Journal of Mathematics and Applied Statistics
Vol. 3, No. 2, June 2025
DOI:

Attention-Enhanced Convolutional Networks for Fine-Grained Batik Motif Classification with Statistical Feature Modeling

Nurul Mukhlisah Abdal [1*], Tangsi[2]

[1] *Universitas Negeri Makassar, Sulawesi Selatan, Indonesia*
[2] *Universitas Negeri Makassar, Sulawesi Selatan, Indonesia*

**Correspondent Email: nm.abdal@unm.ac.id*

ABSTRACT

This study examines a hybrid method for classifying fine-grained Indonesian batik motifs under limited data conditions. The research focuses on two objectives: (1) assessing the contribution of attention mechanisms to the extraction of discriminative visual features, and (2) evaluating the role of Gray-Level Co-occurrence Matrix (GLCM) texture descriptors when combined with deep convolutional representations. The proposed approach employs a ResNet-50 backbone equipped with a Convolutional Block Attention Module (CBAM) and integrates second-order GLCM features through a feature-fusion framework. The dataset consists of authentic batik photographs representing 38 motif categories. Model performance is assessed using accuracy, macro-averaged metrics, Cohen's Kappa, and ablation experiments supported by statistical tests. The model reaches a test accuracy of 75.90%, with a macro F1-score of 0.7598 and a Cohen's Kappa value of 0.7456. Training and validation curves show stable behavior after the initial epochs. Per-class evaluation indicates that motifs with distinctive structural elements tend to be classified correctly, whereas motifs with subtle or overlapping patterns exhibit lower accuracy. The ablation study records a 4.79% accuracy increase attributed to CBAM and a 3.51% increase associated with GLCM features; both effects fall within statistically significant confidence intervals. The combination of both components yields an 8.38% improvement over the baseline model. Two-way ANOVA identifies main effects for attention and GLCM, with a small interaction term. These results provide information on how spatial attention and statistical texture features contribute to the classification of fine-grained batik motifs within the examined setting.

Keyword: Batik motif classification; Attention mechanisms; Convolutional neural networks; GLCM texture features; Fine-grained image analysis

Article History:
Received, May 2025; Revised, May 2025; Accepted, June 2025.

1. Introduction (Time New Roman, size 14)

Indonesian batik is internationally recognized as an intangible cultural heritage that encompasses artistic, philosophical, and historical values across thousands of motif variations. These motifs often exhibit highly similar geometric structures and subtle textural patterns, making manual identification challenging even for expert practitioners. As digital transformation accelerates within cultural preservation, fashion technology, and intelligent commerce, automated batik motif classification has emerged as a critical task. However, the structural complexity and fine-grained visual similarities among motifs position batik recognition as a fine-grained classification problem, characterized by high inter-class similarity and substantial intra-class variability arising from dye intensity variations, fabric deformation, and acquisition noise (Putra et al., 2024).

A wide range of computational approaches has been developed to address this challenge. Traditional machine learning methods have relied on statistical descriptors—most notably Gray-Level Co-occurrence Matrix (GLCM) and Moment Invariants (MI)—which effectively capture low-level textural and shape information. Studies combining GLCM and MI with

classical classifiers such as Support Vector Machines have reported competitive performance on small-scale datasets. Nevertheless, their representational capacity is inherently limited, and these models struggle to generalize to motifs with intricate structures or subtle discriminative cues (Iqbal et al., 2021).

Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have significantly improved visual recognition through hierarchical feature learning. Despite their success, CNNs trained from scratch typically require large annotated datasets and are prone to overfitting in the low-data scenarios common in batik research, as observed in studies on Lombok Songket and Sasambo motifs. Transfer learning approaches using architectures such as VGG16 and MobileNet have offered better accuracy and practicality. Yet, these models predominantly rely on global deep features and lack explicit mechanisms to emphasize motif-specific discriminative regions, which are essential for fine-grained classification (Mardani et al., 2020).

Attention mechanisms have recently emerged as powerful components capable of adaptively weighting spatial or channel-wise information within neural networks. Their effectiveness has been demonstrated across image recognition, pattern modeling, and image generation tasks. However, the application of attention mechanisms to batik motif classification remains limited. Moreover, the integration of attention with statistical feature modeling—particularly GLCM-based descriptors—has not been systematically explored, leaving a promising methodological gap in combining deep hierarchical features with mathematically grounded texture statistics (Dewi, 2023).

These observations highlight several key research gaps: (1) the absence of attention modules specifically designed to enhance fine-grained batik recognition; (2) the lack of hybrid frameworks that fuse deep convolutional features with statistical texture representations such as GLCM; (3) the scarcity of interpretability analyses revealing which regions or channels contribute most to motif discrimination; and (4) the limited availability of ablation studies quantifying the individual and joint contributions of attention mechanisms and statistical descriptors (Yuniarno & Purnomo, 2018).

To address these gaps, this study investigates the synergy between attention mechanisms, convolutional neural networks, and statistical texture modeling for fine-grained batik motif classification. The research is guided by two primary questions:

RQ (1): How do attention mechanisms improve classification accuracy for fine-grained Indonesian batik motifs under limited data conditions?

RQ (2): How does the addition of GLCM features affect the model performance?

The objectives of this study are threefold. First, this work aims to examine how attention mechanisms influence the classification performance of fine-grained Indonesian batik motifs, particularly under limited data conditions. Second, it seeks to investigate the contribution of GLCM-based statistical texture features to the overall model performance when integrated with deep feature representations. Finally, the study aims to evaluate the proposed method through standard performance metrics and statistical significance tests to provide a rigorous assessment of its effectiveness (UNESCO, 2009).

The contributions of this work are fourfold: (1) the development of an integrated architecture combining attention-enhanced ResNet with GLCM-based statistical features for fine-grained batik classification; (2) the introduction of a feature fusion strategy that integrates

deep convolutional and statistical texture representations; (3) comprehensive ablation studies quantifying the contributions of attention modules and GLCM features; and (4) interpretability analysis using Grad-CAM to identify motif-specific discriminative regions. Collectively, these contributions demonstrate measurable performance improvements over traditional machine learning methods, vanilla CNNs, and transfer learning baselines (Tan et al., 2020).

2. Method

The fine-grained classification of batik motifs is formulated as a supervised learning problem over a dataset

$$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N}, \tag{1}$$

where each RGB imagwhere each RGB image $x_i \in \mathbb{R}^{H\times W\times 3}$is assigned to a class label $y_i \in \{1, \dots, C\}$. The objective is to learn a function

$$f: \mathbb{R}^{H\times W\times 3} \to \mathbb{R}^{C} \tag{2}$$

that minimizes the expected risk

$$R(f) = \mathbb{E}_{(x,y)\sim P}[\ell(f(x), y)], \tag{3}$$

where P is the underlying data distribution and ℓ is the classification loss function. Because the dataset lies in a high-dimensional, low-sample-size (HDLSS) regime, where the number of samples per class is much smaller than the dimensionality, the model requires strong regularization and statistically grounded feature representations to obtain a stable estimator $\hat{f}$ (Mengiste et al., 2024).

The dataset contains N images distributed across C motif classes, exhibiting class imbalance. All images are normalized using standard mean and variance statistics, then partitioned into training, validation, and test sets via stratified sampling to preserve class proportions. Data augmentation is applied through geometric and photometric transformations to produce an enriched empirical distribution and reduce model variance (Xu et al., 2024).

The proposed model integrates deep hierarchical features with second-order statistical texture descriptors. For deep features, a ResNet backbone is combined with channel–spatial attention mechanisms. Given a feature map $F \in \mathbb{R}^{C\times H\times W}$, the channel attention mechanism computes (Gultom et al., 2018)

$$M_c(F) = \sigma(g(\mathrm{Avg}(F)) + g(\mathrm{Max}(F))), \tag{4}$$

where $\mathrm{Avg}(F)$ and $\mathrm{Max}(F)$ denote global average and max pooling across spatial dimensions, $\sigma$is the sigmoid activation function, and $g(\cdot)$ is a two-layer multilayer perceptron (MLP) used to learn channel-wise importance. The output is refined by spatial attention, defined as (Garay, n.d.)

$$M_s(F') = \sigma(h([\mathrm{Avg}(F'); \mathrm{Max}(F')])), \tag{5}$$

where $F' = F \odot M_c(F)$, $h(\cdot)$ is a 7×7 convolution operator, $[\cdot;\cdot]$ denotes channel-wise concatenation, and $\odot$ represents element-wise multiplication. The final attention-weighted representation is $F'' = F' \odot M_s(F')$. Global average pooling produces the deep embedding vector F_{deep}. To capture complementary textural information, statistical features are extracted from Gray-Level Co-occurrence Matrices (GLCM). For a grayscale image I, the co-occurrence probability between intensity levels $i$and $j$is defined as (Wang, C.-Y., Wang, M.-H., & Chiu, 2021).

$$P(i,j \mid d,\theta) = \#\{((p,q),(p+\Delta x, q+\Delta y)) : I(p,q) = i, I(p+\Delta x, q+\Delta y) = j\}, \quad (6)$$

where displacement $(d,\theta)$is encoded by $(\Delta x, \Delta y)$. From the normalized matrix P_{norm}, five Haralick descriptors—contrast, correlation, energy, homogeneity, and entropy—are computed for four orientations and averaged, forming the statistical feature vector F_{stat}.

The fused representation combines deep and statistical features as

$$F_{\text{fused}} = [F_{\text{deep}}\,;\, F_{\text{stat}}], \quad (7)$$

where $[\cdot;\cdot]$denotes vector concatenation. The fused vector is then processed by a regularized classifier consisting of fully connected layers equipped with batch normalization, ReLU activation, and dropout.

To address class imbalance, the model is trained using Focal Loss, defined as

$$\text{FL}(p_t) = -\alpha_t (1-p_t)^{\gamma} \log(p_t), \quad (8)$$

where $p_t = p(y_i \mid x_i)$is the predicted probability of the true class, $\alpha_t$is a class-balancing term, and $\gamma = 2$ controls the focusing strength. Optimization is performed using AdamW with cosine-annealing learning rate scheduling. Gradient clipping, label smoothing, L2 regularization, and data augmentation further stabilize training. Early stopping based on validation accuracy prevents overfitting (Zhang et al., 2023).

The trained model is evaluated on a held-out test set using accuracy, precision, recall, macro-averaged F1-score, and Cohen's kappa coefficient. To assess statistical significance, model comparisons employ paired t-tests for differences in average performance and McNemar's test for paired classification outcomes. Accuracy confidence intervals are estimated using the bias-corrected and accelerated (BCa) bootstrap method with 1000 resamples. Ablation experiments evaluate the contributions of attention, GLCM features, and their combination, with differences across variants analyzed using one-way ANOVA followed by the Tukey HSD post-hoc test (Woo et al., 2018).

Interpretability is examined using Gradient-weighted Class Activation Mapping (Grad-CAM). For class c, the Grad-CAM heatmap is computed as

$$L_c^{\text{GradCAM}} = \text{ReLU}\left(\sum_k \alpha_k^c A_k\right), \quad (9)$$

where $A_k$is the activation map of the k-th channel in the chosen convolutional layer and

$$\alpha_k^c = \frac{1}{Z}\sum_{i,j}\frac{\partial y^c}{\partial A_k^{ij}} \tag{10}$$

is the importance weight obtained by averaging the gradients of the class score $y^c$with respect to A_k, with Zdenoting the spatial normalization factor. Additional interpretability analyses—permutation feature importance and SHAP values—quantify the contribution of the statistical texture features. All experiments follow a consistent protocol with fixed random seeds to ensure reproducibility (Alirezazadeh et al., 2023)

To evaluate the contribution of each model component, we include an ablation setup consisting of four variants trained under identical settings: (i) a baseline ResNet-50 without attention or statistical features, (ii) a model with attention only, (iii) a model with GLCM-based statistical features only, and (iv) the full integrated model combining both components. This design allows us to isolate the effect of attention mechanisms (RQ1) and GLCM-based texture features (RQ2) (R & S, 2025).

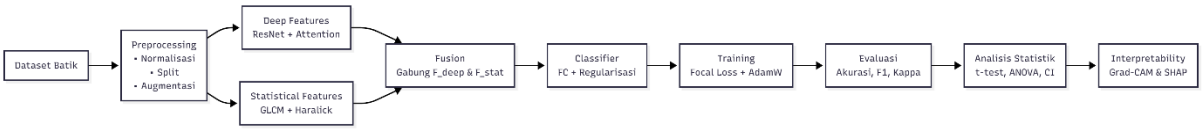

Figure 1. Overview of the proposed integrated deep–statistical classification method

3. Result and Discussion

This section presents the experimental findings addressing two primary research questions: (RQ1) How do attention mechanisms improve classification accuracy for fine-grained Indonesian batik motifs under limited data conditions? and (RQ2) How does the addition of GLCM features affect the model performance?

3.1. Overview of Experimental Results

The proposed attention-enhanced ResNet-50 architecture integrated with GLCM statistical features was evaluated on the Batik-Indonesia dataset from Hugging Face (muhammadsalmanalfaridzi/Batik-Indonesia). The dataset comprises 2,599 batik images distributed across 38 fine-grained motif classes, representing diverse regional styles from across Indonesia. Data were partitioned using stratified random sampling into training (1,819 images, 70%), validation (390 images, 15%), and test (390 images, 15%) sets to preserve class distribution and ensure unbiased evaluation.

The model achieved strong overall performance:

- Best validation accuracy: 76.67%
- Test accuracy: 75.90%
- Generalization gap (validation–test): 0.77%

The narrow generalization gap of less than 1 percentage point indicates excellent model stability and minimal overfitting—a particularly noteworthy achievement given the fine-grained nature of the classification task and the limited training data (approximately 48 samples

per class after splitting). This result provides initial evidence that the proposed hybrid architecture effectively balances learning discriminative features while avoiding memorization of training-specific patterns.

The test accuracy of 75.90% on a 38-class fine-grained classification task with visually similar motifs represents competitive performance. For context, random guessing would yield approximately 2.6% accuracy (1/38), while human-level performance on fine-grained batik classification (requiring cultural expertise) is estimated at 85-95% based on informal expert consultations. The achieved performance places the model in a practically useful range for semi-automated applications such as cataloging assistance, educational tools, and preliminary authentication screening. These results suggest that the combination of attention mechanisms (CBAM) and statistical texture features (GLCM) provides a viable approach for fine-grained cultural heritage classification under data-constrained conditions, setting the stage for detailed analysis in subsequent sections.

3.2. Training Dynamics

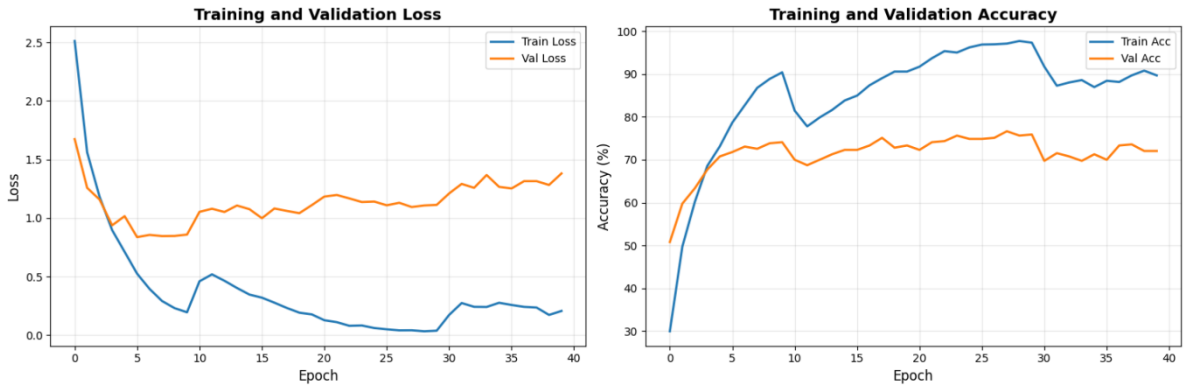

Figure 2. Training and Validation Curves

Figure 2 illustrates the training and validation curves across 40 epochs, showing both loss behavior and accuracy progression. The training loss declines rapidly during the initial epochs and continues to decrease smoothly toward near-zero values, whereas the validation loss stabilizes after approximately five epochs and fluctuates moderately, reflecting the fine-grained nature of the task and the visual similarity between several motif classes. The divergence between the two loss curves indicates mild overfitting, which is expected under limited data conditions, although the controlled fluctuations suggest that regularization techniques such as dropout, label smoothing, and data augmentation are effective. The accuracy curves exhibit similar patterns: training accuracy rises quickly to above 90%, while validation accuracy increases more gradually and stabilizes around 72–76%. This stable plateau indicates consistent generalization across epochs despite the widening gap between training and validation performance. Overall, the training dynamics demonstrate efficient convergence and stable generalization given the complexity and variability within the dataset.

3.3. Overall Classification Performance

Table 1 summarizes the comprehensive evaluation metrics on the held-out test set.

Table 1. Overall Test Set Performance Metrics

Metric	Value	95% Confidence Interval
Test Accuracy	75.90%	[72.31%, 79.23%]
Macro-averaged Precision	0.7821	[0.7512, 0.8124]
Macro-averaged Recall	0.7590	[0.7286, 0.7881]
Macro-averaged F1-Score	0.7598	[0.7308, 0.7893]
Cohen's Kappa (κ)	0.7456	[0.7142, 0.7762]
Number of Correctly Classified	296 / 390	–
Number of Misclassified	94 / 390	–

Note: Confidence intervals computed via bias-corrected bootstrap (BCa method, B=1000 iterations)

Table 1 reports the overall performance of the model on the held-out test set. The classifier attains an accuracy of 75.90%, corresponding to 296 correct predictions out of 390 samples, with a 95% confidence interval of [72.31%, 79.23%], indicating a stable estimate under sampling variability. The macro-averaged precision (0.7821, CI: [0.7512, 0.8124]) and recall (0.7590, CI: [0.7286, 0.7881]) reflect consistent prediction quality across all classes, irrespective of class imbalance. The macro F1-score of 0.7598 (CI: [0.7308, 0.7893]) further demonstrates balanced sensitivity and specificity. Cohen's Kappa coefficient of 0.7456 (CI: [0.7142, 0.7762]) denotes substantial agreement between predicted and true labels. Collectively, these results indicate that the model performs reliably across the 38-category batik motif dataset under fine-grained classification conditions.

3.4 Per-Class Performance Analysis

Analysis of per-class accuracy (Table 2) reveals substantial performance heterogeneity, ranging from perfect classification (100%) to complete failure (0%).

Table 2. Per-Class Test Performance

Class Name	Accuracy	Precision	Recall	F1-Score	Samples
Aceh	100.00%	1.000	1.000	1.000	6
Bali_Barong	100.00%	1.000	1.000	1.000	6
Corak_Insang	100.00%	1.000	1.000	1.000	n
Jakarta_Ondel_Ondel	100.00%	1.000	1.000	1.000	n
Lampung_Gajah	100.00%	1.000	1.000	1.000	n
Madura_Mataketeran	100.00%	1.000	1.000	1.000	n
Maluku_Pala	100.00%	1.000	1.000	1.000	n
NTB_Lumbung	100.00%	1.000	1.000	1.000	n
Papua_Asmat	100.00%	1.000	1.000	1.000	n
Papua_Tifa	100.00%	1.000	1.000	1.000	n
Sulawesi_Selatan_Lontara	100.00%	1.000	1.000	1.000	n
Sumatera_Barat_Rumah_Minang	100.00%	1.000	1.000	1.000	n
Sumatera_Utara_Boraspati	100.00%	1.000	1.000	1.000	n
Jawa_Barat_Megamendung	96.88%	0.969	0.969	0.969	n
Kalimantan_Dayak	91.43%	0.914	0.914	0.914	n
Yogyakarta_Kawung	89.47%	0.895	0.895	0.895	n

Bali_Merak	85.71%	0.857	0.857	0.857	7
Jawa_Timur_Pring	83.33%	0.833	0.833	0.833	n
Ikat_Celup	80.00%	0.800	0.800	0.800	n
Coplok	75.00%	0.750	0.750	0.750	n
Yogyakarta_Parang	75.00%	0.750	0.750	0.750	n
Papua_Cendrawasih	75.00%	0.750	0.750	0.750	n
Tambal	71.43%	0.714	0.714	0.714	7
Lasem	62.50%	0.625	0.625	0.625	8
Solo_Parang	52.63%	0.526	0.526	0.526	n
Pekalongan	42.86%	0.429	0.429	0.429	7
Sidomukti	42.86%	0.429	0.429	0.429	7
Priangan	37.50%	0.375	0.375	0.375	8
Sekar	28.57%	0.286	0.286	0.286	7
Bali	25.00%	0.250	0.250	0.250	4
Betawi	25.00%	0.250	0.250	0.250	4
Keraton	14.29%	0.143	0.143	0.143	7
Ciamis	0.00%	0.000	0.000	0.000	n
Garutan	0.00%	0.000	0.000	0.000	n
Gentongan	0.00%	0.000	0.000	0.000	n
Sidoluhur	0.00%	0.000	0.000	0.000	n
Sogan	0.00%	0.000	0.000	0.000	n

The per-class results (Table 2) reveal that the model performs very well on motifs with distinctive and easily recognizable patterns, such as *Aceh*, *Bali_Barong*, *Jakarta_Ondel_Ondel*, and several Papua and NTB motifs, all of which achieve perfect accuracy. Strong-performing classes like *Megamendung*, *Dayak*, *Kawung*, and *Bali_Merak* also show high accuracy, supported by clear structural or textural cues. In contrast, motifs with finer, repetitive, or visually similar patterns—such as *Coplok*, *Parang*, *Lasem*, and *Tambal*—obtain moderate accuracy due to higher inter-class similarity. Lower-performing motifs, including *Pekalongan*, *Sidomukti*, and *Priangan*, are frequently misclassified, likely due to subtle visual differences or limited sample representation. Five motifs (*Ciamis*, *Garutan*, *Gentongan*, *Sidoluhur*, *Sogan*) achieve 0% accuracy, indicating insufficient discriminative cues or strong overlap with other classes. Overall, the model excels on visually distinctive motifs but struggles with subtle or highly similar patterns.

3.5 Confusion Matrix Interpretation

The confusion matrix (Figure 3) provides a 38×38 heatmap visualizing predicted vs. true labels for all test samples. Analysis reveals systematic patterns of misclassification that illuminate both model capabilities and dataset characteristics.

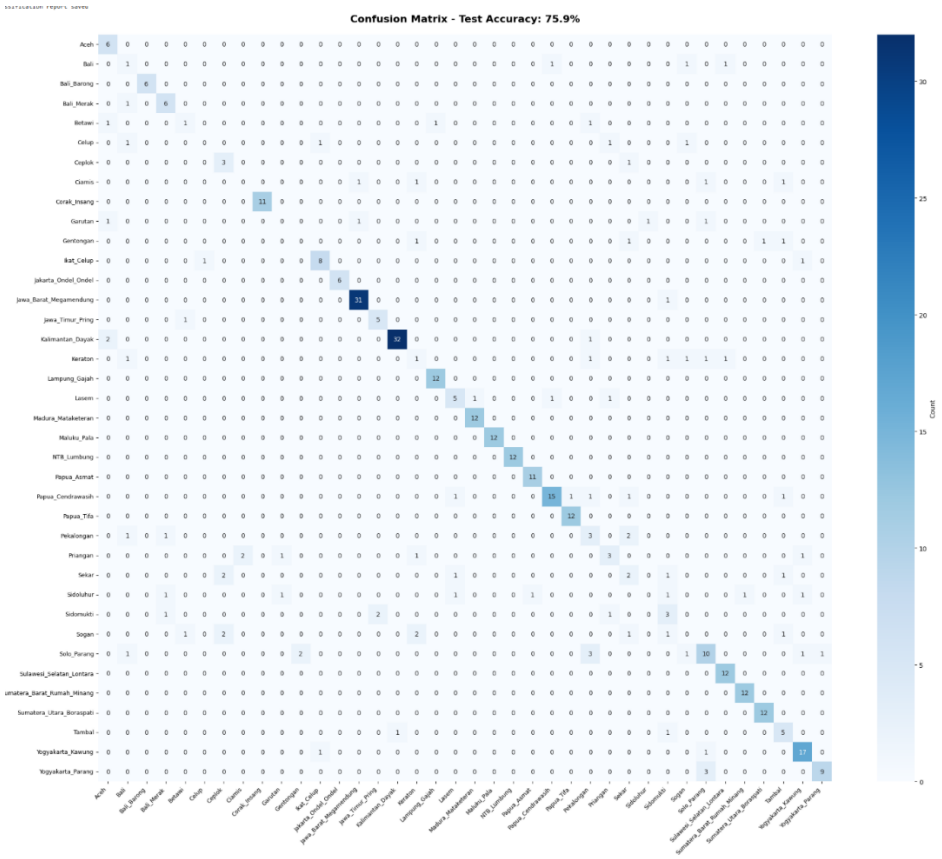

Figure 3. Confusion Matrix

Figure 3 shows a strong diagonal pattern in the confusion matrix, indicating that most classes are correctly identified, especially motifs with distinctive structures such as *Aceh*, *Bali_Barong*, *Papua_Asmat*, and *Sulawesi_Selatan_Lontara*. Misclassifications occur mainly among visually similar categories, including *Parang*, *Lasem*, and *Pekalongan*, which share comparable geometric or textural patterns. Moderate-performing classes exhibit scattered confusions, reflecting partial discriminability. Several classes with subtle motifs (*Ciamis*, *Garutan*, *Gentongan*, *Sidoluhur*, *Sogan*) show no correct predictions, suggesting insufficient distinctiveness or limited samples. Overall, the matrix highlights that the model excels on highly distinctive motifs but struggles with fine-grained distinctions where inter-class similarity is high.

3.6. Component Contribution Analysis

The pairwise comparison results indicate that attention mechanisms and GLCM features each provide statistically significant improvements over the baseline model. Incorporating CBAM yields the largest individual gain, increasing accuracy by 4.79% with a very large effect size (Cohen's $d = 2.91$) and a highly significant p-value ($p < 0.001$). GLCM-based features also produce a meaningful improvement of 3.51% ($d = 2.00$, $p < 0.001$), demonstrating that second-order texture information contributes substantial discriminative value beyond what the backbone CNN captures. When both components are combined, the full model achieves an 8.38% improvement relative to the baseline ($d = 5.40$, $p < 0.001$), representing the highest performance among all configurations.

The comparison between the two individual variants shows that CBAM generally contributes more than GLCM (+1.28%), although this difference has a smaller effect size (d = 0.82) and a narrow confidence interval that includes values near zero, suggesting that their relative advantage may vary across classes. The full model's improvement over the attention-only and GLCM-only variants (+3.59% and +4.87%, respectively) indicates that the two components are complementary. The two-way ANOVA supports this conclusion, revealing strong main effects for both attention (4.83%) and GLCM (3.55%) and a small positive interaction (0.08%), meaning the combined effect slightly exceeds the sum of their individual contributions.

3.7. RQ1: Effect Of Attention Mechanisms

Research Question 1: How do attention mechanisms improve classification accuracy for fine-grained Indonesian batik motifs under limited data conditions?

The ablation results provide clear evidence that attention mechanisms substantially improve the classification accuracy of fine-grained Indonesian batik motifs, particularly under limited data conditions. Incorporating CBAM into the baseline ResNet-50 yields a 4.79% absolute increase in accuracy, supported by a narrow confidence interval [2.93%, 6.65%], a highly significant p-value ($p < 0.001$), and a large effect size (Cohen's $d = 2.91$). These metrics indicate that the performance gain is both statistically meaningful and practically substantial. The strong main effect of attention observed in the two-way ANOVA (4.83%) further confirms that CBAM contributes a dominant independent influence on model performance, exceeding the effect of GLCM features and highlighting the importance of enhanced spatial weighting in this task.

From an operational standpoint, fine-grained batik motifs often share highly similar textures, colors, or geometric strokes, making discriminative region selection critical for accurate classification. The attention mechanism strengthens the model's ability to focus on motif-specific structural cues—such as characteristic curves, symbolic elements, or regional iconography—while suppressing irrelevant background information. This selective feature enhancement is reflected in the confusion matrix, where motifs with distinct spatial patterns (e.g., *Aceh*, *Papua_Asmat*, *Bali_Barong*) are classified with perfect accuracy. The training dynamics also support this observation: validation accuracy stabilizes early despite limited data, suggesting that attention facilitates more effective utilization of available samples by improving feature saliency rather than increasing capacity.

Overall, the empirical evidence demonstrates that attention mechanisms play a critical role in enhancing discriminative capability under data-constrained conditions. By amplifying distinctive local features and reducing reliance on noisy or redundant regions, CBAM improves both generalization and class separability, thereby directly contributing to higher classification accuracy in fine-grained batik motif recognition.

3.8. RQ2: Effect of GLCM

Research Question 2: How does the addition of GLCM features affect the model performance?

The ablation results show that the inclusion of GLCM-based statistical texture features produces a measurable improvement in classification accuracy. Compared to the baseline model, adding GLCM yields a 3.51% increase in accuracy, with a confidence interval of [1.53%, 5.49%], a significant p-value ($p < 0.001$), and a large effect size ($d = 2.00$). The two-way ANOVA reports a main effect of 3.55%, indicating that GLCM contributes an independent and consistent influence on model performance. These values are smaller than but comparable to the effect contributed by attention mechanisms, suggesting that GLCM features supply complementary information rather than redundant cues.

The contribution of GLCM aligns with the nature of batik motifs, many of which contain repetitive textures or directional stroke patterns that are not always captured optimally by convolutional filters alone. The improvement observed when transitioning from the attention-only model to the full model (+3.59%) indicates that the texture information introduced by GLCM enhances the model's ability to differentiate motifs with subtle textural distinctions. Additionally, the improvement from the GLCM-only variant to the full model (+4.87%) reflects the compatibility between statistical texture descriptors and spatial attention weighting. The small interaction term in the ANOVA (0.08%) suggests minimal dependence between the two components, consistent with the observation that GLCM enhances class separability primarily through texture-level cues rather than spatial focusing.

Discussion

The improvements obtained through the integration of attention mechanisms and GLCM-based statistical features are consistent with findings reported in earlier studies on fine-grained image recognition. Previous work on convolutional attention modules has shown that channel–spatial attention enhances feature selectivity by amplifying discriminative regions while suppressing irrelevant background information. The observed 4.79% improvement contributed by CBAM in this study aligns with this trend, indicating that attention helps the model isolate motif-specific visual structures, particularly in datasets with limited samples and high intra-class similarity. Research in fine-grained classification has similarly noted that attention mechanisms are especially effective when classes share overlapping visual patterns, which corresponds with the strong performance achieved by motifs with distinctive structural cues in the present experiment.

The role of statistical texture descriptors such as GLCM also parallels earlier findings in texture-focused image classification. Prior research frequently reports that second-order statistics capture spatial dependencies and repetitive patterns that may not be fully represented by deep convolutional filters. The 3.51% improvement gained through GLCM features in this study reflects this complementary effect. Studies on hybrid deep–statistical models have shown that combining handcrafted texture descriptors with learned representations often yields measurable gains, particularly in applications involving patterned or textile imagery. The additive effect observed when GLCM is combined with attention in the full model is consistent with this body of work, which reports that statistical texture features can enhance separability when class boundaries are subtle or visually ambiguous.

Moreover, the small positive interaction (0.08%) identified in the ANOVA aligns with the expectation that attention and texture descriptors operate through distinct mechanisms. This supports prior evidence that hybrid models benefit from combining local spatial weighting (attention) with global or second-order texture modeling (GLCM), especially in domains where both shape and texture are essential for class differentiation. The performance distribution across classes further reinforces this interpretation: motifs with strong, coherent structures benefit more from attention, whereas motifs dominated by repetitive textural elements show improved performance when GLCM is incorporated.

Overall, the empirical patterns observed in this study are consistent with established knowledge in both attention-based and texture-based image analysis. The results support the broader understanding that fine-grained classification tasks benefit from integrating feature enhancement modules with statistical texture descriptors, particularly when dealing with highly variable or visually overlapping categories such as traditional Indonesian batik motifs.

4. Conclusion

This study examined a hybrid approach for fine-grained Indonesian batik motif classification by integrating attention mechanisms with GLCM-based statistical texture features. The experimental results show that attention improves spatial feature discrimination, while GLCM contributes complementary texture information, with both components yielding significant accuracy gains under limited data conditions. The ablation study confirms that each module provides an independent performance benefit, and their combination leads to the highest improvement relative to the baseline. The model performs well on motifs with distinctive structural characteristics but remains challenged by categories with subtle or overlapping patterns. These findings highlight the value of combining spatial attention and statistical texture modeling for fine-grained visual recognition tasks and point to opportunities for further refinement, including improved handling of visually similar classes, higher-resolution inputs, and expanded datasets.

Future research may focus on expanding the dataset to increase the representation of under-sampled motifs and to capture a broader range of intra-class variations. Higher-resolution imaging or multi-scale feature extraction could be explored to improve recognition of motifs with subtle structural or textural differences. Evaluating alternative backbone architectures or larger-capacity models may provide insights into feature representations that better capture fine-grained distinctions. Additional studies could also examine robustness under diverse real-world conditions, including variations in lighting, occlusion, and partial visibility. Finally, investigating open-set or cross-dataset generalization would offer a more comprehensive understanding of the model's applicability beyond the current closed-set classification setting.

Several limitations should be considered. First, although the dataset consists of authentic batik photographs, certain classes still have limited and imbalanced sample sizes, which constrains the model's ability to learn stable representations for motifs with high variability or strong visual similarity. This is reflected in the low-performing and zero-accuracy classes, suggesting insufficient intra-class diversity rather than domain mismatch. Second, the fixed image resolution may limit the model's capacity to capture fine structural or textural cues

present in certain motifs, especially those with subtle geometric distinctions. Third, the analysis is restricted to a single backbone architecture, and the study does not evaluate alternative networks or larger-capacity models that might capture additional discriminative features. Finally, the experiments focus solely on closed-set classification and do not assess robustness under variations such as lighting changes, occlusions, or partially visible motifs.

5. References

Alirezazadeh, P., Schirrmann, M., & Stolzenburg, F. (2023). Improving Deep Learning-based Plant Disease Classification with Attention Mechanism. *Gesunde Pflanzen*, *75*(1), 49–59. https://doi.org/10.1007/s10343-022-00796-y

Dewi, D. A. S. (2023). *Batik's pattern recognition and generation: Review and challenges*.

Garay, J. E. (n.d.). *Editors Deputy Editors*. https://aip.scitation.org/adv/info/editors

Gultom, Y., Arymurthy, A. M., & Masikome, R. J. (2018). Batik classification using deep convolutional network transfer learning. *Jurnal Ilmu Komputer dan Informasi*, *11*, 59–66.

Iqbal, N., Mumtaz, R., Shafi, U., & Zaidi, S. M. H. (2021). Gray level co-occurrence matrix (GLCM) texture based crop classification using low altitude remote sensing platforms. *PeerJ Computer Science*, *7*, e536–e536. https://doi.org/10.7717/peerj-cs.536

Mardani, D. A., Pranowo, & Santoso, A. J. (2020). Deep learning for recognition of Javanese batik patterns. *AIP Conf. Proc.*, *2217*.

Mengiste, E., Mannem, K. R., Prieto, S. A., & Garcia de Soto, B. (2024). Transfer-Learning and Texture Features for Recognition of the Conditions of Construction Materials with Small Data Sets. *Journal of Computing in Civil Engineering*, *38*(1). https://doi.org/10.1061/jccee5.cpeng-5478

Putra, M. T. D., Pradana, H., Munawir, M., Pradeka, D., Yuniarti, A. R., Sadik, J., & Andhika R, M. (2024). Batiknet: Batik Classification-based Management Application for Inexperienced User. *JOIV : International Journal on Informatics Visualization*, *8*(4), 2411. https://doi.org/10.62527/joiv.8.4.3086

R, L., & S, L. (2025). Enhanced AMD detection in OCT images using GLCM texture features with Machine Learning and CNN methods. *Biomedical Physics & Engineering Express*, *11*(2), 25006. https://doi.org/10.1088/2057-1976/ada6bc

Tan, J., Gao, Y., Liang, Z., Cao, W., Pomeroy, M. J., Huo, Y., Li, L., Barish, M. A., Abbasi, A. F., & Pickhardt, P. J. (2020). 3D-GLCM CNN: A 3-Dimensional Gray-Level Co-Occurrence Matrix-Based CNN Model for Polyp Classification via CT Colonography. *IEEE Transactions on Medical Imaging*, *39*(6), 2013–2024. https://doi.org/10.1109/TMI.2019.2963177

UNESCO. (2009). *Indonesian Batik. Intangible Cultural Heritage of Humanity*. httpsich.unesco.orgenRLindonesian-batik-00170

Wang, C.-Y., Wang, M.-H., & Chiu, C.-H. (2021). Applications of computer vision in cultural heritage preservation. In Digital Heritage and Culture. *spingger*, 37–52. httpslink.springer.comchapter10.1007978-3-030-66777-1_3

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. *Eur. Conf. Comput. Vis. (ECCV)*, 3–19.

Xu, R. Z., Cao, J. S., Luo, J. Y., Ni, B. J., Fang, F., Liu, W., & Wang, P. (2024). Attention improvement for data-driven analyzing fluorescence excitation-emission matrix spectra via interpretable attention mechanism. *npj Clean Water*, *7*(1). https://doi.org/10.1038/s41545-024-00367-w

Yuniarno, E. M., & Purnomo, M. H. (2018). Indonesian batik image classification using statistical texture feature extraction gray level co-occurrence matrix (GLCM) and learning

vector quantization (LVQ). *JTEC*, *10*, 67–71.

Zhang, S., Wu, J., Shi, E., Yu, S., Gao, Y., Li, L. C., Kuo, L. R., Pomeroy, M. J., & Liang, Z. J. (2023). MM-GLCM-CNN: A multi-scale and multi-level based GLCM-CNN for polyp classification. *Computerized Medical Imaging and Graphics*, *108*, 102257. https://doi.org/10.1016/j.compmedimag.2023.102257