ISSN 2087-3336 (Print) | 2721-4729 (Online)
TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika
Vol. 11, No. 1, January, page. 33-40
http://jurnal.sttmcileungsi.ac.id/index.php/tekno
DOI: 10.37373

Red Chili Classification Using HSV Feature Extraction and Naive Bayes Classifier

Hermawan Syahputra[1], Josua Nainggolan[1] , Johanes Apriadi Parlinggoman Sirait[1], Muhammad Fadlan Ikromi[1]. Putri Ameliya Lubis[1]

[1*] Computer Science Study Program, Faculty of Mathematics and Natural Sciences, Medan State University (Universitas Negeri Medan), Indonesia, Jl. Willem Iskandar Pasar V Medan, Sumatera Utara, Indonesia, 20221

*✉ hsyahputra@unimed.ac.id

Submitted: 20/05/2023 Revised: 26/05/2023 Accepted: 27/05/2023

ABSTRACT

In the culinary industry, the classification of red chili pepper types is used to identify varieties that differ in terms of flavor, pungency, or other uniqueness. This enables their proper use in various recipes and meals. In the market, the classification of red chili pepper types helps in pricing, variety selection, or quality standards applied. For this reason, the purpose of this research is to classify red chili peppers using HSV Feature Extraction and Naive Bayes Clasifier. The stages carried out include: data collection, preprocessing, feature extraction and classification. Red chilies are grouped into 4 classes, namely large red chilies, cakplak red chilies, curly red chilies and chili red chilies. The red chili data used is 119 training data and 123 testing data. In the preprocessing, the image is converted to grayscale, then converted to binary image with the thresholding method. Furthermore, feature extraction is done with the HSV method. Finally, classification is done with Naive Bayes. The results of the study provide an accuracy value for training data of 92.43% and for testing data obtained an accuracy of 92.69%. This method is suitable for use in classification because it gives good results.

Keyword: Data mining; classification; chili; naive bayes; thresholding

1. INTRODUCTION

Due to its high market value, agricultural entrepreneurs in Indonesia frequently cultivate the vegetable product known as chili (Capsicum annum L). Chili is a plant-based dietary ingredient that is high in vitamin A, potassium, copper, vitamin C, vitamin B6, and vitamin B6. In general, compared to other plants, chili is a horticulture plant with a brief lifespan (less than a year). Chili vegetables currently come in a wide variety with various textures. However, in general, there are other advantages when seen from other angles, such as vitamins [1].

Red chili is one of the varieties of chili that is most frequently used as a vegetable. The area used for growing chilis has dropped somewhat as a result of the transfer of land uses, but the demand for red chilies has continued to rise [2]. It is necessary to alter the production growth pattern, which is governed by land area, if red chili plants are to become more productive. As a result, red chili varieties need to be categorized. Additionally, the variety of red chili used is crucial in varying the amount of spice in food. As a result, the price depends on the variety of red chili. These red chilies have a wide range of properties, so there needs to be a system to categorize them. Large red chilies, ticks, curls, and other objects will be used as research materials. The characteristics of the object under study are very similar to those of other varieties of red chile. A red chili variety must differ from one another in addition to

this consistency in order to be classified. Red chilies can be distinguished from one another visually by their size, color, and form [2].

The idea of data mining can be applied to categorize the different kinds of red chile. Data mining gained prominence in the 1990s across many industries, including marketing and business, science and engineering, and the arts and entertainment. Some experts claim that data mining, also known as Knowledge Discovery in Databases (KDD), is the practice of examining knowledge in databases [3]. Data mining belongs to the classification category. a Bayesian classification technique that uses the Bayes theorem to forecast the likelihood that a class will contain members of that class. This approach is comparable to that used to categorize neural networks and decision trees. Bayesian classification has demonstrated to have excellent speed and accuracy when used on huge databases [4].

Prior work on identifying objects in digital images using the Naive Bayes method included utilizing the MATLAB program to categorize an object based on its attributes [5][6]. This study used both the thresholding approach and naive Bayes. Thresholding, which automatically applies threshold values, is one method of segmenting images [7]. Thresholding generates an overall number of pixels with a gray level expressed as binary representations of black (0) and white (1). Converting RGB (Red, Green, Blue) photos to grayscale images, extracting image features based on average HSV (Hue, Saturation, and Value) values, identifying object areas, and classifying the resulting images are all examples of image processing.

The growth of red chilies on acidic soil has been studied, among other topics related to red chilies or vegetables [8], Additionally, there is study on categorizing different kinds of fruits and vegetables using the SVM technique [9]. It is believed that a classification system utilizing the Naive Bayes method will be able to classify different types of chili in accordance with their individual traits or traits.

2. METHOD

Data collection, pre-processing, feature extraction, and classification are the processes that were completed in this study.

2.1 Data collection

By photographing red chilies purchased from traditional marketplaces, data on red chilies were immediately gathered. Large red chilies, ticks, curly red chilies, and Lombok red chilies are among the varieties utilized.

Big red chili

The huge red chili, or Capsicum annuum L., is a significant vegetable that is utilized extensively as a food flavoring in Indonesia. Chili peppers not only include protein, fat, and carbs, but also vitamins A, B1, and C, which are vital components for human health. According to information from the Ministry of Agriculture (2009), 1.05 million tons of chili were produced in 2016 from 123,404 hectares of plantings, yielding an average productivity of 8.47 tons per ha-1. In 2017, production climbed by 15.37% over the prior year to 1.2 million tons. However, the 20 tons per ha-1 chile productivity potential is still distant from this [10].

Red chili tick

In Indonesia, the tick chili plant (Capsicum frutescens L.) can be found in both the lowlands and the highlands. Due to their high cost and enduring popularity in the market, chili ticks are highly valuable economically. The great demand for tick chili plants is among the factors that make them ideal for home gardeners [11].

Curly red chili

The Capsicum annum var. longum, sometimes known as the curly red chili, belongs to the Capsium genus. Curly red chilies are sharper and hotter than typical red chilies while being smaller. The fruit's uneven surface and somewhat meandering shape give it the appearance of being curly. This chili is known as a curly red chili because to its physical appearance. Young green and purple fruit are present. The fruit can be considered as either a vegetable or a spice depending on how it is used [12].

Cayenne pepperk

The red chili cultivar known as "Lombok red chili" is native to the Indonesian island of Lombok. This kind is well-known for its potent flavor of spices and distinctive scent. Compared to other red chili varietals, the Lombok red chili is larger.

This study will use two types of image groups, namely 119 digital images as training data and 123 digital images as testing data. The training data consists of:

- 30 digital images of large red chilies;
- 29 digital images of red chili ticks
- 30 digital images of curly red chilies;
- 30 digital images of chili red chilies.

While the data testing consists of:

- 31 digital images of large red chilies;
- 29 digital images of red chili ticks;
- 30 digital images of curly red chilies;
- 33 digital images of chili red chilies.

2.2 Preprocessing and feature extraction

In **Figure 1**, the research starts by entering test images into the program, adding up the total number of digital images entered, initializing variables from the image, reading the total RGB in the image, converting the total RGB image to grayscale and then from grayscale to binary, running complement operations, running morphological operations filling gaps, and finally performing feature extraction by converting RGB images to HSV, extracting each HSV component in the image, changing each HSV component to a different color, and changing the result to a different color.

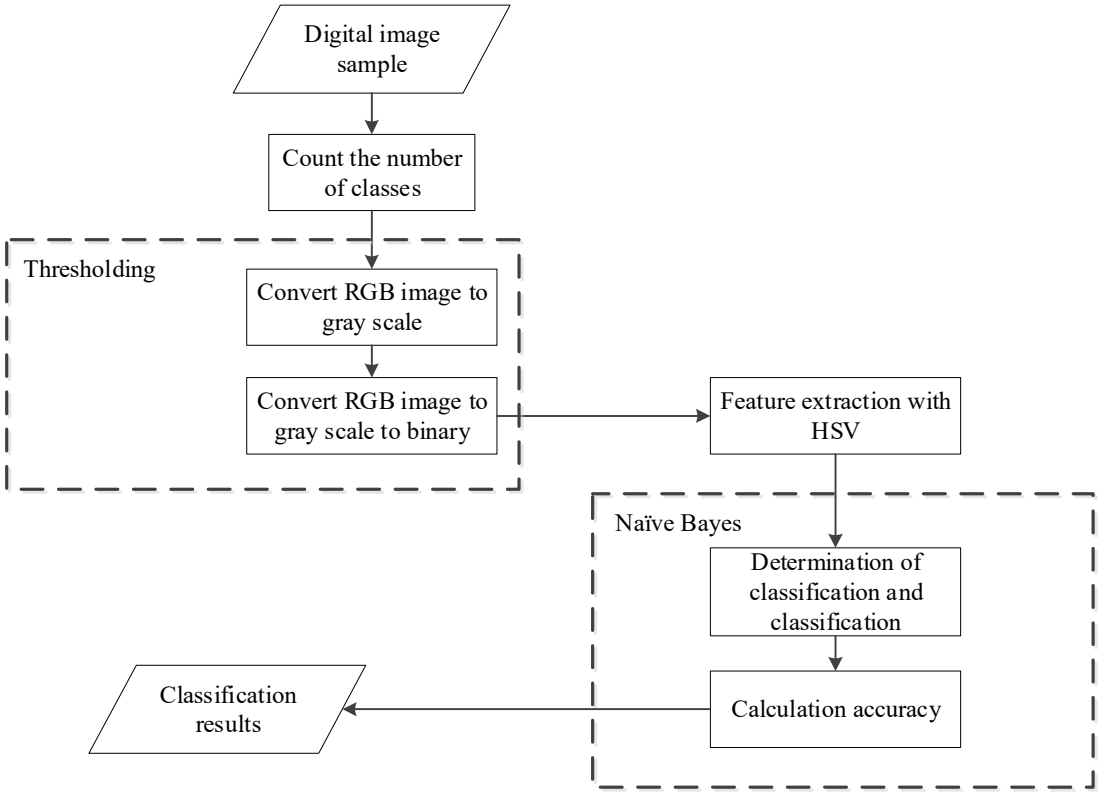

Figure 1. Flow processing

Segmentation

Image segmentation is the following step in the processing of images. This stage seeks to separate the image into key components that represent crucial information; for instance, the application distinguishes between objects and backgrounds. Automatic picture segmentation is the most challenging

task in digital image processing. The precision of the image segmentation process affects the effectiveness of object recognition [13].

Threshold

A threshold, or threshold value, which is utilized automatically, is one method of segmenting a picture. The first task is to generate a histogram that displays the number of pixels for each of the six levels of gray, i through L, which are represented by 0 to 255 in each level. K is used to denote the threshold point to be sought from a grayscale image, and it has a range of 0 to L-1 [14].

The probability of each pixel at level i is written by the equation

$$P_i = \frac{n_i}{N} \quad (1)$$

Where:

P_i = Pixel probability i

n_i = Number of pixels with gray level i

N = The total number of pixels in a digital image [15].

2.2 Classification of Naïve Bayes

One technique for classifying data is the naive Bayes approach. The likelihood of a class's membership can be estimated using a statistical classification technique called Bayesian classification. The capabilities of Bayesian classification are comparable to those of neural networks and decision trees. Bayesian classification is based on Bayes' theorem. Bayesian classification has demonstrated to have excellent speed and accuracy when used on huge databases [16].

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2)$$

P(HX) denotes the posterior probability of H given X. P(H) represents the prior or a priori probability of H (the assertion that X is class data) [17].

MATLAB programming is employed throughout the red chili classification implementation stage. A computer application called MATLAB can assist in resolving a wide range of mathematical issues that people in technical disciplines frequently encounter. Starting with the simplest ones, MATLAB may be used to quickly answer a wide range of numerical issues [18].

3. RESULTS AND DISCUSSION

3.1 Preprocessing and feature extraction results

The outcomes of the red chili data pre-processing, specifically the conversion of RGB to Grayscale images, are segmented by choosing the threshold value, as shown in **Figure 2**.

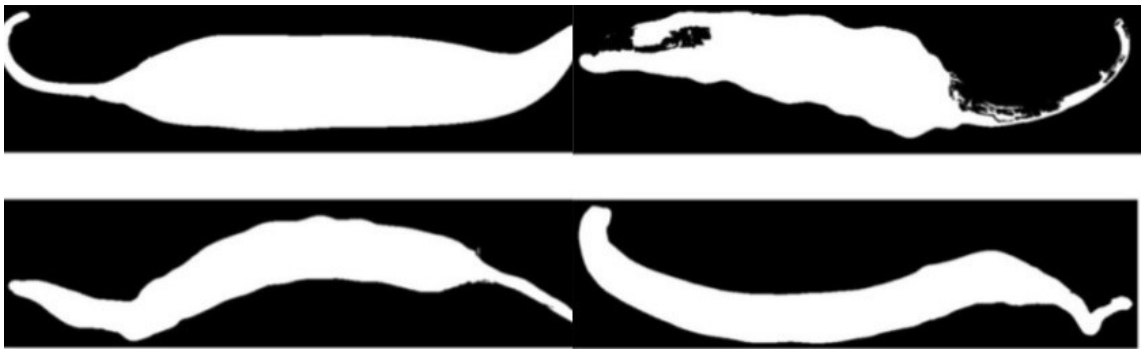

Figure 2. Segmentation results

Data from the threshold segmentation is added to the training set. The information is kept in a folder with the extension ".jpg". Additionally, a number of operations—including the complement operation

and the morphological operation of filling gaps—are performed to produce an ideal binary image. Lastly, some data for HSV color feature extraction is provided.

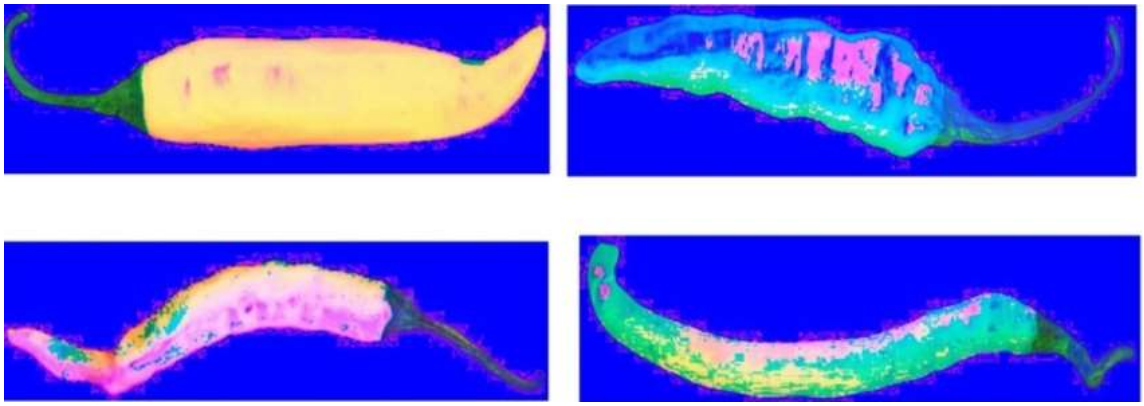

Figure 3. Feature extraction results

The average feature values of the samples used for feature extraction are shown in **Figure 3** in the form of the HSV values derived from the training data. This value was determined using the thresholding segmentation process, which involved converting a number of figures.

3.2 Classification results on data training

Continue building classes on the training data in accordance with the order in the training data folder as indicated in **Figure 4** after obtaining the average value data.

Figure 4. Classification of data training classes

The Nave Bayes approach was used to categorize the training data, and training data was used to determine the system's accuracy in classifying the four types of chili. The categorization of the four types of chili using 119 samples and the accuracy of the training data yielded a result of 92.43%.

3.2 Classification results on data testing

Testing is done on all of the provided testing data after acquiring the image accuracy value on the training data. The calculation and the training data are identical. To conduct this experiment, 31 huge red chilies, 29 tick red chilies, 30 curly red chilies, and 33 red chili peppers were photographed digitally.

After obtaining the average value data for **Figure 5**, continue by creating the class on the testing data in accordance with the folder's order. data analysis This experiment makes use of 33 Lombok red

chili digital photographs, 29 tick red chili digital images, 30 curly red chili digital images, and 31 huge red chili digital images.

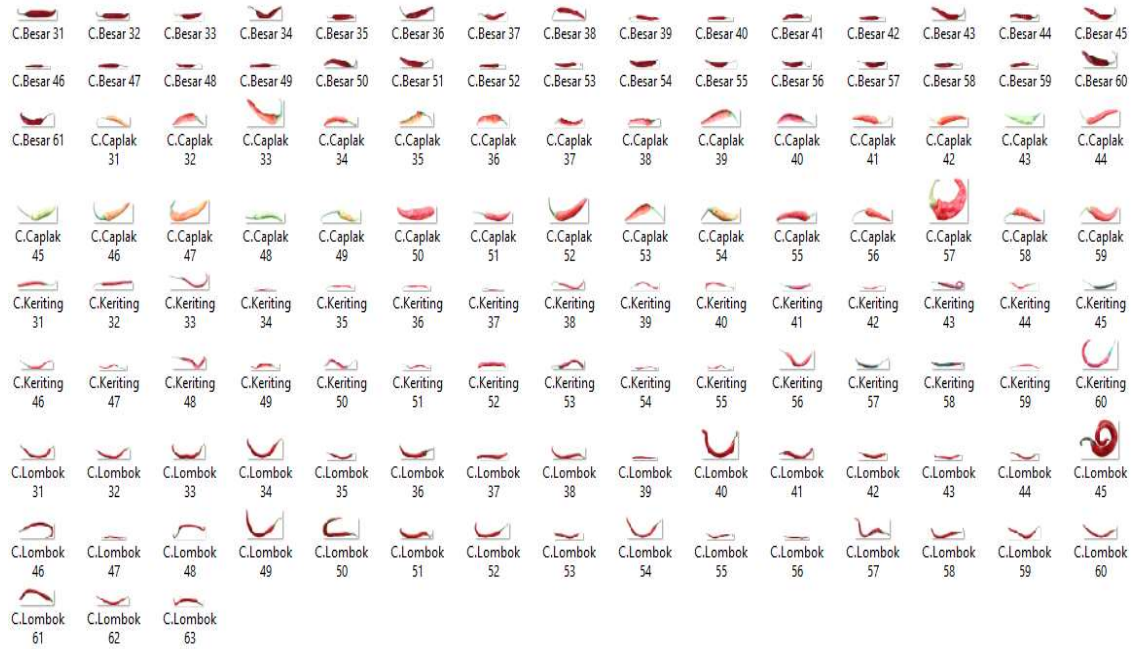

Figure 5. Classification of testing data classes

Figure 6. Results of testing data classification

Figure 6 presents the results of image processing, allowing it to be observed whether the data is consistent with the categorization or not. It also displays the outcomes of data testing. In this system, it is possible to determine, individually, whether the data corresponds to the defined class.

Only 114 photographs from the testing set are accurate, while the remaining 9 images cannot be classified. During testing, some data that is categorized to different red chili types was obtained. Consequently, the testing data's overall accuracy was 92.68%.

3.3 Implementation of a classification system with Matlab's Graphic User Interface (GUI)

After each training and test set's digital image has been processed, the accuracy value for each set has been calculated, and then the "User Interface System" processing has been completed.

In **Figure 7** is the display of the User Interface (UI) system. The way it works is that first click "Open Image" to select the image you want to test, want data from training data or data testing, on the image using the image on the data testing. When the image has been selected, the image is identified by displaying as an RGB image. Then select "Segmentation" to segment the RGB image with threshold segmentation, then the results of the binary image will come out. Then a conversion is performed using

HSV imagery to calculate the Hue, Saturation, Value and area of the object. Then select "Feature Extraction" then the output results are the values from the HSV image conversion.

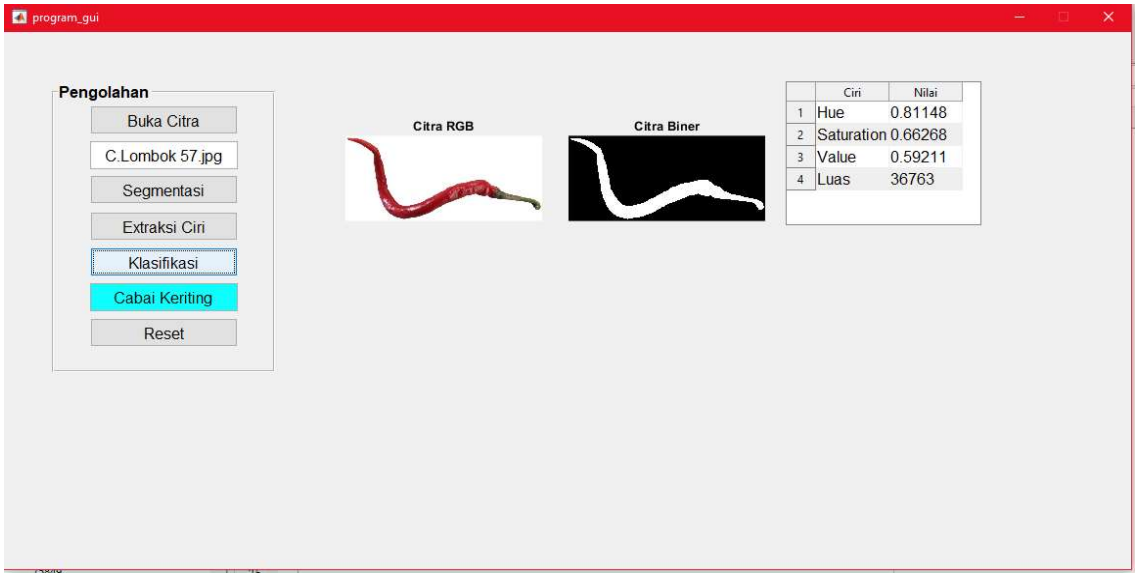

Figure 7. GUI Program results

Pressing "Classification" next produces data that has been classified using the Naive Bayes algorithm. Press "Reset" to remove all displays from the output once all components have been presented. Similar to all data commands, both for training and for testing. Utilizing all of the procedures previously mentioned, the data will be detected and categorized. How effective is this naive Bayes algorithm in classifying huge red chilies, red tick chilies, curly red chilies, or chili red chilies will then be revealed by the results.

4. CONCLUSION

To make it simpler for consumers or distributors to discern between the various types of red chilies, the Naive Bayes approach is excellent for classifying red chilies. The accuracy of the test on the training data is 92.43%, and on the testing data is 92.69%, according to research on the categorization of large red chilies, ticks, curly, and chilies using HSV feature extraction and the Naive Bayes classifier. The Nave Bayes approach is advised for recognizing data in the form of images because the testing data's accuracy results are good.

REFERENCE

[1] F. Nurainy, "Pengetahuan Bahan Nabati I: Sayuran, Buah-buahan, Kacang-kacangan, Serealia dan Umbi-umbian," *Jur. Teknol. Has. Pertan. Fak. Pertan. Univ. Lampung*, pp. 1–61, 2018.

[2] R. Ludovikus, A. Q. Unta, D. A. Pudjiastuti, and K. Yusuf, "EFISIENSI PRODUKSI USAHATANI CABAI MERAH (Capsicum annuum L.) (STUDI KASUS: DI DESA SUMBEREJO, KECAMATAN BATU)," *Buana Sains*, vol. 20, no. 2, pp. 197–208, 2020.

[3] M. A. Muslim *et al.*, *DATA MINING ALGORITMA C4.5*, I. Semarang: Universitas Negeri Semarang, 2019.

[4] S. Mita, Y. Yamazoe, T. Kamataki, and R. Kato, *Metabolic activation of a tryptophan pyrolysis product, 3-amino-1-methyl-5H-pyrido[4,3-b]indole(Trp-P-2) by isolated rat liver nuclei*, vol. 14, no. 3. 1981.

[5] H. B. Kurniasaria, Susilo, and I. Akhlis, "Penerapan Pengolahan Citra Digital Dengan Matlab 7.1 Pada Citra Radiograf," *Unnes Phisics J.*, vol. 1, no. 2252, pp. 1–4, 2012.

[6] T. Puspa, R. Sanjaya, A. Fauzi, A. Fitri, and N. Masruriyah, "Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine Analysis of review sentiment on shopee e-commerce using the naive bayes algorithm and support vector

machine," vol. 4, pp. 16–26, 2023, doi: 10.37373/tekno.v10i2.419.

[7] A. Agung and N. Gunawan, *Anak Agung Ngurah Gunawan PENGOLAHAN CITRA DIGITAL STEP BY STEP MENGGUNAKAN DELPHI 7 Penerbit : ANDI*. .

[8] M. Jamilah, P. Purnomowati, and U. Dwiputranto, "Pertumbuhan Cabai Merah (Capsicum annuum L.) pada Tanah Masam yang Diinokulasi Mikoriza Vesikula Arbuskula (MVA) Campuran dan Pupuk Fosfat," *Biosfera*, vol. 33, no. 1, p. 37, 2017, doi: 10.20884/1.mib.2016.33.1.347.

[9] Y. Yohannes, M. R. Pribadi, and L. Chandra, "Klasifikasi Jenis Buah dan Sayuran Menggunakan SVM Dengan Fitur Saliency-HOG dan Color Moments," *Elkha*, vol. 12, no. 2, p. 125, 2020, doi: 10.26418/elkha.v12i2.42160.

[10] R. Dermawan, M. Farid B. D. R., I. Ridwan Saleh, and R. Syarifuddin, "Respon Tanaman Cabai Besar (Capsicum annuum L.) terhadap Pengayaan Trichoderma pada Media Tanam dan Aplikasi Pupuk Boron," *J. Hortik. Indones.*, vol. 10, no. 1, pp. 1–9, 2019, doi: 10.29244/jhi.10.1.1-9.

[11] A. Lorensyifa and Z. Siregar, "budidaya Tanaman Cabai Caplak guna Meningkatkan Perekonomian Masyarakat Pesisir Dusun Desa Percut Sei Tuan," *Juranal Pengabdi. Masy.*, vol. 5, no. 6, pp. 2331–2337, 2022.

[12] I. G. B. T. Ananta and D. G. A. Anjasmara, "Antioxidant and Antibacterial Potency of Red Chillies Extract (Capsicum annum var. Longum)," *J. Ilm. Medicam.*, vol. 8, no. 1, pp. 48–55, 2022, doi: 10.36733/medicamento.v8i1.3170.

[13] Pranowo, "Buku Pengolahan Citra Berbasis PDE Dengan OpenCV," pp. i–119, 2015.

[14] I. W. A. Heryanto, Artama, M. W. S. Kurniawan, and G. A. Gunadi, "Segmentasi Warna dengan Metode Thresholding," *Wahana Mat. dan Sains*, vol. 14, no. 1, pp. 54–64, 2020.

[15] C. A. B. de Mello, "Image thresholding," *Digit. Doc. Anal. Process.*, vol. 2006, no. Snati, pp. 71–98, 2013, doi: 10.1201/9781003082224-3.

[16] S. L. B. Ginting and R. P. Trinanda, "Penggunaan Metode Naïve Bayes Classifier Pada Aplikasi Perpustakaan," *Expert Syst. Appl.*, pp. 1–7, 2014.

[17] Mustika *et al.*, *Data Mining dan Aplikasinya*. 2021.

[18] Z. Arif *et al.*, "TUTORIAL PRAKTIS BELAJAR MATLAB," *J. Pendidik. Fis.*, vol. 1, no. 1, pp. 1–80, 2021.