Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 9 No. 4 (2025) 737 - 744 e-ISSN: 2580-0760

Comparative Performance of ResNet Architectures for Toraja Carving Image Classification with Data Augmentation

Herman[1*], Muhammad Akbar[2], Haidawati Nasir[3], Herdianti[4], Huzain Azis[5], Lilis Nur Hayati[6]
[1,2,6]Information System, Faculty Computer Science, Universitas Muslim Indonesia, Makassar, Indonesia
[4,5]Informatic, Faculty Computer Science, Universitas Muslim Indonesia, Makassar, Indonesia
[3]Information Technology, MIIT, University Kuala Lumpur, Kuala Lumpur, Malaysia

[1]herman@umi.ac.id, [2]m.akbar@gmail.com, [3]haidawati@unikl.edu.my, [4]herdianti@umi.ac.id, [5]huzain.azis@umi.ac.id

Abstract

The complexity of the motifs and large number of different patterns make the classification of Toraja carvings challenging. The objective of this study is to develop a Convolutional Neural Network automatic classification model using a comparative analysis of the performance of three ResNet architectures. Data augmentation techniques were used to enrich the diversity of the training samples and improve the robustness of the model. The experimental results showed that ResNet101V2 had the highest validation accuracy, which was greater than 97%, followed by ResNet50V2 with more than 96%, and finally, ResNet152V2 with more than 94.74%. These test results indicate that the ResNet101V2 architecture has a better classification performance for complex motifs, with a good balance between precision and recall. However, the confusion matrix and per-class performance metrics indicated that motifs with high similarity, such as Paqdon-Bolu and Paqtedong, remained challenging. This study demonstrated that deeper CNN architectures and data augmentation techniques are effective in improving the classification accuracy of complex carving patterns. Further research should explore hybrid or advanced augmentation methods to improve the overall robustness and accuracy of the model.

Keywords: resnet; classification; toraja carving; data augmentation; cnn

How to Cite: Herman, M. Akbar, H. Nasir, Herdianti, H. Azis, and L. N. Hayati, "Comparative Performance of ResNet Architectures for Toraja Carving Image Classification with Data Augmentation", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 9, no. 4, pp. 737 - 744, Aug. 2025. *Permalink*/DOI: https://doi.org/10.29207/resti.v9i4.6181

Received: November 20, 2024
Accepted: January 26, 2025
Available Online: August 9, 2025

This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia

1. Introduction

Toraja, a region in South Sulawesi Province, Indonesia, is a well-known cultural tourism destination. One of its most prominent cultural heritages is its traditional carvings, which encompass no fewer than 70 motifs with designs and symbolic meanings reflecting the relationship between humans, God, nature, animals, and plants. The complexity and high variability of these motifs make developing image-based classification systems a major challenge in the realm of artificial intelligence and computer vision [1].

In machine learning and pattern recognition, image classification is one of the main focuses, especially in identifying and grouping carved objects or motifs [2]. A fundamental challenge in image classification is addressing the high variability and complexity of visual patterns, as exemplified by traditional Toraja carvings [3]. These carvings, a sophisticated traditional art form from South Sulawesi, Indonesia, feature highly unique and complex motifs [4], [5]. Manually classifying these carvings requires specialized expertise and significant time investment, making automated methods a more efficient alternative [6].

Prior studies indicate that Convolutional Neural Network (CNN) models achieve high accuracy in image classification tasks, including the identification of traditional carvings. CNN architectures such as MobileNet, Inception-v3, VGG16, and VGG19 have been used in the classification of Balinese carvings [7].

Based on the research results, this study uses high-level architectures such as ResNet50, ResNet101, and ResNet152 to solve the problem of classifying Toraja carving motifs [8]-[10]. Deeper architectures allow networks to learn more complex features, which is important for classifying highly complex objects such as Toraja carvings [11], [12].

Given the intricacy of these motifs, conventional CNN models may still struggle to fully capture the fine-grained details. To address this, wavelet decomposition has been employed to enhance CNN performance in other fields, such as medical image classification, offering potential benefits for carving classification as well [13]. Moreover, research indicates that the incorporation of wavelet decomposition in CNN architectures markedly enhances image classification precision for intricate patterns in medical pictures [14]. This approach enables the model to derive distinctive features from images through data augmentation and preprocessing [14]. It can also be utilized for the classification of intricate motifs, such as Toraja sculptures [15].

This research is a continuation, where previous research used CNN. The CNN architecture used is LeNet and the highest accuracy level obtained does not reach 70% [1]. This study focuses on developing an automatic classification model utilizing CNN architecture while evaluating the performance of ResNet architecture in the classification of Toraja carving motifs.

The underlying hypothesis is that deeper CNN architectures will achieve better classification accuracy. This study adds to the limited body of literature on the application of machine learning in traditional carving classification, with a focus on determining the most efficient ResNet architecture for this specific task.

2. Methods

This paper offers a detailed examination of the methodological framework utilized in creating a model based on Convolutional Neural Networks (CNN) for the categorization of Toraja carvings. The document includes a description of the dataset, data augmentation strategies, selected model architectures, and the methodology for evaluating model performance. Figure 1 illustrates the methodology utilized.

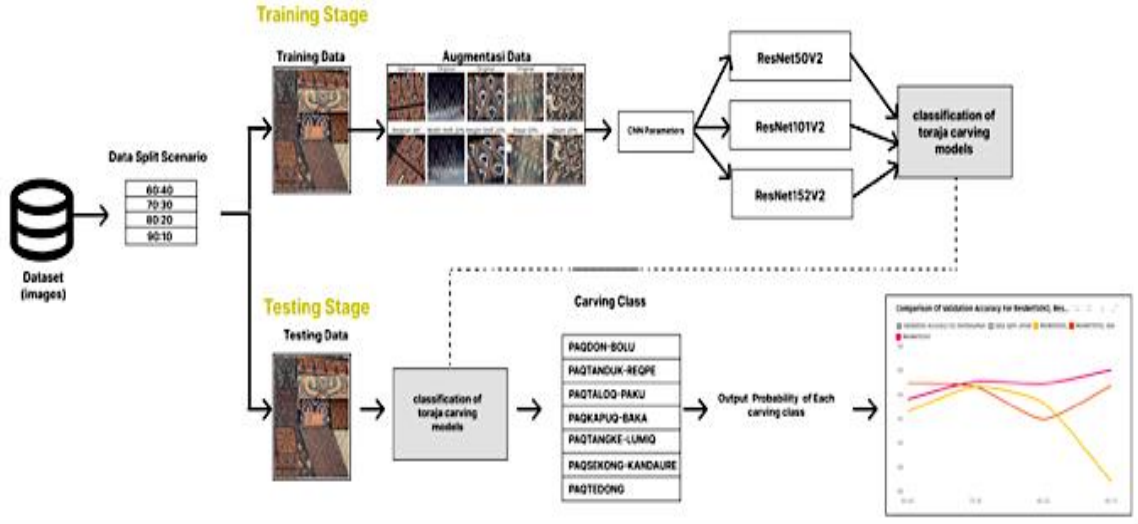


Figure 1. Study Process

Figure 1 illustrates the workflow established for the development and assessment of a CNN-based model aimed at classifying Toraja carvings. The first thing that has to be done is to divide the dataset into subsets that will be used for training and testing. The capability of the model to generalize is improved through the use of data augmentation during the training phase. This is accomplished by increasing the number of training images utilized. Immediately after this, three different ResNet architectures, namely ResNet50V2, ResNet101V2, and ResNet152V2, are trained with the dataset that has been updated. Some of the performance metrics that are evaluated during the testing phase include accuracy, precision, recall, F1-score, and confusion matrix. These metrics are used to evaluate the models. This methodical strategy aims to determine the optimal architecture for precisely categorizing Toraja carving motifs.

2.1 Dataset

This investigation employs a dataset consisting of 700 images of Toraja carvings, obtained from traditional Toraja houses and monuments [16]. The dataset is made up of seven different motif groups, and each group has 100 pictures. Figure 2 shows some pictures from each class that are typical of that class.



Figure 2. Examples of images representing each class of Toraja carving motifs

The dataset was segmented into training and validation sets utilizing four distinct ratios: 90:10, 80:20 70:30, and 60:40 [17]. Table 1 presents the precise allocation of images to the training and test sets for each specified ratio.

Table 1. Distribution of Data for Training and Evaluation

Devide	Training	Test
90:10	630 images	70 images
80:20	560 images	140 images
70:30	490 images	210 images
60:40	420 images	280 images

2.2 Data Augmentation

This study employed various data augmentation techniques to reduce the likelihood of overfitting. These techniques included random rotations of up to ±40 degrees, horizontal and vertical shifts of 20%, shear transformations with a factor of 0.2, and scaling by 20% [18]. The techniques focus on capturing the variety in shapes and orientations of Toraja carvings, which improves the model's capacity to generalize across different angles and image positions.

Shear transformations introduce slight geometric distortions, enhancing robustness to changes in shape, while scaling accounts for size variations in the carvings. To represent the symmetrical characteristics commonly found in various traditional motifs, horizontal image flipping was applied. Subsequently, all images underwent resizing to dimensions of 224×224 pixels and were normalized to a range of 0 to 1, ensuring data consistency throughout the model training phase [19].

2.3. ResNet

This investigation utilized three versions of the ResNet architecture for the creation of the classification model, each incorporating residual blocks and shortcut connections with greater depth to capture more intricate features. Figure 3 illustrates the ResNet50V2 architecture, which consists of 50 layers and employs residual blocks to tackle the vanishing gradient issue commonly encountered in deep neural networks [19]. ResNet101V2, containing 101 layers and capable of learning more intricate visual features, is shown in Figure 4. ResNet152V2, the deepest of the three with 152 layers, incorporates bottleneck layers along with dropout to reduce overfitting, and its structure is presented in Figure 5.

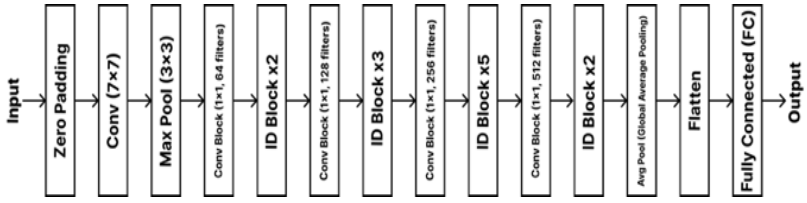


Figure 3. Design of ResNet50V2 Architecture

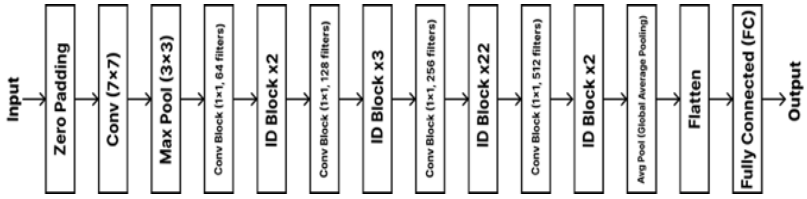


Figure 4. Design of ResNet101V2 Architecture

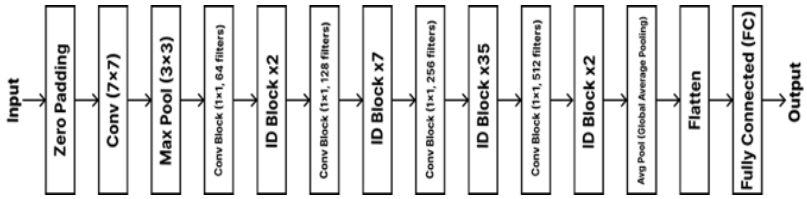


Figure 5. Design of ResNet152V2 Architecture

Throughout the training phase, every input image was resized to dimensions of 224×224 pixels, utilizing a batch size of 32. The use of batch size in this study is based on previous research, which produced the highest level of accuracy compared to other batch sizes [1]. Each model underwent training for a total of 100 epochs, focusing on a classification task that encompassed seven unique classes, each representing specific Toraja carving motifs.

2.4 Assessment of the Model

The evaluation of the CNN models involved three essential metrics: plots of training and validation accuracy, a confusion matrix, and the precision and recall for each class. The metrics offer an in-depth evaluation of the classification capabilities of each model, facilitating thorough comparisons between the different architectures.

Training and Validation Accuracy Visualization: To monitor model performance throughout the training

process, accuracy plots were generated for both training and validation phases of the top-performing models. These visualizations serve to identify potential occurrences of overfitting or underfitting and to evaluate the models' generalization capabilities across training epochs.

The confusion matrix provides a summary of classification performance by displaying both correctly and incorrectly classified instances for each motif class, making it especially valuable in multi-class scenarios. For each top model, the confusion matrix identifies classes that may be frequently misclassified, offering insights into specific classification challenges.

Per-Class Precision and Recall: The evaluation of the model's capability to accurately identify each carving motif involved computing precision and recall for each class. Precision indicates how accurately the model makes positive predictions, whereas recall (or sensitivity) reflects the model's capability to correctly recognize all genuine positive instances, as outlined in Equations 1 and 2.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

In these equations, True Positives (TP) refer to instances that have been accurately identified as positive, while False Positives (FP) denote instances that have been incorrectly classified as positive. Additionally, False Negatives (FN) represent positive instances that have not been identified. The metrics provide a comprehensive perspective on the model's advantages and limitations, emphasizing patterns that might need additional enhancement in classification.

3. Results and Discussions

This study presents its findings by initially summarizing the overall model performance, then proceeding to a detailed analysis of the top-performing model. Precision, recall, and F1-score for every class are evaluated, as are training and validation accuracy, loss values are visualized, and the analysis is comprehensive. Furthermore, confusion matrices are provided to enhance understanding of the model's classification performance across each motif category.

3.1 Performance Comparison Across Different Models

Table 3 presents the mean validation accuracy obtained by each ResNet architecture (ResNet50V2, ResNet101V2, and ResNet152V2) across various data partitioning configurations used in the experiments. This overview helps identify the best-performing architectures and data splits for the Toraja carving classification task.

Table 3. Average Validation Accuracy of ResNet Architectures for Different Data Splits

Architecture	90:10	80:20	70:30	60:40
ResNet152V2	94.74%	91.33%	90.22%	94.31%
ResNet101V2	97.37%	97.33%	96.00%	95.32%
ResNet50V2	93.42%	94.00%	95.56%	96.66%

3.2 Analysis of Top-Performing Models

The highest-performing models from each ResNet architecture were selected for further analysis. The selected models include ResNet50V2 with a training-to-testing data split of 60:40, ResNet101V2 with a 90:10 split, and ResNet152V2 with a 90:10 split. Training and validation accuracy and loss plots are shown in Figures 6, 7, and 8 for each model. These plots emphasize the learning dynamics and generalization capabilities over the training epochs.

With a validation accuracy of 96.66%, ResNet50V2 topped the charts, followed by ResNet101V2 at 97.37%, and ResNet152V2 at 94.74%, according to the visualizations. Each model demonstrates stable training and validation performance over 100 epochs, though slight fluctuations in validation accuracy are observed, especially in ResNet152V2. These findings suggest that deeper network architectures may require further fine-tuning to minimize the risk of mild overfitting.

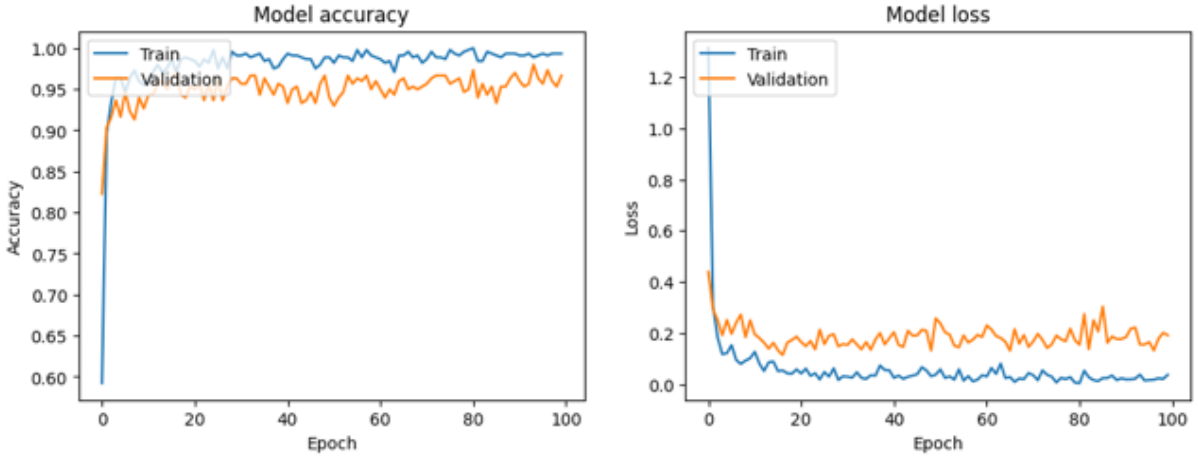


Figure 6. ResNet50V2 Performance Visualization: Training and Validation Accuracy and Loss Trends (60:40 Split)

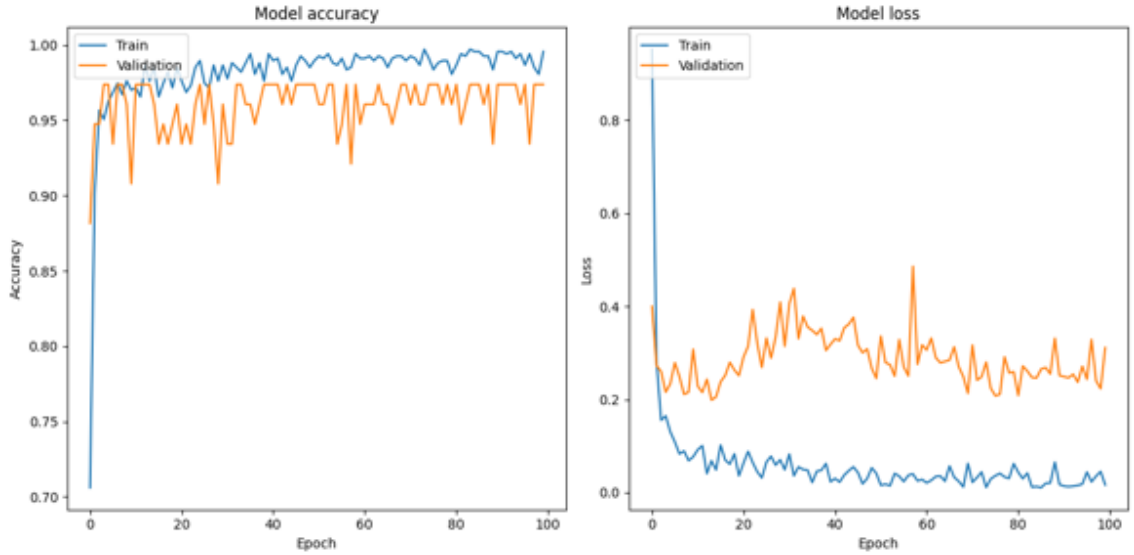


Figure 7. The Visualization of ResNet101V2 Performance: Trends in Loss and Accuracy during Training and Validation (90:10 Split)

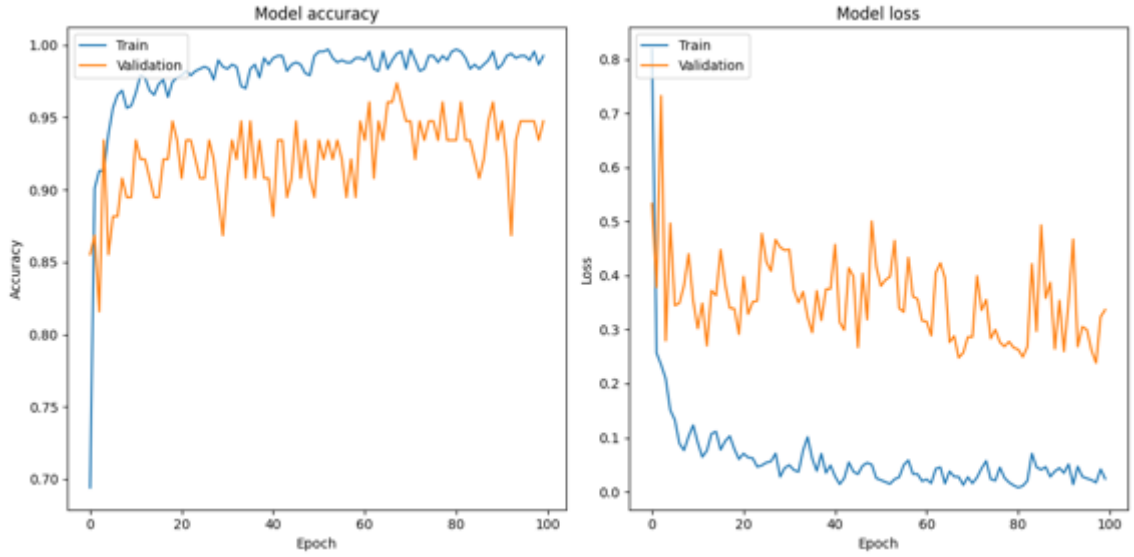


Figure 8. The Visualization of ResNet152V2 Performance: Trends in Loss and Accuracy during Training and Validation (90:10 Split)

3.3 Classification Performance for Each Class

In order to assess classification performance comprehensively, metrics such as per-class precision, recall, and F1-score were computed, as illustrated in Tables 3, 4, and 5, accompanied by confusion matrices for each of the top-performing models. The metrics offer an in-depth insight into the model's performance within each motif class, aiding in the identification of particular motifs that could pose difficulties in accurate classification.

The ResNet50V2 model with a 60:40 data split demonstrates relatively balanced classification performance across all classes, achieving high precision, recall, and F1-scores. Nonetheless, a slight decrease in recall is observed in the Paqdon-Bolu class. These findings suggest that although the model exhibits high accuracy in most classes, there remains a possibility of misclassification for certain motifs.

Table 4. ResNet50V2 Class-Based Recall, Precision, and F1-Score (60:40)

Class	Recall	Precision	F1-Score	Support
Paqkapuq Baka	1.00	0.98	0.99	40
Paqsekong Kandaure	0.95	1.00	0.98	44
Paqtangke Lumuq	0.95	1.00	0.98	42
Paqtoloq Paku	0.97	0.97	0.97	40
Paqdon Bolu	0.94	1.00	0.97	52
Paqtanduk Reqpe	1.00	0.85	0.92	44
Paqtedong	0.95	1.00	0.97	37

ResNet101V2 (90:10) split achieves a perfect score (1.00) for precision, recall, and F1-score in most classes, demonstrating its superior ability to distinguish between different motifs accurately. The *Paqtangke-Lumuq* and *Paqtoloq-Paku* classes exhibit slight variation, achieving F1-scores of 0.91 and 0.90, respectively. This indicates that there are instances of misclassification among motifs that share visual similarities.

Table 5. ResNet101V2 Class-Based Recall, Precision, and F1-Score (90:10)

Class	Recall	Precision	F1-Score	Support
Paqkapuq-Baka	1.00	1.00	1.00	10
Paqsekong-Kandaure	1.00	1.00	1.00	11
Paqtangke-Lumuq	0.91	0.91	0.91	11
Paqtoloq-Paku	0.90	0.90	0.90	10
Paqdon-Bolu	1.00	1.00	1.00	13
Paqtanduk-Reqpe	1.00	1.00	1.00	11
Paqtedong	1.00	1.00	1.00	10

ResNet152V2 (90:10) split demonstrates strong performance, although there is a minor reduction in recall for the Paqsekong-Kandaure and Paqtedong classes. The F1-scores for these classes are slightly lower, indicating that this model, while generally effective, struggles with certain complex patterns.

The confusion matrices in Figures 9, 10, and 11 provide a visual summary of classification errors, showing where the models tend to misclassify motifs with similar features.

Table 5. ResNet152V2 Class-Based Recall, Precision, and F1-Score (90:10)

Class	Recall	Precision	F1-Score	Support
Paqkapuq-Baka	1.00	1.00	1.00	10
Paqsekong-Kandaure	0.82	1.00	0.90	11
Paqtangke-Lumuq	0.91	0.91	0.91	11
Paqtoloq-Paku	0.90	0.90	0.90	10
Paqdon-Bolu	1.00	1.00	1.00	13
Paqtanduk-Reqpe	1.00	1.00	1.00	11
Paqtedong	1.00	0.83	0.91	10

The examination of the confusion matrix reveals that motifs displaying subtle visual similarities, like Paqdon-Bolu and Paqtedong, along with Paqtangke-Lumuq and Paqtoloq-Paku, are prone to a greater incidence of misclassification. This pattern suggests that models may benefit from enhanced feature extraction techniques or additional preprocessing to better differentiate these visually similar motifs.

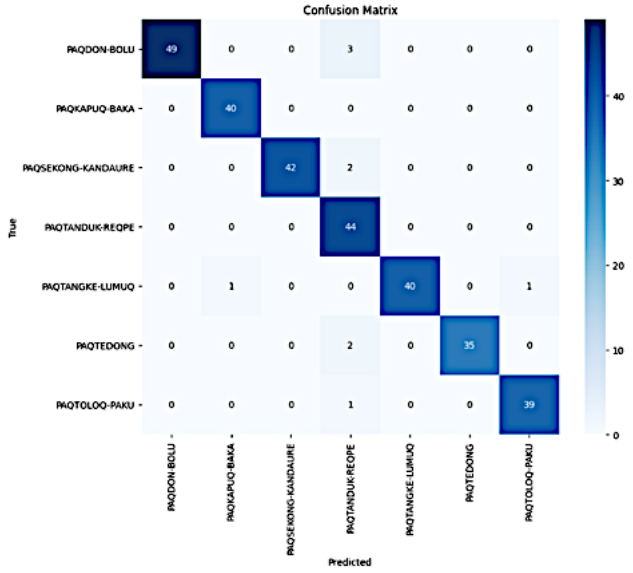


Figure 9. ResNet50V2 Training and Validation Performance Curves (Accuracy and Loss) (90:10)

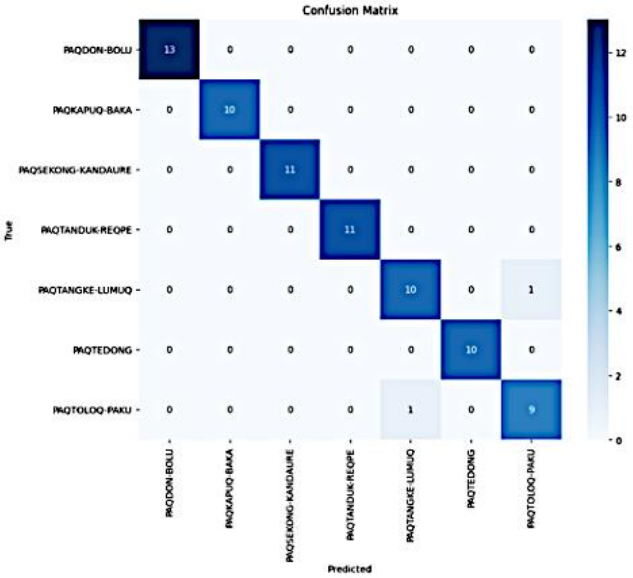


Figure 10. ResNet101V2 Training and Validation Performance Curves (Accuracy and Loss) (90:10)

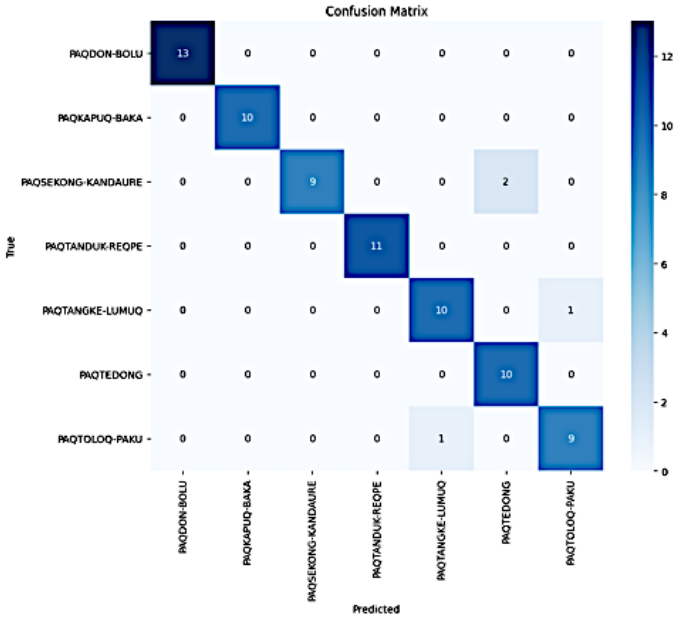


Figure 11. ResNet152V2 Training and Validation Performance Curves (Accuracy and Loss) (90:10)

3.4 Discussion

The results indicate that more intricate network architectures, like ResNet101V2 and ResNet152V2, generally demonstrate enhanced effectiveness in precisely classifying complex Toraja motifs. Specifically, ResNet101V2 exhibits the most favorable balance among precision, recall, and F1-score across all classes, suggesting that a depth of 101 layers may serve as an optimal configuration for extracting complex features while minimizing the risk of excessive overfitting.

This study demonstrates that employing data augmentation techniques can notably improve the model's ability to generalize, as indicated by the minimal difference observed between training and validation accuracy across all assessed architectures. Nevertheless, certain motifs with similar structural characteristics remain a source of misclassification, as reflected in the confusion matrix analysis and per-class metric evaluation. These observations indicate that further optimization may be achieved by exploring additional preprocessing strategies or considering alternative CNN architectures.

Automating the classification of culturally significant motifs, like those seen in Toraja carvings, is a promising area for CNN architectures, particularly the deeper ResNet variants, according to the results of this study. For future development, it is advisable to conduct additional studies to investigate advanced architectural modifications and the implementation of more sophisticated augmentation techniques to enhance classification accuracy.

4. Conclusions

This investigation analyzed the efficacy of three ResNet architectures in the classification of seven unique Toraja carving motifs, emphasizing the assessment of accuracy, precision, recall, and F1-score across different data split ratios. Complicated motifs were classified better by increasingly complicated structures, especially ResNet101V2, which balanced precision and recall. Nonetheless, patterns exhibiting nuanced visual resemblances, like Paqdon-Bolu and Paqtedong, persisted in presenting classification difficulties, as indicated by the confusion matrices and per-class evaluation metrics. The findings suggest that enhancing model performance is possible by implementing more efficient feature extraction methods or incorporating further data augmentation techniques. Generally, deeper ResNet variants within CNN architectures exhibit considerable promise in facilitating the automated classification of culturally significant motifs, like those present in Toraja carvings. Future studies may concentrate on creating hybrid methodologies or applying sophisticated augmentation techniques to improve model resilience and overall classification precision.

References

[1] Herman, H. Nasir, M. N. Megat Mohamed Noor, T. Hasanuddin, D. Indra, and H. B. Lumentut, "Exploration of CNN Parameters to Measure Performance of LeNet-5 Architecture in Toraja Carving Classification," in *2024 IEEE 8th International Conference on Signal and Image Processing Applications, ICSIPA 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICSIPA62061.2024.10686353.

[2] J. Long *et al.*, "A Comprehensive Review of Signal Processing and Machine Learning Technologies for UHF PD Detection and Diagnosis (II): Pattern Recognition

Approaches," *IEEE Access*, vol. 12, pp. 29850–29890, 2024, doi: 10.1109/ACCESS.2024.3368866.

[3] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 11966–11976. doi: 10.1109/CVPR52688.2022.01167.

[4] M. K. Sallata, M. K. Allo, and F. Ansari, "Bamboo forest management practices and local wisdom in Tana Toraja and North Toraja Regency, South Sulawesi," *AIP Conf Proc*, vol. 3001, no. 1, p. 030045, Feb. 2024, doi: 10.1063/5.0184455.

[5] B. V. Remme' and Y. Ba'ru, "GEOMETRIC EXPLORATION IN TORAJA CARVING," *MATEMATIKA DAN PEMBELAJARAN*, vol. 8, no. 2, pp. 122–132, Dec. 2020, doi: 10.33477/mp.v8i2.1594.

[6] X. Liu, H. Wang, Z. Li, and L. Qin, "Deep learning in ECG diagnosis: A review," *Knowl Based Syst*, vol. 227, p. 107187, 2021, doi: https://doi.org/10.1016/j.knosys.2021.107187.

[7] I. W. A. S. Darma, N. Suciati, and D. Siahaan, "Balinese Carving Recognition using Pre-Trained Convolutional Neural Network," in *ICICoS 2020 - Proceeding: 4th International Conference on Informatics and Computational Sciences*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020. doi: 10.1109/ICICoS51170.2020.9299021.

[8] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *Pattern Analysis and Applications*, vol. 24, no. 3, pp. 1207–1220, 2021, doi: 10.1007/s10044-021-00984-y.

[9] S. R. Nayak, D. R. Nayak, U. Sinha, V. Arora, and R. B. Pachori, "Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study," *Biomed Signal Process Control*, vol. 64, p. 102365, 2021, doi: https://doi.org/10.1016/j.bspc.2020.102365.

[10] J. Resti *et al.*, "Classification of Toraja Wood Carving Motif Images Using Convolutional Neural Network (CNN) 'Classification of Toraja Wood Carving Motif Images Using Convolutional Neural Network (CNN),'" *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 4, pp. 486–495, 2026, doi: 10.29207/resti.v8i4.5897.

[11] S. Anwar and N. Barnes, "Densely Residual Laplacian Super-Resolution," *IEEE Trans Pattern Anal Mach Intell*, vol. 44, no. 3, pp. 1192–1204, 2022, doi: 10.1109/TPAMI.2020.3021088.

[12] T. Lawrence, L. Zhang, C. P. Lim, and E.-J. Phillips, "Particle Swarm Optimization for Automatically Evolving Convolutional Neural Networks for Image Classification," *IEEE Access*, vol. 9, pp. 14369–14386, 2021, doi: 10.1109/ACCESS.2021.3052489.

[13] O. N. Oyelade and A. E. Ezugwu, "A novel wavelet decomposition and transformation convolutional neural network with data augmentation for breast cancer detection using digital mammogram," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-09905-3.

[14] D. U. Sarkar and T. Prakash, "A convolutional neural network framework to design power system stabilizer for damping oscillations in multi-machine power system," *Neural Comput Appl*, vol. 36, no. 9, pp. 5059–5075, 2024, doi: 10.1007/s00521-023-09323-0.

[15] L. Malacaria, R. Bruno, G. A. Corrente, D. Armentano, E. Furia, and A. Beneduci, "Experimental insights on the coordination modes of coumarin-3-carboxilic acid towards Cr(III)-, Co(II)-, Ni(II)-, Cu(II)- and Zn(II): A detailed potentiometric and spectroscopic investigation in aqueous media," *J Mol Liq*, vol. 346, p. 118302, 2022, doi: https://doi.org/10.1016/j.molliq.2021.118302.

[16] Rubianus, "Identification of the circle concept in Toraja traditional house carving," *AIP Conf Proc*, vol. 2736, no. 1, p. 110007, Sep. 2023, doi: 10.1063/5.0173924.

[17] H. Miawarni, T. A. Sardjono, E. Setijadi, A. B. Gumelar, Wijayanti, and M. H. Purnomo, "Enhancing Classification of Elderly Fall Detection System using Tuned RBF-SVM," in *2022 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2022, pp. 1–5. doi: 10.1109/IST55454.2022.9827716.

[18] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, "Medical image segmentation using deep learning: A survey," *IET Image Process*, vol. 16, no. 5, pp. 1243–1267, Apr. 2022, doi: 10.1049/ipr2.12419.

[19] T. Rahman and M. S. Islam, "MRI brain tumor detection and classification using parallel deep convolutional neural networks," *Measurement: Sensors*, vol. 26, p. 100694, 2023, doi: https://doi.org/10.1016/j.measen.2023.100694.