The artificial neural network to predict chlorophyll content of cassava (Manihot esculenta) leaf

To cite this article: R Damayanti et al 2020 IOP Conf. Ser.: Earth Environ. Sci. 475 012012

View the article online for updates and enhancements.
The artificial neural network to predict chlorophyll content of cassava (*Manihot esculenta*) leaf

R Damayanti, Sandra and E Dahlena

Department of Agricultural Engineering, Faculty of Agricultural Technology, Universitas Brawijaya, Malang, Indonesia

E-mail: damayanti@ub.ac.id

**Abstract.** Artificial neural network (ANN) based prediction system was presented for predicting the leaf population chlorophyll content from cassava leaf images. As the training of this prediction system relied heavily on how well those leaf green pixels were separated from background noises in cassava leaf images, a global thresholding algorithm and an omnidirectional scan noise filtering coupled with the hue histogram statistic method were designed for leaf green pixel extraction. With the obtained of leaf green pixels, the system training was carried out by applying a back-propagation algorithm. The system was tested to predict the chlorophyll content from the cassava leaf images. The purpose of this research was to find the relationship between the color index Red, Green, Blue (RGB); Hue; Saturation HSV; Value; Saturation HSL and Lightness to chlorophyll, and to find the appropriate form of ANN to predict the highest chlorophyll source in cassava leaf based on digital images. The results showed highest positive regression occurred in the saturation HSL index against the total chlorophyll of cassava leaf as much as 78.6%. The best model produced using ANN methods in predicting total chlorophyll is a network model with 8 inputs, 9 hidden layers and one output layer, in the proportion of training data 75% testing data 25% have value result the smallest MSE testing is 0.092 with regression testing of 0.847. Network model can read the highest source of chlorophyll on cassava leaf with the value of 84.68%.

1. **Introduction**

Cassava leaf is green and often eaten as vegetable because it has a good source of protein. Cassava leaf is also effective in treating rheumatism, helps to restore bones and skin, and prevents the aging process. Green vegetables contain a source of green pigments which are often called chlorophyll. Chlorophyll pigments are very important for health, able as natural cleansers function that encourage detoxification of the body, antioxidants (reduce carcinogens), and antiaging. Chlorophyll is a measure of the amount of antioxidants contained in vegetables and able to be a substitute for the body's dietary supplements. Chlorophyll is a vital pigment primarily responsible for harvesting light energy used in photosynthesis and is therefore an excellent indicator of a crops overall physiological status stressor disease [1, 2], and yield predictions [3, 4]. Chlorophyll can potentially pro-vide an assessment of leaf nitrogen, an essential plant nutrient, due to the close relationship between leaf chlorophyll and leaf nitrogen [5, 6].

Various invasive and noninvasive methods have been proposed earlier to measure the leaf chlorophyll. The conventional methods of measuring chlorophyll such as spectrophotometry that extract pigments from leaf tissue will cause damage to the sample, require a long time, and can not be used in further measurements [7]. Research on Artificial Neural Networks (ANN) and digital images are widely...
implemented in cases that require more accurate and objective decision making, with Back-propagation Neural Network (BPN) training were applied. Digital image processing is related to improving image quality (increasing contrast, transformation, color, image restoration), image transformation (rotation, transformation, geo-mechanical transformation), selecting optimal feature images for analytical purposes, making information withdrawal or object description or imposition of objects contained in the image, perform compression or reduction of data for the purpose of data storage, data transmission, and data processing time. Input from image processing is an image, while the output is image processing results.

There are 3 types of images commonly used in digital image processing, i.e. color image, grayscale image and binary image. Color images are often known as RGB (red, green, blue) [8]. Each component uses 8 bits of data (its value ranges from 0 to 255), so the color that can be presented by one RGB image is 16,580,375 colors. This format is called true color because the number of the colors is large enough, enabling to cover almost all natural colors. Digital images that have been obtained will later be modeled into several color models. The color model is a coordinate system that can map all colors in a system as a point. To make decisions with high accuracy from digital image data, ANN is used. ANN consisted of 3 groups of layers, namely the input layer, the hidden layer and the output layer.

The purpose of this research was to identify the relationship between the color index Red, Green, Blue (RGB); Hue; Saturation HSV; Value; Saturation HSL and Lightness to chlorophyll, and to determine the appropriate form of ANN to predict the highest chlorophyll source in cassava leaf based on digital images.

2. Materials and Method
2.1. Materials
Cassava leaves were picked from tree aged of 25 to 40 days, and third order from the top. Then, cassava leaves were separated from the branches and used 25 for sample. Alcohol 96% was used as a solvent in the analysis of chlorophyll.

The equipment for digital image processing were Nikon Coolpix A10 16.1 Mp digital camera used as shooting media, acquisition image box measuring 50 x 40 x 40 cm as image catcher, 4 pieces of light bulb Foccus power 5 Watts was used as a light source for shooting.

2.2. Data interpretation method
2.2.1. Chlorophyll content in cassava leaf analysis
Analysis of total chlorophyll, chlorophyll a and chlorophyll b levels on vegetable leaves was carried out at the Chemistry Laboratory of Universitas Islam Negeri Maulana Malik Ibrahim Malang, using destructive methods and UV-vis spectrophotometers based on procedures [9] with Optical Density (OD) at wavelengths of 645 nm and 665 nm measurements.

2.2.2. Image processing of cassava leaf method
Cassava leaves were placed into acquisition image box with a white background and 4 lights were installed as light sources and the camera was placed on top of the leaf object. Leaves were taken only once for each sample. Thus, there were 25 image data, which was used as input data, training and validation in programming languages. The image taken from this image was 4608 x 3456 pixels, so it was necessary to reduce the size to 300 x 400 pixels with the bmp extension to make the object easily recognized by MATLAB Software. Before that, the image must be pre-processing by removing the background object with Adobe Photoshop Software.

2.3. Data processing method
The image of cassava leaf was analysed by digital image method, to extract RGB, HSL and HSV color spaces. This color extraction was done by Features Extraction with Thresholding-BG White, which was processed using Visual Basic 6.0 software [9].
Image extraction results consisted of 8 parameters, i.e. Red, Green, Blue, Hue, Saturation HSL, Lightness, Saturation HSV and Value. All parameters were correlated with factual data of total chlorophyll from Lab analysis. This method was applied to find out and predict the relationship between chlorophyll content and 8 parameters of image extraction. Process analysis were based on Determination Coefficient ($R^2$) and Correlation Coefficient ($R$). The Determination Coefficient was used to measure the suitability of linear regression lines between the image results with total chlorophyll, a and b. While the Correlation Coefficient to measure the closeness of the relationship between the variables interpreted. Data from image processing was used in an objective and accurate decision making on the ANN method with 8 input data of Red, Green, Blue, Hue, Saturation HSL, Lightness, Saturation HSV, and Value. Data on chlorophyll content of cassava leaf was the target (output) of ANN.

3. Results and Discussion

The results of chlorophyll analysis showed that chlorophyll content of cassava leaves for the maximum value was 28.64 mg/g (chlorophyll a) and 14.94 mg/g (chlorophyll b) with total chlorophyll of 43.79 mg/g, moreover the minimum value was 20.13 mg/g (chlorophyll a) and 6.59 mg/g (chlorophyll b) with total chlorophyll of 26.87 mg/g. Based on earlier research, the total chlorophyll content of cassava leaf showed a high number, i.e. 27.447 mg/g compared to spinach leaf (23.022 mg/g) and kale leaf (16.767 mg/g). Since the cassava leaf ability to absorb sunlight radiation was more efficient, therefore the photosynthesis rate was higher [10].

3.1. Correlation of total chlorophyll, a and b to RGB colors

Figure 1 shows that the greater of total chlorophyll content, the greater the red mean value follows the equation. The determination coefficient obtained is 0.094. This figure includes a low category correlation where the total chlorophyll content is not able to give a significant effect on the red mean. Read mean color on grape leaf can well represent chlorophyll at high trends, until $R^2$ reaches a value of 0.952 [11]. The higher value of the green color (green), the lower the total chlorophyll content in cassava leaves, with a relatively low $R^2$ is 0.085. The previous study on betel leaf found that the green mean was able to influence the total chlorophyll content until the $R^2$ value reached 0.80 in the decreasing linearity pattern [12]. While the study on grape leaf found that the green mean color could represent chlorophyll well in the rising linearity pattern, until the value of $R^2$ reached 0.953 [11].

![Figure 1. Correlation of total chlorophyll to: a) read mean, and b) green mean](image)
chlorophyll content [13]. Figure 2c shows that the higher the value of saturation HSV means the higher the total chlorophyll content of cassava leaves, with $R^2$ of 0.426. In a study of mango leaf, obtained $R^2$ value of 0.9881 was linearly positive for Saturation HSV Mean. Figure 2d shows that the trend of the resulting pattern is positive linear, in which an increase in the value of the saturation HSL mean resulted in increasing the total chlorophyll content of cassava leaf leaves with $R^2$ value of 0.786. In a study on betel leaf, $R^2$ value for saturation of HSL obtained was about 0.52 positive linear. According to Figure 2c, the trend of positive linear patterns, the greater of chlorophyll content of cassava leaf, the greater of lightness value. Chlorophyll content is not able to influence the lightness mean significantly, this can be seen from the low $R^2$ is 0.029. Study on betel leaf, resulting negative linear trend with $R^2$ 0.87 for the value of lightness [12]. Figure 2f shows a positive linear trend in cassava leaf, chlorophyll content remains unable to influence the mean value. This can be seen from the low $R^2$ value of 0.042. The study on mango leaf obtained a negative linear trend for the mean value, with $R^2$ of 0.915 [13].

![Figure 2c](image1)

![Figure 2d](image2)

![Figure 2e](image3)

![Figure 2f](image4)

**Figure 2.** Correlation of total chlorophyll to color analysis
In general, the overall $R^2$ graph gives a low value trend compared to studies on grape leaf [11], betel leaf [12] and mango leaf [13]. This is due to the use of the UV-vis spectrophotometer method with a wavelength of 645 nm and 665 nm was only able to represent actual chlorophyll a and b content without regard to the color concentrations contained in the leaves. Consequently, the digital image results are less accurate in reading information on chlorophyll a and b. Whereas in the study of grape and betel leaf, atLEAF meters were used which transmit red (660 nm) and infrared (940 nm) light on green leaves so that chlorophyll is easily absorbed, blue mean (blue light) has a wavelength of 450 nm to 495 nm according to the wavelength of blue color of green leaves. In the study of mango leaf, SPAD 502-plus chlorophyll meter used as a provider of chlorophyll values of qualitatively and quantitatively and can estimate of digital images well. Based on low linearity level and the difficulty of digital images reading chlorophyll a and b with laboratory analysis, the RGB, HSV and HSL digital images can be used as a basis for modeling ANN in predicting total chlorophyll on cassava leaf.

3.2. Sensitivity analysis of structures of ANN on chlorophyll
Selection of ANN architecture is done by trial error to find the number of hidden layers and neurons that match the existing problem [14]. The parameters of the pattern forming network system are limited to the learning rate of 0.9; momentum is 0.9 and the maximum iteration is 53000, with an error tolerance of 0.0001. The input data is divided into several dataset proportions so that convergence can occur faster. Until now there is no mathematical equation that can determine exactly the division of datasets on artificial neural networks, some researchers argue that the 80% dataset is able to achieve convergence values, so does the 75% dataset. In this study the dataset is divided into three propositions, namely 75% for training and 25% for testing; 80% training data and 20% testing data; and 60% training data and 40% testing data. This study limits the network pattern by using 8 input neurons in the input layer, 1 output neuron at the output layer with 5, 7, 9 and 11 neurons in the hidden layer 1. Sensitivity analysis of network learning in MATLAB 7.1.

<table>
<thead>
<tr>
<th>Dataset proportion</th>
<th>Network Model</th>
<th>Training %</th>
<th>Testing %</th>
<th>MSE training</th>
<th>MSE testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%</td>
<td>8-5-1</td>
<td>0.922</td>
<td>0.828</td>
<td>0.043</td>
<td>0.367</td>
</tr>
<tr>
<td></td>
<td>8-7-1</td>
<td>0.844</td>
<td>0.670</td>
<td>0.084</td>
<td>0.352</td>
</tr>
<tr>
<td></td>
<td>8-9-1</td>
<td>0.940</td>
<td>0.289</td>
<td>0.042</td>
<td>2.233</td>
</tr>
<tr>
<td></td>
<td>8-11-1</td>
<td>0.813</td>
<td>0.075</td>
<td>0.099</td>
<td>3.542</td>
</tr>
<tr>
<td>75%</td>
<td>8-5-1</td>
<td>0.953</td>
<td>0.818</td>
<td>0.058</td>
<td>0.268</td>
</tr>
<tr>
<td></td>
<td>8-7-1</td>
<td>0.851</td>
<td>0.768</td>
<td>0.089</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>8-9-1</td>
<td>0.870</td>
<td>0.847</td>
<td>0.075</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>8-11-1</td>
<td>0.860</td>
<td>0.075</td>
<td>0.080</td>
<td>0.872</td>
</tr>
<tr>
<td>60%</td>
<td>8-5-1</td>
<td>0.859</td>
<td>0.463</td>
<td>0.082</td>
<td>1.273</td>
</tr>
<tr>
<td></td>
<td>8-7-1</td>
<td>0.854</td>
<td>0.860</td>
<td>0.084</td>
<td>0.796</td>
</tr>
<tr>
<td></td>
<td>8-9-1</td>
<td>0.867</td>
<td>0.604</td>
<td>0.077</td>
<td>0.461</td>
</tr>
<tr>
<td></td>
<td>8-11-1</td>
<td>0.872</td>
<td>0.649</td>
<td>0.074</td>
<td>0.275</td>
</tr>
</tbody>
</table>

According to Table 1, the optimal results of BPN training are shown by network model 8-9-1 in the 75% data proportion with the learning function logsig-purelin-traind, which means the input activation function used is binary sigmoid, the output function is a linear function (identity), with the training function is traind. This network model is reinforced by Haykin [15], that the number of two- to nine-layer hidden neurons has been able to produce the best network in ANN method. The pattern training uses 54 training data, while the testing uses 18 data. Learning is stopped at a maximum iteration of 53000 iterations, with MSE 0.075 in the correlation coefficient ($R$) of 0.870. The $R$ number shows that
training test model has a very strong closeness. Furthermore, testing is carried out ANN which aims to find out how much new test data can be identified by the learning model, until finally the value of R testing is obtained which is quite high at 0.8468. It means that at a percentage of 84.68%, the network model 8-9-1 can predict the total chlorophyll to the maximum.

![Figure 3. The best training performance](image)

Figure 3 shows the best training graph from network modeling 8-9-1. It appears that the graph connects two variables, MSE (Mean Squared Error) and the number of iterations given. The training was stopped at 727 iteration with MSE 0.075164 from 53000 iterations performed. Although the given iteration is quite large, learning will be stopped when it reaches maximum converging at any iterations. As a result, learning displays linear lines along the graph. MSE has reached its maximum point at 727 iterations. The training has shown maximum convergence even though it is not good enough. A good training is reached if MSE decrease along with iteration given, so training graph will show incisively descend.

### 3.3. Chlorophyll analysis of total prediction

The test results show a comparison between the predicted MATLAB data and the actual total chlorophyll data, as shown in Table 2.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Sample</th>
<th>Chlorophyll total (actual)</th>
<th>Chlorophyll total (prediction)</th>
<th>Difference in error</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S1</td>
<td>34.493</td>
<td>38.456</td>
<td>-3.963</td>
<td>Inaccurate</td>
</tr>
<tr>
<td>2</td>
<td>S2</td>
<td>43.792</td>
<td>35.974</td>
<td>7.8179</td>
<td>Inaccurate</td>
</tr>
<tr>
<td>3</td>
<td>S3</td>
<td>32.859</td>
<td>34.024</td>
<td>-1.164</td>
<td>Accurate</td>
</tr>
<tr>
<td>4</td>
<td>S4</td>
<td>39.141</td>
<td>38.058</td>
<td>1.083</td>
<td>Accurate</td>
</tr>
<tr>
<td>5</td>
<td>S5</td>
<td>38.999</td>
<td>37.238</td>
<td>1.762</td>
<td>Accurate</td>
</tr>
<tr>
<td>6</td>
<td>S7</td>
<td>37.043</td>
<td>36.473</td>
<td>0.570</td>
<td>Accurate</td>
</tr>
</tbody>
</table>

Table 2 shows when 6 predictive data compared to the actual data, there are 4 data that are correctly recognized by the network model 8-9-1 as the highest source of chlorophyll. Inaccurate prediction results occur because the error difference is not close to the desired fault tolerance number between 1 and -1, so the new data entered for the test simulation has not been able to be studied (not generalized well) by network model 8-9-1. Nevertheless, this condition shows that this network is quite good in studying patterns based on the MSE produced. The six data have differences in error that are almost
close to the numbers 1 and 1 at the smallest error MSE testing 0.092. Therefore, the data having differences close to the smallest error number can be assumed that training is able to work well.

4. Conclusions
The correlation of 4 digital image parameters, i.e. Red, saturation HSV, lightness and saturation HSL to total chlorophyll, a and b have positive linearity patterns. While Green, blue, hue, and value have negative linearity patterns. The highest regression is shown in the saturation HSL index of the total chlorophyll of cassava leaf by 78.6%. The best ANN model in predicting total chlorophyll is the 8-input model, 9 hidden layers with one output layer, in the proportion of 75% training data and 25% testing data. The activation function used in this condition is the binary sigmoid function in the input and identity functions in the output layer. Training is used traingd (Gradient Descent Backpropagation), obtained the smallest MSE testing value of 0.092 and R testing of 0.8468. The network model was able to maximally read the highest source of total chlorophyll in cassava leaf with value of 84.68%

References