

Identification of Rice Types Based on Shape, Color and Texture using K-Nearest Neighbors Method as Classifier

J Jumi¹, Achmad Zaenuddin²

Department of Business Administration
Politeknik Negeri Semarang
Semarang, Indonesia

Tedjo Mulyono³

Department of Civil Engineering
Politeknik Negeri Semarang
Semarang, Indonesia

Abstract—The rice is one of the superior staples and is very popular and consumed by the public. Rice has many varieties that are difficult for buyers to recognize. However, this can be overcome by identifying it based on the shape, color and texture of the seeds. This research uses the process of Invariant Moment, Hue Saturation Value (HSV) and Local Binary Pattern (LBP) as the extraction of the characteristics of the shape, color and texture of the rice which will then be used as characteristics of the type of rice studied. The K-Nearest Neighbor (K-NN) method is one of the research method on artificial intelligence that can be used to classify the values obtained from the HSV and LBP feature extraction. The data used were 100 images, consisting of 90 training images and 10 test images. The results of the evaluation of the K-Nearest Neighbor method showed that the average value of Precision obtained was more than 85,23%, Recall 95,33%, and an accuracy of more than 86,22%.

Keywords—Texture, Color, Classification, K-Nearest Neighbor, Rice

1. INTRODUCTION

Rice is currently an important component of food for daily consumption in several countries in the world, especially in Asia. The need for rice is also increasing along with the increase in population. The price of rice circulating in the market continues to increase so that many sellers sell rice with poor quality and type of rice, while many consumers do not know how to distinguish between types of rice with good quality or poor quality [15]. Each type of superior rice comes from different growing media [3]. Therefore, digital image processing can be used as a tool to recognize and test the quality of rice and help determine the type of rice [8]; [1]. Another benefit of digital image technology is that it can quickly check the type of rice and accurate [12] and does not damage the object used [20].

Identification on grain objects such as rice has been used to check the quality of rice by comparing inspection by humans which yields an accuracy rate of 90% [18] and assess the quality of rice [1]. Another use is checking the quality of bulk rice which results in a high level of accuracy [22] and identification of rice varieties based on shape and color [11]. Apart from rice, other types that have been tested for quality using digital images are wheat, oats and barley [21]; [19]; [5]; [3]; [17]. However, identification of rice images to identify the type of rice has never been carried out. Another thing that distinguishes is the use of the classification method using K-

NN after three stages of extraction, there are shape extraction, color extraction and texture extraction.

K-NN is called the classification method with the most effective technique because the data is studied in training data [13]. Therefore, K-NN is known to have advantages over other methods, there are the ability to handle data that has a lot of noise and is also more effective in using large training data [14]. While feature extraction in the form of a hue saturation value process as a recognition process with color characteristics and local binary patterns as a recognition process with textural features is able to provide 98% accuracy results in the classification of apples which are the right blend for use in identifying rice types [16]. Meanwhile, the extraction of the invariant moment in identifying the type of rice provides an accuracy of more than 80% [10].

This research is based on the argument that rice is not only a staple food that meets daily needs, but also a source of energy and nutrients necessary for the body to be of high quality. However, visual identification of rice types is still using manual method so it is assumed that errors will occur due to limited human vision [15]. Therefore, technology is needed to help the community precisely and quickly. This study aims to classify the type of rice using an algorithm a K-Nearest Neighbor (K-NN) which is an appropriate classification method by identifying the type of rice based on the basic assumption of data that has an object recognition pattern, there are, sample objects that are similar and close to each other in model space [7].

II. RELATED WORKS

A. Invariant Moment

Invariant moment is a function used in probability theory to analyze the form of a set of moments of a function $f(x, y)$. Moment order $(p + q)$ is a continuous function $f(x, y)$ which can be interpreted that $p, q = 0, 1, 2, \dots, n$ with the following definitions:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad (1)$$

The Central moment is defined in the following equation:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy \quad (2)$$

$$\bar{x} = \frac{m_{10}}{m_{00}} \text{ and } \bar{y} = \frac{m_{01}}{m_{00}} \quad (3)$$

So that the equation for digital images is:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (4)$$

Shape extraction using the invariant moment is illustrated with seven invariant moments from the central moment, there are: μ_{00} , μ_{10} , μ_{01} , μ_{20} , μ_{11} , μ_{02} , μ_{30} , μ_{12} , μ_{21} , μ_{03} by defining normalized central moments as the $n_{pq} = \mu_{pq} / \mu_{00}^{\gamma}$, $\gamma = 0,5(p+q)+1$ untuk $p+q = 2,3$, to seven invariant moments whose value does not change with respect to rotation, translation and scale [9].

$$\begin{aligned} V_1 &= (n_{20} + n_{02}) \\ V_2 &= (n_{20} - n_{02})^2 + 4n_{11}^2 \\ V_3 &= (n_{30} - 3n_{12})^2 + (3n_{21} - n_{03})^2 \\ V_4 &= (n_{30} + n_{12})^2 + (n_{21} + n_{03})^2 \\ V_5 &= (n_{30} - 3n_{12})(n_{30} + n_{12})[(n_{30} + n_{12})^2 - 3(n_{21} + n_{03})^2] + (3n_{21} + n_{03})(n_{21} + n_{03})[3(n_{30} + n_{12})^2 - (n_{12} + n_{03})^2] \\ V_6 &= (n_{20} - n_{02})[(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] + 4n_{11}(n_{30} + n_{12})(n_{21} + n_{03}) \\ V_7 &= (3n_{21} - n_{30})(n_{30} + n_{12})[(n_{30} + n_{12})^2 - 3(n_{21} + n_{03})^2] + (3n_{12} - n_{30})(n_{21} - n_{03})[3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] \end{aligned} \quad (5)$$

B. Hue Saturation Value (HSV)

HSV has the main characteristics of color, among others [4]:

- a) Hue is used to determine redness (redness), greenness (greenness) and etcetera.
- b) Saturation is the purity or strength of color.
- c) The brightness values of the colors range from 0-100%. When the value is 0, the color is black, the greater the value, the brighter and new variations of the color appear. [16]. Value is also called the process of getting the value of each color you want to display through the calculation process by converting the RGB color space to the HSV color space [24]. How to get each defined HSV value with the formula shown in equation (6) to equation (10) [17]. Equation (6) is a formula to find the maximum value of the three RGB color space parameters that will be entered into the max. As for Hue (H), saturation (S) and value (V) are the color spaces in the image being evaluated.

$$\text{Max} = \text{Max} (R, G, B) \quad (6)$$

$$\text{Min} = \text{Min} (R, G, B) \quad (7)$$

$$V = \text{Max} \quad (8)$$

$$S = \frac{\text{Max} - \text{Min}}{\text{Max}} \quad (9)$$

$$H = \begin{cases} \frac{1}{6} \frac{G-B}{\text{Max}-\text{Min}} & V = R \\ \frac{1}{6} \frac{B-R}{\text{Max}-\text{Min}} + \frac{1}{3} & V = G \\ \frac{1}{6} \frac{R-G}{\text{Max}-\text{Min}} + \frac{2}{3} & V = B \end{cases} \quad (10)$$

C. Local Binary Pattern (LBP)

Local Binary Pattern (LBP) is one of the image processing techniques commonly used to determine the texture of an image. LBP labels the pixels in the image by thresholding the neighbors of each pixel and returning the result as a binary number. The process in LBP is similar to the process of thinning operations in image operations. However, for the LBP iteration process the calculation is done by comparing the center pixel with its neighbors. Neighbors that have a pixel value lower than the middle pixel will be replaced with a bit value of 0 while other neighbors will be assigned a value of one [16].

$$\text{LBP}_{P,R}(xc, yc) = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (11)$$

With the the condition:

$$s(x) = \begin{cases} 1, & x < 0 \\ 0, & x \geq 0 \end{cases} \quad (12)$$

xc and yc are the center pixel points, p are circular sampling points, P are the number of sampling points, g_p is the grayscale value of p, g_c is the center pixel and s is the threshold function.

D. K-Nearest Neighbor

The K-Nearest Neighbor (K-NN) algorithm includes a type of classification method that determines the label (class) of a new object based on the majority of classes from the closest k distance in the training data group [8]. The nearest k point neighbor selected from the training set can determine the performance of the forecast results. Therefore, the new point forecast is associated with its k-nearest value [30].

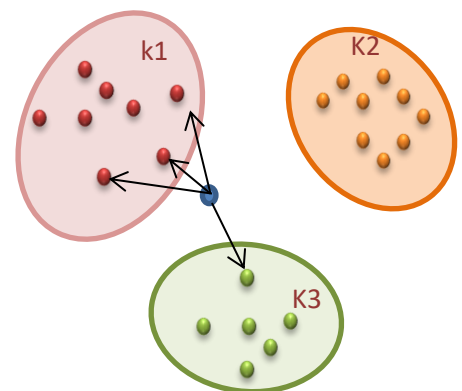


Fig. I. The K-Nearest Neighbors model [30]

k_1, k_2, k_3 is a description of the three classification groups in the training dataset. The arrows connect the proximity of

the four points on x_u , the most minor points classified on x_u are referred to k_1 , so the classification on x_u is k_1 [30]. The Euclidean distance equation is used in the K-NN algorithm to classify the test image into classes with the number of closest members. The working principle of K-NN is to find the closest distance between the data to be evaluated and the k nearest neighbors in the training data [6] using equation (13).

$$P = \sqrt{(j_1 - l_1)^2 + (j_2 - l_2)^2} \quad (13)$$

j is sample data, l is test data and P is distance.

III. METHOD

A. Materials and Tools

The research tool for identifying rice types using digital image data of rice was taken using hardware in the form of a smartphone camera and storage connected to a computer and image identification software. As for the materials for research, the type of rice used in this study were 3 types of rice and 4 colors, there are white rojo catfish type, IR4 white rice type and C4 white rice type and white, red, green and black rice with 90 training data. image training data and 10 image test data.



Fig. 1. Sample Data Training

1) Research Stages

This study uses the stages shown in Fig. 2.

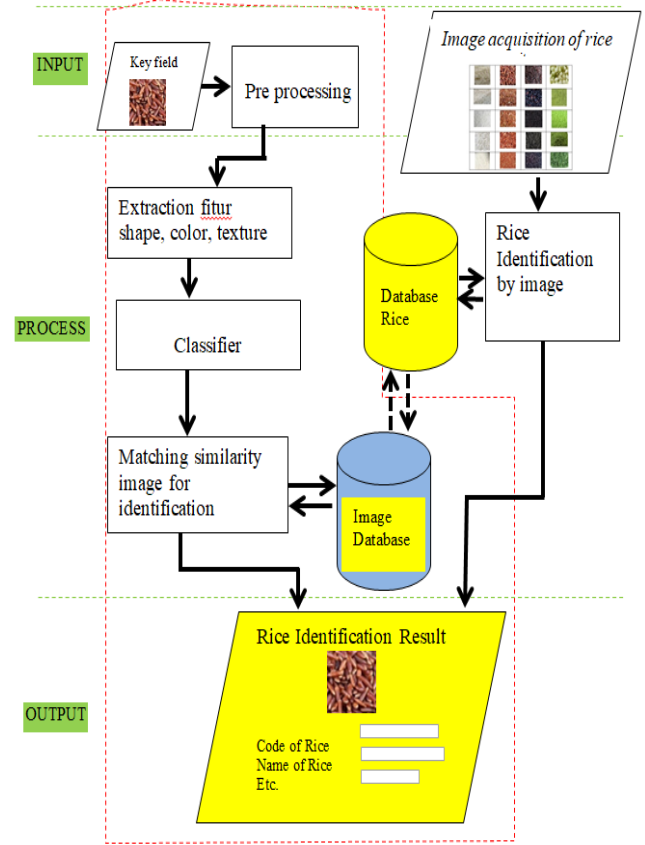


Fig. 2. The process framework for the Rice type identification system

1) Preprocessing

The preprocessing stage is the stage for processing data so that the extraction results have relatively high accuracy. At this stage, the image is resized by 300 x 300 pixels. The image of rice used is white rice, brown rice, black rice and green rice with There are 3 types of rice, there are white rojo catfish type, IR4 white rice type and C4 white rice type. After taking the required pixel image, then separating the object and the image background. After collected, the image is separated between training data and testing data. The next stage is manually labeled for training data totaling 90 images. The remaining 10 images were then prepared as testing data.

2) Feature Extraction

There are 3 feature extraction stages, there are:

a) Shape Feature Extraction

Extraction of shape features in this study used the Invariant Moment method with equation (5). At this stage, the moment value is calculated from moment 1 to moment 7. At the seven moment values describe the features of the extracted rice image object.

b) Color Feature Extraction

At the color feature extraction stage, the Hue Saturation Value (HSV) method is used. HSV feature extraction uses equation (6) to equation (10). This extracted color feature value reflects the color feature of the extracted rice object. In the process, first of all convert the RGB image to the HSV image. Furthermore, for the HSV image the histogram is searched for the image, but in prediction with the HSV statistics the Hue component in the HSV image is removed and replaced with a Grayscale histogram from the image. The

histogram is then searched for the distance with the template image and which label is selected that best matches the identified image.

c) *Texture Feature Extraction*

Extraction of texture features in this study uses Local Binary Pattern (LBP). The texture feature extraction stage uses Local Binary Pattern (LBP), there are by segmenting the input image into a grayscale image. After that, the LBP search process is carried out by equation (11) using the parameters in equation (12). The calculation of the difference in LBP was used using the chi-squared distance from the normalized LBP histogram. The label selected in the material classification using LBP is the label that has the least chi-squared distance value.

3) *Classification Process*

In the K-NN process, the training data is labeled first to mark the new object. New objects or test data will be entered into the class based on the closest distance to k in the training data group using values k = 3 to k = 15. The training data is used to determine the suitable class, while the testing data becomes new data which will be classified by the model that has been created which will later be tested for its accuracy value.

4) *Matching Similarity*

Furthermore, the validation stage is used after the test data is identified as a class that is in the training data. The number of classes in the classification is then tested using a confusion matrix process to find out what percentage of accuracy is in the classification.

IV. RESULTS AND DISCUSSION

Tests in this study used 100 rice images and 300 x 300 pixels in size.

TABLE I. RICE IMAGE ACQUISITION SAMPLE DATA

WHITE RICE	BROWN RICE	BLACK RICE	GREEN RICE
			
WhiteRiceType 1	BrownRiceType1	Black RiceType1	Green Rice Type 1
			
WhiteRiceType 2	Brown Rice Type 2	Black RiceType2	Green Rice Type 2
			
WhiteRiceType 3	Brown Rice Type 3	BlackRiceType 3	Green Rice Type 3
			
WhiteRiceType 4	Brown Rice Type 4	Black RiceType 4	Green Rice Type 4
			

WhiteRiceType 5	Brown Rice Type 5	Black Rice Type 5	Green Rice Type 5
--------------------	-------------------	----------------------	-------------------

In Table I It is an example of rice image data with various types and colors, there are white, red, black and green.

A. *Identification Results*

Identification testing is carried out by varying the number of clusters and feature combinations. The test results with a combination of 3 feature weights and 10 cluster variations are shown in Fig. 5.



Fig. 5. Results of rice identification based on feature shape, color and texture

Fig. 5 shows the results of the identification of the quality and type of rice using the similarity value with the Euclidean distance method. Retrieval has a similarity ranking of 1 to 10 from the rice image database and testing with variations in the number of clusters will produce different retrieval outputs on the same data.

B. *Discussion*

At this stage, the identification accuracy of the rice image database is calculated. Furthermore, analysis of identification accuracy was carried out on the rice image database before and after classification or clustering with the number of clusters varying from 3 to 15 clusters.

C. *Before Clustering*

Analysis of identification accuracy at this stage uses a database of rice images before clustering with a variety of combinations and the number of features used, there are features of shape, color and texture. The accuracy value of the

identification results with various combinations of features can be seen in Table II.

TABLE II. PERCENTAGE OF RETRIEVAL ACCURACY WITH COMBINATION OF FEATURES IN RICE IMAGE DATABASE BEFORE CLUSTERING

Image of Name	Shape Features	Features Shape and color	Color and texture features	Shape and texture features	Features shape, color and texture
B1	60.53%	83.25%	77.67%	74.75%	85.87%
B2	61.77%	81.57%	78.25%	73.87%	85.15%
B3	61%	82.45%	78.785%	74.30%	86.25%
B4	59.45%	85.45%	78.25%	73.75%	86.65%
B5	61.57%	83.25%	77.87%	74.87%	85.76%
B6	61.23%	83.15%	77.65%	74.35%	85.57%
B7	58.44%	81.55%	78.35%	74.85%	85.77%
B8	61.35%	82.47%	77.89%	73.96%	85.58%
B9	62.25%	83%	78.76%	73.95%	85.15%
B10	61.45%	83.24%	78.35%	73.65%	86.58%

Table II shows that the accuracy of identifying rice images with a combination of the three features, there are shape, color and texture features, produces the highest average accuracy and the use of one feature results in the lowest identification accuracy of rice images.

D. After Clustering

Image grouping is based on the similarity, which results in narrowing the search space for image data information. The process is very profitable because it will shorten the identification time and increase retrieval accuracy. The level of identification accuracy with a combination of features and clustering can be seen in Table III.

TABLE III. THE PERCENTAGE OF IDENTIFICATION ACCURACY WITH A COMBINATION OF IMAGE FEATURES COMBINATION WITH CLUSTERING

Number of clusters	Shape Features	Features Shape and color	Color and texture features	Shape and texture features	Features shape, color and texture
3	62.45%	81.25%	74.47%	71.37%	91.67%
4	62.77%	81.26%	74.07%	71.56%	91.33%
5	63.15%	82.36%	75.46%	72%	92.33%
6	63.45%	82.57%	75.12%	72.17%	92.67%
7	64.76%	83.25%	76.67%	73.25%	93.33%
8	64.25%	83.15%	76.45%	73%	93.87%
9	63.45%	82.44%	76.36%	73.25%	93.97%
10	65.87%	85.55%	78.57%	74.55%	95.57%
11	65.35%	83.17%	78.09%	73.36%	93.55%
12	64.75%	84.45%	76.17%	73.09%	93.27%

Table III shows that the combination of the three features, there are features of shape, color and texture as well as variations in the number of clusters 10 resulted in the highest identification accuracy of more than 95%. Meanwhile, using only 1 feature in the number of clusters 3 results in the lowest accuracy, which is less than 63%.

The graph of identification accuracy in the rice image database using a combination of shape, color and texture

features with variations in the number of clusters ranging from 3 to 15 is shown in Fig. 6.

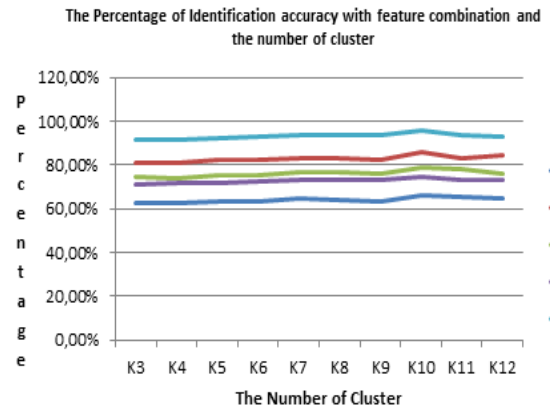


Fig. 6. Graph of identification accuracy with a combination and variation of the number of clusters

In Fig. 6, the accuracy of identification tends to increase until it reaches a variation of the number of clusters of 10 and tends to decrease after the number of clusters is greater than 10.

E. Computing Time

The computation time that takes to identify an image in the rice image database is influenced by the size of the image database and the database management technique. The average computation time for tracing in the image database before clustering is with the variation in the number of clusters shown in Fig. 7.

The time computation of Identification

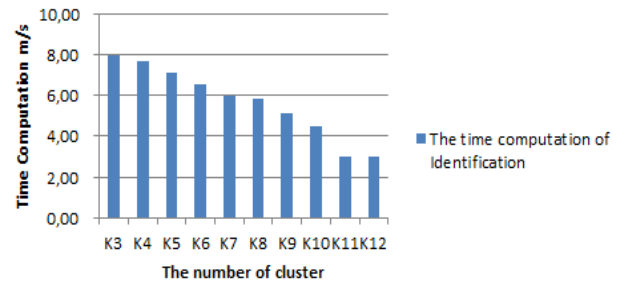


Fig.7. The time computation of identification

Fig. 7. explains that the greater the number of clusters, the faster the computation time.

V. CONCLUSIONS

The test results show that identification using three features, there are shape, color and texture features, has a higher level of accuracy compared to using only one feature or two features. Shape features have the most significant influence on the accuracy of the identification results compared to color and texture features. The use of KNN image classification has a significant effect on increasing the speed of the identification process and the accuracy of identification. The feature extraction method can be

developed so as to produce accuracy, precision and recall values with a higher percentage.

REFERENCES

- [1] Adnan, "Quality Testing of Sintanur Varieties of Rice at Several Types of Quality Using Digital Image Processing and Artificial Neural Networks Based on Color and Texture," Proceedings of the National Seminar on the Acceleration of Innovation-Based Agricultural and Rural Development, 2011, pp. 599-603.
- [2] A.C. Imanda, N. Hidayat, and M.T. Furqon, "Classification of Superior Rice Variety Groups Using Modified KNearest Neighbor". Journal of Information Technology and Computer Science Development, vol. 2, no. 8, pp. 2392-2399, 2018.
- [3] B.S. Anami and D.G. Savakar, "Improved Method for Identification and Classification of Foreign Bodies Mixed Food Grains Image Samples," ICGST-AIML Journal, 9 (1), pp. 1-8, 2009.
- [4] B.Y.B. Putranto, W. Hapsari, and K. Wijana, "Image color segmentation with HSV color detection to detect objects," J. Inform., vol. 2, no. 2, 2018.
- [5] A. Douik and M. Abdellaoui, "Cereal Grain Classification by Optimal Features and Intelligent Classifiers," Int. J. of Computers, Communications & Control, 5 (4), pp. 506-516, 2010.
- [6] E. K. Ratnasari, "The Recognition Of Fruit Types In Images Uses A Classification Approach Based On Lab Color Features And Co-Occurrence Texture," Inf. J. Ilm. Bid. Technol. Inf. and Commun., vol. 1, no. 2, 2016.
- [7] H. Lin, Z. Wang, W. Ahmad, Z. Man, and Y. Duan, "Identification of rice storage time based on colorimetric sensor array combined hyperspectral imaging technology," Journal of Stored Products Research, 85, pp. 1-7, 2020.
- [8] D.M. Hobson, R.M. Carter, and Y. Yan, "Characterization and Identification of Rice Grains through Digital Image Analysis," Instrumentation and Measurement Technology Conference, Warsaw, Poland, May 1-3, 2007.
- [9] M.K. Hu, "Visual Pattern Recognition by Moment Invariants," IEEE Trans. Inf. Theory 12, pp. 179 -187, 1962.
- [10] Jumi, U. T. Sulistyorini, Azizah, "Identification Of Rice Types Through Accuracy Of Rice Features And Color," JUST TI, vol. 11, no. 1, pp. 31-36, 2019.
- [11] C.C. Liu, J.T. Shaw, K.Y. Poong, M.C. Hong, and M.L. Shen, "Classifying Paddy Rice by Morphological and Color Features Using Machine Vision," Cereal Chem., 82 (6), pp. 649-653, 2005.
- [12] W. Liu, Y. Tao, T.J. Siebenmorgen, and H. Chen, "Digital Image Analysis Method For Rapid Measurement Of Rice Degree Of Milling. Cereal Chem," 75 (3), pp. 380-385, 1998.
- [13] J. G. Lopez, S. Ventura, and A. Cano, "Distributed nearest neighbor classification for large-scale multi-label data on spark," Future Generation Computer Systems, pp. 1-49, 2018.
- [14] M. Nanja and P. Purwanto, "The k-nearest neighbor method is based on forward selection to predict the commodity price of pepper," J. Pseudocode, vol. 2, no. 1, pp. 53-64, 2015.
- [15] M. G. Alfianto, R. N. Whidhiasi, and Maimunah, "Rice identification based on color using adaptive neuro fuzzy inference system. Journal of computer science research," Embedded Systems & Logic. 5 (2), pp. 51-59, 2017.
- [16] N. Wijaya and A. Ridwan, "Classification of types of apples using the K-Nearest Neighbors Method," SISFOKOM Journal, vol. 08, no 01. 74-78, 2019.
- [17] A. Pazoki and Z. Pazoki, "Classification system for rain fed wheat grain cultivars using artificial neural network," African Journal of Biotechnology, 10 (41), pp. 8031-8038, 2011.
- [18] S. Sansomboonsuk and N. Afzulpurkar, "The appropriate algorithms of image analysis for rice kernel quality evaluation," The 20th Conference of Mechanical Engineering Network of Thailand, October 2006.
- [19] C. Sun, M. Berman, D. Coward, and B. Osborne, "Thickness measurement and crease detection of wheat grains using stereo vision," Pattern Recognition Letters, 28, pp. 1501-1508, 2007.
- [20] T. Y. Kuo, C. L. Chung, S. Y. Chen, H. A. Lin, and Y. F. Kuo, "Identifying rice grains using image analysis and tenuous representation-based classification," Computers and Electronics in Agriculture, 127, pp. 716-725, 2016.
- [21] N.S. Visen, J. Paliwal, D.S. Jayas, and N.D.G. White, "Image analysis of bulk grain samples using neural networks," Canadian Biosystems Engineering, 46, pp. 711-715, 2004.
- [22] L. Yan, S.R. Lee, S.H. Yang, and C.Y. Lee, " CCD Rice Grain Positioning Algorithm for Color Sorter with Line CCD Camera," The Online Journal on Power and Energy Engineering (OJPEE), 1 (4), pp. 125-129, 2010.
- [23] Y. Dong, X. Ma, and T. Fu, "Electrical load forecasting: A deep learning approach based on K-nearest neighbors," Applied Soft Computing Journal, 1-31, 2020.
- [24] Z. Maula, C. Rahmad, and U.D. Rosiani, "Application development for tomato fruit selection for superior seeds based on color and size using HSV and thresholding," J. Teknol. Inf. Theory. Concept, and Implementation, vol. 7, no. 2, pp. 127-138, 2016.