P-ISSN : 2715-2448 | E-ISSN : 2715-7199
Vol.7 No.1 January 2026
Buana Information Technology and Computer Sciences (BIT and CS)

Classification of Tulungagung Batik Images in Comparison of Convolution Neural Network and Vision Transformer Algorithms

M. Augustiar Mahendra[1]*, Firman Nurdiyansyah[2], Fitri Marisa[3]
[1,2,3]Widya Gama University of Malang, Malang Jl. Taman Borobudur Indah No.35, Telp. 0341-492282, Kec.Lowokwaru, Kota Malang, Jawa Timur 65128
E-mail: augustiar290821@gmail.com[1]*, firmannurdiyansyah@widyagama.ac.id[2]
, fitrimarisa@widyagama.ac.id[3]

Received: *2025-07-27* | ***Revised:*** *2026-01-21* | ***Accepted:*** *2026-01-29*

Abstract

Batik is a significant Indonesian cultural heritage with a vast diversity of motifs, making manual classification a challenging task. This research provides a comparative analysis of two prominent deep learning architectures, the Convolutional Neural Network (CNN), represented by VGG16, and the Vision Transformer (ViT), represented by DeiT, for the classification of Tulungagung batik images. A balanced dataset of 2,400 images, comprising two classes (Bangoan and Majanan), was utilized. The experiment was conducted using three distinct training-to-testing split ratios (80:20, 70:30, and 60:40) to evaluate model robustness. Performance was assessed using accuracy, precision, recall, F1-score, and the confusion matrix. The results indicate that the CNN (VGG16) model consistently outperformed the ViT (DeiT), achieving its peak accuracy of 96% on both the 80:20 and 60:40 split ratios, showcasing high stability. The ViT (DeiT) model was more sensitive to the data split, reaching a peak accuracy of 94% with less consistent performance. We conclude that for this specific classification task, the VGG16 architecture is more robust, stable, and effective than the DeiT architecture.

Keywords: *Batik Tulungagung*, CNN, VGG16, Vision Transformer, DeiT

I. Introduction

Batik, as an intangible cultural heritage of Indonesia recognized by UNESCO, has an extraordinary diversity of motifs, each of which embodies the philosophy and identity of its region of origin [1]. Among this wealth, batik from Tulungagung, East Java, has its own distinctive characteristics, such as the Bangoan and Majanan motifs. However, the high visual similarity between motifs often poses a challenge in the identification process, particularly for the general public or even collectors, which can lead to confusion. To address this issue effectively, deep learning-based image classification technology offers a highly promising solution.

Image classification is the process of analyzing digital images that enables computers to identify objects within them [2]. Two leading architectures in this field are Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). CNNs, represented by models such as VGG16, have proven to be highly reliable in extracting visual features hierarchically, from edges to complex shapes [3], [4]. Meanwhile, ViT, represented by DeiT, is a more modern architecture that uses self-attention mechanisms and demonstrates high data efficiency [5], [6].

Previous studies have applied deep learning methods to batik classification, mostly focusing on CNN architectures [7], [8], [9]. However, there is still a research gap in direct comparisons between established CNN architectures and newer Transformer architectures on specific and controlled batik datasets. Many studies have not evaluated which model is more robust and stable when faced with variations in the amount of training data.

This study aims to answer the question: which of the proven reliable classical architecture (VGG16) and efficient modern architecture (DeiT) is more suitable for the specific task of batik image classification This study conducts a direct comparative analysis of the two models on the Tulungagung batik dataset to evaluate their accuracy, stability, and effectiveness [10], [11].

II. Methods

This chapter describes in detail the systematic steps taken in the research to perform a comparative analysis between the CNN (VGG16) and ViT (DeiT) architectures. This methodology is designed to ensure that the evaluation and comparison of the two models are conducted objectively, validly, and reproducibly [12], [13]. The discussion will cover several key stages, starting with a description and preparation of the dataset used, followed by a detailed explanation of the model architecture, the experimental scenario design including three data splitting ratios, and the definition of evaluation metrics that serve as performance benchmarks.

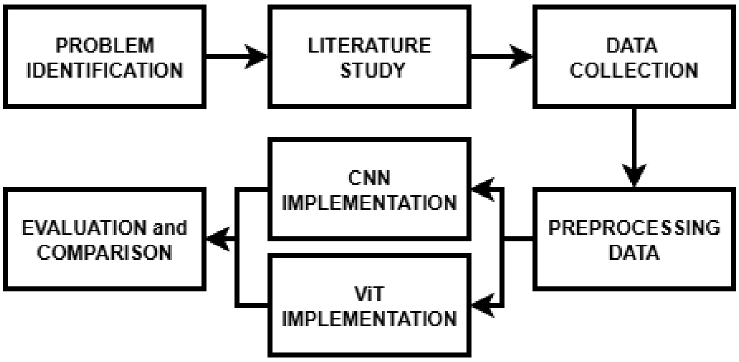

Figure 1. Research Stages

1. Problem Identification

This first stage forms the foundation of the entire study. Here, researchers formulate the main problem to be solved, namely the difficulty in accurately identifying and distinguishing the distinctive batik motifs of Tulungagung (Bangoan and Majanan), especially for the general public. The objective of this stage is to establish the urgency and scope of the study, namely the need for a reliable automatic classification system to assist in cultural preservation and education.

2. Literature Study

Once the main problem has been identified, the next step is to conduct a literature study to review relevant previous research. This stage involves exploring various image classification methods, particularly the application of deep learning architectures such as CNN (VGG16) and Vision Transformer (DeiT) in similar cases. The aim is to understand existing approaches and identify strengths, weaknesses, and gaps in research that can be filled [14].

3. Data Collection

At this stage, the dataset used to train and test the model was collected. For this study, the process involved collecting 2,400 images of Tulungagung batik consisting of two classes, namely Bangoan and Majanan can be seen in figure 2. The data was collected from various digital collections to ensure sufficient diversity [15].

Figure 2. Batik Majanan and Batik Bangoan

4. Data preprocessing

The raw data that has been collected then goes through a preprocessing stage to ensure its quality and readiness before being fed into the model. This process includes cleaning the data of irrelevant images, resizing all images to 224x224 pixels for uniformity, and applying data augmentation techniques (such as rotation, flipping, and zooming) to enrich the dataset's variety. After that, the prepared dataset is then divided into training data and test data using three different ratio scenarios, namely 80:20, 70:30, and 60:40, to test the robustness of the model in various conditions.

5. Implementation of CNN (VGG16) and ViT (DeiT)

This stage is the core of the experiment, where the two architectures to be compared are implemented. The CNN model (VGG16) and the Vision Transformer model (DeiT) are built and trained separately using the prepared dataset. To test the models' robustness and ensure a fair comparison, training is conducted across three distinct data split scenarios: 80:20, 70:30, and 60:40 (training:testing).

CNN (VGG16) uses the standard VGG16 architecture consisting of 13 convolution layers with 3x3 filters and 3 fully-connected layers. This model was chosen because of its simple structure, yet it has proven to be very reliable for image classification and transfer learning tasks.

Vision Transformer (DeiT) uses the DeiT architecture. This model was chosen because it is a ViT variant designed to be efficiently trained on datasets that are not too large (such as ImageNet-1K) through distillation techniques, making it relevant for the dataset in this study.

6. Evaluation

An evaluation was conducted to comprehensively assess and compare the performance of the CNN (VGG16) and ViT (DeiT) models in classifying two types of Tulungagung batik, namely Bangoan and Majanan. The test results were analyzed using the metrics of Accuracy, Precision, Recall, F1-Score, and Confusion Matrix. These metrics provide deep insights into the accuracy level, model stability across various data splitting scenarios, and its ability to recognize each batik class fairly.

The performance of the CNN (VGG16) and ViT (DeiT) models was comparatively evaluated using several key metrics, including accuracy, precision, recall, F1-score, and the confusion matrix, to assess their effectiveness in classifying the two types of Tulungagung batik Bangoan and Majanan.

a. Accuracy is calculated using equation :

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \times 100\% \qquad (1)$$

Accuracy measures the percentage of correct predictions out of the total predictions made, by comparing the number of true positives (TP) and true negatives (TN) against the total number of samples.

b. Precision is computed using equation:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \times 100\% \tag{2}$$

Precision evaluates how accurate the positive predictions are, by comparing the number of correctly predicted positives (TP) to all predicted positives (TP + FP).

c. Recall is calculated using equation:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \times 100\% \tag{3}$$

Recall measures the model's ability to identify all relevant positive samples, by comparing true positives (TP) to the total actual positives (TP + FN).

d. F1-Score is calculated using equation:

$$F1 = 2 \times \frac{precision.recall}{precision+recall} \tag{4}$$

The F1-Score is the harmonic mean of Precision and Recall, providing a balanced evaluation of the model's performance, especially in cases of imbalanced datasets.

The evaluation results were then visualized using bar charts to directly compare the scores of each metric between VGG16 and DeiT, as well as a Confusion Matrix to reveal prediction patterns and the types of errors made by each model. These visual tools are highly effective in identifying which model is superior, understanding where classification errors frequently occur, and serving as a strong foundation for selecting the best model and improving upon it in future research.

III. Results and Discussions

This study developed a system to compare two advanced deep learning architectures for batik classification: the Convolutional Neural Network (CNN), represented by VGG16, and the Vision Transformer (ViT), represented by DeiT. Unlike traditional methods that rely on pre-defined features, these models perform automatic feature extraction. The VGG16 architecture learns hierarchical visual features through its stacked convolutional layers, identifying patterns from simple edges to complex motifs. In contrast, the DeiT architecture utilizes a self-attention mechanism to learn global contextual relationships between different patches of an image.

All images in the dataset were classified into two categories, namely Bangoan and Majanan, and were subsequently resized to 224x224 pixels to meet the input requirements of both models. Furthermore, data augmentation techniques, including rotation and flipping, were applied to enhance the dataset's diversity and improve model generalization. The prepared images were then fed directly into the VGG16 and DeiT models, which internally process the pixel data to learn the distinguishing features for classification.

Figure 3. Augmentation Data

Figure 3. illustrates the data augmentation process, a key step in preparing the image dataset for model training. The process begins with the original batik image. To enrich the dataset and improve the

model's ability to generalize, several augmentation techniques are applied, including horizontal and vertical mirroring. Horizontal flip creates a new image that is a mirror reflection of the original along the vertical axis, effectively simulating a view of the batik from the opposite side. Similarly, vertical flip creates an upside-down version of the image. By applying these transformations, the model is trained on a wider variety of orientations, helping it to learn the core features of each batik motif regardless of its position. This technique significantly increases the effective size of the dataset without requiring new images, thereby reducing the risk of overfitting. To evaluate the models' robustness under various conditions, the total dataset of 2,400 images was divided into training and testing sets using three distinct scenarios. These scenarios, detailed in Table 1, correspond to train-to-test split ratios of 80:20, 70:30, and 60:40. This multi-ratio approach ensures a comprehensive performance comparison.

Table 1. Dataset Distribution Per Ratio

No.	Rasio	Training Sample (per Class)	Testing Sample (per Class)	Total Training Data	Total Testing Data
1.	80:20	960	240	1920	480
2.	70:30	840	360	1680	720
3.	60:40	720	480	1440	960

The implementation of these three distinct scenarios, as detailed in Table 1, is a crucial part of the experimental methodology. This approach was designed to thoroughly test the robustness and generalization capability of each model under different conditions. By evaluating performance across ratios with varying amounts of training data, this study provides a more comprehensive comparison, revealing how each architecture CNN (VGG16) and ViT (DeiT) adapts to data availability. The subsequent sections will describe the specific model architectures and the metrics used for their evaluation.

1. CNN VGG16 Perfomance Analysis

The Convolutional Neural Network (CNN) model used in this study is based on the VGG16 architecture, which is well-regarded for its simplicity and effectiveness in image classification tasks. Leveraging transfer learning, the model was fine-tuned to classify images of Batik Bangoan and Batik Majanan using a balanced dataset. The performance of VGG16 was evaluated under three different data splitting scenarios 80:20, 70:30, and 60:40 where each configuration provides a distinct distribution of training and testing data. This section provides a detailed analysis of the model's performance metrics, including accuracy, precision, recall, F1-score, and confusion matrix results. The objective is to assess the robustness, consistency, and generalization capability of the VGG16 model across varying data availability conditions.

1.1. Split Ratio 80:20

The CNN (VGG16) implementation with an 80:20 split ratio demonstrates exceptionally high and stable performance. During development, a validation set is used to assess the model during training and prevent overfitting, while the final model is assessed using a test set of previously unseen data. This separation is crucial for an objective evaluation of the model's ability to generalize.

The final evaluation on the test set yielded an outstanding accuracy of 0.96, meaning that 96% of the test data was correctly classified. The detailed performance results for this evaluation are shown in Table 2.

Table 2. Evaluation Of The Results CNN VGG16 Split Ratio 80:20

No.	Class	Precision	Recall	F1-Score	Accuracy
1.	Bangoan	0.94	0.98	0.96	0.96
2.	Majanan	0.98	0.94	0.96	

The 80:20 test set evaluation shows the CNN (VGG16) model performs exceptionally well in classifying the two types of Tulungagung batik. For the Bangoan class, the model achieved a Precision of 0.94 and an outstanding Recall of 0.98, resulting in a very high F1-score of 0.96. This indicates the model is extremely effective at identifying almost all Bangoan images. Similarly, for the Majanan class, the performance was also superior, with an excellent Precision of 0.98 and a strong Recall of 0.94, leading to an identical F1-score of 0.96. Overall, the model achieves an outstanding accuracy of 0.96. The macro and weighted averages are also uniformly 0.96 for Precision, Recall, and F1-score, indicating a highly accurate, stable, and well-balanced performance across both classes.

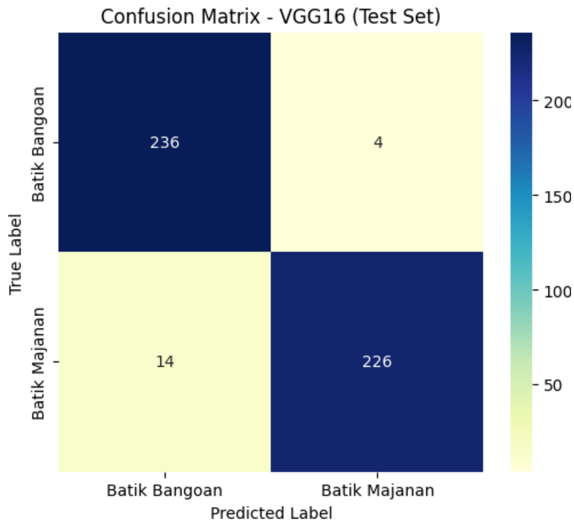

Figure 4. Confusion Matrix CNN VGG16 Split Ratio 80:20

A confusion matrix was used to visualize the performance of the CNN (VGG16) model on the 80:20 data split, which involved a test set of 480 images (240 for Batik Bangoan and 240 for Batik Majanan). The results demonstrated a highly accurate and consistent classification performance. For the Batik Bangoan class, the model performed nearly perfectly, correctly identifying 236 out of 240 images (True Positives) with only 4 misclassifications. Its performance on the Batik Majanan class was also solid, with 226 correct classifications (True Negatives) and 14 incorrect ones. With a total of only 18 misclassifications, the model achieved an overall accuracy of 0.96, proving that the VGG16 architecture is highly reliable and generalizes well for this specific batik classification task.

1.2. Split Ratio 70:30

The CNN (VGG16) application with a 70:30 split ratio demonstrates excellent performance. The test set evaluation results can be seen in Table 3.

Table 3. Evaluation Of The Results CNN VGG16 Split Ratio 70:30

No.	Class	Precision	Recall	F1-Score	Accuracy
1.	Bangoan	0.94	0.95	0.95	0.95
2.	Majanan	0.95	0.95	0.95	

The 70:30 test set evaluation shows the CNN (VGG16) model performs with impressive and exceptionally balanced results in classifying the two types of Tulungagung batik. For the Bangoan class, the model achieved a Precision of 0.94, a Recall of 0.95, and an F1-score of 0.95. Remarkably, the performance for the Majanan class was nearly identical, with a Precision of 0.95, a Recall of 0.95, and an F1-score of 0.95, demonstrating a perfectly symmetrical and unbiased performance. Overall, the model achieves a high accuracy of 0.95. The macro and weighted averages are also uniformly 0.95 for Precision, Recall, and F1-score, indicating a highly consistent and exceptionally well-balanced performance across both classes.

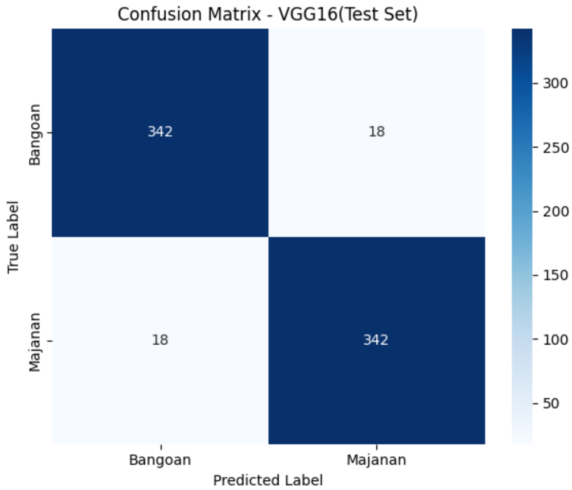

Figure 5. Confusion Matrix CNN VGG16 Split Ratio 70:30

The confusion matrix in Figure 5 demonstrates the model's highly balanced and accurate classification on the 70:30 split ratio, though a small number of misclassifications occurred. For the Bangoan class, out of 360 test images, 342 were correctly classified as Bangoan (True Positives), while only 18 were misclassified as Majanan (False Negatives). Remarkably, an identical pattern of misclassification occurred for the Majanan class, where 342 of 360 images were correctly identified (True Negatives) and exactly 18 were misclassified as Bangoan (False Positives). This perfect symmetry in errors underscores the model's excellent and unbiased overall performance.

1.3. Split Ratio 60:40

VGG16 with a 60:40 split ratio demonstrates outstanding performance, with a final test set accuracy of 96%.

Table 4. Evaluation Of The Results CNN VGG16 Split Ratio 60:40

No.	Class	Precision	Recall	F1-Score	Accuracy
1.	Bangoan	0.95	0.98	0.97	0.96
2.	Majanan	0.98	0.95	0.96	

The evaluation of the test set, as shown in Table 4, reveals the model's outstanding and robust performance. For the Bangoan class, the model achieved a Precision of 0.95, a Recall of 0.98, and an F1-score of 0.97. For the Majanan class, performance was equally impressive, with a nearly perfect Precision of 0.98, a Recall of 0.95, and an F1-score of 0.96. Overall, the model attained a remarkable accuracy of 0.96, with Macro and Weighted averages for Precision at 0.97, and for Recall and F1-score at 0.96, indicating consistently high performance across all classes regardless of sample distribution.

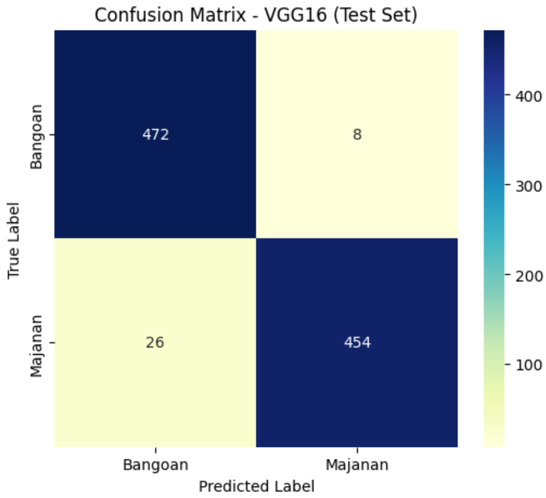

Figure 6. Confusion Matrix CNN VGG16 Split Ratio 60:40

The confusion matrix in Figure 6. highlights the model's robust and highly accurate classification performance on the 60:40 split ratio. For the Bangoan class, 472 out of 480 images were correctly classified, with only 8 images being misclassified as Majanan. For the Majanan class, 454 images were correctly classified, while 26 were misclassified as Bangoan. Although some misclassifications occurred, the overall number of errors is very low, confirming the model's excellent classification capability even on the largest test set.

2. ViT DeiT Perfomance Analysis

The Vision Transformer (ViT) model used in this study is based on the DeiT (Data-efficient Image Transformer) architecture, a modern approach that processes images by analyzing relationships between image patches. This model was applied to classify images of Batik Bangoan and Batik Majanan using a balanced dataset. The performance of DeiT was evaluated under three different data splitting scenarios 80:20, 70:30, and 60:40 where each configuration provides a distinct distribution of training and testing data. This section provides a detailed analysis of the model's performance metrics, including accuracy, precision, recall, F1-score, and confusion matrix results. The objective is to assess the robustness, consistency, and generalization capability of the DeiT model across varying data availability conditions.

2.1. Split Ratio 80:20

The application of DeiT with an 80:20 split ratio shows promising performance. After training and validation, the model was evaluated on the test set, achieving an overall accuracy of 90%.

Table 5. Evaluation Of The Results ViT DeiT Split Ratio 80:20

No.	Class	Precision	Recall	F1-Score	Accuracy
1.	Bangoan	0.85	0.97	0.91	0.90
2.	Majanan	0.97	0.83	0.89	

The evaluation on the test set indicates the model's capability in classifying the two batik types (Bangoan and Majanan), though with notable differences between the classes. For the Bangoan class, the model demonstrated a very high Recall of 0.97 but a lower Precision of 0.85, resulting in an F1-score of 0.91. Conversely, for the Majanan class, it achieved an excellent Precision of 0.97 but a lower Recall of 0.83, yielding an F1-score of 0.89.

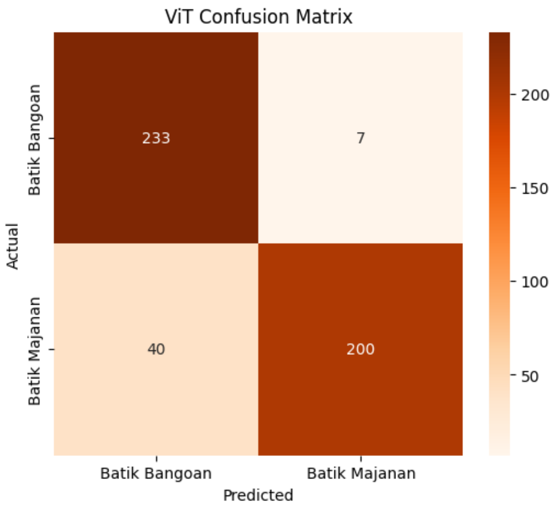

Figure 7. Confusion Matrix ViT DeiT Split Ratio 80:20

The confusion matrix in Figure 7 reveals an imbalanced performance for the DeiT model on the 80:20 split ratio. For the Batik Bangoan class, the model performed exceptionally well, correctly classifying 233 out of 240 images and misclassifying only 7 as Majanan. In stark contrast, the performance on the Batik Majanan class was significantly weaker, with 200 images correctly identified while 40 were misclassified as Bangoan. This significant disparity in errors indicates a predictive bias towards the Bangoan class, highlighting a key area for improvement despite the model's overall capability.

2.2. Split Ratio 70:30

The application of DeiT with a 70:30 split ratio shows a significantly improved and optimal performance. After training and validation, the model was evaluated on the test set, achieving a high overall accuracy of 94%.

Table 6. Evaluation Of The Results Vit DeiT Split Ratio 70:30

No.	Class	Precision	Recall	F1-Score	Accuracy
1.	Bangoan	0.93	0.97	0.95	0.94
2.	Majanan	0.97	0.92	0.94	

The evaluation on the test set indicates the model's strong and well-balanced capability in classifying the two batik types (Bangoan and Majanan). For the Bangoan class, the model demonstrated a high Recall of 0.97 and a strong Precision of 0.93, resulting in an excellent F1-score of 0.95. Similarly, for the Majanan class, it achieved an outstanding Precision of 0.97 and a high Recall of 0.92, yielding a stable F1-score of 0.94.

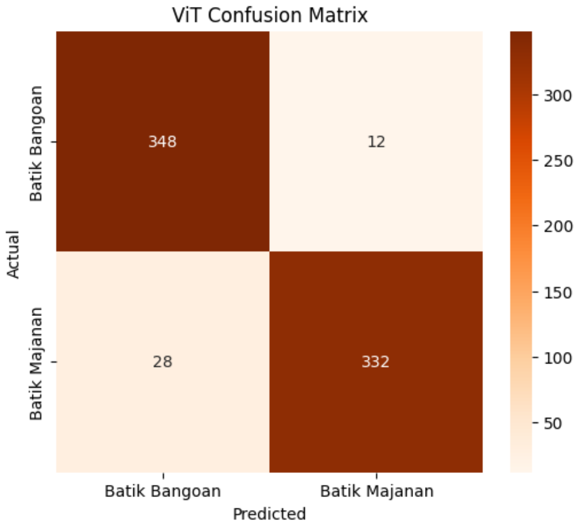

Figure 8. Confusion Matrix ViT DeiT Split Ratio 70:30

The confusion matrix in Figure 8. highlights the model significantly improved and more balanced classification performance on the 70:30 split ratio. For the Batik Bangoan class, 348 out of 360 images were correctly classified, with only 12 images being misclassified as Majanan. For the Batik Majanan class, 332 images were correctly classified, while 28 were misclassified as Bangoan. Although some misclassifications still occurred, the error distribution is far more balanced, confirming the model's enhanced and more stable classification capability in this scenario.

2.3. Split Ratio 60:40

The application of DeiT with a 60:40 split ratio demonstrates robust performance, maintaining its peak accuracy even with a smaller training set. After training and validation, the model was evaluated on the test set, achieving an overall accuracy of 94%.

Table 7. Evaluation Of The Results Vit DeiT Split Ratio 60:40

No.	Class	Precision	Recall	F1-Score	Accuracy
1.	Bangoan	0.91	0.98	0.94	0.94
2.	Majanan	0.98	0.90	0.94	

The evaluation on the test set indicates the model's continued capability in classifying the two batik types (Bangoan and Majanan), though with a slight return to an imbalanced profile. For the Bangoan class, the model demonstrated an excellent Recall of 0.98 but a lower Precision of 0.91, resulting in an F1-score of 0.94. Conversely, for the Majanan class, it achieved a nearly perfect Precision of 0.98 but a lower Recall of 0.90, yielding an F1-score of 0.94.

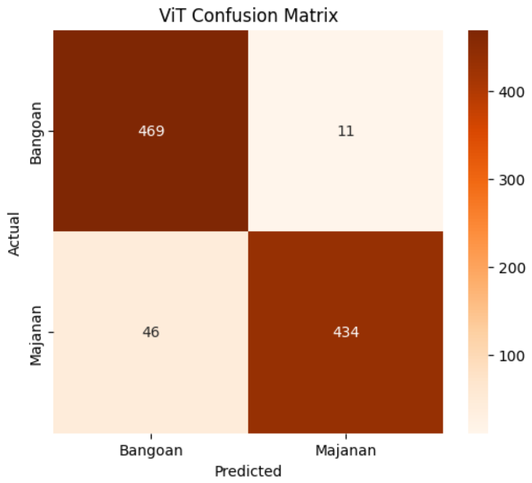

Figure 9. Confusion Matrix ViT DeiT Split Ratio 60:40

The confusion matrix for the 60:40 split ratio reveals the model's robust but imbalanced performance on the largest test set. For the Bangoan class, the model performed exceptionally well, correctly classifying 469 out of 480 images, with only 11 images being misclassified as Majanan. In contrast, for the Majanan class, 434 images were correctly classified, while a higher number, 46, were misclassified as Bangoan. Although the overall error rate remains low relative to the test set size, this disparity confirms the model's predictive bias.

3. Evaluation

The evaluation results of the CNN VGG16 and ViT DeiT model show excellent performance in classifying batik Tulungagung Figure 10.

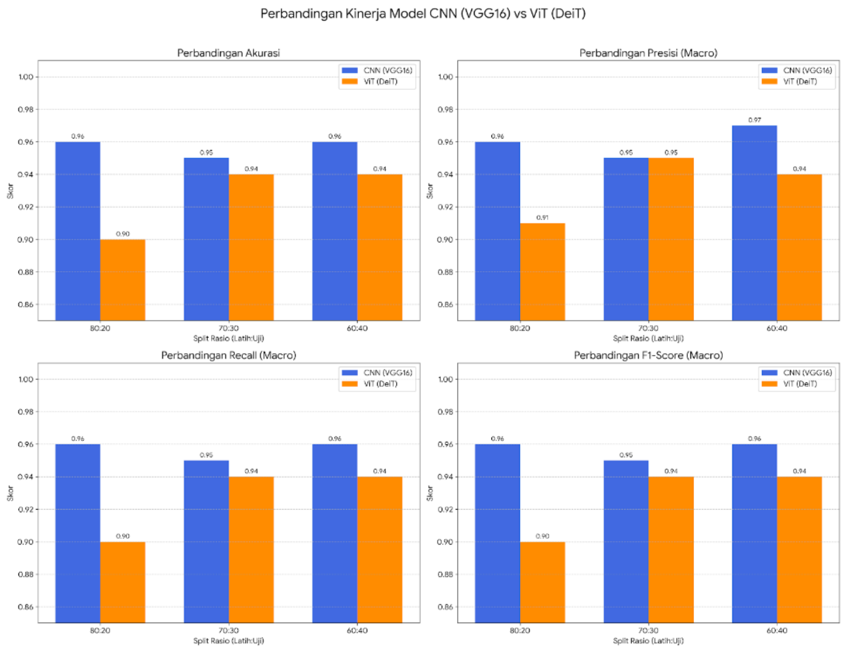

Figure 10. Results graph for each ratio

Based on the comparison graph in Figure 10, a visual analysis confirms a significant performance difference between the architectures, influenced by the data split ratio. In terms of Accuracy, the CNN (VGG16) architecture demonstrates remarkable stability and superiority, achieving a peak performance of 96% in two different scenarios: the 80:20 and 60:40 ratios. In contrast, the ViT (DeiT) model shows greater sensitivity to the data distribution; its performance fluctuates from a low of 90% at the 80:20 ratio, then rises to its peak of 94%.

This trend of superiority extends to the Precision and Recall metrics. Both models were most competitive at the 70:30 ratio, where they both scored a precision of 0.95. However, CNN (VGG16) proved its advantage by achieving the highest precision of 0.97 at the 60:40 ratio, indicating its positive predictions are highly reliable. Similarly, for Recall, VGG16 consistently maintained a high score of 0.96, showcasing its robust ability to identify nearly all relevant samples in each class.

The F1-Score, which harmonizes Precision and Recall to measure balanced performance, reinforces the existing conclusion. The CNN (VGG16) model achieved the best balance with a peak F1-Score of 0.96, proving its ability to effectively balance between minimizing false predictions and maximizing detection. While the ViT (DeiT) model showed good performance with a peak F1-Score of 0.94, it consistently remained below the level of effectiveness achieved by VGG16.

Overall, it can be concluded that the 80:20 and 60:40 ratios were the optimal scenarios for the CNN (VGG16) model, both yielding the highest accuracy of 96%. The ability of the CNN architecture to maintain peak performance even with a smaller amount of training data confirms that its performance is highly robust. In contrast, the ViT's performance was more influenced by the data distribution, confirming that for this batik classification task, the CNN architecture is more effective and reliable.

IV. CONCLUSION

This study conducted a comparative analysis of CNN (VGG16) and ViT (DeiT) architectures for classifying Tulungagung batik. The findings conclusively demonstrate that the CNN (VGG16) architecture is more robust, stable, and accurate for this specific task. It achieved a higher peak accuracy of 96% compared to DeiT's 94% and, crucially, maintained its high performance across all data split ratios, showcasing excellent generalization. The ViT (DeiT) model, while effective, proved to be more sensitive to the data distribution. The superiority of VGG16 suggests that its architectural design, which excels at learning local and textural patterns, is inherently better suited for the classification of complex, artistic motifs like batik on a moderately sized dataset.

V. ACKNOWLEDGMENT

I would like to express his deepest praise and gratitude to God Almighty for the blessings and grace that made the completion of this thesis possible. Sincere gratitude is extended to the research supervisors, Mr. Firman Nurdiyansyah, S.Kom., M.M., as the first supervisor, and Mrs. Fitri Marisa, S.Kom., M.Pd., Ph.D, as the second supervisor, for their invaluable guidance, direction, patience, and constructive feedback throughout this study. Finally, the author wishes to thank his beloved family for their unwavering moral and material support, as well as his friends and colleagues for their encouragement. The author acknowledges that this thesis is far from perfect and hopes that it will contribute positively to the field of study.

REFERENCES

[1] H. R. Fidiastuti, A. S. Lathifah, M. Amin, and Y. Utomo, "Isolasi dan Karakterisasi Bakteri Indigen Pengurai Lemak pada Limbah Cair Batik Tulungagung," *Bioeksperimen: Jurnal Penelitian Biologi*, vol. 6, no. 1, pp. 29–35, Mar. 2020, doi: 10.23917/bioeksperimen.v6i1.10430.

[2] Ò. Lorente, I. Riera, and A. Rana, "Image Classification with Classic and Deep Learning Techniques," May 2021.

[3] D. Achmad, O. V. Putra, and D. Muriyatmoko, "Image classification of leaf disease in corn plants (Zea Mays L.) using the MobileNetV2 method," *Journal of Information System and Application Development*, vol. 2, no. 2, pp. 122–132, Sep. 2024, doi: 10.26905/jisad.v2i2.14004.

[4] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Apr. 2015, [Online]. Available: http://arxiv.org/abs/1409.1556

[5] A. Pangestu, B. Purnama, and R. Risnandar, “Vision Transformer untuk Klasifikasi Kematangan Pisang,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 75–84, Feb. 2024, doi: 10.25126/jtiik.20241117389.

[6] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, “Training data-efficient image transformers & distillation through attention,” Dec. 2020.

[7] 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.

[8] Meranggi, D. G. T., Yudistira, N., & Sari, Y. A. (2022). Batik classification using convolutional neural network with data improvements. JOIV: International Journal on Informatics Visualization, 6(1), 6-11.

[9] Rangkuti, A. H., Harjoko, A., & Putra, A. (2021). A novel reliable approach for image batik classification that invariant with scale and rotation using MU2ECS-LBP algorithm. Procedia Computer Science, 179, 863-870.

[10] Winarno, E., Solichan, A., Ramdani, A. P., Hadikurniawati, W., Septiarini, A., & Hamdani, H. (2025). Enhanced Semarang batik classification using deep learning: a comparative study of CNN architectures. *Bulletin of Electrical Engineering and Informatics*, *14*(5), 3544-3557.

[11] Elvitaria, L., Ahmad, E. F., Samsudin, N. A., Khalid, S. K. A., & Indra, Z. (2025). An Improved Okta-Net Convolutional Neural Network Framework for Automatic Batik Image Classification. JOIV: International Journal on Informatics Visualization, 9(1), 115-119.

[12] Satoto, B. D., Agustiono, W., Ramadhani, N. P., Rafelina, F. L., Rahma, C. A., Hamid, S. B., & Irmawati, B. (2025, November). Classification of batik patterns using graph convolutional networks with data augmentation. In AIP Conference Proceedings (Vol. 3372, No. 1, p. 040001). AIP Publishing LLC.

[13] Fadhilah, H. T., Al Huda, F., & Yudistira, N. (2025, November). Development of Batik Pattern Classification Application Using Convolutional Neural Network Algorithm Using Android-Based Camera. In 2025 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob) (pp. 161-166). IEEE.

[14] Sari, I. P., Elvitaria, L., & Rudiansyah, R. (2025). Data-driven approach for batik pattern classification using convolutional neural network (CNN). Jurnal Mandiri IT, 13(3), 323-331.

[15] Suyahman, S., & Hapsari, A. (2025). VGG-Based Feature Extraction for Classifying Traditional Batik Motifs Using Machine Learning Models. Preservation, Digital Technology & Culture, (0).