Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 10 No. 3 (2026) 776 - 786 e-ISSN: 2580-0760

Texture Analysis and Classification of Lontara Batik Using GLCM and k-Nearest Neighbor

Mohammad Yazdi Pusadan[1], Fuad Mahfud[2], Chairunnisa AR. Lamasitudju[3], Anisa Yulandari[4], Sabarudin Saputra[5]
[1,2,3,4,5]Depertment of Information Technology, Faculty of Engineering, Universitas Tadulako, Palu, Indonesia

*yazdi.diyanara@untad.ac.id

Abstract

Batik is an Indonesian cultural heritage recognized by UNESCO as a Masterpiece of the Oral and Intangible Heritage of Humanity. One distinctive regional variation is Batik Lontara from South Sulawesi, which features motifs derived from the traditional Bugis–Makassar Lontara script. However, the visual similarity between woven and non-woven (stamped) Batik Lontara makes manual identification difficult for the general public. This study proposes an image-based classification approach to distinguish woven and non-woven Batik Lontara using texture analysis and machine learning techniques. Texture features were extracted from batik images using the Gray Level Co-occurrence Matrix (GLCM), including contrast, homogeneity, energy, and entropy. The extracted features were then classified using the k-Nearest Neighbor (k-NN) algorithm with varying values of k (5, 7, and 9). The dataset consisted of Batik Lontara images captured using a smartphone camera and divided into training and testing sets using both hold-out and k-fold cross-validation schemes. Experimental results show that the highest classification accuracy of 86% was achieved using the k-NN algorithm with k = 9 under 5-fold cross-validation, while the hold-out method produced the best accuracy of 87% at k = 5. These results demonstrate that the proposed GLCM and k-NN-based approach is effective for classifying Batik Lontara types and has strong potential for supporting cultural heritage preservation through digital image processing.

Keywords: batik lontara; gray level co-occurrence matrix (GLCM); k-Nearest Neighbor (k-NN)

How to Cite: M. Y. Pusadan, F. Mahfud, C. A. Lamasitudju, A. Yulandari, and S. Saputra, "Texture Analysis and Classification of Lontara Batik Using GLCM and k-Nearest Neighbor ", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 10, no. 3, pp. 776 - 786, Jun. 2026.
*Permalink/*DOI: https://doi.org/10.29207/resti.v10i3.7497

Received: January 15, 2026
Accepted: June 1, 2026
Available Online: June 21, 2026

This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia

1. Introduction

Batik is an Indonesian cultural heritage in which almost every region possesses its own distinctive batik patterns and motifs. In 2009, UNESCO officially designated batik as a world cultural heritage created by the Indonesian people. In Makassar, South Sulawesi, there is a traditional form of batik known as Batik Lontara. Batik Lontara is a Bugis–Makassar batik that is unique to South Sulawesi and is characterized by motifs derived from the Lontara script.

The Lontara[1], [2] script is a traditional writing system used by the Bugis–Makassar community. This script is also considered to represent four major ethnic groups in Sulawesi—Bugis, Makassar, Toraja, and Mandar—all of which historically used the Lontara script. The term *Lontara* itself originates from the local Makassar language and is historically derived from the name of the lontar palm tree (*Borassus flabellifer*).

Batik Lontara began[3] to be developed as a local handicraft industry in 2008, introduced by Adi Pitoyo. Historically, the Lontara script was first created by Daeng Pamatte in the 16th century during his service as a Gowa official under Karaeng Tumaparisi Kallonna, who commissioned the creation of Makassar script. Along with technological advancements and increasing consumer demand, Batik Lontara production techniques have become increasingly modern, enabling artisans and entrepreneurs to produce batik on a larger scale.

However, in general, Indonesian society—particularly in Makassar and South Sulawesi—still has limited knowledge of how to distinguish between woven Batik Lontara and non-woven Batik Lontara fabrics. The identification of batik types should leverage technological advancements, particularly in digital image processing. Therefore, a system capable of

recognizing and classifying Batik Lontara types is required. Based on these issues, this research applies to the GLCM method for feature extraction and the KNN algorithm for image classification.

Based on the Regulation of the Minister of Industry of the Republic of Indonesia No. 74/M-IND/PER/9/2007[4] concerning the use of the Batikmark "Batik Indonesia," the application of batikmark aims to: (i) provide quality assurance for Indonesian batik products; (ii) increase consumer trust, both domestically and internationally, in the quality of Indonesian batik; (iii) legally protect Indonesian batik products from intellectual property and trade-related threats at both national and international levels; and (iv) establish a clear identity that enables the public to easily recognize Indonesian batik products.

Batik not only serves as a symbol of national cultural identity but also plays a significant role in Indonesia's social, economic, and cultural diplomacy sectors. Presidential Decree No. 33 of 2009[5] concerning the National Batik Day marked a major milestone in the formal recognition of batik as a cultural heritage that must be preserved and sustained[6]. This decree was closely linked to UNESCO's recognition of batik as a *Masterpiece of the Oral and Intangible Heritage of Humanity* on October 2, 2009.

Furthermore, the Indonesian government has strengthened the position of batik as a national flagship product through the issuance of Minister of Industry Regulation No. 10 of 2023[7] concerning the Development and Strengthening of the National Batik Industry. This regulation emphasizes batik not only as a cultural product but also as an economic driver with significant potential to create employment, promote regional economic growth, and enhance Indonesia's competitiveness in the global market.

Batik also contributes to the achievement of the Sustainable Development Goals (SDGs) [8-10]. by supporting economic, social, and environmental sustainability. In recent years, batik has gained attention in the context of green entrepreneurship, which integrates sustainability values into business practices. Additionally, the Penta-Helix collaboration model offers an innovative approach to maximizing batik's potential as a driver of inclusive and sustainable community development[11].

While modern Deep Learning (DL) architectures, such as Convolutional Neural Networks (CNN), have become the standard for many computer vision tasks, their application in this study is limited by the available data. Deep Learning[12] models typically require thousands of samples to achieve robust generalization and avoid overfitting. Given that this study utilizes a focused dataset of 150 Batik Lontara images, classical machine learning techniques offer a more reliable alternative. GLCM[13] and k-NN provide a low-complexity framework that is less prone to overfitting on small datasets while maintaining high interpretability of the extracted texture features, which is essential for cultural heritage documentation.

Based on the identified problem and research gap, this study proposes a texture-based classification approach to distinguish woven and non-woven Batik Lontara using Gray Level Co-occurrence Matrix feature extraction and the k-Nearest Neighbor algorithm. The proposed approach extracts four texture features, namely contrast, homogeneity, energy, and entropy, from Batik Lontara images. These features are then classified using k-NN with different neighborhood values to evaluate the effect of parameter variation on classification performance. The model is evaluated using confusion matrix analysis and k-fold cross-validation to provide a more reliable measurement of classification accuracy.

The main contributions of this article are threefold. First, this study addresses a specific and underexplored problem in cultural heritage image analysis, namely the classification of woven and non-woven Batik Lontara. Second, it demonstrates the use of interpretable GLCM texture features to represent measurable differences between the two batik categories. Third, it evaluates the performance of the k-NN classifier under different neighborhood values and validation schemes, thereby providing empirical evidence on the suitability of a simple and explainable machine learning approach for Batik Lontara classification. The remainder of this paper is organized as follows. Section 2 describes the research methodology, including image acquisition, preprocessing, feature extraction, classification, dataset preparation, and evaluation procedures. Section 3 presents the results and discussion. Finally, the conclusion summarizes the findings and proposes future research directions.

Recent technological advancements, particularly in artificial intelligence (AI), have opened new opportunities for preserving and developing batik through innovative approaches. Pattern recognition using AI has emerged as a strategic method to improve efficiency, accuracy, and creativity in the batik industry. In this study, the GLCM method and the k-NN algorithm are employed as AI-based techniques to enhance the accuracy and precision of Batik Lontara motif classification.

2. Methods

The objective of this research is to classify Batik Lontara fabrics using the Gray Level Co-occurrence Matrix (GLCM) [14], [15] method for texture feature extraction and the k-Nearest Neighbor (k-NN) [16] algorithm for classification. The research stages are illustrated in the system flowchart shown in Figure 1.

The system development follows standard digital image processing procedures, consisting of image acquisition, preprocessing, feature extraction, and classification. The stages of system development in this research are

presented sequentially in Figure 2. The proposed system is designed based on digital image processing procedures employing the GLCM[17], [18] method and the k-NN[19], [20] algorithm.

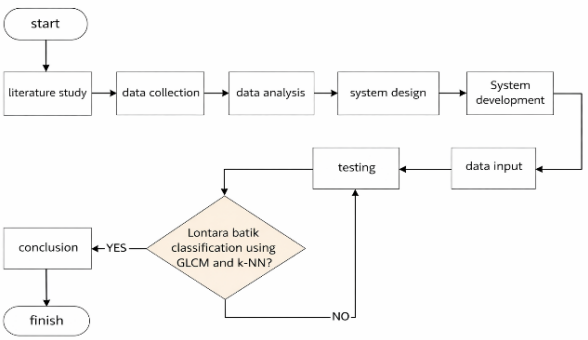


Figure 1.Research Flow Diagram

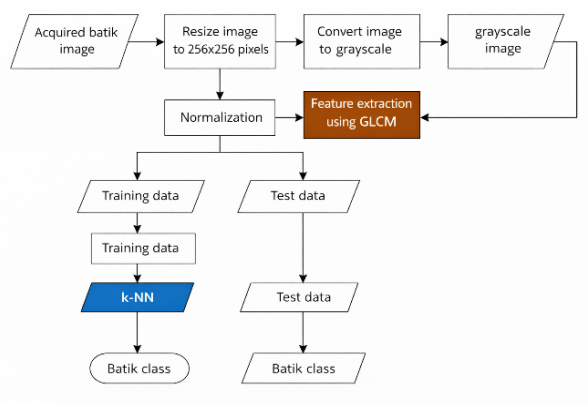


Figure 2. Stages of System Development

Image acquisition[21], [22] is the process of collecting and storing image data in a database. In this study, the acquired images consist of woven Batik Lontara and non-woven Batik Lontara. Image acquisition was performed using a digital camera, with a focus on capturing clear batik motifs.

Preprocessing[23], [24] consists of cropping, resizing, grayscale transformation, and quantization. Cropping removes irrelevant image areas, while resizing normalizes all images to a resolution of 256 × 256 pixels. RGB-to-grayscale transformation simplifies pixel values into a single intensity channel. Quantization is applied to reduce computational complexity.

Feature extraction aims to identify distinguishing characteristics of image objects. This study utilizes texture feature extraction based on the GLCM[18] method. The extraction process includes grayscale-to-matrix conversion, defining spatial relationships between pixels based on distance and angle, matrix symmetrization, normalization, and second-order statistical feature computation. The extracted features include contrast, homogeneity, energy, and entropy. This is illustrated in Figure 3.

The choice of GLCM for feature extraction and k-NN for classification represents a deliberate methodological strategy focused on efficiency and reliability. Unlike 'black-box' Deep Learning[12] models, GLCM allows for the explicit extraction of second-order statistical texture features—such as contrast, homogeneity, energy, and entropy —which directly correspond to the physical characteristics of woven versus stamped fabrics. Furthermore, k-NN is highly effective in low-dimensional feature spaces and does not require a lengthy training phase, making it computationally efficient for smartphone-based applications. This combination ensures that the system remains robust and accurate even with a limited number of training samples, achieving competitive performance without the computational overhead or data requirements of modern neural networks.

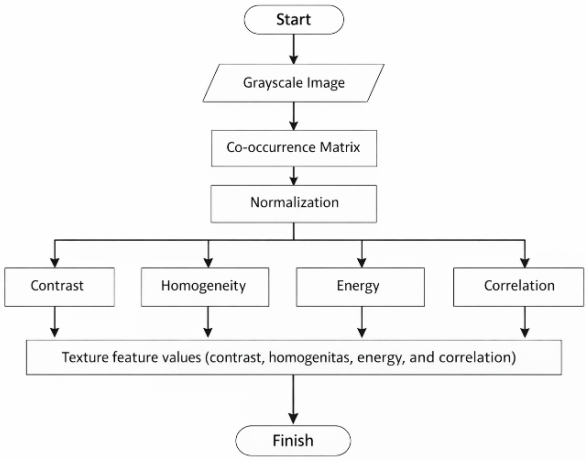


Figure 3. The feature extraction process using the GLCM method

2.1 Classification

The classification stage employs the k-NN algorithm to classify images into woven and non-woven Batik Lontara categories. Classification is based on Euclidean distance, with values of 5, 7, and 9 evaluated to determine optimal performance.

The stages leading to the classification process can be described through the following explanation of the GLCM method: (i) converting the grayscale image into a matrix representation; (ii) determining the spatial relationship between pixels based on predefined distances and angles; (iii) summing the co-occurrence matrix with its transpose to obtain a symmetric matrix; (iv) normalizing the symmetric matrix to transform it into a probability-based matrix; (v) second-order statistical feature extraction is performed using the normalized GLCM matrix. Four second-order statistical features are utilized, namely contrast, homogeneity, energy, and correlation; (vi) classification. At this stage, the algorithm classifies objects into predefined classes based on the extracted feature characteristics using the k-Nearest Neighbor (k-NN) algorithm. The class categories consist of woven Lontara batik and non-woven Lontara batik. The process of the k-NN algorithm is illustrated in Figure 4.

The explanation of the k-NN Algorithm flowchart shown in Figure 4 is as follows: a) determining the value of *k*, which represents the number of nearest

neighbors; b) the k nearest neighbors are identified by calculating the distance between each training data point and the test data using the Euclidean distance formula; c) after computing the distances between the training data and the test data, the distances are sorted in ascending order; and d) the smallest distance values are selected according to the predetermined value of k. Subsequently, the class label is determined based on the majority class among the selected neighbors, which is then assigned as the class of the new test data.

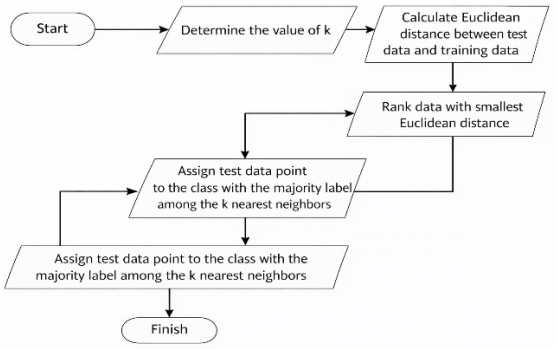


Figure 4. Classification is performed using the k-Nearest Neighbor (k-NN) approach

Data Input, after the system development is completed, the next step is data input into the system. The data entered consist of *Lontara batik* images. Testing, this study employs black-box[25], [26] testing, which is a system testing technique that focuses on verifying the functional specifications of the system. In addition, the K-Nearest Neighbors (K-NN) algorithm is evaluated using test data in the form of *Lontara batik* images to compute the classification accuracy. The accuracy is calculated using a confusion matrix[27-30], as presented in Table 1.

Table 1. Confusion matrix

TC = True Class FC = False Class		Classification C1	 C2
Target Class	C1	TC1	FC1
	C2	FC2	TC2

Description: C1 = Non-woven Lontara Batik (LNT) and C2 = Woven Lontara Batik (LT).

The equations related to the confusion matrix are as follows. Accuracy represents the overall ability of the k-Nearest Neighbor model to correctly classify Lontara Batik images into their respective categories, namely woven and non-woven Lontara Batik. In this study, accuracy indicates the proportion of correctly classified images based on texture features extracted using the Gray Level Co-occurrence Matrix. A higher accuracy value reflects better overall classification performance of the proposed model. The formula used to calculate accuracy is presented in Equation 1.

$$Accuracy = \frac{(TP+TN)}{TP+TN+FP+FN} \quad (1)$$

Error rate represents the proportion of images that are incorrectly classified by the model. It is the complement of accuracy and provides an indication of the classification errors produced by the k-Nearest Neighbor model. A lower error rate indicates better classification performance and greater reliability of the texture-based classification approach. The classification error rate was calculated using the formula presented in Equation 2.

$$Error\ Rate = 100\% - Accuracy \quad (2)$$

Precision measures the reliability of the model in predicting a particular class. In the context of Lontara Batik classification, precision indicates the extent to which images predicted as belonging to a specific class are members of that class. A high precision value implies that the model produces fewer false positive classifications. The precision calculation formula is expressed in Equation 3.

$$Precision = \frac{TP}{(TP+FP)} \quad (3)$$

2.2 Dataset

The batik image dataset consists of a collection of cropped batik image patches belonging to the same class, which may share identical or different motifs as shown in Figure 5. A total of 85 images of woven Lontara batik were obtained from two types of woven Lontara batik fabrics, with each fabric type contributing 42 and 43 images, respectively. Similarly, 85 images of non-woven Lontara batik were collected from two types of non-woven (stamped) Lontara batik fabrics, with each fabric type also contributing 42 and 43 images. Consequently, the data set comprises a total of 170 images**.**

The dataset used in this study was acquired directly using a smartphone camera. It contains two classes, representing two types of batik, namely woven Lontara batik and non-woven Lontara batik**.**



Figure 5. Examples of the dataset: images (1, 2) represent woven batik, while images (3, 4) represent non-woven batik

The entire data set used in this study consists of 170 images. The dataset is divided into two subsets, namely training data and testing data. The distribution of the dataset is presented in Table 2. The data grouped as training data consist of 170 images.

Table 2. The dataset comprises 170 images

No.	Dataset type	Number of images
1	Training data	120
2	Testing data	50
Total		170

The training data consists of 120 images and is used as reference data to provide knowledge or prior learning before performing testing on the testing data. The distribution of the training data based on batik types is presented in Table 3. The data grouped as testing data consists of 60 images. The testing data are used to

evaluate whether the predictions generated by each algorithm are accurate or not.

Table 3. A total of 120 images is used as training data

No.	Batik Type	Number of Training Images
1	Woven Lontara Batik	60
2	Non-woven Lontara Batik	60
Total		120

The number of testing data based on batik types, as presented in Table 4, is 50, consisting of woven Lontara batik and non-woven Lontara batik.

Table 4. Distribution of Testing Data by Batik Type

No.	Batik Type	Number of Testing Images
1	Woven Lontara Batik	25
2	Non-woven Lontara Batik	25
Total		50

The proposed method combines standard and interpretable image-processing techniques. GLCM is a standard statistical texture extraction method that is appropriate for fabric image analysis because it measures spatial relationships between gray-level intensities. This is relevant to Batik Lontara classification because woven and non-woven fabrics differ in surface texture, motif regularity, and local intensity distribution. The k-NN algorithm was selected because it is simple, transparent, and effective for low-dimensional feature spaces. Since this study uses four extracted texture features and a relatively limited dataset, k-NN is suitable because it does not require a complex training process and is less dependent on large-scale data than deep learning models.

This methodological design ensures that the classification process is explainable and reproducible. Each prediction can be traced back to measurable texture features and similarity distances between testing and training samples. Therefore, the proposed GLCM and k-NN approach is appropriate for supporting digital documentation and classification of Batik Lontara, particularly in small-dataset cultural heritage applications.

3. Results and Discussions

This section presents the results obtained from the texture feature extraction and classification processes of Lontara batik images. The discussion focuses on the interpretation of GLCM-based texture features and the performance of the classification model in distinguishing woven and non-woven Lontara batik images.

3.1 Results of Texture Feature Extraction of Lontara Batik Images

Texture feature extraction, including contrast, homogeneity, energy, and entropy, was performed at an angle of 0° with a pixel distance of 1, based on Equations (2.1), (2.2), (2.3), and (2.4). This process yields texture feature values of Lontara batik images, which are subsequently used as discriminative parameters to distinguish between woven Lontara batik images and non-woven Lontara batik images.

The resulting value ranges for each texture feature are as follows: contrast values range from 309,840 to 6,044,580, homogeneity values range from 0.000000202961 to 0.126909, energy values range from 0.00134678 to 0.173532, and entropy values range from 2.17595 to 6.86322. The range of texture feature extraction values for Lontara batik images is presented in Table 5.

Table 5. Range of extracted texture features of Lontara batik images.

Category (a)	Contrast (b)		Homogenity (c)	
	MIN	MAX	MIN	MAX
Non-woven Lontara Batik	389794	5330740	0.000000202961	0.126909
Category	Contrast		Homogenity	
	MIN	MAX	MIN	MAX
Woven Lontara Batik	309840	6044580	0.000000755659	0.00128551
MIN	309840		0.000000202961	
MAX	604458		0.126909	
Category (a)	Energy (d)		Entropy (e)	
Non-woven Lontara Batik	MIN	MAX	MIN	MAX
Category	0.00134678	0.173532	2.17595	6.8622
	Energy		Entropy	
	MIN	MAX	MIN	MAX
Woven Lontara Batik	0.00195	0.0364668	3.77646	6.735
MIN	0.001347		2.17595	
MAX	0.173532		6.86322	

3.2 Results of Lontara Batik Image Classification

The classification of Lontara batik images was performed using the k-Nearest Neighbor (k-NN) algorithm with varying values of k, namely 5, 7, and 9, based on Equation (2.8). This process resulted in different Euclidean distance threshold values for each k. For k = 5, the Euclidean distance values range from 703,878 to 634,368, whereas for k = 7, the distance values range from 703,878 to 3,238,190. The Euclidean distance thresholds for each k value are presented in Table 6.

Table 6. Classification Results of Lontara Batik Images

Category	K = 5		K = 7	
	MIN	MAX	MIN	MAX
Non-woven Lontara Batik	703878	634368	703878	3238190
Woven Lontara Batik	697098	847158	697098	440628
MIN	703878		703878	
MAX	634368		3238190	

3.3. Discussion

The Lontara Batik classification system in this study begins with several stages. The first stage is image feature extraction, in which the Gray Level Co-occurrence Matrix (GLCM) method is employed. To obtain texture features of Lontara batik, direction of 0° and a distance of 1 pixel are used. The extracted texture features include contrast, homogeneity, energy, and entropy.

After the extraction process, Lontara batik classification is performed based on the obtained texture features. The classification process uses the k-Nearest Neighbor (k-NN) algorithm with values of k = 5, 7, and 9. Subsequently, system testing is conducted based on the classification results to determine the accuracy level of the Lontara batik classification system. In this study, system testing is carried out using two approaches, namely Confusion Matrix testing and K-Fold Cross-Validation testing.

The Lontara Batik classification system based on texture features consists of two stages: the training stage and the testing stage. The Lontara Batik image data are divided into training data and testing data with a ratio of 70% for training data and 30% for testing data. This data partitioning method is referred to as the holdout method.

The training dataset consists of 119 images out of 170 images, comprising 60 images of non-woven Lontara batik and 59 images of woven Lontara batik. The testing dataset consists of 51 images out of 170 images, including 26 images of non-woven Lontara batik and 25 images of woven Lontara batik.

The testing results and accuracy calculations in this study are presented in a confusion matrix, as described in Subsection 2.2.8. Based on the explanation in Subsection 2.2.9, a two-class confusion matrix is constructed in accordance with the classified categories. The confusion matrix used in this study is presented in Table 7.

Table 7. Confusion Matrix

Confusion Matrix 2 class		Classification	
		Positif	Negatif
Class Target	Positif	True Positif (TP)	False Positif (FP)
	Negatif	False Negatif (FN)	True Negatif (TN)

The application of the confusion matrix in Lontara batik classification is shown in Table 8.

The confusion matrix consists of two columns representing the predicted classes and two rows representing the target (actual) classes. Each row and column corresponds to one class. The rows and columns are defined as follows: (i) Lontara Non-Woven Column/Row: the class or category of non-woven Lontara batik; dan (ii) Lontara Woven Column/Row: the class or category of woven Lontara batik

Table 8. Confusion Matrix of Batik Classification Results

Confusion Matrix two class		Classification	
		Positif	Negatif
kelas target	Non-woven Lontara Batik	true non woven (tnt)	false non woven (fnt)
	Woven Lontara Batik	false woven(ft)	true woven (tt)

This training process employs the k-Nearest Neighbor (k-NN) algorithm to perform the classification of Lontara batik images into non-woven and woven categories. The values of *k* (number of neighbors) used in this study are 5, 7, and 9. The training phase utilizes 118 training images, consisting of 59 images of non-woven Lontara batik and 59 images of woven Lontara batik. The testing phase uses 52 images, comprising 26 images of non-woven Lontara batik and 25 images of woven Lontara batik.

Thus, the expected number of correctly classified images for each category—non-woven Lontara batik and woven Lontara batik—is 25 images per class. However, several non-woven Lontara batik images are misclassified as woven Lontara batik, and vice versa. Classification errors are influenced by several factors, particularly the similarity or narrow range of the extracted texture feature values between Lontara batik classes. As a result, the computed Euclidean distance between feature vectors becomes small or close to that of other categories, leading to misclassification.

k = 5 (70% Training and 30% Testing): Classification using the k-NN algorithm with k = 5 involves sorting the classification results and selecting the five nearest neighbors based on the smallest Euclidean distances. Based on the classification results with k = 5, the non-woven Lontara batik class, which targets 26 test images, has 4 images misclassified as woven Lontara batik. Consequently, 22 images are correctly classified as non-woven Lontara batik.

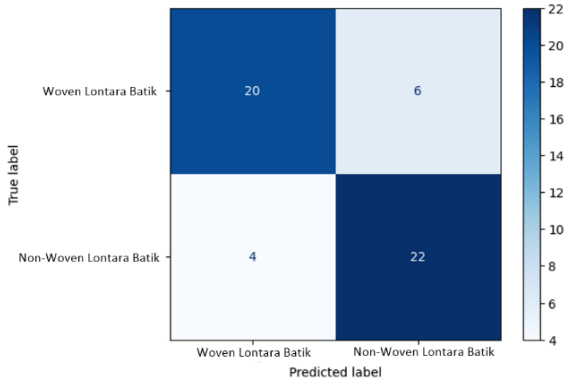


Figure 6. Confusion matrix for the k-NN model with k = 5

For the woven Lontara batik class, which also targets 25 test images, 6 images are misclassified as non-

woven Lontara batik. Therefore, 20 images are correctly classified as woven Lontara batik. The confusion matrix for k = 5 is presented in Figure 7.

Using Equations (1), (2), and (3), the accuracy, error rate, and precision were subsequently calculated as follows: accuracy: 80.77%, error rate: 19.23%, precision of non-woven Lontara: 0.916, and precision of woven Lontara: 0.769. The average precision value obtained was: 84.25%.

KNN classification using k = 5 with 5-fold cross-validation means that the dataset is divided into five folds, and the classification results are obtained from five cross-validation experiments. In this approach, the KNN classifier determines the class label based on the five nearest image samples with the smallest Euclidean distance.

Based on the KNN classification results using k = 5 with 5-fold cross-validation, the accuracy obtained in the first fold was 75.00%, the second fold achieved 87.50%, and the third fold produced an accuracy of 91.66%, which represents the highest accuracy among all folds. Meanwhile, the fourth fold resulted in an accuracy of 79.16%, and the fifth fold achieved 86.95%. The average accuracy across all folds was 84.05%. The accuracy results obtained using k = 5 with 5-fold partitioning are presented in Table 9.

Table 9. k = 5 with 5-fold partitioning

Fold	Accuracy
Fold 1	75.00%
Fold 2	87.50%
Fold 3	91.66%
Fold 4	79.16%
Fold 5	86.95%
Average	84.05%

KNN classification using k = 7 involves sorting the classification results and selecting the seven nearest image samples based on the smallest Euclidean distance. Based on the classification results using k = 7, there were 26 target images belonging to the non-woven Lontara category, of which 4 images were misclassified as belonging to the woven Lontara category. Therefore, 22 images were correctly classified as non-woven Lontara.

For the woven Lontara category, 26 target images were used; however, 6 images were misclassified as belonging to the non-woven Lontara category. As a result, 20 images were correctly classified as woven Lontara.

The confusion matrix for k = 7 is presented in Figure 7. The obtained values are as follows: accuracy: 80.77%, error rate: 19.23%, non-woven Lontara: 0.19, and woven Lontara: 0.769. The average precision value obtained was: 84.25%.

KNN classification using k = 9 involves sorting the classification results and selecting the nine nearest image samples based on the smallest Euclidean distance. Based on the classification results using k = 9, there were 26 target images belonging to the non-woven Lontara category, of which 3 images were misclassified as belonging to the woven Lontara category. Consequently, 23 images were correctly classified as non-woven Lontara.

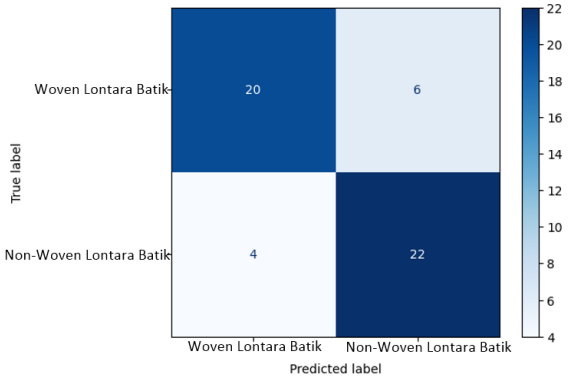


Figure 7. Confusion Matrix untuk k=7

For the woven Lontara category, 26 target images were used; however, 6 images were misclassified as belonging to the non-woven Lontara category. As a result, 20 images were correctly classified as woven Lontara. The confusion matrix for k = 9 is presented in Figure 8.

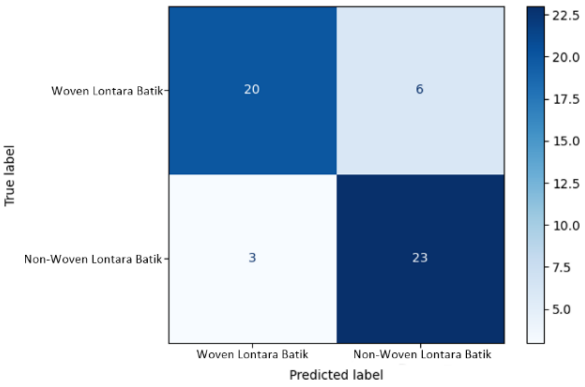


Figure 8. Confusion matrix for k = 9

The obtained values are as follows: accuracy: 82.70%, error rate: 17.30%, non-woven Lontara: 0.884, and woven Lontara: 0.769. The average precision value obtained was: 82.65%.

Based on the confusion matrix calculations for k = 5, k = 7, and k = 9, the accuracy and error rate values obtained for each k value are presented in Table 10.

Table 10. Confusion matrix testing results

K Value in k-NN Algorithm	Accuracy	Error Rate
K = 5	80.77%	19.23%
K = 5 with 5-fold partition (Fold-3)	91.66%	8.34%
K = 7	80.77%	19.23%
K = 9	82.70%	17.30%

The purpose of the confusion matrix testing stage is to determine the accuracy and error rate, which are used as indicators of the performance and reliability of the KNN algorithm in classifying woven Lontara and non-woven Lontara images.

The testing results show that k = 5 produced an accuracy of 80.77%, while k = 5 with 5-fold partitioning (Fold-3) achieved an accuracy of 91.66%. The test using k = 7 resulted in an accuracy of 80.77%, whereas k = 9 produced an accuracy of 82.70%. These results indicate that increasing the value of k generally improves classification accuracy; however, the highest accuracy in this experiment was obtained using k = 5 with 5-fold cross-validation on Fold-3.

Validation of the experimental results was conducted using the k-fold cross-validation method. The dataset consisted of 150 image samples, which were divided into training and testing sets. The value of k = 5 was used, resulting in five experimental iterations, where in each iteration one partition was used as the test set and the remaining partitions were used as the training set.

The number of test samples in each iteration was calculated as: $k = (170/5)$ x 30. Where, k: number of partition (5), N: total dataset (170 samples), and 35 samples were used as test data in each iteration.

The partitioning process was performed randomly; however, the class labels were arranged sequentially to ensure a balanced distribution of categories in each partition. The results of the data partitioning are provided in the appendix.

Validation was further performed by calculating the average accuracy of the KNN algorithm using varying neighborhood values (k = 5, 7, and 9) to determine the optimal k value. Using Equation (2.14), the results of the k-fold cross-validation are presented in Table 11.

Table 11. k-Fold cross-validation results

K Value	Fold	Accuracy
K = 5	Fold-1	75.00%
	Fold-2	87.50%
	Fold-3	91.66%
	Fold-4	79.16%
	Fold-5	88.95%
	Average Accuracy	84%
K = 7	Fold-1	91.16%
	Fold-2	87.50%
	Fold-3	83.33%
	Fold-4	79.16%
	Fold-5	82.60%
	Average Accuracy	85%
K = 9	Fold-1	95.83%
	Fold-2	87.50%
	Fold-3	83.33%
	Fold-4	91.66%
	Fold-5	95.65%
	Average Accuracy	90%

The highest validation accuracy was obtained in Fold-1 using k = 9, while the lowest validation accuracy was observed in Fold-1 using k = 5. The highest average accuracy was achieved using k = 9, with an average accuracy of 90%, whereas the lowest average accuracy was obtained using k = 5, with an average accuracy of 84%. Based on these validation results, the optimal k value for this study is k = 9, using Fold-1 as the test set and the remaining partitions as training data.

The accuracy results obtained using the confusion matrix and those obtained using k-fold cross-validation do not show a significant difference. A comparison between the accuracy obtained using the hold-out method and k-fold cross-validation is presented in Table 12.

Table 12. Accuracy comparison between hold-out method and k-fold cross-validation

Training Method	K Value in KNN Algorithm	Accuracy (%)
K-Nearest Neighbor	K = 5	80.77%
	K = 9 with 5-fold partition (Fold-1)	95.83%
	K = 7	80.77%
	K = 9	82.70%

The highest accuracy was achieved using k = 9, both with and without k-fold cross-validation. This result occurs because a larger neighborhood size provides more comparisons during classification, leading to improved accuracy. However, imbalanced feature ranges across parameters can affect classification performance. Therefore, data normalization or preprocessing is required to improve the accuracy of image dataset classification.

This study contributes to cultural heritage image analysis by addressing a more specific and underexplored problem in batik classification, namely the distinction between woven and non-woven Batik Lontara. Unlike previous studies that mainly focused on motif recognition or general batik pattern classification, this research demonstrates that production-type differences in regional batik can be identified through interpretable texture features extracted using GLCM and classified using k-NN. The findings indicate that classical machine learning remains relevant for small-scale, domain-specific image classification, particularly when large datasets for deep learning are unavailable and when interpretability is required. The proposed method offers a computationally efficient and explainable baseline because classification decisions can be traced to measurable texture characteristics and distance-based similarity. Practically, this approach has potential value for local communities, artisans, consumers, and cultural institutions by supporting digital documentation, education, and the early development of automated Batik Lontara identification systems. However, the results should be interpreted as an initial baseline rather than a final authentication model because the dataset is still limited, image acquisition conditions may affect performance, and the four GLCM features may not fully capture highly similar visual patterns.

Future research should therefore expand the dataset, apply feature normalization, incorporate additional descriptors such as color, shape, Local Binary Pattern, and multi-angle GLCM, compare k-NN with other classifiers, and develop mobile or web-based prototypes to support broader cultural heritage preservation and community-oriented applications.

4. Conclusions

Based on the research conducted using the Grey Level Co-Occurrence Matrix (GLCM) method for texture feature extraction and the k-Nearest Neighbor (KNN) algorithm for classification, the following conclusions can be drawn: (i) the k-Nearest Neighbor algorithm was successfully applied to classify non-woven Lontara and woven Lontara images based on texture features extracted using the Grey Level Co-Occurrence Matrix method in an effective and efficient manner. Four texture features were utilized in this study, namely contrast, homogeneity, energy, and entropy. The neighborhood parameter (k) in the KNN algorithm was varied, with values of k = 5, 7, and 9 being evaluated; and (ii) based on the experimental results obtained using confusion matrix evaluation with the hold-out method, where the dataset was divided into 70% training data and 30% testing data, the highest classification accuracy was achieved at k = 9, with an accuracy of 86%. Furthermore, additional experiments were conducted to improve classification performance by applying k-fold cross-validation. The best result was obtained using k = 5 with a 5-fold partition, achieving an accuracy of 87%.

Although the results demonstrate that the proposed approach performs well, further improvements can be made. Future research should consider increasing the number and diversity of Lontara batik images to develop a more robust classification system. Additionally, implementing data normalization techniques is recommended to improve classification accuracy.

This study demonstrates that the distinction between woven and non-woven Batik Lontara can be objectively identified through quantitative texture analysis rather than relying solely on manual visual observation. The use of GLCM features—contrast, homogeneity, energy, and entropy—successfully captures measurable texture differences between the two production types, while the k-NN algorithm provides a simple, interpretable, and computationally efficient classification mechanism. The findings confirm that texture-based features are relevant for Batik Lontara production-type recognition and that classifier performance is influenced by the selection of neighborhood values and validation strategies, particularly in limited dataset conditions. Scientifically, this study contributes to cultural heritage image analysis by extending batik classification beyond general motif recognition toward a more specific production-based classification problem.

Practically, the proposed approach has potential value for digital documentation, cultural education, consumer information, artisan support, and the early development of automated Batik Lontara identification systems. However, the results should be regarded as an initial baseline rather than a final authentication model because dataset size, image acquisition conditions, motif similarity, and the limited number of texture descriptors may affect classification reliability. Future research should expand the dataset, improve preprocessing and feature normalization, incorporate additional descriptors such as color, shape, Local Binary Pattern, and multi-angle GLCM, compare k-NN with other classifiers, and develop mobile or web-based applications to strengthen the role of digital image processing and interpretable machine learning in preserving and recognizing local cultural heritage.

Further development may also include exploring other feature selection methods to identify texture features that have a greater impact on classification performance. Moreover, future studies can be expanded by incorporating additional feature extraction methods, such as shape features and color features, as well as experimenting with other classification algorithms to enhance the overall performance of Lontara batik classification.

Acknowledgements

The author would like to express sincere gratitude to the Faculty of Engineering, Tadulako University for providing laboratory facilities, infrastructure, and intelligent computing resources that supported the big data processing in this research. The author also extends appreciation to all parties who contributed to the availability of research data, including local artisans and the Regional Tourism Office, for providing access to historical information and datasets used in this study

References

[1] D. T. Afrianto, “Lontara Batik as an Affirmation of Bugis-Makassar Cultural Identity and Legitimacy,” *J. Pakarena*, vol. 3, no. 2, p. 22, 2020, doi: 10.26858/p.v3i2.13062.

[2] A. Asnawi, M. Hadrawi, A. Amrawaty, and S. Nurlaelah, “The public perception of the Bali cattle-Lontara motif cloth product as local wisdom in South Sulawesi,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 788, no. 1, 2021, doi: 10.1088/1755-1315/788/1/012206.

[3] Trisakti Akbar, Muhammad Fajar B, Muhammad Akbar Amir, Andi Akram Nur Risal, Nur Azizah Ayu Safanah, and M. Miftach Fakhri, “Sulsel Typical Batik Motif Classification Using Neural Network Method With Glcm Feature Extraction,” *J. Deep Learn. Comput. Vis. Digit. Image Process.*, pp. 24–33, 2023, doi: 10.61255/decoding.v1i1.49.

[4] Pemerintah RI, “Ministerial Regulation Concerning the Utilization of the ‘Batik Indonesia’ Batikmark,” 2007.

[5] Presiden Republik Indonesia, “Presidential Regulation of the Republic of Indonesia No. 33 of 2009 Regarding the Designation of National Batik Day,” *Pres. Republik Indones.*, 2009.

[6] KEMENTERIAN PERDAGANGAN, “Regulation No. 53/M-DAG/PER/7/2015 of the Indonesian Ministry of Trade Regarding the Export of Batik and Batik-Motif Textile Products,” *PERMENDAG RI No. 53/M-DAG/PER/7/2015,* 2015.

[7] Kementerian Perindustrian, “Regulation No. 10/2023 of the Indonesian Ministry of Industry on Green Industry Standards in the Batik Sector,” 2023.

[8] E. T. Yuniar, H. Susiatiningsih, and F. Eko Wahyudi, “Strategies for Achieving the Sustainable Development Goals in Pekalongan City,” *J. Int. Relations*, vol. 8, pp. 217–231, 2022, doi: https://doi.org/10.14710/jirud.v8i2.33512

[9] S. Sudarmiatin and R. Fitri, “Green Entrepreneurship Model in The Batik Industry: a Solution Towards Sustainable Development goals (SDGs),” *J. Bus. Manag. Rev.*, vol. 4, no. 11, pp. 857–870, 2023, doi: 10.47153/jbmr411.8392023.

[10] T. R. Simanjuntak, R. A. Yanuartha, S. N. Sukmi, and P.

Hergianasari, "Penta-Helix Collaboration Model in Community Development (Batik Making in Sangiran)," *Int. J. Multicult. Multireligious Underst.*, vol. 10, no. 10, p. 120, 2023, doi: 10.18415/ijmmu.v10i10.5103.
[11] KEMENPAREKRAF, "Regulation of the Minister of Tourism and Creative Economy Number 11 of 2022," *Peratur. MENTERI PARIWISATA DAN Ekon.*, 2022.
[12] I. H. Rather, S. Kumar, and A. H. Gandomi, *Breaking the data barrier: a review of deep learning techniques for democratizing AI with small datasets*, vol. 57, no. 9. Springer Netherlands, 2024. doi: 10.1007/s10462-024-10859-3.
[13] Z. Y. Shen, H. De Shen, and F. He, "A hybrid deep learning-based model for enhanced feature representation in image retrieval systems," *Egypt. Informatics J.*, vol. 31, no. February, p. 100717, 2025, doi: 10.1016/j.eij.2025.100717.
[14] D. Wijaya and A. R. Widiarti, "Batik classification using KNN algorithm and GLCM features extraction," *E3S Web Conf.*, vol. 475, pp. 1–13, 2024, doi: 10.1051/e3sconf/202447502012.
[15] Fathul Am and E. I. Sela, "Classification of Pekalongan Batik Based on Image Features Using the GLCM Method and Backpropagation Neural Networks," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 1, pp. 614–621, 2024, doi: 10.35870/jimik.v5i1.532.
[16] A. E. Minarno, M. Y. Hasanuddin, and Y. Azhar, "Batik Images Retrieval Using Pre-trained model and K-Nearest Neighbor," *Int. J. Informatics Vis.*, vol. 7, no. 1, pp. 115–121, 2023, doi: 10.30630/joiv.7.1.1299.
[17] A. E. Minarno, I. Soesanti, and A. Nugroho, "Batik Classification using Microstructure Co-occurrence Histogram," *Int. J. Informatics Vis.*, vol. 8, no. March, pp. 134–140, 2024, doi: 10.62527/joiv.8.1.2152.
[18] Y. Sari, M. Alkaff, and R. A. Pramunendar, "Classification of coastal and Inland Batik using GLCM and Canberra Distance," *AIP Conf. Proc.*, vol. 1977, no. June, 2018, doi: 10.1063/1.5042901.
[19] Darussalam and G. Arief, "Texture Feature Extraction in Grape Image Classification Using K-Nearest Neighbor," *Resti*, vol. 1, no. 1, pp. 19–25, 2018, doi: 10.29207/resti.v6i5.4137.
[20] N. L. Chusna, N. Wiliani, and A. F. Abdillah, "Enhancing Ulos Batik Pattern Recognition through Machine Learning: A Study with KNN and SVM," *J. Ris. Inform.*, vol. 6, no. 3, pp. 175–184, 2024, doi: 10.34288/jri.v6i3.311.
[21] S. Higashiyama *et al.*, "Investigation of the effectiveness of no-reference metric in image evaluation in nuclear medicine," *PLoS One*, vol. 19, no. 11 November, pp. 1–23, 2024, doi: 10.1371/journal.pone.0310305.
[22] K. Zhang *et al.*, "Image acquisition as novel colonoscopic quality indicator: a single-center retrospective study," *Front. Oncol.*, vol. 13, no. May, pp. 1–8, 2023, doi: 10.3389/fonc.2023.1090464.
[23] H. Talebi and P. Milanfar, "Learning to Resize Images for Computer Vision Tasks," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 487–496, 2021, doi: 10.1109/ICCV48922.2021.00055.
[24] K. R. Kulkarni and S. B. Bagal, "Facial expression recognition," *Proc. - IEEE Int. Conf. Inf. Process. ICIP 2015*, pp. 535–539, 2016, doi: 10.1109/INFOP.2015.7489442.
[25] B. Kim and R. F. Barber, "Black-box tests for algorithmic stability," *Inf. Inference*, vol. 12, no. 4, pp. 2690–2719, 2023, doi: 10.1093/imaiai/iaad039.
[26] P. Long and J. Zhao, "Equivalence, identity, and unitarity checking in black-box testing of quantum programs," *J. Syst. Softw.*, vol. 211, 2024, doi: 10.1016/j.jss.2024.112000.
[27] A. F. Alshammari, "Implementation of Model Evaluation using Confusion Matrix in Python," vol. 186, no. 50, pp. 42–48, 2024, doi: 10.5120/ijca2024924236.
[28] J. Erbani, P. É. Portier, E. Egyed-Zsigmond, and D. Nurbakova, "Confusion Matrices: A Unified Theory," *IEEE Access*, vol. 12, no. December, 2024, doi: 10.1109/ACCESS.2024.3507199.
[29] J. Li, H. Sun, and J. Li, "Beyond confusion matrix: learning from multiple annotators with awareness of instance features," *Mach. Learn.*, vol. 112, no. 3, pp. 1053–1075, 2023, doi: 10.1007/s10994-022-06211-x.
[30] A. Arias-Duart, E. Mariotti, D. Garcia-Gasulla, and J. M. Alonso-Moral, "A Confusion Matrix for Evaluating Feature Attribution Methods," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2023-June, no. Cv, pp. 3709–3714, 2023, doi: 10.1109/CVPRW59228.2023.00380.

Appendix

Table of Data Partitioning

Texture Features						
ID	Contrast	Homogeneity	Energy	Entropy	Category	K-Partition
1	1316830	0.000276917	0.0529217	4.43304	Lontara non-woven	1
15	2468880	0.0000564361	0.00365155	6.50231	Lontara non-woven	1
16	1809050	0.000245968	0.0171035	5.56053	Lontara non-woven	2
30	2790680	0.000088834	0.0213049	5.09966	Lontara non-woven	2
31	3019340	0.000112957	0.0413711	4.36982	Lontara non-woven	3
45	1799870	0.0247989	0.0165974	4.75591	Lontara non-woven	3
46	1840820	0.0383048	0.0172184	4.98047	Lontara non-woven	4
60	2996480	0.0397392	0.0116764	5.13391	Lontara non-woven	4
61	2801630	0.0241113	0.00820225	5.39378	Lontara non-woven	5
75	1351160	0.00470926	0.00897832	5.54425	Lontara non-woven	5
Texture Features						
ID	Contrast	Homogeneity	Energy	Entropy	Category	K-Partition
76	1310050	0.00000181182	0.0244166	4.4835	Lontara woven	1
90	380525	0.00000397763	0.0445554	3.77646	Lontara woven	1
91	753650	0.00000389927	0.0364668	4.10494	Lontara woven	2
105	309840	0.00000789033	0.0146035	4.50596	Lontara woven	2
106	524709	0.0000116323	0.00508218	5.76571	Lontara woven	3
120	1116500	0.000057132	0.00552089	5.96666	Lontara woven	3
121	1672580	0.0000115859	0.00499331	5.70271	Lontara woven	4
135	1147780	0.0000799141	0.00679458	5.88886	Lontara woven	4
136	874224	0.0000937287	0.0111929	5.42886	Lontara woven	5
150	612952	0.0000617467	0.00740085	5.61632	Lontara woven	5