A Deep Learning Approach for Automated Detection of Road Pothole

Dept. of Computer Science and Engineering, American International University-Bangladesh (AIUB)
Course: Machine Learning,
Section A

Supervised by Prof. Dr. Md. Asraf Ali

Md. Tanziul Haque
ID: 22-46435-1
22-46435-1@student.aiub.edu

Md. Naimul Islam
ID: 22-47899-2
22-47899-2@student.aiub.edu

Mohammed Ridwan Hasnain
ID: 22-46463-1
22-46463-1@student.aiub.edu

Kaniz Faria Ahmed
ID: 22-46429-1
22-46429-1@student.aiub.edu

***Abstract*—Potholes on road surfaces pose serious safety hazards and contribute to infrastructure degradation. Traditional manual inspection methods are labor-intensive and error-prone, prompting the need for automated solutions. This study presents a deep learning-based approach for the automatic detection of road potholes using Convolutional Neural Networks (CNNs). Four models were implemented and evaluated: a Custom CNN, MobileNetV2, ResNet50, and VGG16. The dataset, comprising labeled road surface images, was preprocessed with augmentation techniques to improve model robustness. Experimental results demonstrate that MobileNetV2 achieved the highest accuracy (89%) , offering an optimal balance between performance and computational efficiency. Overall, the study confirms that pre-trained CNNs can deliver reliable, real-time pothole detection, suitable for deployment on edge devices in smart city infrastructure and vehicle-based systems.**

***Index Terms*—Pothole Detection, Deep Learning, Convolutional Neural Network (CNN), MobileNetV2, Transfer Learning, Image Classification, Road Surface Monitoring, Smart Cities, Real-Time Detection, Edge Computing**

I. Introduction

A pothole is a hole in the surface of a road caused by sustained mechanical damage by traffic and weather. Potholes, a kind of road defect, can damage vehicles and negatively affect drivers' safe driving, and in severe cases can lead to traffic accidents [1]. Potholes and other road surface damages pose significant risks to vehicles and overall traffic safety. Among the common defects, potholes are particularly dangerous, often causing vehicle damage, traffic delays, and even road accidents [8]. Regular maintenance of roadways is vital, but it has always depended on manual inspection and is highly time consuming, tedious, and prone to human error. As a result, engineers and researchers investigated different technologies capable of automated pothole detection to solve these issues.These include vibration-based sensors [6], ultrasonic systems [8], GPS and accelerometer-based approaches [10], computer vision [11], and, most recently, artificial intelligence (AI) techniques such as machine learning (ML) and deep learning (DL). Among these, DL—especially using convolutional neural networks (CNNs)—has emerged as a promising approach due to its high accuracy in visual pattern recognition and its ability to handle large, unstructured image datasets [9].

Several recent studies have explored deep learning-based pothole detection from multiple angles. Dib et al. [2] reviewed anomaly detection approaches and highlighted challenges such as false negatives and environmental noise in vision-based systems. Çınar and Kaya [3] proposed a DenseNet121-based model for pothole detection, achieving a high accuracy of 99.3%, outperforming other deep learning models like ResNet50, InceptionV3, VGG19, and InceptionResNetV2 under the same conditions. Their approach demonstrated strong potential for accurate road surface assessment. Complementing this, Safyari et al. [8] conducted a comprehensive review of vision-based pothole detection methods, noting that hybrid models combining traditional image processing with advanced machine learning techniques yield the highest detection accuracy. They emphasized the growing role of deep learning in improving system adaptability and effectiveness across 2D image processing, 3D point cloud modeling, and smartphone-based sensing systems. Ruseruka et al. [4] used YOLOv5 with in-vehicle cameras, reaching a 96.3% mAP and proving feasible for low-cost, embedded deployment. Ahmed introduced a modified VGG16 using dilated convolutions, which showed improved accuracy with reduced computational load, while YOLOv5 Small excelled in speed [5]. Lee et al. [12] presented a hybrid model combining environmental prediction (temperature, humidity, traffic) with CNN-based visual segmentation, achieving high detection accuracy but

Dept. of Computer Science and Engineering, American International University-Bangladesh (AIUB)

requiring constant retraining. Encoder-decoder CNN models using datasets like KITTI and CamVid have also been used to segment roads, although they focused less on defects. Yik et al. [13] applied YOLO for general road object detection and noted performance trade-offs under different lighting and weather conditions. Despite progress, most existing solutions lack generalizability across datasets and struggle in varied environmental conditions or on resource-limited devices. This indicates a clear research gap in developing real-time, robust, and computationally efficient pothole detection models adaptable to real-world constraints.

To bridge this gap, this study aims to develop a lightweight and accurate deep learning-based model for automated pothole detection that performs reliably across diverse road and environmental conditions. The proposed solution leverages pretrained convolutional neural networks with an optimized model architecture and transfer learning techniques to ensure high detection performance while minimizing computational complexity. Emphasis is placed on real-time feasibility, environmental adaptability, and the potential for deployment on embedded systems such as mobile or edge devices used in smart city infrastructure or vehicle-based monitoring platforms.

II. Methodology

This study presents a deep learning approach for automatically detecting potholes using Convolutional Neural Networks (CNN). The goal is to classify road surface images into two categories—potholes and normal roads—by taking advantage of CNNs' ability to learn spatial features directly from raw images, without relying on manual feature extraction. The system is designed to handle different lighting conditions, road textures, and pothole shapes effectively. Our workflow includes preparing the dataset, preprocessing and augmenting the images, building the CNN model, training it, and evaluating its performance. Each step is discussed in detail below.

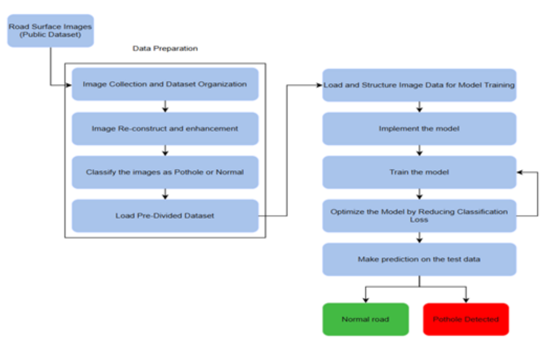

Fig 1. Workflow of the pothole detection pipeline.

A. Data Description

For this research, we used a publicly available dataset hosted on Kaggle titled "Pothole and Plain Road Images" (https://www.kaggle.com/datasets/virenbr11/pothole-and-plain-rode-images). The dataset was accessed in May 12,2025, and it contains labeled images of road surfaces classified into two categories:

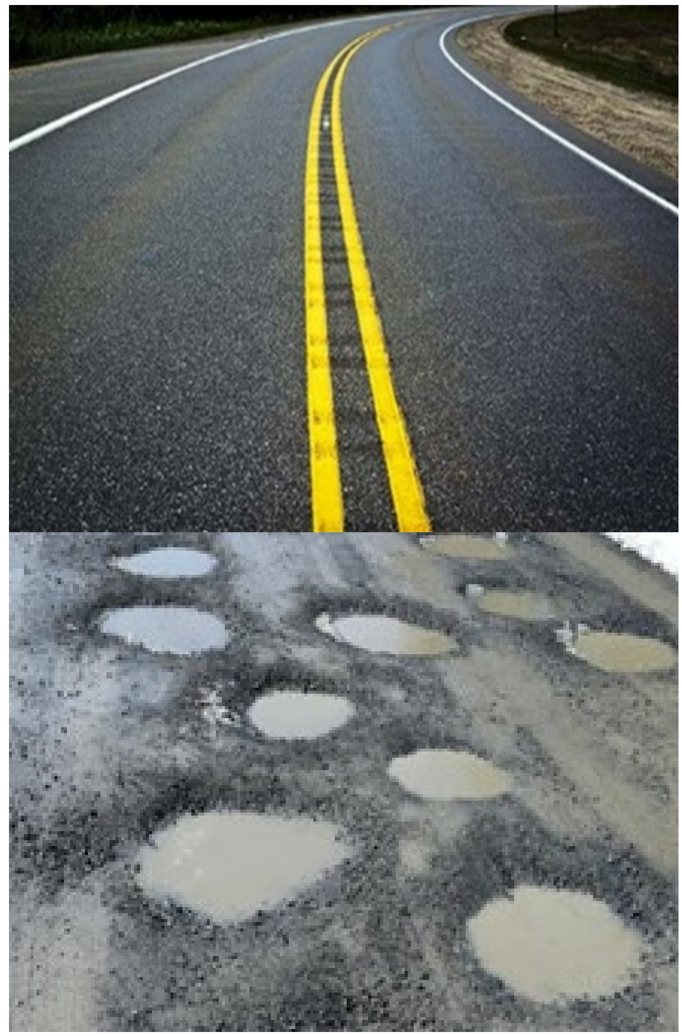

Fig 2. Example of images in datasets Plain road surface & Pothole-filled road surface.

The images are color JPEG files depicting various lighting, shadow, and surface conditions, making the dataset useful for training robust models. Despite the small training set, transfer learning techniques allow us to fine-tune pretrained deep learning models effectively even with limited data.

TABLE I
Summary of the Dataset

Class labels	Training	Testing	Total
Plain	300	75	375
Pothole	292	73	365
Total	**592**	**148**	**740**

B. Data Preparation and Preprocessing

Before feeding images into the classification models, the following preprocessing steps were conducted to ensure consistency and improve model generalization:

- **Image Resizing:** All images were resized to 224 × 224 pixels, compatible with pre-trained CNNs such as MobileNetV2, ResNet50, and VGG16.
- **Normalization:** Pixel values were scaled to the [0, 1] range (for the custom model) or preprocessed using the corresponding pre-trained model's preprocessing function (e.g.,tf.keras.applications.mobilenet_v2.preprocess_input for MobileNetV2).
- **Color Format:** Images were maintained in RGB format.
- **Augmentation:** To prevent overfitting and improve generalization, several augmentation techniques were applied:
  - Random rotations (±30°)
  - Horizontal flipping
  - Zooming
  - Brightness adjustments
  - Width and height shifting
- **Dataset Splitting:** The dataset was split into 80% for training and 20% for testing.
- **Visualization:** A 3×3 grid of augmented images (Figure 3) was visualized to confirm diversity in orientation and lighting conditions.

Fig 3. Sample augmented training images.

C. Feature Extraction and Feature Selection (CNN-Based)

In this study, feature extraction and selection are intrinsically handled by the architecture of Convolutional Neural Networks (CNNs), eliminating the need for manual feature engineering. This approach enables models to automatically identify patterns, textures, and structures crucial for distinguishing potholes from plain road surfaces.

1) Feature Extraction: CNNs consist of multiple convolutional layers that progressively learn hierarchical spatial features from raw image data. Early layers detect low-level features such as edges and textures, while deeper layers learn complex, abstract representations such as pothole shapes and their surroundings [14].

The mathematical operation of a convolutional layer is expressed as:

$$F = f(W * I + b) \tag{1}$$

Where:

- F is the output feature map,
- I is the input image,
- W and b are learnable weights and biases,
- f is a non-linear activation function, typically ReLU.

Each CNN architecture used in this study—Custom CNN, MobileNetV2, ResNet50, and VGG16—employs this hierarchical feature learning mechanism. For example, MobileNetV2 utilizes depthwise separable convolutions to reduce computational complexity while maintaining feature richness [15].

2) Feature Selection: Feature selection is inherently embedded in the CNN pipeline. MaxPooling layers downsample feature maps by retaining the most significant activations, which helps reduce redundancy and focus on salient features [16]. Additionally, dropout layers [17] randomly deactivate neurons during training to prevent overfitting.

Moreover, through backpropagation, CNNs dynamically update weights to emphasize features that most contribute to reducing prediction error [18]. This allows the model to prioritize relevant spatial patterns for accurate classification of potholes.

By combining automated feature extraction with implicit feature selection, CNNs provide a robust foundation for image-based pothole detection without the need for manual preprocessing or handcrafted features.

D. Classification Models

In the pursuit of accurate and efficient pothole detection, four distinct Convolutional Neural Network (CNN) architectures were developed and evaluated: a Custom CNN, MobileNetV2, ResNet50, and VGG16. Each model was designed to leverage the strengths of deep learning in feature extraction and classification, tailored to the specific challenges of identifying potholes in road images.

The Custom CNN model was architected with three convolutional blocks, each comprising convolutional layers followed by activation and pooling layers, culminating in fully connected dense layers. This model was trained from scratch on the pothole dataset, achieving a test accuracy of 85%. Its lightweight nature and tailored design make it suitable for deployment in environments with limited computational resources. This approach aligns with methodologies employed in prior studies focusing on road crack detection using deep convolutional neural networks [19].

MobileNetV2, employed through transfer learning, utilized pre-trained weights from the ImageNet dataset. Its architecture is characterized by depthwise separable convolutions and inverted residuals, which reduce computational complexity while maintaining performance. Fine-tuning the final dense layers on the pothole dataset resulted in an impressive accuracy of 89%, demonstrating high recall for both pothole and non-pothole classes [20].

ResNet50, another transfer learning model, incorporates residual blocks that facilitate the training of deeper networks by mitigating the vanishing gradient problem. Pre-trained on ImageNet, ResNet50 was fine-tuned on the pothole dataset, achieving a test accuracy of 84%. Its deep architecture enables the capture of complex features, making it adept at distinguishing subtle differences in road textures [21].

VGG16, known for its uniform architecture of 16 weight layers, was also utilized through transfer learning. Despite its deeper structure, which can make it prone to overfitting, careful fine-tuning allowed the model to achieve a test accuracy of 86%. The model's performance underscores its capability in feature extraction for complex image classification problems [22].

In summary, each CNN architecture demonstrated unique strengths in the context of pothole detection. The Custom CNN offered a balance between simplicity and accuracy, MobileNetV2 provided efficiency with high performance, ResNet50 delivered depth and robustness, and VGG16 showcased powerful feature extraction capabilities. The selection of an appropriate model depends on specific deployment requirements, such as computational resources, desired accuracy, and real-time processing needs.

The architecture consists of:

- **Input Layer:** Accepts 224 × 224 × 3 images.
- **Convolutional Layers:** Extract spatial features using learnable filters.
- **Activation Function:** ReLU is applied to introduce non-linearity.
- **Pooling Layers:** Reduce dimensionality while retaining key features.
- **Flatten Layer:** Converts 2D feature maps into a 1D feature vector.
- **Fully Connected Layers:** Perform high-level reasoning for classification.
- **Output Layer:** A softmax activation for binary classification (plain vs pothole).

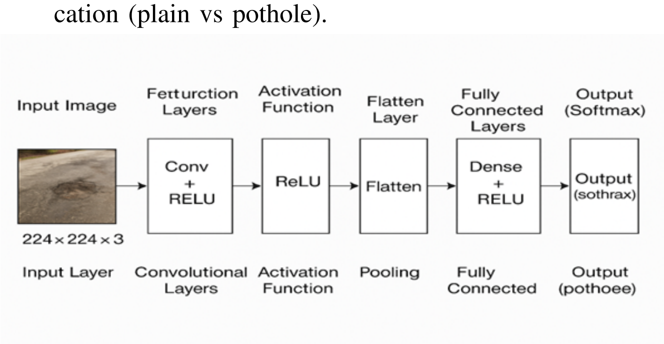

Fig 4. Diagram of CNN Pipeline.

III. Model Evaluation

To rigorously assess the performance of the developed CNN-based models for pothole detection, a suite of classification metrics was employed. Evaluation was based on test predictions using standard statistical measures derived from the confusion matrix. The confusion matrix provides a comprehensive breakdown of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), enabling a detailed analysis of each model's classification behavior.

The following metrics were used for evaluation:

- **Accuracy:** Measures the overall proportion of correctly classified instances:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision:** Quantifies the proportion of positive identifications that were actually correct:

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall (Sensitivity):** Evaluates the proportion of actual positives that were correctly identified:

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1 Score:** The harmonic mean of precision and recall:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

These metrics were computed for each of the four models: Custom CNN, MobileNetV2, ResNet50, and VGG16. Evaluation was performed on a held-out test dataset containing balanced representations of both plain and pothole classes. Threshold tuning (e.g., 0.6) was also applied to optimize the decision boundary for binary classification.

IV. Confusion Matrix Analysis

To ensure reliability, performance scores were supported by confusion matrices, which offer a visual snapshot of correct vs. incorrect predictions.

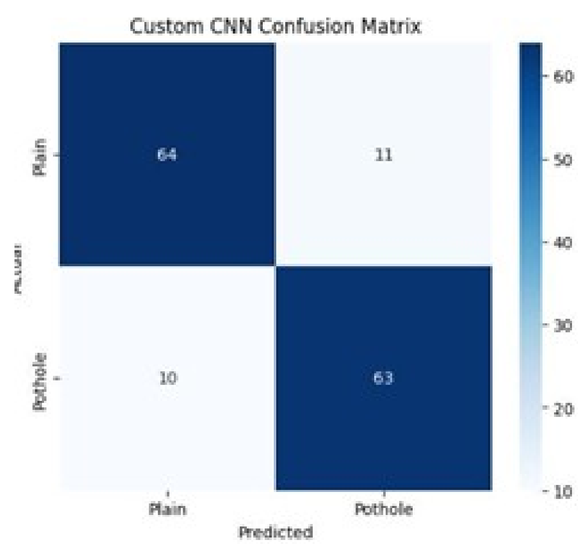

Fig 5. Confusion Matrix – Custom CNN.

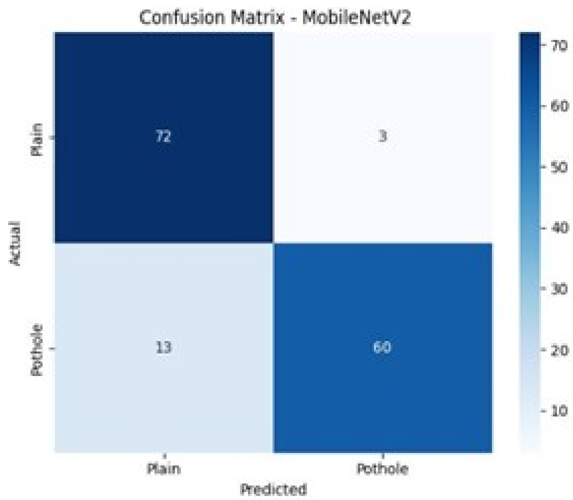

Fig 6. Confusion Matrix – MobileNetV2.

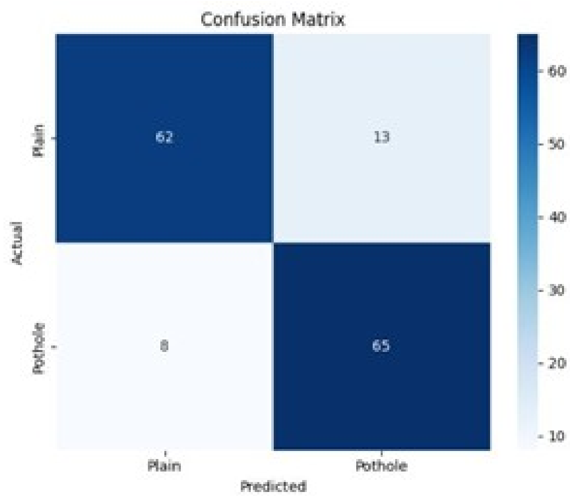

Fig 7. Confusion Matrix – VGG16.

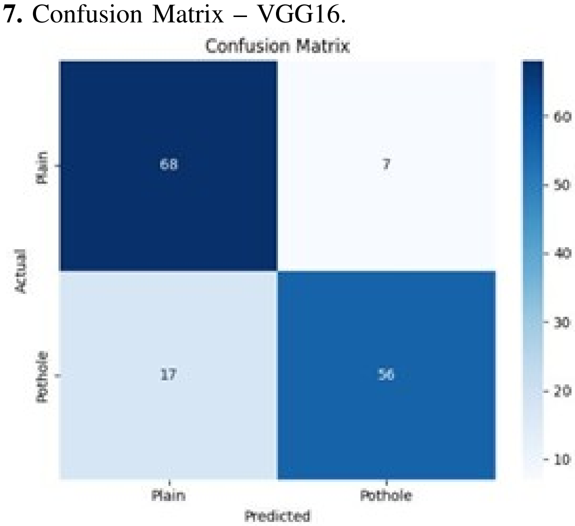

Fig 8. Confusion Matrix – ResNet50.

The Custom CNN (Figure 5) demonstrated a relatively balanced classification ability but showed some confusion between the pothole and plain road classes, leading to a moderate number of misclassifications. MobileNetV2 (Figure 6) outperformed the others in this regard, achieving the fewest misclassifications overall, which indicates strong class separation and a high level of reliability in distinguishing potholes from non-pothole images.

On the other hand, ResNet50 (Figure 7) exhibited a slightly higher number of false negatives for the pothole class, suggesting a lower recall and a tendency to miss some actual potholes. VGG16 (Figure 8) maintained a better balance between precision and recall, although it did produce a few false positives in both classes.

Overall, the confusion matrix analysis reveals that MobileNetV2 provides the most consistent and accurate classification among the evaluated models.

V. Results

This section presents a comparative evaluation of the four classification models: Custom CNN, MobileNetV2, ResNet50, and VGG16. The analysis covers accuracy and loss curves, model-wise performance comparison, training behavior, and trade-offs between complexity and performance.

A. Training Dynamics

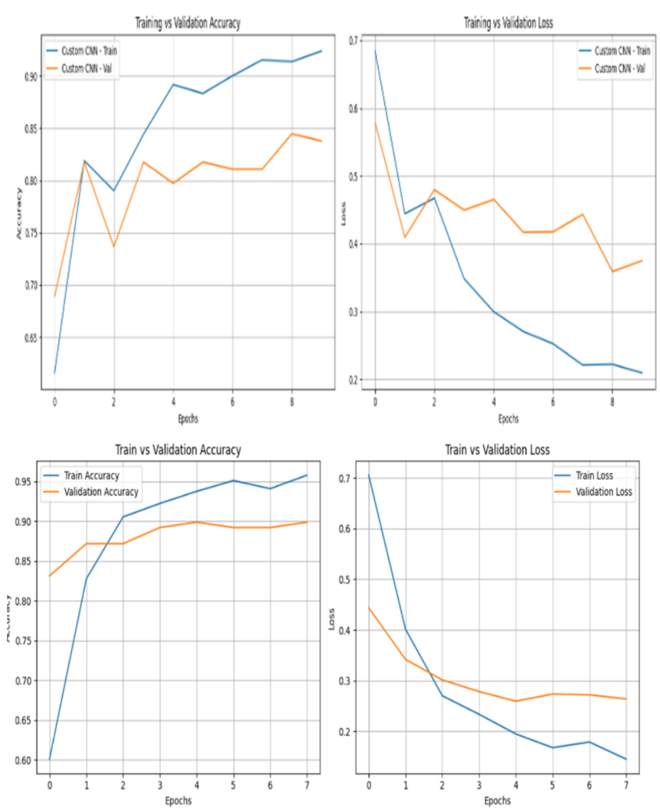

Fig 9. Accuracy and Loss Curve – Custom and Accuracy and Loss Curve – MobileNetV2CNN.

B. Training Dynamics

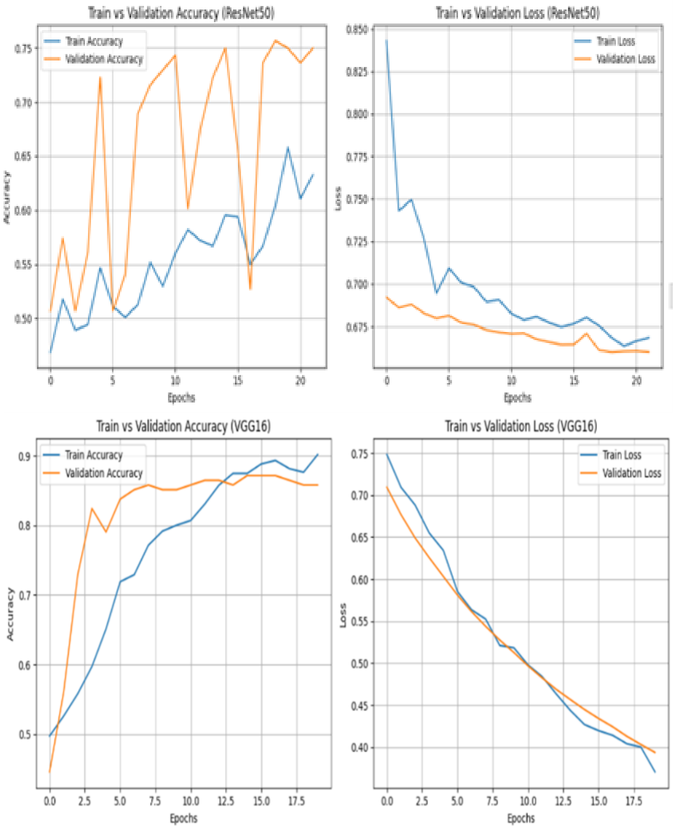

Fig 10. Accuracy and Loss Curve – ReNet50 and Accuracy and Loss Curve – VGG16CNN.

The training and validation curves reveal each model's learning behavior. The Custom CNN (Figure 9) showed smooth convergence with slight overfitting in later epochs. MobileNetV2 (Figuren9) maintained consistent curves, indicating strong generalization. ResNet50 (Figure 10) had some fluctuations, but early stopping prevented overfitting. VGG16 (Figure 10) showed stable training and validation curves, with early stopping triggered by performance plateau. Overall, MobileNetV2 and VGG16 demonstrated the most stable training dynamics.

C. Quantitative Results

Table :2 summarizes the key classification metrics: Accuracy, Precision, Recall, and F1-Score for each model.

TABLE II
MODEL-WISE PERFORMANCE COMPARISON

Model	Accuracy	Precision	Recall	F1-Score
Custom CNN	0.85	0.85	0.85	0.85
MobileNetV2	0.89	0.90	0.89	0.89
ResNet50	0.84	0.84	0.84	0.84
VGG16	0.86	0.86	0.86	0.86

To visually compare the performance, a bar chart is provided (Figure 5), displaying all four metrics side-by-side across the models.

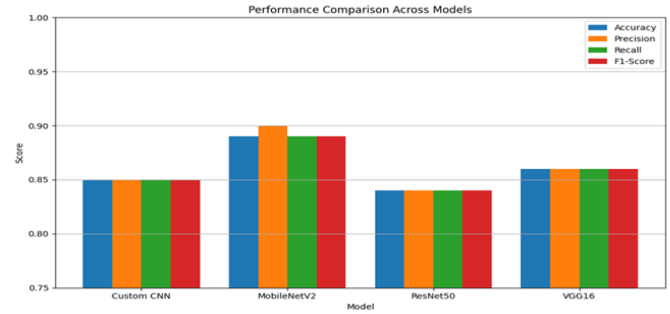

Fig 11. Performance Comparison Bar Chart Across Models.

VI. FINDINGS

This study evaluated the performance of four Convolutional Neural Network (CNN)-based models—Custom CNN, MobileNetV2, ResNet50, and VGG16—for the task of pothole detection using road surface imagery. Each model was assessed using classification metrics including accuracy, precision, recall, and F1-score.

The key findings are summarized as follows:

- **Custom CNN:** Designed and trained from scratch, this model achieved an accuracy of 85%. While computationally efficient, it showed moderate performance in classifying edge cases under varying road textures.
- **MobileNetV2:** Leveraging pretrained weights and depthwise separable convolutions, MobileNetV2 yielded the highest accuracy at 89%. It maintained a strong balance between performance and computational cost, making it well-suited for real-time edge deployment.
- **ResNet50:** Demonstrated an accuracy of 84%. Its deep residual architecture helped in capturing complex spatial patterns, though at the cost of higher training time and resource requirements.
- **VGG16:** Achieved 86% accuracy. Its deeper and more uniform structure required careful regularization but ultimately showed reliable classification capability.

Across all models, MobileNetV2 consistently outperformed others in terms of recall, indicating its robustness in identifying potholes even under variable environmental conditions. Confusion matrices and training curves further validated the classification stability of these models. The results indicate that pretrained CNN architectures, especially MobileNetV2, are effective in balancing accuracy with efficiency in practical pothole detection scenarios.

VII. CONCLUSIONS

This research set out to develop a robust and lightweight deep learning solution for pothole detection that is adaptable to diverse real-world conditions and suitable for deployment on resource-constrained platforms. The experimental results confirmed that transfer learning significantly enhances performance over models trained from scratch.

Among the evaluated models, MobileNetV2 stood out as the most effective architecture, offering high detection accuracy with low computational overhead. The use of augmentation and regularization techniques, along with optimized model selection, ensured improved generalizability and reduced overfitting—addressing a key limitation in many existing pothole detection systems.

The study concludes that CNN-based architectures, particularly those pre-trained on large-scale datasets, are highly capable of automated pothole detection. These models can be effectively integrated into smart city infrastructure, in-vehicle systems, and road monitoring platforms to assist in timely road maintenance and public safety.

Future work may focus on expanding the dataset to include more diverse environmental conditions, optimizing models for real-time embedded deployment, and integrating sensor fusion techniques (e.g., LiDAR or GPS) to further improve robustness and spatial accuracy.

References

[1] M. Kim, J. Lee, and S. Kim, "Deep learning enhanced feature extraction of potholes using vision and LiDAR data for road maintenance," in *Proc. IEEE Int. Conf. Big Data*, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10781387

[2] A. Dib, M. A. Khan, and M. A. Khan, "A review on negative road anomaly detection methods," in *Proc. IEEE Int. Conf. Connected Vehicles and Expo (ICCVE)*, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/9043541

[3] M. Cinar and M. Kaya, "Pothole and plain road classification using adaptive mutation dipper throated optimization and transfer learning for self driving cars," in *Proc. IEEE Int. Conf. Artificial Intelligence and Data Science (AiDAS)*, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9850988

[4] M. Ruseruka, A. M. Ikotun, and A. E. Ezugwu, "Prediction and detection of potholes in urban roads: Machine learning and deep learning based image segmentation approaches," *Machine Learning with Applications*, vol. 10, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666165922000436

[5] A. Ahmed, "Smart pothole detection using deep learning based on dilated convolution," *Sensors*, vol. 21, no. 24, p. 8406, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/24/8406

[6] R. Bharat, A. M. Ikotun, A. E. Ezugwu, L. Abualigah, M. Shehab, and R. A. Zitar, "Augmenting roadway safety with machine learning and deep learning: Pothole detection and dimension estimation using in-vehicle technologies," *Applied Computing and Intelligence*, vol. 6, pp. 750–757, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666827024000239

[7] A. M. Ikotun, A. E. Ezugwu, and R. Bharat, "Review of recent automated pothole-detection methods," *Applied Sciences*, vol. 12, no. 11, p. 5320, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/11/5320

[8] Y. Safyari, M. Mahdianpari, and H. Shiri, "A review of vision-based pothole detection methods using computer vision and machine learning," *Sensors*, vol. 24, no. 17, p. 5652, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/17/5652

[9] M. Asad, M. A. Khan, and A. Khan, "Pothole detection using deep learning: A real-time and AI-on-the-edge perspective," *Advances in Civil Engineering*, 2022. [Online]. Available: https://doi.org/10.1155/2022/9221211

[10] M. Brown, "'Street Bump' app reports potholes," *Wired*, Feb. 16, 2011. [Online]. Available: https://www.wired.com/2011/02/street-bump-app-reports-potholes

[11] D. K. Dewangan and S. P. Sahu, "PotNet: Pothole detection for autonomous vehicle system using convolutional neural network," *Electronics Letters*, vol. 57, no. 2, pp. 53–56, 2020. [Online]. Available: https://doi.org/10.1049/ell2.12062

[12] J. H. Lee, H. Kim, and S. Lee, "A hybrid image segmentation method for accurate measurement of urban environments," *Electronics*, vol. 12, no. 8, p. 1845, 2023. [Online]. Available: https://doi.org/10.3390/electronics12081845

[13] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Image-adaptive YOLO for object detection in adverse weather conditions," *arXiv preprint arXiv:2112.08088*, 2021. [Online]. Available: https://arxiv.org/abs/2112.08088

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012. [Online]. Available: https://papers.nips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[15] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017. [Online]. Available: https://arxiv.org/abs/1704.04861

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [Online]. Available: https://doi.org/10.1038/nature14539

[17] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014. [Online]. Available: https://jmlr.org/papers/v15/srivastava14a.html

[18] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org/

[19] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2016, pp. 3708–3712. [Online]. Available: https://ieeexplore.ieee.org/document/7533052

[20] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 4510–4520. [Online]. Available: https://arxiv.org/abs/1801.04381

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778. [Online]. Available: https://arxiv.org/abs/1512.03385

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*,