Indo Asian Journal of Management and Entrepreneurship (IAJOME)
	ISSN: 2583-2992, Impact Factor 5.007	
	Peer-Reviewed	Open Access	International Journal	
	Volume: 17	Issue: 07	July 2026	www.iajome.com	

Review on Road Damage Detection Using Deep Learning

Ms.Shobana A [1], S. Janani [2],VS. Sujit [3]

[1] Research Scholar, Department of Computer Applications,

Sri Krishna Arts and Science College, Coimbatore

[2,3]Student III BCA-B, Department of Computer Applications,

Sri Krishna Arts and Science College, Coimbatore

***Abstract*—Road transportation is one of the most important parts of modern infrastructure. Safe and well-maintained roads help people travel efficiently and support economic growth. However, road surfaces often develop damages such as cracks, potholes, and surface wear due to heavy traffic, weather conditions, and aging. Traditional road inspection methods are expensive, time-consuming, and require manual labor. Recent developments in Artificial Intelligence (AI) and Deep Learning (DL) have introduced automated systems for road damage detection. This review paper discusses the use of Vision Transformers (ViT) combined with lightweight Convolutional Neural Networks (CNNs) for detecting road damage from images. The proposed approach improves detection accuracy by capturing both local and global image features. The system achieves approximately 94% accuracy and provides an effective solution for transportation safety, infrastructure monitoring, and smart city development.**

Keywords—Artificial Intelligence (AI), Deep Learning, Road Damage Detection, Computer Vision, Convolutional Neural Networks (CNN), Vision Transformer (ViT), Image Processing, Road Infrastructure, Smart Transportation, Pothole Detection.

I. INTRODUCTION

1.1 Importance of Road Infrastructure

Roads are one of the most important components of transportation systems and play a vital role in economic and social development. They facilitate the movement of people, goods, and services between different locations. However, road surfaces deteriorate over time due to heavy traffic, environmental conditions, aging infrastructure, and natural disasters.[1] Common road damages such as cracks, potholes, and surface deformations can reduce driving comfort, increase vehicle maintenance costs, and lead to serious road accidents. Therefore, regular monitoring and timely maintenance of road infrastructure are essential to ensure transportation safety and efficiency.[6]

1.2 The AI Paradigm Shift

Artificial Intelligence (AI), especially Deep Learning (DL), has become an effective technology for automatic road damage detection. Unlike traditional inspection methods that require manual work and expensive equipment, AI systems can analyze road images and identify damages quickly.[3] Models such as Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) help detect cracks, potholes, and other road defects accurately.

Indo Asian Journal of M[illegible]ement and Entrepreneurship (IAJOME)
	ISSN: [illegible]-2992, Impact Factor 5.007	
	Peer-Reviewed	Open Access	International Journal	
	Volume: 17	Issue: 07	July 2026	www.iajome.com	

II. LITERATURE REVIEW

2.1 Traditional Image Processing Approaches:

Early research in road damage detection primarily focused on traditional image processing techniques. These methods include edge detection, thresholding, morphological operations, and texture-based feature extraction. Algorithms such as Canny edge detection and Sobel filters were widely used to identify cracks and surface irregularities in road images.[2]

Although these techniques were simple and computationally efficient, they heavily depended on handcrafted features and fixed rules. As a result, their performance was limited in real-world conditions where lighting variations, shadows, weather changes, and noise significantly affected image quality. These approaches also struggled to generalize across different road environments, making them unsuitable for large-scale deployment.

Fig1(Types of Road Damage)

2.2 Deep Learning-Based Convolutional Neural Networks (CNNs):

With the advancement of Artificial Intelligence, Deep Learning approaches, especially Convolutional Neural Networks (CNNs), have become the dominant method for road damage detection.[7] CNNs automatically learn hierarchical feature representations directly from raw images, eliminating the need for manual feature engineering.

Popular CNN architectures such as VGGNet, ResNet, InceptionNet, MobileNet, Faster R-CNN, SSD, and YOLO have been successfully applied for detecting road cracks, potholes, and surface degradation. Among these, YOLO-based models are widely used for real-time road damage detection due to their fast inference speed and high accuracy.

However, standard CNN models still face limitations in capturing long-range dependencies and global contextual information in complex road scenes.[6] To address these challenges, researchers have introduced lightweight CNN variants optimized for edge devices and mobile applications, improving computational efficiency while maintaining acceptable accuracy.

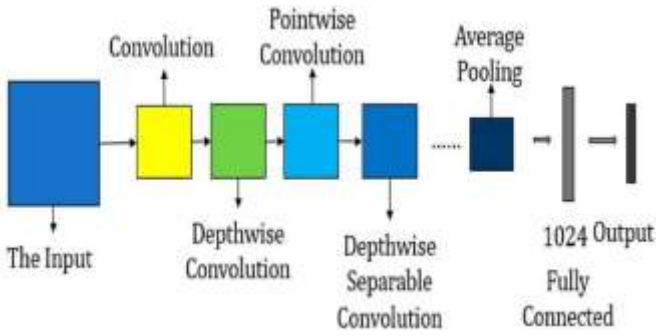


Fig2 (Deep Learning Architectures for Road Damage Detection)

2.3 Road Damage Datasets and Benchmarking:

The development of large-scale datasets has significantly contributed to progress in this field. The Road Damage Dataset (RDD) series is one of the most widely used benchmarks for training and evaluating detection models. RDD2018 initially introduced annotated road images collected from Japan.[7] Later, RDD2020 expanded the dataset by including images from India and the Czech Republic, increasing diversity in road conditions and environments.

The most recent RDD2022 dataset contains more than 47,000 images collected from six different

Indo Asian Journal of Management and Entrepreneurship (IAJOME)
	ISSN: 2319-2992, Impact Factor 5.007	
	Peer-Reviewed	Open Access	International Journal	
	Volume: 17	Issue: 07	July 2026	www.iajome.com	

countries, making it one of the largest publicly available datasets for road damage detection. These datasets include various types of road defects such as longitudinal cracks, transverse cracks, alligator cracks, and potholes.

Fig3 (Sample Images from Road Damage Detection Datasets)

2.4 Transformer-Based and Hybrid Models:

Recent advancements in deep learning have introduced transformer-based architectures for image analysis. Vision Transformers (ViT) utilize self-attention mechanisms to capture global relationships between image patches, enabling better understanding of complex visual patterns in road scenes.[9]

Unlike CNNs, which focus on local features, Vision Transformers excel at modeling long-range dependencies, making them more effective in complex environments. However, they require high computational resources and large datasets for training.[6]

To overcome these limitations, researchers have proposed hybrid models that combine Vision Transformers with lightweight CNN encoders. These hybrid architectures leverage the strengths of both approaches—CNNs for local feature extraction and ViTs for global context understanding—resulting in improved accuracy, efficiency, and robustness for road damage detection systems.[6]

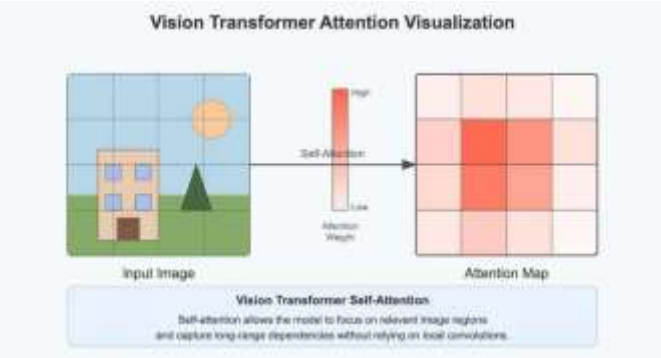


Fig4 (Vision Transformer (ViT) Architecture for Image Analysis)

III. METHODOLOGY

3.1 Data Collection and Dataset Preparation:

The first stage of the proposed system involves collecting and preparing a large and diverse dataset for training the deep learning model.[5] Road images are gathered from publicly available benchmark datasets such as the Road Damage Dataset (RDD2018, RDD2020, and RDD2022), as well as additional images captured using smartphones, vehicle-mounted cameras, and drone-based imaging systems.

The dataset includes various road conditions such as good, satisfactory, poor,

Indo Asian Journal of Management and Entrepreneurship (IAJOME)
	ISSN: 2319-2992, Impact Factor 5.007	
	Peer-Reviewed	Open Access	International Journal	
	Volume: 17	Issue: 07	July 2026	www.iajome.com	

and very poor surfaces. It also contains multiple types of road defects including longitudinal cracks, transverse cracks, alligator cracks, and potholes. This diversity ensures that the model is exposed to different real-world scenarios, improving its generalization capability.[7]

After collection, the dataset undergoes preprocessing steps such as resizing images to a fixed resolution, removing duplicates, and eliminating low-quality or corrupted samples.[8]

Data augmentation techniques such as rotation, flipping, zooming, brightness adjustment, and contrast variation are applied to increase dataset diversity and reduce overfitting.

Finally, the dataset is split into training (70%), validation (15%), and testing (15%) sets to ensure proper model evaluation and unbiased performance measurement.[7]

3.2 Image Preprocessing and Feature Extraction:

Image preprocessing plays a crucial role in improving the performance of the proposed road damage detection system.[10] All input images are resized to a uniform dimension suitable for Vision Transformer (ViT) and Convolutional Neural Network (CNN) architectures. Pixel values are normalized to a standard range to ensure stable and efficient model training.

In the feature extraction stage, a hybrid architecture is used. The lightweight CNN encoder extracts local features such as edges, textures, and fine-grained crack patterns from road images.[3] These features are essential for identifying small and irregular road damages.

Simultaneously, the Vision Transformer processes the same input by dividing the image into fixed-size patches and applying self-attention mechanisms. This allows the model to capture global contextual relationships between different regions of the road surface.[8]

The combination of CNN-based local feature extraction and ViT-based global feature modeling enhances the overall representation capability of the system. This hybrid approach ensures better detection accuracy, especially in complex road environments with varying lighting conditions and damage patterns.[5]

3.3 Model Training Strategy:

- The extracted features are passed into a hybrid Vision Transformer–CNN model for training. The model is trained using supervised learning, where labeled road images guide the learning process.
- The Adam optimizer is used to minimize the loss function efficiently by adjusting learning rates dynamically during training. Categorical Cross-Entropy Loss is applied for multi-class classification of road conditions.
- The training process is carried out over multiple epochs, and performance is monitored using validation accuracy and validation loss. Early stopping techniques are applied to prevent overfitting and ensure optimal generalization.

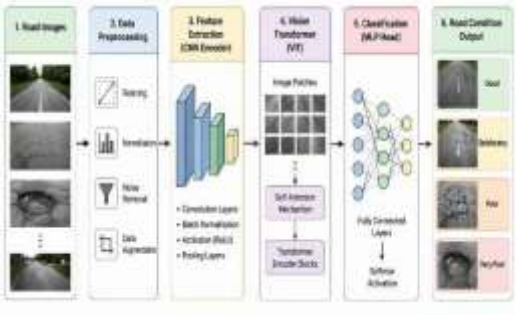
3.4 Model Evaluation and Testing:

- After training, the model is evaluated using the testing dataset, which contains

Indo Asian Journal of Management and Entrepreneurship (IAJOME)
‖ ISSN: 2319-2992, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 17 | Issue: 07 | July 2026 | www.iajome.com ‖

unseen road images. Performance metrics such as accuracy, precision, recall, and F1-score are used to measure the effectiveness of the system.[5]

- A confusion matrix is generated to analyze classification performance across different road condition categories. This helps in identifying misclassification patterns, especially between similar classes such as satisfactory and poor road conditions.
- The final trained model is also compared with existing approaches such as CNN-based models and YOLO-based detectors to demonstrate its improved performance in terms of accuracy and robustness.[7]



Indo Asian Journal of Management and Entrepreneurship (IAJOME)
	ISSN: 2319-2992, Impact Factor 5.007	
	Peer-Reviewed	Open Access	International Journal	
	Volume: 17	Issue: 07	July 2026	www.iajome.com	

Fig5 (Methodology of AI-Based Road Damage Detection)

IV. ALGORITHMS

4.1 Vision Transformer (ViT):

Vision Transformers divide input images into small patches called tokens. These tokens are processed using self-attention mechanisms to understand relationships between different parts of the image.Unlike traditional CNNs that focus on local features, Vision Transformers analyze the entire image globally.[4] This helps the model capture complex road damage patterns and improves classification accuracy.

4.2 Lightweight CNN Encoder:

The lightweight CNN encoder extracts important low-level features such as edges, textures, cracks, and potholes from road images. It reduces computational complexity while maintaining high performance. The extracted features are then forwarded to the Vision Transformer for further analysis.[4]

4.3 Self-Attention Mechanism:

The self-attention mechanism calculates the relationship between image patches and assigns importance scores to each region.[7] This allows the model to focus on critical damage areas while ignoring unnecessary background information. As a result, the system can detect road defects more accurately.

4.4 Optimizer and Loss Function:

The Adam optimizer is used during the training process to update model weights efficiently. It combines adaptive learning rates and momentum techniques to improve convergence speed and overall model performance. For classification, the Categorical Cross-Entropy Loss function is applied to minimize prediction errors and enhance classification accuracy.[5] This enables the model to accurately categorize road conditions into classes such as Good, Satisfactory, Poor, and Very Poor.Improves training efficiency and reduces the time required for model convergence. Enhances the model's ability to generalize and perform accurately on unseen road images. Reduces overfitting and improves the stability of the learning process.[1] Increases classification accuracy by minimizing prediction errors during training

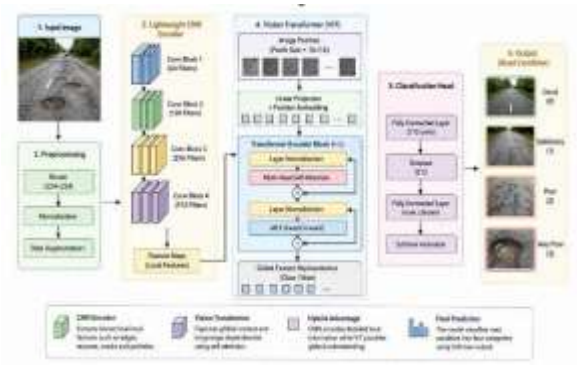


Fig6 (Hybrid Vision Transformer and CNN Architecture for Road Damage Detection.)

V. RESULT

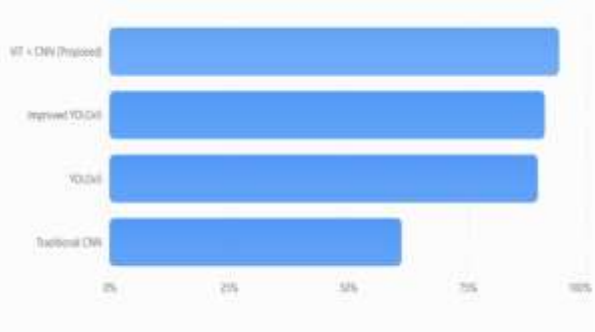


Fig 7. (Comparative Performance of Road

Indo Asian Journal of Management and Entrepreneurship (IAJOME)
	ISSN: 2319-2992, Impact Factor 5.007	
	Peer-Reviewed	Open Access	International Journal	
	Volume: 17	Issue: 07	July 2026	www.iajome.com	

Damage Detection Algorithms)

Model/Algorithm	Accuracy (%)
ViT + CNN (Proposed)	94.00
Improved YOLOv5	91.00
YOLOv5	89.51
Traditional CNN	61.00

Table 1 Result

VI. CONCLUSION

Road damage detection is essential for maintaining safe and efficient transportation systems. Traditional inspection methods are often time-consuming, costly, and dependent on manual efforts. This review highlights the effectiveness of Artificial Intelligence, particularly Vision Transformers (ViT) and lightweight CNN encoders, in automating road damage detection. By combining advanced feature extraction and self-attention mechanisms, the proposed approach achieves high accuracy in identifying different road conditions.[7] The results demonstrate that AI-based systems can improve road safety, reduce maintenance costs, and support smart infrastructure management. Future advancements in deep learning and real-time monitoring technologies will further enhance the reliability and efficiency of road damage detection systems.

REFERENCES

[1] Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashiyama, T., & Sekimoto, Y. (2021). *Deep Learning-Based Road Damage Detection and Classification for Multiple Countries*. Transportation Research Part C: Emerging Technologies, 132, 103377.

[2] Zhang, Y., Wang, H., & Li, X. (2025). *Road Damage Detection Method Based on Improved YOLO-World*. Proceedings of the International Conference on Intelligent Computing (IC-ICC 2025), Ningbo, China.

[3] Kumar, R., Singh, A., & Sharma, P. (2021). *Edge AI-Based Automated Detection and Classification of Road Anomalies in VANET Using Deep Learning*. Computational Intelligence and Neuroscience, 2021, Article ID 6262194.

[4] Patel, S., Verma, K., & Gupta, R. (2023). *Detecting Danger: AI-Enabled Road Crack Detection for Autonomous Vehicles*. E3S Web of Conferences, ICMPC 2023.

[5] Chakurkar, P. S., Vora, D., Patil, S., Mishra, S., & Kotecha, K. (2023). *Data-Driven Approach for AI-Based Crack Detection: Techniques, Challenges, and Future Scope*. Frontiers in Sustainable Cities, 5, 1253627.

[6] Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashiyama, T., & Sekimoto, Y. (2024). *RDD2022: A Multi-National Image Dataset for Automatic Road Damage Detection*. Geoscience Data Journal, 11(1).

[7] Kulambayev, B., Satybaldina, D., & Tleubergenova, M. (2024). *Real-Time Road Damage Detection System on Deep Learning Based Image Analysis*. International Journal of Advanced Computer Science and Applications (IJACSA), 15(9).

[8] Abdelwahed, S. H., Sharobim, B. K., Wasfey, B., & Said, L. A. (2025). *Advancements in Real-Time Road*

Indo Asian Journal of Management and Entrepreneurship (IAJOME)
‖ ISSN: 2319-2992, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 17 | Issue: 07 | July 2026 | www.iajome.com ‖

Damage Detection: A Comprehensive Survey of Methodologies and Datasets. Journal of Real-Time Image Processing, 22(137).

[9] Wan, F., Sun, C., He, H., Lei, G., Xu, L., & Xiao, T. (2022). *YOLO-LRDD: A Lightweight Method for Road Damage Detection Based on Improved YOLOv5s*. EURASIP Journal on Advances in Signal Processing, 2022(98).

[10] Zareen, S., Suha, S. H., Hossain, K., & Bhuiyan, T. (2025). *AI-Powered Road Damage Detection for Enhanced Safety and Life Protection*. World Journal of Advanced Research and Reviews, 27(1), 2169–2180.