Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 2 (2022) 198 - 206 ISSN Media Electronic: 2580-0760

Classification Based on *Machine Learning Methods* for Identification of *Image Matching Achievements*

Rusydi Umar[1], Imam Riadi[2], Dewi Astria Faroek[3]
[1.3] Informatics Engineering Study Program, Ahmad Dahlan University
[2] Information Systems Study Program, Ahmad Dahlan University
[1] rusydi@mti.uad.ac.id, [2]imam.riadi@is.uad.ac.id, [3]dewi1907048014@webmail.uad.ac.id*

Abstract

Classification is one method in image processing. Image processing to search for similar images or with similarity ownership is called image matching or image matching. In the measurement of image matching, the original and fake logo objects are used. Identification of similarity manually with the help of human vision is not necessarily precise and it is difficult to obtain accurate results. Based on this, the identification of image matching of the original and fake logos automatically requires an application, in order to obtain precise and more accurate results. Identification of image suitability is determined through the image segmentation process, and feature extraction is based on the statistics of Red-Green-Blue (RGB), Hue-Saturation-Value (HSV), feature extraction of area, perimeter, eccentricity, and tangent distance measurements. The purpose of this study includes the identification of the achievement of image-matching logo images with comparisons of accuracy between various machine learning methods. The use of machine learning methods in this study includes the k-Nearest Neighbor (kNN), Random Forest (RF), and Multilayer Perceptron (MLP) methods. The use of the dataset includes eighteen training data and eight logo image testing data, divided into genuine and fake classes. The results of the measurement of the accuracy value obtained a value of seventy-five percent with the kNN method or the RF method, while the MLP method obtained an accuracy value of eighty-seven point five percent. Based on these results, it can be concluded that the MLP method with the highest accuracy value was chosen as a classification model from machine learning to identify the achievement of image matching on the original and fake logos. For further development, the system can be developed using other methods or a combination of different methods, in order to obtain better accurate results.

Keywords: *image matching, logo, machine learning, kNN, RF, MLP*

1. Introduction

Classification is one method of image processing. Classification can also be incorporated into analytical methods to differentiate data into classes with the aim of generating relative patterns and value owners [1] . Several *machine learning* classification models, including the *k-Nearest Neighbor* (kNN), *Random Forest* (RF) classification , and *Multilayer Perceptron* (MLP) [2] . These classification models are based on the use of statistical parameter values for RGB, HSV, shape characteristics of *the area, perimeter, eccentricity*, and measurement of *tangent distance values* [3] . Based on this, a classification based on the use of *machine learning methods was carried out* to identify *image matching* on a logo based on the measurement results of the original and fake images.

Based on this background, several *state of the* art related to the use of *machine learning -based classification* as an identifier for *image matching* in a logo. Several previous studies were used as a reference, namely research by Alda Putri [4] on the use of *random forest* classification to identify a vehicle logo in cases of vehicle number plate forgery. In this research, feature extraction is based on the *Local Binary Pattern* (LBP) algorithm and classification with the highest accuracy value of 88.89%. The results of research from Hariyanto [5] regarding the detection of real or fake fingerprints using the *Convolution Neural Network* (CNN) method and based on *a dataset* of comparison results from a training indicated by a *training graph* with a value of 0.97 for validation accuracy of 0.96 and a *loss value of* 0. .05 for a validation *loss of* 0.09. The highest accuracy of the measurement results is given by the comparison value of *training accuracy* to *validation*.

Accepted by Editor: 22-01-2022 | Completed Revision: 29-03-2022 | Published Online: 20-04-2022

The results of research by Bambang [6] are related to the analysis of object compatibility in digital images, using the *Scale Invariant Feature Transform* (SIFT) algorithm and RGB color histograms, but no classification method is used, but the use of these methods results in object matching based on the similarity of *keypoint locations* and number of objects. pixels as an object match hint. The results of Lisdawati's research [7] are related to the use of signatures as research objects for measuring the performance (performance) of back- *progation neural networks* . In this study, *cross validation was* used in the measurement and the resulting performance was 51%, while in the first customer *testing process* a score of 61% was obtained and the second measurement for non-customers was 69%. Based on these results, it is shown that *cross validation in the backpropagation* ANN method is indicated by performance in recognizing customer signature patterns. The results of research from Irfan [8] regarding the vehicle classification system based on image processing using the MLP . method which based on the three types of vehicles, namely cars, buses, and trucks with the evaluation results in the form of an average *accuracy value of 87.60%.*

Based on a number *of state of the art*, this research was conducted. Research on *image matching* of logos has been done, but there has been no research related to the identification of real and fake images based on classification performance. Digital image as a form of ownership of some similarities can be measured based on a parameter. *Image matching* is one part of *image processing* to search for similarities in an image. One of the *essential components* of calculating the image similarity value is the result of an algorithm or method for matching. Comparison in calculating the accuracy of the similarity/similarity values can be seen from the process of measuring the model's performance from an algorithm. The suitability of the image characteristics, one of which can be seen from the use of the distance measurement method.

Logo is part of a *brand* of a product and its use is important for *brand recognition* to consumers and the market [9] . Plagiarism of logo images often occurs so that it can injure a work. The logo becomes a testing tool in this study for measuring the similarity between the original and fake logos. This research is an attempt to compare the performance results of the *machine learning method* used. The method with the best accuracy is used as an automatic classification model in identifying real and fake images based on the *accuracy value* in the measurement process. Based on this description, the research targets were set which include (i) *dataset distribution* , (ii) *preprocessing* , (iii) image segmentation, (iv) feature extraction, and (v) classification and evaluation based on KNN, RF, MLP methods.

2. Research Methods

The research method is an algorithm in carrying out research to obtain each research target [AA, BB] and is made in the form of a flow chart consisting of several stages, namely (i) *dataset division* and *preprocessing,* (ii) image segmentation process, (iii) extraction process characteristics and measurements of *tangent distance,* (iv) classification based on kNN, RF, and MLP methods and (v) evaluation of measurement results. The research method flow chart, as shown in Figure 1.

Based on Figure 1, it can be explained that the process carried out in the form of dividing the *dataset* in this study is divided into *training* and *testing data* which is carried out with *k-fold cross validation. Preprocessing* is done with *CorelDraw editing software.* Image segmentation is assisted by the *k-means clustering method,* feature extraction is carried out with *Matlab* version R2019a, while the classification is done with *Weka tools* version *3.8.5.*

2.1. *Dataset* Sharing and *Preprocessing*

#1) *Dataset*

Research data is in the form of primary data, which is an image of a logo with an *editing process* and stored on a computer. The data is adjusted to the number of *datasets* for *training* and *testing,* which consists of twenty-six logo images and is divided into eighteen *training data* and eight *testing data.* An example of a logo image *dataset,* as shown in Figure 2.

Figure 2. Example *dataset* logo image

#2) *Preprocessing*

Preprocessing as the first step in data processing, in order to produce maximum value accuracy. Data *preprocessing is done by cropping* technique , namely cutting and removing parts of the image that are not needed in the image identification process [10] . The *cropping process* is done with the *CorelDraw Graphic X7 application. The preprocessing* stage of the resulting image is carried out by a conversion process from RGB to L*a*b, in order to achieve ease to the segmentation stage using the *k-means clustering method.*

2.2. Image Segmentation

Segmentation is the process of separating one object from another in an image. The object of the segmentation results can be used as input *for* the next process [11] . *Clustering* -based image segmentation is used on multidimensional data with the aim of grouping image pixels into several clusters *[* 12] . *Clustering* is defined as a similarity search technique in a data group [13] . *K-means clustering* is an algorithm for dividing

DOI: https://doi.org/10.29207/resti.v6i2.3826

objects into groups, determining the initial number of groups by defining the initial *centroid value*. The process was carried out with *k-means* repeatedly [14] . *K-Means* is used in the segmentation process, so that criteria can be assigned to the same object area [15] .

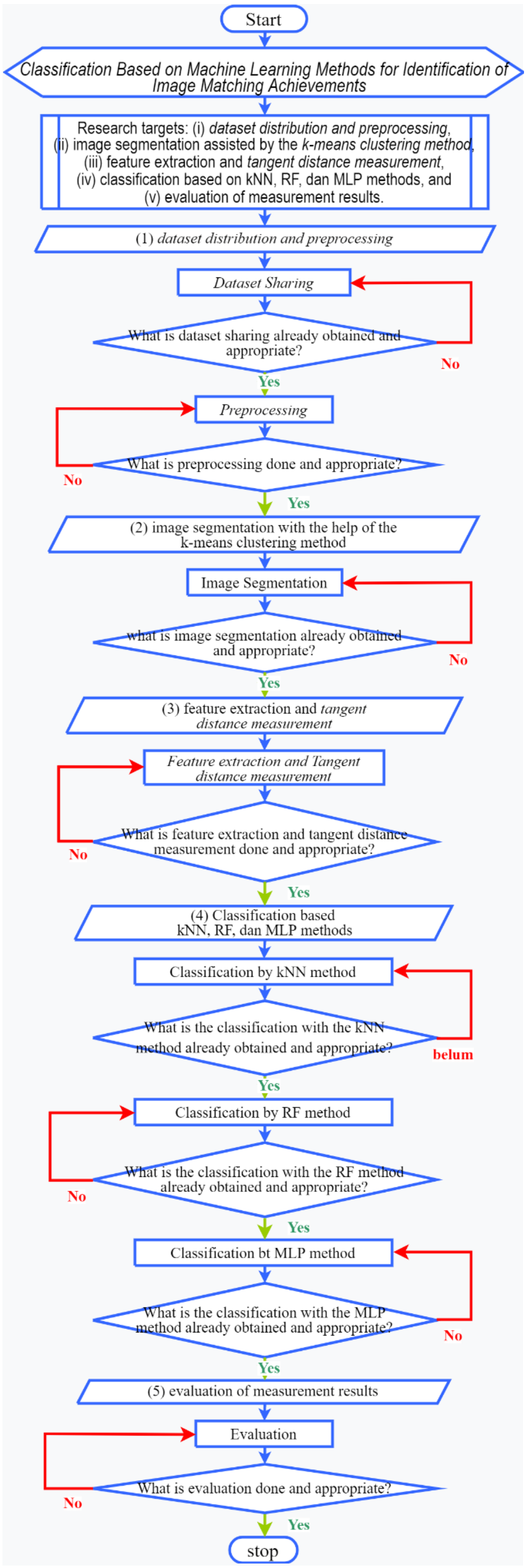


Figure 1. Flowchart of research methods

2.3 Feature Extract

The feature extraction process is a process for obtaining information contained in the image, so that it is easily recognized and distinguished from other images. Characteristics or features are divided into several parts, including features of shape, texture, color, geometry, and size. The feature extraction used in this research is based on RGB, HSV, and *Shape Measurements* (area, perimeter and *eccentricity*) color characteristics.

RGB color with a function as a color presenter in the form of red (R), green (G), and blue (B) components. The reference for the size of the color component is used by an eight-bit system, starting with values ranging from 0 to 255, so that the color component presented is 16,581,375 colors. The RGB color intensity value is in the interval 0 to 255 for each pixel [16] .

HSV colors consist of red, violet, and yellow which are expressed as true colors. Based on that, the *hue is in the* form of reddish and greenish, the *saturation* is the purity or strength of the color, and the *value* is the brightness of the color. The three HSV values range from zero to one hundred percent, while in black with a value of 0 the higher the value, the new color appears [17] .

The HSV color space uses equations (1), (2), and (3) [17] .

$$h\ (hue) = \begin{cases} 0, jika\ \max = min \\ 60^0 x \left(\frac{G-B}{\max - min} mod\ 6\right), jika\ \max = R \\ 60^0 x \left(\frac{B-R}{\max - min} + 2\right), jika\ \max = G \\ 60^0 x \left(\frac{R-G}{\max - min} + 4\right), jika\ \max = B \end{cases} \quad (1)$$

$$s\ (saturation) = \begin{cases} 0\ , jika\ \max = min \\ \frac{\max - min}{V}, otherwise \end{cases} \quad (2)$$

and

$$Value\ (V) = max \quad (3)$$

Shape measurement values consist of *area*, *perimeter*, and *eccentricity values*. *The area* in the image is calculated based on the number of pixels with a value of 1, the *perimeter* is calculated based on the number of the object's outermost pixels, and *eccentricity* is a feature extraction technique that aims to extract or extract the *eccentricity value*. The *eccentricity value* on the object is round or 0, while the oval shape or similar to a straight line is 1.

Tangent distance measurements were carried out to determine the distance value after image transformation and were optimized by gradient descent method for finding the nearest point [18] . The measurement results are used for input values in the classification process.

2.4. Classification Method

The classification process is carried out using three different methods to compare the performance of each

DOI: https://doi.org/10.29207/resti.v6i2.3826

in identifying the image match between the original and fake images. The classification methods in *machine learning* used in this study, namely kNN, RF, and MLP.

2.4.1. *k-Nearest Neighbor* (kNN) Classification

K-Nearst Neighbor (kNN) is one of the guided algorithm methods with the results of a new *query instance* classified based on the majority of the categories. The kNN classification is carried out on the *training data* when taking the value of *k* nearest neighbors, with the number of nearest neighbors being *k* [19] . The best *k value* for this algorithm depends on the data, while a high *k value* has an impact on reducing *noise effects* during classification [20] .

The problem solving step in the kNN classification uses the *k value parameter* which is determined by the number of nearest neighbors, the distance of each *sample* data with the tested data is calculated, the data is sorted from the smallest to the largest distance, then paired to the appropriate class, and the number of classes from the nearest neighbor is determined as the data class being evaluated. The most *voting process* in kNN is used to classify the value of *k* . Equation (4) [21] was used for the voting process.

$$d_1 = \sqrt{\sum_{i=1}^{p}(X_1 - X_2)^2} \quad (4)$$

with:d_1 = distance,X_1 = sample data, X_2 = data *testing*, i = variable, and$p$ = data dimension.

2.4.2. *Random Forest* Classification (RF)

The RF method is one of the best *performing algorithms* for classifying large amounts of data which was developed from the CART method [22] . The term trees grown in the RF method, is an attempt to form as a forest (*forest*), then analyzed in a collection of these trees. Three important aspects of RF are: (i) implementation of *bootstrap sampling* with the aim of constructing a prediction tree, (ii) predicting each decision tree randomly using a *predictor* , and (iii) combining the results of each decision tree by means of *majority vote* for determining the mean value in the regression [23] .

2.4.3. *Multilayer Perceptron* (MLP) Classification

Multilayer Perceptron (MLP) is a linear *classifier with the conversion of a set of inputs* into *outputs.* The basic principle of MLP, has one or more additional *layers* called *hidden units* (hidden units) and serves as a link between *input* and *output units. Additional layers* are called *hidden units*. *Weight layers* are located between adjacent unit *layers (input, hidden*, or *output*).

The learning process in MLP is in the form of *backpropagation* which consists of four stages, namely initialization, activation, weight training, and iteration. The initialization stage of the initial weight value and the *threshold value* is determined randomly within certain limits. The activation stages are given predictable *input* and *output values*. Stages of weight training, in the form of the actual output value compared to the expected expected value, and weight adjustments are made. The second and third stages are repeated processes (iterations) until certain conditions are achieved [24] .

2.5. Measurement Evaluation

The evaluation stage of the classification method is carried out by measuring the accuracy value of the classification results in percent based on equations and for the analysis process a *confusion matrix is used. Confusion matrix* is a tool with a function for the analysis of a *classifier* in the classification process[25]. The prediction value is based on the *confusion matrix,* as shown in Table 2.

Table 2. Prediction values based on *confusion matrix*

Predicted Value	True Value	
	TRUE	*FALSE*
TRUE	*True Positives* (TP): *Correct result*	*False Positives* (FP): *Unexpected result*
FALSE	*False Negatives* (FN): *Missing result*	*True Negatives* (TN): *Correct absence of result*

True Positives (TP) is the number of correctly classified test data. *True Negatives* (TN) is the number of correctly classified other class tests. *False Positives* (FP) is the number of test data that are classified as incorrect. *False Negatives* (FN) is the number of test data for other classes classified incorrectly into that class [26] . Determination of the *accuracy value* is used equation (5)

$$Nilai\ Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (5)$$

Accuracy value or known as recognition rate, is used to analyze the accuracy of the model in the classification.

3. Results and Discussions

This chapter is in the form of appointment of results and discussion according to research objectives based on research methods.

3.1. *Dataset* Sharing and *Preprocessing*

The distribution of *the dataset* in this study is divided into *training* and *testing data* before the classification process. *Training* data as model training data, while *testing data* is used as model test data. The results of the distribution of the *dataset*, as in Table 1.

Table 1. Results of *dataset division*

Category	*Training*	*Testing*
Original Image	9	4
Fake Image	9	4
Amount	18	8

DOI: https://doi.org/10.29207/resti.v6i2.3826

Based on Table 1, it can be explained that this research was divided into two *datasets, namely training* and *testing data* with a ratio of 70% and 30%. The number of *datasets* used is 26 data, which is divided into *training data consisting of 18 original and fake image classes, and testing* data consisting of 8 original and fake classes.

preprocessing stage is in the form of image *cropping, the image dataset* taken is the collection of original and fake logo images. The *cropping process* is done through cutting and removing parts of the image that are not needed for image identification. The result of *cropping* the image of a logo, as shown in Figure 3.

before after

Figure 3. Display of the *cropped* image of a logo

Based on Figure 3, it can be explained that in the logo image before and after the *cropping process is carried out*, to remove the edge areas of the image that are not needed. This process is done with image processing *software* and is done manually.

3.2. Image Segmentation Assisted *by k-Means Clustering Method*

Image segmentation on the original and fake logos is done by using the *k-means clustering method*, so that there is an object confirmation in the follow-up process. The display of the image segmentation results on the original and fake logos, as shown in Table 3.

Table 3. Display of image segmentation results on original and fake logos

Original Image	Fake Image
	

Based on Table 3, it can be explained that the image segmentation process obtained results consisting of converting color images to l*a*b, object *clusters one, two, and three*. This process is carried out in order to obtain the results of the separation between *foreground* and *background objects* that will be used in the follow-up process.

3.3. Feature Extraction and *Tangent Distance Measurement*

The color segmentation image using *k-means clustering is* followed up with a logo feature extraction process and *tangent distance measurement* based on statistical characteristics in the form of *mean values* with RGB color intensity, HSV, and shape characteristics used in the form of *area, perimeter,* and *eccentricity values and tangent* measurements *distance.* The display of the feature extraction process and the measurement of the *tangent distance value*, as shown in Table 4.

Table 4. Display of the results of the feature extraction process and measurement of the *tangent distance value*

No.	Image	Feature Extraction Value
1	 Original Image	R=166.51 G=98.77 B=94.34 H=0.2479 S=0.4613 V=0.6540 Area=66416 Perimeter=1532.68 Eccentricity=0,5797 Tangent Value=0.00184
2	 Fake Image	R=238,70 G=88.69 B=40.56 H=0.0419 S=0.8290 V=0.9361 Area=7384 Perimeter=441.60 Eccentricity=0.8602 Tangent Value=0.00162

Based on Table 4, it can be explained that the results of the feature extraction process and *tangent distance measurements* obtained different values. The value then stored in *Microsoft Excel* (*xlx) with the extension (*csv), so that it can be converted into a *file* (ARFF), so that it can be classified using the Weka application. The display of the results of the collection of feature extraction and *tangent distance measurements,* as shown in Figure 4.

Based on Figure 4, it can be explained that the results of the feature extraction process and the measurement of the *tangent distance* are stored with *Command Separated Value* (CSV) separated by a comma (,) in each data.

DOI: https://doi.org/10.29207/resti.v6i2.3826

	A
1	R,G,B,H ,S,V,Area,Perimeter,Eccentricity,Tan,Class
2	166.51,98.77,94.34,0.2479,0.4613,0.6540,66416,1532.68,0.5797,0.00184,Asli
3	166.44,98.59,94.28,0.2469,0.4620,0.6537,13149,6819.87,0.5794,0.00180,Asli
4	166.43,98.58,94.27,0.2469,0.4620,0.6537,13148,6818.95,0.5794,0.00179,Asli
5	241.15,106.48,69.80,0.2095,0.7147,0.9457,27057,3835.85,0.8554,0.00177,Asli
6	166.45,98.58,94.28,0.2469,0.4621,0.6538,13146,6810.57,0.5791,0.00175,Asli
7	166.43,98.58,94.27,0.2468,0.4620,0.6537,13149,6810.00,0.5791,0.00181,Asli
8	241.13,105.50,68.97,0.2109,0.7181,0.9456,26748,3496.28,0.8547,0.00180,Asli
9	166.43,98.57,94.27,0.2469,0.4620,0.6537,13145,6794.84,0.5794,0.00182,Asli
10	166.43,98.57,94.27,0.2469,0.4620,0.6537,13153,6798.07,0.5795,0.00175,Asli
11	241.60,110.09,82.61,0.4193,0.6718,0.9475,23832,908.94,0.8747,0.00202,Palsu
12	239.43,98.08,76.98,0.4946,0.7128,0.9389,23222,882.74,0.8649,0.00233,Palsu
13	243.58,128.39,90.15,0.1000,0.6302,0.9552,23258,855.70,0.8971,0.00196,Palsu
14	219.36,25.96,117.60,0.8737,0.8944,0.8602,3309,1188.09,0.9539,0.00166,Palsu
15	214.87,54.42,47.75,0.1434,0.7951,0.8426,24543,643.43,0.7628,0.00190,Palsu
16	198.45,41.99,44.10,0.1362,0.7990,0.7794,204599,1601.17,0.2738,0.00187,Palsu
17	235.83,71.72,61.60,0.4892,0.7983,0.9248,22840,861.64,0.8904,0.00297,Palsu
18	224.53,89.89,168.73,0.7379,0.6727,0.8806,13148,433.69,0.0814,0.00177,Palsu
19	222.02,84.92,220.78,0.8357,0.6260,0.8824,6419,544.26,0.6236,0.00189,Palsu
20	

Figure 4. Display of the results of feature extraction collection and *tangent distance measurement*

3.4. Classification

Determination of the classification is carried out by the value of the results of feature extraction and measurement of the *tangent distance* to the *training* data and *testing data, then classification is carried out using machine learning* methods on the selected data. Evaluation is carried out using a *confusion matrix approach,* in order to resulting *accuracy value.* The higher the accuracy value, the better the method used.

#1) Classification by kNN method

Classification using the kNN method is in the form of applying *10-fold cross validation.* The results of the measurement process on kNN obtained *correctly classified results* of 94.4% and *incorrerrectly classified* of 5.55%. The results of the classification process for *training data* using the kNN method *are* as shown in Figure 5.

Classifier output

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          17               94.4444 %
Incorrectly Classified Instances         1                5.5556 %
Kappa statistic                          0.8889
Mean absolute error                      0.1046
```

Figure 5. The results of the classification process for *training data* using the kNN . method

Based on Figure 5, identification of the accuracy of the logo image with the help of the kNN method was carried out and it was successfully carried out, so that two original logos and six fake logos were obtained in the *testing data.* The results of the identification on *testing data* with the kNN method, as shown in Figure 6.

HasilKNN.arff

Relation: datalengkap - Copy_predicted

No.	1: R Numeric	2: G Numeric	3: B Numeric	4: H Numeric	5: S Numeric	6: V Numeric	7: Area Numeric	8: Perimeter Numeric	9: Eccentricity Numeric	10: Tan Numeric	11: prediction margin Numeric	12: predicted Cl... Nominal
1	166....	98.77	94.34	0.24...	0.46...	0.654	664...	1532.68	0.5797	0.00...	0.9	Asli
2	166....	98.57	94.27	0.24...	0.462	0.65...	131...	6794.84	0.5794	0.00...	0.9	Asli
3	198....	41.99	44.1	0.13...	0.799	0.77...	204...	1601.17	0.2738	0.00...	-0.9	Palsu
4	222....	84.92	220....	0.83...	0.626	0.88...	641...	544.26	0.6236	0.00...	-0.9	Palsu
5	238.7	88.69	40.56	0.04...	0.829	0.93...	738...	441.6	0.8602	0.00...	-0.9	Palsu
6	230....	90.06	65.26	0.27...	0.72...	0.90...	158...	1059.89	0.8658	0.00...	-0.9	Palsu
7	223.6	67.0	60.95	0.13...	0.75...	0.87...	230...	623.72	623.72	0.00...	-0.9	Palsu
8	239.5	98.21	76.7	0.491	0.71...	0.93...	230...	901.92	0.8654	0.00...	-0.9	Palsu

Figure 6. Identification results on *testing data* using the kNN . method

Based on Figure 6, it can be explained that the identification results in the *testing data* were successfully shown with two *predicted classes*, original two, and six false. The results of the measurement process for the accuracy value of the *testing data* using the kNN method, as shown in Table 5.

DOI: https://doi.org/10.29207/resti.v6i2.3826

Table 5. The results of measuring the *accuracy value* on data *testing* with the kNN . method

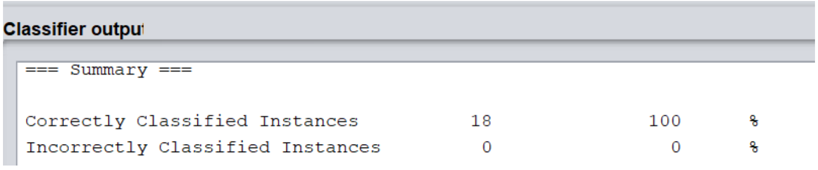
Class Logo	A	P
A = Original	2	2
P = False	0	4
% *Accuracy* $=\frac{6}{8}\ \boldsymbol{x}\ \mathbf{100} = \mathbf{75\%}$		

Referring to Table 5, the percentage *accuracy* value is $=\frac{6}{8}\ x\ 100 = 75\%$.

#2) Classification by RF. method

Classification with the RF method of *training data*, obtained *correctly* by 100% and *incorrectly* by 0%. The results of the identification of the *training data* with the RF method *are* shown in Figure 7.

Based on Figure 7, the accuracy of the logo image using the RF method was carried out and it was successfully carried out, so that two original logos and six fake logos were obtained in the *testing data*. The results of the identification of the *testing data* using the RF method are shown in Figure 8.

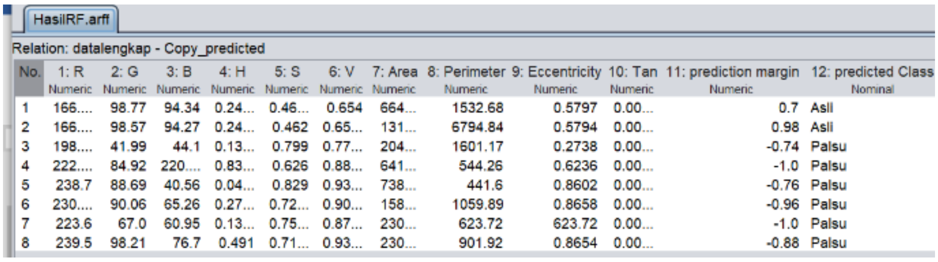

Classifier output

=== Summary ===

Correctly Classified Instances 18 100 %
Incorrectly Classified Instances 0 0 %

Figure 7. Identification results on *training data* using the RF . method

Based on Figure 8, it can be explained that the results of the identification with the RF method were successfully obtained with *predicted class* with the results of the original and fake classes in the data *testing*.


HasilRF.arff

Relation: datalengkap - Copy_predicted

No.	1: R Numeric	2: G Numeric	3: B Numeric	4: H Numeric	5: S Numeric	6: V Numeric	7: Area Numeric	8: Perimeter Numeric	9: Eccentricity Numeric	10: Tan Numeric	11: prediction margin Numeric	12: predicted Class Nominal
1	166....	98.77	94.34	0.24...	0.46...	0.654	664...	1532.68	0.5797	0.00...	0.7	Asli
2	166....	98.57	94.27	0.24...	0.462	0.65...	131...	6794.84	0.5794	0.00...	0.98	Asli
3	198....	41.99	44.1	0.13...	0.799	0.77...	204...	1601.17	0.2738	0.00...	-0.74	Palsu
4	222....	84.92	220....	0.83...	0.626	0.88...	641...	544.26	0.6236	0.00...	-1.0	Palsu
5	238.7	88.69	40.56	0.04...	0.829	0.93...	738...	441.6	0.8602	0.00...	-0.76	Palsu
6	230....	90.06	65.26	0.27...	0.72...	0.90...	158...	1059.89	0.8658	0.00...	-0.96	Palsu
7	223.6	67.0	60.95	0.13...	0.75...	0.87...	230...	623.72	623.72	0.00...	-1.0	Palsu
8	239.5	98.21	76.7	0.491	0.71...	0.93...	230...	901.92	0.8654	0.00...	-0.88	Palsu

Figure 8. Identification results on *testing data* using the RF. method

The results of measuring *accuracy values for testing* data using the RF method are as shown in Table 6.

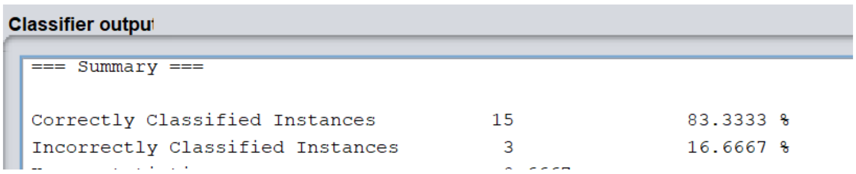
Table 6. The results of measuring *accuracy values for testing* data using the RF method

Class Logo	A	P
A= Original	2	2
P= False	0	4
% *Accuracy* $=\frac{6}{8}\ \boldsymbol{x}\ \mathbf{100} = \mathbf{75\%}$		

Referring to Table 6, the percentage value is % *Accuracy* $=\frac{6}{8}\ x\ 100 = 75\%$.

#3) Classification with the MLP method

MLP classification is done by applying 10 input *neuron layers, 1 hidden layer* with 2 output *layers with Epoch* 500, *learning rate* 0.3, *correctly obtained* by 83.3% and *incorrectly* by 16.6%. The results of the identification of the *training data* with the MLP method, as shown in Figure 9.

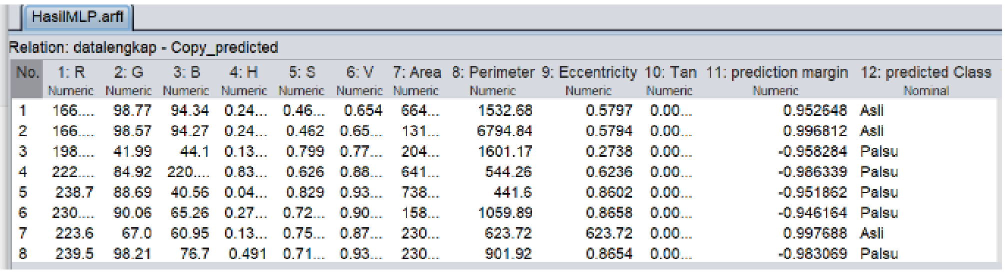

Classifier output

=== Summary ===

Correctly Classified Instances 15 83.3333 %
Incorrectly Classified Instances 3 16.6667 %

Figure 9. Identification results on *training data* using the MLP . method

Based on Figure 9, identification of the accuracy of the logo image with the help of the MLP method was carried out and it was successfully carried out, so that three original logos and five fake logos were obtained in the *testing data* using the MLP method, as shown in Figure 10.



HasilMLP.arff

Relation: datalengkap - Copy_predicted

No.	1: R Numeric	2: G Numeric	3: B Numeric	4: H Numeric	5: S Numeric	6: V Numeric	7: Area Numeric	8: Perimeter Numeric	9: Eccentricity Numeric	10: Tan Numeric	11: prediction margin Numeric	12: predicted Class Nominal
1	166....	98.77	94.34	0.24...	0.46...	0.654	664...	1532.68	0.5797	0.00...	0.952648	Asli
2	166....	98.57	94.27	0.24...	0.462	0.65...	131...	6794.84	0.5794	0.00...	0.996812	Asli
3	198....	41.99	44.1	0.13...	0.799	0.77...	204...	1601.17	0.2738	0.00...	-0.958284	Palsu
4	222....	84.92	220....	0.83...	0.626	0.88...	641...	544.26	0.6236	0.00...	-0.986339	Palsu
5	238.7	88.69	40.56	0.04...	0.829	0.93...	738...	441.6	0.8602	0.00...	-0.951862	Palsu
6	230....	90.06	65.26	0.27...	0.72...	0.90...	158...	1059.89	0.8658	0.00...	-0.946164	Palsu
7	223.6	67.0	60.95	0.13...	0.75...	0.87...	230...	623.72	623.72	0.00...	0.997688	Asli
8	239.5	98.21	76.7	0.491	0.71...	0.93...	230...	901.92	0.8654	0.00...	-0.983069	Palsu

Figure 10. Identification results on *testing data* using the MLP method

Based on Figure 10, it can be explained that the identification results were successful and there were three original *predicted classes* and five false predicted classes.

The results of measuring the accuracy value of *testing data* using the MLP method, like shown in Table 7.

Table 7. Test Results on MLP Classification

Class Logo	A	P
A= Original	3	1
P= False	0	4
$\% \ Accuracy = \frac{7}{8} \ x \ \mathbf{100 = 87,5\%}$		

Referring to Table 7, the *accuracy value is obtained* by $\% \ Accurasy = \frac{7}{8} \ x \ 100 = 87{,}5\%.$

3.5. Evaluation of Measurement Results

Based on the results of the classification with the three methods and two classes of logo images, an evaluation of the measurement results for the three classifications is obtained, namely (i) the kNN method with a predicted value of 94.4% on the *training data, while the measurement results on the testing data are* 75. %; (ii) the RF method with a predicted value of 100% on the *training data, while the value of the measurement results on the testing data is* 75%; and (iii) the MLP method with a predictive value on the *training data* of 83.3%, while the value of the measurement results on the *testing data is* 87.5%. The results of the comparison of the accuracy values on the *training* data and *testing data*, as shown in Figure 11.

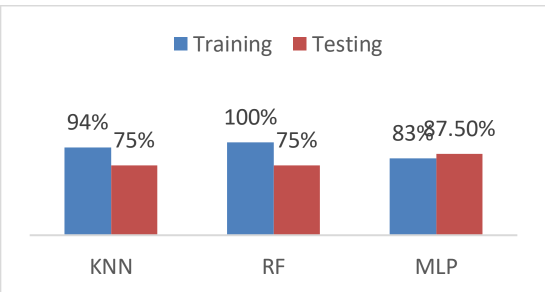


Figure 11. The results of the comparison of the accuracy values on the *training* and *testing data*

Based on Figure 11, it can be explained that the MLP method with a higher level of accuracy of measurement results, when compared to the other two methods, namely kNN and RF.

4. Conclusion

Based on the results and discussion, conclusions can be drawn according to the research objectives. Comparison of the three methods in *machine learning*, namely kNN, RF, and MLP to identify the achievement of *image matching* on the original and fake logo images. The performance of the MLP method with a higher level of accuracy of measurement results, compared to the kNN and RF methods, with the explanation that through the use *of 10-fold cross validation the accuracy* value is 87.5%, while the kNN method has an *accuracy value* of 75% and the *accuracy value* in the method RF is also at 75%. Based on these results, it can be proven that *machine learning methods* can be used for the identification process to achieve *image matching* of the original and fake logo images. Further research can be carried out using several other methods, in order to obtain a feature extraction process in quality and quantity for *image* matching.

References

[1] HD Wijaya and S. Dwiasnati, "Implementation of Data Mining with Naïve Bayes Algorithm in Drug Sales," *J. Inform.* , vol. 7, no. 1, pp. 1–7, Apr. 2020, doi:10.31311/ji.v7i1.6203.

[2] F. Sodik, B. Dwi, and I. Kharisudin, "Comparison of Supervised Learning Classification Methods on Data Bank Customers Using Python," *J. Mat.* , vol. 3, pp. 689–694, 2020, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/article/view/37875.

[3] P. Premaratne, "Feature Extraction," in *Cognitive Science and Technology* , vol. PartF1, 2014, pp. 75–103.

[4] Alda Putri Utami, Febryanti Sthevanie, and Kurniawan Nur Ramadhani, "Recognition of Vehicle Logo Using Local Binary Pattern and Random Forest Methods," *J. RESTI (System Engineering and Information Technology)* , vol. 5, no. 4, pp. 639–646, Aug. 2021, doi:10.29207/resti.v5i4.3085.

[5] Hariyanto, SA Sudiro, and S. Lukman, "Accuracy of Fingerprint Authenticity Detection Using the CNN Method," *Semin. Nas. technol. inf. and Commune. STI&K* , vol. 3, pp. 247–252, 2019, doi: http://dx.doi.org/10.32409/jikstik.3.1.270.

[6] R. Al Caruban, B. Sugiantoro, and Y. Prayudi, "Analysis of Object Match Detection in Digital Images Using the Shift Algorithm and Color Rgb Histogram Method," *Cyber Secur. and Digit Forensics.* , vol. 1, no. 1, pp. 20–27, 2018, doi:10.14421/csecurity.2018.1.1.1235.

[7] A. Lisdawati, P. Wibowo Yunanto, and Widodo, "Performance Testing of Backpropagation Neural Networks in Signature Recognition Models," vol. 2, no. 1, pp. 1–8, 2018, doi:10.21009/pinter.2.1.1.

[8] M. Irfan, BA Ardi Sumbodo, and I. Candradewi, "Vehicle Classification System Based on Digital Image Processing with

DOI: https://doi.org/10.29207/resti.v6i2.3826

Multilayer Perceptron Method," *IJEIS (Indonesian J. Electron. Instrum. Syst.* , vol. 7, no. 2, p. 139, Oct. 2017, doi: 10.22146/ijeis.18260.

[9] S. Januariyansah, "Analysis of Logo Design Based on Theory: Effective and Efficient," vol. 1, no. 1, pp. 13–14, 2018, doi: 10.13140/RG.2.2.20543.97448.

[10] A. Peryanto, A. Yudhana, and R. Umar, "Image Classification Using Convolutional Neural Network and K Fold Cross Validation," *J. Appl. Informatics Comput.* , vol. 4, no. 1, pp. 45–51, May 2020, doi: 10.30871/jaic.v4i1.2017.

[11] RSD Wijaya, Adiwijaya, Andriyan B Suksmono, and Tati LR Mengko, "Cervical Cancer Image Segmentation Using Markov Random Field and K-Means Algorithm," *J. RESTI* , vol. 5, no. 1, pp. 139–147, 2021, doi: 10.29207/resti.v5i1.2816.

[12] AW Kusuma and RL Ellyana, "Application of Compressed Imagery in Image Segmentation Using the K-Means Algorithm," *J. Terap. technol. inf.* , vol. 2, no. 1, pp. 65–74, 2018, doi:10.21460/jutei.2018.21.65.

[13] R. Rulaningtyas, AB Suksmono, TLR Mengko, and GA Putri Saptawati, "Color Image Segmentation Using Patch-Based Clustering Method for Identification of Mycobacterium Tuberculosis," *J. Biosains Postgraduate.* , vol. 17, no. 1, p. 19, 2015, doi:10.20473/jbp.v17i1.2015.19-25.

[14] N. Sivi Anisa and T. Herdian Andika, "Segmentation-Based Leaf Image Identification System Using K-Means Clustering Method," *Aisyah J. Informatics Electr. eng.* , vol. 2, no. 1, pp. 9–17, Feb. 2020, doi:10.30604/jti.v2i1.22.

[15] R. Nurfalah, Dwiza Riana, and Anton, "Identification of Rice Image Using Multi-SVM Algorithm and Neural Network in K-Means Segmentation," *J. RESTI (System Engineering and Information Technology)* , vol. 5, no. 1, pp. 55–62, 2021, doi:10.29207/resti.v5i1.2721.

[16] S. Saifullah, S. Sunardi, and A. Yudhana, "Comparison of Segmentation on Original Image and Wavelet Compression Image for Egg Identification," *Ilk. J. Ilm.* , vol. 8, no. 3, pp. 190–196, Dec. 2016, doi:10.33096/ilkom.v8i3.75.190-196.

[17] GA Pradipta and PD Wulaning Ayu, "Comparison of Image Segmentation of Chicken Eggs Using the Otsu Method Based on Differences in the Rgb and Hsv Color Spaces," *JST (Journal of Science and Technology* , vol. 6, no. 1, pp. 136–147, 2017, doi:10.23887/jst-undiksha.v6i1.9329.

[18] Oschina (2019), *Tangent Distance* . https://my.oschina.net/u/4397001/blog/3421827.(Accessed June 2, 2021)

[19] A. Salam, Sri Suryani Prasetiyowati, and Yuliant Sibaroni, "Prediction Vulnerability Level of Dengue Fever Using KNN and Random Forest," *J. RESTI (Engineering System and Teknol. Information)* , vol. 4, no. 3, pp. 531–536, 2020, doi:10.29207/resti.v4i3.1926.

[20] RR Waliyansyah and C. Fitriyah, "Comparison of the Accuracy of Teak Image Classification Using Naive Bayes and k-Nearest Neighbor (k-NN) methods," *J. Education and Researchers. information.* , vol. 5, no. 2, p. 157, 2019, doi: http://dx.doi.org/10.26418/jp.v5i2.32473.

[21] EF Saraswita, "Accuracy of Digital Scenes RGB Image Classification Using K-Nearest Neighbor and Naive Bayes Models," *Pros. annu. res. Semin.* , vol. 5, no. 1, pp. 978–979, 2019, [Online]. Available: https://seminar.ilkom.unsri.ac.id/index.php/ars/article/view/2131.

[22] GA Mursianto, M. Falih, M. Irfan, T. Sakinah, and D. Sandya, "Comparison of Random Forest and XGBoost Classification Methods and Implementation of SMOTE Technique in Rain Prediction Cases," no. September, pp. 41–50, 2021, doi: http://dx.doi.org/10.24014/ijaidm.v1i1.4903.

[23] A. Primajaya and BN Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indonesia. J. Artif. Intell. Min.* , vol. 1, no. 1, p. 27, Mar. 2018, doi:10.24014/ijaidm.v1i1.4903.

[24] A. Amrin and I. Satriadi, "Implementation of Artificial Neural Networks with Multilayer Perceptrons for Credit Analysis," *J. Ris. computer.* , vol. 5, no. 6, pp. 605–610, 2018, doi: http://dx.doi.org/10.30865/jurikom.v5i6.1006.

[25] Umbar Riyanto, "Comparative Analysis of the Naive Bayes Algorithm and Support Vector Machine in Classifying the Number of Online Article Readers," *IF (Jurnal Inform.* , vol. 2, no. 2, pp. 62–72, Oct. 2019, doi: 10.31000/.v2i2.1521.

[26] R. Umar, I. Riadi, and DA Faroek, "Comparison of Image Matching Using the k-Nearest Neighbor (kNN) Method and the Support Vector Machine (SVM) Method," vol. 4, no. 2, pp. 124–131, 2020, doi: https://doi.org/10.30871/jaic.v4i2.2226.

DOI: https://doi.org/10.29207/resti.v6i2.3826