

# Digital Image Processing using Texture Features Extraction of Local Seeds in Nekbaun Village with Color Moment, Gray Level Co Occurance Matrix, and k-Nearest Neighbor

Yampi R Kaesmetan<sup>1</sup>, Marlinda Vasty Overbeek<sup>2</sup>

<sup>1</sup>Informatic Technique Study Program, STIKOM Uyelindo Kupang, Kupang City, Indonesia

<sup>2</sup>Faculty of Engineering and Informatic, Multimedia Nusantara University, Tangerang, Indonesia

<sup>1</sup>kaesmetanyampi@gmail.com <sup>2</sup>marlinda.vasty@umn.ac.id

Accepted 12 May 2021

Approved 8 August 2021

**Abstract**— The problem in determining the selection of corn seeds for replanting, especially corn in East Nusa Tenggara is still an important issue. Things that affect the quality of corn seeds are damaged seeds, dull seeds, dirty seeds, and broken seeds due to the drying and shelling process, which during the process of shelling corn with a machine, many damaged and broken seeds are found. So far, quality evaluation in the process of classification of the quality of corn seeds is still done manually through visible observations. Manual systems take a long time and produce products of inconsistent quality due to visual limitations, fatigue, and differences in the perceptions of each observer. The selection of local maize seeds in Timor Island, East Nusa Tenggara Province, especially in Nekbaun Village, West Amarasi District with feature extraction with a color moment shows that the mean, standard deviation and skewness features have an average validation of 88% and use the GLCM method which shows the neighbor relationship. Between the two pixels that form a co-occurrence matrix of the image data, namely GLCM, it shows that the features of homogeneity, correlation, contrast and energy have an average validation of 70.93%. The k-Nearest Neighbor (k-NN) algorithm is used in research to classify the image object to be studied. The results of this study were successfully carried out using k-Nearest Neighbor (k-NN) with the euclidean distance and  $k = 1$  with the highest extraction yield of 88% and the results of GLCM feature extraction for homogeneity given 75.5%, correlation of 78.67%, contrast given 65.75% and energy given 63.

**Index Terms**— corn seeds; color moment; digital image processing; gray level co occurrence matrix; k-nearest neighbor

## I. INTRODUCTION

Corn (*Zea mays* L.) is one of the food crops that contain carbohydrates and is a food that is often consumed by the public in general, in addition to wheat, sweet potatoes and rice. Corn is an agricultural

product that is widely planted by farmers because the duration of harvest is faster and plays an important role in the development of the agricultural sector. In Indonesia, especially in East Nusa Tenggara, corn as a staple food is often consumed as a substitute for rice. Where there are sources of protein, carbohydrates, fiber, so that they can help increase endurance, both for consumption of corn that is still not old or that is ready to be harvested. In addition, corn can also be used as food for domesticated livestock such as from young corn stalks, leaves or cobs. Therefore, the demand for corn continues to increase from year to year, this is due to the increase in the economic standard of living of the community and the progress of the animal feed industry so that the quality of corn needs to be considered. There are several factors that affect the quality of corn seeds, one of which is the high level of damage that occurs during the corn shelling process by machine so that many damaged or cracked seeds are found. So far, quality evaluation in the process of classifying corn seed quality is still done manually through direct observation with the naked eye. Visual observation takes a long time so that it can produce products with uneven quality because it has several limitations such as fatigue, and differences in the perception of observers conducting research [1].

Determination of the quality of good corn seed selection has been carried out using supervised learning techniques. One example of a supervised learning method that can be used to determine the quality of corn seeds for seeding is the classification of coffee beans using image processing and fuzzy logic. This research aims to build a digital grading system by using a camera to take pictures of coffee beans samples, then the computer calculates the color and texture. The data test shows that the green color and the entropy, energy and contrast texture characteristics are formed from the co-occurrence matrix with 135

degrees that have been obtained and this is an important parameter of the fuzzy C means in assessing the quality grade of coffee beans [2].

In addition to the classification of coffee beans, there is also research on the determination of the quality of soybeans. Where in the selection of seeds there are broken seeds, pale seeds, dirty and unclean seeds, and broken or cracked seeds due to the drying and firing process that is too long. Soybean quality determination is usually done manually with visual observation. In conducting manual observations can take a long time and produce products with unstable quality because observations are made with the naked eye, fatigue from observers, and different perceptions of each observer when making observations in the field. This research was conducted by using a comparison of image texture extraction using statistical methods of order I (color moment) and statistics of order II (Gray Level Co occurrence Matrix - GLCM) for soybean selection. The first order statistic (color moment) shows the probability that the gray level value of the image pixel will appear, while the second order statistic (GLCM) shows the probability of an adjacency relationship between two pixels that form the cohesion matrix of the image data. classification process in determining soybean seeds, and getting an average accuracy of 70.02% [3].

In addition to soybean seed classification research, there is also research on determining the quality of corn seeds for seeding based on color brightness using a support vector machine. The purpose of this study was to classify corn seeds for seeding, so as to find the right corn seeds for seeding based on the brightness level of color. The results of this study were successfully carried out using a support vector machine with a polynomial kernel function and got the highest accuracy, namely yellow corn by 82% and for white corn by 76% [4].

The problem in determining the selection of corn seeds for re-planting, especially corn in East Nusa Tenggara is still an important issue. The declining price of corn in the market caused by damage of corn seeds that will be used for seeds usually occurs when storing corn seeds that have peeled skin making it easier for fungi to grow quickly, especially from the *Aspergillus* type which has the potential to cause aflatoxins [5]. The selection of corn seeds is often done by farmers by naked eye or manually without looking at the physical characteristics, textures and colors of the corn seeds that will be used for re-seeding. A good selection of corn seeds is corn seeds that are not peeled off, not hollow and black. From the reasons that have been explained, in this research we develop a quality seed selection system using feature extraction based on color moment and GLCM features. While the learning model was developed with supervised learning, k-Nearest Neighbor.

## II. LITERATURE REVIEW

### A. Corn (*Zea mays L.*)

Corn is an annual crop, the growth process of corn plants is completed within three months to six months. The first half of the cycle is the vegetative growth stage and the second half is the generative growth stage. Corn is a type of grain food crop (cereal) from the grass family. The following is a systematic (lineage) of corn plants [6] :

Kingdoms	: <i>Plantae</i> (plant)
Division	: Magnoliophyta
Class	: Liliopsida
Sub Class	: <i>Commoninidae</i>
Order	: <i>Poales</i>
Family	: <i>Poaceae</i> (grass tribe)
Genus	: <i>Zea</i>
Species	: <i>Zea mays I.</i>

### B. Corn Seeds

Seeds are plant material that will be used for replanting so that they can be used as a means of multiplying similar plants. Corn seeds to be used as seeds must go through a process in such a way that they can be used for the replanting process [6]. The selection of corn seeds must be in accordance with the land to be planted, in order to get good quality corn. Corn seeds must be selected carefully, because corn production depends on seed selection. The criteria for seeds that have low quality include defective or damaged seeds, dull colored seeds, dirty seeds, broken seeds, and small seeds. The following is an explanation table regarding each of the physical quality criteria of corn shown in Table 1 [6].

TABLE I. DEFINITIONS FOR CORN SEED PHYSICAL QUALITY CRITERIA

No.	Physical quality	Definition
1	Corn with round seeds	Dried corn kernels that are physically intact without any spots, defects or fungus.
2	Defective seed corn	Corn with defective or damaged seeds due to insect attack or warehouse pests.
3	Corn with broken seeds	Corn with corn kernels that are not intact/damaged due to the threshing or shelling process.
4	Dull seed corn	Colored corn kernels tend to be dirty or dark

### C. Pre-Process

This stage is carried out to obtain data accuracy from the image of corn seeds that will be sampled, this process is to prevent data inaccuracies to get actual data. RGB image or commonly called true color is an image that can represent the color of an object that resembles the original by combining three colors that are often used, namely red (R), green (G) and blue (B). Each pixel of an RGB image has three channels that

represent each component of the basic color [7]. A gray image is a digital image that has only one channel value for each pixel, in other words, the value of the red = green = blue part. This value is used to indicate the level of intensity. The color of the three grayscales is a gray color with various levels from black to white. Grayscale images can be obtained from RGB images. The intensity value of the grayscale image is calculated from the intensity value of the RGB image using the eq. 1.

$$\text{Gray Value} = R + G + B / (\alpha + +) \quad (1)$$

With the value of  $\alpha$  = the value for R (0.35),  $\beta$  = the value for G (0.25),  $\gamma$  = the value for R (0.4), so the value of  $\alpha + \beta + \gamma = 1$ . The image that has been changed to The gray scale will be processed to remove noise using a median filter by finding the average of the image pixel values that have been sorted by the eq.2 :

$$f(y,x) = \text{median}(p,q) \text{csyx} (g(p,q)) \quad (2)$$

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

#### D. Digital Image

Digital images can be interpreted as light intensity on two sides and can be expressed in two dimensions  $f(x,y)$  where is the light intensity in discrete form about the axis  $x$  nor  $y$  which is the position of the coordinate point while  $f$  is the amplitude at position  $(x,y)$  which is often known as intensity or grayscale [8]. In assigning a value of a discrete intensity from 0 to 255, as well as the values of  $x,y$  and  $f(x,y)$  remain in a certain range or area but are in limited quantities. The image taken from the camera and the process of limiting the input of a wide set in discrete form is called a digital image. A digital image is composed of a number of gray level values called pixels at a certain position.

In the translation of light intensity can be calculated by the equation of two dimensions  $f(x,y)$  is:

$$0 < f(x,y) < \infty \quad (3)$$

For example,  $f$  is a 2-dimensional digital image measuring  $N \times M$ . So that it can be spread  $f$  in a matrix can be seen in the figure below, where  $f(0,0)$  is in the upper left corner of the matrix, while  $f(n-1, m-1)$  is in the lower right corner.

$$f(x,y) \begin{bmatrix} f(0,0) & f(0,1) & f(0,M-1) \\ f(N-1,0) & f(N-1,1) & f(N-1,M-1) \end{bmatrix} \quad (4)$$

#### E. Texture Analysis

The textured analysis used is a form of the intrinsic characteristics of an image form and is closely related to the level of roughness, granulation, and regularity of the structural arrangement of pixels. The textural aspects of an image can be used as the basis for segmentation, classification, and image interpretation [9].

So that the image texture can be interpreted as a function of the spatial variation of pixel intensity (gray value) in the image. Based on their shape, textures can be classified into two classification:

##### a. Macrostructure

The shape of the macrostructure has a local pattern repeating periodically in an image area, usually found in man-made patterns and tends to be easy to represent mathematically.

##### b. Microstructure

In microstructural texture, local and repeating patterns do not occur so clearly, so it is not easy to provide a comprehensive definition of texture.

Fig. 1 is an example of a texture that shows the difference between macrostructure and microstructure textures.

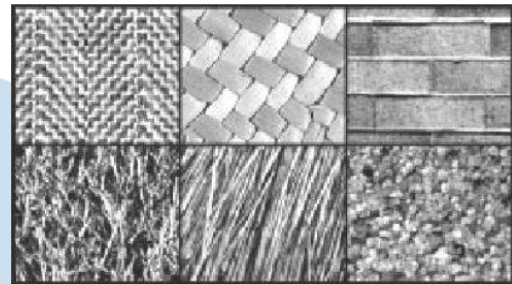


Fig. 1. Examples of visual textures [9]

#### F. Color Moment

Color moment is a form of representation of taking features based on the characteristics of the color of an image. From a histogram, it can show the probability of the occurrence of the gray level value of pixels in an image [9]. Color moments have 3 characteristics, there is:

##### 1). Mean ( $\mu$ )

Mean( $\mu$ ) describes the shape of the size of the dispersion of an image. The form of the equation to calculate the Mean ( $\mu$ ) is:

$$\mu = \sum_n f_n p(f_n) \quad (5)$$

Where ( $\mu$ ) is the average value of a color in the image,  $n$  is the total number of pixels in the image and  $f_n$  is a value of gray intensity, while  $p(f_n)$  shows the histogram value (probability of the occurrence of that intensity in the image).

In developing applications, the authors use the Firebase platform. Firebase is a service from Google that is used to facilitate application developers in developing applications. One of the functionalities of Firebase is a realtime database service that we use to store allergen data. Realtime database has the ability to store and synchronize application data in milliseconds.

The database will be hosted in the cloud. Data will be saved as JSON and then synchronized in realtime to

each client that has been connected so that all clients can receive the latest data updates automatically.

In this study, the database stores allergen data in the form of id, composition, and general name of the allergen. Fig. 5 is an example of a storage format in the database. "Komposisi" is the name of the allergen (processed / technical) which may appear on food packaging, "NamaGeneral" is the name of the main allergen of the ingredient, for example in Fig. 5 ingredients is "lactalbumin" which is processed from milk (susu) so that it has a generalized name "susu".

### 2). Standard deviation or standard deviation

The standard deviation or standard deviation is the most frequently used measure of variation (variation) of statistical data. The standard deviation or standard deviation is the square root of the variance. Form of calculation in standard deviation or standard deviation ( $\sigma$ ) that is:

$$\sigma = \sqrt{\frac{1}{n} \sum_{n=1}^j (f_n - \mu)^2 p(f_n)} \quad (6)$$

### 3). Skewness ( $\alpha_3$ )

Skewness ( $\alpha_3$ ) shows the degree of skewness (a measure of the degree of asymmetry) relative to the histogram curve of an image. Solution to calculate Skewness ( $\alpha_3$ ) that is:

$$\alpha_3 = \frac{1}{\sigma^3} \sum_n (f_n - \mu)^3 p(f_n) \quad (7)$$

### G. Gray Level Co-occurrence Matrix

In analyzing a pixel to obtain statistical characteristics by solving the probability value or often called the probability of an image which can be obtained from the neighboring relationship between two pixels at a certain distance and angle orientation. By using GLCM, the calculation process can work by forming a co-occurrence matrix from image data, then it will be continued by determining the characteristics as a function of the matrix [9].

The number of occurrences of joint calculations or co-occurrence, where it is 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 ( $\theta$ ). Distance is expressed in pixels and orientation is expressed in degrees. Orientation is formed in four angular directions with angular interval  $45^\circ$ , that is  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ . While the distance between pixels is usually set at 1 pixel.

In this case, 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 value of the point ( $p, q$ ) in the oriented co-occurrence matrix contains the probability of occurrence of a value of  $p$  neighbors with a pixel worth to  $q$  at distance  $d$  and the orientation of and ( $180^\circ$ ). The following is an illustration of Order II Statistical Feature Extraction which is shown in Fig 2.

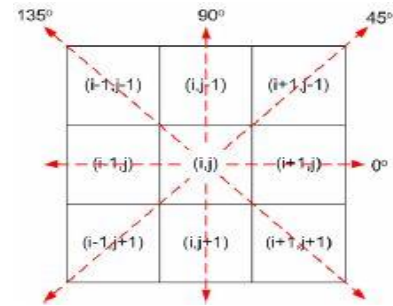


Fig. 2. Illustration of GLCM Feature Extraction [9]

In the form of completion of the GLCM feature extraction, and the co-occurrence matrix is obtained from each angle, the average co-occurrence matrix (Mavg) will be shown in the equation 8:

$$M_{avg} = (M(0) + M(45) + M(90) + M(135)) / 4 \quad (8)$$

GLCM (Gray Level Co-occurrence Matrix) has 14 haralick characteristics or features, but what is taken in this research are 4 characteristics, namely:

#### 1). Angular second moment (homogeneity)

Angular second moment (homogeneity) shows the homogeneity of the image. The form of the equation shown in equation 9:

$$ASM = \sum_i \sum_j \{p(i, j)\}^2 \quad (9)$$

Where  $p(i, j)$  represents the value on the row  $i$  and column  $j$  on the co-occurrence matrix.

#### 2). Contrast

Where in the form of contrast features can show the size of the spread (moment of inertia) of the image matrix elements. If it is located far from the main diagonal, the contrast value is large. Visually, the contrast value is a measure of the variation between degrees of gray in an image area. The form of the equation shown in equation 10:

$$CON = \sum_k k^2 [\sum_i \sum_j p(i, j)] \quad (10)$$

$$|i - j| = k$$

#### 3). Correlation

Where in Correlation can be shown with a linear dependence of the degree of gray in the image that has been obtained and can provide clues to the existence of linear structures in the image. Score  $\mu$  shows the average color in the image shows the square root of the variance (standard deviation). The form of the equation shown in equation 11:

$$COR = \frac{\sum_i \sum_j (ij) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (11)$$

#### 4). Energy

Energy show uniformity. Energy will be high when the pixel values are similar to each other otherwise it will be small indicating the value of the normalized GLCM is heterogeneous. The maximum value of energy is 1, which means that the distribution



of pixels is in a constant condition or in a periodic (not random) shape. The value of  $d$  indicates the distance between two pixels. The form of the equation shown in equation 12:

$$EN = \sum_i \sum_j p^2 d(i, j) \quad (12)$$

#### H. K-Nearest Neighbor

In understanding algorithm *k-Nearest Neighbor* (k-NN) which is a method for classifying objects based on the learning data that is closest to the object and already has a class label. Where data in learning is projected into a multidimensional space, where each dimension represents a feature of the data. This space is divided into sections based on the classification of learning data. A point in this space is marked class  $c$  if class  $c$  is the most common classification found in  $k$  the nearest neighbor of the point. Near or far neighbors are calculated based on Euclidean distance.

When compared with other classification methods, this method has a fairly high level of accuracy because the incoming data will be classified based on the similarity of existing characteristics from the previously classified data. However, on the algorithm *k-NN* needs to determine the value of the parameter  $k$  (number of nearest neighbors) and distance-based learning it is not clear what type of distance to use and which attributes to use to get the best results [10]. The *k-NN* for Euclidean distance is as count as follows:

$$D_{st} = \sqrt{\sum_{j=1}^n (X_{sj} - Y_{tj})^2} \quad (13)$$

With  $d(x,y)$  is a distance between points in training data data  $X_i$  and testing data points  $Y_i$  which will be classified, where  $x=x_1, x_2, \dots, x_i$  and  $y=y_1, y_2, \dots, y_i$  and  $i$  represents the attribute value as well as  $n$  is an attribute dimension.

### III. RESULT AND DISCUSSION

In this research, we used color moment and GLCM as a feature extraction and k-NN as a model classifying. The moment features used are: mean, standard deviation and skewness, while the haralick (GLCM) features used are: homogeneity, correlation, contrast, energy. The analysis of the results of the classification test with the K-Nearest Neighbor in terms of the use of the  $k$  value and the distance measurement method (Euclidean), as well as the analysis of system computational time. This system was developed with Matlab 10th student version.

The image of corn seeds taken in this study was 100 corn seed image data, with initial image dimensions of 2484 x 2134 pixels and saved in JPG file format. The image that will be sampled will be resized to 300 x 300 pixels, so that it gets an area of interest from the image, this image data will be called into Matlab to get the matrix data to be processed. The following are the characteristics of the image of decent

and unworthy corn seeds to serve as seeds. Table 2 shown the explanation of it.

TABLE II. DECENT AND UNWORTHY CORN SEED

No.	Decent seeds	Unworthy seeds
1.	Clean and whole seeds	Dirty seeds
2.	Seeds that are not black	Black seeds
3.	The unbroken seed	Broken seeds
4.	Undamaged seeds	Broken seeds
5.	Seeds that don't have holes	Hollow seeds

#### A. Feature extraction

In this process, the image that has been preprocessed will be extracted using *color* moments and GLCM.

#### B. Color moment

Color moment shows the probability of occurrence of a pixel's gray degree value in an image. The previously 100 x 100 image matrix will be converted into a 1 x 3 vector with a feature extraction process using HSV images. At this stage the image will also calculate the value of each feature. The characteristics used in this extraction include mean, standard deviation and skewness. The amount of data used is 100 corn image data so that the resulting 100 vectors. This vector is the input dataset for the classification process. The following is the syntax or command line in Matlab to get the result of color moment.

```
b=imresize(image1,[300
300]);
gbrHSV = rgb2hsv(b);
c1 =
imresize(gbrHSV,[300
300])
```



#### C. GLCM

GLCM shows the probability of adjacency relationship between two pixels at a certain distance and orientation angle. The previously 100 x 100 image matrix will be converted into a 1 x 1 vector with this extraction process using a grayscale image. At this stage the image will also calculate the value of each feature and also each direction. The characteristics used in this extraction include homogeneity, correlation, contrast and energy, while the directions used are  $0^0$ ,  $45^0$ ,  $90^0$ ,  $135^0$ . The amount of data used is 100 corn image data so that the resulting 100 vectors. This vector is the input dataset for the classification process. Following is the syntax or command line in Matlab to get the results from GLCM.

```
b=imresize(image1,[300
300]);
gbrgray=rgb2gray(b);
c1=imresize(gbrgray,[300
300])
```



#### D. Classification using k-NN

The classification using k-NN is divided into two processes, namely the training and testing process. The training process is used to produce a classification

model with k-NN which will later be used as a reference for classifying the quality of corn seeds to be used as seeds with new raw data. In this study, the authors used the Euclidean distance with k values of 1, 3 and 5.

The data used in this research is Lamuru corn image data typical of Nekbaun Village, West Amarasi District with 100 seed image data for color moment feature extraction and GLCM. The image data is then divided into 75 samples as training data and 25 samples as test data stored in an array. Each image data has dimensions of 1 x 3 for feature extraction of order I (color moment) of each feature (mean, standard deviation and skewness) and dimensions of 1 x 1 for extraction of order II features (GLCM) of each feature (homogeneity), correlation, contrast and energy). Training and testing data through preprocessing and feature extraction stages. The result of feature extraction of each image data has dimensions of 1 x 3 (color moment) and 1 x 1 (GLCM). All data are then arranged based on the composition of the training for k-NN based on the Euclidean distance with k values of 1, 3 and 5. The training data has a vector data dimension of a matrix of 100 x 3 corn seed image for color moment and the second training has a matrix vector data dimension of 100 x 1 corn seed image for GLCM. The target of the results of this study is to find the best results on k-NN based on the Euclidean distance with k values of 1, 3 and 5 nearest neighbors, which will be used to identify each test data, whether it is in class 0 or 1 (unworthy and decent). The scenario model used is 4-fold cross validation. The following is a scenario model in calculating the dataset used in determining the quality of corn seeds shown in Table 3.

TABLE III. MODEL SCENARIOS

Scenario name	Dataset (100 data)				Explanation
	data 1-25	data 26-50	data 51-75	data 76-100	
Scenario I	data 1-25	data 26-50	data 51-75	data 76-100	data 1-25 as test data, data 26-100 as training data.
Scenario II	data 1-25	data 26-50	data 51-75	data 76-100	26-50 data as test data, data 1-25 and data 51-100 as training data.
Scenario III	data 1-25	data 26-50	data 51-75	data 76-100	data 51-75 as test data, data 1-50 data 76-100 as training data.
Scenario IV	data 1-25	data 26-50	data 51-75	data 76-100	data 76-100 as test data, data 1-75 as training data.

### E. System Interface

The interface is a media liaison between the system and the user. System operation will start on this system interface page, making it easier for users to use this application. The following system interface can be seen in Figure 3.

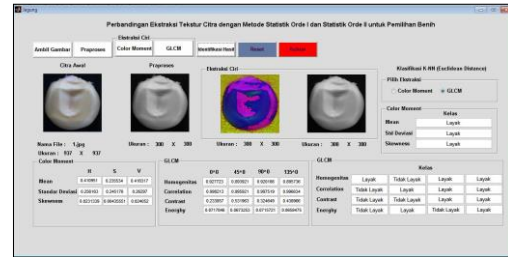


Fig. 3. System interface

### F. Test Result

System accuracy testing is done by measuring the performance of the corn seed quality determination system based on texture. In this research, the image tested is the image of corn. This test is carried out using the k-NN classification (euclidean distance) with k values of 1, 3 and 5. The test is seen from how well the classifier predicts the quality of corn seeds that are decent or unworthy for a seeds. Testing is done by looking for the value of sensitivity, specificity and accuracy of the classification system to determine the accuracy of this system. Based on the results of system testing that has been carried out, the sensitivity, specificity and accuracy values of the system are obtained. The result of this image processing is the generation of numerical data from each image of the corn seeds which will be separated into decent and unworthy corn seeds. The results of the average color moment feature extraction processing can be seen in Table 4.

TABLE IV. AVERAGE RESULT USING COLOR MOMENT FEATURURE EXTRACTION

Characteristic	The average value of suitable corn seeds for planting	The average value of corn seeds is not suitable for planting
mean	0.3516	0.3347
Standard deviation	0.2347	0.2189
Skewness	0.0174	0.0126

In this research, the k-NN classification is used with the distance used is Euclidean distance and uses feature extraction of order I or color moment. The data scenario used is a 4-fold cross validation. The accuracy results obtained from the k-NN classification for the mean is 90%, the standard deviation is 88% and the skewness is 86%. Validation is repeated once for each feature and the average is taken. From the validation data, it was found that image extraction to distinguish viable and unfit corn seeds was feasible to use. For more details, it can be seen in the differences in sensitivity and specificity tests for each feature shown in Figure 4.

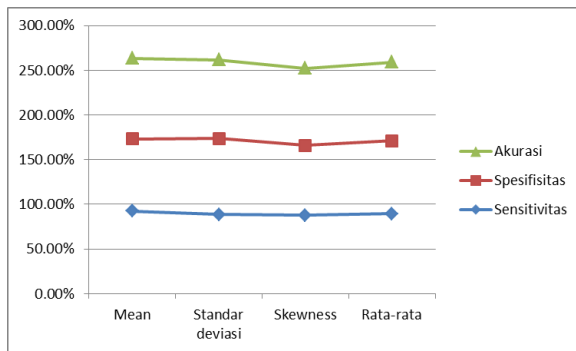


Fig. 4. Average result with color moment

From Figure 4, it can be seen that the mean characteristic given sensitivity is 92.42%, specificity is 80.83%, accuracy is 90%; the standard deviation characteristic given sensitivity is 88.67%, specificity is 85%, accuracy is 88% and the skewness characteristic given sensitivity is 87.76%, specificity of 78.33%, accuracy of 86%. Of the 3 characteristics, the average value for sensitivity is 89.61%, the average value for specificity is 81.38% and the average value for accuracy is 88% so it can be said that the classification technique with k-NN has given good results in classifying images with textures using the color moment feature.

For GLCM feature extraction, given different result if the model classification done with k-NN algorithm. The average result, shown in figure 5.

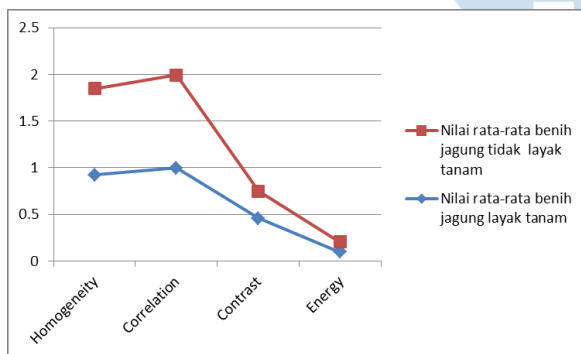


Fig. 5. Average result with GLCM

From Figure 5 it can be seen that the highest average texture feature is correlation because it has an average value of decent corn seeds is 0.9967 and unworthy corn seeds is 0.9969, this means that the size of the linear dependence of the gray level of the image gives an indication of the existence of linear structures in the very high image. While the feature with a low average value is energy because it has an average value of decent corn seeds is 0.0982 and ununworthy corn seeds is 0.1056 this means the uniformity on the texture of the image is less.

The k-NN classification is used with the distance used is the Euclidean distance and uses feature extraction of order II or GLCM. The validation results obtained from the k-NN classification for each feature

are 63.82% for the energy feature, 75.5% for the homogeneity feature, 65.75% for the contrast feature, and 78.67% for the correlation feature. Validation is repeated once for each feature and the average is taken. From the validation data, it was found that image extraction to decent and unworthy corn seeds was feasible to use. For more details, it can be seen in the differences in sensitivity and specificity tests for each feature shown in Tables 5 to 8.

TABLE V. SENSITIVITY AND SPECIFICITY OF HOMOGENEITY FEATURE

Direction	Sensitivity	Specificity	Accuracy
0°	100%	100%	100%
45°	32.10%	70.83%	51%
90°	96.25%	19.17%	81%
135°	96.25%	19.17%	81%
Average	81.15%	52.29%	75.5%

TABLE VI. SENSITIVITY AND SPECIFICITY OF CORRELATION FEATURE

Direction	Sensitivity	Specificity	Accuracy
0°	77.5%	40%	48%
45°	100%	100%	100%
90°	100%	17.5%	83.67%
135°	100%	17.5%	83%
Average	85%	53.12%	78.67%

TABLE VII. SENSITIVITY AND SPECIFICITY OF CONTRAST FEATURE

Direction	Sensitivity	Specificity	Accuracy
0°	41.10%	80.41 %	49%
45°	27.37%	95.83%	42%
90°	100%	100%	100%
135°	98.80%	44.58 %	86%
Average	61.50%	81.04%	65.75%

TABLE VIII. SENSITIVITY AND SPECIFICITY OF ENERGY FEATURE

Direction	Sensitivity	Specificity	Accuracy
0°	41.10%	80.41 %	49%
45°	95.23%	40%	84%
90°	45.59%	76.35%	52.29%
135°	63.69%	95%	70%
Average	61.40%	72.91%	63.82%

From Table 5 to Table 8, it can be seen that the homogeneity feature for sensitivity is 81.15%, for specificity is 52.29%, and accuracy is 75.5%; for correlation feature, for sensitivity is 85%, for specificity is 53.12%, and accuracy is 78.67%; contrast feature that is for sensitivity is 61.50%, for specificity is 81.04%, and accuracy is 65.75%; and energy feature, for sensitivity is 61.40%, for specificity is 72.19%, and accuracy is 63.82%. So from the 4 features in GLCM, it can be said that the classification technique with k-NN has given good results in classifying images with textures using haralick features. The following is a table of averages of sensitivity and specificity tests for GLCM corn seed images, which can be seen in Table 9.

TABLE IX. AVERAGE OF SENSITIVITY AND SPECIFICITY USING GLCM FEATURES

Haralick features	Sensitivity	Specificity	Accuracy
Homogeneity	81.15%	52.29%	75.5%
Correlation	85%	53.12%	78.67%
Contrast	61.50%	81.04%	65.75%
Energy	61.40%	72.91%	63.82%
Average	72.27%	64.84%	70.93%

From table 9 it can be seen that the average sensitivity, specificity and accuracy for GLCM testing is 72.27% for sensitivity, 64.84% for specificity and 70.93% for accuracy.

From the two classification test results in Table 4 and Table 9, the extraction is used first-order statistical features (color moment) and extraction using second order statistical features (GLCM) using the k-NN model classification with euclidean distance for the selection of corn seeds given a good results. Because with color moment feature extraction, given 88% of accuracy and 70.93% accuracy if using GLCM features extraction.

#### IV. CONCLUSION

In this research for the selection of local corn seeds in Nekbaun Village, Amarasi Barat District, East Nusa Tenggara with Color Moment feature extraction, it showed that the mean, standard deviation and skewness characteristics had an average validation of 88% and GLCM feature extraction showed that the homogeneity characteristics, correlation, contrast and energy have an average validation of 70.93%. From these results, it can be concluded that feature

extraction with Color Moment is better than GLCM feature extraction in classifying images into decent seed class and unworthy seed class.

#### REFERENCES

- [1] Supriatna, A. 2009. Identification of Physical Quality of Corn using Digital Image Processing Technology and Artificial Neural Networks. 31st volume. Bogor (ID): Bogor Agricultural Research and Development Agency.
- [2] Maria, 2010. Classification of Coffee Bean Quality using Image Processing and Fuzzy Logic. Thesis. Surabaya (ID): Ten November Institute of Technology Surabaya.
- [3] Kaesmetan Y R, " Comparison of Image Texture Extraction for Selection of Soybean Seeds with Order I and Order II Statistical Methods", Journal of Information Technology Hoaq 2019, pp.996-1006, December,2019.
- [4] Bilaut, CA. 2015. Determining the quality of corn seeds for seeding based on the brightness level of the color using a support vector machine. Thesis. Kupang (ID): Stikom Uyelindo Kupang.
- [5] Agrolology, 2012. Physiological Quality of Corn to Seedlings, Journal Volume 11th. Bogor (ID): IPB Bogor.
- [6] Anggraini, 2008. Estimating the Physical Quality of Corn with Artificial Neural Networks using Color Parameters in Digital Image Processing. Journal Volume 1. Padang (ID): UNAND Padang.
- [7] Gonzalez, CR. Woods, R. 2007. Digital Image Processing. Prentice Hall Publishers. New Jersey (US): USA
- [8] Gonzalez, CR. Woods, R. 2007. Digital Image Processing. Prentice Hall Publishers. New Jersey (US): USA
- [9] Ardi, 2011. Texture Analysis using the GLCM (Gray Level Co-occurrence Matrix) method and feature extraction. Bandung (ID): ITT Bandung
- [10] Zhu, S. Dong, W. & Liu, W. 2015. Hierarchical Reinforcement Learning Based on KNN Classification Algorithms. International Journal of Hybrid Information Technology. Journal Volume 8. Pages 175-184.