Building of Informatics, Technology and Science (BITS)
Volume 5, No 2, September 2023 Page: 523−531
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i2.4192

Tomato Ripeness Detection Using Linear Discriminant Analysis Algorithm with CIELAB and HSV Color Spaces

Rini Nuraini[1,*], Teotino Gomes Soares[2], Popi Dayurni[3], Mulyadi[4]

[1] Informatics Study Program, Faculty of Communication and Informatics Technology, Universitas Nasional, Jakarta, Indonesia
[2] Computer Science Department, School of Engineering and Science, Dili Institute of Technology, Dili, Timor-Leste
[3] Information Technology Education Study Program, Universitas Bina Bangsa, Banten, Indonesia
[4] Information Systems Study Program, Faculty of Computer Science, Universitas Nurdin Hamzah, Jambi, Indonesia
Email: [1,*]rini.nuraini@civitas.unas.ac.id, [2]tyosoares@gmail.com, [3]popi.dayurnipd@binabangsa.ac.id, [4]mulyadiroesly@gmail.com
Correspondence Author Email: rini.nuraini@civitas.unas.ac.id
Submitted: **25/08/2023**; Accepted: **30/09/2023**; Published**: 30/09/2023**

Abstract−Tomatoes have a relatively short ripening period, making it essential to identify their ripeness level before distribution. The ripeness level of tomatoes can be detected based on their color. Therefore, the color of tomatoes serves as a crucial indicator in determining whether they are ripe and of good quality. However, classifying tomato ripeness levels manually has several drawbacks, namely requiring a long process, a low level of accuracy, and being inconsistent. The research aimed at developing a detection model for the ripeness level of tomatoes using the LDA algorithm based on color feature extraction, namely CIELAB (L*a*b) and HSV. The L*a*b and HSV color spaces are applied to obtain information about the color of the object being detected. Furthermore, the information obtained from feature extraction is then grouped by class using the LDA algorithm, which separates information for each class and limits the spread between classes through linear projection searches to maximize the covariance matrix between classes so that members within the class can be identified. This research produces a model that can detect the level of ripeness of tomatoes with an accuracy of 88.194%.

Keywords: Color Features; CIELAB; HSV; L*a*b; Linear Discriminant Analysis; Tomato Ripeness Levels

1. INTRODUCTION

Today, the technology of digital image processing is employed to support human work. Digital image processing is related to the processes of management, analysis, improvement, and other things that are done to get information from an image so that it can be utilized [1]. One of the important parts of a digital image is color, because color is something that is visible visually for the first time when the object is successfully acquired. In an image, there is a color that contains information and represents what is in the image [2]. Image processing can be used to group certain images into several classes, and each class represents entities that have something in common [3]. Digital image processing has been utilized in various fields, including agriculture. Research related to agriculture is currently a frequently discussed topic, for instance, in determining fruit ripeness levels. Tomatoes are one of the fruits whose ripeness level tends to be shorter [4]. For this reason, before distribution, it is necessary to first identify the level of ripeness of the tomatoes, which is generally seen from their color. Thus, the color of the tomatoes is an important indicator in determining the level of ripeness and quality of the tomatoes. The classification of tomato ripeness levels aims to reduce the risk of rotting tomatoes. However, classifying tomato ripeness levels manually has several drawbacks, namely requiring a long process, a low level of accuracy, and being inconsistent. This is due to the subjective determination of workers. There is an automatic grouping of tomato ripeness levels that can be faster and more precise.

Several previous researchers have conducted research on the detection of fruit ripeness, particularly tomatoes. The first research is about detecting the ripeness of tomatoes using HIS color features and the K-Nearest Neighbor (K-NN) approach [5]. This study produced an average value for all test cases, namely 87.80%. This study applies feature extraction with HIS color based on quantization of hue, saturation, and intensity and uses K-NN to classify classes at the three levels of ripeness of the tomatoes used. However, the K-NN approach has a dependency on the feature values obtained, so if the results experience redundancies, it will affect the resulting accuracy [6]. Furthermore, research on the identification of ripeness in tomatoes was carried out by implementing the Support Vector Machine (SVM) approach, and the features used were CIELAB colors [7]. In this study, the ripeness level is divided into six levels with an accuracy of 83.39%. The CIELAB color space can fully represent color, while the SVM classifier algorithm performs grouping based on the optimal hyperplane by considering the distance in each class. However, because SVM compares one class with all other classes, it is less effective to apply to sets with a very large number of dimensions. [8]. Subsequent research is about the classification of ripeness levels in tomatoes using the Learning Vector Quantization (LVQ) algorithm based on HSV color features [9]. This research classifies five levels of ripeness in tomatoes with an accuracy value of 87.25%. This study uses the HSV color feature to represent colors that are like what humans see, and then the LVQ algorithm performs grouping by searching for the appropriate weight values in vector grouping in the initialized destination class.

This research has a difference from previous studies, namely that it grouped data based on the Linear Discriminant Analysis (LDA) algorithm. The LDA algorithm has the advantage of being able to separate data into certain classes by optimizing the values contained in those classes [10]. By using redundant features and moving components from a high-dimensional space to a low-dimensional environment, the LDA method may decrease dimensions [11]. This algorithm has the ability to separate information for each class and limit the spread between classes [12]. The feature extraction used in this study is through the CIELAB color feature (L*a*b color space) and

Building of Informatics, Technology and Science (BITS)
Volume 5, No 2, September 2023 Page: 523−531
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i2.4192

HSV (Hue, Saturation, and value). Because these color features can represent colors in objects, it will be easier to group them [13]. In addition, the detection of the ripeness level of tomatoes used is based on "six ripening stages of tomatoes", which consist of six classes, namely: Mature Green, Breaker, Turning, Pink, Light Red, and Red Ripe [14].

Thus, the purpose of this study was to detect the ripeness level of tomatoes using the LDA algorithm with the color features of the CIELAB (L*a*b) and HSV color spaces. The L*a*b and HSV color spaces are applied to obtain information about the color of the object being detected. Furthermore, the information obtained from feature extraction is then grouped by class using the LDA algorithm, which separates information for each class and limits the spread between classes through linear projection searches to maximize the covariance matrix between classes so that members within the class can be identified.

2. RESEARCH METHODOLOGY

2.1 Research Stages

Before conducting research, it is necessary to arrange a research phase that aims to fulfill the objectives and be organized. This research phase is visualized in Figure 1.

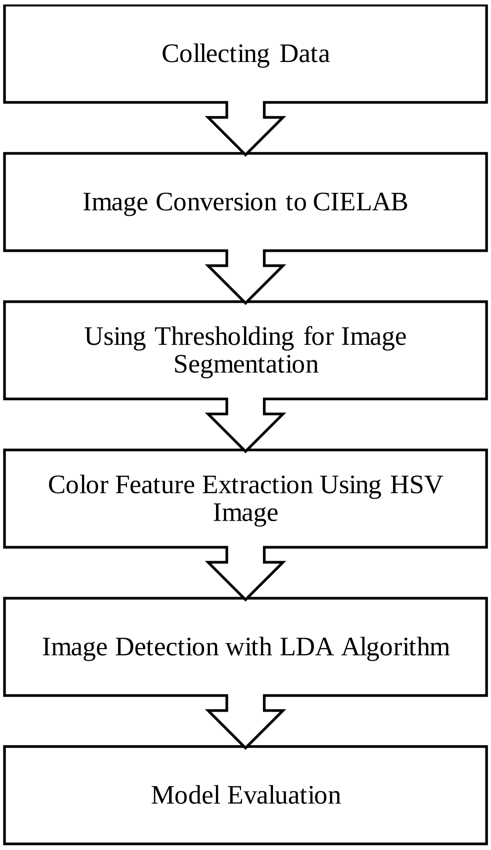


Figure 1. Research Phase

See the section below for a more thorough description of how each stage of the study was completed.

2.1.1 Collecting Data

This data collection is related to the image collection, which will later be applied to the dataset. This dataset is useful for model training and for testing models so that they can learn certain patterns and simulate the use of models in the real world [15]. The existence of a dataset is very important because the dataset is a factor that determines the performance of the model being developed [16], [17]. The detection of the ripeness level of tomatoes used is based on "six ripening stages of tomatoes", which consist of six classes, namely: Mature Green, Breaker, Turning, Pink, Light Red, and Red ripe [14]. The maturity level of the tomatoes used is visualized in Figure 2.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 2, September 2023 Page: 523−531
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i2.4192

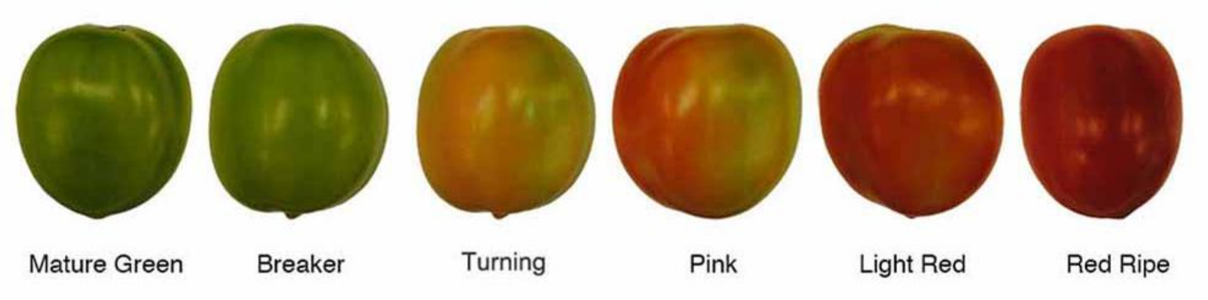


Figure 2. Image Sample Six Stages of Tomato Ripening

The distribution of goods used is 70:30. The percentage for data training is higher than that for data testing so that the model can optimally learn from existing patterns [18]. There are 80 images of tomatoes used for the six classes, so the total number is 480 images. These images are then used as training data, with as many as 336 images and 144 images for data testing.

2.2.2. Image Conversion to CIELAB

The CIELAB color space, sometimes called CIE L*a*b, uses three integer values to represent color: L* stands for light intensity, a* and b* for green-red and blue-yellow components, respectively [19]. This step is done in order to define each color in the object. The identity of a color is determined by the wavelength of light, where the wavelength that can still be captured by the human eye ranges from 380 to 780 nanometers. This partitioning process is carried out through conversion, transforming the image color space from an RGB image to an XYZ color space. In the XYZ color space, some colors are represented as values that are always positive. The calculation for the transformation from RGB to XYZ color space (with a white reference value) is done through the calculation of the transformation matrix. The conversion from RGB to XYZ can be seen in equations (1), (2), and (3).

$$[X] = [0.412453 \quad 0.357580 \quad 0.180423][R] \tag{1}$$

$$[Y] = [0.212671 \quad 0.715160 \quad 0.715160][G] \tag{2}$$

$$[Z] = [0.019334 \quad 0.119193 \quad 0.950227][B] \tag{3}$$

After conversion from RGB to XYZ, the image is then converted into L*, a*, and b* color spaces using equations (4), (5), (6), and (7).

$$L^* = 116\left(\frac{Y}{Y_n}\right)\frac{1}{3}16\, for\ \frac{Y}{Y_n} > 0.008856 \tag{4}$$

$$L^* = 903.3\,\frac{Y}{Y_n}\ otherwise \tag{5}$$

$$a^* = 500\left(f\left(\frac{X}{X_n}\right)\right) - f\left(\frac{Y}{Y_n}\right) \tag{6}$$

$$b^* = 200\left(f\left(\frac{Y}{Y_n}\right)\right) - f\left(\frac{Z}{Z_n}\right) \tag{7}$$

2.2.3. Using Thresholding for Image Segmentation

The goal of the image segmentation procedure is to distinguish particular things from other objects in the image [20]. Separating the objects is done based on the boundaries of the area whose shape and arrangement have similarities. The segmentation method utilized is called thresholding, and it needs a limit value called the threshold value. The thresholding process is related to implementing an image transformation with a degree of gray into a binary image, or in this case, the image will be converted to black and white so that it can be known which area is the foreground and which area is the background [21]. The thresholding approach involves dividing the pixels in the image into two groups based on a certain threshold. Pixels with intensity below the threshold will be collected as one object or part, while pixels with intensity above the threshold will be aggregated into another group. The result is a binary image, where the pixels belonging to the object are marked with a certain intensity value (usually white), while the background pixels are marked with another intensity value (usually black). To transform pixel values into binary images, equation (8) is used.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 2, September 2023 Page: 523−531
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i2.4192

$$g(x,y) = \begin{cases} 1, & if\ f(x,y) \geq T \\ 0, & if\ f(x,y) < T \end{cases} \tag{8}$$

where $f(x,y)$ represents a grayscale image, $g(x,y)$ represents a binary image, and T represents a threshold value.

2.2.3. Color Feature Extraction Using HSV Image

Color feature extraction is useful for obtaining visible color information on objects, which is then used to classify an object [22]. Hue, saturation, and value, which are produced from the conversion of RGB picture colors, are used in the feature extraction process known as HSV [23]. The color that HSV has is almost identical to what the retina of the eye captures. The colors used by the HSV to define redness and greenness are true colors such as red, violet, and yellow, represented by hue. Furthermore, saturation, or chroma, is the purity or strength of the color. The color brightness value expressed by the value ranges from 0 to 100%. Values that are black have a value of 0 and will be white if they have a value of 100. This means that the greater the value, the brighter the image will be. The conversion of RGB images into HSV images can be formulated by equations (9), (10), and (11).

$$H = \tan\left(\frac{3(G-B)}{(R-B)+(R-B)}\right) \tag{9}$$

$$S = 1 - \frac{\min(R,G,B)}{V} \tag{10}$$

$$V = \frac{R,G,B}{3} \tag{11}$$

where, H is hue, S is saturation, V is value. While R is red, G is green and B is blue.

2.2.5. Image Detection with LDA Algorithm

The Linear Discriminant Analysis (LDA) approach is a classifier in which some data from the existing data already have a predefined class or label. To determine the discriminant function, data with known labels are utilized [24]. The goal of the LDA approach is to identify a linear projection, also known as the "Fisher image," that will maximize the inter-class covariance matrix and spread out the class members more widely, hence improving identification success [11]. In the LDA algorithm, the data provided has information about different classes or categories. The main step in the LDA algorithm is to find a new projection that maximizes the differences between class means and, at the same time, minimizes the variation within each class. In other words, the LDA approach seeks to formulate a linear transformation that increases the distance between class means and reduces variation within the same class. The function used to categorize observations into one class, which is known as the discriminant function, is the outcome of this discriminant analysis [25]. Inter-class covariance matrices (S_b) and within-class covariance matrices (S_w) are defined in equations (12) and (13), respectively.

$$S_b = \sum_{i=1}^{k} n_i(m_i - m_o)(m_i - m_o)^T \tag{12}$$

$$S_w = \sum_{i=1}^{k}\sum_{j=1}^{n_i} n_i\,(x_i^{(j)} - m_o)(x_i^{(j)} - m_o)^T \tag{13}$$

The next step is to make a matrix projection that is distributed into classes (S_w) using equation (14).

$$J_2(W) = maxtrace((W^T S_W W)^{-1}(W^T S_b W)) \tag{14}$$

Then proceed with finding the results of the eigenvectors (λ) and the results of the eigenvectors (υ) using equation (15).

$$S_b \upsilon = \lambda S_w \upsilon \tag{15}$$

The eigenvalues (λ) that have been obtained are then sorted from the largest to the smallest value. This process ends by projecting all original images onto the Fisher Base factor through calculations using equation (16).

$$\upsilon^x = V^t x^i \tag{16}$$

Building of Informatics, Technology and Science (BITS)
Volume 5, No 2, September 2023 Page: 523−531
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i2.4192

2.2.6. Model Evaluation

The evaluation phase aims to measure what is achieved by the proposed model [26]. In this phase the model will be evaluated by applying measurements through the confusion matrix. This matrix helps to measure the extent to which the classification model has been successful in classifying data into correct or incorrect classes. Based on the values in the confusion matrix, it can be measured to what extent the model can correctly identify classes and detect discrepancies between predictions and reality. Through the confusion matrix, an assessment of model performance will be carried out through precision, recall, and accuracy [27]. Precision, recall, and accuracy are searched through equations (17), (18), and (19).

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

where TP (True Positive) is positive data that is predicted to be true. Conversely, TN (True Negative) is negative data that is predicted to be true. Meanwhile, FP (False Positive) is negative data but is predicted as positive data. Then, FN (False Negative) is positive data but is predicted as negative data.

3. RESULT AND DISCUSSION

To perform image detection of the maturity level of tomatoes by applying the Linear Discriminant Analysis (LDA) approach, it begins with preparing a dataset that is used to train and test the model. The tomato images used for each class are 80 images, for a total of 480 images. With a 70:30 distribution, 336 images are used to train the model, while 144 images are used to test the model. The detection of the ripeness level of tomatoes used is based on "six ripening stages of tomatoes", which consist of six classes, namely: Mature Green, Breaker, Turning, Pink, Light Red, and Red Ripe. The model is then implemented in MATLAB software, both for training and testing. Training in the developed model begins with the process of converting the color space from RGB images to CIELAB (L*a*b) images. This process is useful for defining any color content in objects. A sample of the converted L*a*b image is presented in Figure 3.

(a) (b)

Figure 3. (a) RGB image; (b) Converted Image to L*a*b Color Space

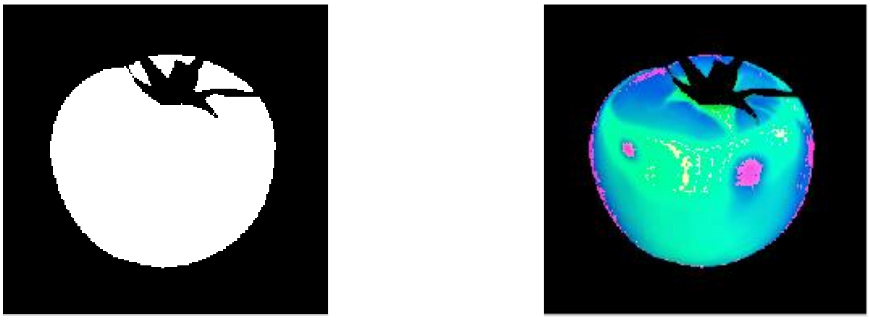
Figure 3 (b) shows the conversion results in the L*a*b color space. These results are then carried out by image segmentation with a thresholding approach, where the image is transformed into a binary image so that the foreground and background can be separated. This process functions to get the foreground object so that it can focus on the object to be detected. An example of segmentation results is presented in Figure 4.

(a) (b)

Figure 4. (a) L*a*b image; (b) Segmentation Image Results

Building of Informatics, Technology and Science (BITS)
Volume 5, No 2, September 2023 Page: 523−531
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i2.4192

As seen in Figure 4 (b), which is a segmented image where the object to be detected is separated from its background. These results are obtained by determining the threshold value so that there are only two values, namely 1 and 0, which represent binary images. After the image has been segmented, the next step is to convert it into the HSV color space so that information can be obtained from the color of the object. The colors used by the HSV to define redness and greenness are true colors such as red, violet, and yellow, represented by hue. Furthermore, saturation, or chroma, is the purity or strength of the color. The results of the segmented image transformation into the HSV color space are presented in Figure 5.



(a) (b)

Figure 5. (a) Segmentation Image Results; (b) HSV Image Results

Figure 5(b) presents an image that has been converted to the HSV color space. Then, in that image, the hue and saturation values are calculated. The hue and saturation values obtained are presented in a table to provide information on the image to be detected. The results of the hue and saturation values of the tested image samples are presented in Figure 6.

	Feature	Value
1	Hue	0.0290
2	Saturation	0.2769

Figure 6. Hue and Saturation values

Figure 6 shows the results of the average hue and saturation values generated for the object to be detected. Then, the information value of the color obtained becomes a parameter to determine the group of the image class using the LDA algorithm. The LDA algorithm will separate data between classes by optimizing the values in each class. By using redundant features and moving components from a high-dimensional space to a low-dimensional environment, the LDA method decreases dimensions. Then this algorithm separates information for each class and limits the spread between classes. Data whose label is known is used to find the discriminant function. Next, a search for linear projections is carried out to maximize the covariance matrix between classes so that the members in the class are more evenly distributed. The discriminant function that is formed can be used to separate, so that a group is divided. This results in a linear transformation that can be used to project data down into a lower space, where the differences between classes are more pronounced. The separation between classes is visualized in Figure 7.

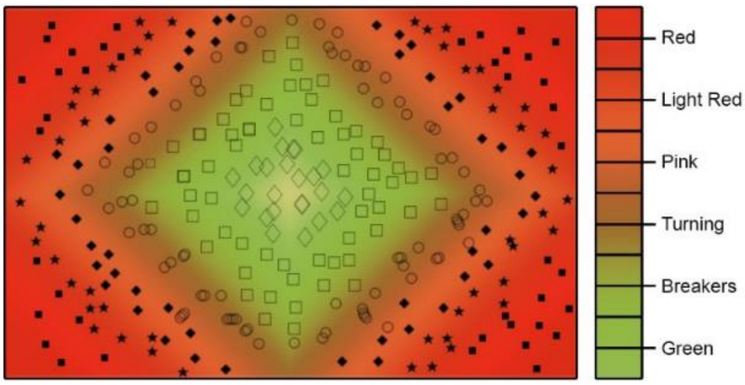


Figure 7. Visualization of Separation Between Classes

Building of Informatics, Technology and Science (BITS)
Volume 5, No 2, September 2023 Page: 523−531
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i2.4192

In Figure 7, you can see the separation between classes for each class or level of ripeness of tomatoes. Each existing class is isolated; this is done so that there is a wide distance between classes so that classes can be formed. The training results from the LDA algorithm have no effect on the number of features produced, which makes the process faster. The model that has been built is then made in the form of a GUI so that it uses MATLAB software that is easy to use and used for testing. The implementation of the model developed in MATLAB is presented in Figure 8.

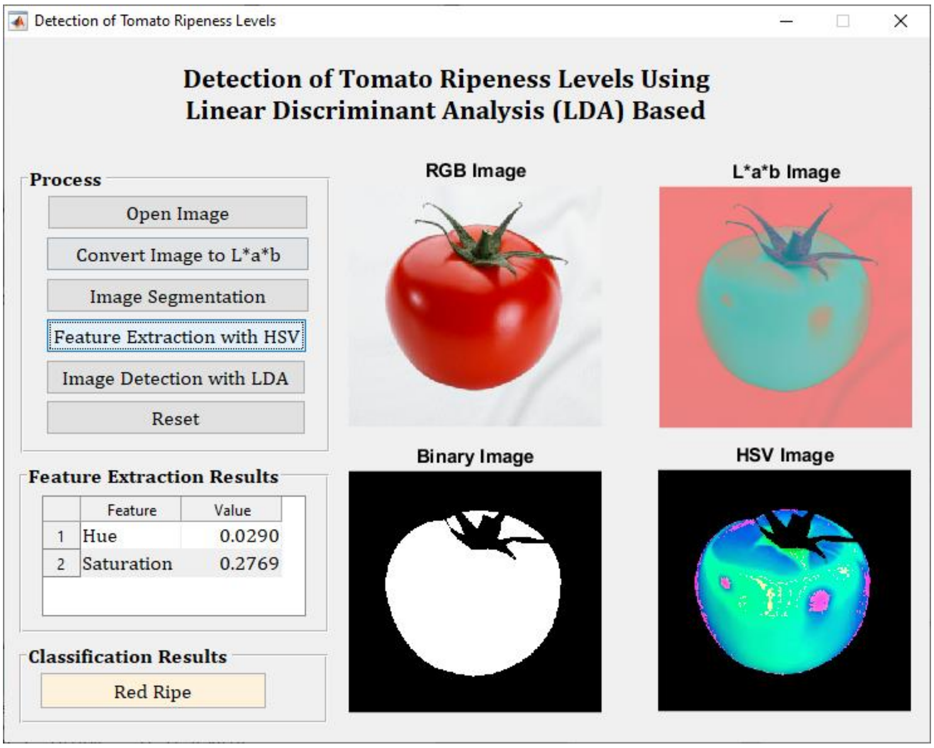


Figure 8. GUI of Tomato Fruit Ripeness Classification System

The model will then be put to the test to gauge its performance via the following model assessment phase. There were 144 images in the image data on the tomato ripeness level utilized for testing. This matrix helps to measure the extent to which the classification model has been successful in classifying data into correct or incorrect classes. Each row in the confusion matrix represents the actual class of data, while each column represents the predicted class by the model. This matrix consists of four cells that represent the combination of the predicted results and the actual class, namely TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative). The findings of the confusion matrix from the experiments that have been carried out are presented in Figure 9.

		Truth data							
		Mature Green	Breaker	Turning	Pink	Light Red	Red Ripe	Classification overall	User's accuracy (Precision)
Classifier results	Mature Green	24	2	0	0	0	0	26	92.308%
	Breaker	2	21	0	0	0	0	23	91.304%
	Turning	0	0	19	2	1	1	23	82.609%
	Pink	0	0	2	20	1	1	24	83.333%
	Light Red	0	0	0	1	20	2	23	86.957%
	Red Ripe	0	0	0	1	1	23	25	92%
	Truth overall	26	23	21	24	23	27	144	
	Producer's accuracy (Recall)	92.308%	91.304%	90.476%	83.333%	86.957%	85.185%		

Figure 9. Results of the Confusion Matrix

Based on the values in the confusion matrix in Figure 9, it is used as a standard to measure how well the model can correctly identify classes and find differences between predictions and reality by making measurements with precision, recall, and accuracy. The precision, recall, and accuracy values are then calculated using the results, and the results are shown in Table 1.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 2, September 2023 Page: 523−531
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i2.4192

Table 1. Test Result

Class Name	Precision	Recall	Accuracy
Mature Green	92.308%	92.308%	88.194%
Breaker	91.304%	91.304%	
Turning	82.609%	90.476%	
Pink	83.333%	83.333%	
Light Red	86.957%	86.957%	
Red Ripe	92.000%	86.957%	

Table 1 presents the results of the test which show that the overall accuracy rate is 88.194%. Then, from the results obtained, it is transformed into the value criteria category with guidelines, namely: if the result is between 76% to 100% it is said to be "Good"; If the result is between 56% to 5% it is said "Enough"; if the result is between 40% to 55% it is said to be "Not Good", and if the result is below 40% it is said to be "Very Bad" [28]. Based on these categories, it can be said that the proposed model meets the "Good" criteria. However, the results of the tests carried out showed that the value of the "Turning" class produced the lowest precision value, namely 82,609%, and the "Pink" class obtained the lowest recall value, namely 83,333%. This is because this class has similarities with other classes when viewed through its color features. However, as a whole, the model built produces better accuracy when compared to previous studies, where the K-Nearest Neighbor algorithm had an accuracy of 87.80% [5], the SVM algorithm had an accuracy of 83.39% [7], and the LVQ algorithm had an accuracy of 87.25% [9]. These results indicate that the LDA algorithm can do good detection because it searches for new projections that can maximize the differences between class averages and, at the same time, minimize variations within each class. In other words, the LDA algorithm attempts to formulate a linear transformation that increases the distance between class means and reduces within-class variation. The result of the LDA approach is a linear transformation that can be used to project data into a lower space, where the differences between classes are more pronounced.

It should be noted that, based on the results of the tests carried out, the error rate reached 11,806%. These results are influenced by several aspects, including: 1) The LDA approach detects by reducing data without regard to other information, so it is vulnerable to outlier influences that can affect the calculation and class separation in the projection; 2) The class used has very identical warrant features, so it is necessary to consider using other features that are more representative; 3) Image data that has a variety of backgrounds with several camera angles makes it difficult for the model to process it, so it needs pre-processing steps before being used to get better data; 4) The dataset used is still on a small scale, so it is necessary to try a larger dataset so that the model gets maximum learning patterns.

4. CONCLUSION

This study has developed a model for detecting the ripeness level of tomatoes using the LDA algorithm with color space features CIELAB and HSV. The CIELAB and HSV color spaces are applied to obtain information about the color of the object being detected. Furthermore, the information obtained from feature extraction is then grouped by class using the LDA algorithm, which separates information for each class and limits the spread between classes through linear projection searches to maximize the covariance matrix between classes so that members within the class can be identified. In other words, the LDA algorithm attempts to formulate a linear transformation that increases the distance between class averages and reduces variation within the same class. Based on the test results, the accuracy value is 88.194% and is in the good category. There are various enhancements that may be done for future study. The LDA algorithm performs detection by reducing data without regard to other information, so it requires a combination of algorithms that are able to utilize the information obtained from the reduction results. In addition, it is necessary to try the application of deep learning to extract abstract features at various levels so as to be able to solve problems with high levels of complexity. This will support a more representative learning pattern, and accuracy can be improved.

REFERENCES

[1] E. R. Dougherty, *Digital Image Processing Methods*. New York: CRC Press, 2020.
[2] V. C. S. Rao, S. Venkratamulu, and P. Sammulal, *Digital Image Processing and Applications*. Singapore: Horizon Books (A Division of Ignited Minds Edutech P Ltd), 2021.
[3] R. I. Borman, F. Rossi, Y. Jusman, A. A. A. Rahni, S. D. Putra, and A. Herdiansah, “Identification of Herbal Leaf Types Based on Their Image Using First Order Feature Extraction and Multiclass SVM Algorithm,” in *International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 2021, pp. 12–17.
[4] P. Skolik, C. L. M. Morais, F. L. Martin, and M. R. Mcainsh, “Determination of developmental and ripening stages of whole tomato fruit using portable infrared spectroscopy and Chemometrics,” *BMC Plant Biol.*, vol. 19, no. 236, pp. 1–15, 2019.
[5] D. Ayuningtyas, E. Suryani, and W. Wiharto, “Identification of Tomato Maturity Based on HIS Color Space Using The K-Nearest Neighbour Method,” in *International Conference on Artificial Intelligence and Computer Science Technology (ICAICST)*, 2021, pp. 73–78. doi: 10.1109/ICAICST53116.2021.9497843.
[6] M. C. Untoro, M. Praseptiawan, M. Widianingsih, I. F. Ashari, A. Afriansyah, and O. Oktafianto, “Evaluation of Decision Tree, K-NN, Naive Bayes and SVM with MWMOTE on UCI Dataset,” in *ICComSET*, 2019, pp. 1–8. doi: 10.1088/1742-

Building of Informatics, Technology and Science (BITS)
Volume 5, No 2, September 2023 Page: 523−531
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i2.4192

6596/1477/3/032005.

[7] M. B. Garcia, S. Ambat, and R. T. Adao, "Tomayto , Tomahto: A Machine Learning Approach for Tomato Ripening Stage Identification Using Pixel-Based Color Image Classification," 2019.

[8] S. Y. Chaganti, I. Nanda, K. R. Pandi, T. G. N. R. S. N. Prudhvith, and N. Kumar, "Image Classification using SVM and CNN," in *International Conference on Computer Science, Engineering and Applications (ICCSEA)*, 2020, pp. 1–5. doi: 10.1109/ICCSEA49143.2020.9132851.

[9] S. Sanjaya, "Application of Learning Vector Quantization in Classifying Levels of Ripeness of Tomatoes Based on Fruit Color," *CoreIT*, vol. 5, no. 2, pp. 49–55, 2019.

[10] C. Deepika, R. P. Gnanamalar, K. Thangaraj, N. Revathy, and A. Karthikeyan, "Linear discriminant analysis of grain quality traits in rice (Oryza sativa L.) using the digital imaging technique," *J. Cereal Sci.*, vol. 109, p. 103609, 2023, doi: 10.1016/j.jcs.2022.103609.

[11] S. Li, H. Zhang, R. Ma, J. Zhou, J. Wen, and B. Zhang, "Linear discriminant analysis with generalized kernel constraint for robust image classification," *Pattern Recognit.*, vol. 136, p. 109196, 2023, doi: 10.1016/j.patcog.2022.109196.

[12] R. Nuraini, "Identification of Freshwater Fish Types Using Linear Discriminant Analysis (LDA) Algorithm," *IJICS (International J. Informatics Comput. Sci.*, vol. 6, no. 3, pp. 147–154, 2022, doi: 10.30865/ijics.v6i3.5565.

[13] R. I. Borman, Y. Fernando, and Y. E. P. Yudoutomo, "Identification of Vehicle Types Using Learning Vector Quantization Algorithm with Morphological Features," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 158, pp. 339–345, 2022.

[14] J. Michael, "Six Ripening Stages Of Tomatoes By The Fruit Surface Color," *FAGS: Organic Farming and Gardening School*, 2020.

[15] A. Mulyanto, E. Susanti, F. Rossi, W. Wajiran, and R. I. Borman, "Penerapan Convolutional Neural Network (CNN) pada Pengenalan Aksara Lampung Berbasis Optical Character Recognition (OCR)," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 7, no. 1, pp. 52–57, 2021.

[16] A. Mulyanto, W. Jatmiko, P. Mursanto, P. Prasetyawan, and R. I. Borman, "A New Indonesian Traffic Obstacle Dataset and Performance Evaluation of YOLOv4 for ADAS," *J. ICT Res. Appl.*, vol. 14, no. 3, pp. 286–298, 2021.

[17] A. Mulyanto, R. I. Borman, P. Prasetyawana, and A. Sumarudin, "2d lidar and camera fusion for object detection and object distance measurement of ADAS using robotic operating system (ROS)," *Int. J. Informatics Vis.*, vol. 4, no. 4, pp. 231–236, 2020, doi: 10.30630/joiv.4.4.466.

[18] A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 Relation Between Training and Testing Sets : A Pedagogical Explanation," *Sch. UTEP*, vol. 2, pp. 1–6, 2018.

[19] J. Pardede, M. G. Husada, A. N. Hermana, and S. A. Rumapea, "Fruit Ripeness Based on RGB , HSV, HSL, L*a*b* Color Feature Using SVM," in *International Conference of Computer Science and Information Technology (ICoSNIKOM)*, 2019.

[20] I. Ahmad, Y. Rahmanto, R. I. Borman, F. Rossi, Y. Jusman, and A. D. Alexander, "Identification of Pineapple Disease Based on Image Using Neural Network Self-Organizing Map (SOM) Model," in *International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 2022.

[21] H. Mayatopani, R. I. Borman, W. T. Atmojo, and A. Arisantoso, "Classification of Vehicle Types Using Backpropagation Neural Networks with Metric and Ecentricity Parameters," *J. Ris. Inform.*, vol. 4, no. 1, pp. 65–70, 2021, doi: 10.34288/jri.v4i1.293.

[22] R. I. Borman, R. Napianto, N. Nugroho, D. Pasha, Y. Rahmanto, and Y. E. P. Yudoutomo, "Implementation of PCA and KNN Algorithms in the Classification of Indonesian Medicinal Plants," in *ICOMITEE 2021*, 2021, pp. 46–50.

[23] U. Erkut, F. Bostancıoğlu, M. Erten, A. M. Özbayoğlu, and E. Solak, "HSV Color Histogram Based Image Retrieval with Background Elimination," in *International Informatics and Software Engineering Conference (UBMYK)*, 2019.

[24] A. Nguyen *et al.*, "A computationally efficient crack detection approach based on deep learning assisted by stockwell transform and linear discriminant analysis," *Structures*, vol. 45, pp. 1962–1970, 2022, doi: 10.1016/j.istruc.2022.09.107.

[25] F. Hariadi and R. R. H. Enda, "Face Detection Using Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) Methods," *JOINCS (Journal Informatics, Network, Comput. Sci.*, vol. 2, no. 1, pp. 1–4, 2019.

[26] Z. Abidin, R. I. Borman, F. B. Ananda, P. Prasetyawan, F. Rossi, and Y. Jusman, "Classification of Indonesian Traditional Snacks Based on Image Using Convolutional Neural Network (CNN) Algorithm," in *International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 2022, pp. 18–23.

[27] R. I. Borman, F. Rossi, D. Alamsyah, R. Nuraini, and Y. Jusman, "Classification of Medicinal Wild Plants Using Radial Basis Function Neural Network with Least Mean Square," in *International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 2022.

[28] R. I. Borman, Y. Fernando, and Y. Egi Pratama Yudoutomo, "Identification of Vehicle Types Using Learning Vector Quantization Algorithm with Morphological Features," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 2, pp. 339–345, 2022, doi: 10.29207/resti.v6i2.3954.