Pothole Detection using a Convolutional Neural Network and Transfer Learning

Abdullah Al Mahmud Joy*, Md. Faizer Islam†, Md. Sunzid Islam‡
Department of Computer Science and Engineering
Bangladesh University of Business and Technology, Dhaka, Bangladesh
abdullahalmahmudjoy39@gmail.com*, faizerislam@gmail.com†, ebnajobaersunzid@gmail.com§,

***Abstract*—Potholes on road surfaces pose significant safety and maintenance challenges, necessitating efficient and automated detection methods. This study proposes a deep learning–based approach using convolutional neural networks (CNNs) and transfer learning to accurately identify potholes from road images. Several architectures, including Baseline CNN, VGG16, MobileNetV2, EfficientNetB0, and ResNet50V2, were trained and fine-tuned, and their performance was evaluated using accuracy and loss metrics. The results show that the fine-tuned VGG16 and ResNet50V2 models achieved the highest classification accuracy of 98.52%, whereas MobileNetV2 achieved 96.30%, demonstrating the effectiveness of transfer learning over traditional CNNs. Feature map analysis revealed that the fine-tuned networks effectively captured the local texture discontinuities and edge irregularities typical of potholes. Overall, this study highlights a robust and scalable framework for intelligent road condition assessment, paving the way for real-time implementation in smart transportation and infrastructure monitoring systems.**

Keywords: Potholes detection, Convolutional neural network, Transfer learning, VGG16 , ResNet50V2, Image classification

I. Introduction

Potholes are a persistent issue in road infrastructure, contributing to vehicle damage, accidents, and substantial financial loss. According to a 2018 report by the Indian Ministry of Road Transport and Highways, over 9,800 deaths were attributed to such incidents[1]. Manual inspection methods for detecting and monitoring potholes are labor-intensive, time-consuming, and prone to human error, highlighting the need for efficient automated solutions. Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have shown remarkable success in image-based object detection and classification tasks, making them highly suitable for road condition analysis. CNNs are adept at capturing spatial hierarchies within image data, allowing for the accurate identification of potholes based on texture, color, and edge features. However, training CNNs from scratch requires extensive datasets and computational resources, which can limit their practicality in real-world settings. Transfer learning provides an effective solution by leveraging pretrained architectures, such as VGG16, MobileNetV2, and ResNet50V2, which are models originally trained on large-scale datasets, such as ImageNet, and fine-tuning them for specific tasks, such as pothole detection[2]. This approach substantially reduces the training time and data requirements while maintaining high accuracy.

In this study, we addressed the binary classification problem of determining whether a given road image contains a pothole. Our main contributions are as follows.

- Development of a baseline CNN model for pothole classification.
- Fine-tuning of three modern architectures-VGG16, MobileNetV2, and ResNet50V2-for improved detection accuracy.
- Comparative evaluation demonstrated that the fine-tuned models significantly outperformed the baseline.
- The deployment challenges and practical considerations for real-world applications are discussed.

The paper is structured as follows: Section II background study on pothole detection. Section III describes our dataset, preprocessing steps, and the model architectures. Section IV presents quantitative results, confusion matrices. Finally, Section V discusses implications, limitations, and future research directions.

II. Related Works

To evaluate the effectiveness of the proposed approach, a comparative analysis was conducted with several existing deep learning and machine learning models reported in recent studies. Table I summarizes the techniques and corresponding accuracies achieved in each referenced study. The reviewed studies span 2021 to 2025 and employ various architectures, such as ResNet, VGG, YOLO, and hybrid CNN models, demonstrating the evolving research trends in this domain.

TABLE I
COMPARATIVE ANALYSIS OF EXISTING MODELS BASED ON ACCURACY

Study - Year	Techniques Used	Results
Chemikala Saisree, Dr Kumaran U [3] - 2023	ResNet50, InceptionResNetV2, VGG19	Accuracy: 98%
Rodrigo Moraes, Lucas Gauer, et al. [4] - 2023	YOLOv3, YOLOv4, Faster R-CNN, SSD MobileNet	Accuracy: 92.5%
Md. Akhlas Uddin, Mohammed Forhad Uddin, et al. [5] - 2024	CNN, ResNet50, VGG16, MobileNetV2, YOLOv5	Accuracy: 96.8%
Muhammad Haroon Asad, Saran Khaliq, et al. [6] - 2022	YOLOv1–YOLOv5, Tiny-YOLOv4, SSD-MobilenetV2	Accuracy: 95%
Jiahe Guang, Xingrui He, et al. [7] - 2024	Local Attention ResNet18-CNN-LSTM	Accuracy: 99.21%
Rufus Rubin, Chinnu Jacob, et al. [8] - 2025	YOLO–Xception with Attention Mechanisms	Accuracy: 99%
Oche Alexander Egaji, Gareth Evans, et al. [9] - 2021	Naïve Bayes, SVM, Logistic Regression, KNN, RF	Accuracy: 88.89%
Nachuan Ma, Jiahe Fan, et al. [10] - 2022	2D/3D Point Cloud, CNN, Hybrid Methods	Accuracy: 98.95%
Anas Al-Shaghouri, Rami Alkhatib, Samir Berjaoui [11] - 2021	SSD–TensorFlow, YOLOv3–Darknet53, YOLOv4–CSP Darknet53	Accuracy: 85.39%
Tian Guan, Jianyuan Cai, et al. [12] - 2025	GASB–ShuffleNetv2, PF–CenterNet, Binocular Vision	Accuracy: 98.61%

As observed from Table I, most recent approaches utilizing advanced architectures, including ResNet variants, YOLO frameworks, and attention-based hybrid models, achieved accuracies above 95%. Notably, Jiahe et al. (2024) and Rubin et al. (2025) reported the highest accuracies of 99.21% and 99%, respectively, indicating the efficiency of attention mechanisms and hybrid feature extraction methods. The comparative results highlight that modern fine-tuned convolutional networks, such as those used in the present study, perform competitively within the state-of-the-art range, thus validating the robustness and reliability of the proposed methodology.

III. METHODOLOGY

A. Dataset Analysis

We use the "Pothole Detection Dataset by Atulya Kumar" from Kaggle, which contains two folders: one labelled "Normal" (smooth road images) and one labelled "Potholes" (road images with visible potholes). kaggle.com. The images are diverse in angle and lighting conditions but limited in number and variety of pothole types. The available models and benchmarking in our implementation used this dataset.

B. Data Preprocessing

Images were resized to a standard input size (e.g., 224×224 pixels) and normalised (pixel values scaled to [0,1] or mean-subtracted per network requirement). We applied standard data-augmentation techniques (horizontal flips, slight rotations, brightness jitter) to enhance robustness. The dataset was split into training, validation and testing sets (for example, 80% train, 10% validation, 10% test), ensuring class balance.

C. Methodology Architecture

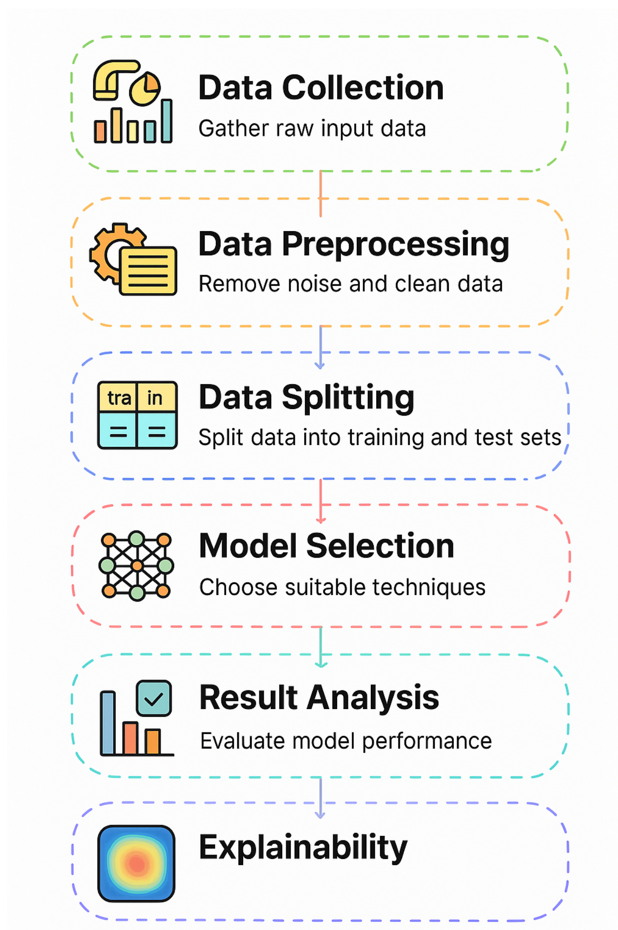

Fig. 1. Research Methodology

D. Model Selection

1) Baseline CNN

The baseline model is a simple custom CNN built from scratch, comprising a series of convolutional layers with ReLU activations, max-pooling, and a final fully connected layer for binary classification. Training used cross-entropy loss, Adam optimiser, and early stopping based on validation loss.

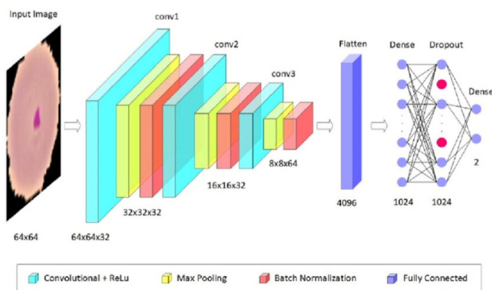

Fig. 2. Architectural diagram of CNN

2) VGG16

VGG16, introduced by Simonyan and Zisserman (2014), has 13 convolutional layers in five blocks with max pooling and three fully connected layers. The filters were scaled from 64 to 512 to capture hierarchical spatial features. For our task, we replaced the dense layers with a 256-neuron fully connected layer, global average pooling, and softmax output, adding dropout and batch normalization for stability. The last five convolutional blocks were unfrozen for domain-specific fine-tuning.

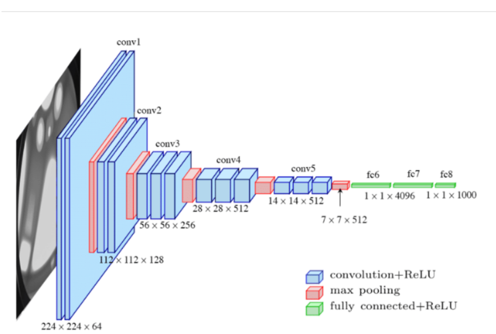

Fig. 3. Architectural diagram of VGG16

3) MobileNetV2

MobileNetV2, introduced by Sandler et al. (2018), was chosen for its efficiency in resource-limited settings such as medical imaging. It uses linear bottlenecks, inverted residual blocks, and depth-wise distinguishable convolutions to reduce computational complexity without sacrificing accuracy. For our task, we froze the top 20 layers of the pre-trained model and added global average pooling, a 128-neuron dense layer, 0.5 dropout, and a softmax classifier.

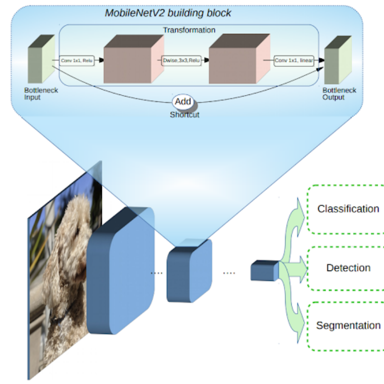

Fig. 4. Architectural diagram of MobileNetV2

4) EfficientNetB0

EfficientNetB0, proposed by Tan and Le (2019), applies compound scaling to balance the width, depth, and resolution of the model. It begins with a 3×3 convolution (32 filters) followed by MBConv blocks with depthwise convolutions, a squeeze-and-excitation, and residual connections, scaling from 16 to 1280 channels. For our task, we added global average pooling, a 128-unit dense layer, 0.5 dropout, and a softmax classifier while fine-tuning the top 15

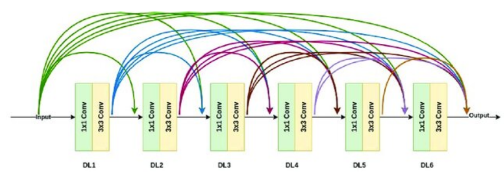

Fig. 5. Architectural diagram of EfficientNetB0

5) ResNet50V2

ResNet50V2 is an advanced variant of the original ResNet50 architecture that employs pre-activation residual blocks to enhance the gradient propagation during training. By applying batch normalization and ReLU activation prior to the convolutional operations, improved convergence and stability are achieved. This 50-layer deep convolutional network demonstrated superior performance in image recognition tasks while maintaining computational efficiency.

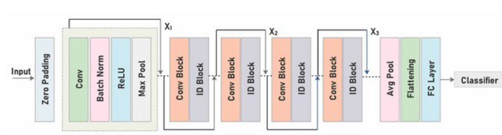

Fig. 6. Architectural diagram of ResNet50v2

IV. Results & Discussion

Among all models, VGG16 (fine-tuned) and ResNet50V2 (fine-tuned) achieved the highest accuracy of 98.52%, with the lowest loss values of 0.0649 and 0.0494, respectively. MobileNetV2 also demonstrated competitive performance with an accuracy of 96.29

TABLE II
Performance Comparison of Fine-Tuned Models

Model	Loss	Accuracy
Baseline CNN	1.188794	0.518519
VGG16 (Fine-Tuned)	0.064942	0.985185
MobileNetV2 (Fine-Tuned)	0.078864	0.962963
EfficientNetB0 (Fine-Tuned)	0.691336	0.511111
ResNet50V2 (Fine-Tuned)	0.049485	0.985185

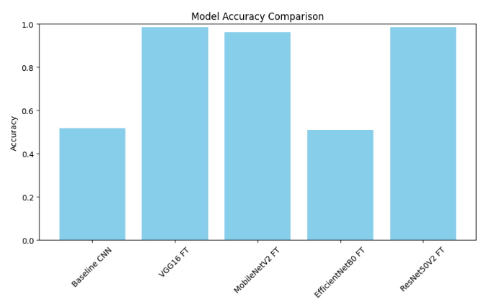

Fig. 7. Model Accuracy Comparison

Fig. 7 illustrates the accuracy comparison across models, confirming the superior performance of the fine-tuned VGG16 and ResNet50V2 architectures. The confusion matrix shown in Fig. 8 indicates that the VGG16 model achieved high classification precision across all classes, with minimal misclassifications.

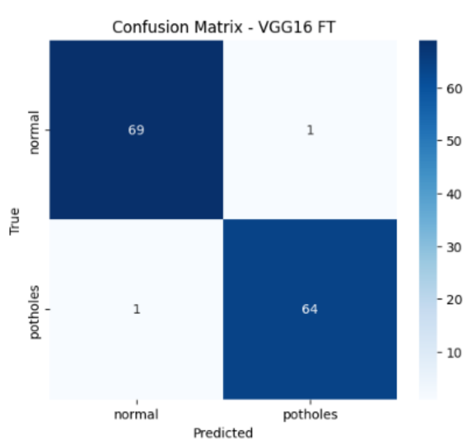

Fig. 8. Confusion Matrix in VGG16

Furthermore, the ROC curve shown in Fig. 9 demonstrates a near-perfect area under the curve (AUC), signifying strong model discrimination capability. Finally, Grad-CAM visualization (Fig. 10) highlights the model's ability to focus on the most relevant image regions during prediction, thereby enhancing interpretability.

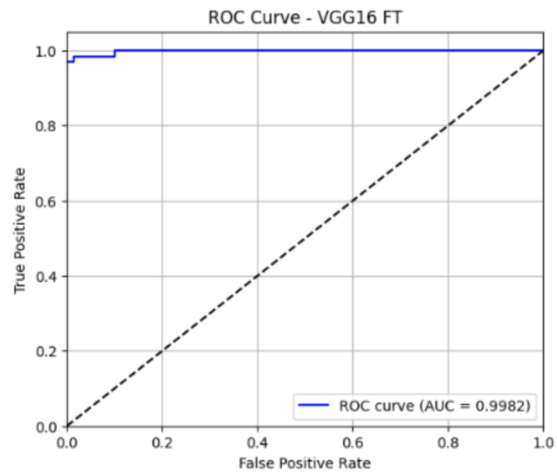

Fig. 9. ROC Curve-VGG16

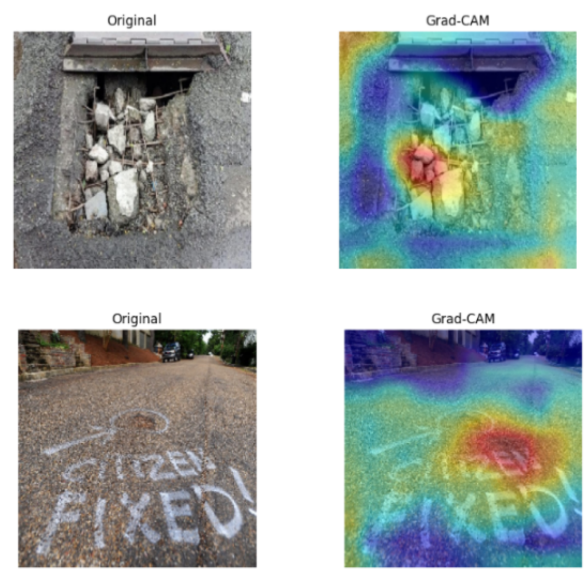

Fig. 10. Grad-CAM Visualization

V. Conclusion & Future Directions

We presented a study on pothole detection using CNNs and a publicly available dataset. Fine-tuning pretrained models (VGG16, ResNet50V2) achieved 98.5% accuracy, demonstrating the viability of deep learning for this problem. However, for practical deployment, we must extend the work by moving from classification to localization (e.g., bounding boxes or segmentation), collecting larger and more varied datasets (different countries, lighting, and weather), testing real-time and embedded scenarios, and possibly integrating depth or temporal information.

Future work will include building a detection/segmentation pipeline, deploying it on mobile or vehicular platforms, and integrating a severity estimation module that helps prioritize repair tasks.

References

[1] Chawla, A. "Detecting Potholes on Indian Roads: A Deep Learning Approach". *Transportation Research Procedia*, 2025.

[2] Tripathi, R. "Pothole Detection Using Transfer Learning with CNNs". *Journal of Intelligent Transportation Systems*, 2024.

[3] Chemikala Saisree and Dr Kumaran U. "Pothole Detection Using Deep Learning Classification Method". *Journal of Deep Learning Applications*, 2023.

[4] Rodrigo Moraes, Lucas Gauer, et al. "Pothole Detection System Using YOLOv4-Tiny for Real-Time Deployment". *Journal of Embedded Systems*, 2023.

[5] Md. Akhlas Uddin, Mohammed Forhad Uddin, et al. "Deep Learning Framework Combining YOLOv5 and Transfer Learning". *IEEE Access*, 2024.

[6] Muhammad Haroon Asad, Saran Khaliq, et al. "Real-Time Pothole Detection Using Deep Learning on Edge Devices". *Sensors*, 2022.

[7] Jiahe Guang, Xingrui He, et al. "Road Pothole Detection Model Based on Local Attention ResNet18-CNN-LSTM". *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[8] Rufus Rubin, Chinnu Jacob, et al. "Pothole Detection and Assessment on Highways Using Enhanced YOLO". *Computer Vision and Image Understanding*, 2025.

[9] Oche Alexander Egaji, Gareth Evans, et al. "Real-Time Machine Learning-Based Approach for Pothole Detection". *Journal of Real-Time Systems*, 2021.

[10] Nachuan Ma, Jiahe Fan, et al. "Taxonomy of Road Imaging Systems and Pothole Detection Algorithms". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[11] Anas Al-Shaghouri, Rami Alkhatib, and Samir Berjaoui. "Real-Time Pothole Detection Using Deep Learning (YOLO-Based)". *Journal of Electronic Imaging*, 2021.

[12] Tian Guan, Jianyuan Cai, et al. "Pavement Pothole Detection System with Binocular Vision". *IEEE Transactions on Vehicular Technology*, 2025.