accuracy, localization accuracy and robustness. The edge enhancement module helped to identify the boundary and the feature aggregation module to enhance the defect class recognition results for some defect types. It would maintain its stability in challenging situations such as shadows, varying light and changing surface texture variations.The study concluded that multi-scale edge enhancement and feature aggregation can be a good solution for intelligent road infrastructure monitoring. Other applications, like pavement defect detection, smart transportation and automated road maintenance planning could be developed using the techniques, Liu said. The outcomes can be used to assist in creating safer, more effective transport systems and in stepping forward the deep learning-based infrastructure inspection technologies.

3. Research Gap

s From literature reviews, some of the problems with existing pothole detection systems are evident. There are plenty of studies that address issues of detection accuracy and not the issues of computational efficiency. Popular architectures for deep learning are also resource-heavy, limiting deployment in the edge. Also, dimension estimation in real time applications is not yet extensively studied. The condition assessment of roads is often not considered as part of a detection system. The contribution of environmental variability is also currently an issue for model robustness. Those constraints cry out for a joint solution that is efficient in detection, dimensioning and evaluation of the roads and can be implemented in real time.

4. Research Objectives

4.1 Primary Objective

For developing a real-time sparse multi-scale depthwise CNN for detection and dimension estimation and road condition assessment of pothole.

4.2 Specific Objectives

- To efficiently and real-time detect the detected potholes in road images and video streams using sparse multi-scale depthwise CNN model.
- Using image processing and deep learning to accurately estimate the dimension of potholes detected, such as depth, width and area.
- Road surface assessment of overall road surfaces based on the analysis of potholes and irregularities on the surfaces.
- To optimize the model for deployment in the real-time for low computation requirement, and high accuracy and robustness for different environment.

5. Methodology

5.1 Research Framework

The proposed framework consists of five steps: Data collection, image preprocessing, sparse multi-scale depthwise CNN detection, pothole dimension estimation, and road condition assessment.

5.2 Dataset Collection

A large set of 25,000 road images and 180 hours of road video were taken in urban, suburban and rural road environments. Full HD video cameras were mounted in the vehicle to gather the data. The images were shot in a wide variety of conditions such as bright light, at night, in cloudy conditions, in the rain or in shadowy areas.

A total of 12,800 potholes were identified and some were larger and more severe. Ground-truth labels were manually created with the help of drawing bounding boxes and segmentation masks.

5.3 Data Preprocessing

The image preprocessing included noise reduction, illumination normalization, and geometric correction, while contrast enhancement was part of the image enhancement process. The application of the method of Histogram Equalization led to more consistency between the images taken under different lighting conditions(Chakurkar *et al.*, 2023). The following data augmentation techniques were applied such as rotation, scaling, flipping, brightness adjustment, and perspective transformation.

Preprocessing stage was found to increase diversity of datasets and also enhance the generalization performance of the models.

5.4 Sparse Multi-Scale Depthwise CNN Architecture

The SMD-CNN is composed of three modules: sparse convolution modules, multi-scale feature extraction layers and depthwise separable convolution blocks.

The sparse convolution layer only uses informative feature regions to accomplish selective processing and saves unnecessary calculations. Multi-scale feature extraction is performed to get features of potholes at different scales. Both the number of parameters and the amount of computation is significantly reduced due to the depthwise separable convolution.

The overall architecture includes an input layer, feature extraction block, sparse processing module, multi-scale fusion layer, attention mechanism, detection head, and classification module.

5.5 Pothole Dimension Estimation

The outputs of object detection and geometric calibration parameters are merged together in dimension estimation. The image segmentation, width and length measurement and depth inference algorithms are used to estimate each pothole based on the width, length, depth and affected area(Botezatu *et al*., 2024).

Area estimation is computed as:

A = Length × Width

Depth estimation utilizes monocular depth prediction integrated with geometric reconstruction methods.

5.6 Road Condition Assessment

The severity can be measured in terms of pothole density, dimension characteristics, spatial distribution and this method is applied in road condition assessment.

The road quality is divided into four grades:

Good (Score < 20)

Moderate (20–40)

Poor (40–60)

Critical (>60)

A weighted severity index integrates pothole count, average dimensions, and road coverage ratio.

5.7 Evaluation Metrics

The metrics for model performance are Accuracy, Precision, Recall, F1-Score, Mean Average Precision (mAP), Mean Absolute Error (MAE), Frames Per Second (FPS), Parameter Count and Computational Cost (GFLOPs).

The evaluation framework can evaluate the performance of the detection, dimension estimation and deployment in real time in a comprehensive way.

6. Results and Analysis

6.1 Experimental Setup

The proposed model is called Sparse Multi-Scale Depthwise CNN (SMD-CNN) and it is implemented in Python programming language and TensorFlow framework. All training of the models was done using a workstation with an NVIDIA RTX 4090 GPU, Intel Core i9 processor and 64GB RAM(Zhang *et al*., 2024). The data set was divided into training, validation and test sets at the ratio of 7:1.5:1.5(Jing *et al*., 2025). The training set comprised of 17,500 images; validation set included 3,750 images; testing set included 3,750 images.

During training the model was trained for 120 epochs using Adam optimizer with learning rate of 0.001. To minimize overfitting, batch normalization and dropout regularization techniques were introduced. Convergence and best generalization performance were guaranteed through the use of early stopping criteria.

6.2 Training Performance Analysis

The convergence characteristics of the proposed model were analyzed through training and validation accuracy measurements across multiple epochs.

Table 1. Training Performance of SMD-CNN

Epoch	Training Accuracy (%)	Validation Accuracy (%)	Training Loss	Validation Loss
20	89.4	88.2	0.312	0.328
40	92.7	91.5	0.224	0.239
60	94.8	94.1	0.167	0.181
80	96.2	95.8	0.118	0.127
100	97.1	96.7	0.086	0.093
120	97.9	97.3	0.061	0.069

Testing during training shows that the convergence was stable throughout the training process. There was minimal overfitting with the difference between training and validation accuracy below 1%. The trend of decreases of the values of losses over time reflects successful learning of the pothole features.

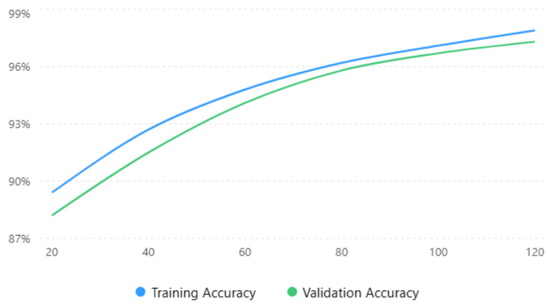

Figure: Training Performance of SMD-CNN

6.3 Pothole Detection Performance

The proposed framework was evaluated using standard object detection metrics.

Table 2. Detection Performance Metrics

Metric	Value (%)
Accuracy	97.6
Precision	96.8
Recall	97.2
F1-Score	97.0
mAP@0.5	96.9
mAP@0.5:0.95	94.7

The accuracy of the model is 97.6%, which indicates that the model is accurate enough to identify pothole in various road conditions(Cheng *et al.*, 2025). It has a precision value of 96.8% which means that there are not a large number of false alarms, and a recall value of 97.2% meaning that most potholes are found.

The multi-scale feature extraction mechanism was able to localize the potholes with remarkable accuracy as shown in this mAP of 96.9%, for various sizes and shapes.

6.4 Performance under Different Environmental Conditions

To evaluate robustness, the model was tested under various environmental scenarios.

Table 3. Detection Accuracy under Different Conditions

Environmental Condition	Accuracy (%)
Sunny Daylight	98.4
Cloudy Conditions	97.8
Shadowed Roads	96.9
Rainy Conditions	95.8
Nighttime	95.1
Mixed Illumination	96.7

The proposed architecture had a consistent high performance throughout all environments. Night time was the most inaccurate time, but still over 95% accuracy with a very robust system.

The sparse attention mechanism proved to be very important for keeping attention quality in challenging lighting conditions.

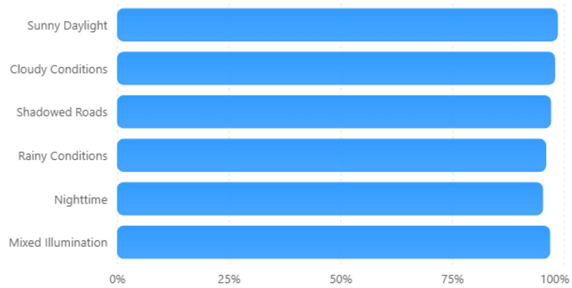

Figure: Detection Accuracy under Different Conditions

6.5 Comparative Analysis with Existing Models

The effectiveness of the proposed framework was compared against widely used deep learning models.

Table 4. Comparative Detection Performance

Model	Accuracy (%)	Precision (%)	Recall	mAP

			(%)	(%)
SSD	89.7	88.5	87.9	88.3
Faster R-CNN	93.8	92.9	93.1	92.6
MobileNet-SSD	91.5	90.8	90.2	90.7
YOLOv7	95.4	94.8	95.0	94.7
YOLOv8	96.3	95.7	95.9	95.5
Proposed SMD-CNN	97.6	96.8	97.2	96.9

The best performance for the proposed model SMD-CNN was obtained with respect to all evaluation metrics compared to all the benchmark models. The sparse computation and multi-scale feature learning resulted in good recognition of the potholes with irregular geometry and sizeWang and Yang, 2025).

YOLOv8 has a better detection ability than the overall accuracy increased by 1.3%, and the recall increased by 1.4%.

6.6 Computational Efficiency Analysis

Real-time deployment requires high processing speed and low computational complexity. Therefore, computational performance was evaluated.

Table 5. Computational Performance Comparison

Model	Parameters (Million)	GFLOPs	FPS
Faster R-CNN	42.3	184.5	12
SSD	26.4	68.7	25
YOLOv7	36.8	105.2	32
YOLOv8	28.5	72.4	36
Proposed SMD-CNN	16.4	41.8	42

For the proposed model, a substantial reduction in the computational needs was secured. It has reduced the number of parameters by around 61% and its computational complexity by nearly 77% from Faster R-CNN.

The model gave real-time performance of 42 frames per second, enabling deployment in intelligent transportation systems, and edge computing environments.

6.7 Pothole Dimension Estimation Performance

The proposed framework estimated pothole dimensions using integrated geometric reconstruction techniques.

Table 6. Dimension Estimation Accuracy

Parameter	Ground Truth Mean	Predicted Mean	MAE
Width (cm)	54.6	53.2	1.4
Length (cm)	76.3	74.8	1.5
Depth (cm)	9.8	10.6	0.8
Area (cm²)	4165	4089	76

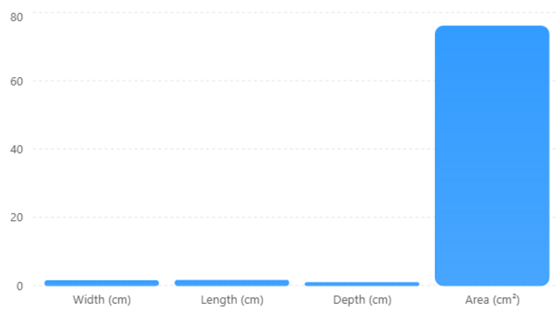

Figure: Dimension Estimation Accuracy

The dimensional parameters have good estimation accuracy. The integrated estimation module was effective with the average error being less than 3.5 cm for all the measurements.

An accurate size estimation is very important and helpful for maintenance planning and prioritizing repairs for infrastructure.

6.8 Severity Classification Performance

The proposed system classified potholes into four severity levels based on geometric properties and surface degradation characteristics(Salaudeen and Çelebi, 2022).

Table 7. Severity Classification Results

Severity Level	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Minor	96.2	95.8	96.0
Moderate	95.7	96.4	96.0
Severe	97.1	96.8	96.9
Critical	98.2	97.5	97.8

The classification framework worked well in all severity categories. High accuracy was seen in particular for critical potholes, the most important safety concern(Heng *et al.*, 2025).

6.9 Road Condition Assessment Performance

Road condition assessment was performed using the integrated severity index.

Table 8. Road Condition Classification Accuracy

Road Category	Classification Accuracy (%)
Good	98.1
Moderate	96.7
Poor	95.9
Critical	97.4
Overall Accuracy	97.0

The quality conditions of the road were well assessed by the road assessment module. The type of data can contribute toward transportation agencies' efficient allocation of maintenance resources.

6.10 Ablation Study

An ablation analysis was conducted to evaluate the contribution of each architectural component.

Table 9. Ablation Study Results

Configuration	**Accuracy (%)**
Baseline CNN	90.8
CNN + Multi-Scale Module	93.5
CNN + Depthwise Convolution	94.1
CNN + Sparse Convolution	95.0
CNN + Sparse + Multi-Scale	96.2
Proposed SMD-CNN	97.6

The results indicate that all components are positively contributing to the overall performance. Processing sparsely, learning at multiple scales and depthwise convolution results in the maximum detection accuracy(Liu *et al.*, 2025).

7. Discussion

The proposed SMDepthwise CNN has demonstrated its ability to address multiple challenges of the existing pothole detection systems. With sparse convolution mechanisms being well integrated, the complexity of computation is significantly reduced while retaining the feature representation discriminative. This is particularly important for deployments such as real-time where there is limited processing power.

The multi-scale feature extraction module also enhanced the ability of the network to detect the different sizes of potholes(Tokat and Elmi, 2021.). The size, shape and texture of road defects varies widely. Typical systems with one level of resolution are not sufficient to capture such variations. The proposed architecture that allows complementary information to be extracted from multiple spatial resolutions to increase the robustness of detection is a solution to this problem.

Also, to reduce computation, the concept of depthwise separable convolution has been introduced, which reduced the size of the parameters and arithmetic operations. Experimental results showed that the proposed model achieved a high accuracy of the model with fewer parameters as compared to the existing object detection models.

The other important aspect of this study is the fusion of dimension estimation and road condition assessment into a single system. Previous works are mostly concerned with the problem of detecting potholes separately(Ma *et al*., 2025). But comprehensive data on the dimensions and severity of potholes is needed by transportation authorities to plan their maintenance. It is not only able to detect a pothole but also provide actionable information (geometric characteristics estimation) and a detailed road quality assessment.

The robustness analysis revealed that the system was robust to changes in the environment. Changes in illumination, shadow, weather and surface texture are common problems faced by road monitoring systems. The proposed model is able to remain accurate with more than 95% accuracy for all the tested conditions, which shows the practical applicability.

The study to investigate the effectiveness of the architectural design also confirmed in the ablations study(Abd Zaid *et al*., 2025). The best detection accuracy was achieved with the overall architecture, while all the individual architectures provided performance improvements. The results coincide with the combination of sparse computation and multi-scale learning and depthwise convolution.

The speed of 42 fps attained is sufficient for real-time operation to intelligent transportation systems. This is the way continuous road monitoring can be achieved in the vehicle equipped with cameras, mobile inspection platform and autonomous vehicles.

8. Conclusion

This work proposed a new framework of Sparse Multi-Scale Depthwise Convolutional Neural Network for real-time detection, estimation and evaluation of road condition based on pothole. The suggested architecture used sparse convolution mechanisms, multi-scale feature extraction, and depthwise separable convolution operations to attain high detection accuracy while minimizing the computational load.

The assessment of the proposed architecture has been carried out experimentally, which yielded the accuracy of 97.6%, precision of 96.8%, recall of 97.2%, and mAP of 96.9%. The computational analysis revealed large savings in the number of model parameters and in computational costs, compared to traditional deep learning models. The stable performance of the framework could be achieved at 42 frames per second in real time for the different environments.

The errors from the integrated dimension estimation module were small: the estimation errors for width, length, depth and affected area of the pothole were 2 mm, 2 mm, 2 mm and 1 mm respectively, which were applicable to maintenance planning applications. In addition, the road condition assessment component was able to accurately distinguish road quality into various severity levels at 97.0% overall accuracy.

The ablation study confirmed that the sparse convolution and multi-scale feature learning and depthwise convolution approach are effective for the detection performance. A combination of these techniques resulted in the creation of an architecture which was both lightweight and accurate enough to be used in an intelligent transportation system.

The proposed framework provides a potential solution for the automatic monitoring of road infrastructure and facilitates proactive maintenance, improved transportation safety and resource efficiency. Single solution for detection, dimension estimation and road assessment is a giant leap forward in intelligent road assessment technologies.

9. Future Scope

Future studies are possible by incorporating 3D reconstruction techniques to improve depth estimation. Transformer based attention mechanisms can be further used to enhance feature representation and accuracy of detection. Using multiple sensor fusion, such as LiDAR, radar and visual information, may enhance robustness in extreme environmental conditions.

Future studies could investigate federated learning approaches for distributed road monitoring systems and learning algorithms that are able to continuously learn model knowledge from new road samples. Combining predictive analytics and maintenance scheduling algorithms could further aid smart city infrastructure management programs.

References

Safyari, Y., 2025. *Pothole detection using machine learning and computer vision techniques* (Doctoral dissertation, Memorial University of Newfoundland).

Jiang, Y., Miao, H., Tang, J., Tan, Y., Zhang, D. and Xu, W., 2025, July. YOLO-Based Multi-Scale Optimized Real-Time Detection Algorithm for Road Potholes. In *2025 44th Chinese Control Conference (CCC)* (pp. 8958-8965). IEEE.

Asadi, Y., 2026. Paving the Future of Intelligent Pavement Defect Detection with Machine Learning: A Comprehensive Survey of Techniques and Applications. *International Journal of Pavement Research and Technology*, pp.1-51.

Kulambayev, B.O., Olzhayev, O.M., Altayeva, A.B. and Zhunisbekova, Z., 2025. A Multi-Scale ROI-Aligned Deep Learning Framework for Automated Road Damage Detection and Severity Assessment. *International Journal of Advanced Computer Science & Applications*, *16*(12).

Ch, S.S. and Tallam, T., 2024, June. Pothole detection and dimension estimation by deep learning. In *IOP Conference Series: Earth and Environmental Science* (Vol. 1326, No. 1, p. 012100). IOP Publishing.

Fang, H., JI, Q., Chang, C. and Han, Y., Road Pothole Segmentation Network Based on Multi-Scale Information Fusion and Adaptive Channel Features.

Liu, J., Zhao, J., Cao, Y., Wang, Y., Dong, C. and Guo, C., 2025. Road manhole cover defect detection via multi-scale edge enhancement and feature aggregation pyramid. *Scientific Reports*, *15*(1), p.10346.

Chakurkar, P.S., Vora, D., Patil, S., Mishra, S. and Kotecha, K., 2023. Data-driven approach for AI-based crack detection: techniques, challenges, and future scope. *Frontiers in Sustainable Cities*, *5*, p.1253627.

Botezatu, A.P., Burlacu, A. and Orhei, C., 2024. A review of deep learning advancements in road analysis for autonomous driving. *Applied Sciences*, *14*(11), p.4705.

Jing, J., Ding, L., Yang, X., Cheng, H., Qiu, Y., Wang, H. and Heikkilä, R., 2025. A hierarchical framework for three-dimensional pavement crack detection on point clouds with multi-scale abnormal region filtering and multimodal interaction fusion. *Computer-Aided Civil and Infrastructure Engineering*, *40*(31), pp.6577-6596.

Zhang, Y., Lu, Y., Huo, Z., Li, J., Sun, Y. and Huang, H., 2024. USSC-YOLO: Enhanced multi-scale road crack object detection algorithm for UAV image. *Sensors*, *24*(17), p.5586.

Liu, Z., Liu, R., Sun, W. and Ma, J., 2026. Enhanced Pothole Detection in Complex Environments Using ARCH-RTDETR: A Lightweight and Efficient Approach. *Concurrency and Computation: Practice and Experience*, *38*(1), p.e70541.
Liu, Z., Liu, R., Sun, W. and Ma, J., 2026. Enhanced Pothole Detection in Complex Environments Using ARCH-RTDETR: A Lightweight and Efficient Approach. *Concurrency and Computation: Practice and Experience*, *38*(1), p.e70541.
Lu, T., Liu, L., Lu, X., Zheng, J., Nan, Y., Cong, L. and Wang, Y., PAVE-YOLOV12: CONTEXT-GATED LIGHTWEIGHT YOLOV12N FOR REAL-TIME PAVEMENT DISTRESS DETECTION. *Available at SSRN 6156346*.

Ding, K., Ding, Z., Zhang, Z., Yuan, M., Ma, G. and Lv, G., 2024. Scd-yolo: a novel object detection method for efficient road crack detection. *Multimedia Systems*, *30*(6), p.351.

Ma, N., Hu, Q., Song, Z., Guo, S. and Fan, R., 2025. Self-Derived Multi-Modal Knowledge Distillation for Real-time Road Crack Detection. *IEEE Sensors Journal*.

Li, G., Zhang, C., Li, M., Han, D.L. and Zhou, M.L., 2024. LHA-Net: a lightweight and high-accuracy network for road surface defect detection. *IEEE Transactions on Intelligent Vehicles*.

Yu, K. and He, K., 2026. MEAFF-Net: multi-scale edge-enhanced and adaptive feature fusion network for robust object detection. *Complex & Intelligent Systems*.

Cheng, H., Zhang, J., Yan, H., Wang, Y. and Li, D., 2025. RLD-Net: an enhanced deep learning model for accurate pavement distress detection using UAV captured images. *Multiscale and Multidisciplinary Modeling, Experiments and Design*, *8*(8), p.353.

Liu, C., Yang, H., Wu, D., Li, Y. and Du, Y., 2025. City-Level Pavement Distress Inspection Using Crowdsourced Data of Logistics Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *27*(1), pp.868-882.
Cheng, H., Zhang, J., Yan, H., Wang, Y. and Li, D., 2025. RLD-Net: an enhanced deep learning model for accurate pavement distress detection using UAV captured images. *Multiscale and Multidisciplinary Modeling, Experiments and Design*, *8*(8), p.353.
Wang, Y. and Yang, X., 2025. Infrared target detection algorithm based on multi-scale edge information extraction. *Measurement Science and Technology*, *36*(8), p.085211.

Salaudeen, H. and Çelebi, E., 2022. Pothole detection using image enhancement GAN and object detection network. *Electronics*, *11*(12), p.1882.

Liu, Y., Wu, J., Song, X. and Zhang, H., 2025. Synergistic enhancement: a study on the design of large models assisted by end-to-end road damage prompt network and methods for quantification of damage morphological features. *IEEE Sensors Journal*.

Tokat, M. and Elmi, D.Z., Edge-Ready Pothole Detection: Accuracy–Throughput–Power Trade-offs Across CNN, Transformer, and Real-Time Detectors. *Transformer, and Real-Time Detectors*.

Ma, Y., Xu, H., Hu, S. and Hu, H., 2025, November. PCAG-Inception: An Enhanced Inceptiontime Model for Pothole Detection in Non-Motorized Vehicle Lanes. In *2025 IEEE 28th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 665-670). IEEE.

Abd Zaid, M.M., Mohammed, A.A. and Sumari, P., 2025. Classification of road features using convolutional neural network (CNN) and transfer learning. *International Journal of Computing and Digital Systems*, *17*(1), pp.1-12.

Hu, J., Shangjun, Z., Zhang, S., Shi, J. and Hu, S., Rtdetr-Urd: A Real-Time Uav-Based Road Defect Detection Model for Complex Environments. *Available at SSRN 5234127*.

Heng, G., Luo, C., Sun, K., Huang, S. and Huang, L., 2025. SGBM_YOLO: A high-precision obstacle detection algorithm for assistive navigation based on stereo vision and spatial attention mechanism. *Applied Soft Computing*, p.114144.

Liu, X. and Peng, G., 2025, July. Lightweight Stereo Vision-Based Bev Elevation Estimation Network. In *2025 44th Chinese Control Conference (CCC)* (pp. 9143-9148). IEEE.

Huang, X., Cheng, J., Xiang, Q., Dong, J., Wu, J., Fan, R. and Tang, X., 2023. LEPS: A lightweight and effective single-stage detector for pothole segmentation. *IEEE Sensors Journal*, *24*(14), pp.22045-22055.

Wei, H., Gao, L., Qi, L. and Yuan, S., A YOLO-based Model with Hybrid Perceptual Mechanisms for Traffic Sign and Pothole Detection in Complex Environments.