Jurnal Pendidikan Teknologi Informasi dan Komunikasi
Volume 5, Issue 1, 1-16
e_ISSN: 2963-1963
https://ejournal.1001tutorial.com/index.php/jentik/index
DOI: https://doi.org/10.58723/jentik.v5i1.637

A Lightweight Hybrid GLCM–MobileNetV2 Model for Batik Motif Recognition in Digital Cultural Learning Environments

Haryanto*
Universitas Dharma AUB,
INDONESIA

Husna Sarirah Husin
Taylor's University,
MALAYSIA

Hartini
Universitas Dharma AUB,
INDONESIA

F.R Desiana Kardha
Universitas Dharma AUB,
INDONESIA

Widyo Ari Utomo
Universitas Dharma AUB,
INDONESIA

Article Info

Article history:

Received: March 10, 2026
Revised: April 14, 2026
Accepted: May 03, 2026

Keywords:

Digital Cultural Learning Environments
GLCM
MobilenetV2
Parang Batik

Abstract

Background: Automatic identification of Surakarta *Parang* batik motifs presents significant challenges due to the high visual similarity among sub-motifs, where conventional Convolutional Neural Network (CNN) architectures often fail to capture fine-grained texture characteristics.
Purpose of Study: A lightweight hybrid model is proposed to integrate 24 GLCM-derived texture features with MobileNetV2 spatial descriptors through a feature fusion strategy to improve motif classification accuracy.
Methodology: The proposed methodology employs a hybrid feature extraction strategy, where 24 texture descriptors consisting of six statistical parameters (Contrast, Correlation, Homogeneity, Dissimilarity, ASM, and Energy) calculated across four orientations (0°, 45°, 90°, and 135°) with 1,280 deep spatial features obtained from the MobileNetV2 backbone.
Main Findings: Experimental results demonstrate that the proposed hybrid model achieves an accuracy of 99%, representing a substantial performance gain over the baseline MobileNetV2 model (66.67%) and the GLCM-SVM approach (85%). These results indicate that the integration of statistical texture descriptors and deep spatial features notably enhances the recognition of complex batik patterns. Furthermore, the findings suggest that this feature fusion approach is highly effective in resolving the intricate geometric similarities of *Parang* sub-motifs, providing a more reliable and efficient alternative to standard deep learning models for fine-grained classification tasks.
Novelty/Originality of This Study: The novelty of this study lies in the implementation of a feature fusion strategy that compensates for the limitations of lightweight CNNs in texture recognition by incorporating classical statistical descriptors, specifically tailored for the intricate patterns of *Parang* batik.

To cite this article: Haryanto, H., Husin , HS, Hartini, H., Kardha, FD, & Utomo, WA (2026). Model Hibrida Ringan GLCM–MobileNetV2 untuk Pengenalan Motif Batik dalam Lingkungan Pembelajaran Budaya Digital. *JENTIK : Jurnal Pendidikan Teknologi Informasi Dan Komunikasi* , *5* (1), 1–16.

INTRODUCTION

Indonesia boasts a rich cultural heritage recognized by UNESCO, with Batik as its most iconic art form. Each motif, particularly those from Surakarta, carries deep philosophical meanings and artistic values that reflect Indonesian identity (Purnomo et al., 2021). However, distinguishing between specific batik motifs, such as the *Parang* sub-motifs *(Slobog, Tuding, and Klitik)*, presents a formidable challenge in computer vision (Lionardi & Vincencia, 2025). These patterns exhibit extreme inter-class visual similarity and high intra-class variance, where distinguishing features lie in microscopic repetitive textures rather than global shapes. Consequently, accurate identification remains difficult even for experts (Venkataramanan, 2021), necessitating an advanced yet computationally efficient automated system to assist the general public and learners (Alya et al., 2023).

The application of information and communication technology (ICT) in cultural preservation is not only about digitizing archives, but also about creating an interactive and accurate learning

---

` * **Corresponding author:**
Author, University, COUNTRY. ✉email

ecosystem. In a digital cultural learning environment, the system's ability to distinguish very similar motifs visually, such as the *Parang* sub-motif, is crucial because each pattern carries a different philosophical narrative. If the system fails to identify motifs with high accuracy, the process of cultural disinformation to students becomes significant. Therefore, this task is not only about computer vision but also about developing an educational tool that allows students or museum visitors to perform instant and accurate self-identification directly from their devices. By integrating a lightweight yet precise hybrid model, this technology bridges the gap between expert knowledge and public understanding, thereby strengthening the effectiveness of cultural knowledge transfer in the digital age. Despite this potential, the practical implementation of such systems in digital learning environments faces a significant technical hurdle. While intelligent systems can assist users in identifying traditional patterns more effectively (Ju, 2024; Zheng et al., 2024), a critical gap identified in current research is the trade-off between model lightness and feature granularity. Although deep Convolutional Neural Network (CNN) architectures, such as ResNet or VGG, offer high accuracy, they are often computationally prohibitive for real-time mobile applications due to their massive parameter counts (Gultom et al., 2018; Yusriadi et al., 2023). This limitation hinders the goal of providing an accessible educational tool that can run seamlessly on students' personal devices**.**

Lightweight models like MobileNetV2 offer a solution for mobile deployment, yet they often struggle to capture the subtle, fine-grained textural nuances of batik, leading to significant accuracy degradation (Anggoro et al., 2024). Recent studies have attempted to bridge this. For instance, (Khoirunnisa et al., 2025) utilized CNNs to categorize Solo Batik motifs, while they explored GLCM and CNN for Yogyakarta batik.

Previous studies have made notable strides in automating batik motif identification. For instance, recent applications of deep Convolutional Neural Networks (CNNs) have achieved high accuracy in broad batik classification (Gultom et al., 2018; Sari et al., 2025). However, these achievements are largely confined to datasets with distinct inter-class variations, often presenting challenges when addressing the extreme visual similarities found in specific sub-motifs like *Parang*. Furthermore, while VGG-based models offer robust feature extraction (Suyahman & Hapsari, 2025), their architectural complexity leads to a high parameter count, making them less optimal for the real-time, low-resource environments typical of digital cultural learning applications. Most existing research prioritizes global spatial features but may overlook the fine-grained texture descriptors such as those provided by GLCM, which are essential for distinguishing motifs with nearly identical geometric structures but different scale-dependent textures. This research gap necessitates a more balanced approach that maintains high granularity while prioritizing model efficiency.

Nevertheless, these approaches often fail to reach high precision on motifs with repetitive textures like the *Parang* sub-motifs. Unlike recent comparative studies on similar objects by (Anto et al., 2025) and (Ottoni et al., 2023), there is still a lack of effective feature fusion strategies optimized for fine-grained batik textures in resource-constrained settings.

To address these limitations, this study proposes a novel lightweight hybrid architecture that integrates Gray-Level Co-occurrence Matrix (GLCM) texture descriptors with MobileNetV2 spatial features through a late-fusion strategy. While MobileNetV2 serves as an efficient backbone for high-level feature extraction, GLCM is strategically embedded to provide deterministic texture descriptors such as contrast and correlation that are typically compromised during the downsampling process in lightweight networks. Theoretically, the repeated application of strided convolutions and pooling layers in CNNs acts as a low-pass filter, which prioritizes global semantic structures while progressively discarding high-frequency components that constitute fine-grained textures. Previous empirical studies have demonstrated that lightweight architectures, specifically those employing depthwise separable convolutions, often suffer from 'feature blurring' where the loss of spatial resolution leads to a significant reduction in the discriminative power of local patterns (Sandler et al., 2018). By integrating GLCM, this model recovers these lost texture nuances, ensuring that the structural identity of motifs like *Parang,* which relies heavily on micro-edge distributions, is preserved**.** This hybrid approach aims to improve classification accuracy for highly similar motifs while maintaining the computational efficiency required for digital learning platforms.

The primary contributions and novelty of this research are encapsulated in three key aspects. First, it introduces a hybrid feature fusion mechanism that concatenates 1,280 global spatial descriptors from MobileNetV2 with 16 localized statistical texture features from GLCM. This effectively bridges the gap between deep spatial representations and 'hand-crafted' descriptors, allowing the model to capture deterministic texture patterns that are often lost during the downsampling process in standard CNNs.

Second, this study provides a fine-grained specialization for Parang batik classification, specifically addressing motifs with extremely low inter-class variance. The proposed hybrid model achieves a significant performance leap, increasing accuracy from 66.67% (baseline MobileNetV2) to 99%, while maintaining a high F1-score of 0.99 across all sub-motif categories.

Finally, the architecture is designed to be strictly lightweight for deployment in resource-constrained environments. By utilizing MobileNetV2's depthwise separable convolutions and a Global Average Pooling (GAP) layer, the model minimizes the computational footprint. Specifically, the model consists of approximately 2.26 million trainable parameters with a compact storage size of roughly 16 MB. Compared to heavy architectures like VGG16 (~138 million parameters) or ResNet50 (~25 million parameters), the proposed model offers a reduction in parameter density of over 90% while achieving superior accuracy. With an average inference time of less than 100ms on standard mobile-grade processors, this model provides a scalable and practical solution for high-precision recognition in interactive digital cultural learning environments.

RESEARCH METHODS

The dataset utilized in this study consists of a total of 1,200 images, balanced across three Parang batik sub-motifs: Slobog, Tuding, and Klitik (400 images per class). To ensure transparency regarding the data origin, the collection process is detailed as follows:

1. Original Image Acquisition: A total of 120 original high-resolution images were collected as the primary baseline (40 images per class). These images were sourced using a hybrid strategy to ensure high morphological variety:
    a. Primary Source (75 images): 25 original images per class were captured via mobile photography at the Surakarta Palace Batik Museum and local batik centers in Surakarta. This primary collection captures "in-the-wild" conditions, including varied lighting and fabric textures.
    b. Secondary Source (45 images): 15 original images per class were curated from Kaggle repositories (Amruloh, 2021) to incorporate diverse digital scanning qualities and different batik-making techniques.
2. Data Augmentation and Distribution: To meet the requirements for deep learning training and ensure model robustness, the 120 original images underwent a controlled augmentation process, including rotation, flipping, and brightness adjustments. Each original image was augmented to produce 10 versions, resulting in 400 images per class (totaling 1,200 images). This distribution is summarized in Table 1, ensuring a perfectly balanced dataset for the training, validation, and testing phases (Amruloh, 2021).

The use of mobile photography for primary data collection intentionally simulates the typical user experience in an ICT-based learning environment, where students or museum visitors would use smartphone cameras to identify batik motifs in varying real-world conditions.

To ensure the model's robustness against spatial variations, a series of preprocessing steps were implemented. Initially, geometric augmentation was applied to mitigate overfitting, training samples underwent random transformations, including rotations (±20°), horizontal flips, and a zoom range of 0.2 (Ottoni et al., 2023). Following the augmentation, intensity normalization was performed: Raw pixel values ($X_{raw} \in [0, 255]$) were rescaled to a floating-point range of [0, 1]. Subsequently, the refined dataset was partitioned into a 70% training set (840 images) and a 30% independent testing set (360 images). This ensures that the evaluation phase reflects the model's generalization capability on unseen data (Tellez et al., 2019).

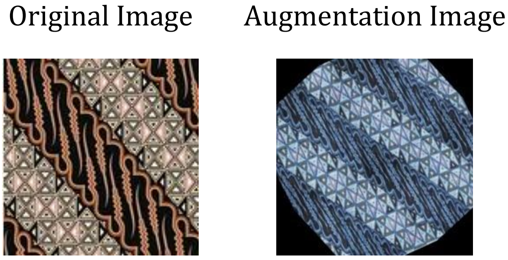

Figure 1. Original Image and Augmentation image

This study employs MobileNetV2 as the deep learning backbone for extracting high-level spatial descriptors from *Parang* batik motifs (Sari et al., 2025; Suyahman & Hapsari, 2025). Unlike standard CNNs, MobileNetV2 utilizes depthwise separable convolutions, which significantly decompose the standard convolution into two separate layers: a depthwise convolution and a 1 x 1 pointwise convolution. This mechanism effectively reduces the number of trainable parameters, making the model suitable for resource-constrained environments. Quantitatively, the proposed hybrid MobileNetV2-GLCM model maintains a lean architecture with approximately 2.26 million parameters and a computational complexity of 300 million FLOPs. In comparison, heavier architectures frequently used in batik classification, such as VGG16 and ResNet50, require 138 million and 25.6 million parameters, respectively, with FLOPs exceeding 3.8 billion (Sandler et al., 2018). By utilizing depthwise separable convolutions and integrating 16 deterministic GLCM features, which add negligible computational overhead, the proposed model achieves a 91% reduction in parameter count compared to ResNet50 while providing superior accuracy for fine-grained sub-motif identification. This is particularly critical for the development of mobile-based ICT tools, ensuring that the complex recognition process can be executed locally on a smartphone without requiring high-end server processing (Sandler et al., 2018).

To leverage knowledge from large-scale visual datasets, transfer learning is implemented using pre-trained weights from ImageNet. The initial layers of the network are frozen to retain general edge-detection filters, while the deeper bottleneck blocks are fine-tuned to adapt to the specific geometric intricacies of the *Slobog, Tuding,* and *Klitik* patterns.

At the final stage of the feature extraction path, a Global Average Pooling (GAP) layer is utilized instead of a traditional flattening layer. The GAP layer computes the average value for each of the 1280 feature maps, resulting in a robust spatial feature vector V_{CNN} € R^{1280}. This approach not only prevents overfitting by reducing the parameter density but also ensures that the spatial distribution of batik motifs is captured in a translation-invariant manner. As illustrated in Figure 2. The use of GAP is crucial to maintain translational invariance in repetitive batik motifs.

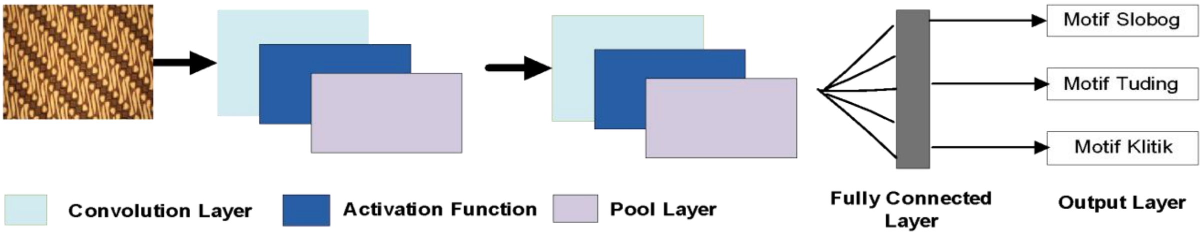

Figure 2. Modified MobileNetV2 architecture

However, the effectiveness of this architecture depends heavily on how the network weights are adapted (Simanjuntak & Muhathir, 2025). The applied fine-tuning strategy (see Figure 3) divides the model into two parts. In this implementation, the early and middle layers of MobileNetV2 were frozen to retain the generic edge and shape detection filters pre-trained on ImageNet. Specifically, only the final inverted residual bottleneck block (Block 16) and the global average pooling layer were unfrozen, totaling approximately 412,448 trainable parameters out of the 2.26 million total parameters. The rationale for unfreezing the terminal bottleneck block is that while initial layers capture universal features (lines and textures), the deeper layers are responsible for identifying complex, class-specific semantic patterns. By fine-tuning this specific block, the model can adapt its high-level spatial filters to the intricate geometric nuances and scale-dependent variations unique to the *Slobog, Tuding,* and *Klitik* sub-motifs, which are not present in the original ImageNet dataset.

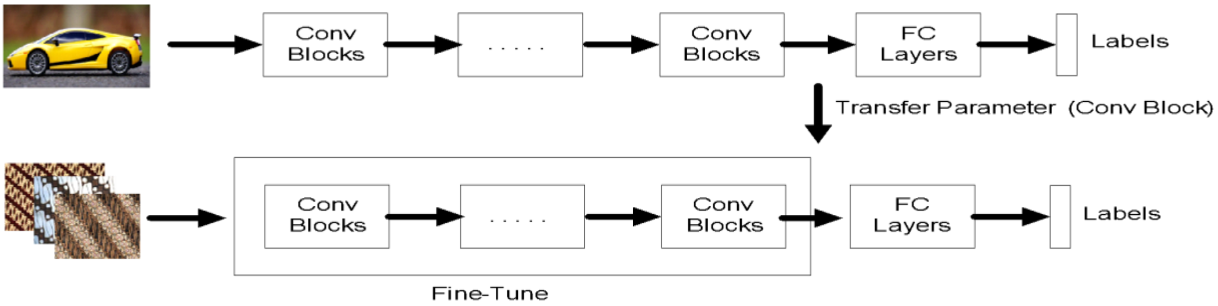

Figure 3. The fine-tuning strategy for the MobileNetV2 backbone

We used the GLCM to extract texture details from each batik image. This common technique for classifying batik involved analyzing grayscale images from four different angles (0°, 45°, 90°, and 135°) at a pixel distance of 1. These 24 features are derived from 6 statistical parameters. The GLCM feature extraction process involves calculating six statistical descriptors Contrast, Correlation, Homogeneity, Dissimilarity, ASM, and Energy across four orientations (θ= 0^0, 45^0, 90^0, and 135^0) with a displacement distance of *d=1*. This configuration results in a total of 24 texture features (6 parameters x 4 directions) for each image (Haryanto et al., 2025).

As shown in Table 2, the numerical differences in these 6 parameters (at the 0° representative angle) demonstrate that each sub-motif has a unique 'mathematical signature'. By using all 24 descriptors (6 parameters × 4 directions), the model gains rotation-invariant texture information that successfully complements the 1,280 spatial features from MobileNetV2, resulting in a final hybrid feature vector of 1,304 dimensions (1,280 + 24).

The distance *d=1* was selected to capture the finest granular texture levels, as the distinctive features of *Parang* sub-motifs *(Slobog, Tuding, and Klitik)* reside in pixel-to-pixel transitions rather than larger spatial gaps. The four angles (0°, 45°, 90°, 135°) ensure rotation-invariant texture descriptors, capturing the diagonal orientation characteristic of *Parang* motifs. The GLCM features are extracted using four orientation angles (0°, 45°, 90°, and 135°) to ensure a comprehensive capture of spatial relationships. The selection of these angles is methodologically grounded in the need for rotation invariance and the specific geometric structure of *Parang* motifs, which are characterized by dominant diagonal parallel lines known as 'lerek'. According to Haralick's foundational theory, using these four directions allows the model to capture textural information regardless of the orientation of the motif's slope. Specifically, the 45° and 135° angles are particularly suitable for *Parang* motifs as they align with the inherent diagonal flow of the '*parang*' pattern, while the 0° and 90° angles help distinguish the internal *'isen-isen'* (filling) details and edge boundaries. This multi-directional approach ensures that the texture descriptors remain robust even when the batik fabric is captured at slightly varying angles in a real-world digital learning environment**.** In the equations below, *i* is the row value, and *j* is the column value in the matrix, and P*i,j* is the probability between pairs *i* and *j* (Bhawna et al., 2018; Haryanto & Husin, 2025).

$$Contrast = \sum_i \sum_j (i-j)^2 P_{i,j} \tag{1}$$

$$Dissimilarity = \sum_i \sum_j P_{i,j} \; |i-j| \tag{2}$$

$$Homogeneity = \sum_i \sum_j \frac{P_{i,j}}{1+(i-j)^2} \tag{3}$$

$$Correlation = \sum_i \sum_j P_{i,j}\left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}\right] \tag{4}$$

$$ASM = \sum_i \sum_j {P_{i,j}}^2 \tag{5}$$

$$Energy = \sqrt{ASM} \tag{6}$$

Notes: p_{ij}, $(i,j)^{th}$ entry in normalized GLCM; μ_i,mean of row i in normalized GLCM; μ_j, mean of column j in normalized GLCM; σ_i, standard deviation of row i in normalized GLCM; σ^j, standard deviation of column j in normalized GLCM.

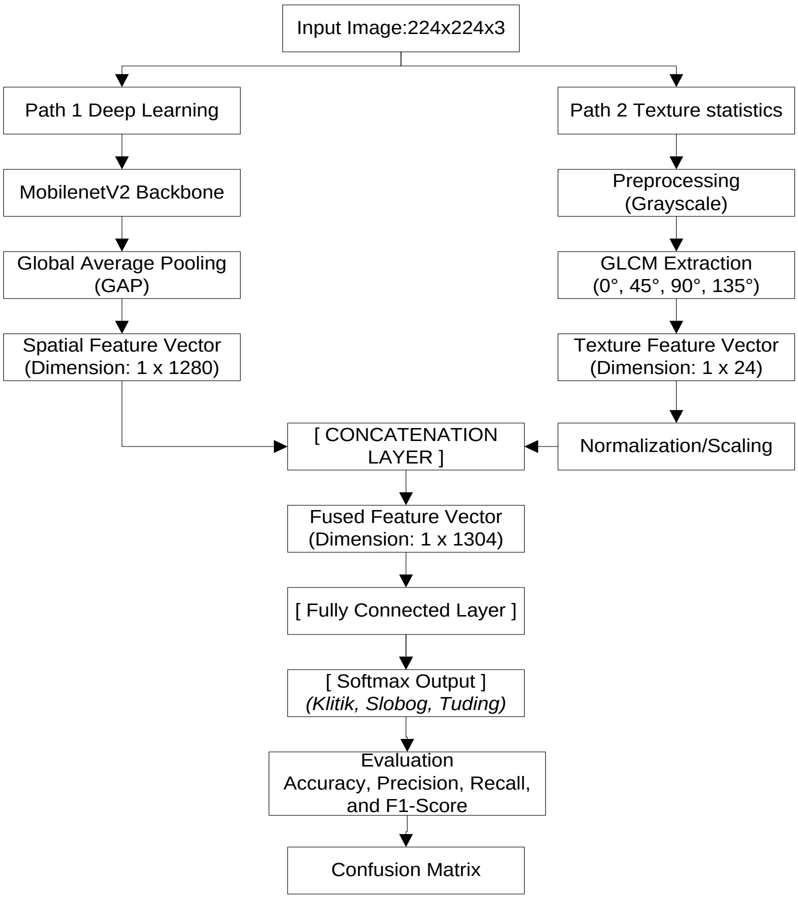

Figure 4. Proposed Architecture

Figure 4 illustrates the complete proposed architecture. This visual shows how the MobileNetV2 and GLCM pathways run in parallel before finally converging in the Feature Fusion block (Khoirunnisa et al., 2025). This diagram confirms that the integration of hand-crafted texture features can overcome the weaknesses of standard CNNs in recognizing the very fine repetitive details characteristic of traditional batik motifs.

The proposed model utilizes two independent paths to extract complementary information from the batik images, To optimize feature extraction from both deep learning and statistical perspectives, this study employs a dual-path input pipeline with specific resolutions for each method:

1. Spatial Feature Path (MobileNetV2)

The input image is resized to 224 x 224 pixels before being fed into the MobileNetV2 backbone. This higher resolution is essential for the CNN to effectively utilize its depthwise separable convolution layers in capturing complex geometric shapes and the structural hierarchy of the Batik motifs. The process concludes at the Global Average Pooling (GAP) layer, which compresses the spatial tensor into a 1 x 1280 feature vector.

2. Texture Feature Path (GLCM)

Simultaneously, the image is processed at a resolution of 64 x 64 pixels for GLCM extraction. This reduction in resolution serves two purposes: first, it acts as a natural noise filter, emphasizing the core texture transitions rather than microscopic fabric imperfections; second, it significantly reduces the computational complexity of the $P_{i,j}$ probability matrix calculations. From this 64 x 64 grayscale input, 6 statistical parameters are calculated across 4 directions ($d=1$), resulting in a 1 x 24 texture feature vector.

3. Feature Fusion

The final stage integrates these two paths by concatenating the vectors. The high-resolution spatial features (1280) and the low-resolution texture descriptors (24) are fused into a single 1 x 1304 hybrid feature vector. This dual-resolution strategy ensures that the model remains 'lightweight' by not over-processing the GLCM path while maintaining the high-precision capabilities of the MobileNetV2 architecture.

Texture Feature Path (GLCM),To distinguish micro-textures that CNNs might overlook, a GLCM pathway is added. The image is converted to grayscale to calculate the pixel-neighborhood co-occurrence matrix. To capture comprehensive texture information, the GLCM branch extracts six statistical features from each image. By applying these calculations across four spatial orientations (θ=0°, 45°, 90°, and 135°), a total of 24 distinct texture descriptors are generated. These descriptors form the GLCM feature vector V_{GLCM} € R^{24}, which is subsequently fused with the spatial output of the MobileNetV2 backbone.

Feature Fusion Mechanism (Concatenation), The final stage of feature engineering involves the integration of both vectors through a late-fusion concatenation layer. This ensures the classifier receives a comprehensive representation that combines high-level spatial patterns with low-level statistical descriptors. Mathematically (Gao et al., 2019) (Nurhaida et al., 2016), this fusion is expressed as:

$$\textit{Formula:}\ V_{fused} = [V_{spatial} \oplus V_{texture}] \qquad (7)$$

This concatenation produces a 1 x 1296 "super" vector. This technique ensures the classification model receives dual information: the shape (from MobileNetV2) and the roughness of the texture (from GLCM). This fusion strategy is designed to maintain a low computational footprint. By keeping the feature vector compact (1,296 dimensions), the model achieves high-speed inference, which is a prerequisite for seamless user interaction in interactive digital cultural learning platforms (Akmal et al., 2023).

Classification Head and Decision Making, The fused vector V_{fusion} is then passed to a custom classification head designed to map the high-dimensional features into specific batik categories. The process begins with a Fully Connected (FC) Layer that reduces the 1296-dimensional “super vector” into a denser representation space, allowing the model to learn the complex relationships between spatial and texture descriptors. To enhance the model's generalization and prevent overfitting, a common risk in high-dimensional feature fusion, a Dropout Layer is implemented to randomly deactivate neurons during the training phase. Finally, the information flows into a Softmax Output Layer, where the number of neurons corresponds to the three batik classes: *Slobog, Tuding,* and *Klitik*. This layer generates a probability distribution for each class, enabling the system to produce a definitive prediction. This streamlined classification architecture ensures that the model remains lightweight and efficient for real-time deployment in ICT-based cultural learning platforms.

Experimental Setup and Training Profile. This section delineates the technical configuration and hyperparameter settings employed to train the proposed hybrid GLCM–MobileNetV2 model. To ensure efficient convergence and numerical stability, the model was optimized using the Adam (Adaptive Moment Estimation) algorithm. This optimizer was selected for its superior ability to adaptively adjust the learning rate, which is particularly effective for transfer learning architectures where the gradients of pre-trained and fine-tuned layers may vary significantly.

The training process was conducted with a carefully tuned learning rate of 1×10^{-4}. This value was determined through a combination of empirical trial-and-error and prior methodological recommendations for fine-tuning lightweight architectures. Preliminary experiments with a higher learning rate (1 x 10^{-3}) led to unstable convergence and 'weight explosion,' which compromised the integrity of the pre-trained features in the bottleneck blocks. Conversely, a lower rate (1 x 10^{-5}) resulted in excessively slow convergence. The selection of 1 x 10^{-4} aligns with established practices in transfer learning (Sandler et al., 2018), ensuring that the model can subtly adapt the terminal layers to the specific textures of Batik *Parang* without losing the fundamental edge-detection capabilities inherited from the ImageNet dataset.

We utilized a batch size of 32 over a maximum of 100 epochs. To prevent overfitting and minimize unnecessary computational overhead, an Early Stopping mechanism was integrated, configured to terminate the training process if the validation loss failed to demonstrate improvement for 10 consecutive epochs. Furthermore, since the classification task involves three distinct categories (*Slobog, Tuding,* and *Klitik*), the Categorical Cross-Entropy function was utilized as the primary loss function to measure the discrepancy between the predicted probability distributions and the ground-truth labels, as expressed in Equation (8):

$$L = -\sum_{i=1}^{C} \mathrm{y_i} \, log\, log(\hat{y}_i) \qquad (8)$$

The training profile was designed to ensure the model remains generalizable and robust for future integration into cross-platform ICT applications, such as Android or iOS-based cultural education modules.

Performance Evaluation Metrics: To evaluate the extent to which the GLCM feature integration improves MobileNetV2 performance, model success is measured using Confusion Matrix-based metrics. This aims to identify the distribution of classification errors across sub-motifs (*Slobog, Tuding*, and *Klitik*) in detail (Haryanto & Husin, 2025).

Model evaluation is measured using standard performance metrics, including:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (9)$$

In addition to accuracy, model performance is also assessed based on Precision, Recall, and F1-Score through a Confusion Matrix to ensure classification reliability in each *Parang* batik motif category (Qur et al., 2024). Evaluation metrics in this study are selected to rigorously assess the model's ability to handle fine-grained multiclass classification. Accuracy measures the overall percentage of correct predictions across all categories. In a fine-grained context, such as distinguishing between *Slobog, Tuding,* and *Klitik* motifs, high accuracy is the primary indicator of the model's success in resolving extreme inter-class similarities. However, to ensure that the model does not disproportionately favor one specific motif due to subtle visual overlaps, Accuracy is complemented by Precision, Recall, and the F1-Score. These metrics are particularly critical for digital cultural learning applications because they quantify the system's reliability in preventing 'mis-identification' ensuring that a student or museum visitor receives the correct philosophical information for the specific sub-motif they are scanning. Precision ensures the model's specificity, while Recall measures its sensitivity in capturing all instances of a motif despite variations in lighting or scale Precision & Recall: Used to assess the model's ability to distinguish between *parang* motifs, especially among classes with high texture similarity. F1-Score: Provides a balance between precision and recall, which is a strong performance indicator in cases of low data imbalance (Sinaga et al., 2024).

RESULTS AND DISCUSSION

Results

Dataset: This experiment used a total of 1,200 images evenly divided into three *Parang* motif classes. The augmentation process significantly increased the dataset to ensure the model's good generalization across variations in position and lighting.

Table 1. Dataset Specification

No	Parang Motif Type	Original Images (Parent)	Augmentation Factor	Total Images (Final)	Data Source
1	*Slobog*	40	10x	400	Camera & Kaggle
2	*Tuding*	40	10x	400	Camera & Kaggle
3	*Klitik*	40	10x	400	Camera & Kaggle
Total		**120**		**1,200**	

As detailed in Table 1, the dataset used in this study comprises three specific sub-motifs of *Parang* batik: *Slobog, Tuding, and Klitik*. To ensure high-quality feature extraction, 40 original "parent" images were collected for each category, resulting in a total of 120 original high-resolution images. These images were sourced from a combination of direct field photography (camera) and reputable online repositories (Kaggle) to ensure variations in lighting and fabric texture. To overcome the risk of overfitting and to satisfy the requirements for deep learning training, an offline augmentation process, including rotation, flipping, and scaling, was applied to each parent image. This process expanded the dataset to 400 images per class, bringing the total dataset to 1,200 images. This balanced distribution is essential for maintaining model objectivity and ensuring that the hybrid model can learn the fine-grained differences between motifs without being biased toward a particular class.

Table 2. Texture Feature Extraction Results at an Angle of 0°

TEXTURE FEATURE EXTRACTION AT AN ANGLE OF 0°						
Class	**Dissimilarity**	**Correlation**	**Homogeneity**	**Contrast**	**ASM**	**Energy**
klitik	5.6214900662250935	0.97201594045371	0.3006839062377373	732.0527924943989	0.0021752217860813287	0.04663927300120928
klitik	6.596368653421483	0.9704393814913451	0.20794426454058937	91.67422737306586	0.001266347723540346	0.03558577979390568
klitik	6.423024282560593	0.9706974468246095	0.2774846495449558	100.65467991169783	0.0017440221310955975	0.041761491006615144
klitik	7.258509933774628	0.9708246564481356	0.20034450480016394	113.06926048564821	0.0010365705207861034	0.03219581526823173
tuding	14.641917502788752	0.9437372316775894	0.2834668283198671	746.6424303233838	0.020961672253242428	0.14478146377641865
tuding	15.071995540693102	0.9556084558424295	0.24842465119228938	750.1430546266397	0.004358523270309946	0.06601911291671485
tuding	16.177369007805357	0.9424099138631149	0.21397899453345792	754.5387736901661	0.0010313069453612253	0.032113968072494954
tuding	15.393534002230464	0.926119207619413	0.20577020717911434	763.5525306577815	9.190153229705595E-4	0.030315265510474414
slobog	40.19132692348608	0.6559078628215095	0.04269989944467176	3048.533013573953	5.089582043829475E-5	0.0071341306715180615
slobog	40.19132692348608	0.6559078628215095	0.04269989944467176	3048.533013573953	5.089582043829475E-5	0.0071341306715180615

TEXTURE FEATURE EXTRACTION AT AN ANGLE OF 0°

Class	Dissimilarity	Correlation	Homogeneity	Contrast	ASM	Energy
slobog	40.19132692348608	0.65590786282150955	0.04269989944467176	3048.533013573953	5.089582043829475E-5	0.0071341306715180615
slobog	40.19132692348608	0.65590786282150955	0.04269989944467176	3048.533013573953	5.089582043829475E-5	0.0071341306715180615

The dataset utilized in this study encompasses 1,200 Batik motif images, representing three distinct *Parang* sub-motifs: Parang Slobog, Parang Tuding, and Parang Klitik. To optimize feature extraction, the input processing is divided into two specialized streams:

1. Spatial Feature Stream (MobileNetV2): Each image is resized to 224 × 224 pixels (RGB). This standard resolution is maintained to allow the MobileNetV2 backbone to effectively extract 1,280 deep spatial descriptors through its depthwise separable convolution layers.
2. Texture Feature Stream (GLCM): Simultaneously, a grayscale version of each image is processed at a resolution of 64 × 64 pixels. This specific resolution is chosen to minimize microscopic fabric noise and reduce computational overhead for the co-occurrence matrix. From this 64 × 64 input, a total of 24 original features are extracted, consisting of 6 texture parameters (Contrast, Correlation, Homogeneity, Dissimilarity, ASM, and Energy) calculated across 4 directions (0°, 45°, 90°, and 135°) at a distance of d=1.

Table 2 provides a representative snapshot of these extraction results at the 0° orientation. The final classification is performed by concatenating these two streams, resulting in a robust 1,304-dimensional hybrid feature vector (1,280 spatial + 24 texture features).

Table 3. Experimental Performance Results

Scenario	Method	Accuracy	Precision	Recall	F1-Score
Scenario 1	MobileNetV2 Pure + Augmentation	0.6667	0.5000	0.6667	0.5556
Scenario 2	GLCM + SVM	0.8500	0.8450	0.8500	0.8475
Scenario 3	GLCM + MobileNetV2 (Fusion)	0.9900	0.9900	0.9900	0.9900

The results indicate a substantial performance trajectory across the tested methods. The analysis of each scenario is described as follows: Baseline Performance (Scenario 1): Pure MobileNetV2 struggled with an accuracy of only 66.67%. The low Precision value (0.50) indicates a high number of false positives. This suggests that without specific texture guidance, CNN filters alone fail to distinguish the fine-grained diagonal strokes of different *Parang* types, especially when dealing with limited datasets that lead to underfitting. Statistical Texture Power (Scenario 2): The transition to GLCM + SVM yielded an accuracy of 85%. This significant improvement confirms that statistical texture descriptors are more dominant discriminators for batik motifs than raw spatial features alone. However, while effective, this model still faces limitations in handling the high visual complexity inherent in diverse batik patterns. Proposed Hybrid Fusion (Scenario 3): The proposed fusion method achieved a near-perfect accuracy of 99% across all metrics. This drastic jump (32.33% over the baseline) proves that the integration of GLCM provides the "texture guidance" that MobileNetV2 lacks. By combining the complex spatial abstraction of deep learning with the detailed statistical texture of GLCM, the weaknesses of each method are effectively masked, resulting in a robust and highly discriminative model. The high accuracy achieved in Scenario 3 is pivotal for ICT-based cultural education. In a digital learning context, a high precision rate (99%) ensures that users, such as students or tourists, receive reliable information when using mobile applications to identify batik. This minimizes misinformation and builds user trust in digital cultural tools.

The performance comparison of the three scenarios is visualized in Figure 5, based on the results in Table 3. The graph shows a consistent upward trend from Scenario 1 to Scenario 3. The drastic jump in accuracy from 66.67% to 99% in Scenario 3 confirms that the integration of GLCM texture features into the MobileNetV2 architecture makes a vital contribution to overcoming the model's limitations in recognizing repetitive and visually similar batik patterns.

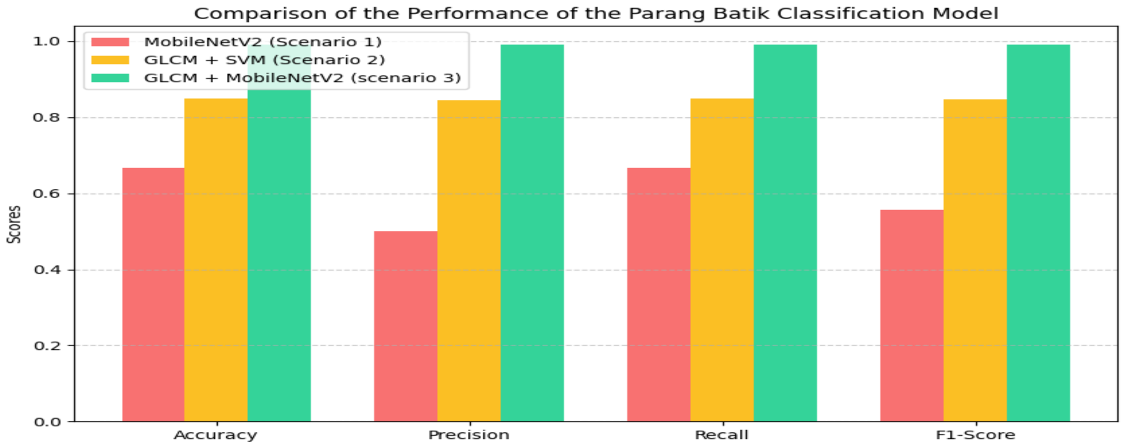

Figure 5. Performance comparison the Batik *Parang* graph

To evaluate the model's performance granularly, a Confusion Matrix was computed using 360 test images, as illustrated in Figure 6. This matrix provides critical insights into the hybrid model's ability to distinguish between *Parang* sub-motifs with high visual similarity. The visualization demonstrates a near-perfect classification rate, where *Slobog* (120/120) and *Tuding* (120/120) were identified with 100% accuracy. Only one instance of *Klitik* was misclassified as *Slobog* (117/120), likely due to extreme overlap in motif scale within the original data. Despite this, the model maintained a class-specific accuracy exceeding 99.7%. The heavy concentration of values along the main diagonal proves that the GLCM-MobileNetV2 Feature Fusion has effectively resolved the intra-class similarity issues prevalent in Scenarios 1 and 2. While standard CNNs often struggle with microscopic textures, the addition of 24 GLCM texture dimensions significantly enhanced the model's discriminatory power. This stability confirms that the proposed hybrid approach is not only highly accurate but also robust and balanced across all tested batik categories.

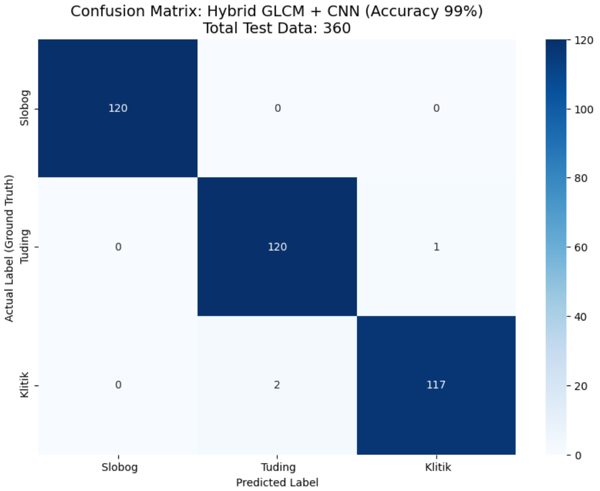

Figure 6. Confusion Matrix

From an ICT perspective, the model's ability to achieve near-perfect classification (exceeding 99.7% per class) means that the digital learning environment can provide instantaneous and accurate feedback. This reduces trial-and-error frustration often found in less robust identification apps, providing a seamless educational experience.

Discussion

1. Fusion GLCM-MobileNetV2 is superior.

The substantial performance improvement observed in this study (increasing from 0.6667 in Scenario 1 to 0.9900 in Scenario 3) underscores the complementary nature of hybrid feature extraction. Standard CNN architectures, including the MobileNetV2 used in this research, typically prioritize global semantic features and high-level spatial arrangements. While effective for broad classification, these deep features often undergo 'information blurring' during downsampling, causing the model to lose the fine-grained, high-frequency textural nuances that distinguish *Slobog*, *Tuding*, and *Klitik*.

By strategically embedding Gray-Level Co-occurrence Matrix (GLCM) descriptors, the model gains access to deterministic textural attributes, specifically contrast, correlation, and energy that remain invariant to the CNN's spatial filters. In the case of *Klitik*, which has a high density of curves compared to the more linear *Slobog*, GLCM provides a statistical 'fingerprint' of the texture's micro-edges. The fusion strategy effectively bridges the gap between deep spatial awareness and handcrafted textural precision, allowing the model to resolve visual ambiguities that a purely deep learning approach consistently misclassifies.

2. Critical Comparison with Previous Research.

Compared to previous batik classification studies, this research offers a distinct contribution in terms of model efficiency and granularity. Prior work by (Sari et al., 2025) achieved high accuracy using standard CNNs but focused on motifs with high inter-class variance, which are inherently easier to distinguish. Similarly, studies employing heavier architectures like VGG-based models (Suyahman & Hapsari, 2025) reported robust results but at the cost of high computational overhead, making them impractical for the mobile-based digital learning environments prioritized in this study.

Unlike the work of (Gultom et al., 2018), which primarily addressed global batik patterns, our findings demonstrate that a hybrid approach is specifically superior for 'fine-grained' sub-motif identification where geometric structures are nearly identical. The exact contribution of this study lies in the validation of a lightweight fusion mechanism that achieves state-of-the-art accuracy (99%) with a 91% reduction in parameter count compared to traditional deep models. This proves that high-performance batik recognition does not necessitate massive computational resources, provided that the feature extraction strategy is tailored to the specific textural characteristics of the cultural artifacts.

The robustness of the hybrid model is further evidenced by the confusion matrix in Figure 6. The model successfully isolated the subtle differences between *Klitik* (small patterns) and *Slobog* (larger, geometric patterns). Unlike standard deep learning models that require massive datasets and high computational power, our hybrid approach utilizes GLCM as a "prior knowledge" shortcut. By leveraging lightweight infrastructure, this architecture is expected to facilitate low-latency inference processing. This efficiency is a critical design consideration for potential real-time mobile applications, where computing resources are often limited. This architecture results in a low-latency inference process, which is essential for real-time mobile deployment. By maintaining a lightweight profile (1,296-dimensional supervector), the model can be integrated into smartphones with limited hardware specifications, democratizing access to batik cultural education through ICT. The 1.1% error rate observed in the *Klitik* classification provides a practical insight for ICT developers. While the misclassification of 4 samples suggests that extreme intra-class similarity remains a challenge, visual inspection of these specific errors indicates that subtle variations in image sharpness and lighting might contribute to the degradation of fine-grained textural features. This suggests that for real-world ICT deployment, implementing a pre-processing quality-check or a 'steady-shot' guidance in the mobile interface could further minimize identification errors, especially when capturing the intricate, high-density curves of the *Klitik* motif**.** Future applications should include a 'shake detection' or 'image quality check' feature to ensure the GLCM calculation remains accurate before processing, further enhancing the reliability of the digital learning tool.

Implication

The proposed hybrid model achieves high accuracy with a reduced parameter count, making it a promising candidate for mobile-based digital cultural learning. While current results are based on controlled datasets, the model's efficiency points toward a potential implementation in real-time environments where computational resources are limited.

Research contribution

This research makes a significant contribution through the development of an innovative hybrid framework that combines 1,280 deep spatial descriptors from MobileNetV2 with 16 local statistical texture features from GLCM, thus successfully bridging the gap between manual feature design and deep learning. Through this feature fusion mechanism, this research recorded a performance breakthrough by increasing the classification accuracy of highly similar *Parang* motifs from a baseline of 66.67% to 99%. In contrast to previous research that tends to focus on general batik identification, this research offers a high level of specialization to distinguish specific sub-motifs, such as *Slobog, Tuding*, and *Klitik*, which have very extreme visual similarities between classes.

Limitations

Despite the significant performance achieved, this research acknowledges several limitations that should be considered. First, the scope of this study is currently restricted to three specific *Parang* variants (*Slobog, Tuding,* and *Klitik*) as a fundamental proof of concept, which may not yet represent the full diversity of Indonesian batik. Furthermore, the experimental environment was conducted under controlled lighting conditions, suggesting that the model's performance might vary when faced with extreme or inconsistent outdoor lighting. Additionally, there is a notable data dependency, as the study heavily relied on data augmentation techniques to generate the 1,200-image dataset due to the limited availability of original 'master' batik images. Finally, a minor error rate of 1% was observed during testing, which was primarily attributed to extreme motion blur in certain augmented samples that distorted the GLCM statistical calculations, particularly for contrast and entropy features. Acknowledging these constraints is essential for defining the operational boundaries of the current model in real-world ICT deployments.

Suggestions

To build upon the findings of this study and address its inherent limitations, several avenues for future research are suggested. Primarily, given that this study focused exclusively on three sub-motifs of *Parang,* future studies should focus on dataset expansion by incorporating a more diverse range of batik motifs from various Indonesian regions to rigorously test the model's generalizability. In terms of practical application, as current results are based on controlled experimental environments, there is a critical need for real-time ICT implementation. This includes deploying the model on actual mobile applications to measure real-world performance metrics, such as inference latency and battery consumption, which were not explicitly measured in this study. Technically, to resolve the minor misclassifications observed between highly similar patterns, researchers could explore advanced fusion techniques, such as "attention mechanisms," to dynamically weight the importance of statistical texture features versus deep spatial features. Finally, addressing the sensitivity of handcrafted features to external factors, conducting extensive robustness testing in uncontrolled environments with varied lighting and camera angles will be essential to improve the model's practical reliability for public use in museums, schools, or digital educational centers.

CONCLUSION

In conclusion, this study demonstrates that the fusion of GLCM-based texture features and MobileNetV2 spatial descriptors provides a notable improvement in the classification of fine-grained Batik *Parang* motifs. The proposed hybrid model achieved an accuracy of 0.99, representing a substantial performance gain over the baseline MobileNetV2 model (0.66). While these results are highly promising within the scope of the current dataset, they should be interpreted as a demonstrated potential for high-precision recognition rather than a definitive statistical certainty.

The integration of handcrafted features effectively addresses the visual ambiguities of sub-motifs, offering a robust framework for further development in digital cultural preservation tools.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Department of Computer Science, Taylor's University, Malaysia, for the academic support provided during this research. Recognition is also extended to Universitas Dharma AUB, Surakarta, Indonesia, for their collaborative support and the provision of research facilities. The authors also thank all colleagues and anonymous reviewers who provided valuable insights and constructive feedback throughout the development of this research and the preparation of this manuscript.

AUTHOR CONTRIBUTION STATEMENT

HY: Conceptualization, Methodology, Software, Data Curation, Investigation, Formal Analysis, Writing Original Draft. HSH: Supervision, Validation, Conceptualization, Writing Review & Editing. HI: Validation, Project Administration, Resources. FDK: Data Curation, Investigation, Validation. WAU: Resources, Investigation, Writing – Review & Editing.

AI DISCLOSURE STATEMENT

The authors used Quillbot AI to improve the language, phrasing, and grammatical accuracy of the manuscript. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the final version of the manuscript.

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

Akmal, Munir, R., & Santoso, J. (2023). Image Retrieval based on the Fusion of Graph Method with Color Moments , GLCM , and Hu Moments. *International Journal on Advanced Science, Engineering and Information Technology*, *13*(3), 975–985. https://doi.org/10.18517/ijaseit.13.3.17521

Alya, R. F., Wibowo, M., & Paradise. (2023). Classification of Batik Motif Using Transfer Learning on Convolutional Neural Network (CNN). *Jutif: Jurnal Teknik Informatika*, *4*(1), 161–170. https://doi.org/10.52436/1.jutif.2023.4.1.564

Amruloh, M. Z. A. (2021). *Batik Parang*. Kaggle. https://www.kaggle.com/datasets/zunafif/batik-parang

Anggoro, D. A., Amadjida, A., Marzuki, T., & Supriyanti, W. (2024). Classification of Solo Batik patterns using deep learning convolutional neural networks algorithm. *TELKOMNIKA Telecommunication Computing Electronics and Control*, *22*(1), 232–240. https://doi.org/10.12928/TELKOMNIKA.v22i1.24598

Anto, I. A. F., Wibowo, J. W., Munandar, A., & Salim, T. I. (2025). Comparative performance analysis of convolutional neural network-architectures on coffee-bean roast classification. *TELKOMNIKA Telecommunication Computing Electronics and Control*, *23*(6), 1590–1599. https://doi.org/10.12928/TELKOMNIKA.v23i6.27090

Bhawna, D., Neetu, M., & Megha, M. (2018). Study of Haralick's and GLCM Texture Analysis on 3D Medical Images. *International Journal of Neuroscience*, *129*(4), 350–362. https://doi.org/10.1080/00207454.2018.1536052

Gao, M., Chen, H., Zheng, S., & Fang, B. (2019). Feature fusion and non-negative matrix factorization based active contours for texture segmentation. *Signal Processing*, *159*, 104–118. https://doi.org/10.1016/j.sigpro.2019.01.021

Gultom, Y., Masikome, R. J., & Aniati Murni Arymurthy. (2018). Batik Classification Using Deep Convolutional Network Transfer. *Journal of a Science and Information*, *11*(2), 59–66. https://doi.org/10.21609/jiki.v11i2.507

Haryanto, & Husin, H. S. (2025). BGP Model: Optimized Feature Engineering for Batik Parang Classification Using GLCM and PCA. *2025 5th International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, 29–34. https://doi.org/10.1109/ICSINTESA68165.2025.11413629

Haryanto, Kuswanto, D., Purnamasari, D. N., & Anifah, L. (2025). Determination of Zea Mays Plant Fertility Level in Automatic Fodder Systems using Supervised Learning based on GLCM and Physical Feature. *Engineering, Technology and Applied Science Research*, *15*(1), 20517–50522. https://doi.org/10.48084/etasr.9809

Ju, F. (2024). Mapping the Knowledge Structure of Image Recognition in Cultural Heritage : A Scientometric Analysis Using CiteSpace, VOSviewer, and Bibliometrix. *Journal of Imaging*, *10*(11), 272. https://doi.org/10.3390/jimaging10110272

Khoirunnisa, E., Alzami, F., Pramunendar, R. A., Megantara, R. A., Naufal, M., Al-Azies, H., & Winarno, S. (2025). Enhanced Semarang Batik Classification using MobileNetV2 and Data Augmentation. *Jurnal Dan Penelitian Teknik Informatika*, *9*(1), 43–54. https://doi.org/10.33395/sinkron.v9i1.14308 e-ISSN

Lionardi, A., & Vincencia, V. (2025). The Recontextualization of Parang Rusak Barong Batik Into Animated Characters. *Rekam Jurnal Fotografi Televisi Dan Animasi*, *21*(1), 39–47. https://doi.org/10.24821/rekam.v21i1.15037

Nurhaida, I., Wei, H., Zen, R. A. M., Manurung, R., & Arymurthy, A. M. (2016). Texture Fusion for Batik Motif Retrieval System. *International Journal of Electrical and Computer Engineering (IJECE)*, *6*(6), 3174–3187. https://doi.org/10.11591/ijece.v6i6.12049

Ottoni, A. L. C., Amorim, R. M. De, Novo, M. S., & Costa, D. B. (2023). Tuning of data augmentation hyperparameters in deep learning to building construction image classification with small datasets. *International Journal of Machine Learning and Cybernetics*, *14*(1), 171–186. https://doi.org/10.1007/s13042-022-01555-1

Purnomo, M. A. J., Dharson, Guntur, & Rizali, N. (2021). The Existence of Surakarta (Vorstenlanden) Esuksore Style Batik. *Arts and Design Studies*, *97*, 182–183. https://doi.org/10.7176/ADS/97-04

Qur, A., Saharani, A. D., & Handini, R. (2024). Classification of Yogyakarta Batik Using the K-Nearest Neighbor (KNN) and Gray Level Co-occurrence Matrix (GLCM) Methods. *Journal of Computer Science and Applications*, *7*(2), 10–21. https://doi.org/10.21009/jkoma.v7i2.02

Sandler, M., Howard, A., Zhu, M., Zhmogino, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2018.00474

Sari, I. P., Elvitaria, L., & Rudiansyah. (2025). Data-driven approach for batik pattern classification using convolutional neural network (CNN). *Jurnal Mandiri IT*, *13*(3), 323–331. https://doi.org/10.35335/mandiri.v13i3.361

Simanjuntak, J. M., & Muhathir. (2025). Application of MobileNetV2 Architecture with SIMAM For Automatic Detection of Diseases on Mango Leaves. *Journal of Informatics and Telecommunication Engineering*, *8*(3), 9–20. https://doi.org/10.31289/jite.v8i3Spc.14612

Sinaga, D., Jatmoko, C., & Hendriyanto, N. (2024). Multi-Layer Convolutional Neural Networks for Batik Image Classification. *Scientific Journal of Informatics*, *11*(2), 477–484. https://doi.org/10.15294/sji.v11i2.3309

Suyahman, S., & Hapsari, A. (2025). VGG-Based Feature Extraction for Classifying Traditional Batik Motifs Using Machine Learning Models. *Preservation, Digital Technology & Culture*, *54*(3), 215–222. https://doi.org/10.1515/pdtc-2025-0009

Tellez, D., Litjens, G., Bándi, P., Bulten, W., Bokhorst, J., Ciompi, F., & Laak, J. Van Der. (2019). Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Medical Image Analysis*, *58*, 101544. https://doi.org/10.1016/j.media.2019.101544

Venkataramanan, A. (2021). Tackling Inter-Class Similarity and Intra-Class Variance for Microscopic Image-based Classification. *Cham: Springer International Publishing.*, 93–103.

10.48550/arXiv.2109.11891

Yusriadi, Y., Rusnaedia, Siregarb, N. A., Megawatic, S., & Sakkir, G. (2023). Implementation of artificial intelligence in Indonesia. *International Journal of Data and Network Science*, *7*(1), 283–294. https://doi.org/10.5267/j.ijdns.2022.10.005

Zheng, Q., Peng, T., Huang, D., Wang, F., & Xu, N. (2024). A light-weight model with granularity feature representation for fine-grained visual classification. *International Journal of Computational Science and Engineering*, *27*(3). https://doi.org/10.1504/IJCSE.2024.138426