Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12006
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Identification of Tempe Fermentation Maturity Using Principal Component Analysis and K-Nearest Neighbor

Istiadi[1)*], Aviv Yuniar Rahman[2)], Alif Dio Raka Wisnu[3)]
[1,2,3)] Widyagama Unversity of Malang, Indonesia
[1)]istiadi@widyagama.ac.id, [2)]aviv@widyagama.ac.id, [3)] alifdio47@gmail.com

Submitted : Dec 15, 2022 | **Accepted** : Dec 28, 2022 | **Published** : Jan 1, 2023

Abstract: Tempe is one of the traditional foods in Indonesia which has nutritional content and benefits that are very much favored by all Indonesian people. To determine the maturity of tempe, it is generally done by fermenting it into tempeh using a certain temperature and usually tempe entrepreneurs are done traditionally. But in this way, tempe producers do not know what temperature and humidity are right for tempeh maturity. In this study, researchers used the MATLAB R2018a application with a total data set of 137 raw data, 137 ripe data and 136 rotten data, totaling 410 tempe image data. The purpose of this research is to produce a system that can detect the ripeness of tempe using the KNN (K-Nearest Neighbor) method which is equipped with GLCM texture feature extraction, with extraction of 8 color features, using the PCA (Principal Component Analysis) selection feature. And compare the results with the same method, namely KNN (K-Nearest Neighbor) without using the PCA (Principal Component Analysis) selection feature with the required running time between the two. KNN with PCA selection feature gets an average accuracy value of 80.63% and takes 1.06 seconds. Compared with the same method, namely KNN without using the selection feature, it gets an average accuracy value of 81.67% with a time of 1.18 seconds.

Keywords: KNN, PCA, maturity, fermentation,

INTRODUCTION

Tempe chips use the main ingredient of tempe which is generally provided by the tempe craftsmen. One of the problems that arise in producing tempeh as a chip material is in the fermentation process of making tempeh. Traditionally the tempe fermentation process is carried out in an open environment which is less than ideal so that the fermentation time becomes slow and erratic (2-3 days tempe fermentation) (Redi Aryanta 2020). **So** far, making tempeh uses Rhizopus sp., which is a type of mold that grows well at an optimum growth temperature of 28-35 0C and humidity below 65-70% (Gunawan and Sukardi 2020). Meanwhile, the average air temperature fluctuates between 18-23 0C which is actually not ideal for the growth of the mold.

Then a study will be carried out to classify tempe fermentation using the K-Nearest Neighbor (KNN) method and use Principal Component Analysis (PCA) as a selection feature to determine the accuracy obtained. From previous research the K-NN method is a classification technique for objects based on training data that is the closest distance or has characteristics similar to the object (Zhang et al. 2018). Near and far neighbors of pixels are calculated by Euclidean distance, while Principal Component Analysis (PCA) is an algorithm that is able to convert a group of data that are initially correlated to data that are not correlated to each other (Principal Component). The number of Principal Components produced is the same as the amount of the original data, but can be reduced to a smaller number and is still able to represent the original data well (Rambe, Tanjung, and Muhathir 2022).

LITERATURE REVIEW

In previous research, Dinda had conducted the Car Vehicle Number Plate Classification System Using Principal Component Analysis and K-NN Classification. From the test results, the accuracy of plate detection was 97.78% or 45 test data were successfully detected from 44 test data (Taningrum, Hidayat, and Hariani 2016). The research conducted by Arie on the Introduction of Palembang Songket Motifs Using Canny Edge Detection, PCA and KNN obtained an accuracy of 91.67% (Hasan and Liliana 2020). Then Siti's research on Classification of Coffee Maturity Levels Based on Color Detection Using the KNN and PCA Methods using 90

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12006
e-ISSN : 2541-2019
p-ISSN : 2541-044X

training data on coffee cherries images and 45 test data on the accuracy results of the classification of coffee pod maturity levels using the KNN method of 97.77% with a value of K = 3 (Raysyah, Veri Arinal, and Dadang Iskandar Mulyana 2021). Subsequent research conducted by Setya on Classification of Maturity Ambon Bananas Using the KNN and PCA Methods Based on Image 8 Color results of the accuracy of the classification of ripeness levels of Ambon bananas using the KNN method is 90.9% with a value of K = 5 (Adenugraha, Arinal, and Mulyana 2022).

METHOD

In the process of tempe fermentation, it is difficult to classify the maturity of tempe fermentation because perceptual limitations to determine the condition of fungus that is in tempe. To overcome this problem, it is necessary to have an automatic tempe fermentation classifier system based on tempe images using machine learning. An image besides having texture features has a variety of color features, so a simplification approach is needed for this diversity. With the method used, namely K-Nearest Neighbor (KNN) with Principal Component Analysis (PCA) as a selection feature that needs to be tested for its accuracy in classifying.

This classification will be compared with KNN which uses all the features. Based on this comparison, the results of the accuracy and time of classification will be seen. Based on this comparison, the results of the accuracy and time of classification will be seen. In addition, the discussion will review the extent of its conformity to previous research.

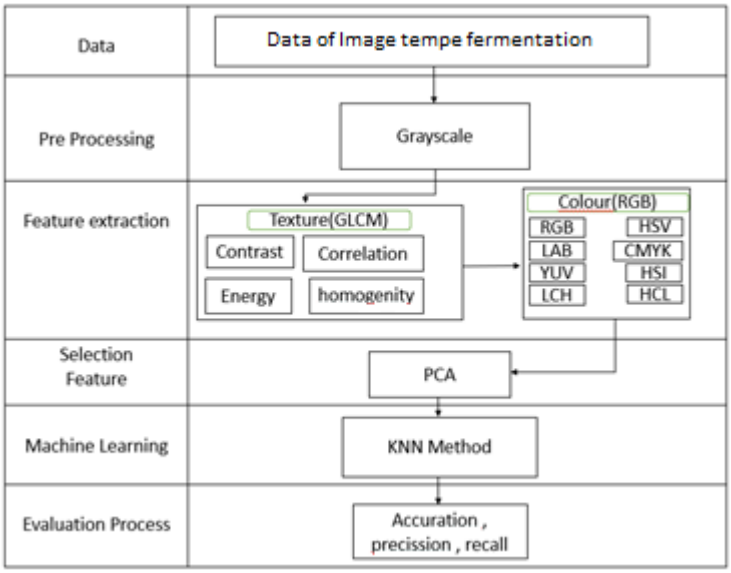


Figure 1 Framework

Further explanation for Figure 1, the first is Data, which is a process for receiving and or collecting information from tempe fermentation image data. The next step is Pre Processing, where a test data and test data are cropped to make an image look clearer and more natural so as to reduce the effect of noise which results in the system not being optimal in segmenting colors and textures. Extraction Feature using the GLCM method. Next is Machine learning, which is the stage where the classification of the type of Tempe will be examined, in this case we use the K Nearest Neighbor method. Then for the selection feature here we use Principal Component Analysis (PCA). Finally, Evaluation Process, which is the stage where the process of evaluating the classification process that has been carried out both in terms of strengths and weaknesses, in this stage everything related to the system is revised and then repaired and produces Accuracy, Precision, Recall. Precision is the level of accuracy between the information requested by the user and the answers given by the system. While recall is the success rate of the system in retrieving information. Accuracy is defined as the degree of closeness between the predicted value and the actual value.

Matlab

Matlab is a software that is used for matrix-based programming, computation, analysis and mathematics. The programming language used is MathWorks Inc. Unite the process of programming, visualization, computing through tools that are easy to use. Matlab has advantages including analysing and exploring data as well as developing algorithms, modelling and simulating in the form of visual plots in 2D and 3D to developing graphical interface applications (L. Galib, S. Tahir, and A. Abdulrahman 2021).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12006
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Image Digital

Digital or analog images are different from thermal-based images, digital or analog images are images that are representations/illustrations of real objects, while thermal-based images are images resulting from temperature detection emitted by objects captured by a thermal camera, so the resulting image is a processed beam object heat is captured and produces certain colors according to the heat emitted (Fadjeri et al. 2022).

Grayscale Image

Grayscale image is a data matrix whose values represent the intensity of each pixel ranging from 0-255. Each pixel requires 8bits of memory. Figure 2 shows a close-up grayscale image with several pixel intensity values taken (Fadjeri et al. 2022).

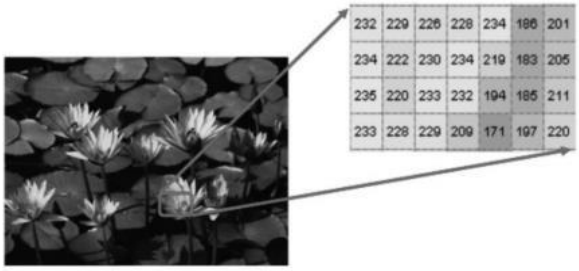


Figure 2 Grayscale Image

GLCM Texture Feature Extraction Concept

Gray Level Co-occurrence Matrix (GLCM) is a method used for texture analysis as part of feature extraction. GLCM is a matrix that describes the frequency of the appearance of a pair of two pixels with a certain intensity in a certain distance and direction in the image (Rambe et al. 2022).

Color Image (RGB)

Color segmentation, there are various color models. The RGB (Red Green Blue) model is a widely used model, one of which is the monitor. In this model, to represent an image using these 3 color components. In addition to the RGB model, there is also an RGB normalization model where this model has 3 components, namely, r, g, b which represent the percentage of a pixel in a digital image (Eriksson and Tabachnikova 2022).

HSV (Hue, Saturation, Value)

HSV defines color in terms of hue, saturation and value. Hue states the actual color, saturation states the purity level of a color, namely by identifying how much white is given to the color, while value is an attribute that states the amount of light received by the color (Yessy Nabella, Arum Sari, and Cahya Wihandika 2019).

LAB (Lightness, A, and B)

LAB is a type of color-opponent color space in the image (red vs green and yellow vs blue) based on the XYZ color space. This color space describes all the colors visible to the human eye, where L identifies light, A identifies the position of red (red) and green (green) and B identifies the position between yellow (yellow) and blue (blue) (Alamsyah and Pratama 2019).

CMYK (Cyan, Magenta, Yellow, Key'black')

Colors that are known in the printing and printing process consist of C = Cyan M = Magenta Y = Yellow K = Key (the result of a combination of several colors), Used to appear balanced against the white background of printed materials such as paper and others, so to reproduce images so that results can be achieved, relatively at least 4 inks are needed, namely: Cyan, Magenta, Yellow, and Black. The four inks are called ink / color (Parolinda and Ramdan 2019).

YUV

YUV Color Moment is a measurement method that can be used to differentiate images based on color features. The YUV model consists of a brightness component (Y) and 2 color content/Chrominance components (U and V) (Arun and Durairaj 2017). YUV is a color model that consists of a lumaY channel which is the brightness of the image and two chrominance channels U and V which describe the specific color of the image.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12006
e-ISSN : 2541-2019
p-ISSN : 2541-044X

HSI (Hue, Saturation, Intensity)

The HSI color model defines color in terms of Hue, Saturation and Intensity. Intensity is an attribute that states the amount of light received by the eye regardless of color. The HSI model is a color system that is closest to how the human eye works. HSI combines information, both color and grayscale from an image. Hue is the angle between the reference color and the S (saturation) vector. The reference color is usually red but it could be other colors (Edha, Sitorus, and Ristian 2020).

HCL (Hue, Chroma, Lightness)

The HCL color space has been developed by retaining the advantages of the HSL and HSV spaces and covering the deficiencies that exist in both. One of the advantages of this color space is that the H component (color) has a constant value even when the light intensity and chroma of the object change. Saturation shows the color of the color intensity of Hue. Chroma color is a term to describe the brightness or dullness of a color, the quality or strength of a color. Colors with full intensity appear very striking and have a firm effect, while colors with low intensity appear softer (Junianto and Zuhdi 2018).

LCH (Local Colour Histogram)

Local Color Histogram divides the image into several parts and then calculates the color histogram for each part, this process requires more computational processing compared to other color features. Color histogram is a way to describe the content or color content by calculating each color that appears in an image (Ilhamsyah, Rahman, and Istiadi 2021).

K-NN (K-Nearest Neighbor)

K-Nearest Neighbor is a classification algorithm for objects based on learning data that is the closest distance or has the most common object characteristics (Yana and Nafi'iyah 2021). There are also several steps to calculate the K-NN algorithm, determine the value of K, calculate the Euclidean distance (query instance) for each object from the training data. Arranges objects into groups that have the smallest Euclidean distance. Collect class Y labels (nearest neighbor classification). Far or near objects can be calculated by Euclidean distance, where two vector distances are of size n, for example x = (x1, x2, x3, ..., xn) and y = (y1, y2, y3, ..., yn) we get Equation (1) as follows:

$$Dist\ (x,y) = \sqrt{\sum_{i=1}^{n}(xi - yi)^2} \quad (1)$$

PCA (Principal Component Analysis)

PCA is a statistical method that has been used in various fields such as pattern recognition, image compression and is used to determine the distribution of data from feature extraction results. This technique is usually used to recognize patterns in high-dimensional data or samples and represent the data by reducing the variables in the features used in the data to determine (Aristo Jansen Sinlae et al. 2022). Calculation analysis using the PCA methodology is a problem of solving the problem of eigen equations because basically PCA calculations are based on more than one eigenvalue. The PCA algorithm in general is as follows:

$$Cov\ (xy) = \frac{\sum xy}{n} - (x)(y) \quad (2)$$

$$(A - \lambda \mathrm{I}) = 0 \quad (3)$$

$$[A - \lambda \mathrm{I}][X] = [0] \quad (4)$$

$$\rho I = \frac{\lambda_i}{\sum_{j=1}^{D} \lambda_j} X\ 100\% \quad (5)$$

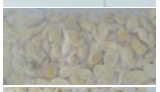
RESULT

The data used in carrying out the classification is tempe image data sourced from tempe producers which are stockpiled to collect training data and test data. With a total dataset of 137 Raw data, 137 Ripe data and 136 Rotten data, a total of 410 tempe image data. In this study the researchers created a system for classifying the maturity level of tempeh. As for this research, it uses the K-Nearest Neighbor algorithm to classify data and uses the Principal Component Analysis algorithm to get the distribution of the dataset. To make this research easier, the researchers used the MATLAB R2018a programming to create a tempeh maturity classification system.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12006
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 1 Class Tempe

No	Class	Definition	Picture
1	rotten	The image is rotten tempeh	
2	ripe	Image of riped tempeh	
3	raw	Image of raw tempeh	

System Implementation

This study implements the KNN method in classifying data and comparing the data using the same method, namely KNN, but using the PCA selection feature. As for the implementation of the KNN method by measuring the shortest distance between the test data and the training data. At this stage, the calculation of the feature extraction value of the training data and test data is carried out, which displays as many as 410 data being tested.

Table 2 KNN Training

Split Ratio	KNN			
	k=2		Data	
	Accuracy	Time/Second	Train	Test
10;90	82.11%	1.165521	41	369
20;80	86.99%	1.192988	82	328
30;70	91.33%	1.171811	123	287
40;60	92.28%	1.142172	164	246
50;50	93.17%	1.182083	205	205
60;40	94.31%	1.17161	246	164
70;30	94.19%	1.173112	287	123
80;20	95.33%	1.104301	328	82
90;10	95.84%	1.215076	369	41

Table 2 explains that the results of the training data using the KNN method with a value of k = 2 obtained the highest accuracy value, namely at a split ratio of 90:10 with an accuracy of 95.84% with a training time of 1.21 seconds. As well as the results of the calculation of the confusion matrix and the distribution of data with the highest value at a split ratio of 90:10 which can be seen in Figure 3.

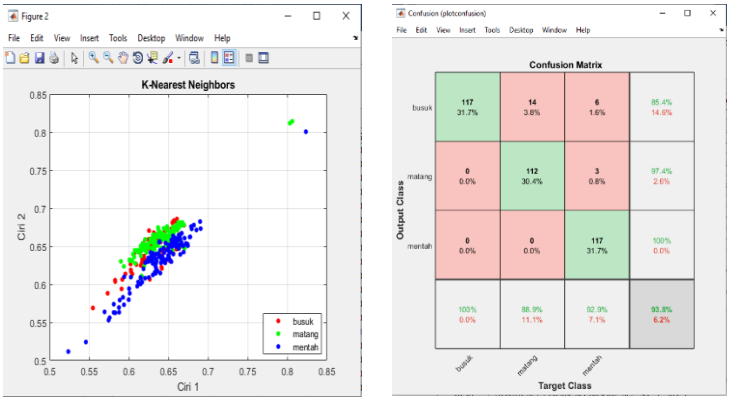


Figure 3. Confusion matrix and distribution of KNN training data with a split ratio of 90:10

Table 3 KNN Testing

Split Ratio	KNN			
	k=2		Data	
	Accuracy	Time/Second	Train	Test

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12006
e-ISSN : 2541-2019
p-ISSN : 2541-044X

10;90	58.99%	1.335533	41	369
20;80	77.03%	1.164216	82	328
30;70	79.79%	1.133188	123	287
40;60	83.20%	1.15116	164	246
50;50	93.17%	1.187153	205	205
60;40	92.28%	1.18351	246	164
70;30	90.24%	1.177131	287	123
80;20	86.99%	1.190485	328	82
90;10	83.74%	1.152974	369	41

Table 3 explains that the results of the test data using the KNN method using the PCA selection feature with a value of k = 2 obtained the highest accuracy value, namely at a split ratio of 50:50 with an accuracy of 93.17% with a test time of 1.18 seconds. As well as the results of the calculation of the confusion matrix with the highest value at a split ratio of 50:50 which can be seen in Figure 4.

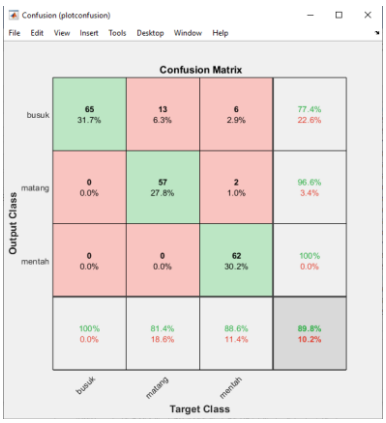


Figure 4. Confusion matrix of 50:50 KNN testing

Table 4 KNN Training with PCA

Split Ratio	KNN PCA			
	k=2		Data	
	Accuracy	Time/Second	Train	Test
10;90	86.99%	1.030898	41	369
20;80	89.43%	1.036452	82	328
30;70	91.87%	1.079751	123	287
40;60	91.87%	1.07452	164	246
50;50	90.89%	1.080466	205	205
60;40	91.87%	1.084465	246	164
70;30	93.50%	1.103249	287	123
80;20	92.48%	1.131063	328	82
90;10	93.50%	1.063575	369	41

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12006
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 4 explains that the results of the training data using the KNN method using the PCA selection feature with a value of k = 2 obtained the highest accuracy value, namely at a split ratio of 70:30 with an accuracy of 93.5% with a training time of 1.1 seconds. As well as the calculation results of the confusion matrix and distribution of training data with the highest value at a split ratio of 70:30 can be seen in Figure 5.

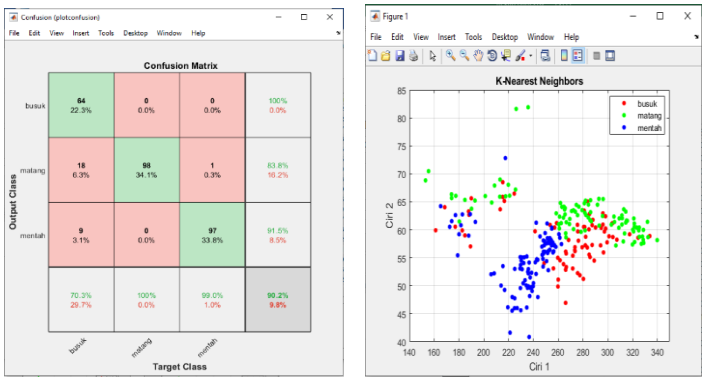


Figure 5. Confusion matrix and distribution of KNN training data with a PCA split ratio of 70:30

Table 5 KNN testing with PCA

Split Ratio	KNN PCA			
	k=2		Data	
	Accuracy	Time/Second	Train	Test
10;90	59.53%	1.054249	41	369
20;80	76.63%	1.050933	82	328
30;70	76.31%	1.049943	123	287
40;60	80.22%	1.057772	164	246
50;50	90.89%	1.025335	205	205
60;40	91.06%	1.068816	246	164
70;30	88.08%	1.050478	287	123
80;20	86.18%	1.0593	328	82
90;10	83.74%	1.057973	369	41

Table 5 explains that the results of the test data using the KNN method using the PCA selection feature with a value of k = 2 obtained the highest accuracy value, namely at a split ratio of 60:40 with an accuracy of 91.06% with a test time of 1.06 seconds. As well as the results of the calculation of the confusion matrix with the highest value at a split ratio of 60:40 which can be seen in Figure 6.

Figure 6. Confusion matrix of KNN testing with PCA 60:40

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

DISCUSSIONS

Based on the number of existing datasets, namely 410 data with details of 137 raw data, 137 mature data, and 136 rotten data, several tests were carried out with the composition of the amount of training data and testing data at a split ratio of 10:90 to 90:10 with a total of 410 data using the KNN method with the PCA selection feature the best accuracy was obtained. In the KNN + PCA method with a value of k = 2 the accuracy obtained for classifying is 91% at a split ratio of 60:40 with a total of 246 training data and 164 test data with a time of 1.06 seconds. By comparing using the KNN method without the PCA selection feature, the accuracy obtained is highest at a split ratio of 50:50 with a k = 2 value of 93% with a total of 205 training data and 205 test data with a time of 1.18 seconds.

Even though the results of the KNN method with PCA are slightly below the KNN without PCA, the results are still said to be high because they are able to produce an accuracy representation of more than 90% (Hasan and Liliana 2020) (Adenugraha, Arinal, and Mulyana 2022). In addition, the use of PCA feature selection is faster in its classification because it can summarize a number of feature variations (Aristo Jansen Sinlae et al. 2022).

CONCLUSION

In the classification, it can be concluded that from the KNN method with the PCA selection feature, an average accuracy value of 80.63%, an average precision of 76.8%, an average recall of 71.11% is obtained in 1.06 seconds. Compared to the same method, namely KNN without using the selection feature, it obtained an average accuracy value of 81.67%, an average precision of 77.17%, an average recall of 72.64% with a time of 1.18 seconds. In this case the PCA selection feature serves to reduce the variables used to be smaller so that the testing process becomes faster with accuracy, precision, and recall which is slightly different from the same method, namely KNN without using PCA but with a difference of only 1% with comparison time is faster with the selection feature with a difference of 0.1 seconds.

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Education Culture Research and Technology (Kemendikbudristek) and LPDP for funding this research through the 2021 Scientific Research Grant.

REFERENCES

Adenugraha, S. P., Arinal, V., & Mulyana, D. I. (2022). Klasifikasi Kematangan Buah Pisang Ambon Menggunakan Metode KNN dan PCA Berdasarkan Citra RGB dan HSV. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *6*(1), 9-17.

Alamsyah, D., & Pratama, D. (2019). Segmentasi Warna Citra Bunga Daisy dengan Algoritma K-Means pada Ruang Warna Lab. *Jurnal Buana Informatika*, *10*(2), 153-163.

Arun, C. H., & Durairaj, D. C. (2017). Identifying medicinal plant leaves using textures and optimal colour spaces channel. *Jurnal Ilmu Komputer dan Informasi*, *10*(1), 19-28.

Aryanta, I. W. R. (2020). Manfaat tempe untuk kesehatan. *Widya Kesehatan*, *2*(1), 44-50.

Taningrum, D. R., Hidayat, B., & Hariyani, Y. S. (2016). Sistem Pengidentifikasian Plat Nomor Kendaraan Mobil Menggunakan Principal Component Analysis Dan Klasifikasi KNN. *E-Proceeding Eng*, *3*(2), 1868-1876.

Edha, H., Sitorus, S. H., & Ristian, U. (2020). Penerapan metode transformasi ruang warna hue saturation intensity (HSI) untuk mendeteksi kematangan buah mangga harum manis. *Coding Jurnal Komputer dan Aplikasi*, *8*(1).

Eriksson, I., & Tabachnikova, N. (2022). "Learning models": Utilising young students' algebraic thinking about equations. *LUMAT: Luonnontieteiden, matematiikan ja teknologian opetuksen tutkimus ja käytäntö*, *10*(2).

Fadjeri, A., Saputra, B. A., Ariyanto, D. K. A., & Kurniatin, L. (2022). Karakteristik Morfologi Tanaman Selada Menggunakan Pengolahan Citra Digital. *Jurnal Ilmiah Sinus (JIS) Vol*, *20*(2).

Galib, S. L., Tahir, F. S., & Abdulrahman, A. A. (2021). Detection Face parts in image using Neural Network Based on MATLAB. *Engineering and Technology Journal*, *39*(1B), 159-164.

Gunawan, B., & Sukardi, S. (2020). Rancang Bangun Pengontrolan Suhu dan Kelembaban pada Proses Fermentasi Tempe Berbasis Internet of Things. *JTEIN: Jurnal Teknik Elektro Indonesia*, *1*(2), 168-173.

Hasan, M. A., & Liliana, D. Y. (2020). Pengenalan Motif Songket Palembang Menggunakan Deteksi Tepi Canny, PCA dan KNN. *vol*, *6*, 1-7.

Ilhamsyah, I., Rahman, A. Y., & Istiadi, I. (2021). Klasifikasi Kualitas Biji Kopi Menggunakan MultilayerPerceptron Berbasis Fitur Warna LCH. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(6), 1008-1017.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Junianto, E., & Zuhdi, M. Z. (2018). Penerapan Metode Palette untuk Menentukan Warna Dominan dari Sebuah Gambar Berbasis Android. *Jurnal Informatika*, *5*(1), 61-72.

Nabella, F. Y., Sari, Y. A., & Wihandika, R. C. (2019). Seleksi Fitur Information Gain Pada Klasifikasi Citra Makanan Menggunakan Hue Saturation Value dan Gray Level Co-Occurrence Matrix. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, *2548*, 964X.

Rambe, A., Tanjung, J. P., & Muhathir, M. (2022). Shafiyyatul Amaliyyah School Student Face Absence Using Principal Component Analysis and K–Nearest Neighbor. *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, *5*(2), 414-422.

Raysyah, S., Arinal, V., & Mulyana, D. I. (2021). Klasifikasi Tingkat Kematangan Buah Kopi Berdasarkan Deteksi Warna Menggunakan Metode Knn Dan Pca. *JSiI (Jurnal Sistem Informasi)*, 88-95.

Sinlae, A. A. J., Alamsyah, D., Suhery, L., & Fatmayati, F. (2022). Classification of Broadleaf Weeds Using a Combination of K-Nearest Neighbor (KNN) and Principal Component Analysis (PCA). *Sinkron: jurnal dan penelitian teknik informatika*, *7*(1), 93-100.

Yana, Y. E., & Nafi'iyah, N. (2021). Klasifikasi Jenis Pisang Berdasarkan Fitur Warna, Tekstur, Bentuk Citra Menggunakan SVM dan KNN. *Journal of Computer, Information System & Technology Management*, *4*(1), 5.

Zhang, S., Li, X., Zong, M., Zhu, X., & Wang, R. (2017). Efficient kNN classification with different numbers of nearest neighbors. *IEEE transactions on neural networks and learning systems*, *29*(5), 1774-1785.

*name of corresponding author