International Journal of Sustainable Applied Sciences (IJSAS)
Vol.2, No.6, 2024: 571-580

Comparative Analysis of Support Vector Machine and Random Forest Algorithms in Indonesian Batik Classification

Rainal Zulian Oktavianto[1*], Muhamad Fatchan[2], Wahyu Hadikristanto[3]
Universitas Pelita Bangsa
Corresponding Author : Rainal Zulian Oktavianto rainalzulian@gmail.com

ARTICLEINFO

Keywords: Indonesian Batik, classification, Support Vector Machine (SVM), Random Forest

Received : 03, May
Revised : 16, May
Accepted: 22, June

ABSTRACT

This study compares the performance of Support Vector Machine (SVM) and Random Forest (RF) algorithms in Indonesian batik image classification. Data collected from four batik categories: Pattern Batik Insang, Pattern Batik, Patterns Batik Dump, and Pattern Megamendung. Image feature extracted using Histogram of Oriented Gradients (HOG). SVM models with linear and RF kernels with 100 decision trees are trained and tested on this dataset. The evaluation results showed that the SVM has an accuracy of 88%, with precision and recall balanced between classes, while RF has an accuracy of 86%, with some classes showing excellent performance. SVM is superior in overall accuracy, but RF offers better interpretability and ease of setting parameters. The conclusions of this study suggest that both algorithms are able to effectively classify bacterial images, but the selection of the algority depends on the specific needs of the application. Further adjustment of parameters and additional preprocessing techniques are recommended to improve model performance. This research provides a strong foundation for further development in the classification of batic images using machine learning.

DOI: https://
ISSN-E: 3025-5597
https://journal.multitechpublisher.com/index.php/ijsas

INTRODUCTION

Indonesia is a country rich in natural resources, culture, and tourism. One of the most famous cultural heritages in Indonesia is Indonesian Batik. Batik is a craft with high artistic value that has long been an integral part of Indonesian culture, particularly in Java (Denny et al., 2019). Batik motifs are created from a cultural art that possesses visual beauty and contains philosophical meanings. Over time, batik motifs have evolved according to the period, place, and accompanying events, as well as the evolving needs of society (Zaman et al., 2021).

Currently, many Indonesians are still unaware of the various names of batik motifs, which are an intellectual heritage recognized by UNESCO (United Nations Educational, Scientific, and Cultural Organization) on October 2, 2009, as one of the world's cultural heritages originating from Indonesia (Wiryadinata et al., 2019). The abundance of batik motifs makes it difficult to distinguish between each motif. Many modern Indonesians cannot identify the types of batik they are wearing, thus classification is needed to address this issue. Classification is a process of image analysis that produces a model to represent the classes present in the image (Deni Akbar & Arie Wijaya, n.d.). Many researchers have conducted studies related to batik classification, such as the research by Hendry Fonda, Yuda Irawan, and Anita Febriani (Fonda et al., 2020). This study applied the CNN method, where the results showed an accuracy of 65% with a loss value of 2.5% and 2.1%.

Another study was conducted by Rizki Mawan (Mawan, n.d.). This research applied the CNN method, with testing results showing a data loss value as high as 1.4 for the test data and 1.1 for the training data. The highest accuracy for the test data was 0.8, and for the training data, it was 0.7 (on a scale from 0.0 to 1.0). When using a combination of CNN and grayscale, the accuracy improved to 70%.

Based on related studies, this research will implement the classification methods of Support Vector Machine (SVM) and Random Forest. The determination or selection of methods is based on several stages, including feature extraction and classification method comparisons, to apply the appropriate method for classifying batik types. Additionally, several related studies have explained that applying preprocessing stages using these methods yields better results.

LITERATURE REVIEW

Batik

The term 'Batik' originates from the Javanese words 'amba' and 'nitik', which mean to repeatedly dot or point on a wide cloth. As a traditional fabric, batik is made using the manual wax-resist dyeing technique with hands. Not only in Indonesia, but the interest in batik has also spread internationally. Many foreign tourists visit Indonesia and often consider batik as one of the must-have souvenirs to bring back to their home countries (Bariyah & Arif Rasyidi, n.d.).

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a technique for making predictions, both in classification and regression cases. SVM has the basic principle of a linear classifier, which means it can separate classes linearly. However, SVM has been developed to work on non-linear problems by introducing the concept of kernels in high-dimensional feature space (Yap, n.d.). The SVM algorithm works by finding an optimal hyperplane by maximizing the distance between classes. The hyperplane in SVM is a feature that can be used to separate one class from another (I Putu Agus Aryawan, n.d.).

Random Forest

The random forest method is a technique that can improve accuracy results because it generates child nodes for each node randomly (Yohannes et al., 2020). Random Forest is an extension of the Classification and Regression Tree (CART) method by implementing Bootstrap Aggregating (Bagging) and Random Feature Selection. The Random Forest method has several advantages, including producing good classification results, generating lower errors, and efficiently handling training data with a very large number of data (Susetyoko et al., n.d.).

METHODOLOGY

The stages of this research implementation can be seen in the following diagram:

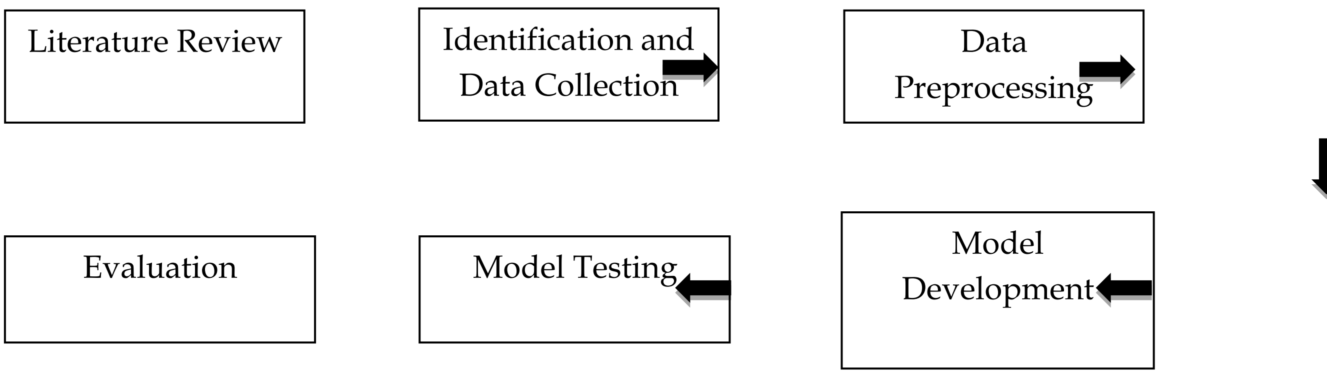

Figure 1. Research Flow

The stages in this research begin with literature review, identification and data collection, data preprocessing, coding to build models using SVM and Random Forest methods. Next, model evaluation and classification model selection are performed.

Data Collection

The data used in this research consists of batik images with several types of patterns, namely Batik Corak Insang, Batik Dayak, Batik Ikat Dip, and Batik Megamendung. These images are divided into three sets: training set, testing set, and validation set. Each set contains images that have been classified based on their batik pattern types.

These batik images were obtained by downloading them from the Kaggle website. The dataset comprises 640 images in jpg format, with varying resolutions, representing 4 different classes or motifs of batik. Below are samples of Indonesian batik motifs used in this study:

Figure 2. Batik motifs used in the study

The literature review includes the study of various image classification techniques, with a focus on Support Vector Machine and Random Forest algorithms. Relevant literature on data preprocessing, feature extraction, and model evaluation techniques are also reviewed to provide a solid foundation for this research.

Data Preprocessing

Data preprocessing involves several important steps, including :

Rescale i.e. All images were rescaled to a uniform size using ImageDataGenerator from Keras with a rescale factor of 1./255. Feature Extraction i.e. Image features were extracted using Histogram of Oriented

Gradients (HOG) from OpenCV. The image was resized to 224x24 pixels before the HOG features were calculated.

Model Development

To develop SVM and Random Forest models, the first step is to use pre-processed training data. This data will be the basis for training both types of models. This process will involve building SVM and Random Forest models using the Python programming language and relevant libraries, such as scikit-learn. The SVM model with a linear kernel is trained using HOG features extracted from batik images and the Random Forest model is trained using 100 decision trees to predict the classification of batik images.

In model development, appropriate parameters will be set and adjusted to achieve optimal performance. After training the model using the training data, the next step is to perform evaluation to validate the reliability and quality of the developed model. By paying attention to these stages in model development, it is expected that the results can make a significant contribution in solving the classification problem at hand.

Model Testing

In the model testing stage, researchers focus on using Train Accuracy and Validation Accuracy as evaluation metrics. We divided the dataset into training data and validation data to evaluate the performance of SVM and Random Forest models in the classification of Indonesian batik. Train Accuracy is used to measure how well the model can learn patterns from the training data, while Validation Accuracy is used to evaluate the model's performance on data that has never been seen before. Using these two metrics, researchers can assess how well the model is able to generalise patterns from training data to validation data. This testing procedure provides an important understanding of the reliability and generalisability of the model in classifying Indonesian batik images.

Evaluation

Evaluation is a crucial stage in this research, where we assess the performance of SVM and Random Forest models in classifying Indonesian batik. We use important evaluation parameters such as Training Accuracy and Validation Accuracy to evaluate both. Training Accuracy is used to measure the extent to which the model can understand patterns from training data, while Validation Accuracy is used to evaluate the model's performance on data that has never been seen before. Researchers also considered the interpretative results of each model to understand their ability to correctly classify batik images. This evaluation provides a deep understanding of the reliability and general capabilities of the models in the context of Indonesian batik classification.

Model evaluation is performed using the following metrics i.e. Precision Measures the accuracy of positive predictions. Recall measures the ability of the model to find all positive cases. F1-Score harmonized average of precision and recall. Accuracy the proportion of correct predictions out of all predictions. Confusion Matrix displays the distribution of correct and incorrect predictions among the classes.

RESULTS

The experiment was conducted using 640 batik images consisting of 4 batik classes at various resolutions. Then, in the preprocessing stage, these batik images were resized to a resolution of 224 x 224. The results of identification and displaying batik images to detect the motif or type of batik being tested are as follows :

Figure 3. Displaying the results of batik identification

In this study, researchers conducted a comparative analysis between the Support Vector Machine (SVM) and Random Forest algorithms in the classification of Indonesian batik. The evaluation process is carried out through several stages to ensure the accuracy of the model in classifying batik patterns.

The evaluation results show that the SVM model has an overall accuracy of 88%. Precision and recall vary among different classes, with a fairly high F1-

score, indicating good performance in batik image classification. The following is a breakdown of the evaluation results of batik classification using the SVM model :

Table 1. Evaluation of Training Results With SVM Model

	Precision	Recall	F1-score	Support
Batik Corak Insang	0.92	0.80	0.86	30
Batik Dayak	0.83	0.97	0.89	30
Batik Ikat Celup	0.82	0.90	0.86	30
Batik Megamendung	0.96	0.83	0.89	30
Accuracy			0.88	120
Macro avg	0.88	0.88	0.87	120
Weighted avg	0.88	0.88	0.87	120

The evaluation results show that the Random Forest model has an overall accuracy of 86%. Although slightly lower than SVM, the precision and recall also varied among the different classes, with some classes performing very well, although there was an imbalance between precision and recall in some classes. The following is a breakdown of the evaluation results of batik classification using the Random Forest model :

Table 2. Evaluation of Training Results with Random Forest Model

	Precision	Recall	F1-score	Support
Batik Corak Insang	0.92	0.73	0.81	30
Batik Dayak	0.69	0.97	0.81	30
Batik Ikat Celup	0.93	0.87	0.90	30
Batik Megamendung	1.00	0.87	0.93	30
Accuracy			0.86	120
Macro avg	0.88	0.86	0.86	120
Weighted avg	0.88	0.86	0.86	120

DISCUSSION

This research focuses on the comparison of two machine learning algorithms, namely Support Vector Machine (SVM) and Random Forest (RF), in the classification of Indonesian batik images. SVM with a linear kernel showed an overall accuracy of 88%. The SVM algorithm is known to be robust in handling high-dimensional data and works well on datasets that have extracted features, such as in this case where features are extracted using Histogram of Oriented Gradients (HOG). SVM showed consistent performance with precision and recall varying between different classes, but generally producing a good

F1-score. This shows that SVM is able to maintain a balance between precision and recall, which is important in multi-class classifications such as this.

On the other hand, Random Forest showed an overall accuracy of 86%. This algorithm works by combining multiple decision trees to improve classification accuracy and reduce the risk of overfitting. Although the accuracy of RF is slightly lower than SVM, it shows some advantages in terms of interpretability and ability to handle variable features. Random Forest tends to be easier to set up and requires less parameter tuning than SVM. However, the imbalance between precision and recall in some classes suggests that RF may require further customization, such as setting the number of trees or tree depth to improve performance in certain classes.

Both models show that they are able to classify batik images with a high degree of accuracy. SVM is slightly superior in terms of overall accuracy, while Random Forest gives very competitive results with advantages in interpretability and robustness against overfitting. The variation in precision and recall shows that both models have strengths and weaknesses in recognizing certain patterns in batik images, which may be related to pattern complexity and similarity between classes. Further adjustments to the model parameters and additional optimization techniques can improve the overall performance of the model.

CONCLUSIONS AND RECOMMENDATIONS

This research concludes that Support Vector Machine (SVM) with linear kernel shows better performance with 88% accuracy compared to Random Forest (RF) which has 86% accuracy. SVM shows a good balance between precision and recall, making it a strong choice for batik image classification. However, RF also shows competitive performance with advantages in interpretability and ease of parameter setting. Both models are able to classify batik images effectively, and the choice between the two depends on the specific needs of the application, such as prioritizing accuracy or interpretability and ease of parameter setting.

To improve the performance of the models, further parameter tuning, increasing the number of images in the dataset, and applying additional preprocessing techniques such as data augmentation are recommended. In addition, considering other algorithms such as Convolutional Neural Networks (CNN) which have been proven to be effective in image classification, may provide better results especially with larger datasets. This research provides a solid foundation for further development in batik image classification using machine learning.

FURTHER STUDY

Every study has limitations that need to be recognised and evaluated for future improvements. In this study, one of the main limitations is the limited size of the dataset, which might affect the ability of the model to generalise to new data. In addition, the variation in batik motifs used in the model training may not cover the entire diversity of Indonesian batik motifs. Therefore, further research is recommended to use larger and more diverse datasets to improve the representation and generalisation of the model. In addition, an in-depth study of data pre-processing techniques such as data augmentation and normalisation can be conducted to improve model performance. Future research can also explore the combination of algorithms or the use of deep learning methods to compare the results and find a more optimal approach in batik classification. By overcoming these limitations, it is hoped that future research results can make a more significant contribution to the development of more effective and accurate batik classification technology.

ACKNOWLEDGMENT

The author would like to express gratitude to all parties who have supported the research process and contributed their thoughts and time. The author also extends thanks to previous researchers.

REFERENCES

Bariyah, T., & Arif Rasyidi, M. (n.d.). Convolutional Neural Network Untuk Metode Klasifikasi Multi-Label Pada Motif Batik Convolutional Neural Network for Multi-Label Batik Pattern Classification Method. In *Februari* (Vol. 20, Issue 1).

Deni Akbar, M., & Arie Wijaya, Y. (n.d.). *KLASIFIKASI MOTIF BATIK JAWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS (KNN)*. https://ejournal.stmikgici.ac.id/

Denny, H., Samuel, K., Didi, K., & Anggun, D. (2019). Klasifikasi Motif Citra Batik Yogyakarta Menggunakan Metode Adaptive Neuro Fuzzy Inference System. *Jurnal Ilmiah Setrum Article In Press*, *8*(2), 229–237.

Fonda, H., Irawan, Y., Febriani, A., Informatika, S., & Pekanbaru, H. T. (2020). KLASIFIKASI BATIK RIAU DENGAN MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORKS (CNN) 1 2 3 Email : 1 2 3. In *JIK* (Vol. 9, Issue 1). http://jik.htp.ac.id

I Putu Agus Aryawan, I. N. P. K. Q. Fredlina. (n.d.). *ANALISIS PERBANDINGAN ALGORITMA CNN DAN SVM PADA KLASIFIKASI EKSPRESI WAJAH*.

Mawan, R. (n.d.). *Klasifikasi motif batik menggunakan convolutional neural network*. https://doi.org/10.36802/jnanaloka

Susetyoko, R., Yuwono, W., Purwantini, E., & Ramadijanti, N. (n.d.). *Perbandingan Metode Random Forest, Regresi Logistik, Naïve Bayes, dan Multilayer Perceptron Pada Klasifikasi Uang Kuliah Tunggal (UKT)*. *7*(1).

Wiryadinata, R., Rofiki Adli, M., Fahrizal, R., Alfanz, R., Teknik Elektro, J., Kunci-Ekstraksi ciri, K., Banten, B., & Vector Machine, S. (2019). *Klasifikasi 12 Motif Batik Banten Menggunakan Support Vector Machine* (Vol. 13, Issue 1). https://jurnaleeccis.ub.ac.id/

Yap, A. J. (n.d.). *KLASIFIKASI KAIN TENUN DI PULAU TIMOR MENGGUNAKAN METODE MULTI SUPPORT VECTOR MACHINE (SVM)*.

Yohannes, Y., Devella, S., & Pandrean, A. H. (2020). Penerapan Speeded-Up Robust Feature pada Random Forest Untuk Klasifikasi Motif Songket Palembang. *Jurnal Teknik Informatika Dan Sistem Informasi*, *5*(3). https://doi.org/10.28932/jutisi.v5i3.1978

Zaman, B., Rifa'i, A., & Hanif, M. B. (2021). Komparasi Metode Klasifikasi Batik Menggunakan Neural Network Dan K-Nearest Neighbor Berbasis Ekstraksi Fitur Tekstur. *Journal of Information Systems and Informatics*, *3*(4). http://journal-isi.org/index.php/isi