Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer
Vol. 25, No. 2, March 2026, pp. 299∼310
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 10/C/C3/DT.05.00/2025
DOI: 10.30812/matrik.v25i2.5818

Identification of the Sub-motifs of Batik Kawung Using Deep Learning

Budi Sunarko[1], Subiyanto[1], Hari Wibawanto[1], Alfanza Rizky Zakaria[1], Alifian[1], Naufal Muhammad[1], Yudha Andriano Rismawan[2]
[1]Universitas Negeri Semarang, Semarang, Indonesia
[2]Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Article Info

Article history:

Received October 10, 2025
Revised December 07, 2025
Accepted February 03, 2026

Keywords:

Batik;
Kawung;
Identification;
Deep learning.

ABSTRACT

Batik is one of Indonesia's cultural heritages, with motifs that are both diverse and intricate. The Kawung motif, characterized by repetitive circular patterns, is divided into sub-motifs such as Kawung Bribil, Kawung Sen, and Kawung Picis. Automatic classification of these sub-motifs is important for digital preservation but remains difficult due to subtle inter-class similarities. The aim of this research is to analyze the performance of VGG, ResNet, and DenseNet and determine the most effective CNN architecture in classifying the sub-motifs of Batik Kawung. The research method is a convolutional neural network-based image classification approach using a dataset of 300 Kawung Batik images evenly distributed across three classes. Preprocessing steps included grayscale conversion, resizing to 256 $\times$ 256 pixels, Canny edge detection, and normalization to the range [0,1]. The dataset was randomly split into 210 training, 60 validation, and 30 testing images. The results of this research are that VGG achieved the highest training accuracy of 97%, but only 67% on the testing set, indicating a tendency to overfit. In contrast, DenseNet achieved the best generalization performance with a testing accuracy of 80%, surpassing both VGG and ResNet. At the class level, DenseNet161 demonstrated consistent performance across all Kawung sub-motifs, with precision ranging from 67% to 91% and F1-scores between 71% and 95%. These results suggest that DenseNet161 not only performed effectively during training but also generalized well to unseen data, establishing it as the most robust architecture for sub-motif Batik Kawung classification. The results underscore the effectiveness of CNNs, particularly DenseNet, in classifying subtle batik sub-motifs. This research contributes to develope a reliable automated system for identifying Kawung batik, leveraging modern technology to support the preservation of Indonesia's cultural heritage.

Corresponding Author:

Budi Sunarko, +62822-2304-4138,
Electrical Engineering, Faculty of Engineering,
Universitas Negeri Semarang, Semarang, Indonesia,
Email: budi.sunarko@mail.unnes.ac.id.

How to Cite:
B. Sunarko, "Identification of the Sub-motifs of Batik Kawung Using Deep Learning", *MATRIK: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, Vol. 25, No. 2, pp. 299-310, March, 2026.

1. INTRODUCTION

As an intangible cultural heritage designated by the United Nations Educational, Scientific and Cultural Organization (UNESCO), Indonesian batik has undergone rapid and diverse motif evolution. This dynamic progression underscores the critical need for a robust automated classification system to facilitate the accurate identification and preservation of this art form [1, 2]. Among the most complex and revered ancient motifs is Kawung. Traditionally, this motif is categorized into three distinct sub-motifs based on the scale of its geometric pattern: Kawung Bribil (resembling a 0.5-cent coin), Kawung Picis (a 10-cent coin), and Kawung Sen or Kawung Beton (akin to a 1-cent coin or a jackfruit seed) [3]. However, the intricate diversity of these Kawung sub-motifs often poses significant challenges for accurate identification, especially for the younger generation. While traditional methods rely on manual pattern size assessment, recent advancements in deep learning offer a more sophisticated and powerful approach for image-based analysis [2, 4–8]. Several studies have applied deep learning approaches to batik classification, with Convolutional Neural Networks (CNN) showing promising results in distinguishing between major motifs such as Kawung, Parang, and Mega Mendung. Rasyidi et al. [2] applied CNN models with ResNet, DenseNet, and VGG architectures on the Batik Tulis, Batik Cap, and Batik Printing datasets, with the best performance achieved by the VGG-13 architecture, yielding an accuracy of 87.61%. Oktarino et al. [4] identified five distinct batik motifs, namely Kawung, Betawi, Megamendung, Cendrawasih, and Parang Batik, using the CNN method with an accuracy of 99.2%. Isnanto and Triwiyatno [6] developed a Neural Network model to classify typical Central Java batik motifs, and obtained an average accuracy of 93.3%. Perdana et al. [8] investigated batik motif classification using convolutional neural networks combined with transfer learning, evaluating models such as MobileNetV2, ResNet50, DenseNet169, DenseNet201, and InceptionV3 on a dataset of approximately 3,000 batik images categorized into Kawung, Mega Mendung, and Parang motifs. The results of the study show that DenseNet achieved the highest classification accuracy, while MobileNetV2 demonstrated greater efficiency with lower computational cost and shorter processing time.

Several studies have explored transfer learning and alternative CNN backbones to improve batik classification performance. Alya et al. [9] utilized VGG-16 and Xception through transfer learning for batik motif recognition, demonstrating improved generalization on limited datasets and reducing computation time compared to the baseline CNN model. Sastyapratiwi et al. [10] used transfer learning and the MobileNet approach, achieving 98% accuracy. Transfer learning enables the reuse of representations learned from large and diverse datasets, thereby mitigating the limitations imposed by the relatively small size of batik image datasets. Preprocessing techniques like edge detection have been reported to enhance structural features in image classification tasks, although their application to repetitive batik motifs remains underexplored [11, 12]. Bakti and Hendrastuty [13] incorporated Canny edge detection with DenseNet121, demonstrating that edge-based preprocessing can improve CNN performance for batik images with complex repetitive patterns. Syafi'i and Khomsah [14] used Canny edge detection to improve the generalization of batik classification. Susanti et al. [15] used Canny edge detection and feature extraction using Gray Level Co-occurrence Matrix (GLCM) to improve the model's ability to recognize and classify batik patterns. These results confirm the relevance of texture representation in batik analysis. There are still unresolved gaps in previous research, namely the absence of focused studies on intra-motif classification of batik, particularly for visually similar sub-motifs such as Kawung Bribil, Kawung Picis, and Kawung Sen. Most existing works concentrate on inter-motif classification (e.g., Kawung, Parang, Mega Mendung, etc) between visually distinct patterns and do not address fine-grained geometric scale variations within a single motif family. Furthermore, comparative evaluations of deep CNN architectures specifically designed for repetitive, scale-sensitive batik sub-motifs remain limited.

To address this challenge, this study proposes a deep learning-based model specifically designed for the automated identification of the sub-motifs of Batik Kawung. This work centers on a comparative performance evaluation of three prominent Convolutional Neural Network (CNN) architectures, namely the Visual Geometry Group (VGG), the Residual Network (ResNet), and the Densely Connected Network (DenseNet), each of which has demonstrated remarkable effectiveness in diverse image recognition domains. These architectures were deliberately selected based on their distinctive structural characteristics and strengths. VGG is widely regarded as a strong baseline due to its simple yet powerful design, which employs sequential stacks of small convolutional filters to capture fine-grained visual details. ResNet introduces residual connections, which facilitate the training of deeper networks by reducing the vanishing gradient problem, thereby improving convergence and stability. DenseNet further advances this concept through dense connectivity, which promotes feature reuse, enhances gradient flow, and reduces parameter redundancy. This comparative approach will identify the most effective architecture for fine-grained classification of the Kawung motif. The difference between this research and the previous one is that this study explicitly addresses fine-grained intra-motif classification of Kawung batik by systematically evaluating VGG, ResNet, and DenseNet architectures for distinguishing sub-motifs of Kawung Batik. Unlike prior studies that focus on broad motif categories, this research emphasizes subtle geometric scale differences within a single traditional motif.

By comparing their model performances, this research aims to identify the most effective CNN architecture for fine-grained recognition of Batik Kawung sub-motifs. Consequently, this research is expected to make a significant contribution by establishing a reliable, automated identification system for sub-motifs of Kawung Batik, providing a benchmark for Kawung sub-motif recognition, and leveraging modern technology to support the preservation of Indonesia's invaluable cultural heritage. The rest of this article is organized as follows: Section 2 presents the research methods, including dataset description, preprocessing techniques, CNN model configurations, and evaluation metrics. Section 3 provides the results and analysis of model performance. Finally, Section 4 concludes the study with key findings and recommendations for future research.

2. RESEARCH METHOD

Figure 1 presents an overview of the methodological steps undertaken in this study. This research begins with the problem definition stage, which involves clearly formulating the research problem by emphasizing the need to accurately and efficiently classify the sub-motifs of Batik Kawung. The primary motivation behind this stage is the recognition that Kawung batik, as one of Indonesia's oldest and most iconic traditional motifs, possesses complex geometric structures and subtle variations that make manual classification challenging and prone to human error. By framing the research problem in this way, the study sets a focused direction for the development of an automated, deep learning-based classification framework. This stage also involves reviewing existing literature on batik motif classification, identifying gaps in current methodologies, and highlighting the need for a comparative analysis of advanced CNN architectures, namely VGG, ResNet, and DenseNet, for fine-grained motif recognition. This foundational stage ensures that the study is grounded in a well-justified rationale and has a clearly articulated objective.

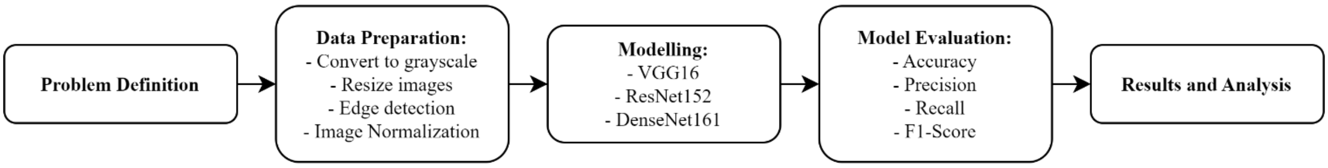

Figure 1. Research design

The second stage is data preparation, which constitutes a crucial step in ensuring the robustness and reliability of the subsequent modeling process. This stage involves collecting an extensive dataset of Kawung batik images from various sources to capture a wide diversity of sub-motif patterns, orientations, and design intricacies. The raw images are then carefully organized and annotated according to their respective sub-motif categories, creating a structured dataset suitable for machine learning tasks. Preprocessing is applied to ensure data uniformity and to enhance the model's learning capacity. Specifically, all images are resized to a standard resolution of 256×256 pixels to achieve dimensional consistency, thereby reducing computational overhead during model training. Furthermore, the images are converted to grayscale, which reduces the complexity of color information and directs the network's focus toward shape and texture features more relevant to structural motif analysis. Additionally, an edge detection technique is applied to the grayscale images. The inclusion of edge detection serves a strategic purpose; it accentuates the contours and structural outlines of the motifs, enabling the CNN to better capture shape-based features and boundary information, which are essential for differentiating between highly similar sub-motifs.

Following data preparation, the research proceeds to the model development stage, where three distinct CNN architectures, VGG, ResNet, and DenseNet, are implemented and evaluated. These models are selected for their proven strengths in image classification: VGG is known for its simplicity and uniform layer structure; ResNet incorporates residual connections that mitigate the vanishing gradient problem in deeper networks; and DenseNet features dense connectivity that promotes feature reuse and improves gradient flow. The training process for each model is carried out using the preprocessed Kawung dataset, with training and validation accuracy closely monitored over multiple epochs to track learning progression and detect potential overfitting. After training, the research advances to the evaluation stage, where the trained models are tested on an independent dataset that was not utilized during training or validation. This ensures an unbiased and objective assessment of model performance. The models were assessed using several evaluation metrics, which collectively provide a comprehensive understanding of each model's effectiveness. Finally, in the analysis and conclusion stage, the results from all three CNN architectures are systematically compared to determine the model that delivers the best overall performance in identifying the sub-motifs of Batik Kawung. This final stage synthesizes the findings and draws conclusions about the suitability of each architecture for this specific cultural and visual recognition task.

2.1. Dataset

This study used several datasets obtained from the Kaggle platform, uploaded by Hermansyah [16], Hendryhb [17] , and Alfan [18]. It comprises 300 kawung batik images, with an equal distribution of 100 images per class: Kawung Bribil, Kawung Sen, and Kawung Picis. This dataset was randomly split into training (120 images), validation (60 images), and test (30 images) sets, thereby ensuring an appropriate balance between model learning, parameter tuning, and unbiased evaluation. Figure 2 illustrates the motif variations among the sub-motifs of Batik Kawung.

(a) Embedded Lena without DCT (b) Embedded Peppers without DCT (c) Embedded Baboon without DCT

Figure 2. Kawung Bribil (a), Kawung Picis (b), and Kawung Sen (c)

2.2. Preprocessing

Preprocessing is an essential stage in preparing the dataset before model training, as it ensures consistency and enhances the quality of the input data for the learning process. In this study, preprocessing involves three primary steps. First, all images were converted into grayscale to reduce data complexity by removing color information while preserving the structural and textural patterns of the kawung batik motifs. The grayscale transformation was performed using a weighted combination of the red, green, and blue (RGB) channels, as defined in Equation 1, where Red(x,y),Green(x,y),Blue(x,y) represent the red, green, and blue channel intensities at the pixel (x,y) respectively. Second, the images were resized to uniform dimensions to meet the input needs of the CNN and to guarantee comparability across samples. The resizing process can be expressed as Equation 2, where I_gray (x,y) denotes the original grayscale pixel intensity at the coordinate (x,y) and s_x,s_y represent the scaling factors along the horizontal and vertical axes, respectively. In this study, all images were resized to a uniform resolution of 256×256 pixels. Third, edge detection was applied using the Canny operator to emphasize motif boundaries and structural contours. The Canny operator combines gradient filtering with dual-threshold hysteresis, enabling effective detection of fine edges while reducing noise [13–15]. The edge detection process can be mathematically formulated as a convolution operation between the grayscale image I_resized (x,y) and an edge detection kernel K [19], as shown in Equation 3. Finally, all processed images were normalized into the pixel intensity range [0, 1] to enhance numerical stability and expedite convergence during training. Normalization was implemented using the min–max scaling formula, as expressed in Equation 4, where I_edge (x,y) is the edge-detected image at pixel (x,y), I_min and I_max are the minimum and maximum intensity values in the image. Through grayscale conversion, resizing, Canny-based edge detection, and normalization, the dataset was standardized and enhanced, thereby facilitating more effective feature extraction and improving the robustness of the subsequent CNN-based classification.

$$I_{gray}(x,y) = 0.2989 \cdot Red(x,y) + 0.5870 \cdot Green(x,y) + 0.1140 \cdot Blue(x,y) \tag{1}$$

$$I_{resized}(x',y') = I_{gray}\left(\frac{x}{s_x}, \frac{y}{s_y}\right) \tag{2}$$

$$I_{edge}(x,y) = I_{resized}(x',y') * K \tag{3}$$

$$I_{normalized}(x,y) = \frac{I_{edge}(x,y) - I_{min}}{I_{max} - I_{min}} \tag{4}$$

2.3. Modelling

Computer vision has undergone a revolution thanks to Convolutional Neural Networks (CNN), a deep learning architecture that enables automatic feature extraction from raw image input. Unlike traditional machine learning approaches, which rely on handcrafted features, CNNs are designed to derive hierarchical feature representations directly from the data through the use of convolutional, pooling, and fully connected layers [20, 21]. The convolutional layers act as localized feature detectors, capturing low-level patterns such as edges, corners, and textures in the earlier layers, and progressively learning more abstract, high-level representations in deeper layers. Pooling layers are interspersed to downsample feature maps, reducing spatial dimensions while preserving the most salient information, which contributes to translational invariance and improved computational efficiency. This hierarchical representation learning capability has made CNNs the dominant paradigm for image classification, object detection, and other vision-related tasks.

In this study, three well-established CNN architectures, namely VGG, ResNet, and DenseNet, were comparatively evaluated for the identification of sub-motifs in Batik Kawung. Kawung is one of the oldest and most iconic traditional batik motifs originating from Java, Indonesia. It is composed of symmetrical circular or elliptical shapes arranged in a repetitive grid pattern. Despite their visual similarity, Kawung sub-motifs such as Kawung Bribil, Kawung Picis, and Kawung Sen differ subtly in the size, spacing, and ornamental details of their components. Accurately classifying these sub-motifs requires fine-grained visual discrimination, which poses a significant challenge even for deep learning models, especially when the available dataset is relatively small. Therefore, the choice of CNN architecture is crucial to determine the success of the classification task.

The first architecture examined is VGG, which has become one of the most widely adopted CNN models in computer vision research due to its simplicity and effectiveness [14, 22]. VGG architectures are characterized by a very uniform and straightforward design: they stack multiple convolutional layers with small 3×3 filters in depth, separated by max-pooling layers, followed by a series of fully connected layers at the end. The small receptive field of the 3×3 filters allows the network to capture fine local details, while deep stacking many such layers enables it to build complex, highly nonlinear feature hierarchies. However, this design results in a very large number of parameters, making VGG models computationally expensive and memory-intensive to train. While VGG has demonstrated strong performance on large-scale datasets such as ImageNet, it is well-documented that it tends to overfit on smaller datasets. This makes it a useful baseline model, but not necessarily the most efficient or generalizable for specialized tasks with limited data, such as Kawung motif classification.

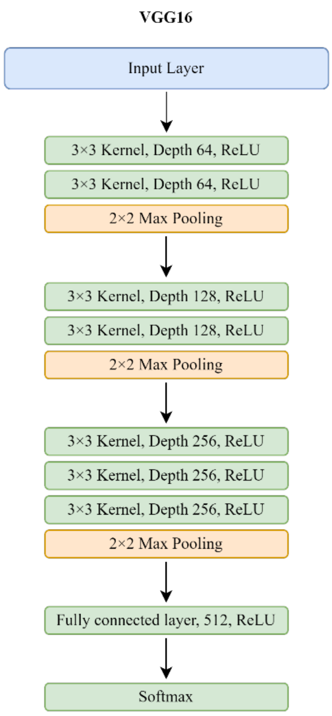

Figure 3. VGG16 architecture

From Figure 3, the VGG16-based model used in this study consists of three convolutional blocks followed by fully connected layers [22]. The first block includes two convolutional layers with 64 filters of size 3×3 using ReLU activation and same padding, followed by a 2×2 max-pooling layer. The second block mirrors this structure with two convolutional layers using 128 filters, also followed by max pooling. The third block consists of three convolutional layers with 256 filters, again followed by max pooling. To reduce model complexity and prevent overfitting on the limited dataset, blocks 4 and 5 from the original VGG16 architecture, which typically use 512 filters, were omitted. After feature extraction, the output is flattened and passed through a fully connected layer with 512 ReLU units, followed by an output dense layer corresponding to the number of classes. The model was compiled using the Adam optimizer with a learning rate of 1×10^{-4}, sparse categorical cross-entropy loss (from logits), and accuracy as the evaluation metric. This configuration preserves the essential hierarchical feature extraction of VGG16 while adapting the architecture for a smaller dataset.

The second architecture, ResNet, was proposed to overcome a major challenge in deep learning: the vanishing gradient problem, which hampers the training of very deep networks [23, 24]. As networks become deeper, gradients propagated back through many layers during training tend to shrink exponentially, leading the initial layers to exhibit significantly reduced learning or to stop learning altogether. ResNet addresses this issue by introducing skip connections or residual blocks, which allow the gradient to flow directly across layers. Instead of learning a full mapping H(x) from inputs to outputs, each residual block learns a residual function F(x)=H(x)-x, and the output is formulated as F(x)+x. This reformulation makes it easier to optimize very deep networks by allowing them to simply learn the identity mapping when needed. ResNet architectures can therefore reach depths of over 100 layers while maintaining good convergence properties and high accuracy. The ResNet152 variant used in this study is a very deep network with 152 layers, offering strong representational power. However, its extreme depth also makes it more data-hungry and susceptible to overfitting when the dataset is not sufficiently large or diverse.

The ResNet152 [23] architecture was implemented using the functional API, with grayscale images first converted into three-channel inputs through a Lambda layer to match the network's expected input format. A pre-trained ResNet152 model with ImageNet weights was employed as the feature extractor, with its layers frozen to retain the learned representations. The extracted features were then processed by a Global Average Pooling layer, followed by a fully connected layer with 512 ReLU units, and a final Dense layer with the same number of classes for classification. The model was compiled with the Adam optimizer and sparse categorical cross-entropy as the loss function, with a learning rate of 1×10^{-4}. Figure 4 shows the architecture of ResNet152 used in this study.

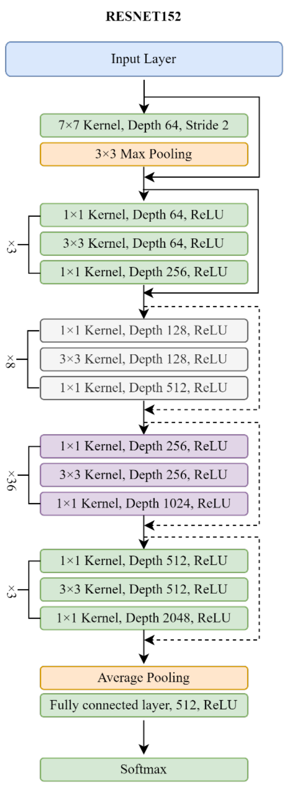

Figure 4. ResNet152 architecture

The third architecture, DenseNet, builds upon the residual learning concept introduced by ResNet and extends it to a more densely connected paradigm [25, 26]. DenseNet establishes direct connections between each layer and every other subsequent layer within the same dense block. This means that the feature maps produced by all preceding layers are concatenated and passed as inputs to each layer, and each layer's output is passed to all subsequent layers. Such dense connectivity offers several advantages. First, it encourages feature reuse, allowing the model to build on previously learned patterns rather than relearning similar features in different layers. Second, it improves gradient flow by allowing gradients to propagate to earlier layers via multiple short paths, alleviating the vanishing gradient problem and making training more stable. Third, because layers can access rich feature sets from earlier layers, each layer can be relatively narrow, significantly reducing the number of parameters compared to conventional deep CNNs. This makes DenseNet more parameter-efficient, less prone to overfitting, and well-suited to tasks with limited training data.

DenseNet161 [25], the variant used in this study, strikes a balance between depth and parameter efficiency, providing a strong architecture for fine-grained classification. In this study, DenseNet201 architecture was employed as the backbone network using TensorFlow's Functional API. Since the dataset consisted of grayscale edge images with a single input channel, the tensor of shape $(256\times256\times1)$ was expanded into three channels by concatenating the grayscale channel, resulting in $(256\times256\times3)$, matching the input requirements of DenseNet pre-trained on ImageNet. The DenseNet201 model was initialized with the top fully connected layer not included and frozen to act as a feature extractor. The output feature maps were processed with a Global Average Pooling 2D layer to aggregate spatial information, followed by a fully connected layer with 512 units and ReLU activation for feature refinement. Finally, a Dense layer with units equal to the number of classes was added to produce the classification logits. The model was trained using the Adam optimizer with a learning rate of 1×10^{-4} and Sparse Categorical Crossentropy loss, ensuring efficient optimization for the Kawung sub-motif classification task. Figure 5 shows the architecture of DenseNet161 used in this study.

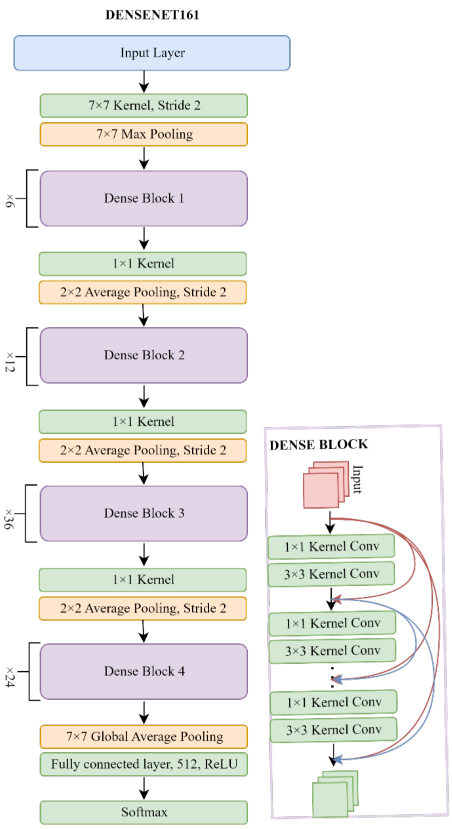

Figure 5. DenseNet161 architecture

By comparing these three architectures, this study aims to identify the most effective CNN model for fine-grained recognition of Batik Kawung sub-motifs. The choice of architecture is critical because it influences not only overall accuracy but also the model's ability to generalize across classes, avoid overfitting, and maintain balanced performance on visually similar categories.

Understanding the strengths and limitations of each architecture in the specific context of cultural heritage image classification contributes to both the technical field of deep learning and the broader goal of preserving and digitizing traditional cultural artifacts. The insights gained from this comparison are expected to guide future work on automated batik motif recognition and similar pattern analysis tasks in the cultural heritage domain. In this study, transfer learning was implemented using ImageNet-pretrained CNN models from the TensorFlow Keras library, enabling the reuse of learned visual features from large-scale natural image datasets.

2.4. Evaluation

True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values were obtained from a confusion matrix to assess the model's performance. Additionally, a number of evaluation criteria were employed, such as accuracy, precision, recall, and F1-score. Equation 5 defines accuracy as the total proportion of cases that are correctly classified. Equation 6 defines precision as the proportion of correctly predicted positive cases among all instances projected as positive. Equation 7 defines recall (sensitivity), the percentage of true positives the model correctly detects. Equation 8 defines the f1-score, the harmonic mean of precision and recall, as a balanced metric between the two. These indicators together provide a comprehensive review of the model's classification performance, with accuracy indicating overall correctness, precision underscoring the reliability of positive predictions, recall emphasizing the ability to identify positive instances, and the F1-score integrating precision and recall into a single measure.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{8}$$

3. RESULT AND ANALYSIS

After preprocessing, normalized grayscale images are produced. The next step is to use these images as input to the CNN model with three different architectures. This research investigates three distinct CNN architectures: VGG16, ResNet152, and DenseNet161 for the classification of the sub-motifs Batik Kawung. Model performance was evaluated using the metrics defined in Equations 5, 6, 7, and 8, enabling a comprehensive comparison across architectures. The findings from this research for each CNN architecture are summarized in Table 1.

Table 1. The Performance Comparison of CNN Architectures

Model	Class	Precision	Recall	F1-Score	Training Accuracy	Validation Accuracy	Testing Accuracy
	Kawung Bribil	58%	70%	64%			
VGG16	Kawung Picis	71%	50%	59%	**97%**	56%	67%
	Kawung Sen	73%	80%	76%			
	Kawung Bribil	75%	60%	67%			
ResNet152	Kawung Picis	50%	50%	50%	65%	49%	57%
	Kawung Sen	50%	60%	55%			
	Kawung Bribil	91%	100%	95%			
DenseNet161	Kawung Picis	86%	60%	71%	95%	75%	**80%**
	Kawung Sen	67%	80%	73%			

VGG16 achieved the highest training accuracy among the three models, reaching 97%. This figure demonstrates the model's strong ability to capture complex visual patterns from the training set. However, this high training performance was accompanied by steep declines in validation (56%) and test accuracy (67%), indicating severe overfitting. Overfitting occurs when a model memorizes the specific patterns of the training data rather than learning generalizable features or patterns that apply to unseen samples. This tendency toward overfitting can be attributed to the structural characteristics of VGG16. As a deep and parameter-heavy architecture, VGG16 requires a large amount of training data to prevent it from fitting to noise or irrelevant variations in the training set. When the available data is relatively limited, as in the present case, the model is prone to fitting spurious correlations that do not generalize well beyond the training set. A closer examination of class-level metrics further underscores this issue. While VGG16 performed reasonably well on the Kawung Sen class (recall 80%, F1-score 76%), its performance on Kawung Picis was notably weaker (recall 50%, F1-score 59%). This discrepancy reveals class-level inconsistency, suggesting that the model fails to learn equally discriminative representations across all categories. Such inconsistency reduces the model's reliability when deployed in real-world settings with diverse class distributions. Therefore, although VGG16 demonstrates strong learning capacity on training data, its limited generalization ability and class inconsistency make it less suitable for the Kawung sub-motif classification task under data-constrained conditions. Future work could explore strategies to alleviate overfitting, such as stronger regularization (e.g., dropout, weight decay), extensive data augmentation, or simplifying the architecture to better match the dataset size.

ResNet152, in contrast, exhibited lower overall performance, with training, validation, and testing accuracies of 65%, 49%, and 57%, respectively. These results indicate that the model struggled to learn meaningful representations from the training data, thereby limiting its performance on unseen data. Despite this generally low performance, ResNet152 achieved moderate results on the Kawung Bribil class (precision: 75%, recall: 60%, F1-score: 67%). However, it struggled to correctly recognize Kawung Picis and Kawung Sen, resulting in low precision and recall for these classes. This imbalance suggests a tendency toward class bias, where the model favors one class (likely the one with more samples) while failing to adequately distinguish between other classes. One plausible explanation is that the dataset may have an imbalanced class distribution, leading the model to be more often exposed to certain classes during training. Consequently, it becomes biased toward predicting the dominant classes, resulting in superficially high precision but poor generalization across minority classes. Additionally, while ResNet152 is theoretically powerful due to its depth and residual connections, its high complexity also makes it more prone to optimization difficulties such as vanishing gradients, especially when trained on limited data or with insufficient training iterations. To address these issues, future studies could employ techniques that directly tackle class imbalance, such as class weighting, oversampling minority classes, or utilizing focal loss to focus the model's learning on hard-to-classify samples. Such strategies could help ResNet152 learn more balanced and discriminative features across all classes, thereby improving its generalization capability.

DenseNet161 outperformed the other two architectures, delivering the most balanced and robust results. It achieved high, consistent accuracy across the training (95%), validation (75%), and test (80%) sets. This consistency indicates that DenseNet161 not only learned useful patterns from the training data but also successfully generalized its knowledge to unseen data. At the class level, DenseNet161 exhibited strong and stable performance across all sub-motifs, with particularly excellent results on Kawung Bribil (precision 91%, recall 100%, F1-score 95%). The high recall shows that nearly all Kawung Bribil samples were correctly identified, while the high precision indicates very few misclassifications into other classes. This uniform performance suggests that the model effectively learned discriminative features that distinguish each sub-motif.

The results of this research are in line with previous studies, suggesting that DenseNet demonstrates superior performance in fine-grained pattern classification compared to VGG and ResNet [8, 13, 25]. The superior performance of DenseNet161 can be attributed to its unique dense connectivity pattern. In DenseNet, each layer receives inputs from all preceding layers and passes their output to all subsequent layers. This design allows feature maps learned in earlier layers to be directly reused in later layers, enhancing feature propagation and reducing information loss. Furthermore, this structure mitigates the vanishing gradient problem, as gradients can flow more easily through short connections back to earlier layers. Another important characteristic of DenseNet is its inherent regularizing effect. Because each layer has access to previously learned features, it does not need to relearn redundant ones, effectively reducing the number of parameters to be optimized. This makes the model less prone to overfitting, especially when trained on relatively small datasets. The relatively small gap between training and testing accuracy in DenseNet161 demonstrates its strong generalization ability, making it the most effective and reliable architecture for classifying the sub-motifs of Batik Kawung in this study.

A key issue across the models in this study is overfitting, particularly in deep architectures such as VGG16 and ResNet152. The considerable difference between the model's performance on the training and testing datasets highlights the challenge of achieving good generalization with a limited dataset. Overfitting limits the model's ability to adapt to unseen data, reducing its reliability in real-world applications. Addressing this issue will be crucial for future research, potentially through techniques such as extensive data augmentation, stronger regularization, or hybrid approaches that integrate handcrafted feature extraction with deep learning.

Although the results of this study demonstrate satisfactory performance, there remains room for further improvement. Kurniawan et al. [27] employed a similar approach using a CNN with the ResNet50 architecture to classify Pekalongan batik motifs, achieving an accuracy of 76%. To enhance performance, they applied contrast enhancement techniques, namely Contrast Limited Adaptive Histogram Equalization (CLAHE) and Histogram Equalization (HE), which successfully increased the accuracy to 83% and 81%, respectively [27]. This approach suggests that contrast enhancement could be an interesting method to implement and evaluate in the context of Kawung batik classification in order to improve model performance. In addition to contrast-based preprocessing, another promising direction is integrating shape and texture feature extraction. Oktarino et al. [4], for instance, utilized the Gray Level Co-occurrence Matrix (GLCM) methods for texture feature extraction on batik motifs such as Betawi, Cendrawasih, Kawung, Megamendung, and Parang, followed by CNN-based classification, achieving an accuracy of 99.2%. This indicates that texture-based features can be highly effective in improving classification accuracy and may serve as a valuable addition to CNN architectures for Kawung batik motif recognition.

4. CONCLUSION

This study compared the performance of three CNN architectures: VGG16, ResNet152, and DenseNet161 for the classification of the sub-motifs of Batik Kawung batik. The findings reveal that VGG16, despite achieving the highest training accuracy, exhibited lower validation and testing performance, indicating an overfitting tendency. ResNet152 produced the weakest overall performance, showing class-level bias and difficulty in achieving balanced generalization. In contrast, DenseNet161 demonstrated the most consistent and reliable results, with the highest testing accuracy and stable precision, recall, and F1-scores across all classes. These outcomes highlight DenseNet161 as the most effective architecture for sub-motifs of Batik Kawung classification in this study.

From a research development perspective, future studies could explore larger and more diverse batik datasets, incorporate data augmentation strategies, and investigate the integration of attention mechanisms to further improve classification accuracy. Moreover, hybrid models that combine CNNs with advanced feature extraction techniques, such as GLCM, wavelet transforms, or texture descriptors, may provide additional robustness in motif recognition. In terms of application prospects, the findings of this study could contribute to the development of automated batik recognition systems that support cultural preservation, digital archiving, and educational platforms. Furthermore, implementing such models in e-commerce and creative industries could assist in motif identification, authentication, and cataloging of batik products, thereby strengthening both cultural heritage appreciation and commercial applications.

5. ACKNOWLEDGEMENTS

Thanks to Universitas Negeri Semarang for funding and supporting this research.

6. DECLARATIONS

AI USAGE STATEMENT

During the preparation of this manuscript, the authors utilized ChatGPT (OpenAI) to enhance the language quality and clarity of the text. The authors subsequently reviewed and revised the content as necessary and assume full responsibility for the final version of the manuscript.

AUTHOR CONTIBUTION

The conceptualization of the study was carried out by Budi Sunarko. The research methodology was developed by Budi Sunarko in collaboration with Alifian. The investigation process was conducted by Hari Wibawanto and Alfanza Rizky. The discussion of the research results was performed by Yudha Andriano Rismawan. The original draft of the manuscript was prepared by Alfanza Rizky Zakaria and Naufal Muhammad, while the review and editing of the manuscript were undertaken by Yudha Andriano Rismawan. The provision of research resources was supported by Budi Sunarko, Alfanza Rizky Zakaria, and Naufal Muhammad. Supervision was conducted by Budi Sunarko, Subiyanto, and Hari Wibawanto. Final approval of the manuscript was granted by Budi Sunarko.

FUNDING STATEMENT

This research was funded by the Research and Community Service Institute of Universitas Negeri Semarang, Indonesia.

COMPETING INTEREST

The authors declare no conflict of interest in this article.

REFERENCES

[1] L. Elvitaria, E. F. A. Shaubari, N. A. Samsudin, S. K. A. Khalid, S. Salamun, and Z. Indra, "A Proposed Batik Automatic Classification System Based on Ensemble Deep Learning and GLCM Feature Extraction Method," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 10, 2024, https://doi.org/10.14569/IJACSA.2024.0151058.

[2] M. A. Rasyidi, R. Handayani, and F. Aziz, "Identification of batik making method from images using convolutional neural network with limited amount of data," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 3, pp. 1300–1307, Jun. 2021, https://doi.org/10.11591/eei.v10i3.3035.

[3] Z. Abidin, M. Supriatna, T. Herman, L. Farokhah, and R. Febriandi, "The geometric patterns in Kawung Surakarta batik motif: An ethnomathematical exploration," p. 020061, 2023, https://doi.org/10.1063/5.0141663.

[4] A. Oktarino, Y. Desnita-Tasri, and A. Efendi, "Identification of Batik Motif Based Deep Learning-Convolutional Neural Network Approach," *The Journal of Ocean, Mechanical and Aerospace -science and engineering- (JOMAse)*, vol. 68, no. 3, pp. 139–147, Nov. 2024, https://doi.org/10.36842/jomase.v68i3.381.

[5] S. Ibrahim, M. F. A. Fauzi, N. N. A. Mangshor, R. Aminuddin, and B. Sunarko, "Automated platelet counter with detection using k-means clustering," *Annals of Emerging Technologies in Computing*, vol. 7, no. 5, pp. 39–49, Oct. 2023, https://doi.org/10.33166/AETiC.2023.05.004.

[6] R. R. Isnanto and A. Triwiyatno, "Classification of traditional batik motifs in central java using gabor filter and backpropagationneural network," *Scientific Journal of Informatics*, vol. 7, no. 2, pp. 279–290, Nov. 2020, https://doi.org/10.15294/sji.v7i2.26215.

[7] B. Sunarko, A. A. Firdaus, Y. A. Rismawan, and A. Nugroho, "Analysis of Facial Areas to Identify CHD Risks Based on Facial Textures," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 14, no. 1, pp. 69–76, Feb. 2025, https://doi.org/10.22146/jnteti.v14i1.13658.

[8] A. A. M. Perdana, M. Fajar B, and A. M. Mappalotteng, "Enhancing Batik Classification Leveraging CNN Models and Transfer Learning," *JOIV : International Journal on Informatics Visualization*, vol. 9, no. 3, p. 1033, May 2025, https://doi.org/10.62527/joiv.9.3.2535.

[9] R. F. Alya, M. Wibowo, and P. Paradise, "Classification of Batik Motif Using Transfer Learning on Convolutional Neural Network (CNN)," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 1, pp. 161–170, Feb. 2023, https://doi.org/10.52436/1.jutif.2023.4.1.564.

[10] H. Sastypratiwi, H. Muhardi, and Y. Yulianti, "Batik Recognition and Classification Using Transfer Learning and MobileNet Approach," *JOIV : International Journal on Informatics Visualization*, vol. 8, no. 4, p. 2400, Dec. 2024, https://doi.org/10.62527/joiv.8.4.2407.

[11] B. Tianhao, M. Lazarou, and T. Stathaki, "Image Edge Enhancement for Effective Image Classification:," pp. 444–451, 2024, https://doi.org/10.5220/0012364900003660.

[12] S. Bhat and P. Szczuko, "Impact of canny edge detection preprocessing on performance of machine learning models for Parkinson's disease classification," *Scientific Reports*, vol. 15, no. 1, p. 16413, May 2025, https://doi.org/10.1038/s41598-025-98356-7.

[13] I. J. Bakti and N. Hendrastuty, "Application of Canny Operator in Batik Motif Image Classification with Convolutional Neural Network Approach," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 3, pp. 789–798, May 2024, https://doi.org/10.52436/1.jutif.2024.5.3.1894.

[14] R. Syafi'i and S. Khomsah, "Classification of Indonesian Batik: A Comparative Study of CNN and VGG16 with Canny Edge Detection," pp. 355–363, Nov. 2024, https://doi.org/10.1109/COMNETSAT63286.2024.10862216.

[15] I. D. Susanti, "Yogyakarta Batik Image Classification Based on Convolutional Neural Network," *Advance Sustainable Science, Engineering and Technology*, vol. 6, no. 1, p. 02401020, Jan. 2024, https://doi.org/10.26877/asset.v6i1.18002.

[16] D. D. Hermansyah, "Indonesian Batik Motifs," *Kaggle*, 2020.

[17] HendryHB, "Batik Nusantara (Batik Indonesia) Dataset," *Kaggle*, 2024.

[18] A. Alfan, "Indonesian Batik Motifs (Corak App)," *Kaggle*, 2020.

[19] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986, https://doi.org/10.1109/TPAMI.1986.4767851.

[20] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaria, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, https://doi.org/10.1186/s40537-021-00444-8.

[21] P. Purwono, A. Ma'arif, W. Rahmaniar, H. I. K. Fathurrahman, A. Z. K. Frisky, and Q. M. U. Haq, "Understanding of Convolutional Neural Network (CNN): A Review," *International Journal of Robotics and Control Systems*, vol. 2, no. 4, pp. 739–748, Jan. 2023, https://doi.org/10.31763/ijrcs.v2i4.888.

[22] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," no. arXiv:1409.1556, Apr. 2015, https://doi.org/10.48550/arXiv.1409.1556.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," pp. 770–778, Jun. 2016, https://doi.org/10.1109/CVPR.2016.90.

[24] A. C. Sani Zulkarnaen, I. Gusti Ngurah Rejski Ariantara Putra, N. F. Reviana, R. Hidayah, N. Ibrahim, N. K. Caecar Pratiwi, and Y. N. Fuadah, "Application of Convolutional Neural Network Method with MobileNet V1 and ResNet-152 V2 Architecture in Batik Motif Classification," vol. 186, pp. 57–68, 2024, https://doi.org/10.1007/978-3-031-46784-4_6.

[25] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," pp. 2261–2269, jul 2017, https://doi.org/10.1109/CVPR.2017.243.

[26] A. Firman Ihsan, "A Study of Batik Style Transfer using Neural Network," pp. 313–319, Aug. 2021, https://doi.org/10.1109/ICoICT52021.2021.9527490.

[27] M. B. Kurniawan and E. Utami, "Comparative Analysis of Contrast Enhancement Methods for Classification of Pekalongan Batik Motifs Using Convolutional Neural Network," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 6, pp. 1779–1787, Dec. 2024, https://doi.org/10.52436/1.jutif.2024.5.6.2621.