

Classification of Tropical Rain Forest East Kalimantan Using Image Processing and Backpropagation Neural Network Algorithm

Masna Wati, Novianti Puspitasari, Herman Santoso Pakpahan, Hario Jati Setyadi, Trias Brata Yudhana Mahmuddin

Abstract—Tropical Rain Forest located in East Kalimantan has a high level of biodiversity, with a high level of biodiversity in east kalimantan then it needs a method to classify the existing plants there. In the research, the researchers tried to classify 5 plants found in tropical rainforests, namely ShoreaBalangeran, Dryobalanopsbeccarii Dyer, Eusideroxylonzwageri, Duriokutejensis, Cerberamanghas. Classification is done by using backpropagation neural network algorithm combined with image processing, where the image used is the image of plant leaf. The result of this research is the classification of 5 species of this plant with precision value above 90% in order to become a supporter of botanical decision in determining the type of plant and become alternative reference to classify plants in tropical rain forest area.

Keywords: Neural Network; Image Processing; Plants; East Kalimantan; Tropical Rain Forests.

I. INTRODUCTION

The tropical rain forests in East Kalimantan are heterogeneous forests, which means the plants are very diverse. Included among them are ShoreaBalangeran, Dryobalanopsbeccarii Dyer, Eusideroxylonzwageri, Duriokutejensi, Cerberamanghas. The high level of biodiversity in tropical rainforest requires a method to assist the work of botanists in classifying the plant with a precision value above 90%.

This research tries to approach by using image processing to extract information from leaf image of every plant then for classification of plants is done by using Backpropagation Neural Network algorithm. Image processing will process the image of the leaves of each plant that brings its own characteristics of each plant [1], [2]. This is because the leaves are one part that is quite unique and the shooting does not require special equipment, because naturally the leaves can be taken its image in 2 dimensional form without

removing the uniqueness brought from the plant. Unlike the other parts of plant, flowers and seeds that have a fairly complex and closed structure that requires equipment and special treatment.

This research is also supported by previous studies that have similarities as Patil, V.R. and R.R. Manza, they do research on the use of Basic Geometric Features and 12 Digital Morphological Features to recognize plants [3]. Then Stephen Wan Gu et al, also used the same technique of classifying plants from leaf images and combining them with the PNN algorithm [4]. Further research conducted by Neeraj Kumar et al, they try to approach by utilizing Neares Neighbor algorithm and combine it with Histograms of Curvature over Scale to extract the information from leaf image [5]. Stefan Fiel and Robert Sablatnig, conducted a study in which they used texture, contrast, correlation, energy, and homogeneity features extracted from the leather image of the poem, which was strongly influenced by moss, then the leaf nail image, and the image of regular leaves, and they also used the SVM algorithm in its classification, and using data derived from Austria [6]. In other literature MohamadFaizalAbJabal et al, conducted a comparative analysis of previous research where in their research they saw that there were some external factors or noise that could affect the final outcome despite the extracted features and algorithms used it has a resemblance [7].

II. EXPERIMENT DETAILS

This research investigates classifying plants that grow in tropical rain forest east Kalimantan, where there are 5 plants that will classified namely: ShoreaBalangeran, Dryobalanopsbeccarii Dyer, Eusideroxylonzwageri, Duriokutejensis, Cerberamanghas. Before doing image processing there are several steps that must be done and these steps are also known as preprocessing, where the leaves will be removed unneeded part so that only remaining part will be extracted that part of the leaf without the stem. The next step is to take the image by using the camera in a closed room and the camera used is Cannon 700D. The distance between the leaf and the camera is set around 40-50 cm by using a tripod or other camera rod, then using two lights that come from lights that are focused towards the leaf and using A4 size

Revised Version Manuscript Received on 10 September, 2019.

MasnaWati, Department of Information and Communication Technology, Informatics Engineering Study Program, Mulawarman University, Samarinda, Indonesia.(Email: masnawati.ssi@gmail.com)

NoviantiPuspitasari,Department of Information and Communication Technology, Informatics Engineering Study Program, Mulawarman University, Samarinda, Indonesia.

Herman SantosoPakpahan,Department of Information and Communication Technology, Informatics Engineering Study Program, Mulawarman University, Samarinda, Indonesia.

HarioJatiSetyadi,Department of Information and Communication Technology, Informatics Engineering Study Program, Mulawarman University, Samarinda, Indonesia.

TriasBrataYudhanaMahmuddin, Department of Information and Communication Technology, Informatics Engineering Study Program, Mulawarman University, Samarinda, Indonesia.

CLASSIFICATION OF TROPICAL RAIN FOREST EAST KALIMANTAN USING IMAGE PROCESSING AND BACKPROPAGATION NEURAL NETWORK ALGORITHM

paper as a base and as a measure of the extent of the image taken so that the concept and configuration image capture will be as in Figure 1.

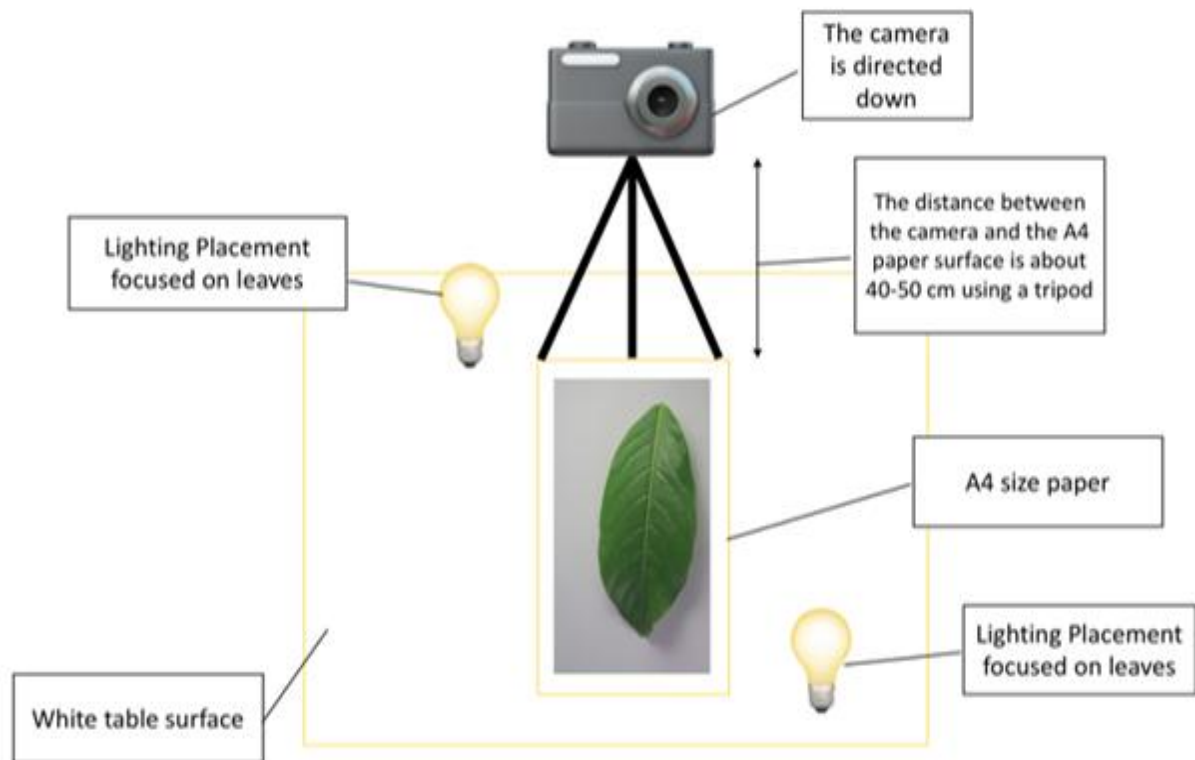


Fig. 1: The concept and configuration of the leaf capture site

After preprocessing and image retrieval, the next step is to extract the leaf features present on the leaf image, by converting the RGB-formatted image into Grayscale format this can be done using the function $\text{Grayscale} = 0.2999 \times R + 0.578 \times G + 0.11 \times B$. ext the image that has been converted into grayscale converted into binary image and given smoothing function with 3x3 kernel to eliminate the noise. Then to bring up the perimeter area of the leaf object from the binary image, it is done by giving the filter laplacian, then

given the morphology transformation function using the closed morphology process with the morphology element of radius 7x7 and rect-shaped, then given the find contours function to form a more neat edge line and thin. The next process, looking for a vein from the leaves by morphology transform with a closed process on the grayscale image of the leaves by using a round element with radius 1,2,3,4, where each radius will create its own image. After all image processing is done the results will be as in Figure 2.

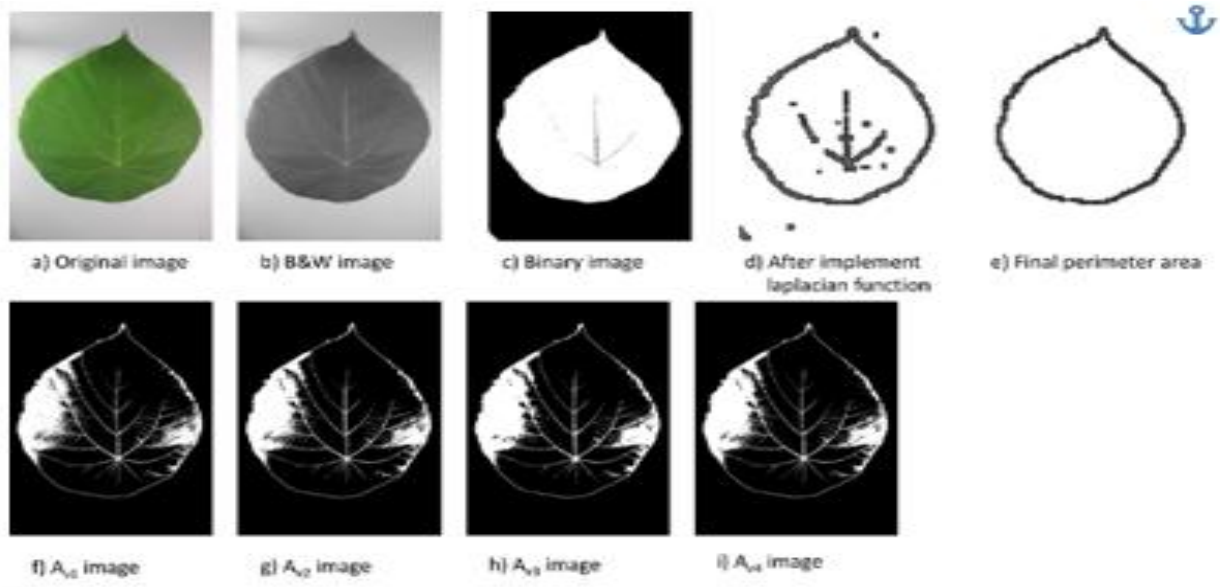


Fig. 2: The end result of image processing to be used in classification

After doing image processing, the next step is to extract the information from the image that has been processed. In this paper using five basic geometric features and Digital Morphological Features with a total number of features as many as 17 features. Each feature can be extracted in Figure 2.

Five basic Geometric Features which are: (a) Longest Diameter: obtained by finding the furthest distance from two pixels of each pixel of the edge line image and will be denoted as D; (b) Physiological Length: obtained by calculating the distance of the point entered by the user and will be denoted as W_p ; (c) Physiological Width: obtained by calculating the distance from two connected pixels and forming an orthogonal line against the Physiological length line with an angle of about $\pm 90^\circ$ and denoted as L_p ; (d) Leaf Area : is the number of white pixels found in binary imagery, and denoted by A; (e) Leaf Perimeter Area : is the number of white pixels found on the image of the outline of the leaf object and is denoted by P;

Digital Morphological Features which are: (a) Smooth factor : is a feature obtained by finding the number of white pixels from the binary image given the smooth function with the 5×5 kernel divided by the same function but the kernel used is 2×2 ; (b) Rectangularity : is a feature obtained by finding the results of physiological Width multiplication with

Physiological Length then divided by leaf area with the following calculation $L_p \cdot W_p / A$; (c) Aspect ratio : is the ratio obtained from the physiological length division by physiological width by the formula L_p / W_p ; (d) Parameter ratio of diameter : is the result of the calculation of the ratio of Leaf perimeter area to longest diameter, and can be formulated with D/P ; (e) Form factor : is the approach taken to determine the difference between a round shape and a leaf shape, to get its value can be done by calculation $4\pi A/P^2$, where A is the leaf area and P is the perimeter area; (f) Narrow factor : is the ratio between area diameter and physiological length, with the following formulation D/L_p ; (g) Parameter ratio of physiological length and physiological width : is a feature obtained by calculating the ratio between the perimeter area divided by physiological length plus physiological width, with the following calculations $P/(W_p + L_p)$; (h) Vein features : is a feature that is divided into 5 functions with each function is the result of the following calculations $f(x) = A_{v1}/A$; $f(x) = A_{v2}/A$; $f(x) = A_{v3}/A$; $f(x) = A_{v4}/A$; $f(x) = A_{v4}/A_{v1}$, where A_{v1} is the leaf area of the vein image with radius 1, A_{v2} is the leaf area of the venous image with radius 2, A_{v3} is the leaf area of the vein image with radius 3, and A_{v4} is the same leaf area of the vein image with radius 4.

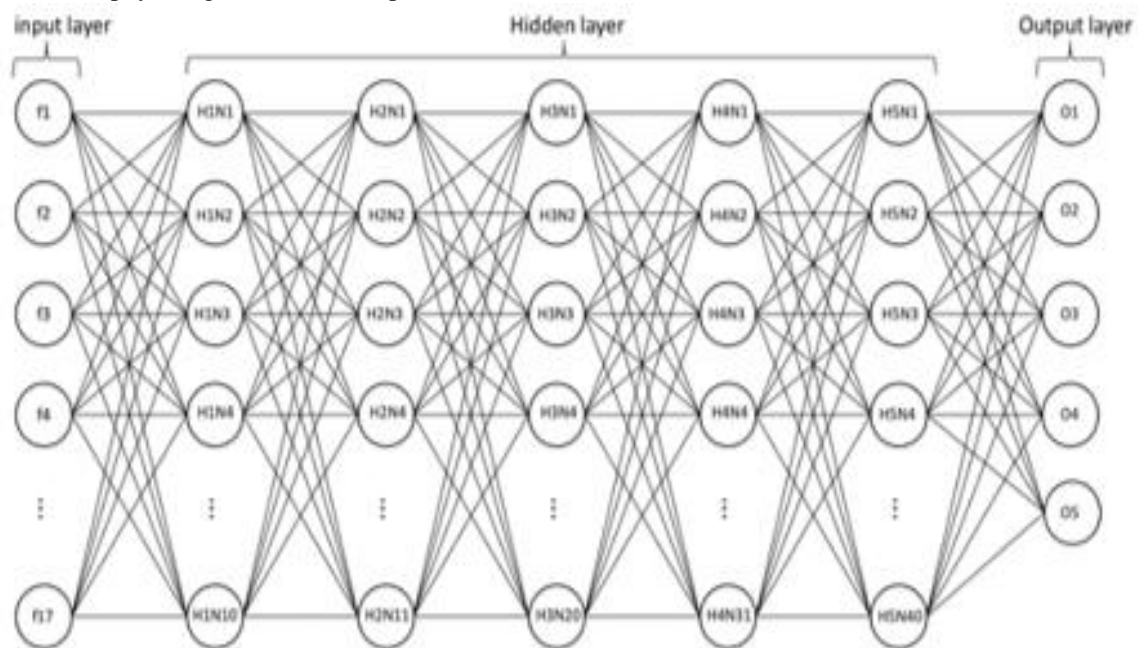


Fig. 3: BPNN algorithm architecture

After successfully obtaining the values of the extracted features of the leaf image, it will then normalize the data with pre-existing data. For the classification and training used backpropagation neural network algorithm, as it is known this algorithm is supervised learning which means that the need for training data so that this algorithm finds the pattern of each plant to be classified. Patterns to be classified are derived from information compiled from feature extraction results. To train this algorithm used 472 data training. In order to limit the course of this algorithm in the classification of the pattern then, this algorithm will do 1000 repetition iteration as well as its epoch as much as 1000. This algorithm initiated initial weight value using Xavier initialization

formula, and has 6 layers of layers that each node is all connected as Figure 3, with the following explanation :

- In the first hidden layer with input 17 and output 10 using hyperbolic tangent activation function;
- The second layer hidden has a total of 10 input parameters and has output 11 and uses the same activation function as the previous layer;
- The 3rd hidden layer has an input parameter number of 11 with output 20 and using the same activation;

CLASSIFICATION OF TROPICAL RAIN FOREST EAST KALIMANTAN USING IMAGE PROCESSING AND BACKPROPAGATION NEURAL NETWORK ALGORITHM

- The 4th Hidden Layer has 20 inputs and output 31 with the same activation function;
- The 5th layer hider has 31 inputs and has 40 outputs with hyperbolic tangent activation function;
- And for the last layer has 40 insert with output 5 with function used in this layer is softmax to determine result of classification.

Because the neural network used backpropagation concept then required a learning rate of 0.01 in order to decrease the value of error or improvement of temporary results in order to achieve predetermined results.

III. RESULT AND DISCUSSION

There are 150 test data in which each plant species has 30

Table 1: Confuse matrix

	ShoreaBalangeran (Kahoi)	Dryobalanopsb eccari Dyer (Kapur/Keladan)	Eusideroxylonz wageri (ulin)	Duriokutuj ensis (elai)	Cerberaman ghas (bintaro)	PR
ShoreaBalangeran (Kahoi)	30	0	0	0	0	100%
Dryobalanopsbeccari Dyer (Kapur/Keladan)	0	30	0	0	0	100%
Eusideroxylonz wageri (ulin)	0	0	30%	0	0	100%
Duriokutujensis (elai)	0	0	3	27	0	90%
Cerberamanghas (bintaro)	0	0	0	0	30	100%
RC	100%	100%	91%	100%	100%	
Precision						98%
Avg. Recall						98%
Avg. Recall						98.18%

test data, then from this test data obtained the results entered into the confusion matrix table to show the level of accuracy, recall and precision. Referring to Table 1 shows that plants in tropical rainforests can be identified with 98.00% (147 correct data and 3 incorrect data) accuracy where the ShoreaBalangeran, Dryobalanopsbeccarii Dyer, Eusideroxylonzwageri, and Cerberamanghas plants have 100% precision. From the results shown in Table 1 there are several factors that cause the precision of Duriokutejensis (90.00% with the correct classification data 27 and the incorrect data of classification 3) are not getting perfect value, this is because the vein of the leaves is not visible clearly and the size of the leaf that has a resemblance to the Eusideroxylonzwageri which also leaves its vein less clear.

IV. CONCLUSION

The conclusions obtained from this research is the use of image processing to obtain information or patterns from tropical rain forest plants can use Five Basic Geometric Features and Digital Morphological Features. For classification using backpropagation neural network algorithm has also been populated to be able to provide accurate information which obtained accuracy rate of 98% and every classified plant also has an average precision level above 90% which means that the decision given from this research is enough helps botanists in recognizing the plants to be classified.

V. ACKNOWLEDGEMENT

The authors would like to thank the Head of Dendrology Laboratory and Forest Ecology as well as his staffs for giving me assistance during the data collection during the writing of this journal.

REFERENCES

1. G. Kaur, S. Kaur, and A. Kaur, "Plant Disease Detection: a Review of Current Trends," *Int. J. Eng. Technol.*, vol. 7, no. 3.4, pp. 874–881, 2018.
2. T. Hemapriya, K. S. Archana, and T. Anupriya, "Comparison of preprocessing techniques for coin recognition using image processing methods," *Int. J. Eng. Technol.*, vol. 7, no. 2.21, pp. 351–354, 2018.
3. V. R. Patil and R. R. Manza, "A method of feature extraction from leaf architecture," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 5, no. 7, pp. 1025–1029, 2015.
4. S. G. Wu, F. S. Bao, E. Xu, Y. Wang, Y. Chang, and Q. Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network," *Int. Symp. Signal Process. Inf. Technol.*, pp. 11–16, 2007.
5. N. Kumar et al., "Leafsnap: A computer vision system for automatic plant species identification," in *European Conference on Computer Vision*, 2012, pp. 502–516.
6. S. Fiel and R. Sablatnig, "Automated identification of tree species from images of the bark, leaves and needles," in *16th Computer Vision Winter Workshop*, 2011.
7. M. F. AbJabal, S. Hamid, S., Shuib, and I. Ahmad, "Leaf features extraction and recognition approaches to classify plant," *J. Comput. Sci.*, vol. 9, no. 10, pp. 1295–1304, 2013.