Identitas Paper

Paper ID / DOI	19851 / **DOI:** https://doi.org/10.18280/ria.370214
Judul Paper	Comparison of Machine Learning Performance Using Naive Bayes and Random Forest Methods to Classify Batik Fabric Patterns
Authors	Abdul Fadlil, Imam Riadi, Izzan Julda D.E Purwadi Putra*
Jurnal	Revue d'Intelligence Artificielle (RIA)
Penerbit	Lavoisier
Negara	France
Quartile dan SJR	Scopus Q3, SJR 0.3 https://www.scimagojr.com/journalsearch.php?q=24847&tip=sid&clean=0
Terbit	Vol 37 No 2 April 2023 pp 379-385
Link Paper	https://www.iieta.org/journals/ria/paper/10.18280/ria.370214

Revue d'Intelligence Artificielle

Q3 Artificial Intelligence

best quartile

SJR 2022
0.3

powered by scimagojr.com

Proses Pengelolaan Paper

Tahap	Tanggal	Proses/Aktivitas
1	26 February 2023	Submission
2	18 April 2023	Revision required
3	19 April 2023	Submit the revision
4	21 April 2023	Accepted
5	27 April 2023 – 8 Mei 2023	Editing process
6	27 Mei 2023	Published

	Quartile	SJR	Nov-22 I	II	III	IV	Dec-22 I	II	III	IV	Jan-23 I	II	III	IV	Feb-23 I	II	III	IV	Mar-23 I	II	III	IV	Apr-23 I	II	III	IV
RIA-apr	Q3	0.30					4	10						30	2	7	24	28								30

Tahap 1
Submission, 26 February 2023

Revue d'Intelligence Artificielle ▾   ⊕ English   👁 View Site   👤 izzanjulda

Submission Library   View Metadata

Comparison of Machine Learning Performance using Naive Bayes and Random Forest Methods to Classify Batik Fabric Patterns
Abdul Fadlil, Imam Riadi, Izzan Julda D.E Purwadi Putra

Tasks 0

Submissions

| Submission | Review | Copyediting | Production |

Submission Files   🔍 Search

▸ 📄 39051-1   izzanjulda, Author, Comparison of Machine Learning Performance using Naive Bayes and   Article Text
                 Random Forest Methods to Classify Batik Fabric Patterns .docx

▸ 📄 39052-1   izzanjulda, Author, Turnitin Similarity Check (2).pdf   Other

                                                                        Download All Files

Pre-Review Discussions   **Add discussion**

Name	From	Last Reply	Replies	Closed
▸ =	izzanjulda	-	0	☐

Tahap 2
Revision required, 18 April 2023

☰ M Gmail   🔍 Telusuri dalam email   ● Aktif ⌄ ? ⚙ ⠿   UNIVERSITAS AHMAD DAHLAN

✏ Tulis   23 dari 283 < >

Mail

Chat   📥 Kotak Masuk 106   **RIA 19851** Eksternal Kotak Masuk ×

          ☆ Berbintang   e   editor.ria iieta.org <editor.ria@iieta.org>   Sel, 18 Apr, 13.03 ☆ ↩ ⋮
Spaces   🕐 Ditunda   kepada saya, imam.riadi@is.uad.ac.id, fadlil@mti.uad.ac.id ▾
Meet   ▷ Terkirim
          📄 Draf   3   🗛 Inggris ▾ > Indonesia ▾ Terjemahkan pesan   Nonaktifkan untuk: Inggris ×
          ⌄ Selengkapnya
                                     Dear author,
Label   +
                                     We have reached a decision regarding your submission to Revue d'Intelligence Artificielle, " Comparison of Machine Learning Performance using Naive
                                     Bayes and Random Forest Methods to Classify Batik Fabric Patterns".

                                     Our decision is: Revisions Required

                                        1. Revise the current paper according to reviewers' comments. Highlight any change or track changes you make. **Response to reviewers is also
                                           required.**
                                        2. Typeset your final paper according to the attached template.

                                     Please return your revised manuscript and the response to reviewers to this e-mail before **22-24 April 2023**. Thanks for your cooperation.

                                     Revue d'Intelligence Artificielle
                                     http://www.iieta.org/Journals/RIA

Review Comments

1. The abstract of this article has made a lot of descriptions about the research background. It's recommended to place the previous part directly in the introduction. The abstract should focus on the innovation of this research method.

2. The introduction part cites some relevant literature, but there are only results or conclusive statements for the literature, and there is no comparison with the relevant literature, so as to highlight the research method of this article, that's, considering the last paragraph of the introduction part (indicating the research focus of this article), the correlation between the previous related status and the literature is not strong.

3. Figure 1 is a flow chart, and the author should insert it directly by editing the flow chart, instead of the screenshots presented in the article. The font format is obviously different from the text, and there are certain problems in clarity.

4. In the feature extraction of machine learning, the author only considers feature extraction from different angles. Does the author consider color, line, pattern newness, etc. for feature extraction and recognition?

5. Considering the sample size of machine learning, 600 samples are mainly aimed at the learning and recognition of three theme patterns, and the number of samples is relatively small for machine learning.

6. Are the chart curves in Figure 3 and Figure 4 extracted through screenshots? From the perspective of clarity, there are similar problems with Figure 1, and the revision should be made.

7. The table names of the tables should usually be placed above the tables. There are some problems with the header placement of Table 3, Table 4, and Table 5, which need to be modified.

8. Considering the typesetting, the position of the figures should be as close as possible to the position of the text description in the main text, and from the relatively large amount of data in Table 3, Table 4, and Table 5, there are not many textual supplementary explanations in the text, so please strengthen the analysis from the comparison of the result data.
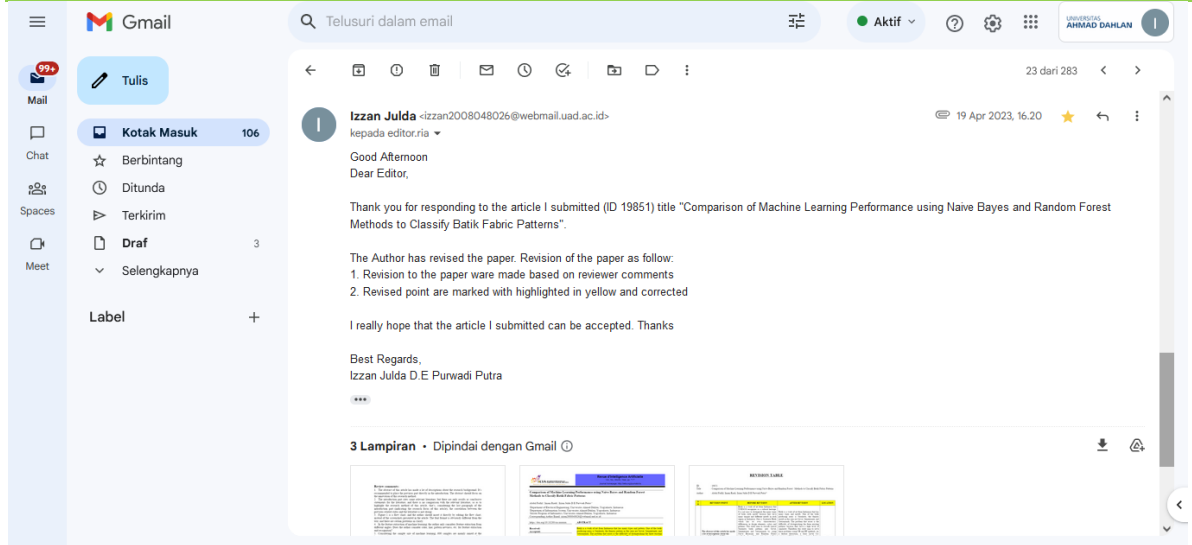
NO	REVISION POINT	BEFORE REVISION	AFTER REVISION	LOCATION
1	**The abstract of this article has made a lot of descriptions about the research background. It's recommended to place the previous part directly in the introduction. The abstract should focus on the innovation of this research method.**	Batik is a work of art from Indonesia that UNESCO recognizes as a cultural heritage. It is not easy to distinguish the many types of batik cloth motifs because they have many unique and different motifs in each region in Indonesia. One is Surakarta Batik which has its own characteristics: differences in design elements, colors and patterns. This study aims to classify typical Surakarta batik patterns into Sawat, Sementrante and Satriomanah by using Naive Bayesian and Random Forest methods. As a feature extraction, a Gray Level Co-occurrence Matrix is used as a texture feature extraction. The research phase includes methods for dataset collection, preprocessing, feature extraction, and classification. These two methods, RF and NB, can be used as methods for batik fabric classification. The most accurate result obtained by the RF method was 97.91% accurate in dataset A, while the NB method was 96.66% accurate on the same dataset. According to	Batik is a work of art from Indonesia that has many types and motifs. One of the batik producing areas is Surakarta, the famous motifs in this area are Sawat, Sementrante, and Satriomanah. The problem that arises is the difficulty of distinguishing the three existing patterns because they have a high level of similarity. Therefore, this study aims to solve these problems using NB and RF methods. As a feature extraction, a Gray Level Co-occurrence Matrix is used as a texture feature extraction. The research phase includes methods for dataset collection, preprocessing, feature extraction, and classification. These two methods, RF and NB, can be used as methods for batik fabric classification. The most accurate result obtained by the RF method was 97.91% accurate in dataset A, while the NB method was 96.66% accurate on the same dataset. According to the research results, it is found that the RF method outperforms the NB method in classifying the types of batik patterns.	**Page 1 (Abstract)**

		the research results, it is found that the RF method outperforms the NB method in classifying the types of batik patterns		
2	**The introduction part cites some relevant literature, but there are only results or conclusive statements for the literature, and there is no comparison with the relevant literature, so as to highlight the research method of this article, that's, considering the last paragraph of the introduction part (indicating the research focus of this article), the correlation between the previous related status and the literature is not strong.**	This study uses the GLCM for feature extraction and two classification techniques, namely NB and RF. This study compared the two methods NB and RF, for classifying sawat, sementrante, and satriomanah types of batik cloth motifs based on similar batik fabric textures.	The fundamental difference from the previous study is that it uses two classification techniques, namely NB and RF, and uses 6 feature extraction features in terms of texture and viewed from 4 angles. Therefore, this study used texture feature extraction (GLCM) and compared two classification techniques using WEKA tools. The dataset obtained uses data augmentation techniques to enrich the image sample. The results of this study contain the accuracy value of each classification technique. The implementation of both classification techniques achieved the highest accuracy of 97.22% in the RF method and 96.66% in the NB method. This paper contains information on the proposed approach to classifying batik motifs including introduction, methodology, results and discussion, and conclusion	**Page 1 (Introduction)**
3	**Figure 1 is a flow chart, and the author should insert it directly by editing the flow chart, instead of the screenshots presented in the article. The font format is obviously different from the text, and there are certain problems in clarity.**	Figure 1. Block diagram of research flow	Correction in Image Clarity	**Page 2**
4	**In the feature extraction of machine learning, the author only considers feature extraction from different angles. Does the author consider color, line, pattern newness, etc. for feature**	Not yet explained	To obtain feature data, this study used texture feature extraction using GLCM	**Page 2**

	extraction and recognition?			
5	Considering the sample size of machine learning, 600 samples are mainly aimed at the learning and recognition of three theme patterns, and the number of samples is relatively small for machine learning.	Not yet explained	600 datasets were used due to limitations on collection and processing, as well as to reduced overfitting	**Page 2**
6	Are the chart curves in Figure 3 and Figure 4 extracted through screenshots? From the perspective of clarity, there are similar problems with Figure 1, and the revision should be made	Figure 3. RF classifier Figure 4. NB classifier	Correction in Image Clarity	**Page 3**
7	The table names of the tables should usually be placed above the tables. There are some problems with the header placement of Table 3, Table 4, and Table 5, which need to be modified.	Not exactly according to the template	Correction according to the template	**Page 4, 5, 6**
8	Considering the typesetting, the position of the figures should be as close as possible to the position of the text description in the main text, and from the relatively large amount of data in Table 3, Table 4, and Table 5, there are not many textual supplementary explanations in the text, so please strengthen the analysis from the comparison of the result data.	1. Not exactly according to the template 2. lack of explanation	1. Correction according to the template 2. Based on Table 5 accuracy, recall, and precision are classification evaluation metrics used to assess the performance of machine learning models.	**Page 4, 5, 6**

RIA 19851 Eksternal Kotak Masuk ×

editor.ria iieta.org
kepada saya, imam.riadi@is.uad.ac.id, fadlil@mti.uad.ac.id ▾

Jum, 21 Apr, 18.03

Inggris ▾ > Indonesia ▾ Terjemahkan pesan   Nonaktifkan untuk: Inggris ×

Dear author,

We have reached a decision regarding your submission to *Revue d'Intelligence Artificielle*,

Manuscript Title: Comparison of Machine Learning Performance using Naive Bayes and Random Forest Methods to Classify Batik Fabric Patterns

Manuscript ID: 19851

Our decision is to: Accept Submission

Before we proceed with the publication of your article, please complete the arrange payment of your article processing charge (US $400) in 15 days by the following ways:

USD Remittance Path
BENEFICIARY NAME: IIETA-RSMC Beijing LTD
ACCOUNT NUMBER: 0200296409116032240
BENEFICIARY ADDRESS: Room 302, Building 3, Jinmao Plaza, Fengtai District, Beijing, PRC
BENEFICIARY BANKER'S NAME: Industrial and Commercial Bank of China Beijing Municipal Branch, Beijing, PRC
SWIFT CODE: ICBKCNBJBJM

IIETA International Information and
Engineering Technology Association
Advancing the World of Information and Engineering

Acceptance Letter

Revue d'Intelligence Artificielle

March 30, 2023

Izzan Julda D.E Purwadi Putra

Master Program of Informatics,
Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Dear Abdul Fadlil1, Imam Riadi2, Izzan Julda D.E Purwadi Putra

MS: Comparison of Machine Learning Performance using Naive Bayes and Random Forest Methods to Classify Batik Fabric Patterns

I am pleased to inform you that as per the recommendation of the editorial board, your above-mentioned manuscript has been accepted for publication in Revue d'Intelligence Artificielle (ISSN 0992-499X).

Please note the following points, and ensure compliance:

1) Page Proofs of your paper will also be sent to you for minor corrections and approval.

2) Provide us with your E-mail ID for future correspondence.

3) Provide complete postal address with correct postal-code.

4) The corresponding author would receive a PDF of the published paper.

5) If your paper is found with plagiarism content after the release of Acceptance Letter, it will not be published in the Journal.
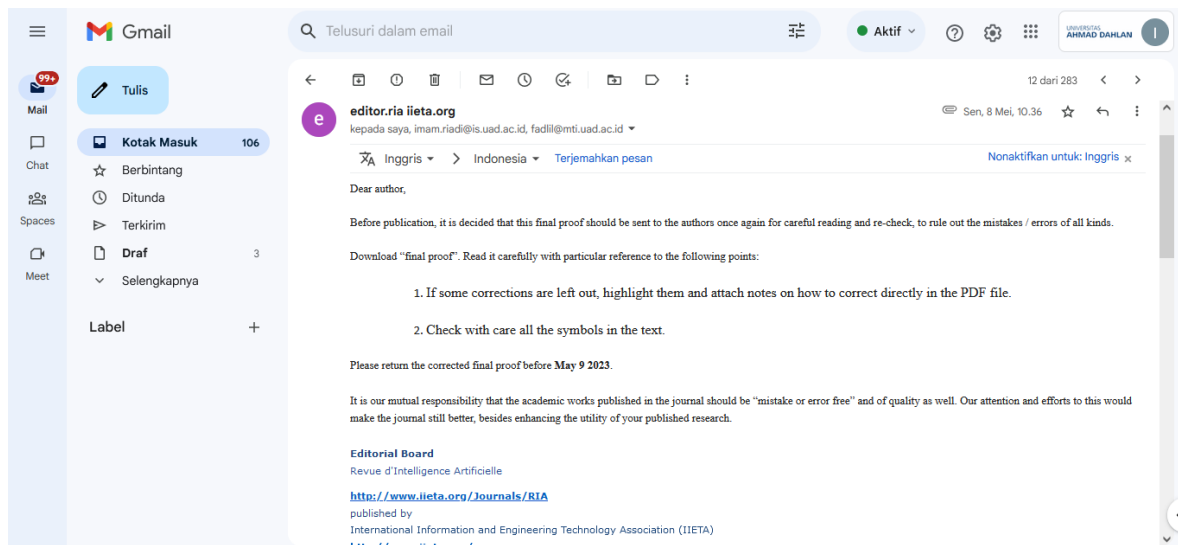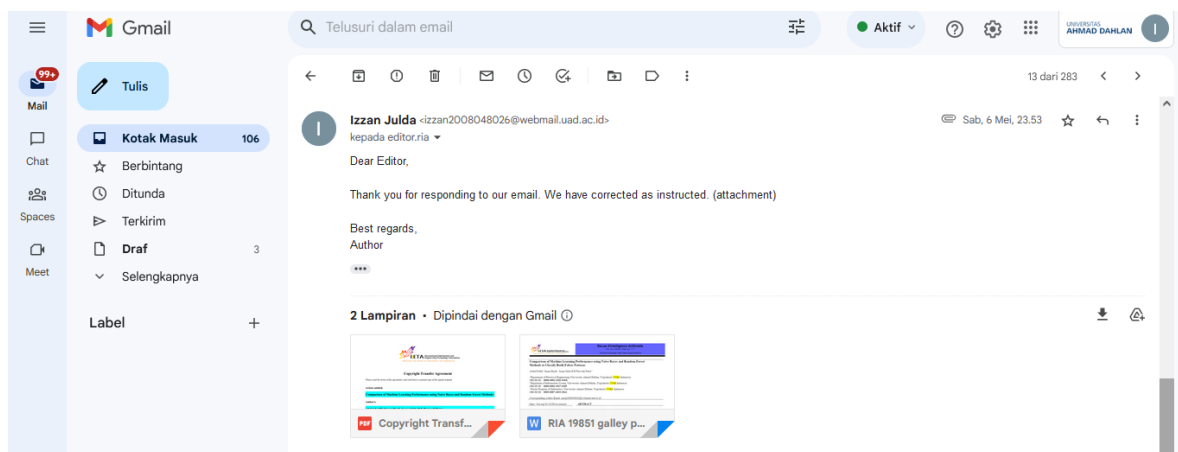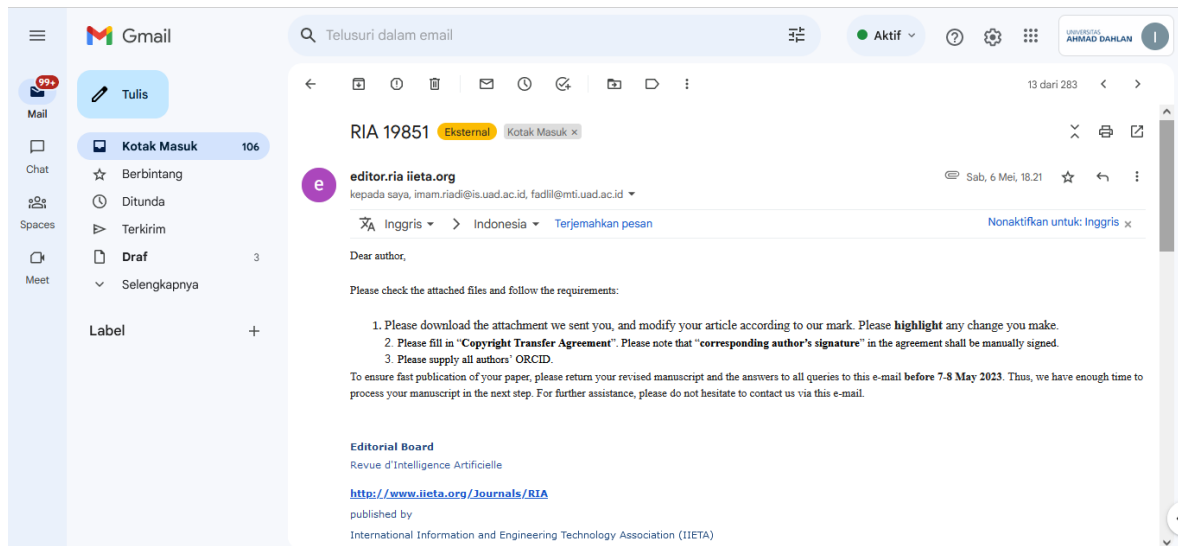
It is recommended that you cite this and other published works from International Information and Engineering Technology Association (www.iieta.org) in your papers to be published in other journals.

Yours sincerely,

Date: March 30, 2023
Place: Edmonton, Canada

RIA 19851 [Eksternal] Kotak Masuk ×

editor.ria iieta.org
kepada saya, imam.riadi@is.uad.ac.id, fadlil@mti.uad.ac.id ▾

文A Inggris ▾ 〉 Indonesia ▾ Terjemahkan pesan   Nonaktifkan untuk: Inggris ×

Dear author,

Please check the attached files and follow the requirements:

1. Please download the attachment we sent you, and modify your article according to our mark. Please **highlight** any change you make.
2. Please fill in "**Copyright Transfer Agreement**". Please note that "**corresponding author's signature**" in the agreement shall be manually signed.
3. Please supply all authors' ORCID.

To ensure fast publication of your paper, please return your revised manuscript and the answers to all queries to this e-mail **before 7-8 May 2023**. Thus, we have enough time to process your manuscript in the next step. For further assistance, please do not hesitate to contact us via this e-mail.

Editorial Board
Revue d'Intelligence Artificielle

http://www.iieta.org/Journals/RIA
published by
International Information and Engineering Technology Association (IIETA)

Izzan Julda <izzan2008048026@webmail.uad.ac.id>
kepada editor.ria ▾

Dear Editor,

Thank you for responding to our email. We have corrected as instructed. (attachment)

Best regards,
Author

•••

2 Lampiran · Dipindai dengan Gmail ⓘ

Copyright Transf...   RIA 19851 galley p...

editor.ria iieta.org
kepada saya, imam.riadi@is.uad.ac.id, fadlil@mti.uad.ac.id ▾

文A Inggris ▾ 〉 Indonesia ▾ Terjemahkan pesan   Nonaktifkan untuk: Inggris ×

Dear author,

Before publication, it is decided that this final proof should be sent to the authors once again for careful reading and re-check, to rule out the mistakes / errors of all kinds.

Download "final proof". Read it carefully with particular reference to the following points:

1. If some corrections are left out, highlight them and attach notes on how to correct directly in the PDF file.

2. Check with care all the symbols in the text.

Please return the corrected final proof before **May 9 2023**.

It is our mutual responsibility that the academic works published in the journal should be "mistake or error free" and of quality as well. Our attention and efforts to this would make the journal still better, besides enhancing the utility of your published research.

Editorial Board
Revue d'Intelligence Artificielle

http://www.iieta.org/Journals/RIA
published by
International Information and Engineering Technology Association (IIETA)

Izzan Julda <izzan2008048026@webmail.uad.ac.id>
kepada editor.ria ▼

Sen, 8 Mei, 14.58

Dear Editor,

Thank you for sending the final manuscript. After we checked the article there were no errors and everything was in order.

Thank you.

Best regards
Author

Satu lampiran · Dipindai dengan Gmail ⓘ

RIA 19851 final.pdf

Tahap 6
Published, 30 April 2023

RIA

> About
> Aims and scope
> Editorial Board
> Instructions for Authors
> Article Processing Charge
> Publication Ethics
> Submission
> Current Issue
> Archive

Comparison of Machine Learning Performance Using Naive Bayes and Random Forest Methods to Classify Batik Fabric Patterns

Abdul Fadlil | Imam Riadi | Izzan Julda D.E Purwadi Putra*

Corresponding Author Email: izzan2008048026@webmail.uad.ac.id

Citation

DOWNLOAD PDF

OPEN ACCESS

Check for updates

Abstract:

Batik is a work of art from Indonesia that has many types and pattern. One of the batik producing areas is Surakarta, the famous pattern in this area are Sawat, Sementrante, and Satriomanah. The problem that arises is the difficulty of distinguishing the three existing pattern because they have a high level of similarity. Therefore, this research aims to solve these problems using NB and RF methods. As a feature extraction, a Gray Level Co-occurrence Matrix is used as a texture feature extraction. The research phase includes methods for dataset collection, preprocessing, feature extraction, and classification. These two methods, RF and NB, can be used as methods for batik fabric classification. The most accurate result obtained by the RF method was 97.91% accurate in dataset A, while the NB method was 96.66% accurate on the same dataset. According to the research results, it is found that the RF method outperforms the NB method in classifying the types of batik patterns.

Username
Password
Remember me
LOGIN

Journal Content

All
Search

Browse
• By Issue
• By Title
• By Author
• Other Journals

Information

• For Readers

Comparison of Machine Learning Performance using Naive Bayes and Random Forest Methods to Classify Batik Fabric Patterns

Abdul Fadlil[1], Imam Riadi[2], Izzan Julda D.E Purwadi Putra[3,*]

[1]Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2]Departmen of Information System, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[3]Master Program of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
Corresponding Author Email: izzan2008048026@webmail.uad.ac.id

ABSTRACT

Batik is a work of art from Indonesia that has many types and pattern. One of the batik producing areas is Surakarta, the famous pattern in this area are Sawat, Sementrante, and Satriomanah. The problem that arises is the difficulty of distinguishing the three existing pattern because they have a high level of similarity. Therefore, this research aims to solve these problems using NB and RF methods. As a feature extraction, a Gray Level Co-occurrence Matrix is used as a texture feature extraction. The research phase includes methods for dataset collection, preprocessing, feature extraction, and classification. These two methods, RF and NB, can be used as methods for batik fabric classification. The most accurate result obtained by the RF method was 97.91% accurate in dataset A, while the NB method was 96.66% accurate on the same dataset. According to the research results, it is found that the RF method outperforms the NB method in classifying the types of batik patterns.

1. INTRODUCTION

Indonesia has many kinds of culture and local arts, which are well known by many other countries worldwide, including Batik. UNESCO recognized that Batik was a cultural heritage from Indonesia on October 2, 2009 [1],[2]. Several studies on the categorization of batik images based on color, type, shape, and texture quality continue to help the batik image documentation process [3]. It is difficult to distinguish the many types of batik fabric pattern because they have many unique and different pattern in each region of Indonesia. One of them is Surakarta batik, which has its characteristics, namely differences in design elements, colors, and patterns [4]. One of the fields used in vision machines in image processing techniques is the identification and examination of data in the form of moving images or still images [5],[6]. Image processing techniques can facilitate better resolution than traditional techniques [7]. Another methods such as Decision Tree, K-Nearest Neighbor, Linear Regression, Back-Propagation, Support Vector Machines, Logistic Regression, Rocchio Method, Naive Bayes (NB), Random Forest, and Neural Network, can also be used as a classification technique using image processing data [8]. Batik image recognition aims to transform and analyze image data into information. Part of image recognition includes data acquisition, image recovery, segmentation, and recognition [9]. There have been several previous studies that have examined batik fabric pattern. One of them is about combining otsu and canny methods to identify the characteristics of Surakarta batik [4]. By applying the method, the data used is 100 image data and obtain an accuracy rate of 93%.

Surya, A.R et al. 2019 [10] researched identifying batik images using the ANN-Backpropagation method and Gray Level Co-Occurrence Matrix (GLCM) feature extraction,

managed to get the highest accuracy rate of 91.2% in Sogan batik. Research conducted by Zaw, T.H et al. 2019 [11] regarding brain tumor detection based on NB classification with an accuracy rate of 81.25%, successfully detected MRI images of brain tumors and predicted the parts of the brain affected by the tumor. Arsa, S and Susila, H [1] use VGG16 method in batik classification, succeeded identifying the type of batik with ±97% accuracy based on the Random Forest (RF) classifier. Bansal, M et al. 2019 [12] conducted image classification using VGG19 with caltech-101 image dataset using ImageNet imagery and several classification techniques, it obtained an accuracy rate of 92.05% in the NB classification and 93.73% in the RF classification.

The fundamental difference from the previous research is that it uses two classification techniques, namely NB and RF, and uses 6 feature extraction features in terms of texture and viewed from 4 angles.

Therefore, this study used texture feature extraction (GLCM) and compared two classification techniques using WEKA tools. The dataset obtained uses data augmentation techniques to enrich the image sample. The results contain the accuracy value of each classification technique. The implementation of both classification techniques achieved the highest accuracy of 97.22% in the RF method and 96.66% in the NB method. This paper contains information on the proposed approach to classifying batik pattern including introduction, methodology, results and discussion, and conclusion.

2. METHODOLOGY

Data collection, pre-processing, feature extraction, and evaluation are all included in the methods proposed by the study. Block diagram in this research is shown in Figure 1.
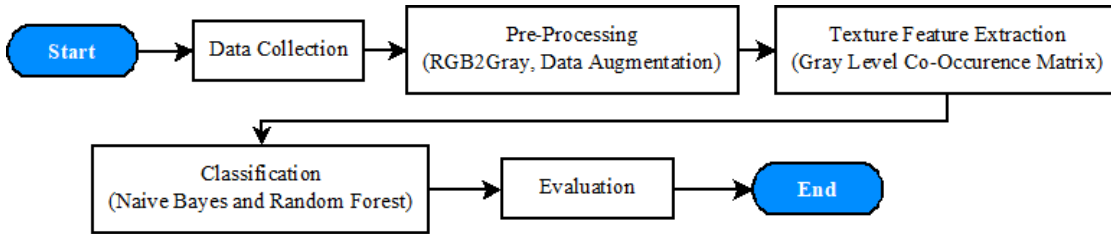
Figure 1. Block diagram of batik pattern classification system

Based on Figure 1, the image is modified after getting the batik image using pixel resizing, data augmentation, and conversion to grayscale from RGB. Pre-processing is followed by classification using NB and RF methods. To find out the success rate of this research, an evaluation was carried out.

2.1 Data Collection

This research, data for classifying Surakarta batik fabric pattern were obtained by scraping data from the internet, especially from websites that provide special types of Surakarta batik. The limitations of the type of batik used are the typical Surakarta batik fabric pattern, namely sawat, sementrante, and satriomanah. Examples of types of batik can be seen in Figure 2.

(a) (b)

(c)
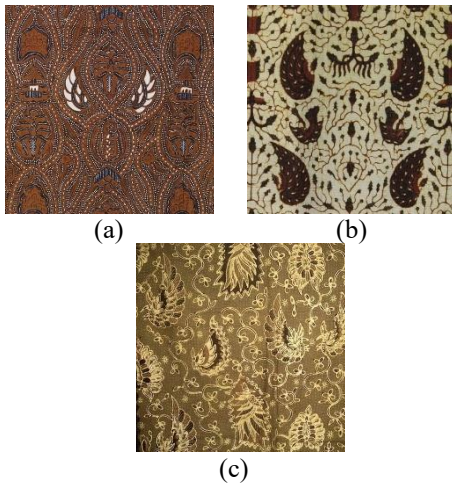
Figure 2. Batik Types (a) Sawat, (b) Semenrante, (c) Satriomanah

In Figure 2 is a type of batik fabric pattern. The pattern was chosen because it has a fairly high level of similarity. The number of datasets is 600 images. 600 datasets were used due to limitations on collection and processing, as well as to reduced overfitting. Divided into 200 pictures for the Sawat class, 200 pictures for the Semenrante class, and 200 pictures for the Satriomanah class. The batik fabric pattern dataset is divided into training and test data with compositions of 60%:40%, 70%:30%, 80%:20%.

2.2 Pre-processing

First, the image of the type of batik fabric is grouped based on the type of pattern. Then the batik image is converted from an RGB image to a grayscale image using equation (1).

$$Gray = (0.289R + 0.587G + 0.114B) \tag{1}$$

In addition, data augmentation was used in this study to diversify data variations. A process called "data augmentation" involves manipulating the rotation, brightness, cropping, and reversal of images [13],[14].

2.3 Feature Extraction

Feature extraction aims to identify pattern differences and differentiate class groupings during further classification [15]. Examples of feature extraction can be identified from color, shape, geometry, size, and texture. [16]. To obtain feature data, this study used texture feature extraction using GLCM. The technique of extracting values from GLCM is performed using the Python programming language. The possibility of a close relationship between two pixels at a certain distance (d) and the right angle (θ) is calculated using this technique to generate statistical values of the 2nd order [15], [17], [18],[19]. In this study, the GLCM feature extraction was seen from 4 angles, namely 0°, 45°, 90°, and 135° and the 6 GLCM features used were dissimilarity (L1), correlation (L2), homogeneity (L3), contrast (L4), ASM (L5), and energy (L6) shown in equations 2 to 7.

L1 measures the dissimilarity of textures in an image, and the resulting value will be large if the pattern is random and small if the pattern is uniform. Equation (2) shows the Equation of dissimilarity.

$$Dissmilarity = \sum_{i,j}^{Levels-1} P_{i,j}|i-j| \tag{2}$$

L2 is an equation for measuring linearity in pixel pairs. The correlation equation is shown in Equation (3).

$$Correlation = \sum_{i,j}^{Levels-1} P_{i,j}\left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}\right] \tag{3}$$

L3 is a measure of image similarity. The value is high if all pixels have the same value. The homogeneity equation is shown in Equation (4).

$$Homogenity = \sum_{i,j=0}^{Levels-1} \frac{P_{i,j}}{1+(i-j)^2} \tag{4}$$

L4 measure of the spatial frequency of the image. The contrast equation is shown in Equation (5).

$$Contrast = \sum_{i,j=0}^{Levels-1} P_{i,j}(i-j)^2 \tag{5}$$

L5 or Angular Second Moment measures the uniformity of pixels in an image. The ASM equation is shown in Equation (6).

$$ASM = \sum_{i,j=0}^{Levels-1} P_{i,j}^2 \qquad (6)$$

L6 is an equation of the gray-level inequality in an image. The energy equation is shown in Equation (7).

$$Energy = \sqrt{ASM} \qquad (7)$$

Where i is the matrix row, j is the matrix column, and $P(i,j)$ is the general matrix element of row (i) and column (j). μi, μj is the average of the matrix's row and column elements. σi, σj are the standard deviations in the rows and columns of the matrix.

The GLCM matrix is calculated using equations 2 to 7 so that it can accurately represent images with fewer parameters. Python supports all available GLCM value extraction techniques. The output of the extraction process is stored in (. CSV) format.

2.4 Classification Techniques

Classification is predicting an unknown data type [20]. This study uses machine learning (ML) to classify the types of batik fabric pattern, namely NB and RF. One of the tools that provide many ML features is WEKA Environment tools. The WEKA tool is a program that provides tools for data preparation, association rule mining, grouping , classification, regression, and visualization in addition to a set of machine learning algorithms for data mining tasks [21],[22].

2.4.1 Random Forest (RF)

RF is one of the ML techniques, a classification of decision trees enhanced from the Classification and Regression Trees (CART) method [23]. RF is capable of handling large datasets and has large input features [24]. Picture of the RF classifier workflow as shown in Figure 3.
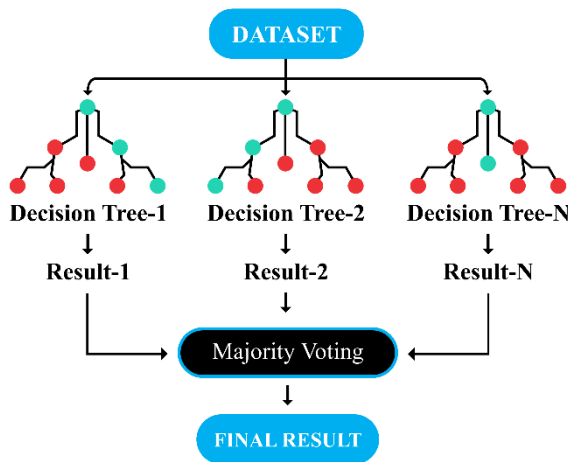
Figure 3. Visualization of the RF classifier

Figure 3 is a dataset that will be trained to produce several decision trees from the decision tree model. Then in each decision tree, the data that has been trained will be categorized as a label. The most votes in voting will determine the final decision on the RF algorithm.

2.4.2 Naïve Bayes (NB)

A probabilistic learning model based on the Bayes theorem, NB is one of the machine learning techniques used to predict the phenomenon of classifying several different classes [22],[25],[26],[27]. Equation for NB classifier shown in Equation 8.

$$P(K|L) = \frac{P(L|K)P(K)}{P(L)} \qquad (8)$$

$P(K)$ represents the prior probability, which indicates the likelihood that K will occur, and $P(L)$ is the marginal probability, which implies the likelihood that Y will occur. The probabilities of K and L were separate quantities that had no impact on one another. The posterior probability, also known as $P(K|L)$, is the likelihood that K will occur after L has already occurred. Last, $P(K|L)$ is identified as the likelihood probability, representing the likelihood that L will occur if K is true [28]. The theorem states that the intrinsic probability, which is calculated based on the facts currently available, is multiplied to determine the probability of an event occurring, as shown in Figure 4.
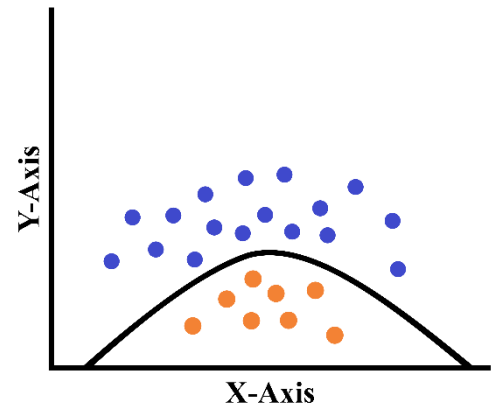
Figure 4. Graphic visualization of NB classifier

Confusion Matrix (CM), a technique used in this study to measure classification accuracy using NB and RF approaches. To evaluate the precision of the classification model, CM contains predictions and actual conditions of classification outcomes [29]. Figure 5 shows the confusion matrix.
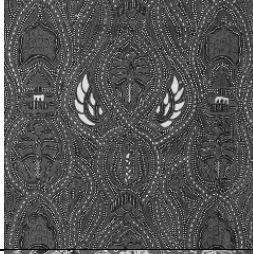

Figure 5. Confusion matrix
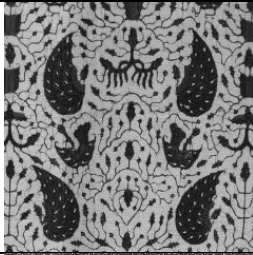
3. RESULT AND DISCUSSION

This section displays the results of the proposed method, which includes pre-processing, feature extraction, and classification. Pre-processing is done using the python programming language and classification techniques are implemented using WEKA tools version 3.8.6.

3.1 Pre-processing

Batik pattern have different types. Different pattern can be used as indicators to detect different batik fabric pattern. The pre-processing results change the color image to grayscale image. Then each image is converted to 50 x 50 pixels to simplify the computation process. The following step after converting pixels is to apply the augmentation data by using rotation, zoom, swipe, flip vertical, and flip horizontal to make the data more diverse. Pre-processing steps performed with the python programming language. Table 1 is the result of the RGB image converted to grayscale.

Table 1. RGB to grayscale image

Type Of Batik	Color Image	Grayscale Image
Sawat		
Semenrante		
Satriomanah		

Based on Table 1, each batik image is changed from a color image to a grayscale image. The primary function of grayscale is to speed up image classification by reducing object features from RGB images that are not needed because the color does not affect the classification of batik fabric pattern [30].

3.2 Feature Extraction

GLCM feature extraction produces 24 features in each image, with six feature parameters, namely L1, L2, L3, L4, and L5. The GLCM feature generates 24 values from each angle. The results of the GLCM feature extraction at an angle of 90° are shown in Table 2.

Table 2. Texture feature extraction results at an angle of 90°

Texture Feature Extraction Results at an Angle of 0°						
Dissimilarity	Correlation	Homogeneity	Contrast	ASM	Energy	Class
45.048	45.60775047	0.047898395	0.055240694	0.094510388	0.068508877	Satriomanah
51.47911111	52.24763705	0.063063819	0.040174667	0.115009638	0.070957903	Satriomanah
49.67777778	51.40926276	0.058520792	0.090811508	0.09869812	0.040074417	Satriomanah
45.79777778	46.73676749	0.097699314	0.07632575	0.095053963	0.061044649	Satriomanah

..	..	..	..	..	..	..
60.14755556	71.09546314	0.015172227	0.010530841	0.211599564	0.00275415	Sawat
59.26444444	67.80718336	0.04139436	0.113103039	0.249128169	0.104301708	Sawat
58.552	65.61672968	0.021875788	0.066433226	0.224280763	0.092154813	Sawat
62.59911111	66.09546314	0.187196081	0.176979542	0.235503639	0.162961042	Sawat
..	..	..	..	..	..	..
41.27644444	41.33506616	0.019199326	0.00355137	0.02194828	0.06858046	Semenrante
43.16222222	40.81758034	0.032989894	0.051611365	0.082826933	0.086246578	Semenrante
45.04177778	45.58837429	0.080022279	0.050227639	0.138111203	0.081224043	Semenrante
36.56222222	37.73771267	0.006272633	0.008842833	0.054262309	0.005443449	Semenrante

Based on Table 2, the test was run using texture feature extraction, namely GLCM. The numerical results of each homogeneity, contrast, ASM, and energy feature in the GLCM, when viewed from a 90° angle, are displayed in each column. The GLCM method extracts the texture of batik fabric for each type of satriomanah, sawat, and semenrante class.

3.3 Classification and Evaluation

The classification technique uses two methods, NB and RF, with a data sample of 600 images from each class. Classification uses WEKA tools by dividing the dataset using a percentage split. The Percentage split used in this research was labeled A, B, and C. A is a percentage split with a dataset of 60% for training data and 40% for testing data. B is a percentage split with a dataset of 70% for training data and 30% for testing data. while C is a percentage split with a dataset of 80% for training data and 20% for testing data. The accuracy results are shown in Figure 5.
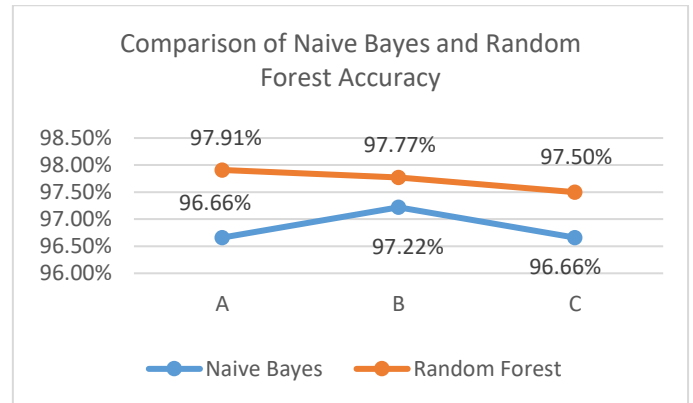

Figure 5. Comparison of NB and RF accuracy chart

Based on Figure 5. The highest accuracy is shown in the RF classification technique with an accuracy of 97.91% in percentage split A, while the lowest accuracy is obtained in the NB classification technique with 96.66% in percentage split A and C. The confusion matrix results, according to Figure 4 with the NB method, are shown in Table 3.

Table 3. Results of NB confusion matrix classification

Dataset Split	Class	Confusion Matrix			Accuracy (%)	Precision	Recall
A	Satriomanah	79	0	0		0.91	1.00
	Sawat	0	80	0	96.66	1.00	1.00
	Semenrante	8	0	73		1.00	0.901
	Avg.				**96.66**	**96.66**	**96.66**
B	Satriomanah	60	0	0		0.92	1.00
	Sawat	0	62	0	97.22	1.00	1.00
	Semenrante	5	0	53		1.00	0.91
	Avg.				**97.22**	**97.22**	**97.22**
C	Satriomanah	41	0	0		0.91	1.00
	Sawat	0	42	0	96.66	1.00	1.00
	Semenrante	4	0	33		1.00	1.00
	Avg.				**96.66**	**96.66**	**96.66**

Based on Table 3, the best classification results among the distribution of datasets using a percentage split were obtained in dataset B. Accuracy, precision, and recall for this model averaged 97.22, 97.22, and 97.22 by using the equation shown in Table 5. The confusion matrix results, according to Figure 4 with the RF method, are shown in Table 4.

Table 4. Results of RF confusion matrix classification

Dataset Split	Class	Confusion Matrix			Accuracy (%)	Precision	Recall
A	Satriomanah	79	0	0		0.94	1.00
	Sawat	0	80	0	97.91	1.00	1.00
	Semenrante	5	0	76		1.00	0.94
	Avg.				**97.91**	**97.91**	**97.91**
B	Satriomanah	60	0	0		0.94	1.00
	Sawat	0	62	0	97.77	1.00	1.00
	Semenrante	5	0	53		1.00	0.96
	Avg.				**97.77**	**97.77**	**97.77**
C	Satriomanah	41	0	0		0.93	1.00
	Sawat	0	42	0	97.50	1.00	1.00
	Semenrante	4	0	33		1.00	0.91
	Avg.				**97.50**	**97.50**	**97.50**

Table 4 shows the best classification results obtained in dataset A. He showed that the average values for accuracy, precision, and recall were 97.91, 97.91, and 97.91 by using the equation shown in Table 5.

Table 5. The equation to calculate the effectiveness of classification

Equation	Focus
$Accuracy = \dfrac{TP + TN}{TP + TN + FP + FN}$	Effectiveness of whole classification
$Precision = \dfrac{TP}{TP + FP}$	The data labels classified as match as the positive labels
$Recall = \dfrac{TP}{TP + FN}$	Defined as the proportion of positive cases that are accurately predicted with all other positive cases

Based on Table 5 accuracy, recall, and precision are classification evaluation metrics used to assess the performance of machine learning models.

The RF method outperforms the NB classifier method in recognizing batik pattern using the GLCM feature, according to the performance value of the classification process in data testing. This result is due to the advantages and disadvantages of both methods.

4. CONCLUSION

The study of classifying the types of batik fabric pattern using the RF and NB methods has been carried out using 600 images of batik images used as a dataset, with each type of batik having 200 data samples. Classifying the types of batik fabric pattern includes collecting batik fabric images, preprocessing, and texture feature extraction, followed by the classification process by dividing the entire dataset using the percentage split method. This research shows that these two methods can classify the types of batik fabric pattern. The RF method has the highest accuracy rate of 97.91% in dataset A, while the NB method in the same dataset has an accuracy rate of 96.66%. Based on the results of this accuracy, it is stated that the RF method is better at classifying the types of batik fabric pattern. This result is also affected because RF does not require a large number of datasets. Future research can use more variety batik datasets, another classification technique, and other data collection techniques with various environmental conditions to be compared with this paper.

REFERENCES

[1] D. M. S. Arsa and A. A. N. H. Susila, "VGG16 in Batik Classification based on Random Forest," *Proc. 2019 Int. Conf. Inf. Manag. Technol. ICIMTech 2019*, vol. 1, no. August, pp. 295–299, 2019, doi: 10.1109/ICIMTech.2019.8843844.

[2] N. D. Girsang, "Literature Study of Convolutional Neural Network Algorithm for Batik Classification," *Brill. Res. Artif. Intell.*, vol. 1, no. 1, pp. 1–7, 2021, doi: 10.47709/brilliance.v1i1.1069.

[3] 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 Comput. Sci.*, vol. 179, no. 2019, pp. 863–870, 2021, doi: 10.1016/j.procs.2021.01.075.

[4] J. Kusanti and A. Suprapto, "Combination of Otsu and Canny Method to Identify the Characteristics of Solo Batik as Surakarta Traditional Batik," *Proc. - 2019 2nd Int. Conf. Comput. Informatics Eng. Artif. Intell. Roles Ind. Revolut. 4.0, IC2IE 2019*, pp. 63–68, 2019, doi: 10.1109/IC2IE47452.2019.8940884.

[5] H. Azarmdel, S. S. Mohtasebi, A. Jafari, and A.

Rosado Muñoz, "Developing an orientation and cutting point determination algorithm for a trout fish processing system using machine vision," *Comput. Electron. Agric.*, vol. 162, no. January, pp. 613–629, 2019, doi: 10.1016/j.compag.2019.05.005.

[6] H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, "Computer vision technology in agricultural automation —A review," *Inf. Process. Agric.*, vol. 7, no. 1, pp. 1–19, 2020, doi: 10.1016/j.inpa.2019.09.006.

[7] A. Yudhana, R. Umar, and S. Saputra, "Fish Freshness Identification Using Machine Learning: Performance Comparison of k-NN and Naïve Bayes Classifier," *J. Comput. Sci. Eng.*, vol. 16, no. 3, pp. 153–164, 2022, doi: 10.5626/JCSE.2022.16.3.153.

[8] K. Nugroho and E. Winarno, "Spoofing Detection of Fake Speech Using Deep Neural Network Algorithm," *Int. Semin. Appl. Technol. Inf. Commun.*, 2022, doi: 10.1109/iSemantic55962.2022.9920401.

[9] F. N. Suteja, E. W. Hidayat, and N. Widiyasono, "Implementation of Image Enhancement Algorithm for Image Forensics using Mathlab," *J. Online Inform.*, vol. 4, no. 2, p. 79, 2020, doi: 10.15575/join.v4i2.314.

[10] R. A. Surya, A. Fadlil, and A. Yudhana, "Identification of Pekalongan Batik Images Using Backpropagation Method," *J. Phys. Conf. Ser.*, vol. 1373, no. 1, 2019, doi: 10.1088/1742-6596/1373/1/012049.

[11] H. T. Zaw, N. Maneerat, and K. Y. Win, "Brain tumor detection based on Naïve Bayes classification," *Proceeding - 5th Int. Conf. Eng. Appl. Sci. Technol. ICEAST 2019*, pp. 1–4, 2019, doi: 10.1109/ICEAST.2019.8802562.

[12] M. Bansal, M. Kumar, M. Sachdeva, and A. Mittal, "Transfer learning for image classification using VGG19: Caltech-101 image data set," *J. Ambient Intell. Humaniz. Comput.*, no. 0123456789, 2021, doi: 10.1007/s12652-021-03488-z.

[13] C. U. Khasanah, E. Utami, and S. Raharjo, "Implementation of Data Augmentation Using Convolutional Neural Network for Batik Classification," *2020 8th Int. Conf. Cyber IT Serv. Manag. CITSM 2020*, pp. 20–24, 2020, doi: 10.1109/CITSM50537.2020.9268890.

[14] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[15] A. Humeau-Heurtier, "Texture feature extraction methods: A survey," *IEEE Access*, vol. 7, pp. 8975–9000, 2019, doi: 10.1109/ACCESS.2018.2890743.

[16] I. M. Achievements, "Classification Based on Machine Learning Methods for Identification of Image Matching Achievements," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 158, pp. 198–206, 2022, doi: doi.org/10.29207/resti.v6i2.3826.

[17] Ş. Öztürk and B. Akdemir, "Application of Feature Extraction and Classification Methods for Histopathological Image using GLCM, LBP, LBGLCM, GLRLM and SFTA," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 40–46, 2018, doi: 10.1016/j.procs.2018.05.057.

[18] Priyanka and D. Kumar, "Feature Extraction and Selection of kidney Ultrasound Images Using GLCM and PCA," *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 1722–1731, 2020, doi: 10.1016/j.procs.2020.03.382.

[19] I. Riadi, A. Fadlil, I. Julda, and D. E. P. Putra, "Batik Pattern Classification using Naïve Bayes Method Based on Texture Feature Extraction," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 9, no. 1, p. 9, 2023, doi: https://doi.org/10.23917/khif.v9i1.21207.

[20] A. Fadlil and D. Prayogi, "Face Recognition Using Machine Learning Algorithm Based on Raspberry Pi 4b," *Int. J. Artif. Intell. Res.*, vol. ISSN, no. 1, pp. 2579–7298, 2022, doi: 10.29099/ijair.v7i1.321.

[21] B. Geyik, K. Erensoy, and E. Kocyigit, "Detection of Phishing Websites from URLs by using Classification Techniques on WEKA," *Proc. 6th Int. Conf. Inven. Comput. Technol. ICICT 2021*, pp. 120–125, 2021, doi: 10.1109/ICICT50816.2021.9358642.

[22] K. Alpan and G. Savaş İlgi, "Classification of Diabetes Dataset with Data Mining Techniques by Using WEKA Approach," *Int. Symp. Multidiscip. Stud. Innov. Technol.*, 2020, doi: 10.1109/ISMSIT50672.2020.9254720.

[23] D. C. Yadav and S. Pal, "Prediction of Thyroid Disease using Decision Tree Ensemble Method," *Human-Intelligent Syst. Integr.*, vol. 2, no. 1–4, pp. 89–95, 2020, doi: 10.1007/s42454-020-00006-y.

[24] A. K. Shrivas, A. K. Dewangan, and S. M. Ghosh, "Robust Text Classifier for Classification of Spam E-Mail Documents with Feature Selection Technique," *Ing. des Syst. d'Information*, vol. 26, no. 5, pp. 437–444, 2021, doi: 10.18280/isi.260502.

[25] A. Yudhana, I. Riadi, and F. Ridho, "DDoS classification using neural network and naïve bayes methods for network forensics," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 11, pp. 177–183, 2018, doi: 10.14569/ijacsa.2018.091125.

[26] B. M. Patil and V. Burkpalli, "A Perspective View of Cotton Leaf Image Classification using Machine Learning Algorithms using WEKA," *Adv. Human-Computer Interact.*, vol. 2021, 2021, doi: 10.1155/2021/9367778.

[27] A. Yudhana, D. Sulistyo, and I. Mufandi, "GIS-based and Naïve Bayes for nitrogen soil mapping in Lendah, Indonesia," *Sens. Bio-Sensing Res.*, vol. 33, p. 100435, 2021, doi: 10.1016/j.sbsr.2021.100435.

[28] C. N. Villavicencio, J. J. E. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, "Covid-19 Prediction Applying Supervised Machine Learning Algorithms with Comparative Analysis using WEKA," *Algorithms*, vol. 14, no. 7, 2021, doi: 10.3390/a14070201.

[29] A. Taheri-Garavand, S. Fatahi, A. Banan, and Y. Makino, "Real-time Nondestructive Monitoring of Common Carp Fish Freshness using Robust Vision-based Intelligent Modeling Approaches," *Comput. Electron. Agric.*, vol. 159, no. August 2018, pp. 16–27, 2019, doi: 10.1016/j.compag.2019.02.023.

[30] A. Abdulrahman and S. Varol, "A Review of Image Segmentation Using MATLAB Environment," *8th Int. Symp. Digit. Forensics Secur. ISDFS 2020*, pp. 8–12, 2020, doi: 10.1109/ISDFS49300.2020.9116191.