Gorontalo Medicinal Plants Image Identification System Using Artificial Neural Network with Back Propagation

Mukhlisulfatih Latief[1], Rampi Yusuf[2]

Abstract—The purpose of this research is to design the application of digital image processing system to identify the image of medicinal plants of Gorontalo region using artificial neural network method using back propagation. This research used a digital image processing method with segmentation and extraction techniques. Segmentation process was carried out using thresholding method. Furthermore, a process of characteristic extraction from medicinal plants drawings was carried out using feature and color feature extractions to obtain the value of metric, eccentricity, hue, saturation and value. these five values were used as parameters for input neurons and one output neuron which denoted the class of the medicinal plants image. Data of this research consisted of 91 images which had been divided into two types, training data and test data. The training data consisted of 80 images and the test data consisted of eleven images. A network architecture was obtained from the training result and it provided the highest accuracy level (100%) and least number of iteration with a number of 50 neurons on hidden layer and 143 epochs. The testing result showed a lower accuracy of 54.54%.

Keywords— Artificial Neural Network, back propagation, image processing.

I. INTRODUCTION

Gorontalo is one of areas that has abundant medicinal plant, i.e. approximately 336 types of medicinal plants [1]. Gorontalo people often use these plants as traditional herbal medicine because they are cheap, easy to obtain and very effective.

Identification and recognition of a type of medicinal plants can be conducted based on its leaf pattern [2]. One of the often-faced obstacles is determining and recognizing the types of those medicinal plants due to the lack of public knowledge about medicinal plants types. Methods to recognize the medicinal plants are to directly observe the plants or by asking the local shaman. Those methods are very subjective so that the identification results are inconsistent.

Image processing field has been widely applied in plants species identification system, including tropical fruits plants identification [3], identification of tea types and quality [4], and medicinal plants identification [5].

Digital image processing to identify tropical fruit plants that have been developed was using the Gray Level Co-Occurrence Matrix (GLCM) feature texture extraction method. The identified plants types were star fruit, guava, mango, jackfruit, and naseberry. The identification algorithm

[1,2] *Lecturer, Informatics Engineering of Engineering Faculty State University of Gorontalo, Jendral Sudirman St. No. 6 (e-mail: mukhlis@ung.ac.id, rampirender1@gmail.com)*

used is the back propagation artificial neural network algorithm. The system yields the highest degree of accuracy of 90% [3].

In other studies, identification of the type and quality of tea plants has been developed using the segmentation thresholding method. Feature extraction is administered using color feature extraction based on the values of Red, Green, Blue, Hue, Saturation, and Intensity. The research uses Learning Vector Quantization (LVQ) artificial neural network as the identification algorithm. The results of the study showed an accuracy of 62.7% in the training process and 42.31% in the testing process [4].

The next study discusses the leaf pattern recognition applications design [5]. Application of leaf pattern recognition was designed using LVQ artificial neural networks for the medicinal plants determination. The designed application is able to properly recognize medicinal plants.

Referring to the previous studies, this research has developed a medicinal plants identification system using thresholding segmentation method, shape and color extractions as well as back propagation artificial neuron network identification algorithms.

This research is expected to minimize the level of misidentification due to the lack of public knowledge about regional medicinal plants, so that it can help the community in determining medicinal plants in a brief time.

II. IMAGE PROCESSING AND ARTIFICIAL NEURAL NETWORK

A. Image Segmentation

Image segmentation is the process aims to get the objects contained in the image or divide the image into several areas in which every object or area has similar attributes [6], [7].

Threshold is a basic technology and often used to conduct image segmentation, it is one of the simplest methods for obtaining sharp image segmentation [6]. Threshold produces binary image, i.e. the object pixel has the value of 1 and the background has the value of 0. Thus, the object will consistently appear brighter or darker than its background. The steps of image segmentation are as follows [8].

1. Read the original image.
2. Convert the image color space which was in RGB color space (Red, Green, Blue) into grayscale. The equation used to convert RGB image into a grayscale image is

$$Grayscale = 0.2989 * R + 0.5870 * G + 0.1140 * B \qquad (1)$$

3. Convert grayscale image to binary image through thresholding operations. Binary image resulted from thresholding operation is defined as

$$g(x,y) = \begin{cases} 1 \ if \ f(x,y) \geq T0 \\ \quad if \ f(x,y) < T \end{cases} \tag{2}$$

with $g(x,y)$ is binary images resulted from tresholding, $fx(,y)$ is grayscale image, and T is the threshold value.

4. Perform an operation complementary to binary image resulted from thresholding so that the object represented by pixel is worth 1 (white) and background represented by pixel is worth 0 (black).

B. Feature Extraction

After the objects in the image are separated from their backgrounds, then the feature extraction stage is performed. The extracted features are used to distinguish between one medicinal plants with others. Feature extraction is administered through shape and color analysis. Shape analysis is conducted based on metric and eccentricity values calculations, as shown in Fig. 1 and formulated in (3), while color analysis is conducted based on calculation of the values of hue, saturation, and value, as in Fig. 2 and formulated in (4). These values are extracted from each image. The equation used in this research is as follows.
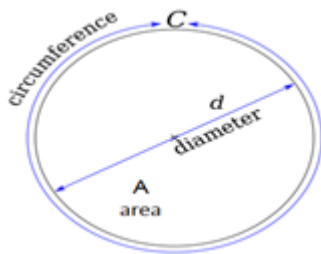

Fig. 1 Eccentricity value.

$$M = \frac{4\pi \times A}{C^2} \tag{3}$$

with

M = Metric
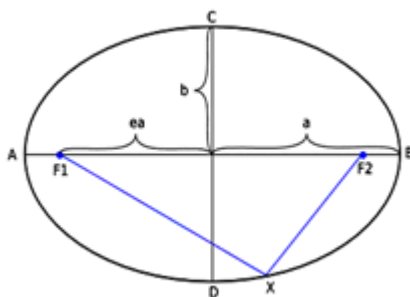A = Area
C = circumference.

Fig. 2 Hue, saturation, and value.

$$e = \sqrt{1 - \frac{b^2}{a^2}} \tag{4}$$

with

e = eccentricity
a = major axis

b = minor axis.

$$r = \frac{R}{(R+G+B)} \ , g = \frac{G}{(R+G+B)} \ , b = \frac{B}{(R+G+B)} \tag{5}$$

$$V = \max(r,g,b) \tag{6}$$

$$S = \begin{cases} 0, & if = 0 \\ 1 - \frac{\min(r,g,b)}{V}, & V > 0 \end{cases} \tag{7}$$

$$H = \begin{cases} 0, & if \ S = 0 \\ \frac{60*(g-b)}{S*V}, & if \ V = r \\ 60 * \left[2 + \frac{b-r}{S*V}\right], & if \ V = g \\ 60 * \left[4 + \frac{r-g}{S*V}\right], & if \ V = b \end{cases} \tag{8}$$

$$H = H + 360 \ \ if \ \ H < 0 \tag{9}$$

C. Artificial Neural Network (ANN)

ANN is a huge spread parallel processor that has a tendency to store experiential knowledge and make them ready to use. ANN resembles the human brain in two ways, i.e. the knowledge obtained by the network through learning process; and the strength of the relationship between nerve cells (neurons) known as synaptic weights which is used to store knowledge [9].

ANN is a computing mechanism able to acquire, represent, and calculate a mapping of a space with many other information variants, and was given a set of data that represents the mapping. Its purpose is to build a model from yielding data process, so that the networks can generalize and predict the outputs and inputs which have not been seen previously. ANN's learning is done by adjusting the weights between neurons in response to errors between actual output value and the output target value [10].
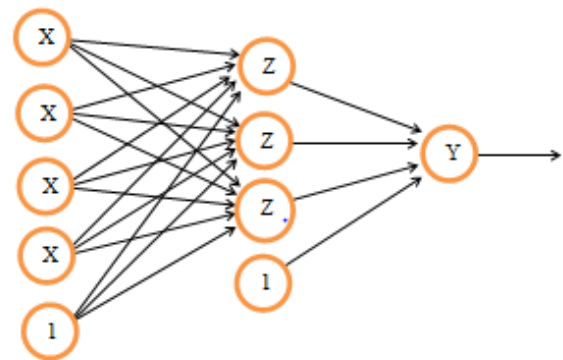

Fig. 3 Back propagation artificial neural network architecture with one hidden layer and one output [11].

D. Artificial Neural Network Learning Process

There are two learning systems in ANN learning process, i.e. learning with supervision and learning without supervision. Supervised learning is a learning process that requires teachers, i.e. something that possesses knowledge about environment through input and output samples

(a)   (b)   (c)   (d)
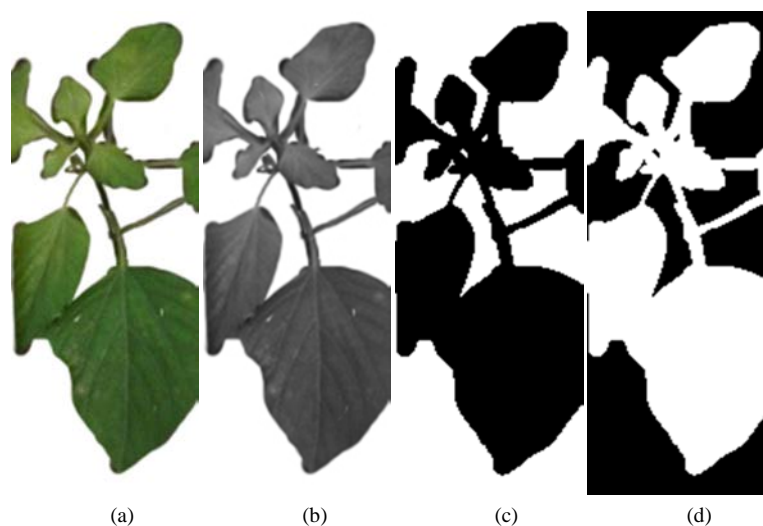
Fig. 4 Image segmentation process, (a) RGB image; (b) Grayscale image; (c) Thresholding result; (d) Complementary results.

TABLE I
THE RESULTS OF FEATURE EXTRACTION

No.	Plant Type	Feature				
		Metric	Eccentricity	Hue	Saturation	Value
1	Indian nettle	0.1754	0.9100	0.1502	0.2774	0.6826
2	Ficus septica	0.1673	0.5727	0.1435	0.1909	0.7900
3	Common bamboo	0.1208	0.6800	0.1516	0.2997	0.7029
4	Goat weed	0.5688	0.8946	0.1176	0.2522	0.7725
5	Begonia	0.5942	0.8053	0.0820	0.2299	0.7835

representation. The network parameter fluctuates based on training vector and error signals between ANN's outputs with the expected result. This repetitive process is carried out so that ANN can have capabilities similar to its teacher, or reference input-output data. It is clear that ANN is trained to be able to map a set of input-output samples with high accuracy. In an area classification, it can be said that landslide is based on obtained reference data, learning process uses supervised system [9].

Fig. 5 Architecture of back propagation artificial neural networks for medicinal plant.

Supervised learning process training has input data to be trained by the network in a form of input and target pairs, until the expected weight is obtained. That data pair serves as teacher in training process to train the network with the best form. At each time of training, the network generated from the input will process and produce output. The difference between the output with the expected output target is the occurring error. A network that can be used in l supervised learning, one of which, is by using back propagation [11].
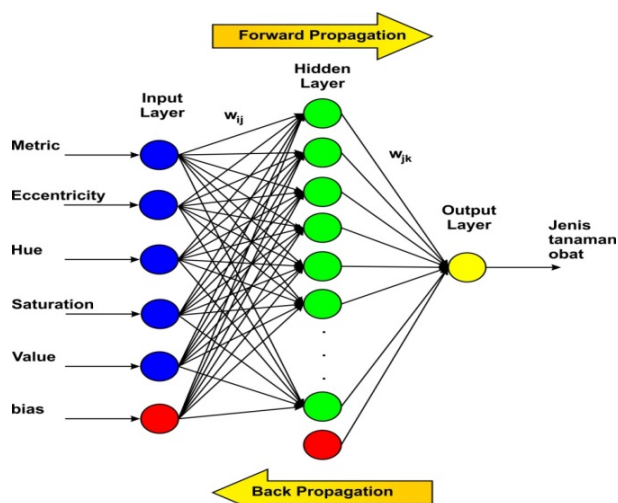
E. Back Propagation

This back-propagation training minimizes the output error which are targeted to be reached between the expected output and actuals values of the network. This training used a set of examples related to input and output values. The algorithm is a layered neural network that consists of input layer, hidden layer and output layer. Process occurring in those layers are to enter the input and to multiply it by the appropriate weight using the nonlinear transfer function. The goal of this training is achieved when neural network provides a model or reference that can predict new target value from the given input value [10].

Fig. 3 shows that the more hidden layers are used, complex problems from input and target training data can be more easily solved, compared to a single layer. The obtained output from training result has a low error value. However, the training process with many hidden layers provides a complex learning and sometimes takes a long time and a lot of
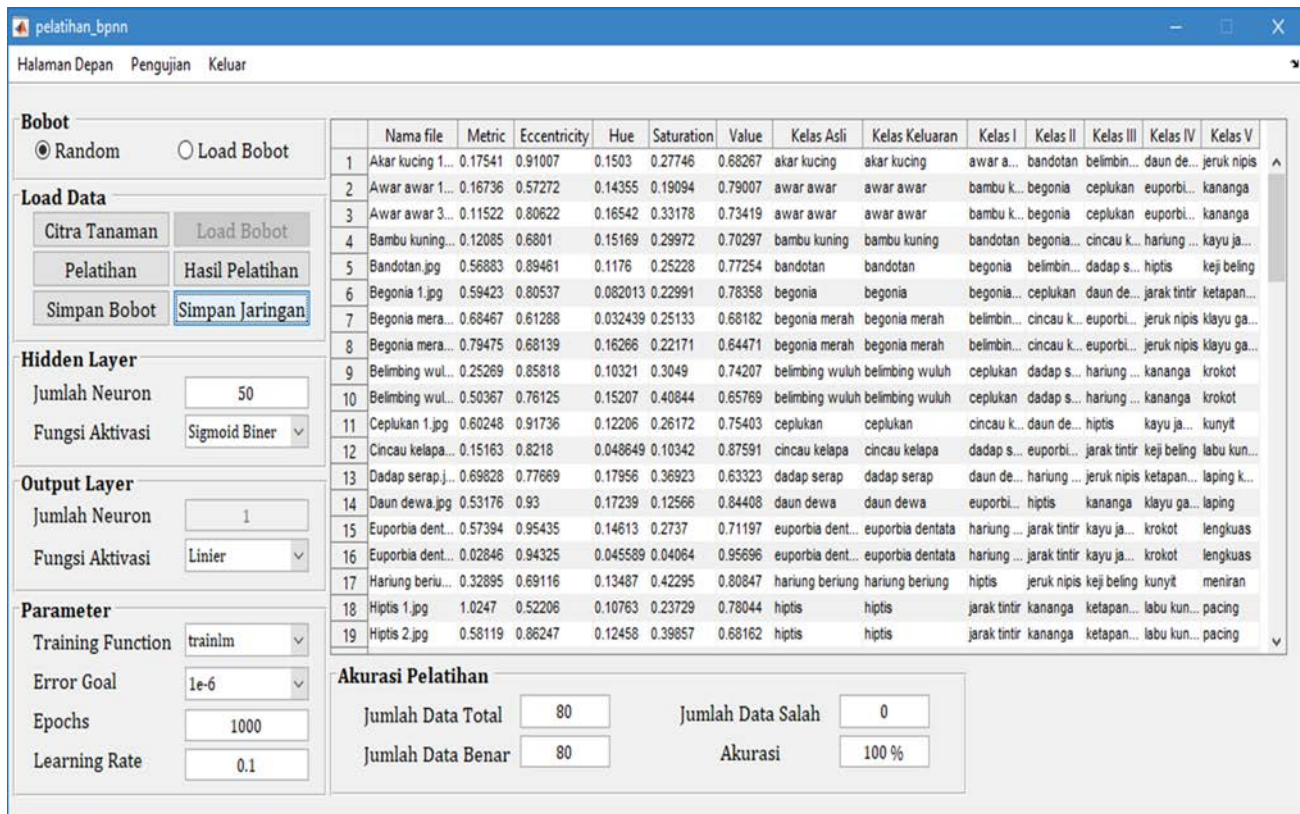
Fig. 6 Network training process.

TABLE II
RESULT OF NETWORK TRAINING WITH VARIOUS PARAMETER VARIATIONS

The number of neurons			The number of epoch	Accuracy (%)
Input layer	Hidden layer	Output layer		
5	10	1	290	23.75
5	20	1	1,000	32.5
5	30	1	1,000	86.25
5	40	1	186	100
5	50	1	143	100
5	60	1	148	100
5	70	1	184	100
5	80	1	587	100
5	90	1	242	100
5	100	1	370	100

iterations to obtain the desired output [11]. Therefore, this algorithm is selected, because it has high accuracy with accuracy value of 87.5% and 90% [12], [13].

III. METHODOLOGY

Materials used in this paper were 91 images of medicinal plants in Gorontalo area. The operated image processing stages were image acquisition, preprocessing using image segmentation, image extraction, and identification stages (training and testing).

This research used ANN algorithm with back propagation which was carried out to identify the medicinal plants images. The initial stage of this research was preprocessing on images using thresholding segmentation. After that, the shape feature and color feature extractions were carried out to obtain the value of metric, eccentricity, hue, saturation, and value.

After the values from preprocessing were obtained then training process using ANN was conducted by varying the number of neurons in the hidden layer to obtain the highest accuracy and the least number of iterations. After the architecture was obtained from the best accuracy and iteration, the architecture is then stored and used in the testing stage or medicinal plants image identification stage.

The training data used in this paper utilized 80 medicinal plants images. This training data was used to see the system's performance by observing the best accuracy and iteration
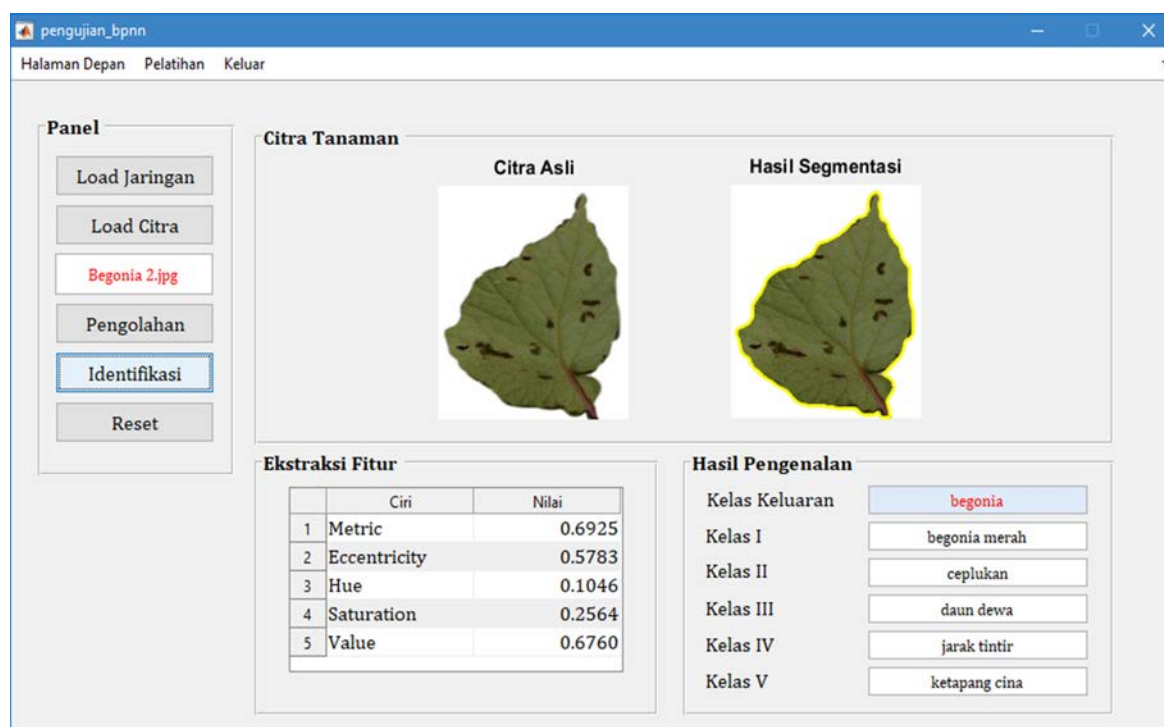
Fig. 7 Network testing process.

TABLE III
MEDICINAL PLANTS IDENTIFICATION RESULT IN THE TESTING PROCESS

No.	Original Class	Output Class	Class 1	Class 2	Class 3	Class 4	Class 5
1	Indian nettle	Ficus septica	Common bamboo	Begonia	Golden berry	Euphorbia dentata	Ylang-ylang
2	Ficus septica	Ficus septica	Common bamboo	Begonia	Golden berry	Euphorbia dentata	Ylang-ylang
3	Begonia	Begonia	Red Begonia	Golden berry	Gynura	Coral bush	Sea almond
4	Golden berry	Golden berry	Grass jelly	Gynura	Iron wort	Indian ash tree	Turmeric
5	Coral bush	Key Lime	Ylang-ylang	Strobilanthes	Verdolaga	Melanolepis	Asthma plant
6	Ylang-ylang	Indian ash tree	Strobilanthes	Garuga	Crookneck	Stonebreaker	Papaya
7	Garuga	Garuga	Verdolaga	Crookneck	Galangal	Asthma plant	Sambong
8	Crookneck pumpkin	Melanolepis multiglandulosa	Melanolepis multiglandulosa	Stonebreaker	Asthma plant	Commelina	Bilimbi
9	Purple bauhinia	Purple bauhinia	Lantana	Hummingbird tree	Ficus septica	Red Begonia	Gynura divaricata
10	Lantana	Curcuma	Hummingbird	Indian nettle	Goat weed	Golden berry	Hariung
11	Curcuma	Curcuma	Hummingbird	Indian nettle	Goat weed	Golden berry	Hariung

values. For the testing data, this research utilized eleven medicinal plants images which were used to calculate the accuracy level from medicinal plants images identification system.

IV. RESULTS AND DISCUSSION

A. Results

The result of image segmentation thresholding method was using MATLAB, the results are shown in Fig. 4. Afterwards, the shape feature extraction was administered to obtain metric and eccentricity values, while the values of hue, saturation, and value were obtained by administering color feature extractions. Some examples of medicinal plants images feature extraction are shown in Table I.

ANN training and testing processes using ANN algorithms were implemented to identify medicinal plants type based on five input values from feature extraction results. Identification process was performed using ANN algorithm with back propagation with architecture as shown in Fig. 5.

The utilized ANN consisted of three layers of one input layer, one hidden layer and one output layer. Training process utilized an activation function in the form of binary sigmoid at the hidden layer and the Levenberg-Marquardt training function.

After the values were obtained from the image extraction, the next stage was image training process. Training stage is useful to seek the parameters and the weights of each the best and the most suitable layer which later be used in the testing stage. The trained parameters are as follow:

a. learning = 0.10,
b. maximum iterations = 1,000,
c. error tolerance limit = 10^{-6},
d. activation function = binary sigmoid, and
e. number of trained data = 80.

ANN training process was carried out by varying the number of neurons in the hidden layer. The obtained results are shown in Table II.

It is shown in Table II that the architecture providing the highest accuracy and the least iteration is the architecture with 50 neurons in the hidden layer. The architecture was then used to propagate the input value towards the output value in the system testing process.

Furthermore, the tested ANN algorithm was then developed or implemented into the user interface display as in Fig. 6 (training process) and Fig. 7 (network testing process).

Table III presents the result of testing that utilized eleven testing images.

Table III shows that of the eleven tested images there are six images are correctly identified, therefore the accuracy resulted from the system in testing process is

$$
\begin{aligned}
(\%) \ & of \ accuracy \\
&= \frac{the \ number \ of \ correctly \ identified \ images}{The \ total \ number \ of \ tested \ images} \ x \ 100\% \\
&= \frac{6}{11} \times 100\% \\
&= 54.54\%.
\end{aligned}
$$

B. Discussion

Output results at the training stage shows a high accuracy, that is of 100%, but at the testing stage the accuracy value is decreased, the accuracy is of 54.54%. This low accuracy result in the testing stage might be as a cause of (1) the number of plants class was quite large, i.e. 61 classes; and (2) the number of training data of each plant class was very small, i.e. only two up to three plants samples. This accuracy value, compared with the previous studies, shows a high accuracy level. i.e. 90% and 80% [3], [14]. It happens as a result of limited number of classes used in this research, i.e. five classes and two different classes. In addition, the use of a lot of training data for each class has resulted the high accuracy value in the testing stage.

V. CONCLUSION

In this paper, the design of identification system of 91 medicinal plants image of Gorontalo area with image processing techniques has successfully been conducted. The stages of conducted image processing were image extraction, feature extraction and identification. Segmentation process was carried out using thresholding method, creating binary images which represents object with white color and background with black color. Feature extraction process was based on the shape and color analysis. Shape analysis was based on metric and eccentricity values, while color analysis was based on the value of hue, saturation, and value. Those five extracted features were used as input value in identification stage. Identification process was conducted using artificial neural network with back propagation. The best artificial neural network architecture used 50 neurons on hidden layers. Accuracy obtained by system reached 100% and testing accuracy in identifying medicinal plants reached 54.54%. It shows that the implemented image processing method can properly identify medicinal plants type.

REFERENCES

[1] N.Y. Kandowangko, M. Solang, and J. Ahmad, "Kajian Etnobotani Tanaman Obat oleh Masyarakat Kabupaten Bone Bolango Provinsi Gorontalo," Univ. Negeri Gorontalo, Research Report, 2011.

[2] G. Tjitrosoepomo, *Taksonomi Umum*, 3rd print, Yogyakarta, Indonesia: Gadjah Mada University Press, 2005.

[3] M.A. Agmalaro, A. Kustiyo, and A.R. Akbar, "Identifikasi Tanaman Buah Tropika Berdasarkan Tekstur Permukaan Daun Menggunakan Jaringan Syaraf Tiruan," *Jurnal Ilmu Komputer Agri-Informatika.*, Vol. 2, No.2, pp. 72-83, Nov. 2013.

[4] M. Effendi, Fitriyah, and U. Effendi, "Identifikasi Jenis dan Mutu Teh Menggunakan Pengolahan Citra Digital dengan Metode Jaringan Syaraf Tiruan," *Jurnal Teknotan*, Vol. 11, No. 2, Agustus 2017.

[5] F. Indrawan, "Aplikasi Pengenalan Pola Daun Menggunakan Jaringan Syaraf Learning Vector Quantification untuk Penentuan Tanaman Obat," *Proc. SemnasIF*, 2010, pp. 16-21.

[6] A. Kadir and A. Susanto, *Pengolahan Citra: Teori dan Aplikasi*, Yogyakarta, Indonesia: ANDI, 2012.

[7] T. Sutoyo, E. Mulyanto, V. Suhartono, O. Nurhayati, and Wijanarto, *Teori Pengolahan Citra Digital*, Yogyakarta, Indonesia: ANDI, 2009.

[8] E. Prasetyo, *Pengolahan Citra Digital dan Aplikasinya Menggunakan MATLAB*, Yogyakarta, Indonesia : ANDI, 2011.

[9] Suyanto, *Artificial Intelligence Searching, Reasoning, Planning, Learning,* Revisi Kedua, Bandung, Indonesia: Informatika, 2014.

[10] B. Pradhan, S. Lee, "Landslide Risk Analysis Using Artificial Neural Network Model Focussing on Different Training Sites," *International Journal of Physical Sciences*, Vol. 4, No. 1, pp. 001-015, Januari, 2009.

[11] J.J. Siang, *Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan MATLAB*, Yogyakarta, Indonesia: ANDI, 2005.

[12] K. Adi, S. Pujianto, R. Gernowo, A. Pamungkas, and A.B. Putranto, "Identification of Plasmadium Falciparum Phase in Red Blood Cells Using Artificial Neural Networks," *International Journal of Applied Engineering Research.*, Vol. 9, No. 23, pp. 13917-13924, 2014.

[13] K. Warman, L. Harahap, and A. Munir, "Identifikasi Kematangan Buah Jeruk dengan Teknik Jaringan Syaraf Tiruan," *Jurnal Rekayasa Pangan dan Pertanian*, Vol. 3, No. 2, pp. 248-253, 2015.

[14] M.M. Sebatubun and M.A. Nugroho, "Ekstraksi Fitur Circularity untuk Pengenalan Varietas Kopi Arabika," *Jurnal Teknologi Informasi dan Ilmu Komputer*, Vol. 4, No.4, pp. 283-289, Des. 2017.