BAREKENG: Journal of Mathematics and Its Applications
September 2026 Volume 20 Issue 3 Page 2681-2692
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss3pp2681-2692

COMPARATIVE EVALUATION OF CNN ARCHITECTURES FOR ROAD DAMAGE CLASSIFICATION USING DIGITAL IMAGES IN SLEMAN REGENCY

Anggun Puspita Sari [1], **Kariyam** [2*], **Feri Wijayanto** [3], **Edy Widodo** [4]

[1,2,4]*Statistics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia*
[3]*Informatics Study Program, Faculty of Industrial Technology, Universitas Islam Indonesia*
Jln. Kaliurang KM 14.5, Sleman, Special Region of Yogyakarta, 55584, Indonesia

*Corresponding author's e-mail: * **kariyam@uii.ac.id***

Article Info

Article History:

Received: 19th October 2025
Revised: 19th January 2026
Accepted: 17th March 2026
Published: 8th April 2026

Keywords:

CLAHE;
Convolutional Neural Networks;
Comparative Evaluation;
Road damage classification;
Sleman Regency.

ABSTRACT

Reliable road condition monitoring is fundamental to maintenance decision-making and transportation safety, particularly in regional contexts where data resources are often scarce. This study presents a comparative evaluation of convolutional neural network (CNN) architectures for classifying road damage types using digital images collected in Sleman Regency. Three widely used CNN architectures, VGGNet-16, InceptionV3, and Xception, were evaluated under a unified experimental framework employing transfer learning, consistent preprocessing, explicit hyperparameter tuning, and four-fold cross-validation. The dataset comprises three road damage categories, alligator crack, corrugation, and pothole, captured under heterogeneous pavement and lighting conditions. Image preprocessing includes resizing, augmentation, and contrast enhancement using Contrast Limited Adaptive Histogram Equalization (CLAHE). To assess the contribution of preprocessing choices, an ablation study was conducted by comparing model performance with and without CLAHE. Experimental results indicate that all evaluated architectures achieve high classification performance. Among them, Xception consistently demonstrates the best overall performance across validation and test sets, achieving the highest accuracy, precision, recall, and F1-score. The ablation analysis further shows that including CLAHE consistently improves performance, particularly in recall and F1-score, indicating enhanced robustness under uneven illumination conditions. Although the contribution of this study is incremental rather than algorithmically novel, the findings provide empirical insights into the comparative behavior of CNN architectures under region-specific road conditions. The results highlight the importance of systematic preprocessing and controlled evaluation in CNN-based road-damage classification and provide practical guidance for selecting suitable architectures for regional road maintenance decision-support systems.

How to cite this article:

A. P. Sari, Kariyam, F. Wijayanto and E. Widodo., "COMPARATIVE EVALUATION OF CNN ARCHITECTURES FOR ROAD DAMAGE CLASSIFICATION USING DIGITAL IMAGES IN SLEMAN REGENCY", *BAREKENG: J. Math. & App.,* vol. 20, no. 3, pp. 2681-2692, Sep, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · **Open Access**

1. INTRODUCTION

Effective monitoring of road conditions is a critical component of transportation safety, infrastructure sustainability, and maintenance decision-making at both national and regional levels. Road surface deterioration, including alligator cracking, corrugation, and potholes, directly affects driving comfort and safety while increasing vehicle operating costs and long-term maintenance expenditures if left unaddressed [1], [2], [3]. In rapidly developing regions, the ability to identify and classify road damage efficiently is essential to support timely maintenance interventions and optimize limited infrastructure budgets.

Conventional road condition assessment methods primarily rely on manual field surveys and visual inspections conducted by trained personnel. Although such approaches are widely used, they are inherently time-consuming, labor-intensive, and prone to subjectivity, particularly when applied across large or heterogeneous road networks [2], [4]. These limitations have motivated increasing interest in automated and semi-automated approaches that leverage digital data to support road condition monitoring. Among these, digital image–based analysis has emerged as a scalable and cost-effective alternative for assessing road surface conditions at the regional level [5], [6]. Early studies in image-based road damage analysis commonly employed classical image processing and statistical techniques, such as texture analysis, edge detection, and dimensionality reduction methods, including principal component analysis (PCA) [7], [8]. While these methods provide useful baseline information, their performance is often sensitive to variations in lighting conditions, pavement texture, and environmental noise, which are common in real-world road environments [7], [9]. As a result, their robustness and generalizability remain limited, particularly when applied to datasets collected under heterogeneous field conditions.

The rapid advancement of deep learning has significantly transformed image analysis, with Convolutional Neural Networks (CNNs) becoming the dominant approach for visual classification. CNNs can automatically learn hierarchical feature representations directly from raw image data, reducing the need for handcrafted feature extraction and improving classification performance [10], [11]. In recent years, CNN-based models have been successfully applied to road damage classification using digital images in both international and Indonesian research contexts [6], [12], [13]. These studies demonstrate that CNNs can effectively capture visual patterns associated with different types of road damage, even under varying pavement conditions. To address data scarcity, a common challenge in regional road monitoring applications, several studies have adopted transfer learning strategies that leverage pretrained CNN architectures such as VGGNet-16. Transfer learning enables models to reuse generic visual features learned from large-scale datasets, thereby improving classification accuracy and training stability when domain-specific data are limited [14], [15]. Despite these advances, many existing studies focus on evaluating a single CNN architecture or report results without a controlled comparative framework. Consequently, it remains difficult to determine which CNN architecture is most suitable for specific regional road characteristics and data constraints [12], [14].

In parallel with classification-based approaches, recent international research has increasingly emphasized object–detection–based methods, particularly variants of the You Only Look Once (YOLO) framework, for real-time road-damage detection and localization [16], [17], [18]. These approaches offer strong performance in identifying both the type and spatial location of road damage and are well-suited for large-scale monitoring and intelligent transportation systems. However, YOLO-based methods typically require large, well-annotated datasets and substantial computational resources to achieve optimal performance [19], [20]. As a result, image-level classification approaches remain relevant, especially for regional applications where the primary objective is damage categorization and decision support rather than precise localization [11], [21].

Despite the growing body of literature on CNN-based road damage analysis, a clear research gap persists in systematically evaluating CNN architectures across region-specific road and environmental conditions. Regional datasets often exhibit heterogeneous pavement materials, uneven illumination, and limited sample sizes, all of which can significantly influence model behavior and performance [22], [23]. Moreover, the role of preprocessing strategies, such as contrast enhancement, in improving classification robustness is frequently underreported or treated as a secondary consideration in existing studies [24], [25]. To address these limitations, this study presents a comparative evaluation of CNN architectures for road damage classification using digital images collected in Sleman Regency, Indonesia. Three widely used CNN architectures, VGGNet-16, InceptionV3, and Xception, are evaluated within a unified experimental framework that incorporates consistent image preprocessing, transfer learning, explicit hyperparameter

tuning, and four-fold cross-validation. In addition, the contribution of contrast enhancement through Contrast Limited Adaptive Histogram Equalization (CLAHE) is systematically examined using an ablation study to assess its impact on classification performance [9], [18].

The primary objective of this research is to assess the relative effectiveness of different CNN architectures for classifying road damage types under regional road and environmental conditions characterized by limited, heterogeneous data. By providing a controlled, transparent comparative evaluation, this study aims to offer empirical guidance for selecting suitable CNN architectures for regional road maintenance decision-support systems. Furthermore, the findings complement detection-oriented studies by highlighting the continued relevance of classification-based approaches for infrastructure monitoring in resource-constrained settings.

2. RESEARCH METHODS

This study employed transfer learning by initializing the CNN architectures with ImageNet pre-trained weights. During the initial training phase, the base layers of each architecture were frozen to preserve learned generic features, while the newly added classification layers were trained. Fine-tuning was subsequently applied by unfreezing selected upper convolutional layers to better adapt the models to road-damage features. This research was conducted through several stages, as shown in Fig. 1. The first stage is planning, which involves identifying the problem to determine the research focus, in this case, road damage detection. Next, a literature review was conducted to identify prior research on road image detection and the theoretical basis for appropriate CNN architectures. Data collection was conducted using a sampling design, with particular attention to regional representation.

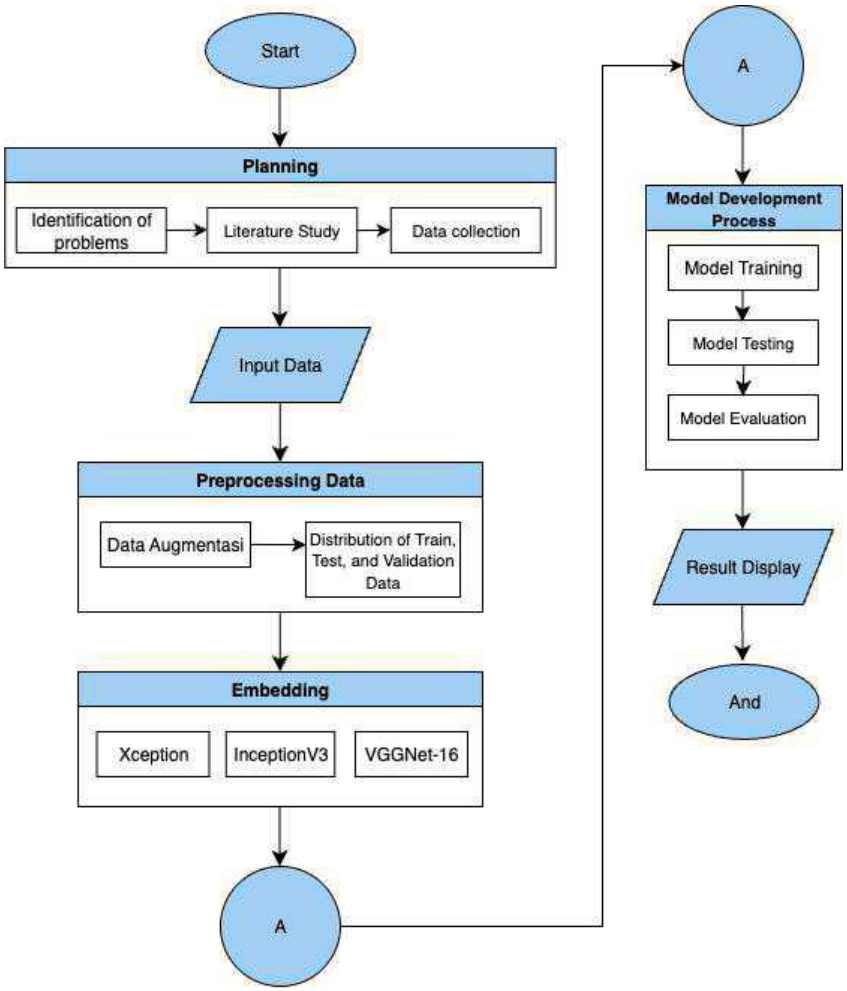

Figure 1. Stages of Research Methods

The next stage is data input, where all collected data is fed into the system to prepare it for processing. Pre-processing is performed using data augmentation and data division into training, validation, and test sets.

Data augmentation is used to increase image variation through transformations such as rotation, flipping, zooming, and lighting adjustments, thereby improving model generalization. Each dataset is divided based on predetermined proportions to ensure balanced model evaluation. The next stage is embedding, the process of extracting image features using several CNN architectures, including Xception, InceptionV3, and VGGNet-16. Each architecture contributes to generating feature representations that are then used in the model development stage.

The embedding results serve as input for the model development phase, which is built through three main processes: training, validation, and evaluation. In the training phase, the model learns from the training data to adjust the network's weights and parameters. The testing phase uses test data to assess the model's generalization ability to new data. Model evaluation is performed by calculating performance metrics such as accuracy, precision, recall, and F1-score to determine the best performance of each CNN architecture. The final stage is the presentation of results, where model performance is visualized in graphs, tables, and comparative analysis. These evaluation results are used as a basis for drawing conclusions and providing recommendations for further model development.

2.1 Data

The data used in this study consists of digital images of road damage. The data were collected from five sub-districts in Sleman Regency. The images include three types of road damage, namely alligator crack, corrugation, and pothole, each exhibiting distinct visual patterns and structures, as shown in Fig. 2. In Fig. 2 (a), Alligator Crack is shown, characterized by interconnected cracks resembling the pattern of an alligator's skin. Fig. 2 (b) depicts Corrugation, a series of regular, wave-like ridges on the road surface that can affect vehicle comfort. In Fig. 2 (c), Pothole damage is evident, with cavities or depressions forming on the asphalt surface, which may pose safety risks for vehicles.

(a) (b) (c)

Figure 2. Types of Road Damage (a) Alligator, (b) Corrugation, (c) Pothole

The dataset consists of a total of 600 original road damage images, evenly distributed across three classes: alligator crack, corrugation, and pothole, with 200 images per class. To address data scarcity and improve model generalization, data augmentation was applied exclusively to the training data (180 or 90%), generating 1,800 augmented images per class via geometric and photometric transformations. As a result, the augmented training set contains 5,400 image variations, while all validation and test samples remain original, non-augmented images. This balanced class distribution ensures that performance differences across classes are not influenced by sample size imbalance.

Road damage image data collection was carried out in a planned manner, accounting for spatial representation and variations in road conditions across five study areas. The selection of five sub-districts as sampling locations was based on several criteria: (1) varying levels of traffic density, (2) differences in road network characteristics (main roads and neighborhood roads), and (3) variations in road damage levels observed based on reports from relevant agencies and preliminary surveys. This approach aims to ensure that the collected data represents a diverse and realistic range of road damage conditions. Image acquisition was carried out directly in the field under relatively controlled conditions. Images were taken with a camera-to-road distance of approximately 1–1.5 meters from the road surface, and the shooting angle is close to perpendicular (± 60-90°) to the road surface to minimize perspective distortion. Lighting provides natural light, with shooting times from morning to noon to avoid extreme shadows and low lighting conditions. Variations in lighting conditions are maintained to increase the model's robustness to real conditions in the field.

2.2 Image Preprocessing

Image preprocessing is a critical step in image-based road damage classification, as variations in pavement texture, illumination, and acquisition conditions can significantly influence model performance. Effective preprocessing improves visual quality, enhances discriminative features, and facilitates more accurate feature learning by convolutional neural networks (CNNs), particularly when working with limited and heterogeneous datasets [8], [24].

In this study, image enhancement was first performed using Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve local contrast in road surface images. CLAHE was selected because road damage patterns, such as cracks and surface irregularities, often exhibit low intensity differences under uneven outdoor lighting conditions. Unlike global histogram equalization, CLAHE enhances contrast locally while limiting excessive noise amplification, making it well-suited for road damage imagery [9], [15]. CLAHE was applied with a clip limit of 2.0 and a tile grid size of 8 × 8, where the clip limit controls contrast amplification, and the tile grid divides the image into small regions for localized histogram computation [8].

Furthermore, CLAHE was applied either to grayscale images or to the Lightness (L) channel in the LAB color space, using a histogram resolution of 256 intensity levels, which corresponds to standard CNN input representations. This configuration enhances luminance information without altering chromatic components, allowing clearer visualization of crack boundaries and surface textures while preserving overall image quality [18], [24]. Fig. 3 illustrates the visual impact of CLAHE, where fine crack patterns that are difficult to observe in the original image become more distinguishable after contrast enhancement. This visual improvement supports more effective feature extraction during CNN training [9].

(a) (b)

Figure 3. Original Image (a) and CLAHE Results (b)

After contrast enhancement, all images were resized to 299 × 299 pixels to match the input requirements of the pretrained CNN architectures. Images were then cropped to focus on damaged regions, reducing background interference and improving learning efficiency. To increase data variability and strengthen model generalizability, data augmentation was applied exclusively to the training data, including horizontal and vertical flipping, ±90° image rotation, zooming, and shifting. These augmentation strategies are commonly used in CNN-based road damage classification to mitigate overfitting and improve robustness under limited data conditions [14], [21].

Finally, the dataset was divided into a held-out test set (10%) and a training–validation set (90%). Four-fold cross-validation was applied only to the training–validation subset, and performance metrics were averaged across folds to obtain robust estimates. To prevent data leakage, augmentation was applied solely to the training folds, while validation and test sets consisted only of original images [14].

2.3 CNN Model Implementation

This study employs transfer learning to improve classification performance under limited data conditions, which is a common challenge in regional road damage analysis [14], [24]. Three widely used convolutional neural network architectures, VGGNet-16, InceptionV3, and Xception, were utilized, all initialized with ImageNet-pretrained weights rather than being trained from scratch. This strategy enables the models to leverage generic visual representations learned from large-scale datasets while adapting to the specific characteristics of road damage imagery [10], [11]. For each architecture, the original classification layers were removed and replaced with a custom classification head designed for three road damage classes: alligator crack, corrugation, and pothole. The added head consists of a global average pooling layer, followed

by a dense layer, a dropout layer for regularization, and a final softmax output layer to generate class probabilities. This design is widely adopted in CNN-based road damage classification to reduce model complexity while maintaining discriminative feature learning [12], [21].

Model training was conducted using the Adam optimizer with the categorical cross-entropy loss function, which is appropriate for multi-class classification problems [10], [24]. Hyperparameters such as learning rate, dropout rate, and the number of dense units were optimized empirically based on validation performance. To mitigate overfitting, dropout regularization was applied primarily within the classification head, while during the initial training stage, the convolutional base of each pretrained network was frozen to preserve learned feature representations. Subsequently, fine-tuning was applied selectively to higher-level layers to further adapt the models to the road damage dataset, as is commonly practiced in transfer learning–based image classification studies [14], [15]. Model performance was evaluated using accuracy, precision, recall, and F1-score, all derived from the confusion matrix. During four-fold cross-validation, these metrics were computed separately for each fold using the corresponding validation set, and the reported results are macro-averaged across all folds to ensure balanced evaluation across classes [14], [24]. Accuracy was calculated as the overall proportion of correctly classified samples. After model selection, the final trained model was evaluated once on a held-out test set using the same metrics. In addition, confusion matrices were analyzed to examine class-wise predictive behavior and identify common misclassification patterns, which is particularly important for understanding errors between visually similar road damage categories [9], [11].

3. RESULTS AND DISCUSSION

This section presents the experimental results obtained from applying different convolutional neural network architectures to classify road damage types. The analysis focuses on the effect of image preprocessing and the comparative performance of VGGNet-16, InceptionV3, and Xception under the same training and validation framework.

3.1 Image Preprocessing Results

Image preprocessing plays an important role in improving the quality of input data for CNN-based classification. In this study, preprocessing was consistently applied across all experiments to ensure a fair comparison between architectures. The preprocessing pipeline includes image resizing, sharpening, and data augmentation. All images were resized to 299 × 299 pixels to match the input requirements of the pretrained CNN architectures and to ensure computational efficiency. Image sharpening was applied to enhance visual clarity, particularly to emphasize crack edges and pothole boundaries. Data augmentation techniques, including rotation, horizontal flipping, zooming, and shifting, were applied only to the training data to increase variability and improve model generalization without introducing new road segments.

The dataset was split into a held-out test set (10%) and a training–validation subset (90%), with four-fold cross-validation applied exclusively to the training–validation data, as described in the Methodology section. This strategy ensures robust performance estimation while preventing data leakage. Qualitatively, preprocessing, especially contrast enhancement and augmentation, improves the visibility of pavement distress patterns under varying lighting conditions. These improvements are reflected in the classification performance discussed in the following subsection.

3.2 Model Configuration and Hyperparameter Tuning

Hyperparameter configuration was designed to optimize model performance while preventing overfitting under limited data conditions. To ensure a fair comparison across architectures, all models were trained with the same baseline settings: a batch size of 32, the Adam optimizer, and a categorical cross-entropy loss function. Hyperparameter tuning focused on the parameters with the greatest impact on generalization performance: the learning rate, dropout rate, and number of dense units in the classification head. Due to the limited dataset size, tuning was conducted using manual, iterative adjustment, guided by validation accuracy and a macro-averaged F1-score obtained from four-fold cross-validation. The learning rate was evaluated over the range [1×10^{-4}–1×10^{-3}], dropout rates over [0.3–0.5], and dense units over [128–256].

The final hyperparameter configuration was selected based on the best average validation performance across folds and was applied consistently to all CNN architectures evaluated in this study. This configuration

balances learning stability and model flexibility, enabling effective feature learning without excessive sensitivity to training data. These hyperparameters are detailed in Table 1.

Table 1. Final Hyperparameter Configuration

Hyperparameter	Value
Batch size	32
Epochs	10
Optimizer	Adam
Learning rate	0.001
Dropout rate	0.3-0.5
Dense units	128-256
Activation (output)	ReLU, Softmax

3.3 Model Development Process

Fig. 4 illustrates the training and validation performance of the Xception architecture using four-fold cross-validation.

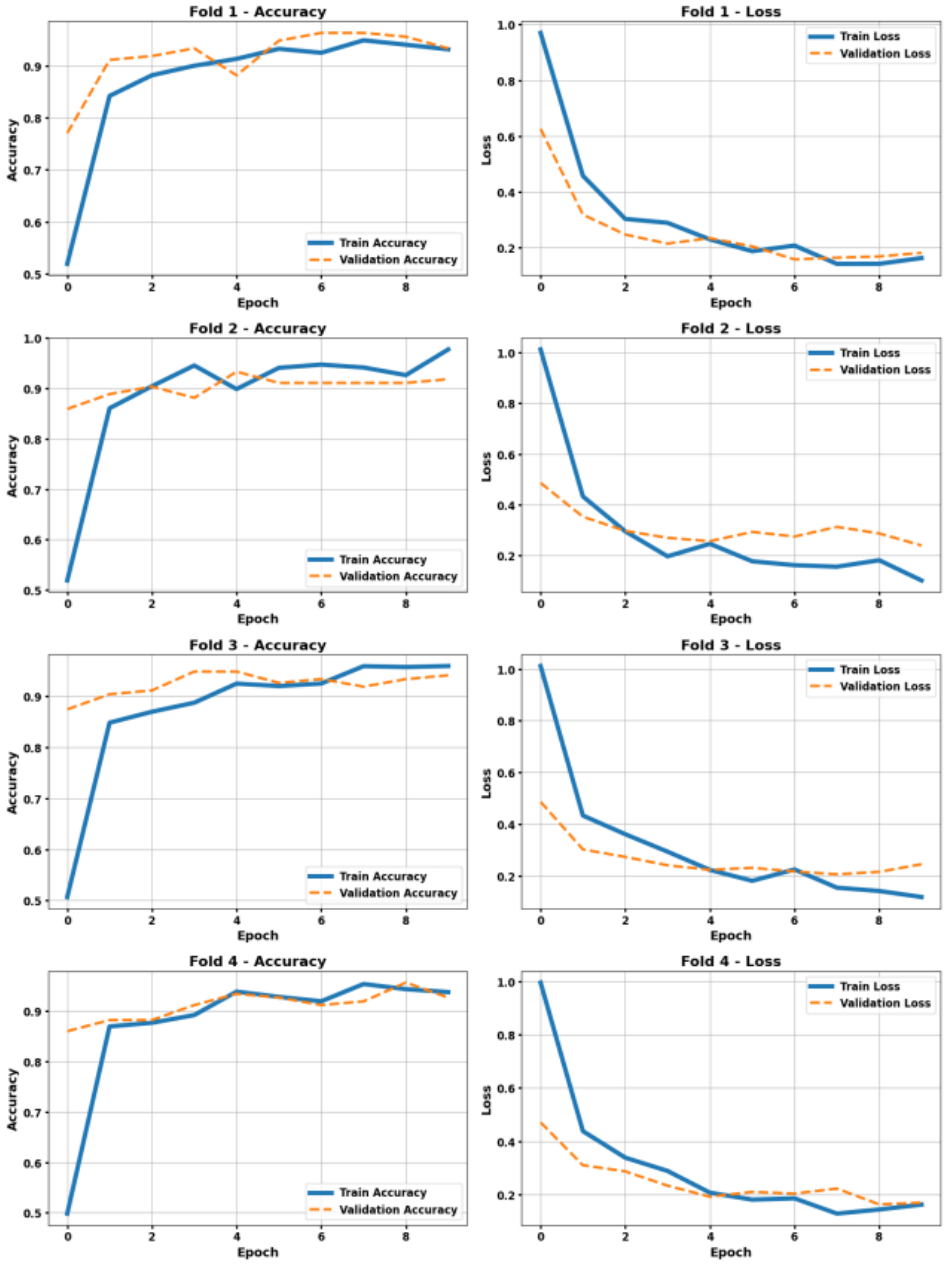

Figure 4. Xception Architecture Graph

Across all folds, both the accuracy and loss curves exhibit stable, consistent convergence. Training and validation accuracy increased rapidly during the early epochs and stabilized at 0.92–0.96, while the

corresponding loss values decreased steadily and converged to below 0.2. The close alignment between training and validation curves indicates that the model generalizes well and does not exhibit noticeable overfitting.

Class-wise predictive behavior was further examined using confusion matrices for the validation and test datasets. Figure 5 presents the confusion matrix obtained from the validation data during four-fold cross-validation using the Xception architecture. The matrix shows a strong diagonal concentration of samples, indicating that the model achieves high classification accuracy across all three road damage classes during validation Fig. 5 shows that the model learns discriminative features effectively across classes, with misclassifications primarily occurring between visually similar damage types. The relatively balanced error distribution and absence of class dominance indicate stable learning behavior and good generalization across validation folds.

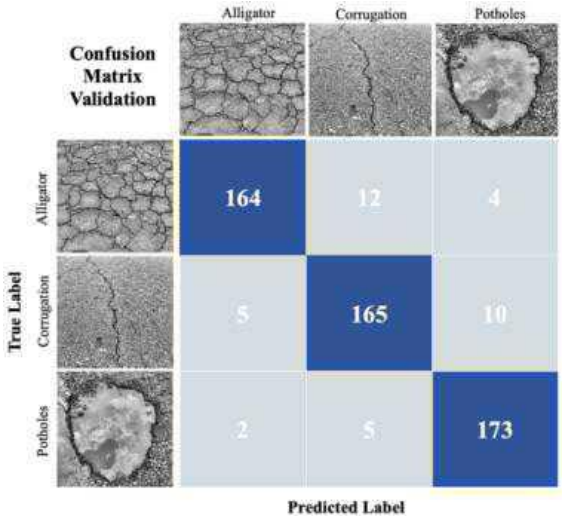

Figure 5. Confusion Matrix of The Validation Data

Fig. 6 presents the confusion matrix obtained from the held-out test dataset using the Xception architecture. This dataset was completely excluded from training, validation, and hyperparameter tuning, providing an unbiased assessment of the model's generalization capability on unseen data. Overall, the test confusion matrix shows a strong diagonal dominance with very limited misclassification across all classes. Compared to the validation results, the error rate on the test data is similarly low, indicating stable decision boundaries and no evidence of overfitting. The consistency between validation and test confusion matrices confirms that the proposed preprocessing and training strategy enables robust generalization for road damage classification in regional settings.

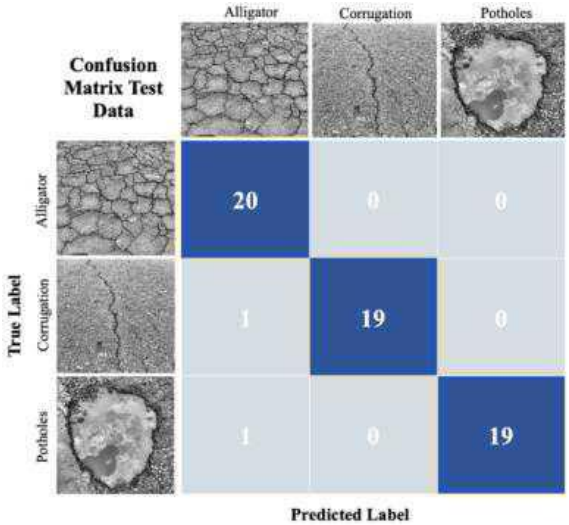

Figure 6. Confusion Matrix of The Testing Data

A closer examination of the misclassification patterns reveals that most errors occur between visually similar road damage types, particularly alligator cracks and corrugations, as well as occasional confusion between corrugations and potholes. These misclassifications are primarily observed when crack density is low, damage boundaries are partially occluded, or surface textures appear shallow, thereby reducing the visual distinction between damage types. In the validation data, corrugation exhibits the highest misclassification rate, which can be attributed to its transitional visual characteristics that overlap with both crack-based and depression-based damage patterns. In contrast, potholes demonstrate the lowest error rate, reflecting their more distinctive geometric and textural features, such as localized cavities and high-contrast boundaries. The significantly lower error rate in the test set compared to the validation set indicates that the learned decision boundaries remain stable and generalize well to unseen data.

A comparative performance analysis of all evaluated architectures is summarized in Table 2. Among the three models, Xception achieved the highest performance, with a cross-validation accuracy of 92.96% and a test accuracy of 96.67%. InceptionV3 closely followed, with cross-validation accuracy of 92.4% and test accuracy of 95%, while VGGNet-16 achieved 92% and 91.7%, respectively.

Table 2. Cross-Validation Accuracy and Test Confusion Matrix

Architecture	Accuracy	
	Cross-Validation Accuracy	**Test Confusion Matrix**
InceptionV3	92.4%	95%
Xception	92.96%	96.67%
VGGNet-16	92%	91.7%

The superior performance of Xception can be attributed to its use of depthwise separable convolution, which enables efficient feature extraction with fewer parameters compared to conventional convolutional layers. This architectural design is particularly advantageous in limited-data settings, enabling the model to capture discriminative road-damage features while reducing the risk of overfitting. Although InceptionV3 demonstrated comparable performance, its slightly lower accuracy suggests that Xception offers a better balance between model complexity and generalization for regional road damage classification.

3.4 Ablation Study on CLAHE Preprocessing

To assess the contribution of Contrast Limited Adaptive Histogram Equalization (CLAHE) in the preprocessing pipeline, an ablation study was conducted focusing on the Xception architecture, which demonstrated the best overall performance in previous experiments. Two configurations were evaluated: Xception with CLAHE enabled and Xception without CLAHE. All other components, including image resizing, normalization, data augmentation policy, number of training epochs, optimizer, learning rate, and the four-fold cross-validation protocol, were kept identical to ensure a fair and controlled comparison. Table 3 summarizes the average performance across cross-validation folds for both configurations.

Table 3. Comparison the Average Performance for Xception

Setting	Accuracy	Precision	Recall	F1-Score
Xception with CLAHE	96.67%	0.9697	0.9667	0.9671
Xception without CLAHE	91.67%	0.9235	0.9167	0.9176
Delta $\Delta = with - without$	+5%	+0.0462	+0.0500	+0.0495

The results indicate that applying CLAHE consistently improves all evaluation metrics, including a 5% absolute increase in accuracy. More importantly, recall and F1-score show the largest relative gains, highlighting CLAHE's positive impact on class-balanced recognition. This suggests that CLAHE improves the model's ability to detect road damage instances that might otherwise be missed under uneven illumination conditions.

A class-wise inspection using confusion matrices provides further insight into this improvement. Without CLAHE, most misclassifications occur between corrugation and alligator crack classes, as well as between pothole and corrugation, particularly when crack patterns are shallow or partially occluded. By enhancing local contrast in pavement textures, CLAHE reduces these ambiguities, leading to fewer false negatives and improved recall across all damage categories. From a practical perspective, this improvement is critical for road maintenance applications, where undetected damage may delay necessary repair actions.

Furthermore, this ablation study demonstrates that the observed performance gains are not solely attributable to architectural design, but also to appropriate preprocessing choices. Even with a modern, efficient CNN architecture such as Xception, systematic preprocessing optimization remains crucial for improving robustness and reliability. These findings reinforce the importance of integrating contrast enhancement techniques into CNN-based road damage classification pipelines, particularly when working with limited and heterogeneous datasets.

Model evaluation using precision, recall, and F1-score for each CNN architecture is visualized in Fig. 7. The results indicate that Xception consistently outperforms the other architectures across all evaluation metrics. Xception achieved an accuracy of approximately 97%, with precision, recall, and F1-score values all reaching 0.97, demonstrating a well-balanced classification performance in terms of both correctness and sensitivity across damage classes.

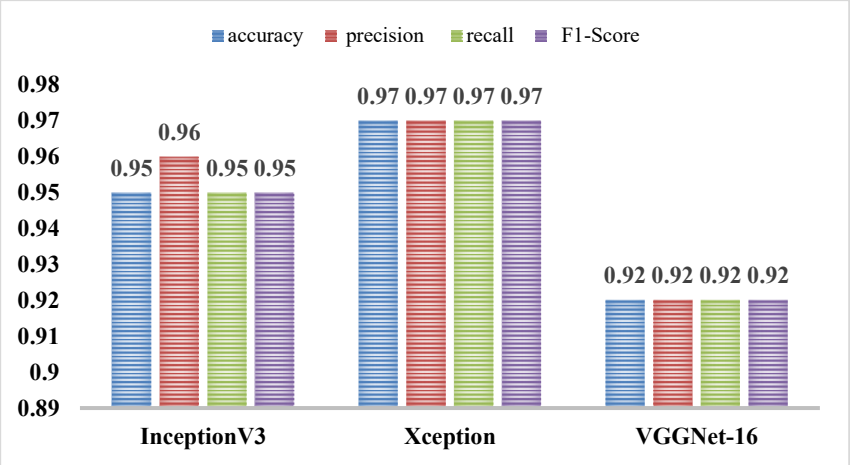

Figure 7. Architectural Comparison

InceptionV3 also performed well, achieving 95% accuracy, with a precision, recall, and F1-score of 0.96, 0.95, and 0.95, respectively. Although slightly lower than those of Xception, these results indicate that InceptionV3 remains highly effective at capturing multiscale features of road damage patterns. In contrast, VGGNet-16 recorded the lowest performance among the three architectures, with an accuracy of 92% and precision, recall, and F1-score values all at 0.92. Nevertheless, these results still reflect a satisfactory level of classification capability, particularly considering the limited dataset size.

Overall, the analysis confirms that all three CNN architectures achieve high classification performance; however, Xception demonstrates the most stable and accurate performance across the validation and test stages. This advantage is primarily attributed to its use of depthwise separable convolution, which enables efficient feature extraction while reducing model complexity. Such characteristics are particularly beneficial for handling variations in pavement texture and damage patterns under limited data conditions. Accordingly, Xception is identified as the most suitable architecture for the image-based road damage classification framework proposed in this study.

Despite these promising results, several limitations should be acknowledged. The dataset size remains relatively small, which may constrain the generalizability of the findings. Additionally, variations in lighting conditions, camera viewpoints, and damage scales were not exhaustively examined. Future research should consider incorporating larger, more diverse datasets, evaluating robustness across varying environmental conditions, and extending the framework to integrated detection and classification systems to enhance real-world applicability.

4. CONCLUSION

This study evaluated the effectiveness of several convolutional neural network (CNN) architectures, VGGNet-16, InceptionV3, and Xception, for classifying road damage types under the specific road and environmental conditions of Sleman Regency. Using a controlled experimental framework that integrates consistent preprocessing, explicit hyperparameter tuning, and four-fold cross-validation, the study aimed to

identify an architecture that delivers robust performance under limited, heterogeneous data conditions. The experimental results demonstrate that all evaluated CNN architectures achieve high classification performance. Among them, Xception consistently delivered the most stable and accurate results across validation and testing stages, achieving the highest accuracy, precision, recall, and F1-score. This advantage is largely attributed to the use of depthwise separable convolution, which enables efficient feature extraction while reducing model complexity—an important property when working with relatively small datasets. The ablation study further confirms the importance of preprocessing choices. The inclusion of CLAHE in the preprocessing pipeline resulted in consistent improvements across all evaluation metrics, with particularly notable gains in recall and F1-score. These improvements indicate that CLAHE enhances the model's ability to detect road damage patterns under uneven illumination, reducing false negatives and improving class-balanced recognition. This finding highlights that performance gains are not solely determined by architectural design but are also strongly influenced by systematic preprocessing strategies. Despite these promising results, several limitations must be acknowledged. The dataset is relatively small and includes only three damage categories, which may limit the generalizability of the findings to other regions or damage types. In addition, the effects of extreme lighting variations, camera viewpoints, and damage scale were not exhaustively explored. Furthermore, this study focuses on image-level classification and does not address spatial localization or real-time inference, which are key strengths of object detection–based approaches. Future research should expand the dataset across multiple regions and damage categories, evaluate robustness under diverse environmental conditions, and explore hybrid frameworks that integrate classification and detection paradigms. Nevertheless, within a regional context, this study's findings demonstrate that CNN-based classification models, particularly Xception, can provide reliable and practical support for road damage identification when combined with appropriate preprocessing and validation strategies. These results offer empirical guidance for selecting deep learning architectures in region-oriented road maintenance decision-support systems, especially in scenarios where data availability is limited.

Author Contributions

Anggun Puspita Sari: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - Original Draft. Kariyam: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing - Review and Editing.. Feri Wijayanto: Software, Validation, Writing - Review and Editing. Edy Widodo: Data Curation, Validation, Writing - Review and Editing. All authors discussed the results and contributed to the final manuscript.

Funding Statement

The author would like to express gratitude to the Ministry of Higher Education, Science and Technology of the Republic of Indonesia for the financial support that has been provided, as stated in Decree Number 126/C3/DT.05.00/PL/2025; 0498.01/LL5-INT/AL.04/2025; and the research contract letter in the framework of the Implementation of the Fundamental Research Program - Regular Research Scheme Number: 029/ST-DirDPPM/70/DPPM/PFR-KEMDIKTISAINTEK/VI/2025.

Acknowledgment

The author would like to express gratitude to the Directorate of Research and Community Service of the Universitas Islam Indonesia for providing assistance throughout the application process, from submission to receipt of the research grant.

Declarations

The authors declare that they have no conflicts of interest to report for this study.

Declaration of Generative AI and AI-assisted technologies

Generative AI tools (e.g., ChatGPT) were used solely for language refinement (grammar, spelling, and clarity). The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. The authors reviewed and approved all final text.

REFERENCE

[1] K. Ceni and E. Ina, "IDENTIFIKASI JENIS DAN PENANGANAN KERUSAKAN JALAN (STUDI KASUS JL. G. OBOS XII, JL. SAMUDIN AMAN, JL. JATI KOTA PALANGKA RAYA)," *Narotama Jurnal Teknik Sipil*, vol. 5, no. 2, pp. 28–36, Nov. 2021, doi: https://doi.org/10.31090/njts.v5i2.1567

[2] M. S. Karim, A. T. Handayani, and H. P. Astutik, "KINERJA RUAS JALAN SAAT KONDISI NEW NORMAL (STUDI KASUS JALAN LAKSDA ADISUTJIPTO, YOGYAKARTA KM 6,3-6,8)," *EQUILIB*, vol. 2, no. 1, pp. 13–20, 2021.

[3] S. Elmaningtyas and S. Andayani, "FUZZY TOPSIS APPLICATION TO DETERMINE PRIORITIES OF ROAD MAINTENANCE IN SLEMAN REGENCY," *Jurnal Kajian dan Terapan Matematika*, vol. 8, no. 2, pp. 138–148, Jul. 2022, [Online]. Available: http://journal.student.uny.ac.id/ojs/index.php/jktm

[4] A. Kusnadi and Ranny, "KERUSAKAN JALAN FLEXIBLE PAVEMENT DENGAN MENGGUNAKAN ALGORITMA PCA," *ULTIMATICS*, vol. 8, no. 2, pp. 125–130, Dec. 2016, doi: https://doi.org/10.31937/ti.v8i2.521

[5] A. M. Wira, I. Ruslianto, and D. M. Midyanti, "KLASIFIKASI KERUSAKAN JALAN PADA CITRA JALAN RAYA PONTIANAK DAN SEKITARNYA MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK," *Coding: Jurnal Komputer dan Aplikasi*, vol. 11, no. 1, 2023, doi: https://doi.org/10.26418/coding.v11i1.57905

[6] R. Octavia, "IMPLEMENTASI CONVOLUTIONAL NEURAL NETWORK UNTUK KLASIFIKASI KERUSAKAN JALAN BERBASIS CITRA DIGITAL (STUDI KASUS: KERUSAKAN JALAN ASPAL DI BANYUMAS)," 2024, *Yogyakarta, Indonesia*.

[7] D. Arya *et al.*, "DEEP LEARNING-BASED ROAD DAMAGE DETECTION AND CLASSIFICATION FOR MULTIPLE COUNTRIES," *Autom. Constr.*, vol. 132, p. 103935, Dec. 2021, doi: https://doi.org/10.1016/j.autcon.2021.103935

[8] L. Manoni, S. Orcioni, and M. Conti, "RECENT ADVANCEMENTS IN DEEP LEARNING TECHNIQUES FOR ROAD CONDITION MONITORING: A COMPREHENSIVE REVIEW," *IEEE Access*, vol. 12, pp. 154271–154293, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3481649

[9] B. Omarov and B. Kulambayev, "DEEP LEARNING-BASED IMAGE PROCESSING FOR REAL-TIME DETECTION OF ROAD SURFACE DAMAGE," *Procedia Comput. Sci.*, vol. 251, pp. 609–614, 2024, doi: https://doi.org/10.1016/j.procs.2024.11.157

[10] L. Alzubaidi *et al.*, "REVIEW OF DEEP LEARNING: CONCEPTS, CNN ARCHITECTURES, CHALLENGES, APPLICATIONS, FUTURE DIRECTIONS," *J. Big Data*, vol. 8, no. 1, 2021. https://doi.org/10.1186/s40537-021-00444-8

[11] Y. Jiang, "ROAD DAMAGE DETECTION AND CLASSIFICATION USING DEEP NEURAL NETWORKS," *Discover Applied Sciences*, vol. 6, no. 8, p. 421, 2024, doi: https://doi.org/10.1007/s42452-024-06129-0

[12] S. Umar, Maimunah, and H. A. Meidar, "KLASIFIKASI JALAN RUSAK MENGGUNAKAN TRANSFER LEARNING ARSITEKTUR VGG16," *Journal of Information Systems and Informatics Engineering*, vol. 8, no. 1, pp. 75–85, Jun. 2024.

[13] T. J. W. Adi, P. Suprobo, and Y. E. P. R. Waliulu, "IRODD (INTELLIGENT-ROAD DAMAGE DETECTION) FOR REAL-TIME INFRASTRUCTURE PRESERVATION IN DETECTION, CLASSIFICATION, CALCULATION, AND VISUALIZATION," *Journal of Infrastructure Policy and Development*, vol. 8, no. 11, Oct. 2024, doi: https://doi.org/10.24294/jipd.v8i11.6162

[14] S. Swain and A. K. Tripathy, "AUTOMATIC DETECTION OF POTHOLES USING VGG-16 PRE-TRAINED NETWORK AND CONVOLUTIONAL NEURAL NETWORK," *Heliyon*, vol. 10, no. 10, p. e30957, 2024, doi: https://doi.org/10.1016/j.heliyon.2024.e30957

[15] X. Ren, S. Huang, Y. Hu, K. Ye, Z. Chen, and Z. Wang, "A LIGHTWEIGHT CONVOLUTIONAL NEURAL NETWORK FOR DETECTING ROAD CRACKS," *Signal Image Video Process.*, vol. 18, no. 10, pp. 6729–6743, 2024, doi: https://doi.org/10.1007/s11760-024-03347-2

[16] S. Zhang, Z. Liu, K. Wang, W. Huang, and P. Li, "OBC-YOLOV8: AN IMPROVED ROAD DAMAGE DETECTION MODEL BASED ON YOLOV8," *PeerJ Comput. Sci.*, vol. 11, p. e2593, 2025, doi: https://doi.org/10.7717/peerj-cs.2593

[17] Y. Li, C. Yin, Y. Lei, J. Zhang, and Y. Yan, "RDD-YOLO: ROAD DAMAGE DETECTION ALGORITHM BASED ON IMPROVED YOU ONLY LOOK ONCE VERSION 8," *Applied Sciences*, vol. 14, no. 8, p. 3360, 2024, doi: https://doi.org/10.3390/app14083360

[18] Z. Zhang *et al.*, "ARDS-YOLO: INTELLIGENT DETECTION OF ASPHALT ROAD DAMAGES AND EVALUATION OF PAVEMENT CONDITION IN COMPLEX SCENARIOS," *Measurement*, vol. 242, p. 115946, 2025, doi: https://doi.org/10.1016/j.measurement.2024.115946

[19] S. Youwai, A. Chaiyaphat, and P. Chaipetch, "YOLO9TR: A LIGHTWEIGHT MODEL FOR PAVEMENT DAMAGE DETECTION UTILIZING A GENERALIZED EFFICIENT LAYER AGGREGATION NETWORK AND ATTENTION MECHANISM," *J. Real. Time. Image Process.*, vol. 21, no. 5, p. 163, 2024. https://doi.org/10.1007/s11554-024-01545-2

[20] M. W. Khan, M. S. Obaidat, K. Mahmood, B. Sadoun, H. M. S. Badar, and W. Gao, "REAL-TIME ROAD DAMAGE DETECTION USING AN OPTIMIZED YOLOV9S-FUSION IN IOT INFRASTRUCTURE," *IEEE Internet Things J.*, vol. 12, no. 11, pp. 17649–17660, 2025, doi: https://doi.org/10.1109/JIOT.2025.3537640

[21] Z. Demirel, S. T. Nasraldeen, Ö. Pehlivan, S. Shoman, M. Albdairi, and A. Almusawi, "COMPARATIVE EVALUATION OF YOLO AND GEMINI AI MODELS FOR ROAD DAMAGE DETECTION AND MAPPING," *Future Transportation*, vol. 5, no. 3, p. 91, 2025, doi: https://doi.org/10.3390/futuretransp5030091

[22] P. I. Wayan and others, "KLASIFIKASI CITRA MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN) PADA CALTECH 101," *Teknik ITS*, vol. 5, 2016, doi: https://doi.org/10.12962/j23373539.v5i1.15696

[23] J. P. Tanjung, "CLASSIFICATION OF WHEAT SEEDS USING NEURAL NETWORK BACKPROPAGATION ALGORITHM," *Journal of Informatics and Telecommunication Engineering*, vol. 4, no. 2, pp. 335–342, Jan. 2021, doi: https://doi.org/10.31289/jite.v4i2.4449

[24] N. Shakhovska, V. Yakovyna, M. Mysak, S. A. Mitoulis, S. Argyroudis, and Y. Syerov, "REAL-TIME MONITORING OF ROAD NETWORKS FOR PAVEMENT DAMAGE DETECTION BASED ON PREPROCESSING AND NEURAL NETWORKS," *Big Data and Cognitive Computing*, vol. 8, no. 10, p. 136, 2024, doi: https://doi.org/10.3390/bdcc8100136

[25] C. S. Wijaya, I. Wijaya, M. T. Shidqi, and D. Novita, "ANALISIS IMPLEMENTASI ARTIFICIAL INTELLIGENCE UNTUK BISNIS: SYSTEMATIC LITERATURE REVIEW," *Computer Science and Information Technology*, vol. 4, no. 2, p. 133, 2023, doi: https://doi.org/10.46576/device.v4i2.4037