Selection of Superior Rice Seed Features Using Deep Learning Method

Dinda Ayusma Tonael[1], Yampi R. Kaesmetan[2], Marinus I. J. Lamabelawa[3]
[1,2,3]Department of Informatics Engineering, STIKOM Uyelindo Kupang
Email: [1]dindaayusma11@gmail.com, [2]kaesmetanyampi@gmail.com, [3]ignaslamabelaw4@gmail.com

Article Info
Article history:
Received Aug 28th, 2021
Revised Sept 28th, 2021
Accepted Sept 30th, 2021

Keyword:
Classification
GLCM
KNN
SVM
Rice Seeds

ABSTRACT
Indonesia is a tropical country known as an agricultural country, where 88.57% of the population works in the agricultural sector (BPS Indonesia, 2020). Indonesia is rich in agricultural products such as rice, soybeans, corn, peanuts, cassava and sweet potatoes. Rice (Oryza sativia L) is one of the most dominant food commodities for the people of Indonesia. The carbohydrate content per 100 grams of rice reaches 79.34 grams. The main benefit of rice is as a source of carbohydrates and a source of energy for the body. Seed is one of the factors that play a role as a carrier of technology in advanced agriculture, therefore the seeds used must be of good quality. Farmers tend to equate rice seeds from previous harvests, the rice seed classification process is carried out manually through visual observation and soaking rice seeds in a container filled with water, submerged and floating rice seeds are selected for use, and those that float are discarded. But in reality it still produces less than optimal results, for example rice that is less dense and cracked. This study uses a color moment to be extracted using GLCM (gray level co-occurence matrix) then classified with k-NN to determine the class, then uses the SVM model to display the best hyperplane line to separate the two classes, namely superior and non-superior classes after that system tested with confusion matrix. With a continuous and more intense work process, the research entitled Selection of Superior Rice Seed Features Using Deep Learning Methods. The output of this research leads to a conclusion which rice seeds are superior and which are not superior, aiming to optimize the yield of rice with better quality. The research was successfully carried out using the deep learning method with the highest accuracy of 92.85%.

Corresponding Author:
Yampi R Kaesmetan,
Study Program Informatics Technique, STIKOM Uyelindo Kupang
Jl. Perintis Kemerdekaan I, Kayu Putih, Oebobo, Kota Kupang, NTT
Email: kaesmetanyampi@gmail.com
DOI: http://dx.doi.org/10.24014/ijaidm.v2i2. 13947

1. INTRODUCTION

Rice (Oryza sativia L) is one of the most dominant food commodities for the people of Indonesia, where rice that is processed into rice is useful as a source of carbohydrates for the body. The carbohydrate content per 100 g of rice reaches 79.34 grams. That is why Indonesians do not feel full if they have not eaten rice, even though they have consumed other foods. Rice is the fastest energy processed by the body, so it can be concluded that the main benefits of rice are as a source of carbohydrates and a source of energy for the body [13]. In 2019, rice productivity in Indonesia decreased by 4.60 million tons or 7.76% [2].

Seed is one of the factors that play a role as a technology conductor in advanced agriculture, contained in the genetic potential of varieties (BB Padi 2009). Ministry of Agriculture No 12 of 2018 article 27 paragraph 1 states that to determine the suitability of seed quality in the form of seeds, laboratory tests are carried out. Seed quality testing in the form of seeds in the laboratory is usually done visually. But observation requires special skills. According to Mulsanti et.al (2013), direct visual morphological observations tend to be subjective in data collection between analysts and the efficiency level is low. The time used is also relatively long to identify seeds based on shape. Farming communities tend to cultivate their own rice seeds by obtaining rice from the previous harvest. When the rice planting season approaches, farmers select which seeds to choose. During the soaking process there will be separation of rice seeds which will be divided into 3 groups, namely the most basic or sinking part, while the second part that floats and the second part that floats. that floats on the surface. The submerged and floating rice seeds were selected for planting. Floating rice seeds are not recommended. The Department of Agriculture and Food Security categorizes rice varieties into three categories, namely hybrid rice varieties, superior rice varieties and local rice varieties, each of which has its own advantages and disadvantages. In producing good quality rice, it is necessary to select superior rice seeds because they have an impact on the vigor and viability of the rice seeds themselves[8].

There are several previous studies conducted by Liantoni F. (2015) applying the K-Nearest Neighbor method for leaf classification by improving image features. In this study, leaf pattern recognition was carried out by recognizing leaf structural characteristics such as the shape and texture of each leaf. The results of the study using the K Nearest Neighbor method obtained an accuracy value of 86.67% [6]. Further research was conducted by Neneng et.al (2016), with the title Support Vector Machine for image classification of meat types based on texture using Gray Level Co-Occurrence Matrix (GLCM) feature extraction. This study uses the SVM method to classify images of goat meat, buffalo meat, horse meat, and beef with a distance of 20 cm, 30 cm, and 40 cm, using four GLCM directions, namely 0^o, 45^o, 90^o, and 135^o. This study resulted in the best recognition rate of 87.5% at a distance of 20 cm with a neighboring pixel distance of d=2 in the GLCM 135o direction [9]. The next research was conducted by Taek (2017), with the title Sorting the Quality of Rice Seed in Kupang Regency with the k-Nearest Neighbor Classification. In this study, feature extraction of 100 rice seed images was carried out using statistics (color moment) with characteristics such as mean, standard deviation and skewnees. Furthermore, classification will be carried out using the k-Nearest Neighbor method with two ways of measuring similarity, namely based on Euclidean and Minkowski to produce suitable and unfeasible rice seeds. This study resulted in an accuracy rate of 82% [15]. The research was conducted by Hasanah and Nafi'iyah (2020), with the title Classification of Bulbs Based on Image Using SVM and KNN. The inability of farmers' memory for each shape and texture of this variety of sweet potato is the reference in this study. In this classification process, a grayscale process will be carried out (changing a color image to grayish), a thresholding process (separating objects from the image background), then a shape feature extraction process will be carried out with features such as (Area, Perimeter, Metric, Major Axis, Minor Axis, Eccentricity) and texture feature extraction using the GLCM (Gray Level Cooccurrence Matrix) method with characteristics such as (Contrast, Correlation, Energy, Homogeneity), after that the classification process is carried out with SVM and KNN [5]. Thus farmers can easily identify the type of tubers.

Therefore, to find out which rice seeds are superior and not superior, it is necessary to develop technology for classifying the shape, texture, color of rice seeds based on shape, texture, color that is more objective and efficient, so that more accurate and consistent analysis results are needed. This study uses a color moment to be extracted using GLCM then classified with k-NN to determine the class, then uses the SVM Model to display hyperplane graphs from the two different classes and is evaluated with a confusion matrix, with this ongoing process this will lead to a conclusion which rice seed is the best. The research that will be conducted is entitled Feature Selection of Superior Rice Seeds Using Deep Learning Methods.

2. RESEARCH METHOD

This research aims to classify superior rice seeds based on the texture of rice seeds so as to optimize yields with better quality. This research uses a deep learning approach with a combination of several methods. The details of the research work step by step can be seen in Figure 1.

2.1. Literature Study

Looking for materials from sources in the form of books, journals, websites, internet and other sources that can be used as author references.

2.2. Rice Seed Image Data

Rice seed image data was obtained from data collection through observation and interviews to obtain sample images. The criteria for seeds that have superior quality include seeds that are intact or contain solid, bright seeds, seeds that are not easily broken. The following is an explanation table regarding each seed criteria at table 1.

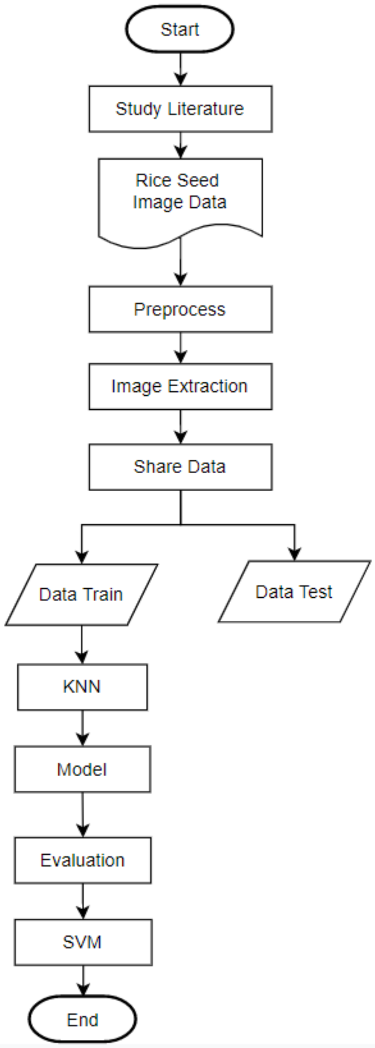


Figure 1. Research Flowchart

Table 1. Seed Characteristics

No	Physical Quality	Definition
1	Whole Rice Seeds	Seeds that are physically full and bright overall without any spots or mottling
2	Broken Rice Seeds	Seeds that are not intact, broken, damaged by the process
3	Dull Rice Seeds	Rice seeds that tend to be dark brown-black in color
4	Empty Rice Seeds	Rice seeds that do not contain or are light

Rice seed images were taken from superior rice varieties consisting of two types of rice seeds, namely Super Ciherang and Inpari 39. The rice seed sample will be carried out in an acquisition process, namely the transfer from an analog image to a digital image using a Nikon DSLR d7100 digital camera with a shooting distance of 60 cm, ISO -500, exposure bias -0.7 step, flash mode condition on godox tt600. The image of rice seeds placed on a black flaxen cloth. Rice seed image taken with initial dimensions of 1336 x 1221 pixels with JPG file format.

2.3. Preprocess

The purpose of preprocessing is to convert the data into a format that is easier and more effective to process to obtain more accurate values and reduce computation time for large-scale problems, thereby making the data values smaller without changing the information contained in the image data [4]. The process carried out at this stage is resizing or changing the dimensions to a smaller size with the aim that the system can easily detect the texture of rice seeds.

Furthermore the RGB image conversion (red, green, blue) is converted into image form grayscale so that it can be processed by a computer. An image is represented in 8 bits, it means that in the image there are

2^8 or 256 grayscale levels. Usually a value of 0 – 255, where 0 indicates the darkest intensity level and 255 indicates the lightest intensity, [10]. Formulated by the equation: Gray value = $\alpha R + \beta G + \delta B$

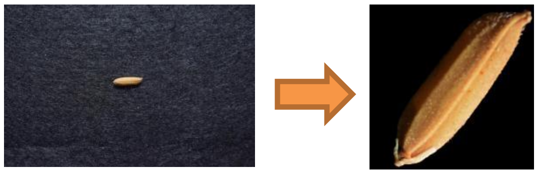


Figure 2. Image before and after cropping

with value α = value for R(0.35), β = value for G (0.25), δ = value for B (0.4), so that the value of

$$\alpha + \beta + \delta = 1 \quad (1)$$

Then proceed to the filtering stage to calculate the median value. The median filter serves to smooth and reduce noise or interference in the image. In the median filter, a window or filter containing an odd number of pixels is shifted point by point over the entire image area. The values in the window are sorted in ascending order and then the median value is calculated using the equation:

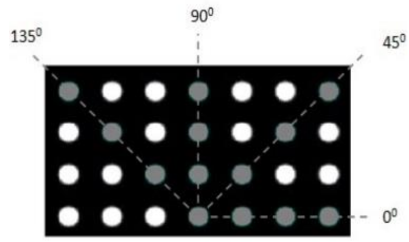
$$f(y,x) = median(p,q)\varepsilon syx(g(p,q)) \quad (2)$$

where $f(y,x)$ = result weight on position (y,x),g(p,q) = gauss kernel matrix element at position (p,q).

2.4. Image Extraction

Feature extraction technique using Gray Level Coocurency Matrix (GLCM) shows the probability of the neighboring relationship between two pixels that make up a co-occurrence matrix of the image. This process is related to the quantization of image characteristics into a group of appropriate feature values. The characteristics obtained are in the form of a homogeneous matrix, which will be input data into the algorithm before the classification process is carried out [1].

The co-occurrence matrix is formed from an image by looking at paired pixels that have a certain intensity. The use of this method is based on the hypothesis that in a texture there will be a repetition of configurations or pairs of gray levels. Co-occurrence means co-occurrence, namely the number of occurrences of one level of neighboring pixel values with one level of other pixel values within a certain distance (d) and angle orientation (θ). Distance is expressed in pixels and orientation is expressed in degrees. The orientation is formed in four angular directions with angular intervals of 45°, namely 0°, 45°, 90°, and 135° [11].



Figur 3. GLCM with angles 0^o , 45^o , 90^o , and 135^o

While the distance between pixels is usually set at 1 pixel, 2 pixels, 3 pixels and so on. The co-occurrence matrix is a square matrix with the number of elements as much as the square of the number of pixel intensity levels in the image. Each point (i,j) in the oriented co-occurrence matrix contains the probability of the occurrence of a pixel of value i being neighbors with a pixel of value j at a distance d and orientation (180−θ). Texture analysis is better at presenting image textures in measurable parameters, such as energy, contrast, homogeneity and correlation. The calculation of the four features can be described as follows:

a) Energy

Represents a measure of the concentration of pairs with a certain gray intensity in the matrix.

$$Energy = \sum i,j \, p_d^2(i,j) \quad (3)$$

where p(i, j) represents the value in row i and column j in the co-occurrence matrix.

b) Contrast
Is the result of calculations related to the amount of variation in the intensity of gray in the image.

$$f_3 = \sum_i \sum_j (i-j)^2 p_d(i,j) \tag{4}$$

c) Homogeneity

$$f_4 = \sum_i \sum_j \frac{p_d(i,j)}{1+|i-j|} \tag{5}$$

d) Correlation
Shows a measure of the linear dependence of the gray level of the image so that it can provide an indication of the existence of a linear structure in the image.

$$F_5 = \sum_i \sum_j \frac{ijPd(i,j) - U_x U_y}{\sigma_x \sigma_y} \tag{6}$$

2.4. Share Data

Image data that has gone through the preprocessing and feature extraction stages, then divided by k-fold cross validation is a scenario where the data is divided to be used as testing data for each fold. The image of rice seeds is 140 inpari 39 seeds and super ciherang, the k used is 10 fold cross validation. The data set is divided into training data and testing data by 90% for training data and 10% for testing data [4]. Figure 10-fold cross validation is shown in the image below:

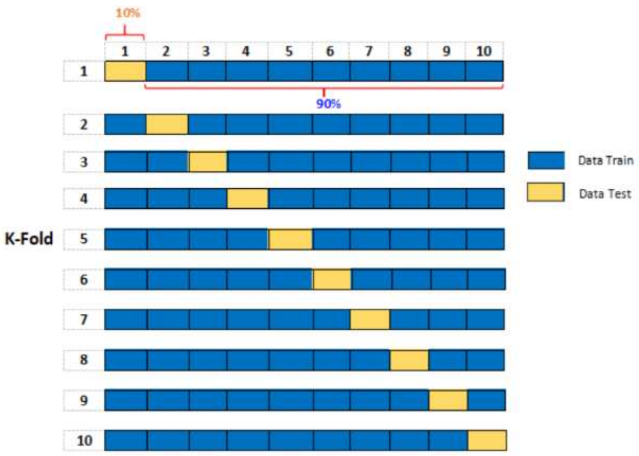


Figure 4. 10-fold cross validation

2.5. K- Nearest Neighbor

The k-Nearest Neighbor (k-NN) algorithm is a method for classifying objects based on the learning data that is closest to the object. Learning data is projected into a multidimensional space, where each dimension represents a feature of the data.

The decision making of the KNN algorithm is to find a test sample with k closest neighbors or the most similar training samples in the feature space, then the test sample is assigned to a majority vote of k closest neighbors. Distance is a commonly used approach to achieve image search. Its function is to determine the similarity or dissimilarity of two feature vectors. One way to calculate the distance near or far from neighbors is using the euclidean distance method. euclidean distance serves to test the size that can be used as an interpretation of the proximity of the distance between two objects. [18]. The form of the k-NN equation for Euclidean is as follows:

$$dist(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{7}$$

where,
x = sample data
y = test data
dist = distance

2.6. Support Vector Machine (SVM)

The results of the classification with k-NN will also be displayed in the form of a graph or figure where using the Support Vector Machine (SVM) model, which is to find the best hyperplane that separates the two superior and non-superior classes by reading the two variables x and y.

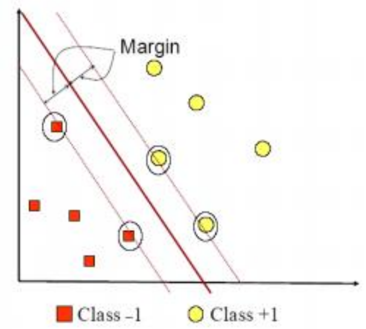


Figure 5. Support Vector Machine

Available data is denoted as $\vec{x}_i \in \Re^d$, while the respective labels are denoted $y_i \in \{-1,+1\}$ for i=1,2,3 …. l. which l is the amount of data. Assuming both classes –1 dan +1 can be completely separated by a d-dimensional hyperplane, which is defined:

$$\vec{w}.\vec{x} + b = 0 \tag{8}$$

Pattern $\vec{w}$ which includes class –1 (negative sample) can be formulated as a pattern that satisfies the inequality

$$\vec{w}.\vec{x} + b \le -1 \tag{9}$$

Whereas pattern $\vec{w}$ which includes class +1 (positive sample)

$$\vec{w}.\vec{x} + b \ge +1 \tag{10}$$

The results of the extraction there are four variables that will be randomized to get the best hyperplane between energy and contrast, homogeneity and correlation, homogeneity and energy, contrast and correlation.

2.7. Evaluation

Evaluation of determining the quality of rice seeds based on image texture is carried out to predict how well the label classification is in the tuple class. The results of the classification using k-NN will be tested using a confusion matrix. This method uses the matrix table below if the data set only consists of two classes, one class is considered positive and the other negative. With the provision that true positive (TP) is a positive tuple that has a true value on the classification label. True negative (TN) is a positive tuple which is interpreted as negative. Sensitivity and specificity are also called positive tuples which are true, while specificity is negative tuples which are true. How to find the value of sensitivity and specificity can be formulated by [4].

		Predicted Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

Figure 6. Confusion Matrix

3. RESULTS AND ANALYSIS

3.1. Data Acquisition

The data used are 140 rice image data. Image data will be divided into two parts, 126 images for training images and 14 for testing images, then inputted into Matlab to obtain matrix data to be processed.

Image data consists of two classes, namely superior image and inferior image. The following is the average data from feature extraction using GLCM.

Table 2. Data on the average value of GLCM extraction

	Contrast	Correlation	Energy	Homogeneity
1	0.0631	0.9867	0.5993	0.9834
2	0.1025	0.9879	0.4688	0.9721
3	0.0696	0.9806	0.7092	0.9839
4	0.0505	0.9893	0.6178	0.9848
5	0.0425	0.9851	0.6559	0.9869
	:	:	:	:
140	0.0543	0.9835	0.6149	0.9834

3.2. Implementation System

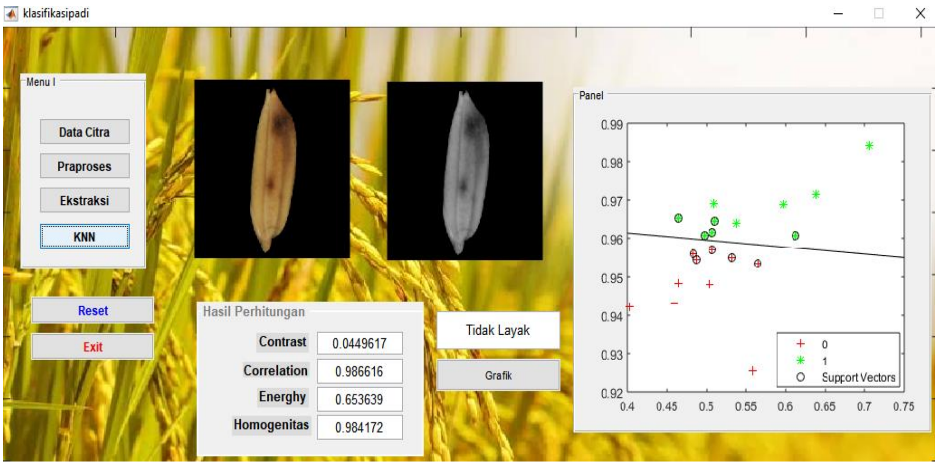


Figure 7. Classification Form

Based on Figure 7 above, the system is able to correctly classify the new test data into the non-excellent class and the graph with the SVM model is able to find the best hyperplane that can separate the two superior and non-excellent classes.

3.3. System Testing

Tests are carried out to determine how much accuracy k-Nearest Neighbor is in classifying classes. The K value used is 5 with 140 data samples. The test uses a confusion matrix to find the level of accuracy, sensitivity and specificity. Meanwhile, k-fold cross validation is used to divide training data and testing data with a ratio of 1:9 between training data and testing data. The test was carried out 10 times. The following are the results of the tests carried out, the results of the comparison in test 1 are entered into the confusion matrix to calculate the accuracy, sensitivity and specificity of the system.

Table 7. Fold Validation 7

Fold 7			
		Prediction Class	
Actual Class		Class 1	Class 0
	Class 1	6	1
	Class 0	0	7

Accuracy $$\frac{6+7}{14} = \frac{13}{14} = 0.92 \text{ x } 100 = 92.85\%$$

Sensitivity $$\frac{6}{6+0} \text{x}100 = 100\%$$

Specificity $$\frac{7}{7+1} \text{x}100 = 87.5\%$$

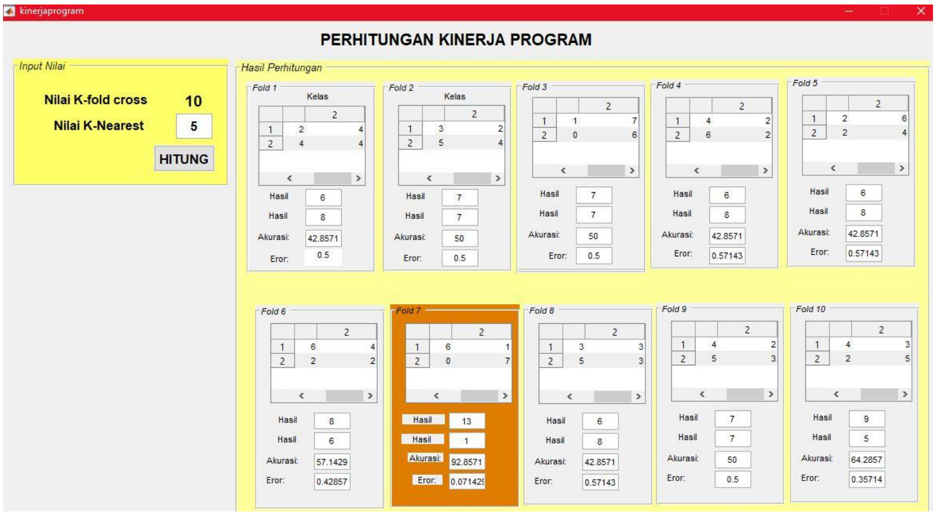


Figure 8. System Test Form

Based on the results of the evaluation with the confusion matrix divided into 10 fold cross validation by calculating the sensitivity, accuracy and specificity values, which can be seen in the graph below;

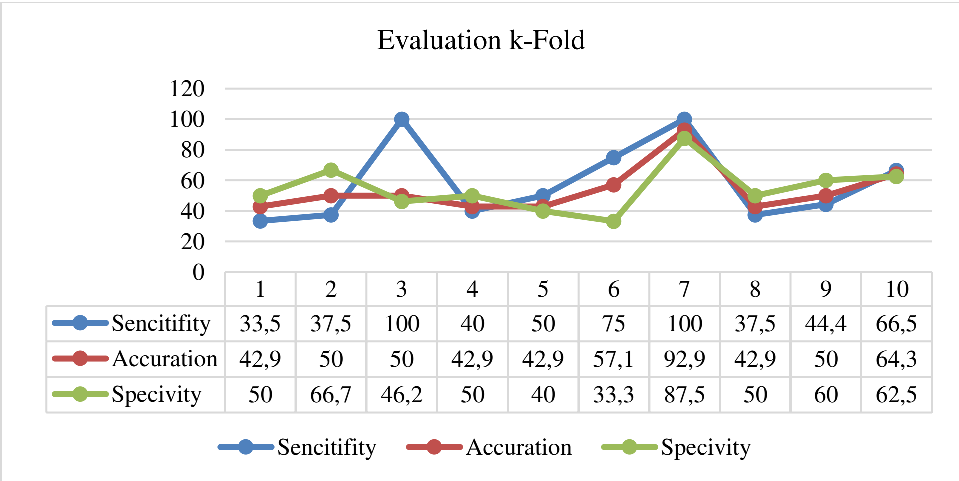


Figure 9. Calculation results of accuracy, sensitivity and specificity

From table 8 above, it can be seen that the lowest sensitivity results are in the 1st fold, which is 33.33% and the highest in the 7th fold, which is 100%. The lowest accuracy results in the 1st fold is 42.85% and the highest in the 7th fold is 92.85% and the lowest specificity results in the 6th fold is 33.33% and the highest in the 7th fold is 87.5%. Based on the program testing, it can be concluded that the program was successful in classifying superior and not superior rice seeds until it reached an accuracy of 92.85%.

3 CONCLUSION

Based on the results of this study, it can be concluded that the selection of superior rice seeds with feature extraction (contrast, correlation, energy and homogeneity) with the k-Nearest Neighbor classification obtained a validation test for sensitivity of 100%, specificity of 87.5% and accuracy of 92.85% in the 7th fold. From these results, it can be concluded that the program succeeded in selecting superior and not superior rice seed classes.

REFERENCES

[1] Agmalaro, M. A., Kustiyo A. and Akbar, A. R., 2012. Tropical Fruit Plants Identification Based on Leaf Surface Texture Image Using Artificial Neural Network. IPB Journal of Computer Science and Agri-Informatics. [Internet]. [Accessed 04 March 2021]; 2(2); 73-82. Available at: https://journal.ipb.ac.id/index.php/jika/article/view/7996/6283

[2] Badan Pusat Statistik. 2020. Statistik Tenaga Kerja. Jakarta (ID): Badan Pusat Statistik

[3] Gonzalez, C.R. and Woods, E.R. 2013. Digital Image Processing. New Jersey (US): Boundary Prentice Hall.

[4] Han, J., Kamber, M. and Pei, J. 2012. Data Mining Concepts and Techniques Third Edition. America (US): Morgan Kaufmann.

[5] Hasanah, A. and Nafi'iyah, N., 2020. Classification of Bulbs Based on Image Using SVM and KNN. Journal of Spirit: STMIK Yadika Journal of Computing and Cybernetic System. [Internet]. [Downloaded November 21, 2020]; 12(1); 48-51. Available at: https://jurnal.stmikyadika.ac.id/index.php/spirit/article/view/17

[6] Liantoni, F. 2015. Leaf Classification With Image Feature Improvement Using K-Nearest Neighbor Method. Journal of Ultimatics. [Internet]. [Downloaded November 21, 2020]; 7(2); 2085-4552. Available at: https://ejournals.umn.ac.id/index.php/TI/article/view/356

[7] Mediatani. 2018. Types of Rice. Department of Food Security, Food Crops and Horticulture. [Internet]. [Accessed November 20, 2020]. Available at: https://www.dinastph.lampungprov.go.id/detail-post/tipe-tipe-padi

[8] Mulsanti, I. W., Surahman, M., Wahyuni, S. and Utami, D. W., 2013. Identification of Hybrid Rice Parental Lines With Specific SSR Markers and Their Utilization in Seed Purity Test. Journal of Agricultural Research and Development. [Internet]. [Downloaded November 21, 2020]; 32(1); 2541-5166. Available at: http://ejurnal.litbang.pertanian.go.id/index.php/jpptp/article/view/3012

[9] Neneng., Adi, K. and Isnanto, R. R., 2016. Support Vector Machine for image classification of meat types based on texture using Gray Level Co-Occurrence Matrix (GLCM) feature extraction. Journal of Business Information Systems. [Internet]. [Downloaded November 21, 2020]; (6); 1-10. Available at: https://ejournal.undip.ac.id/index.php/jsinbis/article/view/10510/pdf

[10] Putra, D. 2010. Digital Image Processing. Yogyakarta (ID): Andi Yogyakarta.

[11] Rizal, R. A., Gulo, S. and Della, O., 2019. Analysis of Gray Level Co-Occurrence Matrix (GLCM) in Recognizing Facial Expression Images. Mantik Journal. [Internet]. [Downloaded January 20, 2021]; 3(2); 31-38. Available at: https://iocscience.org/ejournal/index.php/mantik/article/view/497/302

[12] Santosa, B. 2007. Data Mining: Data Utilization Techniques for Business Purposes. Yogyakarta (ID): Graha Ilmu.

[13] Suliartini, N.W.S., Sadimantara, G. R., Wijayanto, T. and Muhidin., 2011. Testing of Anthocyanin Levels in Upland Rice of Red Rice from Germplasm Collection of Southeast Sulawesi. Crop Argo Scientific Journal of Cultivation. [Internet]. [Downloaded November 20, 2020]; 4(2); Available at: https://www.cropagro.unram.ac.id/index.php/caj/article/download/100/83

[14] Syahri. and Somantri, R. U., 2016. The use of superior varieties that are resistant to pests and diseases supports increasing national rice production. Journal of Agricultural Research and Development. [Internet]. [Downloaded December 20, 2020]; 35(1); 25-36. Available at: http://ejurnal.litbang.pertanian.go.id/index.php/jppp/article/view/3466/2954

[15] Taek, M. Y. 2020. Sorting Quality of Rice Seed in Kupang Regency with k-Nearest Neighbor Classification. [Essay]. Kupang (ID): Informatics Engineering, STIKOM Uyelindo Kupang.

[16] Tampani, A. Y. 2016. Comparison of Image Texture Extraction with Statistical Methods of Order I and Order II Statistics for Corn Seed Selection. [Essay]. Kupang (ID): Informatics Engineering, STIKOM Uyelindo Kupang.

[17] Utama, M.Z.H. 2015. Rice Cultivation on Marginal Land Tips to Increase Rice Production. Yogyakarta (ID): Andi Yogyakarta.

[18] Zhu, S,. Dong, W. and Liu, W., 2015. Hierarchical Reinforcement Learning Based on KNN Classification Algorithms. International Journal of Hybrid Information Technology. [Internet]. [Downloaded February 20, 2021]; 8(8); 175-184. Available at: https://pdfs.semanticscholar.org/01e7/7a073529d86a7d3eb161267056a98a173877.pdf?_ga=2.258547018.1337612652.1616398487-136 5745971.1616 398487

[19] Agmalaro, M. A., Kustiyo A. and Akbar, A. R., 2012. Tropical Fruit Plants Identification Based on Leaf Surface Texture Image Using Artificial Neural Network. IPB Journal of Computer Science and Agri-Informatics. [Internet]. [Accessed 04 March 2021]; 2(2); 73-82. Available at: https://journal.ipb.ac.id/index.php/jika/article/view/7996/6283

BIBLIOGRAPHY OF AUTHORS

Dinda Ayusma Tonael is a student at STIKOM Uyelindo Kupang who recently earned his S.Kom degree

Yampi R. Kaesmetan, S.Kom, M.Kom, now works at STIKOM Uyelindo Kupang as a lecturer in the Informatics Engineering Study Program.

Marinus I. J. Lamabelawa, S. Kom., M.Cs. now works at STIKOM Uyelindo Kupang as the Chairperson of STIKOM Uyelindo Kupang and a lecturer at the Informatics Engineering Study Program.