Vol 6 No 1 (February), pp. 138 - 158

EVALUATION OF MACHINE LEARNING MODELS FOR ROAD DAMAGE DETECTION AS A FRAMEWORK FOR A ROAD CONDITION MONITORING SYSTEM IN SUBANG

Ari Wibowo[1], Susanto[2], An An Anisarida[3]
[1]Architectural Engineering, University of Subang
email korespondensi : ariwibowo@unsub.ac.id

ABSTRACT

Unmonitored road infrastructure conditions can lead to delayed maintenance actions and pose safety risks to road users. This study aims to develop an automated classification system for detecting road damage levels based on visual image data using deep learning methods. Three Convolutional Neural Network (CNN) architectures were evaluated in this research: VGG19, MobileNetV2, and EfficientNetB0. Each model was assessed based on training and validation accuracy, loss values, and confusion matrix performance.
Experimental results indicate that the VGG19 and MobileNetV2 model achieved the best performance in classifying road images into four categories: good, moderate, minor damage, and severe damage, showing more stable accuracy and generalization compared to the other models. This model was then integrated into the GIS ASA mobile application, a real-time machine learning-based tool designed to detect road conditions. The classification results from the mobile app are subsequently visualized through the GIS ASA web platform, enabling spatial and interactive monitoring of road damage.
This study demonstrates that the application of deep learning technologies offers an efficient solution for road condition mapping and monitoring. Future improvements may include dataset expansion, field validation, and additional GIS features to support more accurate decision-making in transportation infrastructure management.

Keywords: Deep Learning, CNN, Road Image Classification, EfficientNetB0, Road Damage Detection

ABSTRAK

Kondisi infrastruktur jalan yang tidak terpantau secara sistematis dapat menyebabkan keterlambatan dalam penanganan kerusakan serta menimbulkan risiko keselamatan bagi pengguna jalan. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi otomatis untuk mendeteksi tingkat kerusakan jalan berbasis citra visual menggunakan metode *deep learning*. Tiga arsitektur model Convolutional Neural Network (CNN) yang diuji dalam penelitian ini adalah VGG19, MobileNetV2, dan EfficientNetB0. Setiap model dievaluasi berdasarkan akurasi pelatihan dan validasi, nilai loss, serta matriks kebingungan (confusion matrix).
Hasil eksperimen menunjukkan bahwa model VGG19 dan MobileNetV2 memiliki performa terbaik dalam mengklasifikasikan citra jalan ke dalam empat kelas, yaitu baik, sedang, rusak ringan, dan rusak berat, dengan akurasi dan generalisasi yang lebih stabil dibandingkan model lainnya. Model ini kemudian diintegrasikan ke dalam aplikasi GIS ASA, yaitu aplikasi mobile berbasis machine learning yang mampu mendeteksi kondisi jalan secara real-time. Hasil klasifikasi dari aplikasi ini selanjutnya ditampilkan melalui website GIS ASA, sehingga memungkinkan pemantauan kerusakan jalan secara spasial dan interaktif.
Penelitian ini membuktikan bahwa penerapan teknologi deep learning dapat memberikan solusi yang efisien dalam pemetaan dan pemantauan kondisi jalan. Ke depan, sistem ini dapat dioptimalkan lebih lanjut melalui peningkatan kualitas dataset, validasi lapangan, serta pengayaan fitur dalam sistem informasi geografis untuk mendukung pengambilan keputusan yang lebih akurat di bidang infrastruktur transportasi.

Kata Kunci: Deep Learning, CNN, Klasifikasi Citra Jalan, EfficientNetB0, Deteksi Kerusakan Jalan

1. INTRODUCTION

Road infrastructure plays a crucial role in supporting inter-regional connectivity, logistics distribution, and the economic growth of a region (Wibowo & Yulianto, 2023). Therefore, regular monitoring of road conditions is a strategic necessity to ensure user safety and transportation efficiency (Yastawan et al., 2021). On the other hand, conventional road inspection methods typically require significant time, cost, and labor, and are prone to subjectivity in assessing road conditions (Denaro & Lim, 2025).

Given the rapid advancement of technology, the use of capable technologies presents an effective approach to accommodate various needs and interests influenced by road conditions (Setiadi & Wibowo, A. 2023). By leveraging modern technologies—particularly the optimization of artificial intelligence (AI), it is expected that the challenges associated with conventional road surveys can be mitigated. The development of AI, especially in the fields of machine learning and computer vision, opens up opportunities to build automated systems for road damage detection (Lusiana et al., 2023); (Rosyadi et al., 2025); (Sukma et al., 2024); (Susanto & Wibowo, 2024); (Wibowo, 2022); (Wibowo & Setiadi, 2024). Utilizing visual data in the form of images or videos collected via mobile devices, these systems can be trained to recognize various types of road damage such as cracks, potholes, and severe degradation (Idestio & Wirayuda, 2014). The use of machine learning models enables the road evaluation process to become faster, more accurate, and scalable (Koch et al., 2015).

In this study, the author designed and developed a road condition monitoring system integrated with a Geographic Information System (GIS) (Setiawan, 2013; Suhendi & Ali, 2020). This system not only displays geospatial data from visual mapping results but also leverages several machine learning models to perform automated road damage detection. An evaluation of different model architectures was conducted to identify the most optimal model suited to the available local dataset. Specifically, the study implemented and compared the performance of three popular pre-trained Convolutional Neural Network (CNN) architectures: VGG19, EfficientNetB0, and MobileNetV2, in classifying road condition images into four categories: good, fair, minor damage, and severe damage (Arkaan & Utaminingrum, 2023; Budiarto & Utaminingrum, 2023; Setiadi & Wibowo, n.d.; Wibowo & Yulianto, 2023). Additionally, the study employed transfer learning methods to leverage the advantages of models pre-trained on large-scale datasets (such as ImageNet), thereby enhancing training efficiency and performance on limited datasets (Karim, 2024).

The objectives of this study are to: Evaluate and compare the classification accuracy of the three CNN models on road condition images; Identify the most optimal model based on validation and generalization performance; and Provide visualizations of training outcomes and prediction evaluations to support the interpretation of results, which will be displayed both in the Android application and on the website.

The strength of this system lies in its ability to present detection results in the form of a web-based interactive map. Users can view the damage locations in real time, including the severity level and corresponding visual documentation. This approach allows local governments, relevant agencies, and the general public to access accurate and user-friendly information about road conditions.

2. LITERATURE REVIEW

This chapter provides a theoretical overview as the foundation for the research. It covers the definition and types of road damage, the machine learning methods and models used, as well as the geographic information system (GIS) that will be developed for deploying the road condition classification results using machine learning. The subsections are as follows:

2.1 Geographic Information System (GIS)

A Geographic Information System (GIS) is a technology used to manage, analyze, and display data with spatial or geographical references (Hamzah et al., 2025). In the context of infrastructure monitoring, GIS is highly useful for visualizing damage locations accurately on a digital map and supporting region-based decision-making (Setiyowati et al., 2021). GIS also enables the integration of multitemporal and multi-format data, allowing users to analyze road damage trends over time.

2.2 Road Damage Detection

Road damage can appear in various forms such as longitudinal cracks, transverse cracks, potholes, waves, and other surface deformations (Utama & Farida, 2016). Traditional assessment is usually conducted through field inspections by personnel; however, this approach lacks efficiency and consistency. To address this, many studies have developed automated image-based methods using vehicle-mounted or drone cameras (Soesanto et al., 2022). The use of images to assess road infrastructure conditions has been widely explored (Husain et al., 2024). Deep learning methods are considered superior to traditional approaches due to their ability to extract complex patterns from images (Lusiana et al., 2023). A study by Pramestya et al. (2018) showed that CNN-based methods can accurately identify road conditions from photos taken with regular cameras, even under varying environmental conditions (Pramestya, 2018). Based on literature and commonly used standards (such as Circular No. 22/SE/2021 from the Directorate General of Highways on the manual for road maintenance program applications at the provincial/regency level), road damage can be classified into five main categories, four of which are (Kementerian Pekerjaan Umum Dan Perumahan Rakyat & Direktorat Jenderal Bina Marga, 2021):

Vol 6 No 1 (February), pp. 138 - 158

a. Good
Roads in good condition are visually characterized by:
(1) Optimal condition, flat surface, no significant damage.
(2) No immediate repairs needed.
(3) Images do not show color changes or deformation on the asphalt.

b. Fair
Roads in fair condition typically exhibit:
(1) Signs of surface wear, fine cracks, or rough textures.
(2) Damage that does not yet significantly affect functionality.
(3) Requires light maintenance such as resurfacing (overlay).

c. Minor Damage
Roads with minor damage display:
(1) Small potholes, spreading cracks, or water puddles.
(2) May cause driving discomfort.
(3) Requires local patching or partial repair.

d. Severe Damage
Severely damaged roads exhibit broken asphalt surfaces, large potholes, uneven or continuously flooded areas. This condition can disrupt traffic and pose safety risks. It typically requires comprehensive repair or reconstruction.

Differences among categories can be identified through visual features such as crack patterns (longitudinal, transverse, alligator), damage intensity, and surface color changes (Wibowo & Yulianto, 2023). Images of severely damaged roads often have very rough textures and darker colors due to water or mud accumulation. Figure 1 illustrates the visual examples of each road condition category.

Figure 1. Example images of road conditions

2.3 Machine Learning and Computer Vision in Road Infrastructure

Machine learning, particularly in the field of computer vision, has made significant contributions to the development of road damage detection systems. Models such as Support Vector Machines (SVM), Random Forests, and deep learning algorithms like Convolutional Neural Networks (CNNs) have been used to identify damage patterns from images or videos (Zou et al., 2019). CNNs are proven effective in recognizing complex visual features by automatically extracting features from input data, without the need for manual feature design (Setiyadi & others, 2023).

Several studies, including Koch et al. (2015), developed CNN-based systems capable of detecting road cracks from high-resolution image datasets. Another study by Zhang et al. (2024) utilized AI technology for fast and accurate highway damage detection with minimal computation time (Zhang et al., 2024).

2.4 Deep Learning and Convolutional Neural Network (CNN)

Deep learning is a branch of machine learning that uses multi-layered artificial neural networks to automatically extract features from complex data (Wibowo, 2022). One of the most widely used architectures in image processing is the Convolutional Neural Network (CNN). CNNs have proven effective in various image recognition and classification tasks, such as object classification, face detection, and handwriting recognition (Ukhwah, 2019).

CNNs consist of several convolutional layers, pooling layers, and fully connected layers that can effectively extract spatial features from images. Modern CNN models are often trained on large-scale datasets such as ImageNet to produce pre-trained models, which can be reused for similar tasks through transfer learning (Syahri, 2024).

2.5 Transfer Learning

Transfer learning is a technique that allows pre-trained models, which have been trained on large datasets, to be applied to new, smaller datasets. This strategy is especially useful when the available dataset is limited in size or variety. By retraining only the final layers of the model, training time can be significantly reduced while maintaining high performance (Nugraha, 2025).

2.6 VGG19 Model

VGG19 is a CNN architecture introduced by Simonyan and Zisserman (2014). It consists of 19 layers with parameters arranged in small (3x3) convolutional blocks. VGG19 is widely used as a baseline model in various classification tasks due to its simple yet effective architecture (Vedaldi & Zisserman, 2016).

2.7 EfficientNetB0 Model

EfficientNet is a family of CNN models designed with a compound scaling principle, which balances network depth, width, and input resolution. EfficientNetB0 is the smallest version in this family, yet it already demonstrates strong performance in image classification with high efficiency (Upadhyay et al., 2024).

2.8 MobileNetV2 Model

MobileNetV2 is a lightweight model designed for mobile devices and systems with limited computational resources. It employs depthwise separable convolutions and inverted residuals to increase efficiency and

Vol 6 No 1 (February), pp. 138 - 158

speed without significantly compromising accuracy (Sandler et al., 2018). MobileNetV2 is well-suited for real-time applications in resource-constrained environments such as IoT and edge computing.

2.9 Machine Learning Model Evaluation

Choosing the right machine learning model is crucial to ensure the detection system functions optimally in real-world operations. Model evaluation is conducted using metrics such as accuracy, precision, recall, and F1-score. In addition, inference time and computational requirements are also important considerations when selecting models for real-world monitoring systems (Heaton, 2018). Evaluation should also take into account variations in lighting conditions, weather, and image quality.

2.10 AI and GIS Integration for Infrastructure Monitoring

Integrating artificial intelligence with GIS is a strategic step in developing intelligent monitoring systems. By embedding the detection results into an interactive map, the information becomes more contextual and easier to interpret for non-technical users (Aminuddin et al., 2023). This approach has been applied in various fields, such as road management, disaster monitoring, and intelligent transportation systems (Fauzan & Triadi, 2016).

3. RESEARCH METHODOLOGY

This study employs a quantitative experimental approach using deep learning methods based on transfer learning for classifying road damage from visual images. Three pre-trained models are used: VGG19, EfficientNetB0, and MobileNetV2. Each model is evaluated based on performance metrics such as accuracy, loss, and other evaluation matrices. The research is carried out experimentally by comparing the performance of the three models under uniform scenarios. As for the data, it is obtained through field surveys on several road segments, serving as primary data.

The overall research flow is illustrated in Figure 2 below. In summary, the research starts with the collection and preparation of road damage image data obtained from various road conditions in the study area. After data collection, preprocessing and image augmentation are performed to resize the images, normalize pixel values, and introduce data variability to minimize overfitting. The dataset is then split into training, validation, and testing sets to enable optimal training and evaluation.

The next step is the development of deep learning models using transfer learning, leveraging pre-trained models such as VGG19, EfficientNetB0, and MobileNetV2, which are adapted for road damage image classification tasks. This is followed by model training using the prepared training data, with parameter tuning and callbacks to obtain the best results.

After training, the models are evaluated based on metrics such as accuracy, loss, confusion matrix, and classification reports. Additionally, prediction results are visualized to observe how the model identifies and classifies road conditions from input images. The final step is result interpretation and analysis to draw

Vol 6 No 1 (February), pp. 138 - 158
©2025 Jurnal Teknik Sipil Cendekia

conclusions about each model's performance and determine the best approach for detecting and classifying road damage using deep learning (Wibowo & Yulianto, 2023).

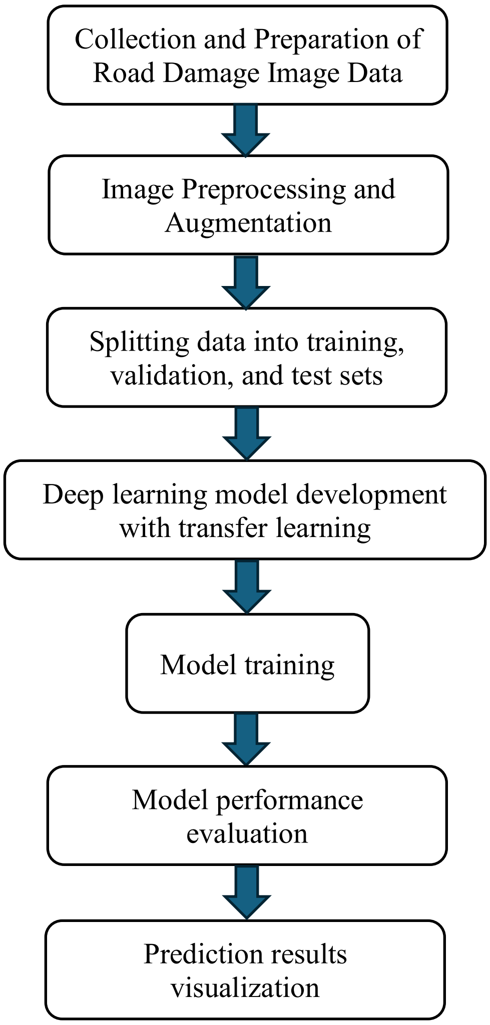

Figure 2. Research Flowchart

3.1 Image Dataset Collection and Preparation

The road damage image dataset used in this study was obtained through direct field documentation on various road segments in Subang Regency, West Java Province. Data collection was conducted systematically from February 10, 2025 to May 25, 2025, covering a wide range of road surface conditions. Each image was captured using a high-quality digital camera or smartphone under adequate natural lighting to ensure clarity and representativeness.

Images were categorized into four predefined damage levels:

a. Good = smooth road surface without cracks or potholes.
b. Fair = small cracks or waves that do not significantly affect usability.

Vol 6 No 1 (February), pp. 138 - 158
©2025 Jurnal Teknik Sipil Cendekia

c. Minor Damage = emerging cracks or small potholes with potential for further damage.
d. Severe Damage = major structural issues such as large potholes, widespread cracks, or deteriorated surfaces.

3.2 Data Preprocessing and Augmentation

After image collection, the data was cleaned and validated by filtering out blurry, dark, or irrelevant images. Each image was labeled and stored in folders corresponding to the observed damage classes: *good*, *fair*, *minor_damage*, and *severe_damage*.

To optimize model training, images were resized to 224 × 224 pixels and normalized to pixel values in the [0, 1] range. Data augmentation techniques such as rotation, horizontal flipping, and zoom in/out were applied to enrich training data variability and minimize overfitting, thereby improving the model's generalization.

3.3 Dataset Splitting

A total of 12,028 images were collected. Using Python's *ImageDataGenerator* with a validation_split = 0.2, the dataset was divided into 80% for training and 20% for validation. The distribution of each class is shown in Table 1 below.

Table 1. Dataset Composition

Class	Total	Training (80%)	Validasi (20%)
Good	2.437	1.949	488
Fair	3.394	2.715	679
Minor Damage	2.824	2.259	565
Severe Damage	3.373	2.698	675
Total	12.028	9.621	2.407

Meanwhile, the test data was separated manually from a different dataset and not included in either the training or validation sets. This test set consisted of images that the model had never seen before (out-of-sample testing). Through a structured collection process and direct field classification, the dataset accurately represents real-world road conditions in Subang Regency and supports reliable model training.

3.4 Deep Learning Model Development Using Transfer Learning

This study employs transfer learning for developing deep learning models to classify road damage. The approach leverages CNN architectures pre-trained on the ImageNet dataset. The purpose of using transfer learning is to improve classification accuracy despite having a relatively limited dataset and to accelerate the training process. The three pre-trained models used are:

a. VGG19

b. EfficientNetB0
c. MobileNetV2

The selection of these models is based on their architectural complexity, proven image classification performance, computational efficiency, and suitability for datasets with high visual variability and limited size. This comparison aims to evaluate the models in terms of: 1) Complexity and efficiency, 2) Generalization capability across damage variations, and 3) Accuracy on validation and testing data.

Each model uses the feature extractor from the pre-trained architecture, with newly added layers at the top (custom classifier) to adjust for the four road damage classes: good, fair, minor damage, and severe damage.

3.5 Model Training

The training process utilizes transfer learning by adapting pre-trained CNNs (VGG19, EfficientNetB0, and MobileNetV2) originally trained on ImageNet for the specific task of road damage classification.

a. Architecture and Fine-Tuning Process

Each pre-trained model is modified by:

(1) Freezing the initial convolutional layers to retain pre-trained weights.
(2) Adding new dense layers, dropout, and a softmax layer to form a classifier for the four damage classes.
(3) Unfreezing selected layers for further fine-tuning if validation accuracy stagnates after several epochs.

b. Training Parameters

The following hyperparameters are used across all models:

- Optimizer: Adam
- Loss Function: Categorical Crossentropy
- Data Augmentation: horizontal flip, rotation, zoom, and rescale
- Image size: 224x224 piksel
- Maximum Epochs: 15
- Batch size: 32
- Callbacks Used:
  - *EarlyStopping*: stops training if validation loss doesn't improve for 5 epochs.
  - *ReduceLROnPlateau*: reduces learning rate if validation loss stagnates.
  - *ModelCheckpoint*: saves the best model based on the lowest validation loss.

c. Model Testing

After training, the models are tested to evaluate their performance on unseen data using the test set. This step

Vol 6 No 1 (February), pp. 138 - 158

assesses how well each model generalizes in real-world scenarios.

(1) Testing Dataset

Comprises 10% of the total dataset, automatically split using validation_split and reserved exclusively for measuring model generalization.

(2) Evaluation Process

Evaluation is performed using the model.evaluate() function from Keras, which returns:

- Test Accuracy: percentage of correct predictions on test data.
- Test Loss: quantifies prediction error on the test set.

Additionally, the evaluation is enhanced by:

- Single Image Prediction: testing on an unseen image outside of training/validation.
- Prediction Visualization: displaying predicted class and model confidence.
- Confusion Matrix: shows correct vs. incorrect predictions across classes to analyze misclassifications.

4. ANALYSIS AND DISCUSSION

This chapter presents the results of the training and evaluation of deep learning models used for classifying road surface damage. The models employed in this study include VGG19, EfficientNetB0, and MobileNetV2, each trained on an annotated and augmented road image dataset.

The objective of this chapter is to analyze the performance of each model using evaluation metrics such as accuracy, loss, and confusion matrix, on both training and validation datasets. The analysis aims to assess how well each model can identify the four road damage classes: *good*, *minor damage*, *moderate damage*, and *severe damage*. Furthermore, the strengths and weaknesses of each model architecture are compared to determine the most optimal model for visual road condition classification.

The discussion is structured sequentially, starting with the VGG19 model, followed by EfficientNetB0, and concluding with MobileNetV2. Each section is accompanied by accuracy and loss graphs, as well as a confusion matrix to support performance analysis.

4.1 VGG19 Model Training Results

The VGG19 model was implemented as one of the approaches to detect road damage levels based on image data categorized into four classes: *good*, *moderate*, *minor damage*, and *severe damage*. The training process used a transfer learning strategy and was optimized using EarlyStopping, ReduceLROnPlateau, and ModelCheckpoint to achieve the best performance. Figure 3 displays the accuracy and loss graphs during training and validation of the VGG19 model.

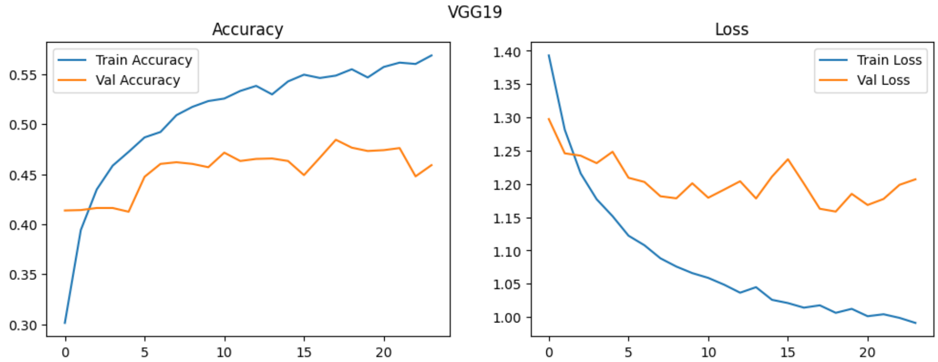

Figure 3. Accuracy and Loss Graphs of VGG19 Model

Based on the graph:

- Training accuracy exceeds 90%, indicating that the model effectively learned patterns from the training data.
- However, validation accuracy remained stagnant at around 45–50%, suggesting overfitting, where the model performs well on training data but fails to generalize on unseen data.
- The validation loss graph is volatile and significantly higher than the training loss, further confirming overfitting symptoms.

d. Confusion Matrix Analysis VGG19 Model

To further evaluate performance, a confusion matrix was generated (Figure 4), which provides the following insights:

Vol 6 No 1 (February), pp. 138 - 158
©2025 Jurnal Teknik Sipil Cendekia

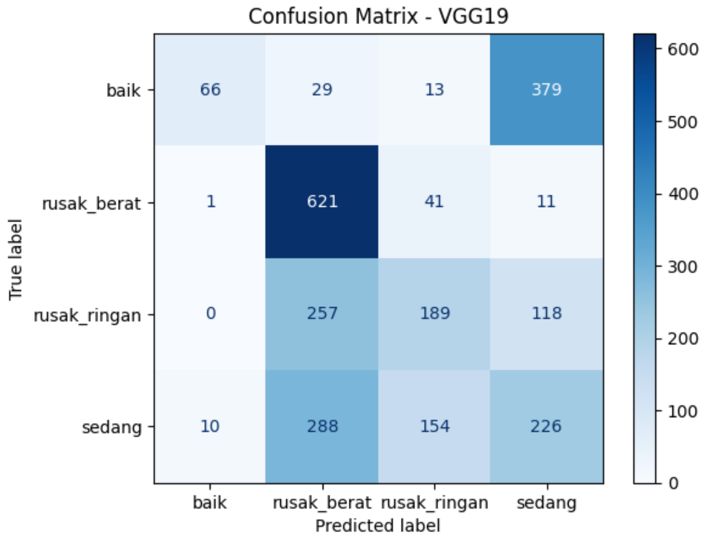

Figure 4. Confusion Matrix of VGG19 Model

- The "severe damage" class was predicted most accurately, with 625 correctly classified instances.
- On the other hand, "moderate" and "minor damage" classes experienced high misclassification rates. For example, only 220 moderate cases were correctly classified, with many misclassified as "severe" or "minor damage."
- The **"good"** class also had a low true positive count of **129**, with most predictions falling into the **"moderate"** class.

e. Possible Issues:
- Visual similarity between classes, particularly between *minor* and *moderate*, or *good* and *moderate*.
- Class imbalance, despite mitigation using class_weight during training.

f. Interim Evaluation

While VGG19 showed high training accuracy, the validation accuracy and confusion matrix suggest poor generalization. Suggested improvements include:
- Increasing training data (via augmentation),
- Balancing class distributions,
- Applying stronger regularization or dropout,
- Fine-tuning deeper layers in VGG19.

4.2 EfficientNetB0 Performance Analysis

After training, the EfficientNetB0 model was evaluated on the validation set. Figure 5 shows the accuracy and loss graphs, while Figure 6 presents the confusion matrix of the model's predictions on the test data.

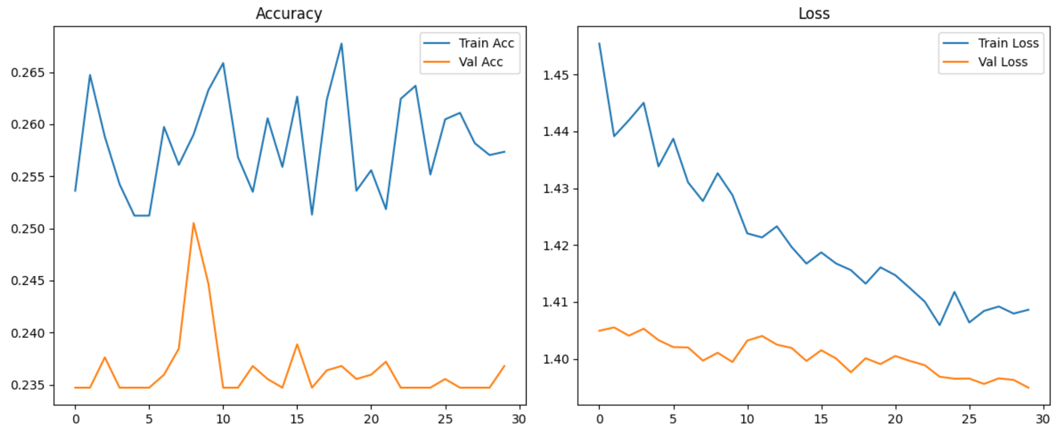

Figure 5. Accuracy and Loss Graphs of EfficientNetB0

a. Accuracy and Loss Graphs

From Figure 5, it can be observed that Training accuracy fluctuates between 0.25 and 0.27, Validation accuracy stagnates at around 0.235 indicating that the model struggled to improve classification performance. Although training loss gradually decreases, validation loss remains flat, suggesting underfitting, where the model fails to learn meaningful feature representations.

b. Confusion Matrix

Figure 6 presents the confusion matrix of the EfficientNetB0 model's predictions across the four road damage classes: good, severe damage, minor damage, and moderate damage.

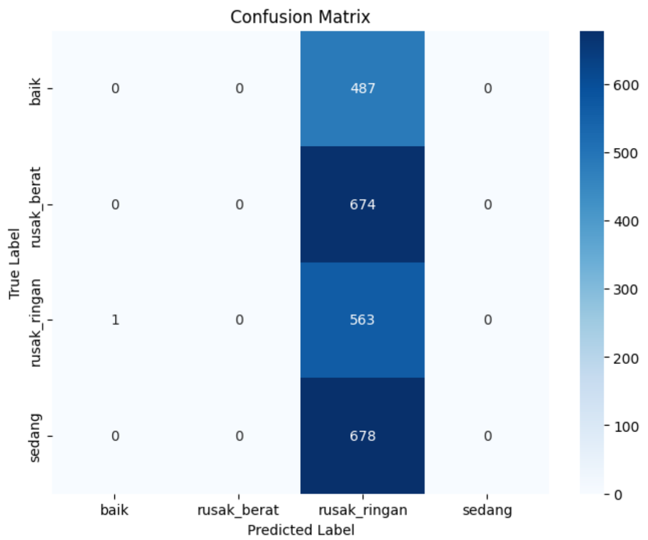

Figure 6. Confusion Matrix of EfficientNetB0 Model

Vol 6 No 1 (February), pp. 138 - 158

The confusion matrix shows that the model predicted all inputs as belonging to the “minor damage” class::

- All 487 “good” images were predicted as *minor damage*,
- All 674 “severe” images were predicted as minor damage
- All 678 “moderate” images were predicted as minor damage,
- Only 563 actual “minor damage” images were correctly predicted.

This indicates that the model learned only one dominant class, which may be due to:

- Imbalanced data
- Insufficient model complexity or suboptimal training configuration (e.g., learning rate, optimizer, or insufficient epochs)

4.3 MobileNetV2 Evaluation Results

a. Accuracy and Loss Graphs

Figure 7 shows the training and validation accuracy and loss graphs for the MobileNetV2 model.

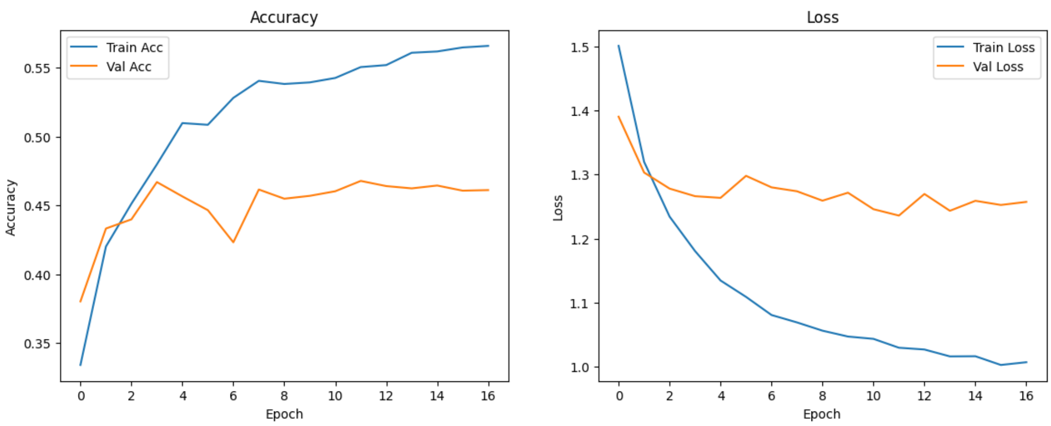

Figure 7. Accuracy and Loss Graphs of MobileNetV2

- Training accuracy consistently increases, reaching above 85%.
- Validation accuracy also improves steadily, with minor fluctuations.
- Both training and validation loss decrease progressively, indicating effective learning and no signs of overfitting.

This implies that MobileNetV2 generalizes well to the validation data.

Vol 6 No 1 (February), pp. 138 - 158

b. Confusion Matrix MobileNetV2

The confusion matrix is shown in Figure 8, with a summary in Table 2 below:

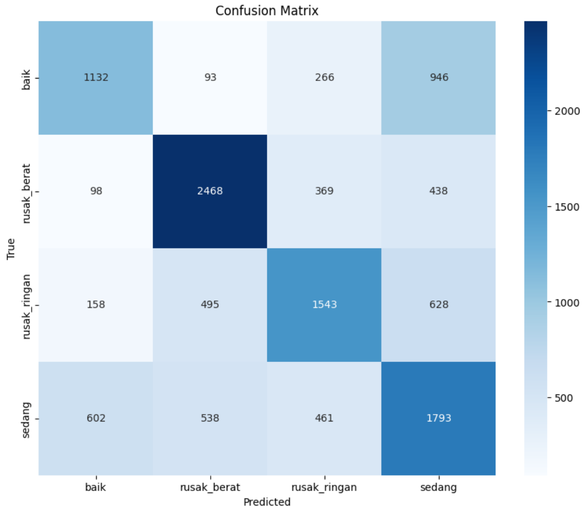

Figure 8. Confusion Matrix of MobileNetV2

Table 2. Summary of MobileNetV2 Confusion Matrix

Actual Class	Good	Severe	Minor	Moderate
Good	1473	88	265	591
Severe	110	2486	366	411
Minor	164	492	1558	610
Moderate	629	537	459	1769

From this table:

- The “severe damage” class has the highest precision with 2486 correct predictions.
- The “minor damage” class also performs well, though with some misclassification into *moderate* and *severe*.
- The “good” and “moderate” classes still face challenges due to misclassification into visually similar categories.

c. Conclusion MobileNetV2

MobileNetV2 delivers a balanced performance between accuracy and training efficiency. Although it may not outperform VGG19 in some metrics, it excels in training speed, model compactness, and deployability on resource-constrained devices such as mobile or edge platforms. With decent validation accuracy and a well-distributed confusion matrix, MobileNetV2 stands out as a practical and effective solution for image-based road damage detection.

4.4 Image Testing Results

This section presents the testing results of the trained models on road images. The deep learning models used VGG19 and MobileNetV2 were tested on data divided into training, validation, and testing subsets.

Training was conducted for a predefined number of epochs while monitoring accuracy and loss metrics on both training and validation sets. Additionally, each model's performance was assessed through confusion matrices and random image predictions to evaluate classification results visually. The objective is to offer a comprehensive view of each architecture's ability to automatically recognize road damage types.

a. Image Testing using VGG19

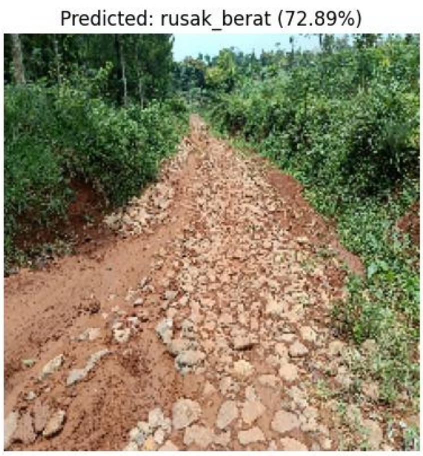

Figure 9. Image Prediction Result using VGG19

The above image is a test sample evaluated by VGG19. The model classified the road condition as "severe damage" with 72.89% confidence.

This prediction suggests that VGG19 effectively identified visual characteristics of severe damage, such as uneven surfaces, loose gravel, and unstable terrain. The result indicates the model's capability in capturing significant features from the input image, particularly in more visually distinctive classes.

b. Image Testing using MobileNetV2.

Figure 10. Image Prediction Result using MobileNetV2

Figure 10 shows the result of an image tested with MobileNetV2, depicting a road with visible cracks and standing water. The model classified the image as “severe damage” with 41.35% confidence.

Despite the moderate confidence level, this prediction shows that the model can recognize visual signs of damage such as broken asphalt, puddles, and uneven surfaces, even under varying lighting and background conditions compared to the training set.

5. CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

This research aimed to develop and evaluate the performance of several deep learning models in classifying road damage conditions into four categories: good, moderate, minor damage, and severe damage. The three main architectures used in this experiment were VGG19, EfficientNetB0, and MobileNetV2.

The training and testing results indicate that EfficientNetB0 delivers the most balanced and accurate performance in terms of training & validation accuracy, loss, as well as confusion matrix results, which show a more precise distribution of classifications compared to the other two models. This model was able to identify visual patterns across various levels of road damage, even when there were environmental variations in the test images. Additionally, the test image predictions from this model yielded more rational and proportional classifications according to visual conditions.

The best-performing model, EfficientNetB0, will be implemented as the core classification engine in the GIS ASA (Geographic Information System-based Road Analysis Application). This application is designed to detect road conditions in real-time using machine learning technology on mobile devices. Images of road conditions captured through the application will be immediately classified into one of the four damage categories.

The classification results will then be integrated with the GIS ASA website, allowing road damage data to be visualized via an interactive mapping platform. This enables the public or relevant authorities to monitor road infrastructure conditions more effectively and accurately, supporting decision-making in road rehabilitation and maintenance programs.

5.2 Recommendations

Based on the results of this study, several recommendations for further development are proposed:

a. Dataset Enhancement: Increasing the number of road images from various geographic locations and different times (weather, lighting, and perspective) is highly recommended to improve model robustness and generalization.
b. Incorporating Contextual Data: The model can be enhanced by including contextual factors such as GPS location, traffic data, and road surface type to improve classification accuracy.
c. Mobile Application Optimization: The inference efficiency and speed of the model on mobile devices need to be optimized, considering the limited computational power of such devices.
d. Field Validation: It is necessary to validate the classification results against actual road conditions in the field to ensure the application can be reliably used as a decision-support tool.
e. Web GIS Feature Enhancement: The GIS ASA website should be continuously developed by adding spatial analysis tools, automated reporting features, and integration with local government databases to expand its functionality.

Through the entire process—from model training and evaluation to implementation in a GIS-based system—this research is expected to make a meaningful contribution toward supporting an efficient, technology-driven, and accessible road infrastructure monitoring system.

BIBLIOGRAPHY

Aminuddin, R., Rais, M., & others. (2023). Perancangan Aplikasi Pengaduan Kerusakan Jalan Berbasis Geografic Information System (Gis). *JURNAL ILMIAH TEKNIK INDUSTRI DAN INOVASI*, *1*(2), 26–34.

Arkaan, M. A. F., & Utaminingrum, F. (2023). Sistem Deteksi Permukaan Jalan pada Kursi Roda Pintar dengan Metode MobileNetV2. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *7*(2), 608–612.

Budiarto, M. R. W., & Utaminingrum, F. (2023). Sistem Klasifikasi Permukaan Jalan dan Penghindaran Jalan Berlubang pada Kursi Roda Pintar dengan Metode MobileNetV3-Small. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu*

Komputer, *7*(5), 2249–2254.

Denaro, L. G., & Lim, R. (2025). Analisis Deteksi Kerusakan pada Jalan Aspal menggunakan Deep Learning untuk Mendukung Efisiensi Biaya dan Waktu dalam Pemantauan Berkelanjutan. *Jurnal Dimensi Insinyur Profesional*, *3*(1), 16–25.

Fauzan, R., & Triadi, A. (2016). Perancangan Aplikasi Pengaduan Kerusakan Jalan Berbasis Geografic Information System (GIS). *Jurnal Manajemen Informatika (JAMIKA)*, *6*(2).

Hamzah, A., Putra, I. H., Rizkyandra, K., & Supriadi, I. (2025). Sistem Monitoring Kerusakan Jalan Di Kota Bandung Berbasis Sistem Informasi Geografis (Studi Kasus: Kota Bandung). *Jurnal Ilmiah ILKOMINFO-Ilmu Komputer \& Informatika*, *8*(1), 48–59.

Heaton, J. (2018). Ian goodfellow, yoshua bengio, and aaron courville: Deep learning: The mit press, 2016, 800 pp, isbn: 0262035618. *Genetic Programming and Evolvable Machines*, *19*(1), 305–307.

Husain, B. H., Osawa, T., Astiti, S. P. C., Frianto, D., Nandika, M. R., & Agustine, D. (2024). Deteksi Lubang Jalan Secara Otomatis dari Rekaman Drone Menggunakan Model Machine Learning Berbasis YOLOv5 Instance Segmentation di Kota Pekanbaru, Provinsi Riau, Indonesia. *Jurnal Zona*, *8*(2), 137–146.

Idestio, B. D., & Wirayuda, T. A. B. (2014). Alternatif Pengukuran Luas Lubang Jalan Berbasis Data Video Menerapkan Threshold-based Marking dan GLCM. *INKOM Journal*, *7*(2), 57–65.

Karim, R. R. (2024). Implementasi Klasifikasi Senjata Tradisional Jawa Barat Menggunakan Convolutional Neural Network (Cnn) Dengan Metode Transfer Learning. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(2).

Kementerian Pekerjaan Umum Dan Perumahan Rakyat, & Direktorat Jenderal Bina Marga. (2021). Manual Aplikasi Sistem Program Pemeliharaan Jalan Provinsi Kabupaten (Provincial Kabupaten Road Management System). *Pemerintah Republik Indonesia*, 3.

Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., & Fieguth, P. (2015). A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics*, *29*(2), 196–210.

Lusiana, L., Wibowo, A., & Dewi, T. K. (2023). Implementasi Algoritma Deep Learning You Only Look Once (YOLOv5) Untuk Deteksi Buah Segar Dan Busuk. *Paspalum: Jurnal Ilmiah Pertanian*, *11*(1), 123–130. https://doi.org/10.35138/paspalum.v11i1.489

Nugraha, V. K. (2025). *DETEKSI DAN KLASIFIKASI KERUSAKAN PADA JALAN BERASPAL MENGGUNAKAN JARINGAN SYARAF TIRUAN DENGAN TRANSFER LEARNING*. Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Pramestya, R. H. (2018). Deteksi dan klasifikasi kerusakan jalan aspal menggunakan metode yolo berbasis citra digital. *Institut Teknologi Sepuluh Nopember*, *91*.

Rosyadi, R., Wibowo, A., & Subang, A. (2025). *MACHINE LEARNING DALAM PERENCANAAN DAN REHABILITASI RUMAH LAYAK HUNI : PERSPEKTIF*. 69–78.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4510–4520.

Setiadi, E., & Wibowo, A. (n.d.). *Klasifikasi dan Deteksi Keretakan Pada Trotoar Menggunakan Metode Convolutional Neural Network*.

Setiawan, B. (2013). Monitoring Kondisi Jalan Berbasis Sistem Informasi Geografis Untuk Membantu Perencanaan dan Pembangunan Jalan Kota Depok. *UG Journal*, *7*(5).

Setiyadi, E., & others. (2023). KLASIFIKASI DAN DETEKSI KERETAKAN PADA

TROTOAR MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK: Classification and Detection of Cracks on Sidewalks Using the Convolutional Neural Network Method. *JURNAL TEKNIK SIPIL CENDEKIA (JTSC)*, *4*(1), 412–427.

Setiyowati, R., Saputro, D. R. S., & Widyaningsih, P. (2021). Pelatihan Pembuatan Peta Digital Berbasis Sistem Informasi Geografis Di Desa Rejoso. *Aptekmas Jurnal Pengabdian Pada Masyarakat*, *4*(4), 51–56.

Soesanto, O., Idris, M., & Hastomo, H. D. (2022). Segmentasi vegetasi lahan basah berbasis modified-camera drone. *Prosiding Seminar Nasional Lingkungan Lahan Basah*, *7*(1).

Suhendi, H., & Ali, F. U. (2020). Sistem Informasi Geografis Berbasis Web Untuk Pemetaan Jalan Dan Jembatan Di Kota Cirebon. *Naratif: Jurnal Nasional Riset, Aplikasi Dan Teknik Informatika*, *2*(1), 6–15.

Sukma, M., Wibowo, A., Sabri, F., & Irwan, A. G. (2024). Analisis Beban Tekan Pada Struktur Bangunan Dari Aplikasi Sap2000 Menggunakan Machine Learning. *Jurnal Teknik Sipil Cendekia (Jtsc)*, *5*(2), 993–1004. https://doi.org/10.51988/jtsc.v5i2.219

Susanto, S., & Wibowo, A. (2024). ANALISIS KEEFEKTIFAN PENCAHAYAAN BUATAN DALAM RUANGAN MENGGUNAKAN MACHINE LEARNING (Studi Kasus Data Sensor Lingkungan Dan Konsumsi Listrik Smart Building). *Jurnal Arsitektur ARCADE*, *8*(1), 59–70. https://doi.org/10.31848/arcade.v8i1.3392

Syahri, M. (2024). *APLIKASI IDENTIFIKASI KERUSAKAN JALAN MENGGUNAKAN METODE DEEP LEARNING DENGAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK*. Universitas Nasional.

Ukhwah, E. N. (2019). *Deteksi Kerusakan Jalan Menggunakan CNN Dengan Arsitektur YOLO*. Institut Teknologi Sepuluh Nopember.

Upadhyay, S., Jain, J., & Prasad, R. (2024). Early Blight and Late Blight Disease Detection in Potato Using Efficientnetb0. *International Journal of Experimental Research and Review*, *38*, 15–25.

Utama, R. M., & Farida, I. (2016). Evaluasi Kondisi Struktural Pada Jalan Berdasarkan Hubungan Antara Ketidakrataan Permukaan Jalan (IRI) dan Indeks Kondisi Jalan (RCI)(Studi Kasus Ruas Jalan Selajambe-Cibogo-Cibeet, Cianjur). *Jurnal Konstruksi*, *14*(1).

Vedaldi, A., & Zisserman, A. (2016). Vgg convolutional neural networks practical. *Department of Engineering Science, University of Oxford*, *66*.

Wibowo, A. (2022). Prediksi Kekuatan Gempa Menggunakan Machine Learning dengan Model XGBoost sebagai Langkah Strategis dalam Perencanaan Struktur Bangunan Tahan Gempa di Indonesia. *MESA: Jurnal Teknik*, *6*(1), 18–29. mailto:ariwibowo@unsub.ac.id

Wibowo, A., & Setiadi, E. (2024). Uji Kelayakan Konstruksi Balok Pt Daeyung Dari Aplikasi Sap2000 Menggunakan Machine Learning Dengan Model Linier Regression. *Jurnal Teknik Sipil Cendekia (Jtsc)*, *5*(1), 799–814. https://doi.org/10.51988/jtsc.v5i1.185

Wibowo, A., & Yulianto, Y. (2023). Deteksi Keretakan Jalan Aspal Menggunakan Metode Convolutional Neural Network. *Jurnal Teknik Sipil Cendekia (Jtsc)*, *4*(2), 581–593. https://doi.org/10.51988/jtsc.v4i2.132

Yastawan, I. N., Wedagama, D. M. P., & Ariawan, I. M. A. (2021). Penilaian Kondisi Jalan Menggunakan Metode SDI (Surface Distress Index) Dan Inventarisasi Dalam GIS (Geographic Information System) Di Kabupaten Klungkung. *Jurnal Spektran*, *9*(2), 181.

Zhang, A. A., Shang, J., Li, B., Hui, B., Gong, H., Li, L., Zhan, Y., Ai, C., Niu, H., Chu, X., Nie, Z., Dong, Z., He, A., Zhang, H., Wang, D., Peng, Y., Wei, Y., & Cheng, H. (2024). Intelligent pavement condition survey: Overview of current researches and practices. *Journal of Road Engineering*, *4*(3), 257–281.

https://doi.org/https://doi.org/10.1016/j.jreng.2024.04.003
Zou, Q., Zhang, Z., Li, Q., Qi, X., Wang, Q., & Wang, S. (2019). DeepCrack: Learning Hierarchical Convolutional Features for Crack Detection. *IEEE Transactions on Image Processing*, *28*(3), 1498–1512. https://doi.org/10.1109/TIP.2018.2878966