Vernonia Amygdalina Chlorophyll Content Prediction by Feature Texture Analysis of Leaf Color

R Damayanti¹, D F A Riza¹, A W Putranto¹, R J Nainggolan¹

¹Department of Agricultural Engineering, Faculty of Agriculture Technology, Universitas Brawijaya, Malang, Indonesia

Corresponding author’s email address: damayanti@ub.ac.id

Abstract. Vernonia amygdalina has been scientifically proven to have activity against various diseases because it contains high antioxidants. The antioxidant content can be related to the chlorophyll content in leaves. Chlorophyll levels increase when the leaves are fully developed which is accompanied by an increase in antioxidants. So, chlorophyll detection by non-invasive sensing can be used to estimate the antioxidant content. An artificial neural network (ANN) was used to model RGB color as input and leaf chlorophyll content as output. Performance comparisons in each ANN model were carried out to find the best model in predicting leaf chlorophyll content, indicated by the smallest prediction error value. This study aims to model the chlorophyll content of Vernonia amygdalina with ANN analysis. The results showed that the chlorophyll content could be identified using 9 selected color texture features through the filter method feature selection with the best attribute of correlation. The selected ANN structure produces R training of 0.98522, R validation of 0.93417, MSE training of 0.0067, and MSE of validation of 0.0322. The results showed that digital image processing and ANN models have the potential as sensors in detecting the percentage of chlorophyll content of Vernonia amygdalina.

Keywords: Antioxidants; ANN Analysis; Chlorophyll Content; Model RGB Color

1. Introduction

Vernonia amygdalina has been scientifically proven to have activity against various diseases because it contains high antioxidants. One of the antioxidant compounds that have been shown to have a very strong biological effect on Vernonia amygdalina leaves are flavonoids [1]. Luteolin, Luteolin-7-O-glucuronoside, Luteolin-7-O-β-glucosides which are from the flavonoid class can be used as antioxidants [2]. Some of the benefits of this leaf in the medical field i.e. to cure dysentery [3], gastrointestinal disorders, fever, malaria, hepatitis, scabies, headaches, stomach aches, and coughs. This herb has also been scientifically proven to have activity against various diseases i.e. anti-inflammatory [2], antimicrobial [4], antioxidant [5], and anti-allergic [6]. The antioxidant effect of this compound is due to the scavenging of free radicals through hydrogen atom donors from the flavonoid hydroxyl groups [7].

Chlorophyll contained in many leaves has the ability as an anti-oxidant, anti-inflammatory, and substances that heal wounds [8]. Chlorophyll levels will increase with age until the leaves are fully developed and then chlorophyll levels decrease as the leaves get older. When the leaves are old, it indicates that there are other compounds that act as the main barrier to oxidation reactions, namely flavonoids [9]. The flavonoid content is still low in young leaves, then it increases with the aging of the leaves, where photosynthesis occurs optimally [10]. Detection of chlorophyll content in the leaves...
of *Vernonia amygdalina* by non-invasive sensing is very important to measure the effectiveness of the plant photosynthesis process of *Vernonia amygdalina* in real-time without having to damage the leaves. Real-time leaf chlorophyll detection can also be used to estimate the antioxidant content so that the readiness of the leaves of the *Vernonia amygdalina* plant can be determined for harvest.

Chlorophyll analysis methods are usually carried out in the laboratory using a spectrophotometric instrument. Laboratory analysis provides more accurate results than human sensory analysis, but laboratory analysis is destructive in nature or sample destruction so it takes a long time and is a high cost [11-13]. One of the non-invasive sensing methods that practical, cheap, effective, and accurate is computer vision [14].

Computer vision is closely related to image processing, which analyzes images through Artificial Intelligence (AI) [15]. AI analyzes more and deeper data using a neural network that has many hidden layers. The method used in the AI application in this study is the Artificial Neural Network (ANN) method [16]. The relationship between color change and chlorophyll content is used as the basic data for building a software system that is processed in learning using ANN.

The purpose of this study was to predict the chlorophyll content of *Vernonia amygdalina* leaves using computer vision. ANN is used to model RGB as input and leaf chlorophyll content as output. Performance comparisons in each ANN model were carried out to find the best model in predicting leaf chlorophyll content, indicated by the smallest prediction error value.

2. Methods

2.1. Material

The main material used in this research is *Vernonia amygdalina* leaves, taken from a plantation in The Laboratory of Agro-Industry Mechatronics Tools and Machinery, Universitas Brawijaya, Malang, East Java. Leaf images were taken in an acquisition box equipped with 4 white RoHS LED lights 0.75 watts, 12 volts, 20mA. The tool used for leaf image acquisition is the Canon Kiss x7i camera, which is connected to a laptop. The software used is the Windows 10 64-bit operating system, with color-based visual basic 6.0 software and texture analysis for the feature extraction process, equipped with an application for the Feature Selection process WEKA 3.8., and ANN modeling software using Matlab R2014a. The comparison tool for measuring the chlorophyll content of *Vernona amygdalina* leaves was chlorophyll-meter Konica Minolta SPAD 502.

2.2. Method

Leaf samples were taken from several different plants. To distinguish leaf color, leaf samples were taken from three levels of development i.e. young leaves at the top, leaves that were still developing in the middle, and mature leaves at the bottom. The leaves included in the sample in this study were taken from 3-4 leaves from the top and bottom, while the leaves at the development stage or the middle part were taken from the leaves between the top and bottom of the plant. The leaves were taken twice by plucking. In one take, there were 30 leaves plucked, every 10 leaves from each stage of development. The total leaves prepared for data collection were 60 leaves. From the collected data sample, the samples were divided into two sets i.e. one set for training and one set for testing (validation). The training set was used to learn the data and create the model, while the validation set was used to evaluate the model's predictive performance. Each sample went through two data acquisition processes. The first was digital image acquisition using a Canon Kiss x71 camera. The second data acquisition was the measurement of leaf chlorophyll content using a chlorophyll-meter to measure the greenish index of leaves (leaf chlorophyll).

The leaf image is taken through a black acquisition box which is connected to an electric current so that the LED light is on and is used as a light source in the box. The camera is placed on top of the perforated box to take an image of the leaf. The resulting image has a Bitmap format (bmp). The image acquisition system can be seen in Figure 1.
A total of 480 images were obtained from the image acquisition stage, which then converted the image format to Bitmap (BMP). The result of feature extraction was the color co-occurrence matrix (CCM) in each color group Red, Green, and Blue (RGB). Then from RGB colour, it can be converted to grey [17]. From the CCM then the texture value was calculated using 10 Haralick texture equations [18] which includes energy, entropy, contrast, homogeneity, inverse difference moment (IDM), correlation, sum mean, variance, cluster tendency, and maximum probability.

The results of the data acquisition method of digital image analysis were 480 images as input data. Image data was divided into 75% as training data and 25% as validation data. Separation of training data and validation data was the initial stage in the ANN process. Training data was applied to update weights, bias, and learn data patterns. Meanwhile, validation data was used to determine the ability of ANN or modeling results to identify new data patterns. The results of measuring chlorophyll content using a chlorophyll meter were used as output data. The imbalance value range can affect the quality of data mining results. Therefore, it was necessary to transform data with the normalization process. ANN topology design in this study used sensitivity analysis and the Backpropagation algorithm. Variations in the sensitivity analysis applied were learning rate, momentum, learning function, activation function, number of hidden layer nodes, and the number of hidden layers. Sensitivity analysis was performed to obtain the best ANN model with the lowest predictive error or validation mean square error (MSE).

3. Results and Discussion
The chlorophyll content increases as the leaves age. The mean chlorophyll content of each leaf development level was 41,015 CCI in the shoots, 47,655 CCI s in the middle, and 51,405 CCI s in the lower leaves (Figure 2). If observed from the color change, the leaves at the top have a yellowish-green color, the middle leaves are light green, while the lower leaves have a darker green color (dark green). The chlorophyll in the leaves at the shoots is usually still a protochlorophyllide, which when developing will turn green [19]. These results are consistent with research conducted by Pratama and Laily [20], which states that the chlorophyll content of gandasuli leaves will increase at the base. It was emphasized that the chlorophyll content will increase as the leaves age until the leaves are fully developed and then the chlorophyll decreases as the leaves get older [21]. This is because the older the leaves are, the ability to photosynthesize will decrease so that it will damage chlorophyll, where chlorophyll cannot perform its function properly [22].
Feature identification is used in ANN modeling to identify relevant features and clean up data that has redundant and irrelevant noise. The results of the texture selection feature obtained 9 features that have a high correlation in determining the chlorophyll content in *Vernonia amygdalina* leaves, such as $b^*_{(Lab)}$ energy, $b^*_{(Lab)}$ entropy, $b^*_{(Lab)}$ correlation, value maximum probability, $a^*_{(Lab)}$ energy, $a^*_{(Lab)}$ cluster, $a^*_{(Lab)}$ variance, $L^*_{(Lab)}$ correlation, and saturation $(HSV)$ cluster.

Figure 2. Examples of the image acquisition of *Vernonia amygdalina* leaves: (a) young leaves with chlorophyll content 41.015 cci; (b) mature leaves with chlorophyll content of 47.665 cci; (c) old leaves with chlorophyll content of 51.405 cci

Figure 3 shows that the color texture features are negatively correlated with leaf chlorophyll content with a low determination coefficient value of 0.1751. These results illustrate that from the image analysis obtained, the image of *Vernonia amygdalina* leaves with high chlorophyll content has a low color texture value. The energy texture value of CCM $b^*$ has the same results on the energy
texture of saturated CCM (HSV), CCM a*, and CCM hue, which results in higher values for sago starch texture analysis compared to other color texture analyses [23].

The entropy texture feature analyzes the texture of the image by measuring the randomness of the pixels in the image (intensity distribution). Entropy is the randomness or level of disturbance contained in the image [24]. Therefore, a homogeneous image will produce a lower entropy value, while an inhomogeneous (heterogeneous) area will produce a higher entropy value [25]. The b*(Lab) entropy texture feature value in this study is the randomness value of the b*(Lab) chromatic color randomness. The more random (heterogeneous) pixel pairs, the higher the entropy value. Figure 3b shows the relationship between b*(Lab) and entropy color texture features positively correlated with the chlorophyll content of Vernonia amygdalina leaves with a determinant coefficient value of 0.0086 which is slightly larger than the previous color texture. If you pay attention to the distribution of data in the image, it can be seen that the entropy texture in the leaf image has a high value, namely from a range of 1 to 2. So it can be concluded that the chromatic color texture b* a leaf image has a random texture.

The correlation test states the measure of the linear relationship between the gray level of one pixel relative to other pixels at a certain position [26]. The correlation to the level of leaf content is positively correlated with a high determinant coefficient value of 0.1054. The higher the chlorophyll content of leaves has the high of the textural value. This relationship illustrates that the textures of the Vernonia amygdalina leaf image have the same element value or a high degree of similarity. If you look at the data distribution in the image, it can be seen, the high correlation texture of leaves is in the range of 20000 to 70000.

![Figure 4. Value Maximum Probability features vs chlorophyll](image)

The correlation between the color texture feature Value Maximum Probability and the chlorophyll content in Vernonia amygdalina leaves has a negative correlation with a coefficient of determination of 0.051 (Figure 4). The lower the chlorophyll content, the higher the texture value. Judging from the distribution of data located in the range of low color texture values, i.e 0 to 0.08, it can be seen that the leaf image has a low level of image pixel regularity. In the research of Chiou et al [27], the maximum probability texture analysis measures the maximum value in pixel pairs. The maximum probability texture value will be high if the most dominant pixel pair in an image is high. Through this analysis, it can be seen the level of order in an image. The higher the maximum probability texture value, the more regular the image will be.

The a* chromatic color feature in the Lab color space represents a color change from green to red. The value on the energy texture feature is generated by measuring the number of repeated pixel pairs, this is also related to the similarity or homogeneity of pixels at the gray level of the image (Figure 5). The more homogeneous or similar to an image pixel, the higher the energy feature results, otherwise the low energy texture value indicates that the image is heterogeneous [28]. The color texture features a*(Lab) Energy has a determinant coefficient value of 0.0002. Through the image, it can be seen that the data distribution is located at a low texture value, which is in the range 0.2 to 0.5, which can be seen
that the texture of the Vernonia amygdalina leaf image has a low level of homogeneity of image pixels.

Variance has a negative correlation with the chlorophyll content of Vernonia amygdalina leaves as seen from the determinant coefficient of 0.185. The texture is measured from the CCM a* (Lab) Value, when the leaf image has a high chlorophyll content, the resulting texture has a low degree of heterogeneity, whereas the low chlorophyll content is indicated by the higher texture value. The distribution of data in the image shows the heterogeneous texture of the leaf image shown from the texture value range, namely at 20 to 30. The texture value a* (Lab) This variance will be remembered if the difference between the chromatic color values a* (Lab) and the associated global average is increasing [27].

The clusture texture characterizes the tendency of pixels to clusture and is a measure of asymmetry, the higher the texture value, the more asymmetric the image is [27]. The color texture of a* (Lab) cluster has a negative correlation with the chlorophyll content of leaves with a determinant coefficient value of 0.1833. These results illustrate that the texture of the clusters will be lower when the chlorophyll content of leaves increases. When seen in the figure, the data is spread over the range 80-120 which shows an asymmetrical image of leaves because it has a high cluster texture value.

![Figure 5. a* (Lab) Features vs chlorophyll; (a) energy; (b) variance; (c) clusture](image)

The color texture feature L* (Lab) correlation shows the degree of similarity of CCM L* (Lab) in the row or column direction [28]. In this research, the relationship between L* (Lab) color texture and chlorophyll content can be seen in Figure 6, the large change in the energy saturation color texture (HSL) value has a determinant coefficient of 0.0134. This result illustrates the degree of similarity of CCM L* which is higher when the leaf chlorophyll content increases. In general, when viewed from the distribution of data in the image, which is mostly located close to low texture values, namely in the range 0 to 2000000, it can be seen that the image of leaves has a low degree of CCM similarity.
The saturation of a color is a measure of how pure the color is. This color component is usually a value from 0 to 1 (or 0 to 100%) and shows a grayish color value where 0 indicates gray and 1 indicates a pure primary color. The relationship between the HSV cluster and the chlorophyll content has a positive correlation with the determinant coefficient of 0.0964. The higher the leaf chlorophyll content, the higher the texture value. It can be seen that the leaf image has an asymmetrical texture if it contains higher chlorophyll.

Several parameters affecting the ANN topology were varied with predetermined values such as the number of hidden layers (1,2), the number of neurons (nodes) (10, 20, 30, 40), learning rate and...
momentum (0.1, 0.5, 0.9). ANN models with different network topologies are trained and evaluated (validation) by trials and error methods to select the best network topology. The correlation coefficient (R) and MSE values from training and validation data for different ANN topologies can be seen in Table 1. The best topology is generated from two hidden layers with 30 nodes and 40 nodes, 0.1 learning rate and 0.5 momentum. This topology stops at the 18th iteration with a time of 2 seconds, when the minimum error value is reached.

Figure 8 shows the regression plots for the training simulation and the validation simulation. The value of R shows the correlation between the input and output variables. The closer the value to 1, the stronger the correlation. Through the training regression plot image, the data distribution is close to the linear conformity line which shows the predicted results are close to the true value, as seen from the correlation coefficient that is close to one, which is 0.98522. While the validation regression plot can be seen from the data distribution points that deviate from the linear fit line, causing a low validation correlation coefficient of 0.93417. Based on the results of the study, the coefficient value close to 1 indicates a strong relationship between image texture and chlorophyll content as output.

4. Conclusions
The texture features are selected to the identification of Vernonia amygdalina chlorophyll content, such as b*<sub>(Lab)</sub> energy, b*<sub>(Lab)</sub> entropy, b*<sub>(Lab)</sub> correlation, value maximum probability, a*<sub>(Lab)</sub> energy, a*<sub>(Lab)</sub> cluster, a*<sub>(Lab)</sub> variance, L*<sub>(Lab)</sub> correlation, and saturation (HSV) cluster. The best ANN model with a structure of 9-30-40-1 with a learning rate of 0.1 and momentum 0.5. The selected ANN structure produced an R training of 0.98522, R validation of 0.93417, MSE training of 0.0067, and MSE of validation of 0.0322. The results showed that digital image processing with color texture analysis and ANN models had the potential to act as sensors in detecting chlorophyll content of leaves.

References
[1] Sukmawati, Hadi H, Aminah, 2017 Potensi senyawa flavonoid daun afrika (Vernonia amygdalina Del.) asal ternate sebagai antioksidan As-Syifaa 9 2 195-200.
[2] Igile G, Oleszek W, Jurzysta M, Burda S, Fafunso M, Fasa<br>nmade A 1994 Flavonoids from Vernonia amygdalina and their antioxidant activities J. Agric. Food Chem 42 11 2445-2448.
[3] Moundipa F P, Kamini G, Melanie F, Bilong F C, Bruchhaus I 2006 In vitro amoebic activity of some medicinal plants of the Bamun region (Cameroon) Afr. J. Tradit. Cam 62 113-121.
[4] Adetunji C O, Olaniyi O O, Ogunkunle A T 2013 Bacterial activity of crude extracts of Vernonia amygdalina on clinical isolates J. Microbiol. Antimicrob 5 6 60-64.
[5] Ho Y G, Liang W S, Yeap B K, Yousr A H, Alitheen N B 2012 In-vitro and in-vivo antioxidant activity of Vernonia amygdalina water extract Afr. J Biotechnol 11 17 4090-4094.
[6] Ngatu N R, Okajima M K, Yokogawa M, Hirotada R, Takaishi M, Eitoku M, Muzembo B A, Sabah A B, Saruta T, Miyamura M, Kaneko T, Sano S, Suganuma N 2012 Anti-Allergic Effects of Vernonia amygdalina Leaf Extracts in Hapten-Induced Atopic dermatitis-like Disease in Mice *Allergy International* **61** 4 597-607

[7] Neldawanti, Ratnawulan, Gusnadi 2013 Analisis Nilai Absorbsi dan Penentuan Kadar Flavonoid untuk Berbagai Jenis Daun Tanaman Obat *Pillar of Physics* **2** 76-83.

[8] Wigmore A 1985 *The wheatgrass book: How to Grow and Use Wheatgrass to Maximize Your Health and Vitality Penguin United State of America*

[9] Setiawati, T, Saragih I A, Nurzaman M, Mutaqin A Z 2016 Analisis Kadar Klorofil dan Luas Daun Lampeni (*Ardisia humilis Thunbergh*) Pada Tingkat Perkembangan yang Berbeda di Cagar Alam Pengandaran. *Prosiding Seminar MIPA Peran Penelitian Ilmu Dasar dalam Menunjang Pembangunan Berkelanjutan.* Universitas Padjadjaran. Jatinangor.

[10] Deay A K, Manisha S, Meshram R 2016 An analysis of leaf chlorophyll measurement method using chlorophyll meter and image processing technique. Department of Electronics and Telecommunication, Bhilai Institute of Technology. India.

[11] Huck M, Löttker P, Böhle U R, Heymann E 2005 Paternity and kinship patterns in polyandrous moustached tamarins (*Saguinus mystax*). *Am. J. Phys. Anthropol* **127** 449-464

[12] Mamun M D, Kwon S, Kim J E, An K G 2020 Evaluation of algal chlorophyll and nutrient relations and the N:P ratios along with trophic status and light regime in 60 Korea reservoirs *Science of The Total Environment* **741** 140451.

[13] Ortuno M M, Mora P S, Hernandez R H, Andres V, Falco P C 2017 A new tool for direct non-invasive evaluation of chlorophyll a content from diffuse reflectance measurements *Science of The Total Environment* **609** 370-376.

[14] Hendrawan Y, Riza D F A, Murase H 2014 Applications of intelligent machine vision in plant factory *IFAC Proceedings* **47** 3 8122-8127.

[15] Hendrawan Y, Murase H 2011 Determining an ANN pre-treatment algorithm to predict water content of moss using RGB intensities *Engineering in Agriculture, Environment and Food* **4** 4 95-105.

[16] Hendrawan Y, Fauzi M R, Khirunnisa N S, Andreane M, Hartianti P O, Halim T D, Umam C 2019 Development of colour co-occurrence matrix (CCM) texture analysis for biosensing *IOP Conference Series: Earth and Environmental Science* **230** 012022.

[17] Rotterman Y, Porat M 2006 Colour image coding using regional correlation of primary colours *Image Vision Comput* **25** 637–651.

[18] Haralick R M, Shanmugam K, Dinstein 1973 Textural features for image classification *IEEE Trans. Syst. Man Cyb* **3** 6 610–621.

[19] Sumenda L 2011 Analisis kandungan klorofil daun mangga (*Mangifera indica L.*) pada tingkat perkembangan daun yang berbeda *Jurnal Bios Logos* **1** 1 1-5.

[20] Pratama A J, Laily A N 2015 Analisis Kandungan Klorofil Gandasuli (*Hedychium gardnerianum Shephard ex Ker-Gawl*) pada Tiga Daerah Perkembangan Daun yang Berbeda *Seminar Nasional Konservasi dan Pemanfaatan Sumber Daya Alam* 216–219.

[21] Setiawati T I A, Saragih M, Nurzaman, Mutaqin A Z 2016 Analisis Kadar Klorofil dan Luas Daun Lampeni (*Ardisia humilis Thunbergh*) Pada Tingkat Perkembangan yang Berbeda di Cagar Alam Pengandaran. *Prosiding Seminar MIPA Peran Penelitian Ilmu Dasar dalam Menunjang Pembangunan Berkelanjutan.* Universitas Padjadjaran: Jatinangor.

[22] Musyarofah N, Susanto S, Aziz S A, Kartosoewarno S 2007 Respon Tanaman Pegagan (*Centella asiatica L. Urban*) Terhadap Pemberian Pupuk Alami di Bawah Naungan. *Jurnal Agronomi Indonesia* **35** 3 217–224

[23] Sumarlan S H, Ariyanti D A 2017 Identifikasi Jenis Tepung dengan Machine Vision Menggunakan Metode Artificial Neural Network ( ANN ) *Jurnal Keteknikan Pertanian Tropis dan Biosistem* **5** 5 163–169

[24] Zayed N, Elnemr H A 2015 Statistical Analysis of Haralick Texture Features to Discriminate
Lung Abnormalities *International Journal of Biomedical Imaging* **2015** 1-7

[25] Yang X, Tridandapani S, Beitler J J, Yu D S, Yoshida E J, Curran W J, Liu T 2012 *Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: An in vivo study of late toxicit*. American Association of Physicists in Medicine America

[26] Ahsani A F, Sari Y A, Adikara P P 2019 Temu Kembali Citra Makanan Menggunakan Ekstraksi Fitur Gray-Level Co-occurrence Matrix dan CIE L*a*b* Color Moments Untuk Pencarian Resep Masakan. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* 3 3 2261-2268

[27] Chiou H, Yeh C, Hwang H, Liao Y 2019 Efficacy of Quantitative Muscle Ultrasound Using Texture-Feature Parametric Imaging in Detecting Pompe Disease in Children. *Entropy* 21 7 14 1-16

[28] Dong Z, Men Y, Li Z, Zou Q, Ji J 2019 Chlorophyll fluorescence imaging as a tool for analyzing the effects of chilling injury on tomato seedlings. *Scientia Horticulturae Elsevier* 246 490–497