Classification of Macronutrient Deficiency in Chili Leaves using *Support Vector Machine*

Deffa Rahadiyan[1] Sri Hartati[1,*] Wahyono[1] Andri Prima Nugroho[2] Lilik Sutiarso[2]

[1] *Department of Computer Science and Electronics, Universitas Gadjah Mada (UGM), Yogyakarta, Indonesia*
[2] *Department of Agricultural and Biosystems Engineering, Universitas Gadjah Mada (UGM), Yogyakarta, Indonesia*
[*]*Corresponding author. Email: shartati@ugm.ac.id*

ABSTRACT

Chili is a horticultural crop that has high economic value in Indonesia. The productivity level of chili in the country is not proportional to the level of consumption, one of the causes is malnutrition. Each plant requires different amounts of macronutrients and micronutrients to support plant growth and development. Chili plants that lack or excess macronutrients show different visual symptoms. *Digital Image Processing* is a non-destructive method that is useful for determining plant health conditions based on visual symptoms of chili leaves. The combination of digital image processing and learning methods such as the *Support Vector Machine* (SVM) helps classify the types of macronutrient deficiencies in order to obtain a nutrient solution. In this study, there are several stages in determining macronutrient deficiencies in chili plants, namely image acquisition, pre-processing, feature extraction, to classification using SVM with several kernels. Based on the experimental results in this study, the SVM method can help modern farmers to determine the health condition of plants non-destructively with 97.76% accuracy using a Polynomial kernel. Applying this system to an intelligent farming system is expected to support the realization of precision agriculture in Indonesia.

Keywords: *Image Processing, Machine Learning Algorithm, Nutrient Deficiency, Color Feature Extraction,Classifier.*

1. INTRODUCTION

Chili is a horticultural crop with high economic value in Indonesia [1]. However, the productivity level of chili in the country is not proportional to the level of consumption [2]. One of the causes is crop failure due to pests and plant diseases due to the provision of nutrientsthat are not following plant needs [3], [4].

Plants need macronutrients and micronutrients to support plant growth and development [5]. There are 4 phases: the embryonic phase, the juvenile phase, the production phase and the senile phase [6]. If the nutrients given are not following the growth and development phase, the plant cannot grow properly and even die [7]. There are two methods to determine the health condition of plants based on the lack of macronutrients, namely destructive and non-destructive methods. Destructive methods such as testing in the laboratory have weaknesses in the form of a length of time, queue length, number of repetition samples, and human error. Therefore, non-destructive methods are needed to determine the health condition of plants quickly without damaging the chili plants [8].

Digital Image Processing (DIP) is a non-destructive method that is useful in the case of classifying plant diseases based on their lack of macronutrients [8]. DIP processes certain information in an image. The information is in the form of object color information, object shape, object geometry, and others [9]. Chili plants with nutrient deficiency show specific visual characteristics such as changes in leaf color, texture, leaf tip, and others [2]. Leaf color is the most easily observed visual feature, even in the early stages of macronutrient deficiency [10], [11].

Leaf color information has been used to classify diseases and macronutrient deficiencies in various plants [12], [13]. Several studies have identified nitrogen deficiency in rice, tomato, and maize using the RGB and HSV color spaces[10], [12]–[14]. In addition, Cucumber and Olive tree plants also use statistical values from the RGB, HSV, and YGB color spaces[8]. Color space channel information can also represent

A. D. Saputro et al. (Eds.): ICOSEAT 2022, ABSR 26, pp. 564–575, 2023.
https://doi.org/10.2991/978-94-6463-086-2_77

other features such as texture, such as using RGB values to identify macronutrient deficiencies in Mango plants [15], [16].

Several machine learning methods can help to determine the type of macronutrient deficiency. They are Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Naïve Bayes, and Support Vector Machine (SVM)[8], [11], [17]. SVM provides promising results compared to several other ML methods in identifying macronutrient deficiencies in multiple plants based on color statistical information[18],[19], [20].

The proposed of this study is to do classification of deficiency of 4 types of macronutrients in chili plants based on the image processing in determining the type of macronutrients deficiency, namely data acquisition, preprocessing, image segmentation of chili plant leaves, feature extraction with various color channels, and classification using Support Vector Machine (SVM) as machine learning methods. This study selects the best channel and SVM kernel based on their accuracy for classifying macronutrient deficiencies in chili plants in the form of healthy, phosphorus deficiency, magnesium deficiency, and sulfur deficiency.

2. RELATED WORK

Classification of macronutrient deficiencies based on leaf image data has been widely carried out [6], [8]. Some studies use features such as color, shape, texture, morphology, and so on [9], [21]. Color is the useful information because it can be seen visually [10], [22]. Inaddition, several digital image processing methods in collaboration with machine learning have also been widely used to classify macronutrient deficiencies in plants based on color information [8], [23], [24].

Healthy leaves are usually standard green in color [13]. However, leaves that lack macronutrients such as nitrogen has discoloration such as turning yellow [25]. Research [10], [13] identifies nitrogen deficiency in rice based on RGB and HSV colors. The result is that the HSV color model produces a higher accuracy, up to 86%. Research [26] used the HIS/HSL color model. Research xx identified diseases on banana leaves by comparing the YUV, CIELAB, YCbCr, and HSV color space models[27]. The result is YUV color space followed by a final conversion to RGB shows the best result. Another study converts RGB to grayscale YCbCr and HIS. The best component in HIS is the I component, while in YCbCr, it is the Cr component [20].

RGB and HSV color values can be directly processed using the rule-based method [10], [13]. However, this method is challenging to apply to high-dimensional data. One method that can be used is the learning method. Learning methods, decision trees, and Naïve Bayes were used to identify olive trees based onthe RGB, HSV, and YGB color models. As a result, ML produces the highest accuracy, reaching 97% [8], [24], [17] study about the classification of macronutrient deficiency and MLP give the best performance, with 83.33% of accuracy. RFT, Naive Bayes, and SVM have been compared in other studies and SVM deliver the best performance [28]. In addition, research [20] also compares the use of deep learning methods with and without a combination of SVM. The result is that the combination with SVM increases accuracy by up to 12%.

SVM with various kernels has been tested. SVM works very well on high-dimensional datasets. SVM using kernel technique must map the original data from its original dimension to another relatively higher dimension. SVM does not involve all training data in the training process [29]. SVM has several kernels that can be used. At [30] Polynomial kernel with a soft margin produces the best performance with 95% correct classification compared to the other types of Kernel function. The RBF kernel produced the best result in xx with 80% accuracy [31].

Therefore, this paper tries to find a color channel that produces high performance based on our data. Then also look for the best SVM kernel based on specific color channels and its hyperparameters to produce high accuracy. The purpose is to classify four types of plant conditions such as healthy, phosphorus, magnesium, and Sulphur. Plant condition can be known well even thoughit is only based on color characteristics.

3. PROPOSED METHOD

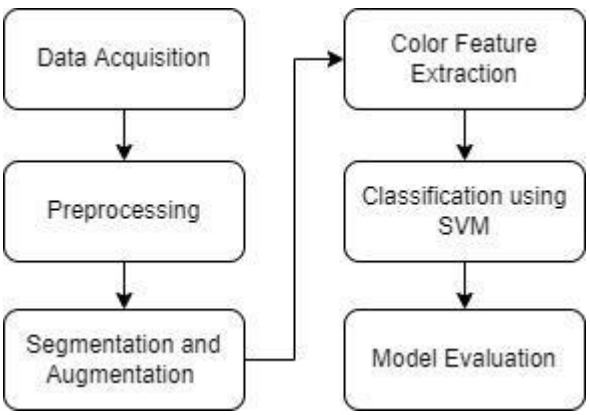


Figure 1 Macronutrient deficiency classification stages

Classification of types of macronutrient deficiencies based on the image of chili leaves can be done through several stages. This study carried out six stages: data acquisition, preprocessing, segmentation and

augmentation, color feature extraction, classification using Support Vector Machine (SVM), and model evaluation using Confusion Matrix. The flow of these stages is shown in Figure 1 and is explained in detail in the following sub-section.

3.1. Data Acquisition and Preprocessing

Data acquisition is the process of collecting image data in certain environmental conditions [2], [8]. In this study, leaf data was obtained with a specific scenario. There are four classes: Healthy, Phosphorus, Magnesium, and Sulphur. Leaf data for each class can be seen in the following figure.

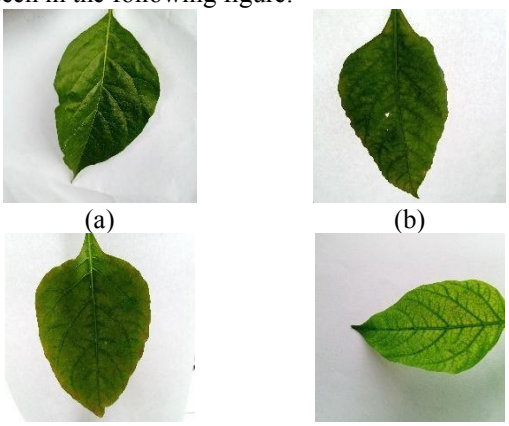


Figure 2 Data acquisition result of : (a) Healthy leaf, (b) Mg deficiency, (c) Sulphur deficiency, and (d) Phospohor deficiency.

Data acquisition took place in daylight, with bright and not foggy conditions. The single leaf of the plant is placed on a white background in varying lighting in the range of 8000-10000 Klux. The 32 MP of smartphone camera was used for acquisition. The distance and angle of retrieval of each data are the same. In this research, preprocessing is done by rotating some non-uniform data. The example in Figure 2 (d) is rotated to the left by90 degrees to normalize the leaf position.

3.2. Segmentation and Data Augmentation

Segmentation is the process of separating objects from the background [15], [32]. In this study, segmentation was performed using the colorspace method. The colorspace method in segmentation is a method of grouping colors on a particular object [33]. The flow chart of the segmentation process is shown in Figure 3 below.

Chili leaf RGB data is converted to L*a*b. Then the colorspace is done with the lower limit value of 120 and the upper limit value of 235. In this study, the information taken only comes from channel *a. The results of channel *a are dilated with a kernel size of 5x5 for three iterations. Then erosion is carried out for five times iterations. Pixel values less than 45 will be considered 0, while more than 45 will be considered 1. The result will be *bitwise_and* and masked with the original image. The segmentation results will be saved as a new image file.

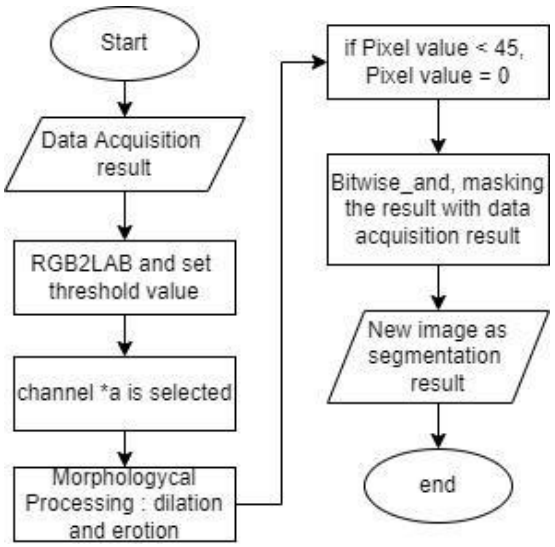


Figure 3 Segmentation flow chart

Augmentation is the process of multiplying data by performing certain variations [34], [35]. In this study, each class was augmented until each class reached 1000 data. Augmentation takes place using tools in the form of ImageDataGenerator [2], [34], [36]. Several variations were made: rotation up to 40 degrees, shear range of 0.2, the zoom range of 0.2, and brightness in the range of 0.5 to 1.10. The results of the augmentation data can be observed in Figure 4. below.

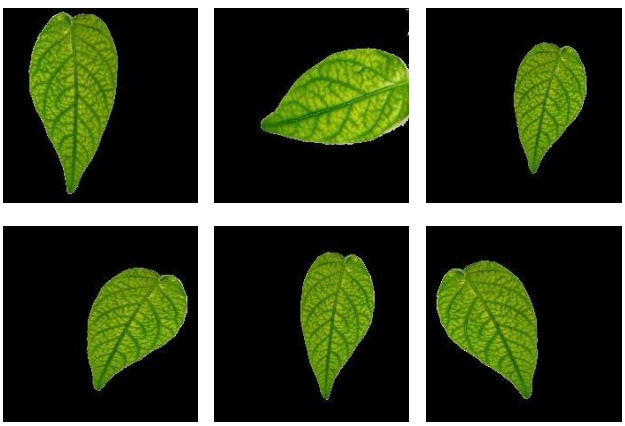


Figure 4 Augmentation data result

3.3. Color Feature Extraction

Chili plants that are deficient in certain macronutrients show different visual characteristics. Color is a feature that can be used as information on the visual characteristics of an object. In this study, there are eight color space models compared. The eight-color space models are as follow.

3.1.1. RGB

The camera type commonly used for capturing images in farming environments is RGB [8]. The RGB image colour extraction process uses Eq. (1)-(3). It shows that R is the value of the red component, G is the value of the green component, and B is the value of the blue component.

$$R' = \frac{R}{R+G+B} \quad (1)$$

$$G' = \frac{G}{R+G+B} \quad (2)$$

$$B' = \frac{B}{R+G+B} \quad (3)$$

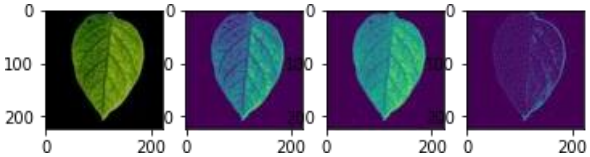


Figure 5 RGB channel result

3.1.2. HSV

Each color component in the HSV color space has main characteristics: Hue, Saturation, and Value [37]. According to Crane, Hue refers to colors known toman, such as red and green. Saturation states the degree of purity of a color or how much white light is mixed with Hue. Value states the intensity of the object's reflection received by the eye [31], [38]. The components of HSV are obtained from the RGB color conversion that has been searched as shown in equations(4) – (7).

$$V = \max(R', G', B') \quad (4)$$

$$S = \begin{cases} \frac{V - \min(R', G', B')}{V}, & if\ V \neq 0 \end{cases} \quad (5)$$

$$H = \begin{cases} \frac{60(G' - B')}{V - \min(R', G', B')}, & if\ V = R \\ 2 + \frac{60(B' - R')}{V - \min(R', G', B')}, & if\ V = G \\ 4 + \frac{60(R' - G')}{V - \min(R', G', B')}, & if\ V = B \end{cases} \quad (6)$$

$$if\ H < 0, then\ H = H + 360 \quad (7)$$

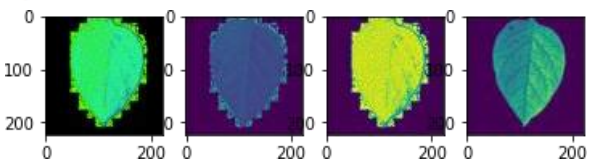


Figure 6 HSV channel result

3.1.3. HSI/HSL

The HIS color components H, S, and I represent Hue, Saturation and Intensity [39]. The difference between HSL and HSV is that a color with maximum lightness in HSL is pure white, but a color with maximum value/brightness in HSV is analogous to shining a white light on a colored object (e.g. shining a bright white light on a red object causes the object to appear still red, just brighter and more intense, while shining a dim light on a red object causes the object to appear darker and less bright [40].

$$H = \begin{cases} \theta_{rgb}, if\ B \leq G \\ 360 - \theta_{rgb}, if\ B > G \end{cases} \quad (8)$$

$$S = 1 - \frac{3}{R+G+B} \quad (9)$$

$$I = \frac{R+G+B}{3}, \text{where} \quad (10)$$

$$\theta_{rgb} = cos^{-1}\left\{ \frac{\frac{2}{\sqrt{(R-G)+(R-B)}}}{\frac{2}{\sqrt{(R-G)^2+(R-G)(R-B)}}} \right\}$$

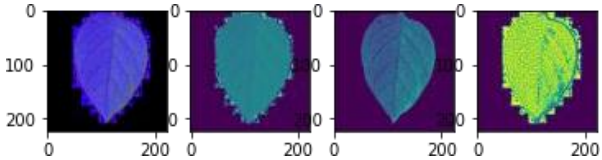


Figure 7 HSL/HSI channel result

3.1.4. YCrCb

Y′ is the luma component and Cr are the blue-difference and red-difference chroma components. Y′ (with prime) is distinguished from Y, which is luminance, meaning that light intensity is nonlinearly encoded based on gamma-corrected RGB primaries. The formula to convert the YCrCb value from RGB is shown in Equation (11) – (13).

$$Y = 16 + (0.275 * R + 0.504 * G + 0.009 * B) \quad (11)$$

$$C_b = 128 + (-0.148 * R - 0.291 * G + 0.439) \quad (12)$$

$$C_r = 128 + (0.439 * R - 0.368 * G - 0.071 * B) \quad (13)$$

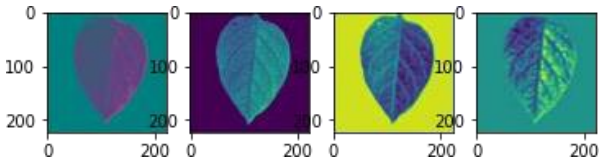


Figure 8 YCrCb channel result

3.1.5. YUV

The YUV model defines one luminance component(Y), that means physical linear-space brightness, and two chrominance components called U (blue projection)and V (red projection) respectively. It can be used to convert to and from the RGB model, and with different color spaces [27]. The formula convert RGB color model to YUV shown in Equation (14) - (16).

$$Y = 0.257 * R + 0.504 * G + 0.098 * B + 16 \quad (14)$$

$$U = -0.148 * R - 0.291 * G + 0.439 * B + 128 \quad (15)$$

$$V = 0.439 * R - 0.368 * G - 0.071 * B + 128 \quad (16)$$

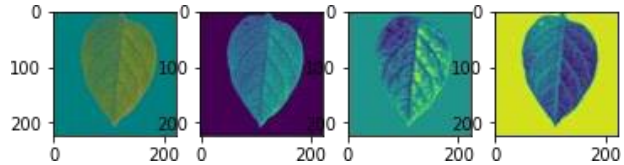


Figure 9 YUV channel result

3.1.6. XYZ

Three-dimensional cartesian coordinate systemaxes visualize XYZ labelled X, Y, and Z. In the XYZ color space, Y corresponds to relative luminance; Y also carries color information related to the eye's "M" (yellow-green) cone response. X and Z carry additional information about how the cones in the human eye respond to light waves of varying frequencies. The formula to convert the RGB color model to XYZ is shown in Equation (17) - (19).

$$X = 0.41245 * R + 0.3575 * G + 0.18042 * B \quad (17)$$

$$Y = 0.21267 * R + 0.71516 * G + 0.72169 * B \quad (18)$$

$$Z = 0.01934 * R + 0.11919 * G + 0.95022 * B \quad (19)$$

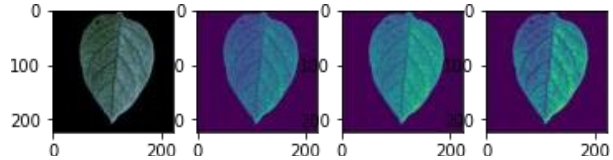


Figure 10 XYZ channel result

3.1.7. LUV

LUV color space is simply the brightness and chroma. For example, white is a bright color, while gray is considered a less bright white. In other words, the chromaticity of white and gray are the same, but the brightness is different. In this color space, "L" represents brightness, "U" and "V" represent chroma [4]. The chroma range is -100 to +100, and the range of brightness is 0 to 100. The formula to convert LUV from XYZ show on (20)-(23).

$$L = \begin{cases} \frac{\left(\frac{29}{3}\right)^3 Y}{Y_n}, if\ Y/Y_n \leq \left(\frac{6}{29}\right)^3 \\ 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}}, if\ \frac{Y}{Y_n} > \left(\frac{6}{29}\right)^3 \end{cases} \quad (20)$$

$$U = 13 \cdot L \cdot (u - u_n) \quad (21)$$

$$V = 13 \cdot L \cdot (v - v_n) \quad (22)$$

$$where, \quad (23)$$

$$u = \frac{4X}{X+15Y+3Z}\ ;\ v = \frac{9Y}{X+15Y+3Z}$$

$$u_n = \frac{4X_n}{-2X_n+12Y_n+3}\ ;\ v_n = \frac{9Y_n}{-2X_n+12Y_n+3}$$

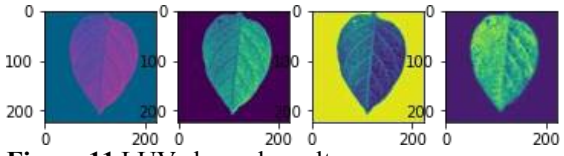


Figure 11 LUV channel result

3.1.8. Lab

Lab expresses the color in three values. L* for the luminosity from black (0) to white (100), a* from green (-) to red (+) and b* from blue (-) to yellow (+). Each a* and b* are found in the range of ± 127[27]. CIELAB was designed so that the same amount of numerical change in these values corresponds to roughly the same amount of visually perceived change [42]. The formula to covert Lab from XYZ color model is showon (24) - (26).

$$L = 116\, f\left(\frac{Y}{Y_n}\right) - 16 \quad (24)$$

$$a = 500\, (f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)) \quad (25)$$

$$b = 500\, (f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)) \quad (26)$$

Where :

$$f(n) = \begin{cases} n^{1/3} - 16, if\ n > 0.008856 \\ 7.787n + \frac{16}{116}, if\ n \leq 0.008856 \end{cases}$$

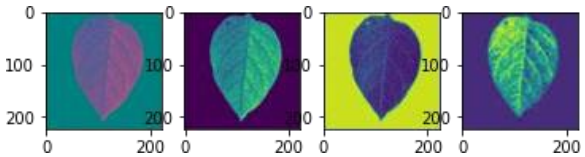


Figure 12 Lab channel result

Each color space is analyzed for statistical characteristics. Statistical characteristics are calculated in the form of Mean, standard deviation, and skewness for each channel [43]. The Mean is the average value, the standard deviation is the statistical value used to determine the closeness of the statistical sample to the average of the data, and skewness is the level of asymmetry or the distance of symmetry from a distribution. The formula for calculating Mean, standard deviation, and skewness is shown in Equation (27) - (29).

$$\mu = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} I_{ij} \quad (27)$$

$$\sigma = \sqrt[2]{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_{ij} - Mean\right)^2} \quad (28)$$

$$Skewness = \sqrt[3]{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_{ij} - Mean\right)^3} \quad (29)$$

The result of the feature extraction process is nine statistical information from a leaf object. The statistical

characteristics for each channel are the mean, standard deviation, and skewness. Examples of feature extraction results and labels are shown in Table 1. Where label 1 represents a healthy condition, label 2 represents a phosphorus deficiency condition, label 3 represents a magnesium deficiency condition, and label 4 represents a sulfur deficiency condition.

Table 1. Statistical feature as a feature extraction result

Channel R			Channel G			Channel B			Label
μ	σ	Skewness	μ	σ	Skewness	μ	σ	Skewness	
8.796	20.791	1.269	17.729	35.931	1.480	2.944	9.341	0.945	1
23.234	39.447	1.767	37.321	58.474	1.914	3.302	8.976	1.104	2
18.653	29.416	1.902	32.388	48.152	2.017	4.157	8.558	1.457	3
10.446	26.449	1.184	14.952	34.176	1.312	4.157	12.431	1.003	4

3.4. Support Vector Machine (SVM)

SVM is a method for classifying by separating data based on a hyperplane. An illustration of hyperplane to separate dataset is shown in Figure 12. Many machine learning techniques have been developed with the assumption of linearity so that the resulting algorithm is limited to linear cases [30]. The linear function of SVM:

$$g(x) = sign(f(x)) \quad (30)$$

$$with\ f(x) = w^T x + b \quad (31)$$

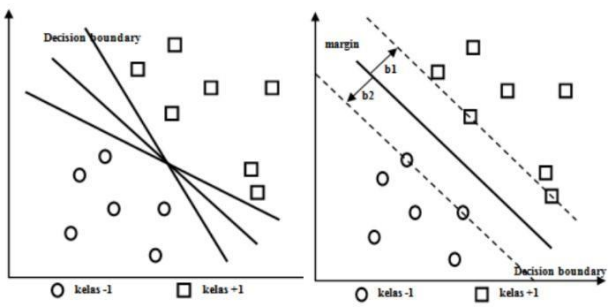


Figure 13 Possible decision limits for the dataset [30]

SVM can work on non-linear data by using a kernel approach to the initial data features of the data set. Kernel function maps a data set's initial dimension (lower dimension) to the new dimension (relatively higher dimension) [29], [44]. This study compares three kernels for each color channel statistical information: Linear kernel, Radial Basic Function (RBF), and Polynomial kernel.

The linear Kernel function works for linearly separable data, while polynomial kernel functions for hard and soft margins deal with nonlinear separable patterns. Different Kernel function gives different optimal hyperplane [30]. The linear and polynomial kernels have the same formula as shown in Equation (32). When $d = 1$, it becomes the linear kernel.

$$K(x_i, x_j) = ((x_i \cdot x_j) + c)^d \quad (32)$$

RBF is the default kernel used within the sklearn's SVM classification algorithm tools and can be described with the following Equation (33) below

$$K(x, x') = e^{-\gamma||x-x'||^2} \quad (33)$$

$$\text{Where} : \gamma = \frac{1}{\text{n features} * \sigma^2}$$

$||x - x'||^2$ is the squared Euclidean distance between two feature vectors (2 points). *Gamma* (γ) is a scalar that defines how much influence a single training example (point) has.

3.5. Confusion Matrix for Model Evaluation

The confusion matrix is a representation used to evaluate model learning methods [2]. Table 2 shows the confusion matrix representation.

The True Negative (TN) value is data that is correctly classified as negative or false output. True Positive (TP) is data that is correctly classified as positive or actual output. False Positive (FP) is data that is classified incorrectly if the output is positive or true. False Negative (FN) is data that is classified incorrectly. Parameters that can be calculated based on these criteria in this study are as follows.

Table 2. Confusion matrix representation

		Prediction Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (35)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (36)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (37)$$

Table 3. Classification result of macronutrient deficiency

Color model	SVM Kernel	Accuracy		Precision		Recall	
		Train	Test	Train	Test	Train	Test
HSL	Linear	93.21%	92.55%	93.11%	92.96%	92.88%	91.35%
	Polynomial	98.62%	97.75%	99.05%	98.32%	98.05%	97.51%
	RBF	97.84%	91.80%	98.73%	91.18%	98.47%	90.25%
HSV	Linear	87.81%	88.05%	87.86%	88.42%	88.53%	89.82%
	Polynomial	90.23%	89.53%	91.24%	90.27%	91.18%	90.16%
	RBF	90.01%	85.37%	97.87%	87.24%	97.82%	86.73%
LAB	Linear	87.81%	88.06%	88.51%	88.25%	88.91%	89.54%
	Polynomial	88.25%	88.38%	89.39%	89.41%	88.49%	88.04%
	RBF	86.79%	81.67%	87.20%	82.93%	87.38%	81.27%
LUV	Linear	86.94%	86.14%	87.03%	86.95%	87.85%	86.65%
	Polynomial	95.56%	92.99%	96.65%	93.72%	95.71%	92.99%
	RBF	84.81%	80.48%	85.19%	80.38%	85.09%	80.33%
RGB	Linear	86.94%	86.14%	87.47%	86.40%	87.63%	86.85%
	Polynomial	95.63%	93.29%	95.71%	93.19%	95.29%	92.17%
	RBF	90.34%	78.99%	87.35%	79.31%	86.65%	79.48%
YCbCr	Linear	85.93%	84.05%	86.68%	84.77%	86.17%	83.79%
	Polynomial	92.80%	92.10%	93.20%	92.54%	93.44%	92.82%
	RBF	88.96%	84.35%	89.64%	84.12%	89.72%	84.61%
YUV	Linear	88.25%	87.63%	89.31%	88.48%	88.38%	87.43%
	Polynomial	95.30%	94.91%	95.77%	95.23%	95.32%	95.37%
	RBF	90.78%	87.68%	91.53%	87.99%	91.56%	87.55%
XYZ	Linear	88.21%	87.96%	88.02%	88.84%	89.25%	88.61%
	Polynomial	95.41%	93.29%	95.21%	93.65%	95.88%	93.29%
	RBF	91.01%	84.50%	91.64%	84.01%	91.51%	84.14%

4. EXPERIMENTAL RESULT ANDDISCUSSION

The main purpose of this study is to classify the plant health condition using digital image processing and SVM as a machine learning method. The first step of the classification stage is to split processing corn leaf images for training and testing. Four classes are analyzed, each consisting of 950-1050 data. Training and testing data are divided by a ratio of 80% fortraining and 20% for testing [17].

Eight types of color space models were compared using the SVM method. The eight-color spaces are HSL, HSV, LAB, LUV, YUV, RGB, XYZ, and YCbCr. Each color model was trained using SVM with parameter values C=1, degree $= 3$, and γ =0.1. Then the three types of SVM kernels are compared to see the best performance in each color space. The results of the model performance are shown in Table 3.

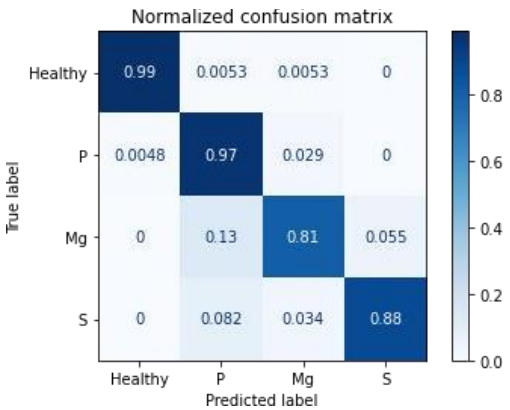
Based on table 3, each color model produces a different performance, as well as the use of the kernel type. The HSL color model generally performs better than the other seven color space models. The test accuracy is 92.55% for the linear kernel, 97.75% for the polynomial kernel, and 91.80% for the RBF kernel. HSL's lightweight dimensions resemble varying amounts of black or white paint in the mix. These saturated colors have a brightness of 0.5 in HSL. Unlike HSV, where color with maximum value/brightness is analogous to shining a white light on a colored object, the color with maximum brightness in HSL is pure white [39]. The data in this study was taken during the daywhen there is more white light. Therefore, HSL produces better performance compared to HSV and other color models.

The polynomial kernel produces the best performance compared to the linear and RBF kernel. Itis because of polynomial is an SVM kernel that can solve problems with non-linear data [30]. Even the polynomial kernel outperforms the other two kernels in each class. For the Healthy class, the polynomial kernel produces the same accuracy as Linear, which is 0.99. In the Phosphorus deficiency class, the Polynomial kernel excels with an accuracy of 1.00, followed by two other kernels producing the exact value of 0.97. For the Magnesium deficiency class, the poly kernel produces accuracy much different from the other two kernels, reaching 0.14. In the Sulphur class, the polynomial kernel still outperformed at 0.95, followed by two other kernels with the same value of 0.88. However, there are weaknesses in the polynomial kernel. The disadvantage of the polynomial kernel compared to other kernels is that the kernel requires a longer

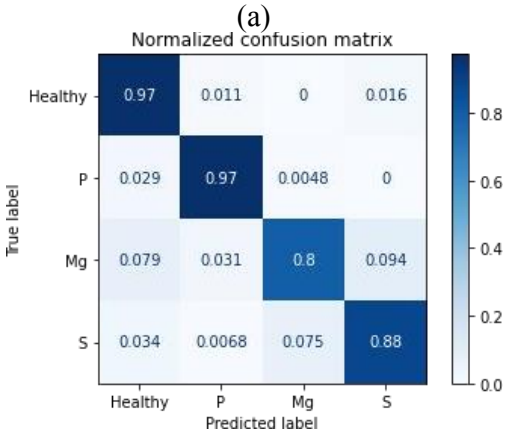
computation time.

Table 4. Combination of hyperparameter value result using Polynomial kernel

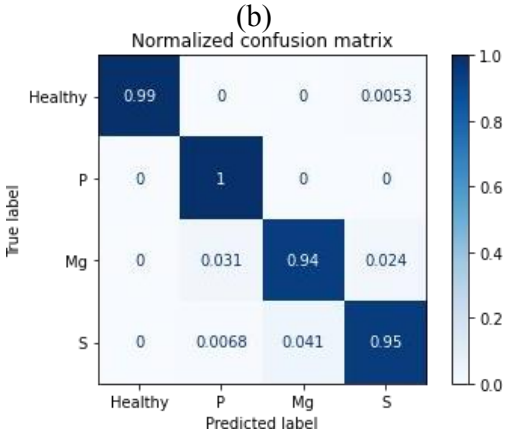
Color Model	Parameter		Accuracy		Precision		Recall	
	degree	C	Train	Test	Train	Test	Train	Test
HSL	3	1	98.69%	97.76%	99.38%	98.29%	98.14%	97.91%
	3	0.5	98.58%	97.72%	98.26%	98.25%	98.08%	97.83%
	3	0.1	98.69%	97.75%	99.15%	98.17%	98.03%	97.77%
	2	1	98.21%	97.75%	98.27%	97.75%	97.81%	97.02%
	2	0.5	98.13%	97.61%	96.33%	97.18%	97.41%	96.90%
	2	0.1	98.24%	97.17%	97.96%	97.43%	97.78%	96.32%



(a)



(b)

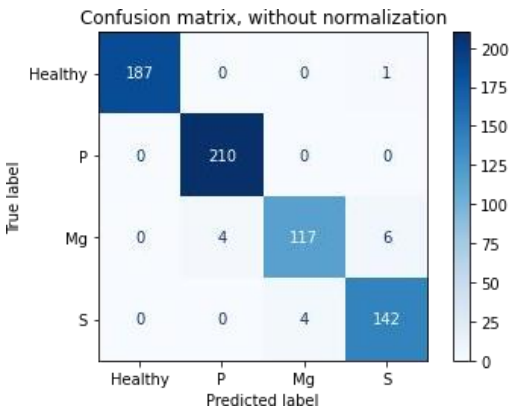


(c)

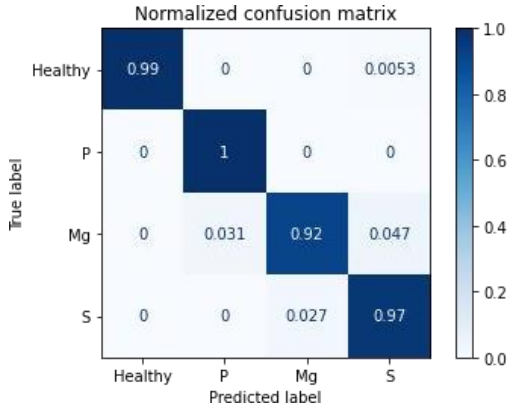
Figure 14 Confusion matrix result of: (a) Linear kernel, (b) RBF kernel, and (c) Polynomial kernel.

The next test is the variation of parameter values in SVM using the Polynomial kernel. Two varied parameters are the value of C and degree. The value of C was varied in 3 values, namely 1, 0.5, and 0.1, while the degree value was varied in 2 values, namely 2 and 3. The results of the variation in parameter values are shown in Table 4.

Based on table 4, variations in kernel values in this study affect changes in the values of accuracy, precision, and recall but are not significant. Even in the variation of the value of degree= 2 and degree=3 with parameter C = 1 hasa similar value of testing accuracy, which is 0.01% different. In addition, based on the results of variations, the parameters degree = 3 and C = 1 produced the best results in terms of accuracy, precision, and recall compared to other variations. Even the testing precision reaches 98.29%. Although degree=3 and C=1 show the best results, other parameters can still be used because the difference in results is not too far. The confusion matrix for the results of the degree = 3 and C = 1 can be observed in Figure 14 below.



(a)



(b)

Figure 15 Confusion matrix result of macronutrient deficiency classification for C=1 and degree=3, (a) without normalization, (b) with normalization.

5. CONCLUSION

Chilli plants' classification conditions can be overcome by color characteristics using the *Support Vector Machine* (SVM). This study compares eight color models: RGB, HSV, HSL, YCbCr, L*a*b, YUV, XYZ, HSL, and LUV. Based on experiments, HSL produces the highest result compared to other color models. This study also compares several SVM kernels. They are linear, polynomial, and RBF. The best performance is given to the polynomial kernel SVM method with an accuracy of 97.76% using degree=3 and C=1. For further study, we plan to analyze the combination of features to produce higher performance in future research.

AUTHORS' CONTRIBUTIONS

Deffa Rahadiyan: Writing - original draft, Investigation, Software. **Sri Hartati**: Conceptualization, Supervision. **Wahyono**: Methodology, Writing - review & editing, **Andri Prima Nugroho**: Formal analysis, Visualization. **Lilik Sutiarso**: Validation.

ACKNOWLEDGMENTS

This research was funded by Directorate of Research and Community Service, Deputy for Strengthening Research and Development, Ministry of Research, Technology/National Research and Innovation Agency of the Republic of Indonesia in the PMDSU program.

REFERENCES

[1] H. Waskito, A. Nuraini, and N. Rostini, Respon pertumbuhan dan hasil cabai keriting (Capsicum annuum L.) CK5 akibat perlakuan pupuk npk dan pupuk hayati, *Kultivasi*, vol. 17(2), 2018, pp. 676–681, doi: 10.24198/kultivasi.v17i2.17856.

[2] D. Rahadiyan and S. Hartati, Design of an Intelligent Hydroponics System to Identify Macronutrient Deficiencies in Chili, vol. 13(1), 2022, pp. 137–145.

[3] S. Sivagami and S. Mohanapriya, Automatic detection of tomato leaf deficiency and its result of disease occurrence through image processing, *Int.J. Innov. Technol. Explor. Eng.*, vol. 8(11), 2019, pp. 4165–4172, doi: 10.35940/ijitee.K1539.0981119.

[4] T. Amirtha, T. Gokulalakshmi, P. Umamaheswari, and T. R. M. Tech, Machine Learning Based Nutrient Deficiency Detection in Crops, *Int. J. Recent Technol. Eng.*, vol. 8(6), 2020, pp. 5330–5333, doi: 10.35940/ijrte.f9322.038620.

[5] V. Aleksandrov, Identification of nutrient deficiency in bean plants by prompt chlorophyll fluorescence measurements and Artificial Neural Networks, *bioRxiv*, no. June, p. 664235, 2019, doi: 10.1101/664235.

[6] M. P. S. da Silva, M. S. Mendonça Freitas, P. CesarSantos, A. J. C. de Carvalho, and T. S. Jorge, Capsicum annuum var. annuum under macronutrients and boron deficiencies: Leaf content and visual symptoms, *J. Plant Nutr.*, vol. 42(5), 2019, pp. 417–427, doi: 10.1080/01904167.2018.1544255.

[7] C. Jae-Won, T. T. Trung, T. Le Huynh Thien, P. Geon-Soo, C. Van Dang, and K. Jong-Wook, A nutrient deficiency prediction method using deep learning on development of tomato fruits, *iFUZZY 2018 - 2018 Int. Conf. Fuzzy Theory Its Appl.*, pp. 338–341, 2018, doi: 10.1109/iFUZZY.2018.8751688.

[8] D. Rahadiyan, S. Hartati, Wahyono, and A. P. Nugroho, an Overview of Identification and Estimation Nutrient on Plant Leaves Image Using Machine Learning, *J. Theor. Appl. Inf. Technol.*, vol. 100(7), 2022, pp.1836-1852.

[9] M. Lukic, E. Tuba, and M. Tuba, Leaf recognition algorithm using support vector machine with Hu moments and local binary patterns, *SAMI 2017 - IEEE 15th Int. Symp. Appl. Mach. Intell. Informatics, Proc.*, no. June 2020, 2017, pp. 485–490, doi: 10.1109/SAMI.2017.7880358.

[10] M. V. Latte, S. Shidnal, and B. S. Anami, Rule Based Approach to Determine Nutrient Deficiency in Paddy Leaf Images, *Int. J. Agric. Technol.*, vol. 13(2), 2017, pp. 227–245.

[11] L. Tan, J. Lu, and H. Jiang, Tomato Leaf Diseases Classification Based on Leaf Images: A Comparison between Classical Machine Learning and Deep Learning Methods, *AgriEngineering*, vol. 3(3), 2021, pp. 542–558, doi: 10.3390/agriengineering3030035.

[12] S. G. Brahmini and K. J. Rani, NPK deficiency detection in paddy leaf images, *Int. J. Futur. Revolut. Comput. Sci. Commun. Eng.*, vol. 3(9),

2017, pp. 102–107.

[13] M. V. Latte and S. Shidnal, Multiple nutrient deficiency detection in paddy leaf images using color and pattern analysis, *Int. Conf. Commun. Signal Process. ICCSP 2016*, 2016, pp. 1247–1250, doi: 10.1109/ICCSP.2016.7754352.

[14] M. Sardogan, A. Tuncer, and Y. Ozen, Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm, *UBMK 2018 - 3rd Int. Conf. Comput. Sci. Eng.*, 2018, pp. 382–385, doi: 10.1109/UBMK.2018.8566635.

[15] A. Jose, S. Nandagopalan, V. Ubalanka, and D. Viswanath, Detection and classification of nutrient deficiencies in plants using machine learning, *J. Phys. Conf. Ser.*, vol. 1850(1), 2021, doi: 10.1088/1742-6596/1850/1/012050.

[16] N. Sabri, N. S. Kassim, S. Ibrahim, R. Roslan, N. N. A. Mangshor, and Z. Ibrahim, Nutrient deficiency detection in maize (Zea mays L.) leaves using image processing, *IAES Int. J. Artif. Intell.*, vol. 9(2), 2020, pp. 304–309, doi:10.11591/ijai.v9.i2.pp304-309.

[17] K. A. Myo Han and U. Watchareeruetai, Black Gram Plant Nutrient Deficiency Classification in Combined Images Using Convolutional Neural Network, *2020 8th Int. Electr. Eng. Congr. iEECON 2020*, 2020, doi: 10.1109/iEECON48109.2020.229562.

[18] A. Siedliska, P. Baranowski, J. Pastuszka-Woźniak, M. Zubik, and J. Krzyszczak, Identification of plant leaf phosphorus content at different growth stages based on hyperspectral reflectance, *BMC Plant Biol.*, vol. 21(1), 2021, pp. 1–17, doi: 10.1186/s12870-020-02807-4.

[19] L. N and K. K. Saju, Classification of Macronutrient Deficiencies in Maize Plant Using Machine Learning, *Int. J. Electr. Comput. Eng.*, vol. 8(6), 2018, pp. 4197–4203, doi: 10.11591/ijece.v8i6.pp4197-4203.

[20] F. Jiang, Y. Lu, Y. Chen, D. Cai, and G. Li, Image recognition of four rice leaf diseases based on deep learning and support vector machine, *Comput. Electron. Agric.*, vol. 179, no. April, p. 105824, 2020, doi: 10.1016/j.compag.2020.105824.

[21] A. Kadir, L. E. Nugroho, A. Susanto, and P. I. Santosa, Leaf Classification Using Shape, Color, and Texture Features, no. May 2014, 2013, [Online]. Available: http://arxiv.org/abs/1401.4447.

[22] S. B. Sulistyo, W. L. Woo, and S. S. Dlay, Regularized Neural Networks Fusion and Genetic Algorithm Based On-Field Nitrogen Status Estimation of Wheat Plants, *IEEE Trans. Ind. Informatics*, vol. 13(1), 2017, pp. 103–114, doi: 10.1109/TII.2016.2628439.

[23] A. Shah, P. Gupta, and Y. M. Ajgar, Macro-Nutrient Deficiency Identification in Plants Using Image Processing and Machine Learning, *2018 3rd Int. Conf. Converg. Technol. I2CT 2018*, 2018, pp. 1– 4, doi: 10.1109/I2CT.2018.8529789.

[24] J. Drdsh, D. Eleyan, and A. Eleyan, A Prediction Olive Diseases Using Machine Learning Models, Decision Tree and Naïve Bayes Models, *J. Theor. Appl. Inf. Technol.*, vol. 99(18), 2021, pp. 4231–4240.

[25] D. P. Putra, P. Bimantio, T. Suparyanto, and B. Pardamean, Expert system for oil palm leaves deficiency to support precision agriculture, *Proc. 2021 Int. Conf. Inf. Manag. Technol. ICIMTech 2021*, no. August, 2021, pp. 33–36, doi: 10.1109/ICIMTech53080.2021.9535083.

[26] Z. Chen and X. Wang, Model for estimation of total nitrogen content in sandalwood leaves based on nonlinear mixed effects and dummy variables using multispectral images, *Chemom. Intell. Lab. Syst.*, vol. 195, no. July, p. 103874, 2019, doi: 10.1016/j.chemolab.2019.103874.

[27] R. Guerrero, B. Renteros, R. Castaneda, A. Villanueva, and I. Belupu, Detection of nutrient deficiencies in banana plants using deep learning, pp. 1–7, 2021, doi: 10.1109/icaacca51523.2021.9465311.

[28] J. Sosa, J. Ramírez, L. Vives, and G. Kemper, An Algorithm for Detection of Nutritional Deficiencies from Digital Images of Coffee Leaves Based on Descriptors and Neural Networks, *2019 22nd Symp. Image, Signal Process. Artif. Vision, STSIVA 2019 - Conf. Proc.*, 2019, pp. 3–7, doi: 10.1109/STSIVA.2019.8730286.

[29] R. A. Wijayanti, M. T. Furqon, and S. Adinugroho, Penerapan Algoritme Support Vector Machine Terhadap Klasifikasi Tingkat Risiko Pasien Gagal Ginjal, *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2(10), pp. 3500–3507, 2018, [Online]. Available: http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/2647/991/.

[30] H. M. Asraf, M. T. Nooritawati, and M. S. B. S. Rizam, A comparative study in kernel-based Support Vector Machine of oil palm leaves nutrient disease, *Procedia Eng.*, vol. 41, no. Iris, 2012, pp. 1353– 1359, doi: 10.1016/j.proeng.2012.07.321.

[31] Y. Sari, M. Maulida, R. Maulana, J. Wahyudi, and A. Shalludin, Detection of Corn Leaves Nutrient Deficiency Using Support Vector

Machine (SVM), *Proc. - 2021 4th Int. Conf. Comput. Informatics Eng. IT-Based Digit. Ind. Innov. Welf. Soc. IC2IE 2021*, 2021, pp. 396–400, doi: 10.1109/IC2IE53219.2021.9649375.

[32] S. B. Sulistyo, D. Wu, W. L. Woo, S. S. Dlay, and B. Gao, Computational Deep Intelligence Vision Sensing for Nutrient Content Estimation in Agricultural Automation, *IEEE Trans. Autom. Sci. Eng.*, vol. 15(3), 2018, pp. 1243–1257, doi: 10.1109/TASE.2017.2770170.

[33] S. Jeyalakshmi and R. Radha, a Review on Diagnosis of Nutrient Deficiency Symptoms in Plant Leaf Image Using Digital Image Processing, *ICTACT J. Image Video Process.*, vol. 7(4), 2017, pp. 1515–1524, doi: 10.21917/ijivp.2017.0216.

[34] L. A. Wulandhari *et al.*, Plant nutrient deficiency detection using deep convolutional neuralnetwork, *ICIC Express Lett.*, vol. 13(10), 2019, pp. 971–977, doi: 10.24507/icicel.13.10.971.

[35] D. Kuznichov, A. Zvirin, Y. Honen, and R. Kimmel, Data augmentation for leaf segmentation and counting tasks in rosette plants, *arXiv*, 2019.

[36] M. A. Islam, N. Rahman Shuvo, M. Shamsojjaman, S. Hasan, S. Hossain, and T. Khatun, An Automated Convolutional Neural Network Based Approach for Paddy Leaf Disease Detection, *Int. J. Adv. Comput. Sci. Appl.*, vol. 12(1), 2021, pp. 280–288, doi: 10.14569/IJACSA.2021.0120134.

[37] C. Junhua and L. Jing, Research on color image classification based on HSV color space, *Proc. 2012 2nd Int. Conf. Instrum. Meas. Comput. Commun. Control. IMCCC 2012*, no. 3, 2012, pp. 944– 947, doi: 10.1109/IMCCC.2012.226.

[38] K. Ayuningsih, Y. A. Sari, and P. P. Adikara, Klasifikasi Citra Makanan Menggunakan HSV Color Moment dan Local Binary Pattern dengan Naïve Bayes ClassifierAyuningsih, Karunia, Yuita Arum Sari, and Putra Pandu Adikara. 2019. Klasifikasi Citra Makanan Menggunakan HSV Color Moment Dan Local Binary Patt, *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol.3(4), 2019, pp. 3166–3173.

[39] O. S. Android, Detection of Plant Leaf Disease Using Image Processing Approach, *Int. J. Sci. Res. Publ.*, vol. 6(2), 2016, pp. 73–76.

[40] S. Sahurkar, P. B. J. Chilke, and S. M. Tech, Assessment of Chlorophyll and Nitrogen Contents of Leaves Using Image Processing Technique, *Int. Res. J. Eng. Technol.*, vol. 4(7), 2017, pp. 2243–2247.

[41] M. P. Rico-Fernández, R. Rios-Cabrera, M. Castelán, H. I. Guerrero-Reyes, and A. Juarez-Maldonado, A contextualized approach for segmentation of foliage in different crop species, *Comput. Electron. Agric.*, vol. 156, no. December 2018, 2019, pp. 378–386, doi: 10.1016/j.compag.2018.11.033.

[42] R. Rulaningtyas, A. B. Suksmono, T. L. R. Mengko, and G. A. Putri Saptawati, Segmentasi Citra Berwarna dengan Menggunakan Metode Clustering Berbasis Patch untuk Identifikasi Mycobacterium Tuberculosis, *J. Biosains Pascasarj.*, vol. 17(1), 2015, pp. 19, doi: 10.20473/jbp.v17i1.2015.19-25.

[43] N. S. A. M. Taujuddin *et al.*, Detection of plant disease on leaves using blobs detection and statistical analysis, *Int. J. Adv. Comput. Sci. Appl.*, vol. 11(8), 2020, pp. 407–411, doi: 10.14569/IJACSA.2020.0110852.

[44] H. Al Azies, D. Trishnanti, and E. Mustikawati P.H, Comparison of Kernel Support Vector Machine (SVM) in Classification of Human Development Index (HDI), *IPTEK J. Proc. Ser.*, vol. 0(6), 2019, p. 53, doi: 10.12962/j23546026.y2019i6.6339.