VEGETABLE TYPE CLASSIFICATION USING *NAIVE BAYES* ALGORITHM BASED ON IMAGE PROCESSING

Hanny Nurrani [1], Andi Kurniawan Nugroho[2], Sri Heranurweni[3], Eko Supriyanto[4], Generousdi[5]

[1,2,3] *Electrical Engineering Study Program, Electrical Engineering Departement, Semarang University, Central Java, Indonesia*

[4] *Electrical Engineering Departement, State Polytechnic of Semarang, Central Java, Indonesia*

[5] *Mechanical Engineering Study Program, National College of Technology, Jambi, Indonesia*

Abstract— There are so many different varieties of vegetables in Indonesia that the sorting procedure presents difficulties. In an effort to expedite the introduction of smart farming in Indonesia, more agricultural assistance techniques will be created. Utilizing the Naive Bayes algorithm is one way that may be used to advance agriculture in Indonesia. Image processing consists of converting RGB images to grayscale images, segmenting images using the thresholding method, collecting image features based on the HSV average value and object area, and classifying pictures using the Naive Bayes algorithm. This research seeks to use image processing technologies to agricultural products, particularly vegetables. The system is comprised of a single picture captured by a digital camera. There were eight varieties of vegetables employed for the picture data, with a total of eighty consisting of 64 training data and 16 test data. Spinach, green chilies, red chilies, chayote, cucumber, eggplant, tomatoes, and carrots were the vegetables categorized. The categorization findings indicate that 87.5 % of the test values produced using this approach are accurate. This study demonstrates that the Naive Bayes method has a high degree of accuracy for the categorization of vegetables based on image processing. It is anticipated that the findings of this study would promote the implementation of smart farming 4.0 in Indonesia.

Keywords— image processing, vegetables, nave bayes, classification

1. Introduction

This is due to the fact that the demand for food commodities to satisfy human requirements will continue to grow with the human population. Along with the advancement of technology and science, and in light of the agricultural sector's issues, the conventional farming method is no longer particularly efficient and effective. There are several improved innovative methods and systems in the agriculture business. Smart farming technology is anticipated to boost agricultural commodities. Image processing is one use of artificial intelligence technology. One of the technologies that must be developed is image processing. Vegetables such as spinach, chili, carrots, eggplant, cucumber, tomato, and chayote must be classified as agricultural commodities. Each variety of vegetable has a distinctive form and hue. Among the variety, there are various varieties of vegetables with nearly identical forms and hues [1]. Vegetables are currently classed based on the visual capabilities of the human eye, which have limits. This classification procedure has a number of flaws, such as using more energy to sort, not guaranteeing human consistency in identifying vegetable varieties, and requiring a great deal of time to sort [2].

On the foundation of their qualities, technological advancements can assist in classifying vegetables. Diverse vegetable properties necessitate a technique for classifying the varieties. The classification of vegetables is a challenging field of study due to the resemblance of certain varieties' forms. Nonetheless, it should be emphasized that these qualities frequently share commonalities, making them almost identical. Different forms and colors can be used to classify vegetables. The identification of vegetables can be aided by labeling based on the examination of particular characteristics.

Numerous studies on image processing techniques have been widely used for the classification of vegetable and fruit varieties. One of the research on the tomato ripeness detection system employing Color Histogram and k-Nearest Neighbor yielded an accuracy of 95%, precision of 96%, and recall of 94.1%. This study demonstrates that the performance of color histogram and k-Nearest Neighbor may be utilized to detect and categorize unripe or ripe tomatoes [3].

Using the Naive Bayes method, the further study on the classification of fruit recognition has been conducted. The objective of this study is to examine the use and performance of the Naive Bayes algorithm in the categorization of apple varieties. This software's technique includes picture capture, preprocessing, segmentation, analysis, and categorization of apple types. Accuracy is 81 percent, sensitivity is 73 percent, precision is 100 percent, and specificity is 70 percent, according to the findings of this study. This study demonstrates that naive Bayes is capable of identifying apple cultivars that are non-destructive. [4]

Based on Bayesian probability theory, Naive Bayes is a machine learning technique for categorization tasks. It is utilized for text categorization using large training data sets. The Naive Bayes algorithm is an algorithm that investigates the likelihood that an item with specific qualities belongs to a particular group or class [5]. It is, in brief, a probabilistic classifier. It is recognized not just for its simplicity, but also for its efficacy. The Naive Bayes method [6] allows for the rapid construction of models and the generation of predictions.

Based on the need for classification and sorting of vegetables, a system is needed that can help simplify, speed up, and lighten human work.

2. Methods

In this work, MATLAB R2020b software was used for the system design phase. The flow of research comprises a literature review, the collection of information, the definition of research objectives, the creation of software system procedures, the collection of data, and the production of findings.

The sorts of vegetables defined throughout the study phase include spinach, red chili, green chili, carrot, eggplant, cucumber, tomato, and chayote. Red, Green, and Blue pictures are converted to grayscale images prior to image segmentation using the thresholding method, image feature extraction based on the average Hue Saturation Value (HSV) and object area, and image classification utilizing the Naive Bayes algorithm. [7].

Figure 1 depicts the steps of the study to be conducted based on the vegetable picture.

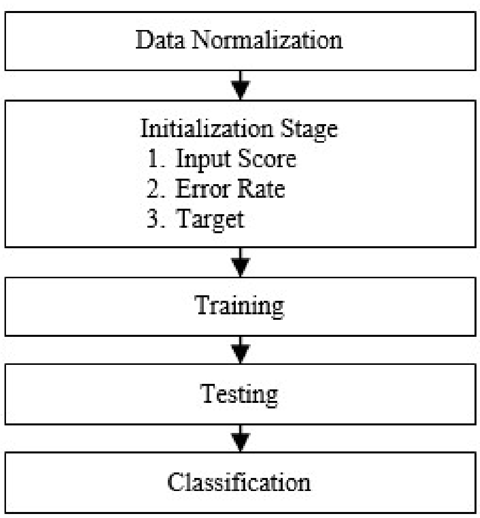

Figure 1. System Flowchart

From the normalized data, it is possible to quickly define the pattern that will be utilized for training and testing. The data is then loaded into the system, beginning with training, accuracy testing, and testing, to provide data or information output. Output data or information may be utilized for subsequent analysis or reprocessing. Figure 2 provides a flowchart representation of this study's methodology.

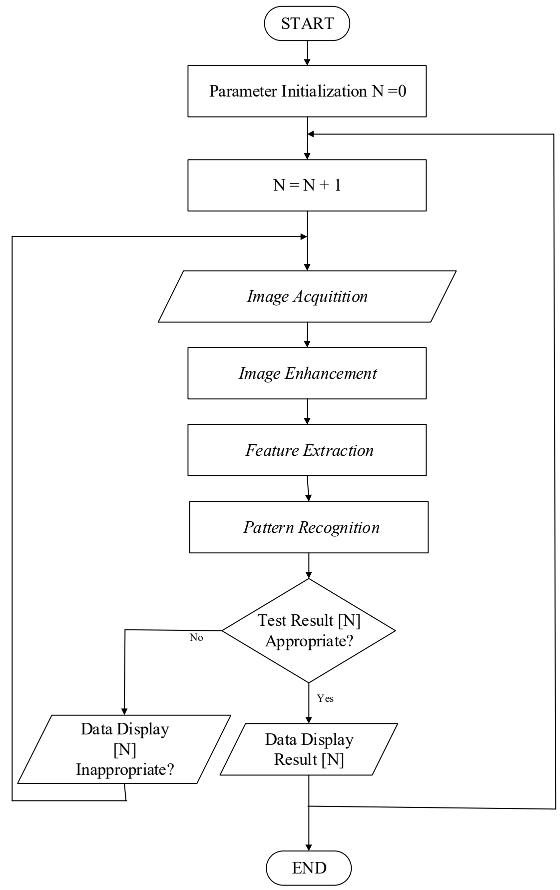

Figure 2. System Flowchart

First, the process begins with the acquisition of a vegetable image where the image is taken by a camera with a resolution, the catch is then improved in the quality of the vegetable image, including improving the quality by adjusting the brightness, either brightness or contrast, as well as removing noise, noise, and spots in the resulting image. generated by the camera. This is pre-processing in order to obtain good images so as to facilitate the next process, after going through image enhancement, feature extraction is carried out, namely extracting image features through several processes such as RGB to grayscale conversion, grayscale to black and white, edge detection and so on. . After obtaining feature extraction, pattern recognition is carried out using the nave Bayes algorithm method, at this stage the training process starts from various images so that the results are obtained from a conclusion in the form of vegetable classification. From the existing data, classifications will be obtained such as distinguishing spinach, carrots, tomatoes and so on. The training and testing process is carried out using a single image data. The vegetable image data used are 8 types of vegetables with a total of 80 consisting of 64 training data and 16 test data. Figure 3 below is the training data.

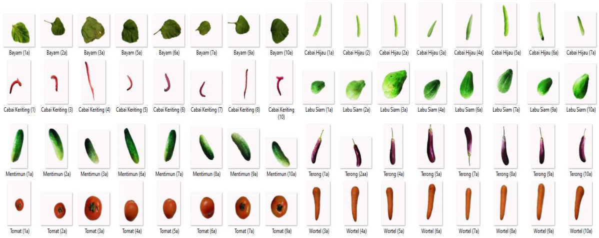

Figure 3. Image Training Data

And the data used for testing is shown in Figure 4 below.

Figure 4. Test Image Data

After designing the system, the next step is to test the vegetable image to find out the classification of vegetable types. Figure 5 below is the system made with a GUI or Graphical User Interface.

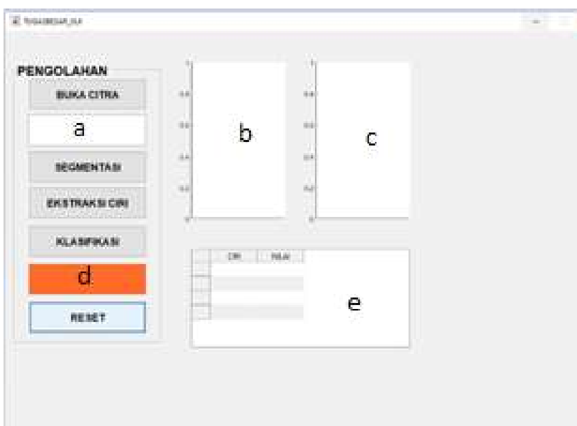

Figure 5. System Main Display

Regarding the primary view of the vegetable categorization system, the image above depicts the primary view. The first step is to open the image, wherein the image to be evaluated is selected and the name of the file to be evaluated is displayed in section a, the image to be evaluated is displayed in section b, and the image segmentation process is executed, with the segmentation results displayed in section c. After the segmentation procedure is complete, feature extraction is performed, and the results of extracted features and their values are shown in section e. Following feature extraction is classification, with the final categorization dependent on the extracted value. The reset button has the capability of resetting the display.

3. Result and Analysis

This investigation into the categorization of vegetables using the Naive Bayes algorithm and a graphical user interface yields very good results. The system can categorize photos of vegetables, including spinach, green chilies, red chilies, chayote, cucumber, eggplant, tomatoes, and carrots, with a training data accuracy rate of 100 percent and a test data accuracy rate of 87.5%. Two of the sixteen test data for vegetables had incorrect test findings.

The results of the vegetable categorization test using the Naive Bayes method are provided below. The test results for vegetables are depicted in the image below.

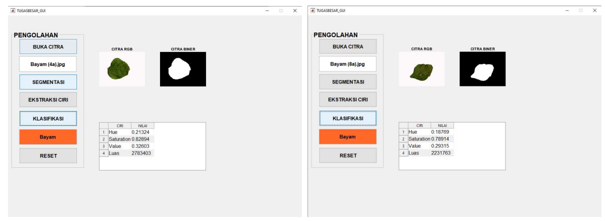

Figure 6. Results of Classification of Spinach Vegetables

In the test results of Figure 6, the resulting classification is appropriate.

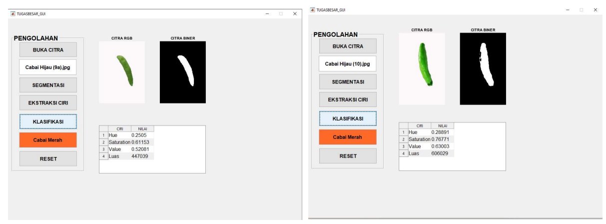

Figure 7. Results of Classification of Green Chili Vegetables

In Figure 7's test results, the categorization of green chili veggies is incorrect; the picture of a green chili is incorrectly identified as a red chili. In terms of form, the results of segmentation and feature extraction for green chili and red chili are comparable. The algorithm utilized in this study was unable to categorize more precisely.

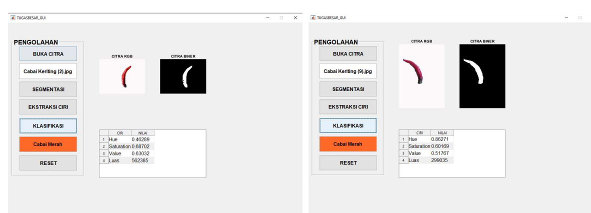

Figure 8. Results of Classification of Red Chili Vegetables

In the test results of Figure 8, the resulting classification is appropriate.

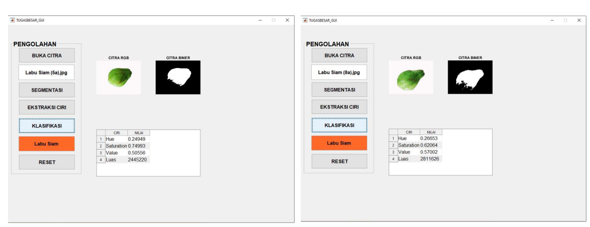

Figure 9. Results of Classification of Siamese Pumpkin Vegetables

In the test results of Figure 9, the resulting classification is appropriate.

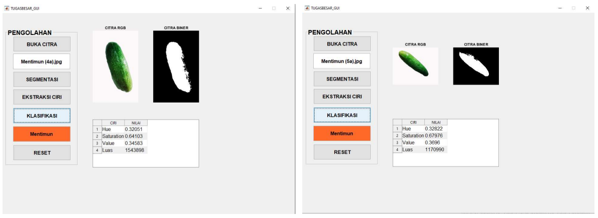

Figure 10. Cucumber Vegetable Classification Test Results

In the test results of Figure 10, the resulting classification is appropriate.

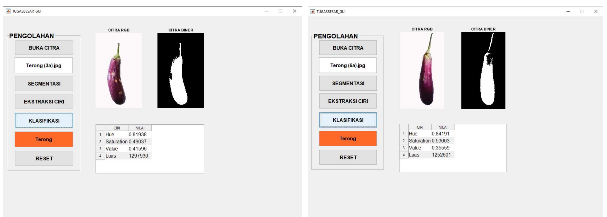

Figure 11. Eggplant Vegetable Classification Test Results

In the test results of Figure 11, the resulting classification is appropriate.

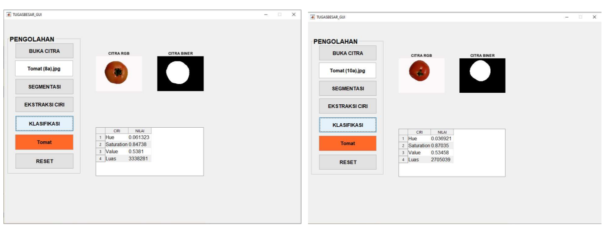

Figure 12. Tomato Vegetable Classification Test Results

In the test results of Figure 12, the resulting classification is appropriate.

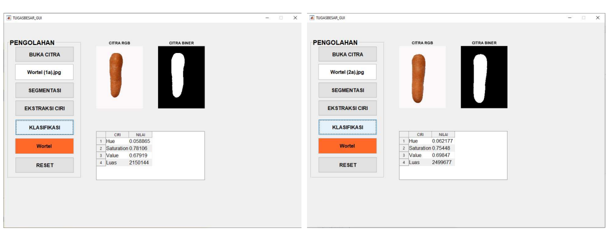

Figure 13. Carrot Vegetable Classification Test Results

In the test results of Figure 13, the resulting classification is appropriate.

The results from Figures 6 to 13 show that from 16 test data, there are 2 green chili vegetable data whose test results do not match, this is because the results of the green chili binary image fall into the red chili binary image segmentation range. This study measures the performance of the system only based on accuracy. The resulting accuracy value is quite high, but this system has not been able to detect the same type of vegetables more specifically. In order for the system to make specific classification differences, it is possible to add an analysis of color differences for the same type of vegetable. It is assumed that the system has a good performance in classifying several types of vegetables.

4. Conclusion

The results of this study obtained satisfactory results. Vegetable image classification can be done well using the proposed method. Image processing includes converting Red Green Blue images into grayscale images, image segmentation using the thresholding method, image feature extraction based on the average value of HSV (Hue, Saturation, Value) and object area, and image classification using the Naive Bayes algorithm. Based on the level of accuracy of the resulting training data is 100% and the level of accuracy of the resulting test data is 87.5%. In this study, it was developed using a GUI. The GUI system used is able to classify the types of vegetables based on the results of segmentation and feature extraction values. This research can be concluded that the classification system using the Naive Bayes algorithm has a fairly good performance. This research can be the basis for other research in building a system to classify other objects.

References

[1] Y. Takaoka and N. Kawakami, "Fruit and vegetable consumption in adolescence and health in early adulthood: a longitudinal analysis of the Statistics Canada's National Population Health Survey," *BMC Public Health*, vol. 13, no. 1, p. 1206, Dec. 2013, doi: 10.1186/1471-2458-13-1206.

[2] 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, Oct. 2020, doi: 10.26418/elkha.v12i2.42160.

[3] Jannah M, Rochmanto R.A, and Afandi M.A, "Tomato Maturity Detection System Using Color Histogram and Nearest Neighbor," *Journal of Applied Information and Communication Technologies*, vol. Vol.7, No.1, 2022, pp. 1–5, 2022, doi: 10.32497/jaict.v7i1.3074.

[4] A. Saputra, "KLASIFIKASI PENGENALAN BUAH MENGGUNAKAN ALGORITMA NAIVE BAIYES," *Jurnal RESISTOR (Rekayasa Sistem Komputer)*, vol. 2, no. 2, pp. 83–88, Oct. 2019, doi: 10.31598/jurnalresistor.v2i2.434.

[5] E. Utama, F. Yapputra, and G. Gasim, "Identifikasi Jenis Mangga Berdasarkan Bentuk Menggunakan Fitur HOG dan Jaringan Syaraf Tiruan," *Jurnal Ilmiah Informatika Global*, vol. 9, no. 1, Aug. 2018, doi: 10.36982/jig.v9i1.437.

[6] Y. Pratama, F. Utaminingrum, and W. Kurniawan, "Implementasi Background Subtraction Untuk Klasifikasi Keripik Kentang Berbasis Raspberry Pi Menggunakan Metode Naive Bayes," 2019. [Online]. Available: http://j-ptiik.ub.ac.id

[7] F. Y. Manik and K. S. Saragih, "Klasifikasi Belimbing Menggunakan Naïve Bayes Berdasarkan Fitur Warna RGB," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 11, no. 1, p. 99, Jan. 2017, doi: 10.22146/ijccs.17838.