Journal of Telecommunication, Electronic and Computer Engineering
ISSN: 2180 – 1843 e-ISSN: 2289-8131 Vol. 16 No. 4 (2024) 23 – 30
jtec.utem.edu.my
DOI: https://doi.org/10.54554/jtec.2024.16.04.004

Classification of Malaysian and Indonesian Batik Designs Using Deep Learning Models

Nurulfajar Abd Manap[1]*, Lee Xiao Xuan[1], Koushlendra Kumar Singh[2], Akbar Sheikh Akbari[3] and Azma Putra[4]

[1]Centre for Telecommunication Research & Innovation, Fakulti Teknologi Dan Kejuruteraan Elektronik Dan Komputer (FTKEK), Universiti Teknikal Malaysia Melaka (UTeM), Melaka, 76100, Malaysia.
[2]Machine Vision and Intelligence Lab, National Institute of Technology, Jamshedpur, India.
[3]School of Built Environment, Engineering and Computing, Leeds Beckett University, Leeds, LS1 3HE, United Kingdom.
[4]School of Civil and Mechanical Engineering, Curtin University Kent St., Bentley, Australia.

Article Info

Article history:
Received Sep 13th, 2024
Revised Nov 18th, 2024
Accepted Dec 10th, 2024
Published Dec 20th, 2024

Index Terms:
Deep Learning
Batik Classification
Convolutional Neural Network
Cultural Heritage Preservation

Abstract

Batik is a traditional textile art form native to Southeast Asia, especially prominent in Malaysia and Indonesia, where unique patterns reflect significant cultural value. The intricate designs of batik, often embodying floral, geometric, and symbolic elements, make automated classification challenging and time intensive. This study presents a method for classifying Malaysian and Indonesian batik patterns using deep learning models. A curated dataset of 1,825 batik images was compiled, consisting of 949 Indonesian batik images and 876 Malaysian batik images. Three popular Convolutional Neural Network (CNN) architectures: MobileNet v2, YOLO-v8, and LeNet-5 were evaluated based on classification accuracy, loss, and training efficiency. Results show that YOLO-v8 achieved the highest accuracy at 98.80%, followed by MobileNet v2 with 97.79%, and LeNet-5 with 92.94%. These findings indicate that CNN models can effectively distinguish between Malaysian and Indonesian batik designs, offering valuable applications in cultural preservation and industry documentation. Future work could focus on refining these models for real-time use and expanding the dataset to capture additional regional variations in batik design.

*Corresponding Author: nurulfajar@utem.edu.my

I. INTRODUCTION

Batik is a traditional textile art form deeply rooted in the cultural heritage of Malaysia and Indonesia, renowned for its intricate patterns and designs created through a wax-resist dyeing technique. These designs often draw inspiration from nature, mythology, and daily life, reflecting the rich cultural values of the region. Despite increasing interest in batik as a sustainable and eco-friendly textile, the classification of batik designs remains a significant challenge. Traditional methods rely on the expertise of trained individuals to visually inspect and analyze these complex patterns, making the process both time-consuming and subjective.

While research on Indonesian batik has benefited from the availability of numerous datasets, such as Batik Nitik 960 and other regional collections [1][2], there is a notable lack of comprehensive datasets for Malaysian batik. This disparity limits research opportunities and hinders the development of automated systems capable of recognizing and classifying Malaysian batik patterns. Addressing this gap is critical for advancing the understanding, preservation, and promotion of Malaysian batik within the broader field of cultural heritage studies.

This study seeks to address these challenges by developing a deep learning-based classification system to distinguish between Malaysian and Indonesian batik designs, thereby contributing to the preservation and promotion of this cultural heritage. Manual classification demands substantial expertise and time, and current automated systems lack comprehensive datasets, particularly for Malaysian batik. To bridge this gap, a new dataset of batik images has been compiled, and deep learning techniques have been employed to facilitate accurate classification of these designs.

In recent years, artificial intelligence (AI), particularly deep learning, has demonstrated considerable promise in automating and enhancing various aspects of cultural heritage preservation and analysis. The classification of batik patterns presents an ideal case for leveraging these computational methods due to the complexity and variety of batik designs. This study specifically addresses two key challenges in this domain: the lack of a comprehensive, standardized dataset of Malaysian batik designs, and the need for a reliable classification system capable of distinguishing between Indonesian and Malaysian batik styles.

To tackle these challenges, a dataset of 1,825 batik images, representing both Indonesian and Malaysian designs, was compiled. The study implemented and compared three popular Convolutional Neural Network (CNN) architectures: MobileNet v2, YOLO-v8, and LeNet-5, each selected for their unique characteristics in terms of computational efficiency and classification capability. MobileNet v2 was chosen for its efficiency in mobile and embedded vision applications, YOLO-v8 for its speed and suitability for real-time classification, and LeNet-5 as a foundational benchmark for comparative analysis.

This study contributes to the field by introducing a curated dataset of Indonesian and Malaysian batik images, addressing the scarcity of publicly available resources for batik classification research. The comparative analysis of three CNN architectures provides insights into their performance, highlighting their respective strengths and limitations in batik pattern recognition. Furthermore, this research demonstrates the potential of CNN models to accurately distinguish between Malaysian and Indonesian batik designs, offering practical applications in cultural heritage preservation, education, and industry documentation. By demonstrating the effectiveness of deep learning for batik classification, this study lays the groundwork for future research and the development of real-time classification systems.

II. RELATED WORKS

Batik is a traditional textile art form that holds significant cultural and historical value in both Malaysia and Indonesia. In Malaysia, batik is primarily produced in the states of Kelantan and Terengganu, characterized by floral and geometric motifs in vibrant colors. These designs are often simpler and larger compared to the intricate patterns of Indonesian batik. The creation of Malaysian batik predominantly uses the "canting" method, which involves applying hot wax onto the fabric with a small copper tool. On the other hand, Indonesian batik, particularly from Java, is renowned for its complexity and diversity. Its designs often incorporate motifs inspired by natural elements, folklore, and religious symbols. The "tjap" method, which employs copper stamps for repetitive patterns, is frequently used in Indonesian batik production. Recognizing its cultural importance, UNESCO inscribed Indonesian batik as a Masterpiece of Oral and Intangible Heritage of Humanity.

The distinct differences between Malaysian and Indonesian batik extend to their motifs and color palettes. Malaysian batik commonly features flora motifs and bright colors, such as pink, purple, and green. In contrast, Indonesian batik is often designed with darker, earthy tones, including brown, gold, and black, and exhibits a broader variety of motifs influenced by different provinces [3]. In Malaysia, there are four main styles of batik design: hand-drawn batik, stamped batik, stenciling batik, and dip (dye) batik. Artisans often combine these techniques to create unique and distinctive patterns. A sample of Malaysian and Indonesian batik designs is shown in Figure 1.

Despite the wealth of cultural and artistic diversity in batik, the availability of datasets for automated classification research is unequal. Indonesian batik datasets, such as Batik Nitik 960, provide a valuable foundation for image processing and classification studies. This dataset comprises 960 images of 60 Nitik patterns from Yogyakarta, Indonesia, and has been pivotal in advancing research on Indonesian batik [1][2]. However, no comparable dataset exists for Malaysian batik, limiting the scope of research and the development of automated classification systems for this art form. Addressing this gap, this study introduces a curated dataset comprising 1,825 images of Malaysian and Indonesian batik designs to facilitate further research and development in batik classification.

Previous studies highlight the effectiveness of deep learning techniques in classifying batik patterns. For instance, Tiwari [4] demonstrated that applying Convolutional Neural Networks (CNNs) with transfer learning achieved high accuracy, precision, and recall in classifying batik motifs, showcasing the utility of CNNs in cultural preservation and economic applications. However, many studies focus on specific types of batik, such as Nitik or Demak batik, which limits the generalizability of these models to diverse designs from other regions. This constraint hinders the broader application of automated classification systems across the batik industry [5].

(a)

(b)

Figure 1. Sample of (a) Malaysia batik and (b) Indonesia batik

A. Batik Classification Methods

Traditional batik classification methods relied heavily on manual inspection by experts. While this approach provided valuable insights into the cultural and artistic nuances of batik patterns, it was labor-intensive, subjective, and inefficient for large-scale applications. The complexity of batik designs, particularly Indonesian patterns known for their intricate and overlapping motifs, further complicated manual classification. This prompted researchers to explore automated classification techniques to enhance accuracy and efficiency in distinguishing batik patterns.

Initial attempts at automated batik classification focused on basic image processing techniques such as edge detection, texture analysis, and color histograms. These methods were moderately successful in identifying simple patterns but struggled with the intricate and diverse nature of batik motifs. For example, overlapping designs and complex textures frequently caused inaccuracies in classification, limiting the scalability of these techniques.

More advanced machine learning methods were subsequently introduced to improve the classification of batik patterns. Algorithms such as Support Vector Machines (SVM) and k-Nearest Neighbors (KNN) utilized feature extraction techniques like Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF) to analyze and classify batik designs. KNN, combined with Gray Level Co-occurrence Matrix (GLCM) features, achieved a maximum accuracy of 96%, demonstrating the effectiveness of statistical feature extraction [6]. Similarly, Minarno [2] compared SVM and KNN using Multi Texton Histogram (MTH) feature extraction, reporting an optimal accuracy of 82% for KNN and 76% for SVM. However, both SVM and KNN struggled with the complex textures of batik motifs, as their reliance on simpler statistical features often failed to capture the intricacies of these designs.

Azhar et al. [6] utilized SIFT features with an SVM classifier, achieving an average accuracy of 97% in distinguishing batik patterns. While promising, these methods required extensive manual feature engineering, limiting their adaptability to diverse or unseen batik patterns. Similarly, the Naïve Bayes classifier performed well in classifying batik motifs based on texture features, achieving 97.22% accuracy with a 70-30 training-testing data split [7]. However, the Naïve Bayes approach assumes feature independence, which may not hold for batik designs where spatial relationships between features are critical, leading to reduced performance in complex patterns.

The MU2ECS-LBP algorithm by Rangkuti [8] demonstrated exceptional accuracy in batik classification when combined with KNN and Artificial Neural Networks (ANN). While the results achieved high precision, they were based on specific datasets, limiting their generalizability across other batik designs or larger datasets with more varied patterns and textures. These findings highlighted a recurring challenge in traditional and machine learning-based methods: their reliance on feature extraction and dataset specificity, which restricts their ability to generalize across diverse batik styles.

The limitations of manual feature engineering, sensitivity to dataset variability, and inability to capture complex spatial relationships in traditional machine learning methods underscore the need for more robust approaches. This has led to the adoption of deep learning techniques, which eliminate the need for manual feature extraction and are better equipped to handle the complexity and diversity of batik designs.

B. Deep Learning and CNN in Image Classification

The advent of deep learning, particularly Convolutional Neural Networks (CNNs), has significantly advanced the field of image classification. Unlike traditional methods, CNNs can automatically learn hierarchical features directly from raw pixel data, making them especially suitable for tasks involving complex visual patterns, such as batik design classification. Deep learning models have been applied successfully to various cultural heritage preservation tasks, including the classification of traditional textile patterns. In the context of batik, CNNs have outperformed traditional machine learning methods by effectively capturing intricate patterns and subtle differences across various batik designs.

Recent studies have explored different CNN architectures for batik classification. Agastya and Setyanto [9] employed data augmentation to improve the generalization of CNN models for Indonesian batik classification, achieving high accuracy even with limited datasets. Similarly, Winarno et al. [10] utilized a hybrid approach combining Artificial Neural Networks (ANNs) and supervised learning, achieving an accuracy of 99.76%. However, their method faced challenges with image misclassification due to poor image quality.

Arsa and Susila [11] demonstrated the effectiveness of the VGG16 architecture combined with a Random Forest classifier, achieving approximately 97% accuracy, precision, and recall in classifying batik patterns. While effective, their reliance on a small dataset raised concerns about the model's ability to generalize across diverse batik designs. Alya et al. [12] applied transfer learning using pre-trained VGG16 models, improving accuracy and computational efficiency compared to training models from scratch. However, pre-trained models trained on general datasets such as ImageNet might not fully capture the cultural and stylistic intricacies unique to batik designs.

To address the challenges of small datasets, Khasanah et al. [13] implemented data augmentation, expanding their dataset and improving classification accuracy from 95.8% to 98.9%. Meranggi et al. [14] employed a patch-based method, achieving 88.8% accuracy. These studies underscore the importance of high-quality and diverse datasets for enhancing CNN performance in batik classification tasks.

This study evaluates three CNN architectures: MobileNet v2, YOLO-v8, and LeNet-5, selected for their diverse characteristics in terms of network depth, computational efficiency, and suitability for specific applications.

MobileNet v2 is designed for efficiency in mobile and embedded vision applications. It employs depthwise separable convolutions to reduce parameters and computational requirements while maintaining high accuracy [15]. This lightweight architecture makes it ideal for deployment in resource-constrained environments. Jamil et al. [16] demonstrated MobileNet's effectiveness in real-time applications, achieving a balance between speed and accuracy, which aligns well with the requirements of batik classification.

YOLO-v8 is a real-time object detection system that divides images into regions and predicts bounding boxes and probabilities simultaneously. Its speed and accuracy make it suitable for applications requiring rapid classification [17]. For example, Nawarathne et al. [17] applied YOLO-v8 for jellyfish classification, achieving 99.5% accuracy. Similarly, Kumar et al. [18] improved YOLOv3 by integrating MobileNet, achieving high precision and sensitivity for rapid object recognition, demonstrating its potential for batik classification.

As one of the earliest CNN architectures, LeNet-5 was originally developed for digit recognition. Its simplicity and shallow architecture make it a useful benchmark for comparing modern CNN models. Pitsun et al. [19] compared LeNet with AlexNet and MobileNet for emotion classification, highlighting its limitations in terms of accuracy and efficiency compared to newer architectures. Despite its simplicity, Deepti and Deepthi [20] achieved 95% accuracy using a modified LeNet for skin disease classification, indicating its potential for resource-constrained applications.

This study contributes to the existing literature by addressing the lack of datasets for Malaysian batik and conducting a comparative analysis of MobileNet v2, YOLO-v8, and LeNet-5 for batik classification. By demonstrating the high accuracy achievable with these models, this research highlights the potential for CNNs to automate and enhance cultural heritage preservation. Furthermore, the introduction of a curated dataset of Malaysian and Indonesian batik designs provides a valuable resource for future research.

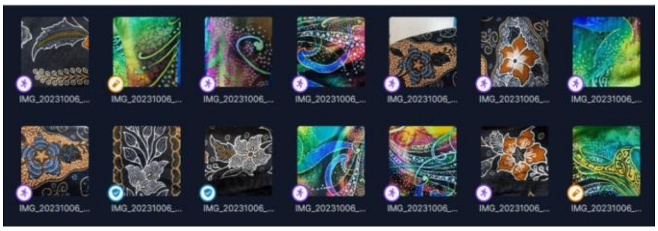
III. METHODOLOGY

The methodology for this study was structured into four key phases: (1) Dataset Compilation, (2) Data Preprocessing, (3) Model Selection and Implementation, and (4) Model Evaluation. Each phase was carefully designed to ensure the accuracy and efficiency of the batik classification system using deep learning algorithms.

A. Dataset Compilation

The first phase of this study involved compiling a comprehensive dataset of batik designs from both Malaysia and Indonesia. The dataset consisted of 1,825 images, comprising 949 images of Indonesian batik and 876 images of Malaysian batik. These images were sourced from diverse origins, including online repositories, digital archives, and direct contributions from batik artisans and collectors.

The dataset was curated to represent a wide range of batik patterns, with a particular focus on those created using the stamped method. This method was selected due to its prevalence in both Malaysian and Indonesian batik production. Additionally, efforts were made to include diverse designs, colors, and patterns, ensuring a robust training set that captured the richness and variety of batik art from both countries. Samples from the dataset, showcasing batik designs from Indonesia and Malaysia, are presented in Figure 2.



(a)

(b)

Figure 2. Sample of (a) Indonesia batik dataset and (b) Malaysia batik dataset

B. Data Preprocessing

Before training the deep learning models, the dataset underwent several preprocessing steps to enhance model performance and ensure consistency. These steps included image resizing, normalization, and data augmentation.

All images were resized to a uniform size of 576×576 pixels to match the input requirements of the selected CNN models. This resizing ensured compatibility with the network architectures and streamlined the training process. Normalization was then applied to the images by scaling the pixel values to the range [0, 1] through division by 255. This standardization improved the efficiency of the training process, enabling faster convergence and reducing computational overhead.

To further address the limited size of the dataset and improve the models' ability to generalize, data augmentation techniques were employed. Augmentation methods such as rotation, flipping, zooming, and shifting were used to artificially increase the dataset size by generating modified versions of the existing images. Following augmentation, the dataset size expanded to 2,507 images. This augmented dataset was then split into 80% for training (1,995 images) and 20% for validation and testing (512 images). These preprocessing steps not only enriched the diversity of the dataset but also helped mitigate overfitting by exposing the models to varied input conditions.

C. Model Selection and Implementation

This study evaluated three popular CNN architectures: MobileNet v2, YOLO-v8, and LeNet-5. Each model was selected for its unique characteristics and suitability for the task of batik classification, providing a comparative analysis across varying levels of complexity and application scenarios.

MobileNet v2 was chosen for its lightweight architecture, which makes it highly suitable for mobile and embedded applications. The model, pre-trained on the ImageNet dataset, employed transfer learning to fine-tune its parameters on the curated batik dataset. This approach allowed MobileNet v2 to leverage pre-existing feature representations while adapting to the specific patterns and textures of batik designs.

YOLO-v8, renowned for its real-time object detection capabilities, was selected for its ability to classify batik patterns quickly and accurately. For this study, YOLO-v8 was configured to treat each batik pattern as a distinct object within the image, enabling efficient detection and classification.

LeNet-5, one of the earliest CNN architectures, served as a baseline model. Despite its simpler and relatively shallow design, LeNet-5 provided a valuable benchmark for assessing the performance improvements offered by more advanced architectures like MobileNet v2 and YOLO-v8.

Each model was trained using a supervised learning approach, where batik images were labeled according to their country of origin (Malaysia or Indonesia). The training process involved backpropagation to optimize the model weights, utilizing categorical cross-entropy as the loss function to minimize classification error. This setup ensured that all three models were evaluated under consistent training conditions for a fair performance comparison.

D. Model Evaluation

The performance of the CNN models was evaluated using three key metrics: accuracy, loss, and the confusion matrix. These metrics provided a comprehensive assessment of the models' effectiveness in classifying batik designs.

Accuracy measured the proportion of correctly classified images out of the total number of images and served as a primary indicator of the model's overall performance. For the loss metric, the categorical cross-entropy loss function was used to evaluate prediction error during training and testing. Lower loss values corresponded to better model performance, indicating the model's ability to minimize errors in its predictions.

The confusion matrix was utilized to assess the classification performance of the models by analyzing the relationships between true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). This analysis provided insights into areas where the models performed well and where they struggled, allowing for the identification of misclassification patterns. The confusion matrix is structured as shown in Table 1.

Table 1
Confusion Matrix

	Positive (1)	Negative (0)
Positive (1)	True Positive (TP)	False Positive (FP)
Negative (0)	False Negative (FN)	True Negative (TN)

From the confusion matrix, additional metrics such as precision, recall, F1-score, and accuracy were calculated to provide further insights into the models' classification capabilities. These metrics are defined as follows:

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

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

$$F1\ Score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \quad (3)$$

$$Accuracy = \frac{TP}{TP + FN} \quad (4)$$

These metrics provided a holistic evaluation of the models' classification performance by balancing precision, recall, and overall accuracy.

The models were trained and tested on the compiled batik dataset, which was divided into training, validation, and testing subsets. Data preprocessing and augmentation were performed using Roboflow to enhance the quality and diversity of the dataset. This step was crucial in improving the models' generalization to unseen data.

The models were implemented using Python and popular deep learning libraries, including TensorFlow and PyTorch. Training and testing were conducted in a high-performance computing environment equipped with GPUs to accelerate the learning process. The training process involved multiple runs with hyperparameter optimization, adjusting factors such as learning rate, batch size, and the number of epochs. Early stopping and model checkpointing techniques were employed to prevent overfitting and ensure that the best-performing models were saved during training.

Finally, the performance of MobileNet v2, YOLO-v8, and LeNet-5 was compared using the metrics. This comparative analysis aimed to identify the most effective architecture for automated batik classification and highlight the strengths and weaknesses of each model.

IV. RESULTS AND DISCUSSION

This section presents the outcomes of the experiments conducted to evaluate the performance of the three CNN architectures: MobileNet v2, YOLO-v8, and LeNet-5 in classifying Malaysian and Indonesian batik designs. The results are analyzed in terms of accuracy, loss, training time, and confusion matrices. Additionally, the implications of these findings are discussed in relation to the study's objectives and the existing literature.

A. Model Performance Metrics

The performance of each CNN model was assessed using three primary metrics: classification accuracy, loss (categorical cross-entropy), and training time. The models were trained and validated on the compiled dataset of 1,825 batik images, with an 80-10-10 split for training, validation, and testing, respectively.

The MobileNet v2 model demonstrated consistent improvement in both training and validation accuracy over the 10 epochs. The model started with a training accuracy of 76.29% and a validation accuracy of 88.30% in the first epoch, and gradually improved to a training accuracy of 97.79% and a validation accuracy of 96.78% by the 10th epoch. The training loss decreased from 0.449 to 0.067, while the validation loss decreased from 0.284 to 0.096 over the same period. These results indicate that the model effectively learned the features of the training data while generalizing well to the validation data. The observed reduction in loss and increase in accuracy over the epochs supports this conclusion. The results from the model are visualized into graphs as shown in Figures 3 and 4.

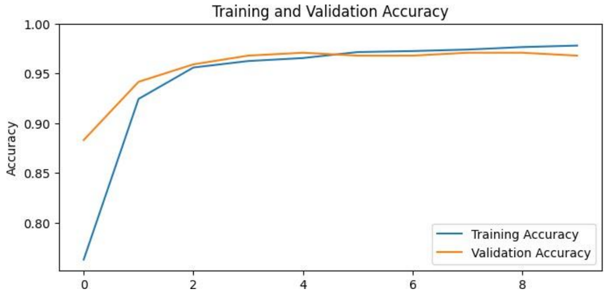


Figure 3. Training and Validation Accuracy for MobileNet v2

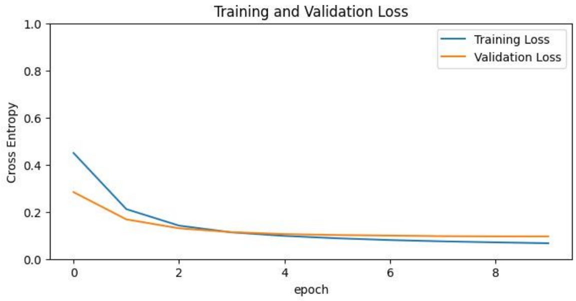


Figure 4. Training and Validation Loss for MobileNet v2

The YOLO-v8 model achieved high performance from the initial epoch, starting with both training and validation accuracies at 96.47%. By the 20th epoch, these values increased to 98.82% for both metrics. The training loss decreased significantly from 0.448 to 0.004, while the validation loss decreased from 0.375 to 0.324 over the same period. These findings demonstrate the model's efficiency in learning from the dataset while maintaining high accuracy. The training process results for YOLO-v8 are illustrated in Figure 5.

The LeNet-5 model exhibited a consistent enhancement in both training and validation accuracy. The training accuracy started at 74.03%, and the validation accuracy was 80.36% in the initial epoch, increasing to 90.29% training accuracy and 89.88% validation accuracy by the 20th epoch. Concurrently, the training loss decreased from 1.4699 to 0.2675, and the validation loss decreased from 0.673 to 0.231 over the 20 epochs. These results signify the model's adeptness to learn the training data and generalize effectively to the validation data, as evidenced by the decreasing loss and increasing accuracy throughout the epochs. The results are shown in Figures 6 and Figure 7.

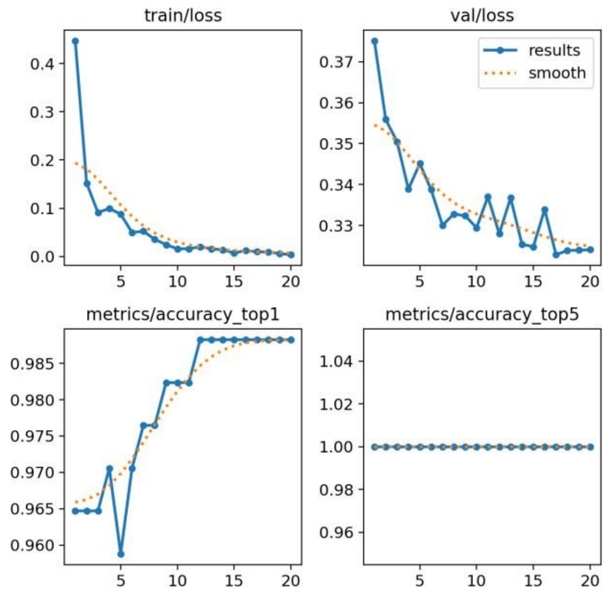


Figure 5. Training and Validation of Accuracy and Loss for YOLO-v8

Figure 6. Training and Validation Accuracy for LeNet-5

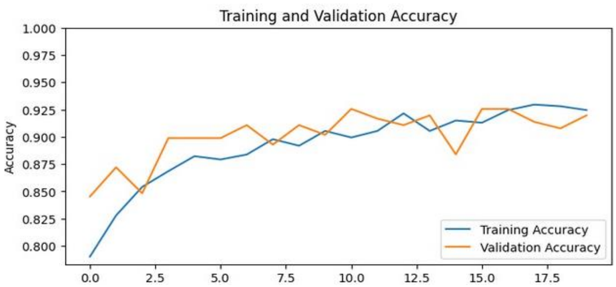


Figure 7. Training and Validation Loss for LeNet-5

The YOLO-v8 model demonstrated the highest classification accuracy, achieving 98.80%, compared to 97.79% for MobileNet v2 and 92.94% for LeNet-5. These results indicate that YOLO-v8 is particularly well-suited for applications requiring high precision and robust classification performance, making it a strong candidate for developing a batik classification system where accuracy is paramount. MobileNet v2, while slightly less accurate than YOLO-v8, offers significant advantages in terms of computational efficiency. Its lightweight architecture and reduced resource requirements make it highly suitable for mobile and embedded application development. This trade-off between accuracy and efficiency positions MobileNet v2 as an optimal choice for scenarios where computational resources are constrained or real-time processing is required. LeNet-5, despite being one of the earliest CNN architectures, achieved a respectable accuracy of 92.94%. Its simpler architecture provides a useful benchmark, highlighting the advancements made by more modern architectures like MobileNet v2 and YOLO-v8. While its performance is adequate for basic classification tasks, LeNet-5 may be less effective for more complex batik patterns or applications demanding higher precision. Table 2 summarizes the key performance metrics of the three models during their initial epochs, offering a comparative view of their baseline capabilities.

Table 2
Performance Metrics of CNN Models

Model	Training Accuracy (%)	Validation Accuracy (%)	Training Loss	Validation Loss
MobileNet v2	76.29	88.30	0.45	0.28
YOLO-v8	96.47	96.47	0.45	0.37
LeNet-5	74.03	80.36	1.47	0.67

B. Confusion Matrix Analysis

The confusion matrix provided a detailed view of the classification performance by visualizing the relationships between true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each model. It served as a valuable tool for summarizing the predictions of a classification model by comparing them to the actual class labels of the test data. In the matrix, rows represented the predicted class, and columns represented the actual class. The diagonal elements indicated true positive predictions, while the off-diagonal elements represented misclassifications. Figures 8, 9, and 10 display the confusion matrices for MobileNet v2, YOLO-v8, and LeNet-5, respectively.

As shown in Figure 8, the confusion matrix for MobileNet v2 revealed a high number of true positives and true negatives, along with a minimal number of false positives and false negatives. This indicated a high level of accuracy and precision, with minimal confusion between Malaysian and Indonesian batik designs. Only a small number of misclassifications were observed, highlighting MobileNet v2's capability in reliably distinguishing intricate patterns.

For YOLO-v8 (Figure 9), the confusion matrix showed slightly higher misclassification rates compared to MobileNet v2. Although the model maintained a high level of accuracy, some overlap was observed in distinguishing similar patterns. This was expected, as YOLO-v8 was primarily optimized for object detection rather than pure classification tasks. Despite this, the model remained effective, particularly in scenarios requiring rapid classification.

In contrast, the confusion matrix for LeNet-5 (Figure 10) demonstrated a higher rate of misclassifications, particularly for complex batik designs with intricate patterns. The model struggled to accurately distinguish between certain Malaysian and Indonesian batik designs, reflecting its limitations due to a simpler architecture and lower capacity to capture complex features. These challenges reinforced the need for more advanced CNN architectures when addressing complex classification tasks like batik design differentiation.

The results highlighted that advanced CNN architectures, especially MobileNet v2 and YOLO-v8, significantly outperformed traditional models like LeNet-5 in terms of accuracy and reliability. MobileNet v2's efficient balance between precision and recall made it suitable for real-world applications, especially those requiring resource efficiency and accurate classification. Table 3 summarizes the performance of the three models in terms of precision, recall, and F1-score, metrics derived from the confusion matrices.

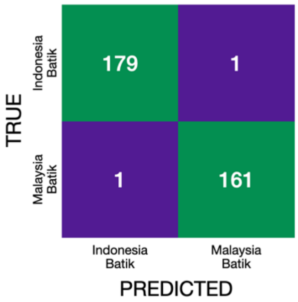


Figure 8. Confusion Matrix for MobileNet v2

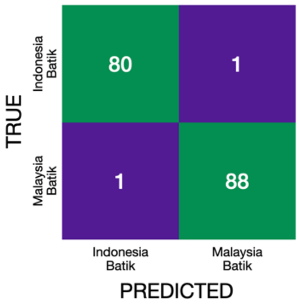


Figure 9. Confusion Matrix for YOLO-v8

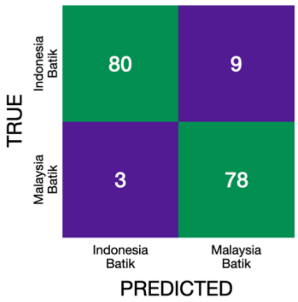


Figure 10. Confusion Matrix for LeNet-5

Table 3
Performance Metrics based on Precision, Recall and F1-score of CNN Models

Model	Precision	Recall	F1-score
MobileNet v2	0.994	0.994	0.994
YOLO-v8	0.988	0.988	0.985
LeNet-5	0.899	0.964	0.997

Precision measured the proportion of correctly predicted positive cases relative to the total predicted positives. MobileNet v2 achieved the highest precision of 0.994, indicating its reliability in minimizing false positives. YOLO-v8, with a precision of 0.988, also demonstrated strong performance. However, LeNet-5, with a precision of 0.899, exhibited a higher tendency for false positives, reflecting its limitations in discerning complex patterns.

Recall, or sensitivity, measured the proportion of correctly identified positive cases relative to the total actual positives. LeNet-5 achieved the highest recall at 0.964, emphasizing its sensitivity in detecting positive cases. However, this came at the cost of increased false positives. MobileNet v2 and YOLO-v8 demonstrated more balanced recall values of 0.994 and 0.988, respectively, maintaining high sensitivity while minimizing misclassifications.

F1-Score, the harmonic mean of precision and recall, provided a comprehensive measure of model performance. LeNet-5 achieved the highest F1-score at 0.997, driven by its high recall. However, MobileNet v2, with an F1-score of 0.994, emerged as the most balanced model, excelling in both precision and recall. YOLO-v8's F1-score of 0.985 demonstrated its robustness, especially in real-time scenarios requiring speed and accuracy.

C. *Implication of Findings*

The findings of this study underscore the effectiveness of advanced deep learning models in automating the classification of Malaysian and Indonesian batik designs. MobileNet v2, with its balanced precision (0.994), recall (0.994), and F1-score (0.994), emerged as the most suitable model for general-purpose classification tasks. Its lightweight architecture and high accuracy make it an ideal candidate for deployment in real-world applications, particularly in scenarios where computational efficiency is critical, such as mobile or embedded systems.

YOLO-v8, while slightly lower in precision (0.988) and recall (0.988) than MobileNet v2, demonstrated robust performance, achieving an F1-score of 0.985. Its real-time processing capabilities make it particularly suitable for applications that require rapid classification, such as interactive cultural exhibitions or dynamic cataloging systems. Despite its slightly higher misclassification rates compared to MobileNet v2, YOLO-v8 remains highly effective in distinguishing between batik patterns.

LeNet-5, with its simpler architecture, achieved a recall of 0.964 and an F1-score of 0.997. These results highlight its sensitivity in detecting positive cases, making it a potential candidate for scenarios, where identifying as many batik patterns as possible is more important than minimizing false positives. However, its lower precision (0.899) suggests limitations in handling intricate designs, reinforcing the need for more advanced architectures for tasks demanding high accuracy.

The high accuracy achieved by the models, particularly MobileNet v2, demonstrates the potential of deep learning in supporting cultural preservation efforts. Automated classification systems can significantly enhance the documentation, analysis, and promotion of traditional art forms, reducing reliance on manual expertise and providing scalable solutions for cataloging diverse batik designs. These tools can also be integrated into educational platforms to raise awareness about the cultural significance of batik, thereby contributing to its preservation and appreciation in the modern era.

Despite the promising results, several limitations should be acknowledged. The dataset, while comprehensive, does not fully capture the diversity of batik patterns found in Malaysia and Indonesia. Additionally, the models were trained on static images, which may not entirely reflect the nuances of batik created using different techniques and materials. Future research should focus on expanding the dataset to include a broader range of designs and exploring the use of ensemble models or advanced architectures, such as transformers, to further improve classification accuracy.

Moreover, integrating interpretability techniques, such as feature visualization, could provide deeper insights into how the models distinguish between batik patterns, paving the way for more refined and culturally informed classification systems.

V. CONCLUSION

This study developed a deep learning-based framework to classify Malaysian and Indonesian batik designs, utilizing a dataset of 1,825 images and evaluating three CNN architectures: MobileNet v2, YOLO-v8, and LeNet-5. Among these, YOLO-v8 achieved the highest accuracy at 98.80%, while MobileNet v2 balanced accuracy (97.79%) and computational efficiency, making it ideal for lightweight applications. LeNet-5, despite its simpler architecture, achieved 92.94% accuracy, demonstrating its utility in resource-constrained environments. However, the study has certain limitations. While the dataset is comprehensive, it could be further expanded to include more diverse batik patterns from various regions to enhance model generalization. Additionally, real-world deployment may introduce challenges such as variations in lighting, image quality, and pattern diversity, which could impact the models' performance.

This research lays the groundwork for future exploration into automated systems for batik classification. Future work could focus on integrating the models into a fully automated system with a graphical user interface (GUI) for practical applications in the batik industry. Additionally, the dataset could be expanded, and advanced techniques such as transformers, hybrid models, or transfer learning could be explored to improve classification performance and efficiency. These efforts will further facilitate the documentation, study, and promotion of batik art, ensuring its relevance and preservation in the modern era.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest regarding the publication of this paper.

AUTHOR CONTRIBUTION

The authors confirm contribution to the paper as follows: study conception and design: Nurulfajar Abd Manap, Lee Xiao Xuan; data collection: Lee Xiao Xuan; analysis and interpretation of findings: Nurulfajar Abd Manap, Lee Xiao Xuan; provision of deep learning expertise: Koushlendra Kumar Singh, Akbar Sheikh Akbari; draft manuscript preparation: Nurulfajar Abd Manap. All authors reviewed the findings and approved the final manuscript.

REFERENCES

[1] A. E. Minarno, I. Soesanti, and H. A. Nugroho, "Batik nitik 960 dataset for classification, retrieval, and generator", Data, vol. 8, no. 63, 2023.

[2] A. E. Minarno, A. S. Maulani, A. Kurniawardhani, F. Bimantoro, and N. Suciati, "Comparison of methods for batik classification using multi texton histogram", Telkomnika (Telecommunication Computing Electronics and Control), vol. 16, pp. 1358–1366, 2018.

[3] J. Rodrigo, "3 batik types by colours, patterns, & techniques in Malaysia", Accessed: June 06, 2023, [Online] Available: https://www.eksentrika.com/batik-types-colours-patterns-technique-malaysia/

[4] R. G. Tiwari, A. K. Agarwal, V. Jain and A. Kumar, "Batik classification in Indonesia: exploring its significance on tourism and economy", in Proc. of International Conference on Sustaining Heritage: Innovative and Digital Approaches (ICSH), 2023, pp. 119-124.

[5] S. Adyani, and C. A. Sari, "A web-based for Demak batik classification using vgg16 convolutional neural network", Journal Advance Sustainable Science, Engineering and Technology, vol. 6, no. 4, 2024.

[6] R. Azhar et al., "Batik image classification using sift feature extraction, bag of features and support vector machine," Procedia Computer Science, vol. 72, pp. 24-30, 2015.

[7] I. Riadi, A. Fadlil, and I. Putra, "Batik pattern classification using naïve bayes method based on texture feature extraction", Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika, vol. 9, no. 1, 2023.

[8] A. H. Rangkuti, A. Harjoko and A. Putra, "A novel reliable approach for image batik classification that invariant with scale and rotation using MU2ECS-LBP algorithm", Procedia Computer Science, no. 179, pp. 863-870, 2021.

[9] I. M. A. Agastya and A. Setyanto, "Classification of Indonesian batik using deep learning techniques and data augmentation", in Proc. of IEEE International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE), 2018, pp. 27-31.

[10] Winarno et al., "The hybrid features and supervised learning for batik pattern classification", ACM Journal on Computing and Cultural Heritage, vol. 12, pp. 1-13, 2024.

[11] D. M. S. Arsa and A. A. N. H. Susila, "VGG16 in batik classification based on random forest", in Proc. of International Conference on Information Management and Technology (ICIMTech), 2019, pp. 295-299.

[12] R. F. Alya, M. Wibowo, and P. Paradise, "Classification of batik motif using transfer learning on convolutional neural network (CNN)", J. Tek. Inform. (JUTIF), vol. 4, no. 1, pp. 161-170, 2023.

[13] C. U. Khasanah, E. Utami and S. Raharjo, "Implementation of data augmentation using convolutional neural network for batik classification", in Proc. of 28th International Conference on Cyber and IT Service Management (CITSM), 2020, pp. 1-5.

[14] D. Meranggi, N. Yudistira, and Y. Sari, "Batik classification using convolutional neural network with data improvements", International Journal on Informatics Visualization (JOIV), vol. 6, no. 1, 2022.

[15] R. O. Ogundokun, S. Misra, A. O. Akinrotimi, and H. Ogul, "MobileNet-SVM: a lightweight deep transfer learning model to diagnose BCH scans for IoMT-based imaging sensors", Sensors, vol. 23, 2023.

[16] N. Jamil, M. H. M. Fadhil, R. Hamzah, and M. I. Ramli, "Comparing MobileNet-SSD and YOLO v3 learning architecture for real-time driver's fatigue detection", International Journal of Academic Research in Business and Social Sciences, vol. 11, no. 12, pp. 2538–2549, 2021.

[17] U. M. M. P. K. Nawarathne, H. M. L. S. Kumari and H. M. N. S. Kumari, "Comparative analysis of jellyfish classification: a study using YOLOv8 and pre-trained models", in Proc. of International Research Conference on Smart Computing and Systems Engineering (SCSE), 2024, pp. 1-6.

[18] S. S. Kumar, S. K N, G. Sunil, V. P and S. S. Shankar, "A rapid object recognition scenario based on improved YOLO-V3 network", International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT), 2023, pp. 1-5.

[19] O. Pitsun et al., "Comparative analysis of CNN architecture for emotion classification on human faces", in Proc. of the First International Workshop of Young Scientists on Artificial Intelligence for Sustainable Development, 2024.

[20] K. Deepti and S. A. Deepthi, "Comparison of machine learning methods for classification and diagnosis of skin disease", in Proc. of IEEE 3rd International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC), 2023, pp. 1-5.