Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

Balinese Statue Image Classification Using Transfer Learning: A Comparative Study of MobileNetV3 and EfficientNetV2

Dwi Wulandari[1], Ida Bagus Ary Indra Iswara[*2], I Gede Made Yudi Antara[3], I Made Dwi Putra Asana[4], I Kadek Dwi Gandika Supartha[5]

[1,2,3,4,5]Informatics, Institut Bisnis dan Teknologi Indonesia, Bali, Indonesia

Email: [2]indraiswara@instiki.ac.id

Received : Dec 10, 2025; Revised : Dec 20, 2025; Accepted : Dec 20, 2025; Published : Apr 16, 2026

Abstract

Balinese sculpture is an important form of cultural heritage that exhibits high visual diversity in terms of shape, structure, and carving style, which makes manual identification and documentation challenging. Previous studies on automated statue classification have generally focused on limited sculpture categories and therefore do not fully represent the visual diversity of Balinese sculptures. This study aims to develop an automatic image classification model capable of recognizing multiple Balinese statue categories using transfer learning and fine-tuning strategies. The proposed approach compares two convolutional neural network architectures, MobileNetV3 and EfficientNetV2, across eight statue classes: Dewa, Dewi, Mitologi, Penabuh, Pengapit, Punakawan, Raksasa, and Wanara. A dataset of 8,400 images was constructed from three-dimensional video documentation to capture multiple viewing angles of each statue. The images were processed through frame extraction, resizing, normalization, data augmentation, and dataset splitting. Model training was conducted in two stages, consisting of transfer learning followed by fine-tuning using reduced learning rates. Experimental results indicate that both models achieve high classification performance on the test dataset. MobileNetV3 obtained the highest test accuracy of 99.64% with a loss value of 0.0119, while EfficientNetV2 achieved an accuracy of 98.56% with a loss of 0.0613. These findings demonstrate that lightweight architectures can deliver competitive performance when supported by appropriate training strategies. This study provides a comparative evaluation of efficient deep learning models for cultural heritage image classification and supports the development of more reliable and systematic digital documentation of Balinese sculptures.

Keywords : *Balinese Sculpture Classification, Convolutional Neural Networks, Cultural Heritage Preservation, EfficientNetV2, MobileNetV3, Transfer Learning.*

1. INTRODUCTION

Bali is widely recognized for its diverse artistic heritage. One of the most prominent and enduring art forms is sculpture [1]. Sculpture is an artistic creation composed of multiple elements arranged in balanced proportions to produce aesthetic value and distinctive characteristics that can be appreciated from various viewpoints [2]. It combines solid and empty spaces and offers freedom in determining form, production methods, and modes of expression. Balinese sculpture represents a traditional artistic practice that reflects the cultural and spiritual values of local communities and functions as a medium for conveying beliefs and traditions that remain embedded in daily life [3]. External influences, particularly through interactions with tourists and Western artists, have encouraged a transformation from sacred sculptural forms to more liberated and expressive styles. This shift has resulted in the emergence of increasingly diverse sculptural types in terms of style, function, and form. These developments also broaden the visual complexity of Balinese sculptures and strengthen the need for more systematic identification methods.

The diversity of styles and forms that characterizes the evolution of Balinese sculptures enriches their aesthetic value while simultaneously presenting challenges in categorization and identification.

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

Several sculptures exhibit highly similar visual attributes, making manual recognition difficult. Addressing this problem requires an automatic classification approach capable of distinguishing sculptures based on subtle visual differences. Image classification refers to the process of grouping pixels or picture elements within an image into classes, where each class represents a specific entity with identifiable characteristics [4]. With increasing access to image processing technologies, such approaches have become particularly important for cultural objects with substantial variation in shape and visual detail, such as Balinese sculptures. Research on the classification of Balinese sculptures has been conducted previously. In [5], Convolutional Neural Networks (CNN) were applied to classify Dewa and Dewi sculptures into five classes. Their study involved constructing a CNN model and comparing it with AlexNet and ResNet. The constructed model achieved the highest accuracy of 97.14%, while AlexNet and ResNet achieved 24.44% and 33.33%, respectively. Although these results demonstrate the potential of deep learning for Balinese sculpture classification, the study was limited to a single sculpture category and a relatively narrow classification scope. This limitation highlights the need for broader and more representative classification studies that can capture the diversity of sculptural forms found in Bali. Therefore, the present study expands the classification scope to include a wider range of sculpture types commonly found in Bali, divided into eight classes: Dewa, Dewi, Mitologi, Penabuh, Pengapit, Punakawan, Raksasa, and Wanara. This broader approach enables a more comprehensive analysis of the diverse visual characteristics of Balinese sculptures.

In recent years, advances in computer vision and deep learning have increasingly been applied to cultural heritage documentation and analysis, with transfer learning becoming a common strategy for coping with limited labeled data and visual complexity in such domains [6]. Lightweight CNN architectures like MobileNet and EfficientNet are particularly valued for balancing performance and efficiency in heritage classification tasks [7], [8]. Recent studies have demonstrated deep learning approaches for classification of heritage buildings and artifacts using transfer learning and data augmentation [9]. In addition, systematic reviews highlight the broad applicability of pretrained CNNs in visual heritage analysis [10].

To develop the classification model, this study employs deep learning methods using Convolutional Neural Networks (CNN), an architecture specifically designed for processing two-dimensional data [11]. A CNN consists of feature extraction through convolutional operations and classification using the conceptual framework of neural networks that emulate the human nervous system [12]. This research develops an automatic classification model for Balinese sculptures by comparing the performance of EfficientNetV2, which is designed with progressive learning mechanisms and optimized scaling to achieve high accuracy with more efficient training [13]. In [14], MobileNetV3 was developed through Neural Architecture Search to produce lightweight yet accurate models for image classification tasks. In [15], EfficientNetV2 achieved a test accuracy of 94.29% at the sixth epoch and increased to 95.89% by the fifteenth epoch in a bird species classification task. These results indicate that the EfficientNetV2 architecture is capable of rapid convergence while maintaining high performance on visually complex datasets. In [16], MobileNetV3-Large was evaluated for histopathological breast cancer image classification using the BreakHis_v1 dataset. The model achieved an F1-score of 0.98 for binary classification and the highest accuracy of 0.95 for ductal carcinoma subtypes. This result shows that MobileNetV3, despite its lightweight and computationally efficient design, can deliver strong classification performance without extensive preprocessing. Overall, these findings suggest that EfficientNetV2 is well suited for visually complex datasets, while MobileNetV3 offers a favorable trade-off between accuracy and computational efficiency.

The model in this study is developed using transfer learning, a machine learning technique in which a model previously trained for one task is repurposed for a related task [17]. In recent studies, deep transfer learning has been shown to improve classification performance in complex image domains

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

by leveraging pre-trained features that generalize across visual tasks, particularly when labeled data are limited [18], [19]. This approach accelerates training while maintaining high accuracy because the model has already learned relevant representations. Fine-tuning is also applied to adjust the pre-trained parameters so the model can better accommodate the distinct visual patterns of Balinese sculptures [20], [21]. Prior work on cultural heritage image classification using transfer learning demonstrates that extracting and adapting high-level features from pre-trained CNN backbones significantly enhances the representation of complex visual structures inherent to heritage objects [22]. The combination of transfer learning and fine-tuning enables the model to capture more specific visual patterns in Balinese sculptures while preserving the inherent strengths of the underlying architecture. This is supported by [23], who demonstrated that fine-tuning significantly improves a pre-trained model's ability to adapt to new visual domains and enhances classification accuracy.

Another important aspect of this study lies in the characteristics of the dataset. The images of Balinese sculptures were extracted from three-dimensional video documentation, allowing the dataset to capture sculptures from multiple viewing angles. Compared to conventional single-view image acquisition, this approach introduces greater variability in visual appearance, which is expected to improve the robustness of the trained models when distinguishing between visually similar sculpture classes [24]. Research exploring photogrammetric 3D model-based synthetic dataset generation suggests that leveraging multi-view 3D reconstructions can augment diversity in training data and enhance deep learning model robustness for complex visual tasks [25]. Based on the limitations identified in previous studies and the recent developments in deep learning for cultural heritage analysis, this study aims to evaluate the effectiveness of transfer learning combined with fine-tuning for multi-class Balinese sculpture classification. Recent studies indicate that transfer learning is particularly effective for cultural heritage image classification tasks characterized by limited labeled data and high visual complexity [26]. In addition, systematic comparative evaluations of lightweight convolutional neural network architectures have been emphasized as an important step to balance classification accuracy and computational efficiency in real-world applications [7]. Through a controlled comparative analysis of MobileNetV3 and EfficientNetV2 under identical training configurations, this study seeks to provide a clearer understanding of the suitability of lightweight CNN architectures for classifying visually complex cultural heritage objects, while supporting more reliable digital documentation practices [10].

2. METHOD

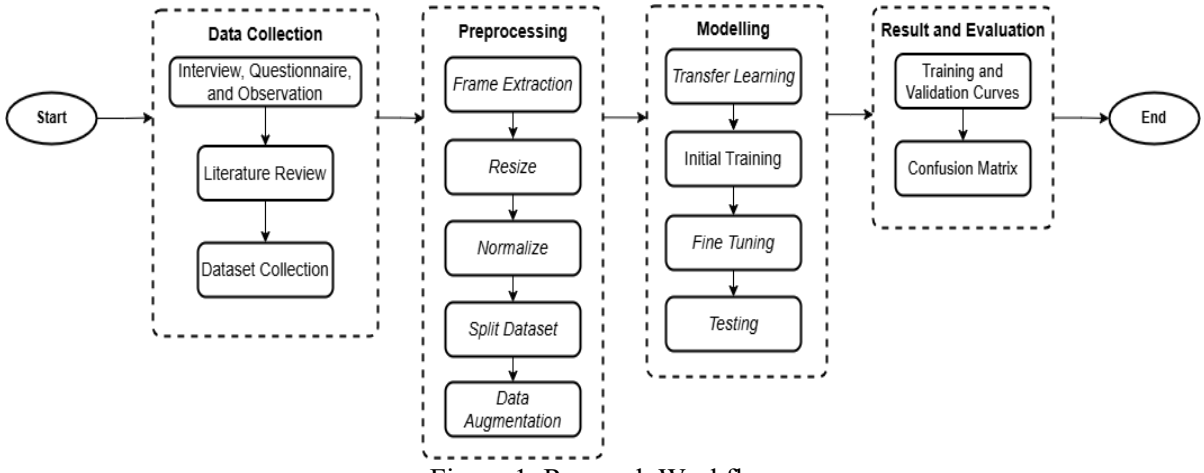

Figure 1. Research Workflow

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

The research begins with a preliminary phase that includes selecting the topic, identifying the problem, defining the objectives, and determining the scope of the study. This is followed by data collection conducted directly in the field, accompanied by a literature review. The collected dataset then undergoes preprocessing, which consists of several steps to prepare the data for classification, including frame extraction, resizing, normalization, dataset splitting, and data augmentation. The next stage involves the classification process using EfficientNetV2 and MobileNetV3 models through the application of transfer learning and fine-tuning. The final stage is the evaluation of the developed models. The overall workflow of the proposed method is illustrated in Figure 1, which outlines each stage from data collection to model evaluation in a sequential manner.

2.1. Data Collection

The dataset used in this study was collected from 240 Balinese stone sculptures located in various environments across Bali, including sculpture workshops, temples, public parks, and open cultural spaces. Three-dimensional video documentation was captured using the Polycam application, allowing each sculpture to be recorded from multiple viewing angles. During the acquisition process, particular attention was given to ensuring that each sculpture was fully captured from all visible sides to avoid missing structural parts, thereby preserving the completeness of the object representation in the dataset. From the recorded three-dimensional videos, two-dimensional image frames were extracted to generate the final dataset. This multi-view acquisition strategy was adopted to capture variations in viewpoint, orientation, and illumination, which are commonly encountered in real-world cultural heritage documentation scenarios. By deriving images from three-dimensional recordings rather than single-view photography, the dataset inherently incorporates visual diversity that supports the robustness and generalization capability of the classification models. In total, 8,400 images were obtained and categorized into eight sculpture classes: Dewa, Dewi, Mitologi, Penabuh, Pengapit, Punakawan, Raksasa, and Wanara. Each class consists of images generated from 30 distinct sculptures, resulting in approximately 1,050 images per class and ensuring balanced representation across all categories. Because no significant class imbalance was present, no additional class reweighting or resampling techniques were applied during model training, ensuring that classification performance was not biased toward any particular sculpture class. Sample images from all eight classes are presented in Figure 2.

Figure 2. Dataset Sample

2.2. Pre-processing

The preprocessing stage prepares the image data before it is fed into the model [27]. Its main purpose is to ensure that the input aligns with the model requirements so the network can learn effectively and produce accurate predictions [28]. In this study, preprocessing consisted of frame extraction, resizing, normalization, data splitting, and data augmentation.

2.2.1. Frame Extraction

The dataset used in this study consists of two-dimensional images generated from three-dimensional statue videos recorded using Polycam. These videos were converted into sequences of static

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

images through a frame extraction process implemented using a Python-based program. A Tkinter-based graphical user interface was developed to facilitate efficient and consistent conversion of video files into image frames. Each 8-second video was decomposed into 35 frames, producing a set of images that captured various viewing angles and structural details of the statues. To maintain proportional consistency during subsequent resizing, all extracted frames were cropped to a 1:1 aspect ratio.

2.2.2. Data Split

Before the data processing stage, the dataset was divided into three subsets using a 70:20:10 ratio, consisting of 70% for the training set, 20% for the validation set, and 10% for the test set[29]. The training set was used to learn feature representations, while the validation set was employed to monitor performance during training and detect potential overfitting [30], [31] The test set was reserved exclusively for final performance evaluation on unseen data.

To ensure reproducibility of the data partitioning process, a fixed random seed was applied during dataset splitting. The same seed value was consistently set across the Python random module, NumPy, and TensorFlow, guaranteeing identical data splits across repeated experimental runs. This deterministic splitting strategy ensures that performance differences observed between models are attributable to architectural and training variations rather than randomness in data allocation.

2.2.3. Resize

Resizing was performed to ensure that all images meet the dimensional requirements of the model [32]. Resizing was performed to ensure that all images meet the dimensional requirements of the model [33], [34]. Standardizing image size ensures that both models process uniform inputs and prevents unnecessary variability caused by inconsistent resolutions. Figure 3 shows an example of the resized images.

Figure 3. Resized Image

2.2.4. Normalization

Data normalization is the process of scaling numerical values into the 0 - 1 range to enable more stable computations and prevent excessively large outputs[35]. In general, normalization serves to standardize data that originate from different value scales or distributions, ensuring that they can be processed effectively by the same algorithm or model [36]. EfficientNetV2 utilizes its built-in preprocessing function for input scaling, while MobileNetV3 applies ImageNet-based normalization to align input pixel values with the statistical properties of its pretrained weights. This ensures compatibility between the pretrained backbone and the target dataset.

2.2.5. Data Augmentation

Data augmentation is a process of applying random transformations to an image to generate new variations while preserving the original label or category [37]. This approach increases the size of the dataset without requiring additional data collection in the field. Data augmentation also helps reduce the

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

risk of overfitting, a condition in which the model learns the training data too closely, causing it to capture unnecessary and irrelevant details [38]. The augmentation techniques applied in this study include rotation, horizontal flipping, width and height shifting, and zooming, as summarized in Table 1. Examples of augmented images are shown in Figure 4.

Tabel 1. Data Augmentation Methods Used

No	Augmentation	Value
1	Flip	*Horizontal*
2	Width shift	0.1
3	Height Shift	0.1
4	Zoom	0.1
5	Rotation	15

Figure 4. Examples of Augmented Images

2.3. Modelling

The modelling stage aims to build and train deep learning models capable of classifying Balinese statues into eight predefined classes. Transfer learning is applied using EfficientNetV2 and MobileNetV3, implemented in TensorFlow and Keras.

2.3.1. Load Pretrained Model

The pretrained EfficientNetV2 and MobileNetV3 models were loaded along with their ImageNet-trained weights. These models provide feature extraction layers capable of recognizing general visual patterns. The original top layers were removed and replaced with a custom classifier consisting of fully connected layers and an 8-class softmax output, allowing adaptation to the Balinese statue dataset. This step ensures that pretrained visual representations are retained while enabling learning of task-specific features.

2.3.2. Training and Fine-tuning

The training regimen was executed in two distinct phases. Initially, only the newly added classification output layer was permitted to learn, accomplished by freezing the parameters of the foundational network layers. Subsequently, a fine-tuning step was implemented; this involved selectively unfreezing some of the deeper layers. This crucial action allowed the pre-trained weights of the model to adapt and specialize to the specific visual features inherent in the Balinese statue dataset. Across all trials, models underwent 50 training cycles (epochs) with a batch size of 32. The optimization was managed using the Adam algorithm, coupled with a sparse categorical cross-entropy objective function. Finally, an analysis was performed by comparing the classification effectiveness of models that utilized this weight adaptation strategy (fine-tuning) versus those that did not.

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

2.3.3. Testing

After training, the models were evaluated using the unseen test set consisting of 840 images distributed across eight classes. Each image was passed into the EfficientNetV2 and MobileNetV3 models to generate predictions. The evaluation configuration is summarized in Table 2, which lists the learning rates and fine-tuning settings for each experiment.

Table 2. Testing Scenario for EfficienNetV2 and MobileNetV3

Data and Model	Learning Rate
Without Fine-Tuning	1e-4
Without Fine-Tuning	1e-3
With Fine-Tuning	1e-3 + 1e-5
With Fine-Tuning	1e-4 + 1e-6

2.4. Result & Evaluation

The evaluation stage measures model performance under different learning rate and fine-tuning configurations. Accuracy and loss curves across epochs are visualized to assess convergence behavior and detect signs of overfitting. In this study, model performance is primarily assessed using top-1 accuracy and class-wise metrics, including precision, recall, and F1-score, to maintain consistency with previous image classification studies in cultural heritage analysis. Although higher-order metrics such as top-k accuracy can be used to account for inter-class visual similarity, the experimental evaluation in this research focuses on top-1 accuracy and confusion-matrix-based metrics to ensure direct comparability with prior studies. This choice allows for direct comparison with prior works and emphasizes exact class prediction performance. The potential use of top-k accuracy, such as top-3 accuracy, is acknowledged as a future extension to further analyze classification ambiguity among visually similar sculpture classes.

In this research, the confusion matrix is used to examine the classification behavior of the model in more detail. True Positive (TP) denotes statue images that are correctly assigned to their actual class, indicating successful recognition. True Negative (TN) refers to images that are accurately identified as not belonging to a particular class. False Positive (FP) represents instances where the model incorrectly assigns an image to a class it does not belong to, while False Negative (FN) captures cases in which the model fails to classify an image into its correct class. These four components provide a technical foundation for calculating accuracy, precision, recall, and F1-score, while also helping to identify specific patterns of misclassification within the dataset [39]. From the confusion matrix, several key evaluation metrics are derived:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Accuracy represents the overall correctness of the model's output. It is precisely defined as the quotient of correctly classified samples divided by the entire set of samples evaluated. Essentially, this metric quantifies the degree of alignment between the model's predictions and the true underlying labels, where a greater resulting value signifies a superior operational performance [40] (1).

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Precision is a key metric used to assess the reliability of a model's positive classifications. It is defined as the ratio of correctly predicted positive observations to the entire pool of observations that were predicted as positive. Consequently, this value reflects the overall trustworthiness of the positive predictions generated by the system. [40] (2).

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Recall measures the system's ability to successfully identify relevant positive samples. It is defined as the ratio of correctly predicted positive observations to the total number of actual positive cases. This metric fundamentally assesses how well the model avoids False Negative errors. [40] (3).

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

The F1-score is an evaluation metric that effectively consolidates precision and recall into a singular measure. It is defined as the harmonic mean of these two quantities. This single value is utilized to represent the overall effectiveness and balance of a classification system's performance[41](4).

3. RESULT

This study evaluates the performance of two CNN architectures utilizing transfer learning, EfficientNetV2 and MobileNetV3, across four training configurations: two without fine-tuning (learning rates 1e-3 and 1e-4), and two with fine-tuning using two-stage learning rate schedules (1e-3 followed by 1e-5, and 1e-4 followed by 1e-6). Each configuration was trained for 50 epochs with a batch size of 32, and evaluated using the separate test set consisting of 840 images across eight statue classes. The presentation of results is organized into training behavior, accuracy–loss curves, confusion matrices, and detailed per-class performance.

3.1. EfficientNetV2 Performance Model Results

Both EfficientNetV2 and MobileNetV3 were trained under four configurations that differed in learning rate and the use of fine-tuning. Although the training setup is identical for both models, each architecture responds to these configurations in its own way. To provide a clearer picture of how the learning process unfolds, the behavior of each configuration is described before the corresponding figure. This approach allows readers to relate the narrative directly to the visual patterns observed in the accuracy and loss curves.

In the first configuration, EfficientNetV2 was trained for 50 epochs without fine-tuning, meaning that all pretrained layers remained frozen and only the classification head was updated. The training curve shows a steady and gradual rise in accuracy, accompanied by a smooth decline in loss throughout the entire training process. Because the model learns in a single continuous phase without structural changes, the trend remains stable from start to finish. The training and validation curves remain close to each other, suggesting that the model fits the data well without exhibiting notable overfitting. The corresponding accuracy and loss curves for this configuration are presented in Figure 5.

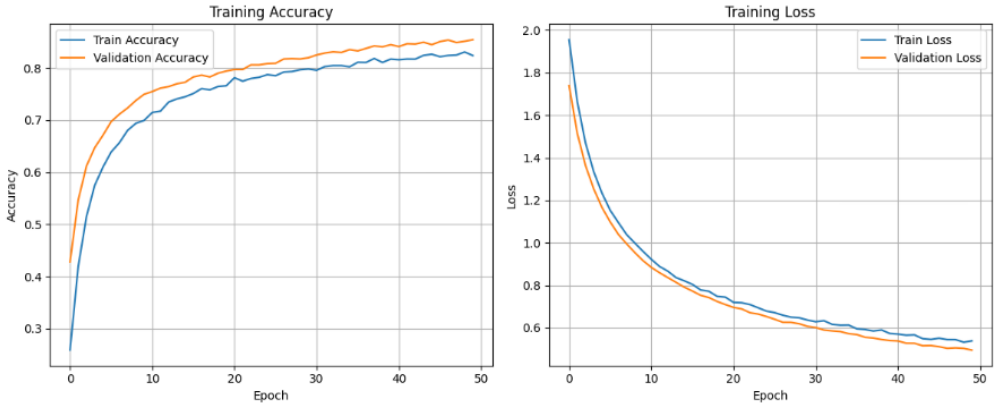

Figure 5. EfficientNetV2 without fine-tuning (LR = 1e-4)

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

The second non–fine-tuning configuration adopts a higher learning rate of 1e-3, which leads to noticeably faster improvement during the early epochs. Both accuracy and loss curves exhibit sharper changes as the model adapts more aggressively. Despite the faster learning pace, the validation curves remain broadly consistent with the training curves, indicating that the model manages to generalize well without overfitting. The behavior reflects the effect of the larger learning rate, which drives the optimization process more rapidly while remaining stable. The corresponding accuracy and loss curves for this configuration are presented in Figure 6.

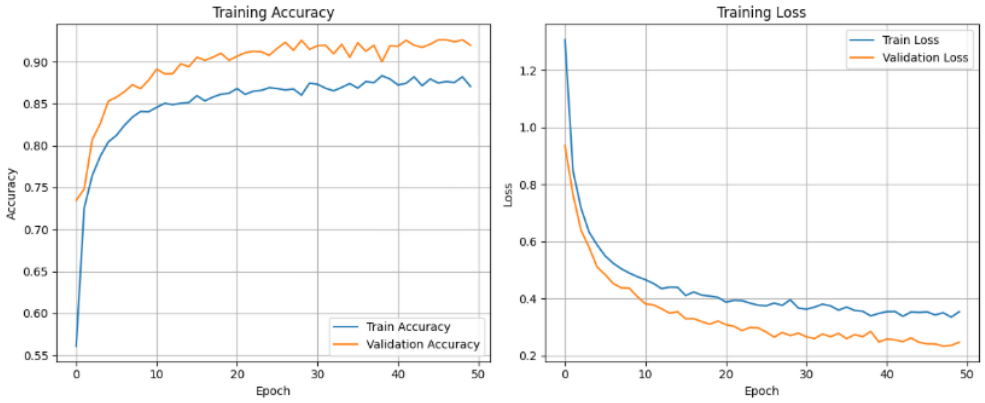

Figure 6. EfficientNetV2 without fine-tuning (LR = 1e-3)

The training curve shows a clear two-phase learning pattern. During the initial transfer-learning stage, both training and validation accuracy increase rapidly and remain close, indicating stable learning. When fine-tuning begins, a brief fluctuation appears visible as a small dip in validation accuracy and a short spike in validation loss. This behavior is typical when deeper layers are unfrozen, as the model readjusts its internal representations. After a few epochs, both curves stabilize again and continue improving steadily, showing better convergence and stronger generalization in the later stages. The corresponding accuracy and loss curves for this configuration are presented in Figure 7.

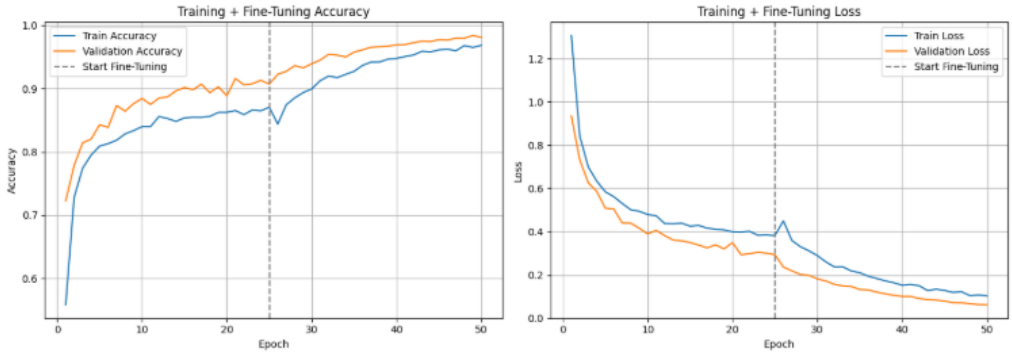

Figure 7. EfficientNetV2 with fine-tuning (LR = 1e-3 → 1e-5)

With lower learning rates, the curves appear smoother and more gradual throughout training. The transition into fine-tuning produces only minor changes in accuracy and loss, suggesting a stable adaptation process with minimal disruption when deeper layers are unfrozen. Improvements occur slowly but consistently during the fine-tuning phase, leading to gradual refinement without large oscillations. This configuration favors safer convergence and controlled learning dynamics, making it suitable for tasks that require conservative parameter updates. The corresponding accuracy and loss curves for this configuration are presented in Figure 8.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

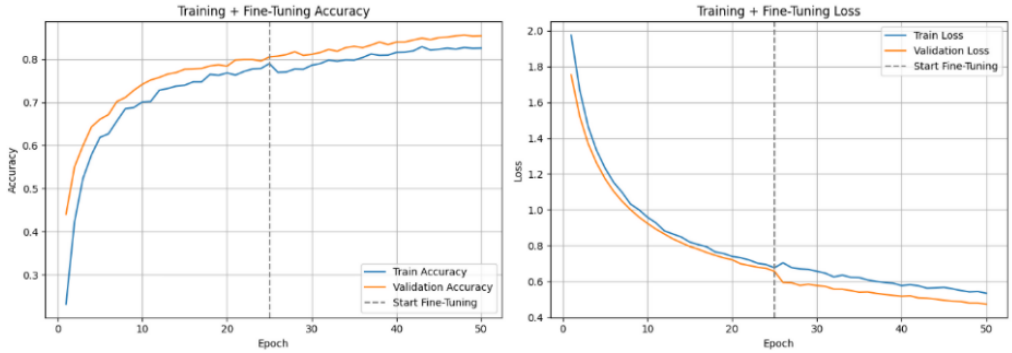

Figure 8. EfficientNetV2 with fine-tuning (LR = 1e-4 → 1e-6)

When fine-tuning was applied, EfficientNetV2 consistently achieved higher validation accuracy and lower validation loss compared to the non–fine-tuning configurations. The two-stage learning-rate schedules allowed the model to refine its pretrained representations more effectively, resulting in smoother convergence patterns and stronger generalization on the validation set. Among the four scenarios, the fine-tuning configuration with a learning rate of 1e-3 followed by 1e-5 produced the best performance, reaching the highest validation accuracy and the lowest validation loss. A summary of the best training and validation performance for all configurations is presented in Table 3.

Table 3. Results Accuracy with EfficientNetV2

Data and Model	Learning Rate	Train-Acc	Train-Loss	Val-Acc	Vall-Loss
Without Fine-Tuning	1e-4	82.14%	0.5386	85.42%	0.4951
Without Fine-Tuning	1e-3	87.81%	0.3414	92.38%	0.2332
With Fine-Tuning	1e-3 + 1e-5	96.72%	0.0985	97.86%	0.0648
With Fine-Tuning	1e-4 + 1e-6	82.34%	0.5321	85.42%	0.4725

After identifying the fine-tuning configuration with a learning rate of 1e-3 followed by 1e-5 as the best-performing setup, the model was evaluated on the independent test set consisting of 840 images from eight statue classes. This evaluation reports class-level precision, recall, and F1-score, as well as the overall accuracy, to provide a comprehensive view of the model's generalization performance. The confusion matrix and its normalized version are presented in Figure 9, illustrating how well the model distinguishes among visually similar statue categories and where occasional misclassifications occur.

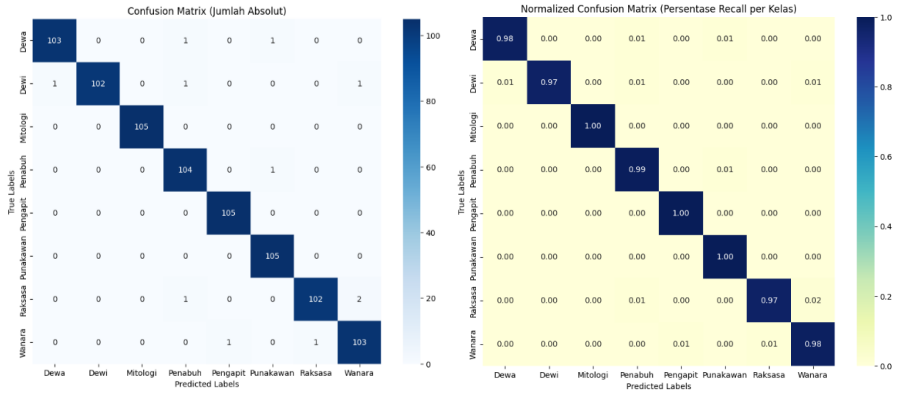

Figure 9. Confusion Matrix and Normalized Confusion Matrix

The complete test results for each class are listed in Table 4. Overall, the model demonstrated strong and consistent performance across all categories, reflected in the high precision, recall, and F1-scores.

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

Table 4. Test Metrics for EfficientNetV2

Class	Precision	Recall	F1-Score
Dewa	0.99	0.98	0.99
Dewi	1.00	0.97	0.99
Mitologi	1.00	1.00	1.00
Penabuh	0.97	0.99	0.98
Pengapit	0.99	1.00	1.00
Punakawan	0.98	1.00	0.99
Raksasa	0.99	0.97	0.98
Wanara	0.97	0.98	0.98
Accuracy			0.98
Macro Avg	0.98	0.98	0.98
Weighted Avg	0.98	0.98	0.98

The EfficientNetV2 model reached an overall accuracy of 98.56% on the test dataset. Both macro and weighted averages were 0.98, reflecting stable performance across all classes. Most statue categories achieved near-perfect precision and recall, while minor misclassifications occurred in classes such as Dewa, Dewi, and Raksasa, likely due to subtle visual similarities in their carvings. These results indicate that EfficientNetV2 can reliably distinguish between the different statue types, demonstrating strong generalization to unseen images.

3.2. MobileNetV3 Performance Model Results

MobileNetV3 was trained using the same four configurations applied to EfficientNetV2, but the model exhibited its own learning characteristics due to differences in architectural complexity and parameter capacity. As with the previous model, each configuration is explained prior to the corresponding figure to help readers interpret the accuracy and loss patterns more clearly.

In the first configuration, MobileNetV3 was trained without fine-tuning using a learning rate of 1e-4. The accuracy curve increases steadily with a relatively gentle slope, while the loss curve decreases in a smooth and consistent manner. Compared to EfficientNetV2, the progression appears slightly slower, which aligns with MobileNetV3's compact architecture that limits the rate of feature adaptation. The training and validation curves remain closely aligned, indicating good generalization and an absence of overfitting. The corresponding accuracy and loss curves for this configuration are presented in Figure 10.

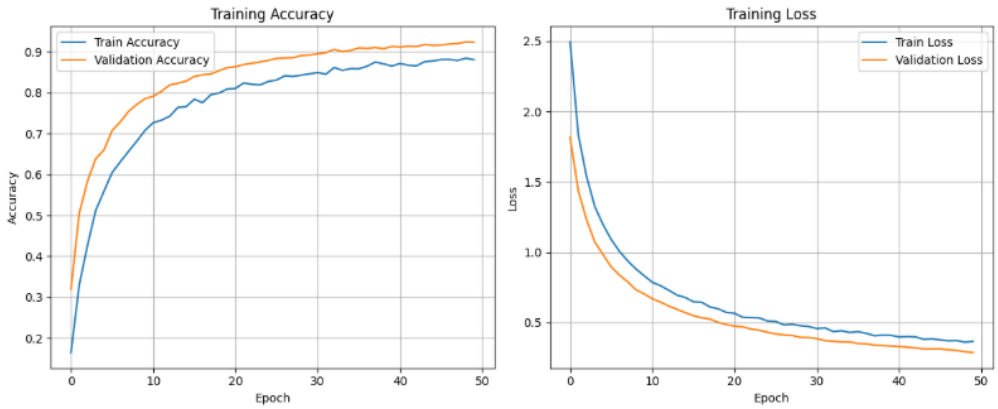

Figure 10. MobileNetV3 without fine-tuning (LR = 1e-4)

The second configuration, which uses a higher learning rate of 1e-3, shows a noticeably faster rise in accuracy during the initial epochs. The steeper learning trajectory indicates that MobileNetV3 adapts more aggressively under this setting. While this configuration accelerates early learning, minor

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

fluctuations appear in the validation curve, suggesting that the model becomes more sensitive to shifts in the optimization process. Nevertheless, both curves remain broadly consistent, showing that the model remains stable despite the faster learning rate. The corresponding accuracy and loss curves for this configuration are presented in Figure 11.

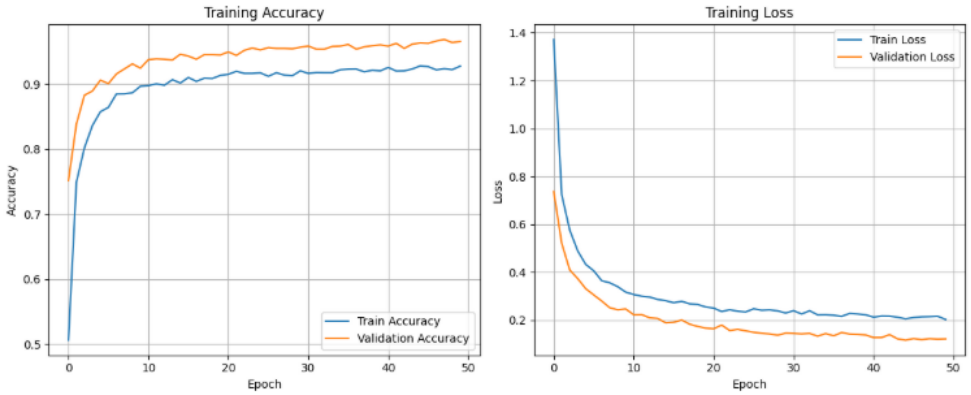

Figure 11. MobileNetV3 without fine-tuning (LR = 1e-3)

The third configuration introduces fine-tuning by training the classification head for 25 epochs with a learning rate of 1e-3, followed by unfreezing selected backbone layers and continuing training for another 25 epochs with a reduced rate of 1e-5. The transfer-learning phase produces rapid accuracy gains, while the transition to fine-tuning introduces a brief dip in accuracy and a momentary spike in loss as previously frozen backbone layers are unfrozen and begin updating. After this short adjustment, both curves stabilize and continue improving steadily, reflecting effective refinement of features during the fine-tuning stage. The corresponding accuracy and loss curves for this configuration are presented in Figure 12.

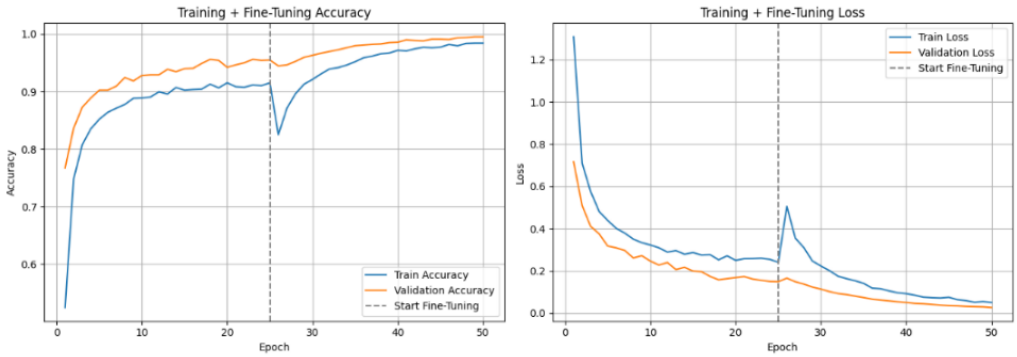

Figure 12. MobileNetV3 with fine-tuning (LR = 1e-3 → 1e-5)

The fourth configuration adopts a more conservative approach, using a learning rate of 1e-4 during the transfer learning phase and 1e-6 during fine-tuning. The accuracy increases steadily throughout training, with curves that are notably smooth and stable. The transition between the two phases is subtle, reflecting the effect of the small learning rates that limit abrupt parameter updates. This results in stable convergence, though the improvement gained during fine-tuning is smaller in scale compared to the previous configuration. The corresponding accuracy and loss curves for this configuration are presented in Figure 13.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

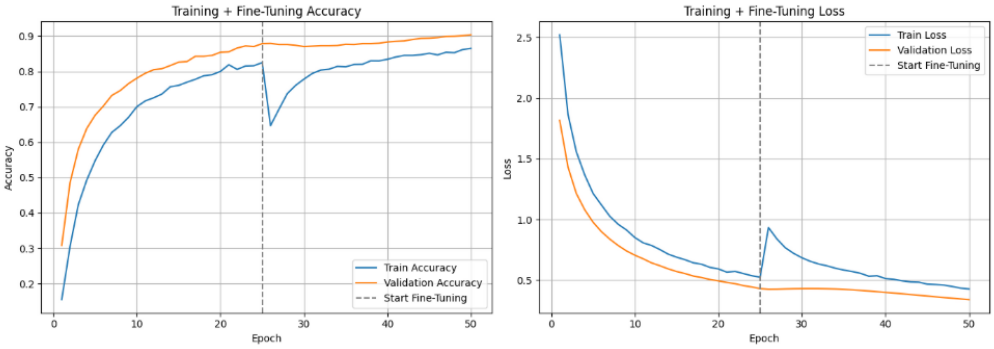

Figure 13. MobileNetV3 with fine-tuning (LR = 1e-4 → 1e-6)

Among the four configurations, MobileNetV3 achieved its strongest validation performance in the fine-tuning scenario that used learning rates of 1e-3 followed by 1e-5. This setting produced the highest validation accuracy and the lowest validation loss across the model's training runs. A summary of the best performance for each configuration is presented in Table 5.

Table 5. Results Accuracy with MobileNetV3

Data and Model	Learning Rate	Train-Acc	Train-Loss	Val-Acc	Vall-Loss
Without Fine-Tuning	1e-4	88.62%	0.3541	92.38%	0.2934
Without Fine-Tuning	1e-3	92.80%	0.2013	96.31%	0.1157
With Fine-Tuning	1e-3 + 1e-5	98.10%	0.0577	99.52%	0.0266
With Fine-Tuning	1e-4 + 1e-6	85.82%	0.4373	90.36%	0.3388

The best MobileNetV3 configuration was evaluated on the independent test set of 840 images across eight statue classes. The class-level precision, recall, and F1-score, along with overall accuracy, are reported in Table 5. The confusion matrix and its normalized version are shown in Figure 14, highlighting the model's ability to distinguish visually similar categories and the locations of occasional misclassifications.

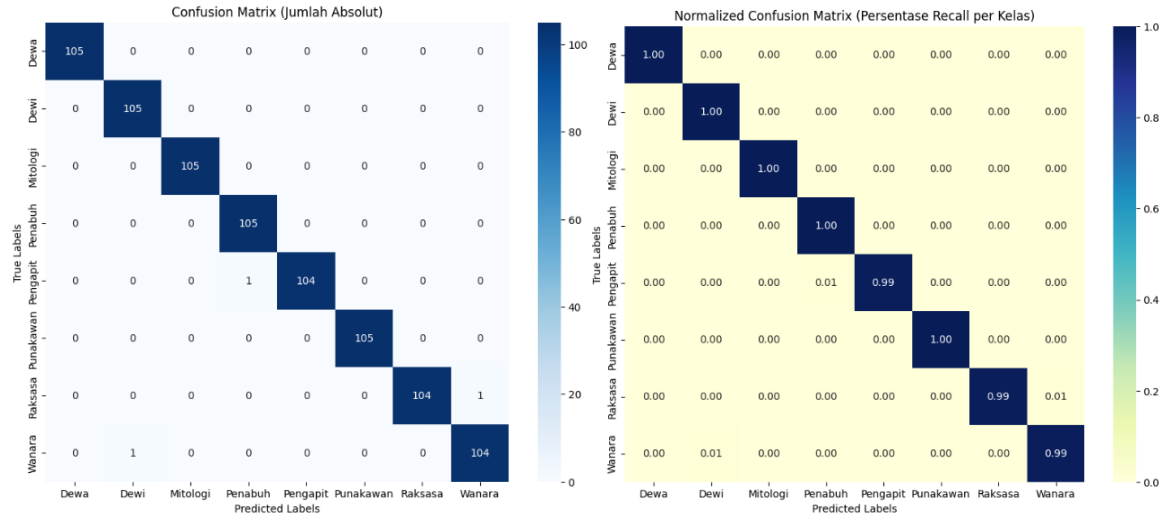

Figure 14. Confusion Matrix and Normalized Confusion Matrix

The complete test results for each class are listed in Table 6. Overall, the model demonstrated strong and consistent performance across all categories, reflected in the high precision, recall, and F1-scores.

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

Table 6. Test Metrics for MobileNetV3

Class	Precision	Recall	F1-Score
Dewa	1.00	1.00	1.00
Dewi	0.99	1.00	1.00
Mitologi	1.00	1.00	1.00
Penabuh	0.99	1.00	1.00
Pengapit	1.00	0.99	1.00
Punakawan	1.00	1.00	1.00
Raksasa	1.00	0.99	1.00
Wanara	0.99	0.99	0.99
Accuracy			0.99
Macro Avg	0.99	0.99	0.99
Weighted Avg	0.99	0.99	0.99

The MobileNetV3 model obtained an overall test accuracy of 99.64%, with both macro and weighted averages also at 0.99. This indicates that the model performed consistently across all eight statue classes. Most classes achieved high precision and recall values, although minor deviations were observed in the Dewi, Pengapit, Raksasa, and Wanara categories, reflecting occasional misclassifications likely due to subtle visual similarities among the statues.

3.3. Analysis

The comparative analysis between EfficientNetV2 and MobileNetV3 was conducted based on their best-performing configurations, both utilizing fine-tuning with a two-stage learning rate schedule (1e-3 followed by 1e-5). MobileNetV3, designed as a compact and lightweight network, demonstrates rapid adaptation during early training, especially under higher learning rates (1e-3). The training and validation curves indicate faster initial improvements in accuracy compared to EfficientNetV2, which aligns with MobileNetV3's efficient feature extraction and lower computational overhead. This fast adaptation is beneficial for scenarios requiring quick convergence. However, due to its smaller representational capacity, the magnitude of improvement during fine-tuning is more modest relative to EfficientNetV2. The two-stage learning rate schedule (1e-3 followed by 1e-5) allows MobileNetV3 to refine features carefully during the second phase, achieving a peak validation accuracy of 99.52% and stable loss curves, demonstrating controlled learning and minimal overfitting.

EfficientNetV2, on the other hand, exhibits smooth and stable convergence across all configurations. Its deeper architecture and higher parameter capacity enable precise feature extraction and strong generalization, as evidenced by consistently high validation accuracies and low loss values. While its adaptation during the initial epochs is slightly slower under lower learning rates (1e-4), the model benefits more from fine-tuning, particularly with the two-stage learning rate schedule, reaching 97.86% validation accuracy. The accuracy–loss curves for EfficientNetV2 show a gradual but steady improvement, reflecting careful weight updates and resilience to fluctuations, which is advantageous for datasets containing visually similar classes.

The influence of learning rate is clearly observable. Configurations with an initial high learning rate followed by a smaller rate allow rapid convergence initially, while the lower rate in fine-tuning reduces the risk of overfitting and stabilizes validation loss. Conversely, using a consistently low learning rate throughout results in slower adaptation and smaller overall gains during fine-tuning. To provide a clearer overview of these performance patterns, the comparative results for all configurations are summarized in Table 7.

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

Table 7. Comparison Results

Model	No Fine-Tune (1-e4)	No Fine-Tune (1-e3)	No Fine-Tune (1-e3 & 1-e5)	No Fine-Tune (1-e4 & 1-e4)
EfficientNetV2	85.42%	92.38%	97.86%	85.42%
MobileNetV3	92.38%	96.31%	99.52%	90.36%

On the test set, MobileNetV3 reached an overall accuracy of 99.64% with a loss of 0.0119, slightly surpassing EfficientNetV2 at 98.56% with a loss of 0.0613. Both models maintain high precision, recall, and F1-scores across most classes. Minor misclassifications occurred in visually similar categories such as Dewi, Pengapit, Raksasa, and Wanara, reflecting the challenge of subtle differences in statue details. A consolidated comparison of these configurations is provided in Table 8.

Table 8. Comparison Evaluation confusion matrix

Model	Accuracy	Precision	Recall	F1-Score
EfficientNetV2	0.98	0.98	0.98	0.98
MobileNetV3	0.99	0.99	0.99	0.99

Based on the experimental results, MobileNetV3 demonstrates rapid adaptation to the dataset, achieving relatively quick increases in accuracy and high computational efficiency, making it suitable for scenarios that require fast learning or lightweight deployment. In contrast, EfficientNetV2 exhibits smoother and more stable convergence, with consistent generalization across all classes, providing robustness against subtle visual variations in the images. Although the overall performance difference is relatively small, these characteristics suggest that model selection should be guided by specific needs: MobileNetV3 for speed and efficiency, and EfficientNetV2 for stability and precision in feature representation.

4. DISCUSSIONS

The experimental results indicate that both MobileNetV3 and EfficientNetV2 perform well in classifying Balinese statue images, with accuracy values above 98% on the test set. MobileNetV3 reached the highest performance with a validation accuracy of 99.52% and a test accuracy of 99-64%, while EfficientNetV2 achieved a validation accuracy of 97.86% and a test accuracy of 98.56%. These results show that both architectures are capable of learning the visual characteristics of stone sculptures, which generally involve consistent textures, repetitive structural patterns, and shape-dominant features. Under these conditions, MobileNetV3 displayed slightly better performance, particularly in configurations using a two-stage learning rate schedule.

The difference between the two models becomes clearer when considering how each architecture responds to the characteristics of the dataset. The dataset in this study contains 8,400 images collected from real sculpture environments, with significant variation in angle and carving details. Despite the relatively large dataset, EfficientNetV2 did not surpass MobileNetV3. This outcome is consistent with findings in [42], which reported that MobileNetV3 outperformed EfficientNetV2 in American Sign Language classification under similar transfer learning settings, despite the larger dataset of 13,000 images. Their study highlights that EfficientNetV2 is more sensitive to hyperparameter settings and tends to require more extensive fine-tuning to reach peak performance. A comparable pattern is also visible in this study, where EfficientNetV2 showed stable convergence but gained less improvement from fine-tuning compared to MobileNetV3.

Meanwhile, several previous studies emphasize that EfficientNetV2 remains a strong architecture when trained on visually complex datasets with high intraclass variation. For example, [43] reported that EfficientNetV2-Large achieved 99.9% accuracy in cervical-cell image classification, outperforming smaller variants due to its capacity to capture fine-grained visual cues. Similarly, [15] demonstrated that

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

EfficientNetV2 can reach high accuracy in bird-species classification, with performance improving steadily as training progresses. These findings suggest that EfficientNetV2 excels when detailed texture information is essential. However, in the Balinese statue dataset used in this research, the discriminative features are dominated more by shape and structural form than by fine texture variations. This condition appears to favor MobileNetV3, whose lightweight architecture is optimized through Neural Architecture Search (NAS) to efficiently represent shape-based features.

MobileNetV3 also demonstrated advantages in earlier related work. In [16], it was reported that MobileNetV3-Large achieved an F1-score of 0.98 in histopathological image classification despite its compact size, suggesting that the architecture can adapt well even in domains requiring high-level feature abstraction. This aligns with the present study's finding that MobileNetV3 converges more rapidly and benefits greatly from the two-stage learning rate strategy. Taken together, the literature indicates that MobileNetV3 often provides a favorable balance between accuracy and computational efficiency, whereas EfficientNetV2 tends to reach its best performance when deeper fine-tuning and larger-scale optimization are applied.

When compared to previous research on Balinese sculpture classification, such as the study by [5], which reported a highest accuracy of 97.14% using a custom CNN on five sculpture classes, the present study demonstrates improved generalization with a broader set of eight sculpture categories. This suggests that combining transfer learning with fine-tuning enables more robust feature extraction compared to constructing a CNN from scratch, particularly when the dataset includes wide variation in form and carving intricacies.

Beyond model accuracy, the dataset size and diversity play a critical role in determining the generalization capability of deep learning models. Although the dataset used in this study comprises 8,400 images derived from real-world cultural heritage environments, questions regarding whether this size is sufficient for broader generalization remain relevant. Recent studies in cultural heritage image analysis indicate that dataset diversity particularly variations in viewpoint, illumination, and object morphologycan be as influential as dataset size in achieving robust model performance [9]. In addition, systematic reviews of deep learning applications in cultural heritage emphasize that pretrained convolutional neural networks are well suited for scenarios where labeled data are limited but visually complex, reinforcing the appropriateness of transfer learning for this task [6].

In addition, qualitative inspection of misclassification patterns using confusion matrices indicates that most classification errors occur between visually similar statue classes, such as Dewi and Raksasa, which share comparable pose structures and ornamental elements. This observation suggests that misclassification is primarily driven by inter-class visual similarity rather than model instability. Data augmentation strategies, particularly rotation-based transformations, are therefore important for exposing the model to varied viewing angles and mitigating confusion arising from pose-dependent similarities. Such augmentation helps the network learn more invariant representations of statue geometry across different orientations.

Although Class Activation Mapping (CAM) visualizations were not explicitly included in this study, prior research demonstrates that CAM-based analysis can provide insight into which regions of an image contribute most strongly to a model's prediction. In the context of sculpture classification, CAM techniques may help verify whether the model focuses on semantically meaningful regions such as facial structure, posture, or symbolic ornaments, rather than background artifacts. Incorporating CAM analysis in future work would therefore enhance model interpretability and support more transparent cultural heritage documentation.

From a broader computer vision perspective, emerging architectures such as Vision Transformers (ViT) and hybrid CNN–Transformer models have recently been explored for artistic and cultural image classification, showing promising performance in handling complex visual semantics. While such

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

models were not evaluated in this study, they represent a potential direction for future comparative analysis in cultural heritage classification tasks [44]. Finally, ethical considerations must also be acknowledged when applying artificial intelligence to cultural heritage documentation. Automated classification systems inherently encode the assumptions embedded in their training data, which may influence how cultural artifacts are represented and interpreted. Careful dataset construction, transparent labeling, and collaboration with cultural stakeholders are therefore essential to mitigate potential cultural bias and ensure respectful digital preservation practices.

Overall, the results highlight that model performance is influenced not only by architectural characteristics but also by the interaction between learning rate scheduling, dataset properties, and the depth of fine-tuning. MobileNetV3 demonstrates advantages when discriminative features rely primarily on global shape and structural patterns, whereas EfficientNetV2 provides stable learning dynamics and consistent predictions across classes. The relatively small performance gap between the two models indicates that both architectures are well suited for Balinese statue classification. Future research may explore deeper fine-tuning strategies, alternative learning rate schedulers, CAM-based interpretability analysis, and comparisons with transformer-based architectures to further advance cultural heritage image classification.

From a computer science perspective, this study contributes to the field of deep learning and computer vision by providing a controlled comparative benchmark of lightweight convolutional neural network architectures for fine-grained image classification tasks involving visually complex objects. The findings demonstrate how architectural design choices, learning rate scheduling, and fine-tuning strategies interact with dataset characteristics to influence model performance. In particular, this work highlights the suitability of efficient CNN models such as MobileNetV3 for shape-dominant classification problems, which are common in real-world visual recognition scenarios beyond cultural heritage domains. By bridging methodological advances in transfer learning with practical documentation needs, this research supports the development of scalable, accurate, and computationally efficient visual classification systems, thereby contributing to broader applications in digital documentation, pattern recognition, and applied computer vision.

5. CONCLUSION

This study successfully implemented an image classification approach to recognize eight categories of Balinese stone sculptures using transfer learning and fine-tuning. The workflow covered image extraction from 3D video documentation, preprocessing, model training, and performance evaluation. The experimental results reaffirm that fine-tuning plays a crucial role in refining pre-trained models, especially when the target domain exhibits structural and stylistic characteristics different from those found in standard image datasets.

Both CNN architectures MobileNetV3 and EfficientNetV2 demonstrated strong capability in learning the visual patterns of Balinese sculptures. MobileNetV3 showed advantages in efficiency, rapid convergence, and responsiveness to learning-rate scheduling, making it well-suited for scenarios that require lightweight models. EfficientNetV2 offered stable learning dynamics and consistent predictions across classes, although it required deeper parameter adjustments to fully utilize its representational capacity. Overall, MobileNetV3 achieved slightly better generalization in this research setting, likely due to the dataset's emphasis on shape-dominant and structural visual cues.

The findings indicate that both architectures have meaningful potential to support digital documentation and automated cataloging of cultural heritage artifacts. From a computer science perspective, this research contributes empirical evidence on the effectiveness of lightweight and scalable deep learning architectures for complex visual classification tasks in the cultural heritage domain, particularly when data are derived from three-dimensional documentation and exhibit high inter-class

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

similarity. Future research can be directed toward improving the reliability of the classification model in more realistic conditions. This includes expanding the dataset with additional statue samples, particularly those with higher variation in surface wear, lighting, and viewing distance. Another potential improvement lies in examining other fine-tuning depths or alternative learning rate strategies to determine whether the performance gap between MobileNetV3 and EfficientNetV2 can be further reduced. In addition, integrating the trained models into mobile or web-based applications for cultural tourism, museum documentation, or heritage management systems represents a practical direction to enhance public accessibility and real-world impact.

Further research may also explore privacy-aware or federated learning approaches to enable collaborative model training across data from different artisans, institutions, or regions without requiring centralized data sharing, thereby respecting cultural data ownership while improving model generalization. In addition, testing the models on field-captured images outside the controlled dataset may provide stronger evidence of generalization, ensuring that the classification system remains dependable when applied to real documentation activities.

REFERENCES

[1] I. W. Arissusila, I. G. A. Nilawati, and I. P. I. G. P. Sumardiana, "Dinamika Kerajinan Patung Kayu Dalam Mendukung Pariwisata Budaya Bali," *Dharmasmrti: Jurnal Ilmu Agama dan Kebudayaan*, vol. 20, no. 2, pp. 154–165, Oct. 2020, doi: 10.32795/ds.v20i2.1039.

[2] M. O. Luruk, A. Y. Rahman, and F. Marisa, "Klasifikasi Jenis Patung Menggunakan Metode Convolutional Neural Network (CNN)," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 12, no. 1, pp. 201–208, Mar. 2025, doi: 10.35957/jatisi.v12i1.9247.

[3] P. T. Janu Budi Utama, I. K. Sudita, and A. Sudarmawan, "Patung Padas Dan Ragam Hias Yang Ada Di Pura Gunung Sekar Desa Adat Sangsit Dangin Yeh,Kecamatan Sawan, Buleleng, Bali," *Jurnal Pendidikan Seni Rupa Undiksha*, vol. 13, no. 2, pp. 151–167, Oct. 2023, doi: 10.23887/jjpsp.v13i2.65676.

[4] A. Herdiansah, R. I. Borman, D. Nurnaningsih, A. A. J. Sinlae, and R. R. Al Hakim, "Klasifikasi Citra Daun Herbal Dengan Menggunakan Backpropagation Neural Networks Berdasarkan Ekstraksi Ciri Bentuk," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 2, p. 388, Apr. 2022, doi: 10.30865/jurikom.v9i2.4066.

[5] N. L. G. P. Suwirmayanti, I. M. B. Sentana, I. K. G. D. Putra, M. Sudarma, I. M. Sukarsa, and K. Budiarta, "Deep Learning Implementation Using CNN to Classify Bali God Sculpture Pictures," *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, vol. 15, no. 02, pp. 87–98, Oct. 2025, doi: 10.24843/LKJITI.2024.v15.i02.p02.

[6] F. Ju, "Mapping the Knowledge Structure of Image Recognition in Cultural Heritage: A Scientometric Analysis Using CiteSpace, VOSviewer, and Bibliometrix," *J Imaging*, vol. 10, no. 11, p. 272, Oct. 2024, doi: 10.3390/jimaging10110272.

[7] S. Gokak, P. Khichade, N. Heggalagi, A. Billowria, and S. Hegde, "Defect Detection and Classification of Cultural Heritage Buildings Using Deep Learning," in *Proceedings of the 3rd International Conference on Futuristic Technology*, SCITEPRESS - Science and Technology Publications, 2025, pp. 619–626. doi: 10.5220/0013633700004664.

[8] Z. Liu, W. Hong, R. Long, Y. Zhu, and X. Zhang, "A lightweight enhanced EfficientNet model for Chinese eaves tile dynasty classification," *npj Heritage Science*, vol. 13, no. 1, p. 494, Oct. 2025, doi: 10.1038/s40494-025-02049-3.

[9] A. L. Carvalho Ottoni and L. T. Cordeiro Ottoni, "A deep learning approach for cultural heritage building classification using transfer learning and data augmentation," *J Cult Herit*, vol. 74, pp. 214–224, Jul. 2025, doi: 10.1016/j.culher.2025.06.010.

[10] J. Yang, T. Liu, Y. T. Luo, and P. C.-I. Pang, "Deep learning in cultural imagery dissemination: a systematic scoping review of AI-driven visual transmission mechanisms," *Front Commun (Lausanne)*, vol. 10, Sep. 2025, doi: 10.3389/fcomm.2025.1645168.

[11] F. M. Qotrunnada and P. H. Utomo, "Metode Convolutional Neural Network untuk Klasifikasi Wajah Bermasker," *Prosiding Seminar Nasional Matematika*, vol. 5, no. 5, pp. 799–807, 2022.

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

[12] L. Hakim, H. R. Rahmanto, S. P. Kristanto, and D. Yusuf, "Klasifikasi Citra Motif Batik Banyuwangi Menggunakan Convolutional Neural Network," *Jurnal Teknoinfo*, vol. 17, no. 1, p. 203, Jan. 2023, doi: 10.33365/jti.v17i1.2342.

[13] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," *in International Conference on Machine Learning*, pp. 10096–10106, Jun. 2021, doi: 10.48550/arXiv.2104.00298.

[14] A. Howard *et al.*, "Searching for MobileNetV3," *in Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1314–24, Nov. 2019, doi: 10.48550/arXiv.1905.02244.

[15] G. Singh, K. Guleria, and S. Sharma, "A Pre-trained EfficientNetV2 Deep Learning Model for Birds Species Classification," in *2024 3rd International Conference for Advancement in Technology, ICONAT 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICONAT61936.2024.10774953.

[16] K. DeVoe, G. Takahashi, E. Tarshizi, and A. Sacker, "Evaluation of the precision and accuracy in the classification of breast histopathology images using the MobileNetV3 model," *J Pathol Inform*, vol. 15, Dec. 2024, doi: 10.1016/j.jpi.2024.100377.

[17] A. E. Putra, M. F. Naufal, and V. R. Prasetyo, "Klasifikasi Jenis Rempah Menggunakan Convolutional Neural Network dan Transfer Learning," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 9, no. 1, p. 12, Apr. 2023, doi: 10.26418/jp.v9i1.58186.

[18] P. Yan *et al.*, "A Comprehensive Survey of Deep Transfer Learning for Anomaly Detection in Industrial Time Series: Methods, Applications, and Directions," *IEEE Access*, vol. 12, pp. 3768–3789, 2024, doi: 10.1109/ACCESS.2023.3349132.

[19] P. D. Alfano, V. P. Pastore, L. Rosasco, and F. Odone, "Top-tuning: A study on transfer learning for an efficient alternative to fine tuning for image classification with fast kernel methods," *Image Vis Comput*, vol. 142, p. 104894, Feb. 2024, doi: 10.1016/j.imavis.2023.104894.

[20] Z. Han, C. Gao, J. Liu, J. Zhang, and S. Q. Zhang, "Parameter-Efficient Fine-Tuning for Large Models: A Comprehensive Survey," *arXiv preprint arXiv:2403.14608*, Sep. 2024, doi: 10.48550/arXiv.2403.14608.

[21] N. L. W. S. R. Ginantra, T. Hendrawati, and D. A. P. Wulandari, "Penerapan Metode Stable Diffusion Dengan Fine Tuning Untuk Pola Endek Bali," *TEMATIK*, vol. 11, no. 2, pp. 141–147, Nov. 2024, doi: 10.38204/tematik.v11i2.2069.

[22] A. Rahujo, D. Atif, S. A. Inam, A. A. Khan, and S. Ullah, "A survey on the applications of transfer learning to enhance the performance of large language models in healthcare systems," *Discover Artificial Intelligence*, vol. 5, no. 1, p. 90, Jun. 2025, doi: 10.1007/s44163-025-00339-0.

[23] T. B. Sasongko, H. Haryoko, and A. Amrullah, "Analisis Efek Augmentasi Dataset dan Fine Tune pada Algoritma Pre-Trained Convolutional Neural Network (CNN)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 4, pp. 763–768, Aug. 2023, doi: 10.25126/jtiik.2024106583.

[24] Y. Li, M. Ge, M. Li, T. Li, and S. Xiang, "CLIP-Based Adaptive Graph Attention Network for Large-Scale Unsupervised Multi-Modal Hashing Retrieval," *Sensors*, vol. 23, no. 7, p. 3439, Mar. 2023, doi: 10.3390/s23073439.

[25] G. Patrucco and F. Setragno, "Enhancing Automation Of Heritage Processes: Generation Of Artificial Training Datasets From Photogrammetric 3d Models," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-M-2–2023, pp. 1181–1187, Jun. 2023, doi: 10.5194/isprs-archives-XLVIII-M-2-2023-1181-2023.

[26] E. Nerantzis, L. Malletzidou, E. Kyratzopoulou, N. C. Tsirliganis, and N. A. Kazakis, "From Contemporary Datasets to Cultural Heritage Performance: Explainability and Energy Profiling of Visual Models Towards Textile Identification," *Heritage*, vol. 8, no. 11, p. 447, Oct. 2025, doi: 10.3390/heritage8110447.

[27] M. E. Prasetyo, M. R. Faza, R. Pratama, S. N. H. Alhabsy, H. Purwanti, and A. P. A. Masa, "Klasifikasi Ragam Kendaraan Menggunakan Metode Convolutional Neural Network (Cnn)," *Adopsi Teknologi dan Sistem Informasi (ATASI)*, vol. 2, no. 2, pp. 142–148, Dec. 2023, doi: 10.30872/atasi.v2i2.1156.

Vol. 7, No. 2, April 2026, Page. 1589-1608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5556

[28] J. E. Widyaya and S. Budi, "Pengaruh Preprocessing Terhadap Klasifikasi Diabetic Retinopathy dengan Pendekatan Transfer Learning Convolutional Neural Network," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 1, Apr. 2021, doi: 10.28932/jutisi.v7i1.3327.

[29] R. D. Armelia, R. Andrian, and A. Junaidi, "Perbandingan Kinerja Backpropagation dan Convolutional Neural Network untuk Klasifikasi Citra Batik Lampung," *Jurnal Komputasi*, vol. 12, no. 1, pp. 11–18, Apr. 2024, doi: 10.23960/komputasi.v12i1.248.

[30] W. Musu, A. Ibrahim, and H. Heriadi, "Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4.5," *SISITI : Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, vol. 10, no. 1, pp. 186–195, Mar. 2021, doi: 10.36774/sisiti.v10i1.802.

[31] B. Nugroho and E. Y. Puspaningrum, "Kinerja Metode CNN untuk Klasifikasi Pneumonia dengan Variasi Ukuran Citra Input," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 3, pp. 533–538, Jun. 2021, doi: 10.25126/jtiik.2021834515.

[32] H. K. Putri, B. H. Iswanto, and H. Suhendar, "Klasifikasi Kualitas Cangkang Telur Ayam Menggunakan Efficientnet Berbasis Citra Digital," *Prosiding Seminar Nasional Fisika (E-Journal)*, vol. XIII, 2024, doi: 10.21009/03.1301.FA13.

[33] A. M. Marhelio, M. Munir, and Y. Wihardi, "Klasifikasi Pose Kepala Siswa Menggunakan EfficientNetV2 dengan Seat Position Embedding," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 12, no. 3, Sep. 2025, doi: 10.35957/jatisi.v12i3.12987.

[34] J. Krisna Putra and Y. Yoannita, "Transfer Learning dengan MobileNetV3 untuk Deteksi Serangan Spoofing Wajah pada Foto," *Jurnal Algoritme*, vol. 5, no. 2, pp. 218–230, Apr. 2025, doi: 10.35957/algoritme.v5i2.10954.

[35] M. N. M. Hakim, A. B. Nugroho, and A. E. Minarno, "Prediksi Tumor Otak Menggunakan Metode Convolutional Neural Network," *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, vol. 17, no. 1, p. 48, Jul. 2023, doi: 10.30872/jim.v17i1.5246.

[36] A. Simarmata, A. Putra, and A. Husein, "Penerapan Metode Computer Vision Dalam Klasifikasi Buah Jeruk Menggunakan Teknik Image Pre-Processing," *Data Sciences Indonesia (DSI)*, vol. 3, no. 2, pp. 108–114, Jun. 2024, doi: 10.47709/dsi.v3i2.4010.

[37] A. Akbar and D. I. Mulyana, "Optimasi Klasifikasi Batik Betawi Menggunakan Data Augmentasi Dengan Metode KNN Dan GLCM," *Jurnal Aplikasi Teknologi Informasi dan Manajemen (JATIM)*, vol. 3, no. 2, pp. 92–101, Nov. 2022, doi: 10.31102/jatim.v3i2.1577.

[38] W. A. Kurniawan and A. Salam, "Penggunaan Feature Space SMOTE Untuk Mengurangi Overfitting Akibat Imbalance Dataset," *Jurnal Transformatika*, vol. 22, no. 2, pp. 140–149, Jan. 2025, doi: 10.26623/transformatika.v22i2.8305.

[39] Y. Amrozi, D. Yuliati, A. Susilo, N. Novianto, and R. Ramadhan, "Klasifikasi Jenis Buah Pisang Berdasarkan Citra Warna dengan Metode SVM," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 11, no. 3, pp. 394–399, Dec. 2022, doi: 10.32736/sisfokom.v11i3.1502.

[40] N. J. Hayati, D. Singasatia, and M. R. Muttaqin, "Object Tracking Menggunakan Algoritma You Only Look Once (YOLO)v8 untuk Menghitung Kendaraan," *Komputa : Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 2, pp. 91–99, Nov. 2023, doi: 10.34010/komputa.v12i2.10654.

[41] R. R. Adhitya, Wina Witanti, and Rezki Yuniarti, "PERBANDINGAN METODE CART DAN NAÏVE BAYES UNTUK KLASIFIKASI CUSTOMER CHURN," *INFOTECH journal*, vol. 9, no. 2, pp. 307–318, Jul. 2023, doi: 10.31949/infotech.v9i2.5641.

[42] J. Hartanto, S. M. Wijaya, Anderies, and A. Chowanda, "Performance Evaluation of EfficientNetB0, EfficientNetV2, and MobileNetV3 for American Sign Language Classification," in *ICEEIE 2023 - International Conference on Electrical, Electronics and Information Engineering*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICEEIE59078.2023.10334648.

[43] D. P. Sidik, F. Utaminingrum, and L. Muflikhah, "Penggunaan Variasi Model pada Arsitektur EfficientNetV2 untuk Prediksi Sel Kanker Serviks," *J-PTIIK*, vol. 7, no. 5, pp. 2116–2121, 2023, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12656

[44] K. Ma, S. Lee, X. Ma, and H. Chen, "Fine art image classification and design methods integrating lightweight deep learning," *Sci Rep*, vol. 15, no. 1, p. 33006, Sep. 2025, doi: 10.1038/s41598-025-18420-0.