Journal of Science and Applied Engineering (JSAE)
Vol. 6, No. 1, June 2023
ISSN Print : 2621-3745
ISSN Online : 2621-3753
DOI : 10.31328/jsae.v6i1.4655
(Page.22-30)

IMAGE CLASSIFICATION OF TEMPE FERMENTATION MATURITY USING NAÏVE BAYES BASED ON LINEAR DISCRIMINANT ANALYSIS

Dio Amin Putra[1], *Istiadi*[2,*], *Aviv Yuniar Rahman*[3]

[1,2,3]*Department of Informatics Engineering, Universitas Widyagama Malang, Malang, Indonesia*
**Email: istiadi@widyagama.ac.id*

Submitted : 16 May 2023; *Revision :* 13 June 2023; *Accepted :* 20 June 2023

ABSTRACT

One of the foods in Indonesia that has a lot of nutritional content and benefits, one of which is tempe. Tempe is usually made by fermenting soybeans with mold under special conditions to become tempe. In the fermentation process, tempe producers need to monitor the maturity of the tempe until it is suitable for consumption. To detect this maturity requires a separate effort, so that an image processing approach is proposed in this study with the support of feature selection. An image allows for various features to be taken, such as texture features using GLCM and various color features including RGB, HSV, LAB, CMYK, YUV, HCL, HIS, LCH. With so many features, it is necessary to do a selection so that computation in its classification becomes efficient. This study aims to produce and classify tempe fermented images using the Naive Bayes method with Linear Discriminant Analysis (LDA) feature selection for GLCM features and eight color features and compare systems using LDA and not using LDA. This study uses image processing which requires many features to improve research accuracy and uses feature selection to summarize it in order to speed up tempe detection calculations. Tempe fermentation image is divided into three classes, namely raw, ripe and overripe. Based on the experimental results, the average accuracy in the test is 84.06% while the Naïve Bayes feature without selection features is 79.63%, a difference of 4.43%.. In testing the fastest time is 1.87 seconds and the longest is 2.20 seconds. This shows that the classification of fermented tempe maturity with Naive Bayes with LDA feature selection can work well.

Keywords Tempe; Naïve Bayes; LDA; Color Features
Paper type Research paper

INTRODUCTION

Tempe is one of Indonesian traditional food that has many nutritional contents and benefits that are very popular with the people of Indonesia, so many people use tempe to be processed into other products. Used to optimize people's income and to prosper the surrounding community. There are so many people who turn tempe into tempe chips, to find out the doneness of tempe, it is usually produce by fermentation into tempe and using specifically temperature and Usually Tempe entrepreneurs traditionally do that by drying at room temperature and hanging on the wall [1]. Usually all a long this time fermented tempe made from the fungus Rhizopus sp is a well-developed mold with an perfect growth climate of 28-35 °C and evaporation below 65-70%. The optimum temperature and humidity to support tempe fermentation should be between 30°C-35°C and 60%-70% RH[1]. Meanwhile, the temperature in Malang between 18°C and 23°C is not conducive to mold growth. And the way to solve this problem is to research using incubators designed to regulate temperature and humidity [2][3]. Tempe fermentation with temperatures between 25 °C -37 °C with the best relative humidity at 70-80 % [4], but it still has to be checked so that the doneness of the tempe is as desired.

Research in [5], which detected tempe maturity using K-Nearest Neighbors Algorithm (KNN) and Principal component analysis (PCA) resulted in an accuracy of 80.63% and a time spent of 1.06 seconds which has the potential to improve performance. Research in [6], which detects tempe maturity using YOLO Tiny V4 has a maximum mean average precision (mAP) of 1.00 but requires heavy computation so it is necessary to use Google Colaboratory. Therefore, the authors propose another approach, namely the use of the Naive Bayes Classifier Method. Naïve Bayes Classifier is a classification method uses probability and statistical methods, namely predicting future opportunities based on previous experience.

Furthermore, in image processing for classification, feature extraction is required.. To detect the maturity of tempe, researchers can use a method that involves feature extraction using GLCM (Gray-Level Co-occurrence Matrix) and color selection. GLCM is a powerful feature extraction method that can capture textural information from an image. Color selection, on the other hand, can help to distinguish between different maturity levels of tempe based on color changes. Using various color spaces such as HSV, LAB, CMYK, YUV, HCL, HIS, and LCH can help to improve the accuracy of tempe maturity detection. However, it can also lead to the problem of the curse of dimensionality [5], where the feature space becomes too large. To overcome this problem, dimensionality reduction techniques such as Linear Discriminant Analysis (LDA) can be used to select the most informative features that contribute the most to the classification task.

LDA is a linear transformation technique that can reduce the dimensionality of the feature space while preserving the maximum amount of class separability. It can select the features that maximize the between-class scatter and minimize the within-class scatter, thus providing the most discriminative power for the classification task. Comparing LDA to Principal Components Analysis (PCA), LDA tends to provide better accuracy since it takes into account the class labels and their distribution, whereas PCA is an unsupervised method that only considers the variance of the data. Therefore, LDA is more suitable for classification tasks where the goal is to distinguish between different classes. After the feature selection process using LDA, the Naive Bayes method can be used to assess the accuracy of the classification model. This study aims to classify the image of tempe fermentation maturity using the Naïve Bayes method using LDA.

METHOD

Stages of research based on Figure 1 is to involves assembling data or images of tempe at different stages of maturity. The second step is preprocessing, where the images are cropped and processed to remove noise and make them sharper. The third step is feature extraction, where color features such as RGB, HSV, LAB, CMYK, YUV, HCL, HIS, and LCH are used to segment the image and obtain binary values that can be used for processing. The color features are used to detect the maturity of tempe, and LDA is used to analyze and select the most relevant features. The Naïve Bayes method is used to assess the accuracy of the selected features. Overall, it seems that this process involves a combination of image analysis and machine learning techniques to classify the maturity of tempe based on color features. It may be useful in quality control or production processes for tempe production.

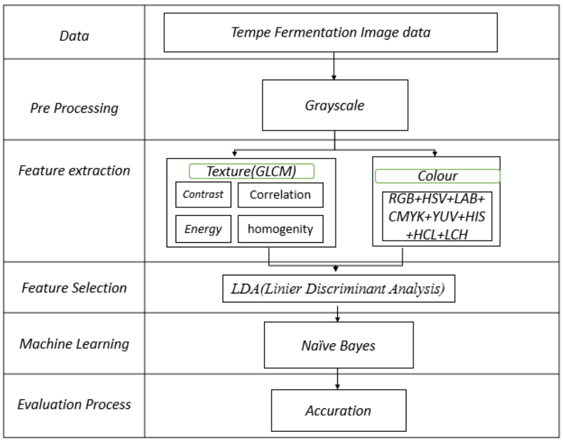


Figure 1. Research Framework

In Figure 1, it is explained that the first step is to collect data in the form of tempe images. The tempe image data is grouped into raw (*mentah*), ripe (*matang*), rotten (*busuk*) classes. The second step is Preprocessing where the data is cropped so that the image looks clearer and natural so as to reduce the effect of noise which results in the system not optimally segmenting textures using GLCM.

Then feature extraction is a process in which segmentation and color feature testing is carried out to get the binary value contained in an image to be processed. At this stage color features are used in the form of RGB, HSV, LAB, CMYK, YUV , HCL, HIS, LCH to detaxify color images on tempe. This is to find out the type of tempe maturity by adding the expected color selection so that it can be an additional parameter for the selection feature being tested.

The next stage is feature selection, which is the process of selecting several features that are used to select the parameters that have the most influence on the features that have been tried using Linear Discriminant Analysis (LDA). This stage is the part where the classification determines the type of tempe to be studied, in this case the Naïve Bayes machine learning method is used. The last one is evaluating, which is the stage where the process of assessing the classification process that has been carried out both in terms of strengths and weaknesses, at this stage everything related to the system is revised and then repaired and produces accuracy.

Tempe Fermentation Image Data

TABLE I. SAMPLE IMAGE

Raw			
Ripe			
Rotten			

Examples of tempe images used for this study are shown in Table I which contains raw tempe, cooked tempe, and rotten tempe. The data used to classify tempe image data includes 137 raw data, 137 ripe data, as well as 136 rotten data as many as 410 tempe data. Data from tempe producers is stored to retrieve training data and test data.

Image Preprocessing

The first thing to do is to take a picture of tempe which is then extracted into grayscale and calculated the value. Grayscale is a matrix data with value represents the concentration of each pixel between 0 and 255. All pixel represents 8 bits of memory in a grayscale image near to some of the captured pixel values. Since each pixel in a grayscale image is represented by 8 bits of memory, the total amount of memory required to store a grayscale image is typically equal to the number of pixels multiplied by 8 bits.

Features Extraction

The RGB model is the percentage of the total visible (human-visible) color spectrum that can be formed or represented by mixing red, green, with blue colors with illumination scales in different proportions, rates and intensities (ovarlap). In extension to the RGB , there is also an RGB normalize model where this model having three components, there is r the red amount, g is the green amount, and b is the blue amount which represents one hundred pixels in a digital image. [7] in the form of RGB and will be lowered and produce other colors, namely, HSV, LAB, CMYK, YUV, HCL, HIS, LCH.

One of the various shadeation structures human beings use to pick out colorations is **HSV** (Hue, Saturation, Value). This shadeation version is regularly utilized in area of the RGB version due to the fact it's far towards how the human eye describes and perceives colors. The HSV color area version is the form transformation of an RGB color block alongside the grey axis (the diagonal axis of the mixture of black and white dots), developing a cone or cone on the palette. Moving along the vertical axis (gray), the hexagonal layer whose dimensions are perpendicular to the axis is volume. (Hue) H is represented by the position angle of the color on the axis of the cone circle. The price of

V is measured alongside the vertical axis of the cone. Saturation S shows the diploma of saturation (white mild content) or the color purity price alongside the radian cone [8].

LAB As a three-dimensional model, it can only be described in three-dimensional space [9], and when slices of components a* and b* are taken, then you get a chroma histogram a*b*. With CIELAB, meaning is given from each formative dimension:

a The magnitude of CIE_L* describes the intensity of a color, 0 for black and L* = 100 for white).
b The size of CIE_a* describes the type of green red color, negative number a*: green color; CIE_a* positive indicates a red color.
c The CIE_b* dimensions for the blue-yellow color type, the negative number indicates blue and the positive CIE_b* indicates yellow.

CMYK color space includes cyan, magenta, yellow, and black. C,M,Y color values can be expressed as decimals between 0 and 1, and Black color values can be expressed as numbers between 0 and 255 [10]. **YUV** color space consists of Y which is the luma appreception and you and V as chroma reception appreception. All three colors of YUV are independent of each other [10]. **HIS** color features are defined as Hue (H), Saturation (S), Intensity (I) terminology. The primary color is described by Hue, which is determined by the dominant wavelength in the spectral distribution of the light wavelength. [11].

HCL color space (Hue, Chroma, Luminance) was advanced by accomodating the improvement of the HSV and HSL color while compensating for the shortcomings of both. The advantage of this color area is that the H components (color) remains conssistent alike when the luminous intensity or saturation of the matter changes. [12]. **Local Color Histogram** divides the picture into numerous elements after which obtains a color histogram for every part. LCH contains more image information, but this method requires more computer processing. When extracting using a local color histogram, the first step is to group the images into blocks and get a local color histogram for each block. Letter two compare blocks at the same position of two images (the distance between two blocks is the distance between their color histograms) third Calculate the sum of the distances of all blocks [13].

Linear Discriminant Analysis (LDA) Feature Selection

LDA is a classifier where from the existing data there is some data that is already known to be the class or label. Data already known labels are used to find the discriminant function. The purpose of the LDA method Is to locate linear projections (referred to as 'fisherimage') to boost the between-class covariance matrix (between-class covariance matrix) so that class members are more involved in their roles and can ultimately increase the success of the introduction [8].

The formed discriminant function serves to separate groups. The output of this discriminant analysis is to obtain a function that is used to group observations into one of the cllasses, which is then called the discriminant function. The covariance matrices in the class (Sw) and the inter-class covariance matrix (Sb) are defined as equations (1) and (2).

$$S_W = \sum_{i=1}^{C} \sum_{X_k \in_i} (X_k - \mu_i)(X_k - \mu_i)^{T} \qquad (1)$$

$$S_B = \sum_{i=1}^{C} N_i(\mu_i - \mu)(\mu_i - \mu)^{T} \qquad (2)$$

Naïve Bayes Classifier

Naïve Bayes Classifier Is a separate method and can be called the Bayes theorem which has been considered to be able to predict the future based on the past [14]. Naïve Bayes has its own advantages and disadvantages, the advantages of using the Naïve Bayes classifier in classifying documents are seen from the way it works which take action based on available data. Therefore, the document classification with their application can be personalized in another sense, the way document classification works can be conditioned according to the characteristics and needs of each person. This advantage can be felt clearly in the implementation of spam filtering that has been tested previously [15]. One of the machine learning methods that uses probabilistic computing is Naïve Bayes which utilizes simple statistics along with probability methods that assume that one another is not interdependent and can be expressed from equation (3) X is unknown class data. H is the data hypothesis X which is a specific class. Next P(H| X) is the probability hipotesis I based on condition X (Posteriori probability). P(H) is the probability of hypothesis H (prior probability). P(X|H) Probability based on conditions on hypothesis H:

$$P(H|X) = \frac{P(H|X).P(H)}{P(X)} \tag{3}$$

Evaluation Process

Evaluation is a method or process of achieving measurable value. This method is used to predict the predictability of the prediction model during its run in practice [16]. Accuracy is the scale of accurate guess (negative and positive) on the total data, (TP) is true positive, (TN) is true negative, (FP) is false negative, (FN) Is false negative.

$$Accuracy = \frac{(\text{TP} + \text{TN})}{(\text{TP+FP+FN+TN})} x100\% \tag{4}$$

RESULT AND CONCLUSION

The data used to classify tempe image data includes 137 unripe data, 137 ripe data, also 136 overripe data as many as 410 tempe data figure from tempe producers are stored for collection. to retrieve training data and test data.

TABLE II. TRAINING NAÏVE BAYES

Split Ratio	Naïve Bayes			
	Data		Result	
	Training	Testing	Accuracy	Time/seconds
10;90	41	369	82.11%	2.2
20;80	82	328	83.73%	1.99
30;70	123	287	90.86%	1.95
40;60	164	246	89.43%	2.06
50;50	205	205	90.56%	1.97
60;40	246	164	92.68%	2.04
70;30	287	123	90.94%	2.47
80;20	328	82	90.04%	2.12
90;10	369	41	91.86%	2.08

The conclusion of the training data using the Naïve Bayes approach and can be ilustrated in Table II. The Table shows that the outcome of the training input data using the Naïve Bayes Uniform approach obtain the lowest exact value, at ratio divided by 10:90 with 82.11% accuracy with a learning time of 2.2 seconds. The highest accuration value is 60:40 with an accuracy of 92.68% and a learning time of 2.04 seconds can be identify in Figure 2.

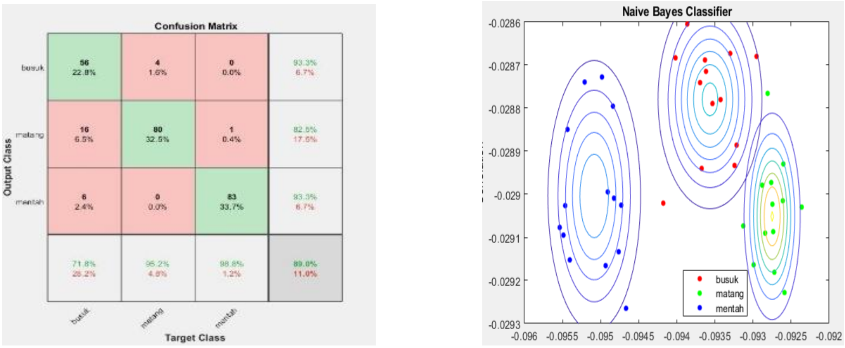
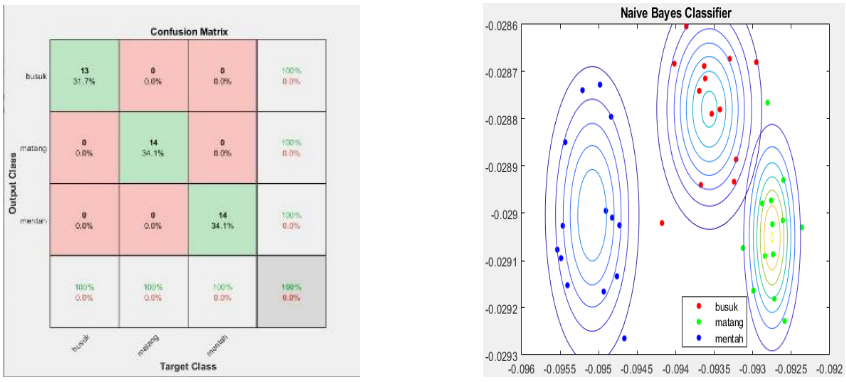


Figure 2. Matrix confusion 60:40 and naïve bayes data spreading

TABLE III. TRAINING NAÏVE BAYES WITH LDA

Split Ratio	Naïve Bayes			
	Data		Result	
	Training	Testing	Accuracy	Time/seconds
10;90	41	369	100.00%	1.99
20;80	82	328	92.68%	1.98
30;70	123	287	90.86%	2.02
40;60	164	246	93.90%	2.07
50;50	205	205	92.84%	2.07
60;40	246	164	93.22%	1.99
70;30	287	123	92.79%	2.21
80;20	328	82	92.27%	2
90;10	369	41	92.23%	2.12

The conclusion of training data using Naïve Bayes approach applying LDA and can be ilustrated in Table III. In the Table explain that the results of training data accepting Naïve Bayes approach obtain the lowest exact value, namely at a division of 30:70 with an accuracy of 90.86 and a test time of 2.02 seconds. The highest accuracy value is at 10:90 with an accuracy of 100% and a test time of 1.99 seconds shown in Figure 3.



Confusion Matrix

Output Class	busuk	matang	mentah	
busuk	13 31.7%	0 0.0%	0 0.0%	100% 0.0%
matang	0 0.0%	14 34.1%	0 0.0%	100% 0.0%
mentah	0 0.0%	0 0.0%	14 34.1%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%

Figure 3. Matrix confusion 10:90 and naïve bayes spreading

TABLE IV. TESTING NAÏVE BAYES

Split Ratio	Naïve Bayes			
	Data		Result	
	Training	Testing	Accuracy	Time/seconds
10;90	41	369	58.62%	2.05
20;80	82	328	73.37%	2.00
30;70	123	287	75.14%	2.03
40;60	164	246	81.57%	2.13
50;50	205	205	90.56%	1.97
60;40	246	164	90.24%	2.01
70;30	287	123	83.73%	2.00
80;20	328	82	81.30%	2.05
90;10	369	41	75.60%	2.00

The result of testing data using Naïve Bayes approach applying LDA and can be ilustrated in Table IV. The Table explain that the conclusion of the test data applying the Naïve Bayes approach obtain the highest, specific, accurate value is at ratio of 50:50 with an accuracy of 90.56% and a attempt time of 1.97 seconds. The lowest accuracy amount is at a division ratio of 10:90 along with an accuracy of 58.62% and a test time of 2.05 seconds can be examine in Figure 4.

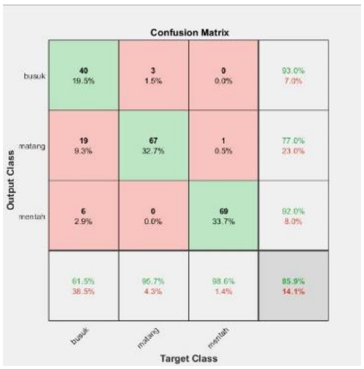


Figure 5. Matrix confusion 50:50 without LDA

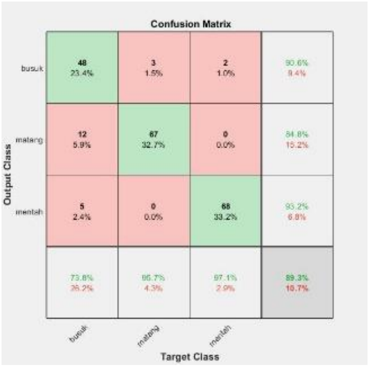


Figure 4. Matrix confusion 10:90 with LDA

TABLE V. TESTING NAÏVE BAYES WITH LDA

Split Ratio	Naïve Bayes			
	Data		Result	
	Training	Testing	Accuracy	Time/seconds
10;90	41	369	70.37%	1.90
20;80	82	328	82.72%	1.95
30;70	123	287	82.57%	1.94
40;60	164	246	86.17%	1.99
50;50	205	205	92.84%	1.95
60;40	246	164	91.05%	1.93
70;30	287	123	87.53%	1.95
80;20	328	82	84.55%	1.96
90;10	369	41	78.86%	1.94

The conclusion of testing data applying the Naïve Bayes and can be seen in Table V. The Table explain the test data results using the Naïve Bayes approach obtain the most accurate value, namely at 50:50 with an accuracy of 92.84% and a test time of 1.95 seconds. The lowest accuracy value corresponding to a division of 10:90 with an accuracy of 70.37% and a testing time of 1.9 seconds can be viewed in Figure 5.

TABLE VI. TIME COMPARISON

Split Ratio	Naïve Bayes	
	With LDA	Without LDA
	Time/seconds	Time/seconds
10;90	1.90	2.05
20;80	1.95	2.00
30;70	1.94	2.03
40;60	1.99	2.13
50;50	1.95	1.97
60;40	1.93	2.01
70;30	1.95	2.00
80;20	1.96	2.05
90;10	1.94	2.00

The results obtained from making comparisons written in Table VI. Show that research using LDA requires a smaller time estimate than the method without using LDA, the lowest time is found at a ratio of 60; 40 1.93 second and the longest is only 1.99 seconds at a ratio of 40; 60. All ratios using LDA is less than those not using LDA.

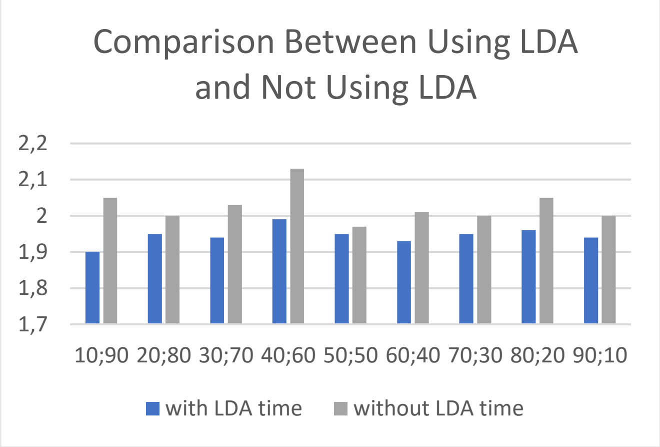


Figure 6. Graphic comparison between using LDA and without LDA

The comparison using LDA and without LDA can be identify on Figure 6. At a ratio of 10;90 to 90;10, it can be compared that LDA use is very influential and takes less time than those who do not use LDA, although there is very little difference, but LDA is superior by a fraction of a second less. The use of LDA in this study resulted in no more than 2 seconds at all ratios.

CONCLUSION

The accuracy in the training is found in the Naïve Bayes feature which is added with a selection feature worth 100% and which is not added a selection feature worth 92.68% at a ratio of 60:40. In the test dataset that has been made, in the Naïve Bayes classification, the average accuracy percentage of accuracy in the test is 84.06%, while the Naïve Bayes feature without selection features is 79.63%, a difference of 4.43%.The average training time of the Naïve Bayes classification without LDA is 2.07 seconds while training using LDA is 2.03 seconds. For tests from LDA usage is 1.97 seconds while testing without LDA is 2.00 seconds.

REFERENCES

[1] R. S. Putri, M. I. Fanani, I. I. Kurniawan, E. P. O. Danawan, K. I. F. Sugiarto, and Istiadi, “Penerapan Teknologi Pengendali Fermentasi Tempe Bagi Usaha Krudel Lariso Kelurahan Purwantoro Kota Malang,” *Conf. Innov. Appl. Sci. Technol. (CIASTECH 2018)*, vol. 9, no. September, pp. 353–361, 2018.

[2] A. H. Rahmawati and D. Harmantyo, “Pola Spasial Suhu Permukaan Daratan di Kota Malang Raya , Jawa Timur,” *Ind. Res. Work. Natl. Semin. Politek. Negeri Bandung*, pp. 548–559, 2017.

[3] B. Gunawan and S. Sukardi, “Rancang Bangun Pengontrolan Suhu dan Kelembaban pada Proses Fermentasi Tempe Berbasis Internet of Things,” *JTEIN J. Tek. Elektro Indones.*, vol. 1, no. 2, pp. 168–173, 2020, doi: 10.24036/jtein.v1i2.63.

[4] D. Wijanarko and S. Hasanah, “Monitoring Suhu Dan Kelembaban Menggunakan Sms Gateway Pada Proses Fermentasi Tempe Secara Otomatis Berbasis Mikrokontroler,” *J. Inform. Polinema*, vol. 4, no. 1, p. 49, 2017, doi: 10.33795/jip.v4i1.144.

[5] I. Istiadi, A. Y. Rahman, and A. D. R. Wisnu, “Identification of Tempe Fermentation Maturity Using Principal Component Analysis and K-Nearest Neighbor,” *Sinkron*, vol. 8, no. 1, pp. 286–294, 2023, doi: 10.33395/sinkron.v8i1.12006.

[6] B. Setyawan, A. Y. Rahman, and Istiadi, “Detection of Tempe Maturity Quality Using Yolo Tiny V4,” pp. 456–460, 2023.

[7] D. L. Parolinda and A. M. Ramdan, “Perbandingan Kualitas Citra BMP Steganografi dengan Ruang Warna RGB dan CMYK,” no. December, 2019.

[8] R. P. Sari, U. D. Rosiani, and A. R. Syulisttyo, “Implementasi Metode Linear Discriminant Analysis Untuk Deteksi Kematangan Pada Buah Stroberi,” no. 2013, pp. 395–401, 2020.

[9] A. S. Sinaga, “SEGMENTASI RUANG WARNA L*a*b,” *J. Mantik Penusa*, vol. 3, no. 1, pp. 43–46, 2019.

[10] D. Hernando, A. W. Widodo, and C. Dewi, “Pemanfaatan Fitur Warna dan Fitur Tekstur untuk Klasifikasi Jenis Penggunaan Lahan pada Citra Drone,” vol. 4, no. 2, pp. 614–621, 2020, [Online]. Available: http://j-ptiik.ub.ac.id.

ISSN Print : 2621-3745
ISSN Online : 2621-3753
DOI : 10.31328/jsae.v6i1.4655

[11] H. S. Value, “Identifikasi Kematangan Daun Teh Berbasis Fitur Warna Hue Saturation Intensity (HSI) dan Hue Saturation Value (HSV) (Identification Maturity Tea Leaves Based on Color Feature Hue Saturation Intensity (HSI) and Hue Saturation Value,” vol. 8, no. November, pp. 217–223, 2020.

[12] E. Junianto and M. Z. Zuhdi, “Penerapan Metode Palette untuk Menentukan Warna Dominan dari Sebuah Gambar Berbasis Android,” *J. Inform.*, vol. 5, no. 1, pp. 61–72, 2018, doi: 10.31311/ji.v5i1.2740.

[13] T. Y. Prahudaya and A. Harjoko, “Metode Klasifikasi Mutu Jambu Biji Menggunakan Knn Berdasarkan Fitur Warna Dan Tekstur,” *J. Teknosains*, vol. 6, no. 2, p. 113, 2017, doi: 10.22146/teknosains.26972.

[14] S. T. Wulan *et al.*, “OPTIMASI SELEKSI FITUR KLASIFIKASI NAÏVE BAYES MENGGUNAKAN ALGORITMA GENETIKA UNTUK PREDIKSI RISIKO KREDIT KONSUMEN (Studi Kasus : PT. Finansia Multi Finance (KreditPlus) Tanjungpinang),” pp. 1–17.

[15] F. Y. Manik and K. S. Saragih, “Klasifikasi Belimbing Menggunakan Naïve Bayes Berdasarkan Fitur Warna RGB,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 11, no. 1, p. 99, 2017, doi: 10.22146/ijccs.17838.

[16] F. Febriana *et al.*, “Perbandingan Klasifikasi Naive-Bayes dan KNN untuk Mengidentifikasi Jenis Buah Apel dengan Ekstraksi Ciri LBP dan HSV,” no. September, pp. 191–201, 2021.