Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 4 (2022) 624 - 631 ISSN Media Electronic: 2580-0760

Detecting Diseases on Clove Leaves Using GLCM and Clustering K-Means

Mila Jumarlis[1], Mirfan[2]
[1]Islamic Broadcasting and Communications, STAIN Majene
[2]Informatics Engineering, Universitas Handayani Makassar
[1]milajumarlis.mirfan@yahoo.com, [2]irfan_phapros@yahoo.co.id.

Abstract

The detection of disease in clove plant leaves is generally carried out by diagnosing the symptoms that appear on clove plants. This diagnosis is conducted by clove farmers only by relying on their experience or even having to seek information from other clove farmers. This is because the agricultural sector has no disease detection system for clove leaves by utilizing digital image processing technology to detect diseases in clove leaves. In this study, the researchers applied two methods to make it easier for clove farmers to diagnose diseases in their clove plants. Those methods were the imaging system using Gray Level Co-Occurrence Matrix (GLCM) and disease clustering using the K-Means algorithm. The objective of this study was to design and build image pattern recognition by utilizing 4 features of the Gray Level Co-Occurrence Matrix (GLCM): energy, entropy, homogeneity, and contrast. These 4 features were used to obtain the extraction value from an image. The outcomes were then used to cluster the clove plant diseases using the K-Means method. In making the software, the researchers used Javascript, HTML, CSS, PHP, and MySql to create a database. The output in this study was an information system application that provides disease-type clustering using the K-Means algorithm. The results of the Gray Level Co-occurrence Matrix (GLCM) concerning extracting images of clove plant leaves affected by disease indicated that the created system can be used to help clove farmers in diagnosing what diseases are infecting their plants by only uploading photos from affected leaves of the clove plant. Furthermore, the results of the K-Means calculation on the examined data showed several categories of *Anthracnose* leaf spot diseases. In addition, sample number #40 was included in cluster 2 status, in which the average values for energy, entropy, homogeneity, and contrast were 0.583, 0.175, 0.939, and 0.175, respectively.

Keywords: K-Means, GLCM, Image Processing, Clove Plants, Diagnosis.

1. Introduction

One of Indonesia's economic sources is agriculture and farmers are professions that are mostly occupied by Indonesians, in which Indonesia is an agricultural country. Agricultural productivity is based on how farmers increase yields in terms of quality and maximum quantity by using fertilizers, minimal drugs, and existing natural resources. One of the plants that are quite widely developed in Indonesia is the clove plant.

Clove is a plantation plant very suitable to be cultivated in Indonesia. Cloves can grow in loose soil so that they can bind water well [1].

Clove is one of the tropical plants that can thrive in Indonesia. It can be processed further to make clove oil which is used in several industries, such as pharmaceutics, cosmetics, and others. In addition, cloves are also used as raw materials in the manufacture of cigarettes [2].

Cloves are widely used in the industrial sector as an ingredient for making kretek cigarettes. One of the clove-producing areas in Indonesia is Selayar Islands Regency, South Sulawesi, as shown in Table 1.

Table 1. Data on Clove Production in Selayar Islands Regency

No.	Year	Yield (in ton)
1	2016	178.35
2	2017	102.49
3	2018	535.09

In 2016, Selayar Islands Regency produced cloves with a total production of 178.35 tons. However, there was a decline in 2017, in which the number of cloves produced was only in the range of 102.49 tons. This was due to a large number of diseased clove leaves and the difficulty of farmers distinguishing between diseased and healthy ones. For this reason, a digital image processing system is needed to help farmers detect diseases that attack their clove plants. One of the advantages of digital image

Accepted: 20-04-2022 | Received in revised: 12-07-2022 | Published: 30-08-2022

processing is that it can extract certain features from images for analysis so that diseases can be distinguished from one another only by the image [6]. In 2018, clove production increased again with total production reaching 535.09 tons [3].

A study by Sofwan, Hamid, and Kadir reveals that disease in clove plants is one of the factors that can affect the growth and productivity of cloves [4]. There are several symptoms in clove plants when being attacked by diseases, such as sudden leaf fall, yellow leaves, spots in the leaves, and others [5]. By recognizing the symptoms of the disease that occurs in clove plants, it is expected that prevention or treatment measures can be carried out quickly and precisely based on the symptoms that appear.

Currently, many technologies have been used to identify diseases in various plants. Among the technologies, image processing is one of the technologies that is developing rapidly in identifying or diagnosing a disease in plants [6]. Image processing is needed because it can extract certain features from an image to be analyzed further so that diseases can be distinguished from one another only from an image.

Two of the image processing methods that are often used to identify leaf images are GLCM and K-means. GLCM is a method for extracting features contained in the images statistically, while K-means is a method for classifying. These features produce a certain value and form an angle pattern that is often used in retrieving features in the image processing stage [7]. A study conducted by Sivi Anisa and Herdian Andika focuses on feature extraction, namely taking features on the leaves of betel, cassava, salam, rambutan, and papaya with perfect leaf conditions (without defects) [7]. Their study concentrates on how to distinguish the images of the 5 leaves to describe the distinctive characteristics of those objects. What distinguishes their study from ours is the development of the classification of clove leaves contaminated with disease.

There have been many previous studies attempting to identify the types of diseases in plants by digital image processing [6],[8],[9],[10].

A study conducted by Efrilla *et al*. by utilizing Artificial Neural Networks (ANN) to detect disease in strawberry leaves shows that the best ANN architecture is to use two hidden layers with input parameters in the form of 12 image parameters, namely mean R, mean G, mean B, contrast, correlation, energy, entropy, homogeneity, area, perimeter, eccentricity, and metrics [6].

Another study conducted by Ariesdianto, Fitri, Madjid, and Imron to identify diseases on Siamese citrus leaves using K-NN reveals that the K-Nearest Neighbor (K-NN) classification method can classify and identify diseases in Siamese citrus leaves with an accuracy of 70% with variations in the K value of 21 [8]. The study examines 150 training data and 30 test data consisting of three classes, namely the healthy Siamese citrus leaves, the Siamese citrus leaves with cancer, and the Siamese citrus leaves attacked by the peliang caterpillars.

Similarly, a study conducted by Haris shows that the results of testing the implementation of a disease identification system in rice plants by combining shape and texture characteristics indicate that the accuracy of 70 training data is 100%, while the accuracy of 30 test data is 93% [9]. Meanwhile, the results of the test using fuzzy clustering generate a percentage of 83%, while the results of the test using distortion generate a percentage of 67%.

In addition, a study conducted by Nurhasanah on the classification of rice leaf diseases based on the results of the 4-angle interval GLCM feature extraction shows that the system's accuracy for identifying diseases in rice leaf images reaches 80% [10].

However, those previous studies focus only on identifying diseases in strawberry, Siamese citrus, and rice plants using various methods (i.e., ANN, K-NN, fuzzy clustering, and GLCM). In contrast with those studies, this study aimed at detecting clove plant disease based on leaf images using GLCM feature extraction. After that, the researchers classified the obtained results with the K-means algorithm. In performing GLCM feature extraction, the researchers used accuracy testing by examining 4 features: energy, entropy, contrast, and homogeneity. Furthermore, to classify clove plant diseases based on clove cultivation guidelines released by the plant research center, the researchers divided the obtained results in the previous stage into 5 classes of diseases that infect clove leaves, namely *Anthracnose*, *Coniella*, *Pestalotia*, *Sooty Mold*, and *Cephaleuros*.

In principle, to find out the presence of those five diseases based on the value of the GLCM feature, the researchers examine the representation of values for variations in color levels by the cooccurrence matrix, which is called contrast. After that, the researchers measure the level of homogeneity or similarity of variations in the color intensity of the image, which is called homogeneity. Then, the researchers look at the level of texture uniformity, which is called energy. Finally, the researchers measure the degree of randomness of the color levels of an image in the matrix, which is called entropy [10].

2. Research Methods

This section consisted of the running system (that is, when clove farmers find symptoms that appear on their clove plants, they will seek information or consult with experts about these symptoms), the proposed system,

DOI: https://doi.org/10.29207/resti.v6i4.4033

image captures, RGB to grayscale conversion, GLCM implementation, and K-Means implementation.

2.1. The Running System

In the running system, an analysis was carried out with the illustration shown in Figure 1 below.

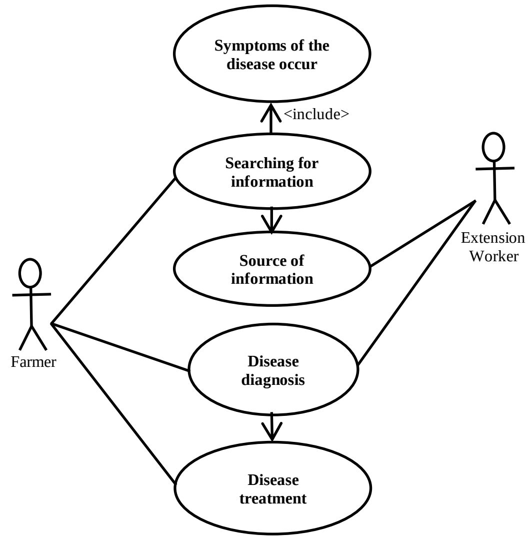


Figure 1. Use Case Diagram

In the use case diagram, there are only two actors who play a role manually without any use of the computer system. Here, farmers need the information to diagnose diseases and treat diseases in clove plants, while extension workers provide information to farmers about the types of diseases that attack clove plants.

2.2. The Proposed System

After analyzing the running system, the proposed new system design was carried out. The main objective of this new system was to improve the old or existing system to make it easier for users. The system design process described the flow of the system process, making it easier and more efficient for work performance and activities.

The proposed system was the relationship between components, variables, and system parameters. In this case, the admin can carry out the disease classification process by uploading an image of the impact that occurs on the leaves. After that, the image is processed and extracted with the image features using the GLCM method, which will then be classified using the K-Means method. Only after that, the admin can see the results of the disease classification from the previously uploaded image. Similarly, users (clove farmers) can also carry out the disease classification process by uploading pictures of the impact that occurs on the leaves. The image is then processed and extracted using the GLCM method, which will be classified further using the K-Means method and can see the results of the classification of diseases based on previously uploaded images.

It can be seen more clearly in Figure 2 below.

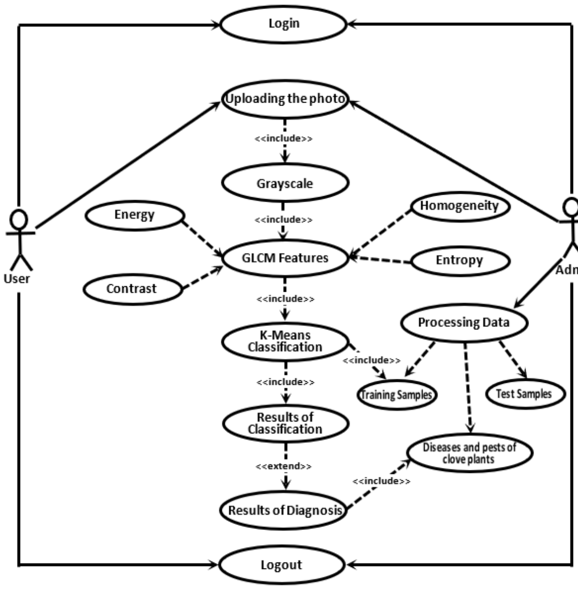


Figure 2. Use Case Diagram of the Proposed System

The use case diagram above has two actors: admin and user. Here, the admin can log in to access and enter the system that has been created, manage disease data, organize training and test samples, carry out the disease classification process by uploading images of clove leaf samples, and find out the results of disease classification from uploaded sample images.

Shortly, users can input clove leaf images to find out the diseases contained in the leaves and find out the results of the classification of those clove leaves.

2.3. Image Capture

At this stage, the images of the clove plant leaves were taken using a camera. The data used in this study were 115 clove leaf images, which were divided into 95 training data and 20 test data. These images consisted of several types of clove leaf diseases, such as *Glorosporium piperantum*, *Cylibdrocladium quinaseptatum*, and root rot disease. Image capture must be as detailed as possible with good lighting so that the image can be maximized at the processing stage. The total number of clove leaf diseases examined in this study was 5, namely *Anthracnose*, *Coniella*, *Pestalotia*, *Sooty Mold*, and *Cephaleuros*.

The type of camera used for this stage was a digital camera. The images were pictures of clove leaves that had a symptom of a disease. The images had a plain white background with an image size of 150×150 pixels. The image format was PNG. Furthermore, the captured

DOI: https://doi.org/10.29207/resti.v6i4.4033

images were uploaded into the system that had been built.

2.4. Converting Images to Grayscale

A grayscale image is a digital image that has only one channel value for each pixel with an 8-bit format for each pixel. Grayscale images have a value range of 0-255.

$$Gray_{image} = Wr * R + Wg * G + Wb * B \quad (1)$$

In equation 1, R, G, and B are the monochrome color values (red, green, and blue) in a linear RGB color image; Wr is the coefficient (fixed weight) of the red color, which is 0.299; Wg is the coefficient (fixed weight) of the green color, which is 0.587; Wb is the coefficient (fixed weight) of the blue color, which is 0.114 [11].

2.5. GLCM Implementation

This process was carried out to determine the matrix value of the image pixels by calculating 4 GLCM features in the image. Those features were energy, entropy, homogeneity, and contrast.

The formula for those 4 features used in extraction can be seen in the following equation.

Energy

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

In equation 2 (energy), *j* is the column of the matrix; *i* is the row of the matrix; *p*(*i*,*j*) is the element of the co-occurrence matrix in row *i* and column *j*. Energy is the intensity of the spread of pixels of an image in grayscale. If the values of energy of all pixels are relatively similar, it indicates a high category [12].

Entropy

$$= \sum_i \sum_j p(i,j).^2 \log\ p(i,j) \quad (3)$$

In equation 3 (entropy), *j* is the column of the matrix; *i* is the row of the matrix; *p*(*i*,*j*) is the element of the co-occurrence matrix in row *i* and column *j*. Entropy is the degree of uncertainty or randomness in a digital image [13].

Contrast

$$= \sum_i \sum_j (i,j)^2 \ \ p(i,j) \quad (4)$$

In equation 4 (contrast), *j* is the column of the matrix; *i* is the row of the matrix; *p*(*i*,*j*) is the element of the co-occurrence matrix in row *i* and column *j*. Contrast is a measure of the presence of the value of the gray level around the image area [14].

Homogeneity

$$= \sum_i \sum_j \frac{p(i,j)}{1+|i-j|} \quad (5)$$

In equation 5 (homogeneity), *j* is the column of the matrix; *i* is the row of the matrix; *p*(*i*,*j*) is the element of the co-occurrence matrix in row *i* and column *j*. Homogeneity represents the similarity value of the sample variation whether it comes from the same population. The homogeneity value has a maximum value if all pixels have the same value [15].

GLCM is one of the feature extraction methods to obtain feature values by calculating the probability value from the results of the relationship that exists between two pixels at a certain distance and angle orientation.

Each of these features has a co-accuracy matrix with degrees (image angles) of 0, 45, 90, and 135 [16]. Figure 3 shows the image angles of GLCM.

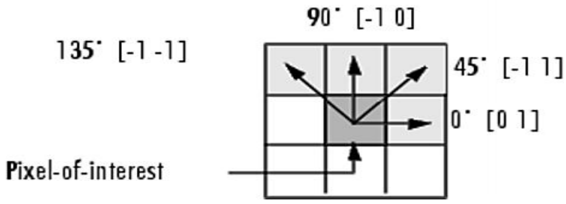


Figure 3. The GLCM Image Angles

2.6. K-Means Implementation

The clustering technique used in this study was the K-Means algorithm. Clustering with K-Means was carried out to find similarities in the characteristics of data and then classify those data.

K-Means is a data analysis method by performing the modeling process without direction (unsupervised). In addition, K-Means also uses a partition system in data grouping. The data are classified into two or more groups. The data are divided by the mean value at the center of the cluster [17].

The steps in the implementation of K-Means are as follows [18]. At the beginning of the selection iteration, the cluster center is chosen freely by several members of the dataset, which is then adjusted to the number of clusters (k) that have been determined. The next step is to calculate the distance of each cluster center from each data. The value of k must be positive. Furthermore, in calculating the distance to each cluster center, the researchers employed the Euclidean formula, as follows.

$$d_{ik} = \sqrt{\sum_{j=1}^{m} (x_{ij} - c_{kj})^2} \quad (6)$$

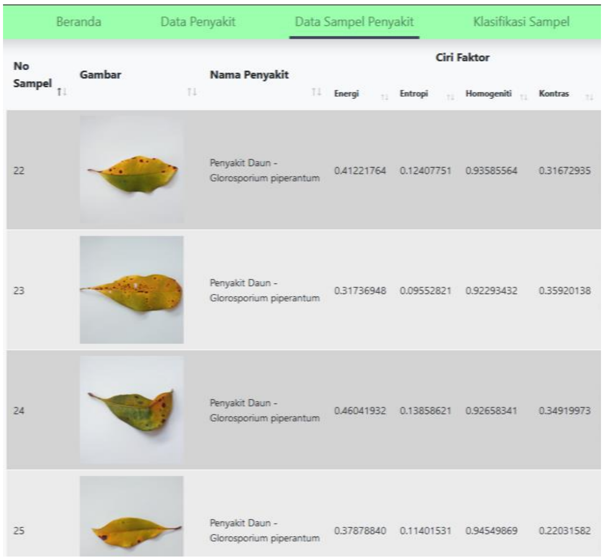
Where: d_{ik} is the distance between the *i*-th data and the *k*-th cluster center; *m* is the number of attributes; x_{ij} is the *i*-th data; C_{kj} is the *k*-th cluster center. Data may enter into a cluster if the data have the smallest distance between the existing clusters. Then, the data that are members of each cluster will be grouped. After that, it is to update the cluster center value, which is calculated using the mean value. This value is adjusted to the number of members in each cluster according to the following formula.

$$c_{kj} = \sqrt{\frac{\sum_{h=1}^{p} y_{hj}}{p}};\ y_{hj} = x_{ij} \in \text{cluster ke} - k \quad (7)$$

Where: C_{kj} is the center of the cluster; p is the total of all cluster members; h is the initial number of cluster members; y_{hj} is the number of data; x_{ij} is the i-th data.

2. Results and Discussion

The data in this study were 115 clove leaf images which were divided into 2 groups, namely 95 in training data and 20 in test data. The image data consisted of several types of clove leaf diseases, such as *Glorosporium piperantum*, *Cylibdrocladium quinaseptatum*, and root rot diseases. Figure 4 shows samples of training data.



No Sampel	Gambar	Nama Penyakit	Ciri Faktor: Energi	Entropi	Homogeniti	Kontras
22		Penyakit Daun - Glorosporium piperantum	0.41221764	0.12407751	0.93585564	0.31672935
23		Penyakit Daun - Glorosporium piperantum	0.31736948	0.09552821	0.92293432	0.35920138
24		Penyakit Daun - Glorosporium piperantum	0.46041932	0.13858621	0.92658341	0.34919973
25		Penyakit Daun - Glorosporium piperantum	0.37878840	0.11401531	0.94549869	0.22031582

Figure 4. Samples of Training Data

Determination of the centroid in this study was carried out by taking the mean value of each training sample data with the same disease, as seen in Table 2 below.

Table 2. Training Data and Test Data

No. Sample	Disease	Features: Energy	Entropy	Homogeneity	Contrast
22	Glorosporiu piperantum	0.583	0.175	0.939	0.175
23	Glorosporiu piperantum	0.736	0.221	0.962	0.272
24	Glorosporiu mpiperantu	0.637	0.191	0.948	0.333
25	Glorosporiu mpiperantu	0.603	0.181	0.913	0.840
26	Glorosporiu mpiperantu	0.149	0.044	0.920	0.412
27	Glorosporiu mpiperantu	0.414	0.124	0.926	0.479
28	Glorosporiu mpiperantu	0.414	0.124	0.913	0.481
29	Glorosporiu mpiperantu	0.392	0.118	0.924	0.406
30	Glorosporiu mpiperantu	0.463	0.139	0.921	0.751
31	Cylindrocla diquinqesep tatu	0.442	0.133	0.931	0.663
32	Cylindrocla diquinqesep tatu	0.252	0.075	0.916	0.589
33	Cylindrocla diquinqesep tatu	0.430	0.129	0.915	0.740
34	Cylindrocla diquinqesep tatu	0.579	0.174	0.914	0.862
35	Cylindrocla diquinqesep tatu	0.3920	0.1179	0.9273	0.30511
36	Penyakit Busuk Akar	0.3911	0.1177	0.9208	0.35168
37	Penyakit Busuk Akar	0.3036	0.0914	0.9389	0.26931
38	Penyakit Busuk Akar	0.2852	0.0858	0.9379	0.22961

Based on the stages described in the method section, the initial steps taken were inputting and converting the original image to grayscale, carrying out GLCM feature extraction, and conducting classification with K-Means, which were described in the following.

3.1. Inputting and Converting Original Image to Grayscale

The original images of the clove leaves were input to the system with a size of 150×150 pixels and a format of PNG. The inputted images were a single leaf (not a collection of several leaves) that had diseases. Images were taken as close as possible and in good lighting to get the best image quality. After that, the images were converted into grayscale. An example of original images can be seen in Figure 5.

Figure 5. An Original Image of Clove Leaf

The result of the conversion of the original image to grayscale is shown in Figure 6.

Figure 6. The Result of the Conversion to Grayscale

DOI: https://doi.org/10.29207/resti.v6i4.4033

3.2. GLCM Extraction

The initial step in the Gray Level Co-Occurrence Matrix (GLCM) extraction is to create a co-occurrence matrix, followed by determining the spatial relationship between the reference pixels and neighboring pixels. This spatial relationship is based on angle θ and distance d.

The next step is to create a symmetrical matrix by adding a co-occurrence matrix with its transpose matrix. After that, it is to normalize the symmetric matrix. This normalization is carried out by calculating each probability of each matrix element.

The last step is to calculate the four features in GLCM: energy, entropy, homogeneity, and contrast. Each feature is calculated by one-pixel distance in four types of degrees: 0°, 45°, 90°, and 135°.

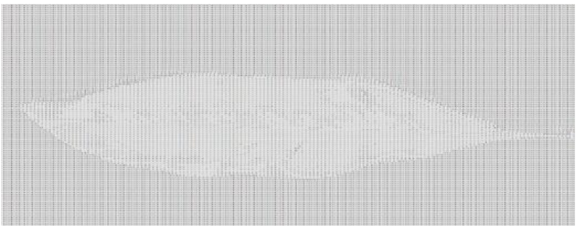


Figure 7. Image with Pixel Matrix

After the matrix calculation is complete from the total pixels in Figure 7 using 4 GLCM features (energy, entropy, homogeneity, and contrast) with four types of degree as shown in Figure 3, the results of the GLCM values of the image are shown in Table 3.

Table 3. Results of GLCM values

Features	Value
Energy	0.583
Entropy	0.175
Contrast	0.532
Homogeneity	0.939

Table 4 shows the results of the extraction calculation with 4 GLCM features.

Table 4. Results of GLCM Extraction Calculation

No. Sample	Disease	Features of GLCM			
		Energy	Entropy	Homogeneity	Contrast
40	Antraknose	0.583	0.175	0.939	0.175
46	Coniella	0.736	0.221	0.962	0.272
47	Coniella	0.637	0.191	0.948	0.333
48	Pestalotia	0.603	0.181	0.913	0.840
49	Sooty Mold	0.149	0.044	0.920	0.412
62	Cephaleuros	0.414	0.124	0.926	0.479
63	Cephaleuros	0.414	0.124	0.913	0.481
65	Cephaleuros	0.392	0.118	0.924	0.406
75	Pestalotia	0.463	0.139	0.921	0.751
76	Pestalotia	0.442	0.133	0.931	0.663
77	Cephaleuros	0.252	0.075	0.916	0.589
80	Pestalotia	0.430	0.129	0.915	0.740
81	Cephaleuros	0.579	0.174	0.914	0.862

3.3. Clustering with K-Means

Clustering with K-Means is carried out by choosing the center of the cluster freely by several members of the dataset at the beginning of the iteration and then is adjusted to the number of clusters (k) that have been set. It is followed by calculating the distance of each cluster center from each data. The value of k must be positive. Furthermore, in calculating the distance to each cluster center, the Euclidean formula is used. The results of this first calculation can be seen in Table 5.

Table 5. Distance Values of the First Iteration

No.	No. Sample	Distance Value			Cluster Status
		Cluster 1	Cluster 2	Cluster 3	
1	22	0.026	0.098	0.077	1
2	23	0.132	0.081	0.075	3
3	24	0.039	0.103	0.136	1
4	25	0.116	0.189	0.079	3
5	26	0.195	0.289	0.285	1
6	27	0.065	0.052	0.124	2
7	28	0.066	0.058	0.073	2
8	29	0.078	0.167	0.178	1
9	30	0.151	0.047	0.139	2
10	31	0.106	0.148	0.206	3
11	32	0.238	0.142	0.192	2
12	33	0.083	0.044	0.138	2
13	34	0.094	0.031	0.094	2
14	35	0.049	0.104	0.053	1
15	36	0.059	0.058	0.081	2
16	37	0.148	0.160	0.046	3
17	38	0.182	0.204	0.084	3
18	40	0.872	0.807	0.895	2

After performing the first iteration, the next process is to update the cluster center value which can be calculated by finding the mean value according to the number of members of each cluster. After that, it is to re-calculate the distance of the data to be classified with the centroid. The results of the calculation of the first center or centroid can be seen in Table 6.

Table 6. First Centroid Values

New Centroid	Energy	Entropy	Homogeneity	Contrast
Klaster 1	0.478	0.143	0.933	0.308
Klaster 2	0.372	0.112	0.850	0.460
Klaster 3	0.360	0.108	0.934	0.292

After the above calculation is complete, it is continued by recalculating the distance of each data with each cluster center using the Euclidean formula again based on the sequence of the previous process. It is carried out until the process of updating the cluster center value is finished by calculating the mean value according to the number of members of each cluster. The results of the calculation of the second iteration are shown in Table 7.

After carrying out the second iteration, we gain the second centroid as shown in Table 8.

Because there is no change in the cluster status in the first and second iterations, the iteration loop stops with the final results as shown in Table 9.

DOI: https://doi.org/10.29207/resti.v6i4.4033

Table 7. Distance Values of the Second Iteration

No.	No. Sample	Distance Value			Cluster status
		Cluster 1	Cluster 2	Cluster 3	
1	22	0.069	0.172	0.059	3
2	23	0.175	0.137	0.081	3
3	24	0.045	0.162	0.118	3
4	25	0.136	0.258	0.074	3
5	26	0.151	0.337	0.267	1
6	27	0.094	0.123	0.110	1
7	28	0.109	0.134	0.062	3
8	29	0.036	0.221	0.160	1
9	30	0.189	0.092	0.138	2
10	31	0.084	0.188	0.188	1
11	32	0.279	0.148	0.200	2
12	33	0.109	0.111	0.125	1
13	34	0.134	0.110	0.087	3
14	35	0.090	0.174	0.035	3
15	36	0.101	0.131	0.068	3
16	37	0.186	0.222	0.064	3
17	38	0.216	0.263	0.100	3
18	40	0.882	0.715	0.894	2

Table 6. Second Centroid Values

New Centroid	Energy	Entropy	Homogeneity	Contrast
Klaster 1	0.497	0.149	0.932	0.347
Klaster 2	0.333	0.100	0.723	0.606
Klaster 3	0.369	0.111	0.930	0.313

Table 9. Distance Values of the Last Iteration

No.	Distance Value	Distance Value			Cluster status
		Cluster 1	Cluster 2	Cluster 3	
1	22	0.095	0.369	0.045	3
2	23	0.189	0.318	0.071	3
3	24	0.040	0.354	0.102	1
4	25	0.178	0.448	0.094	3
5	26	0.153	0.508	0.263	1
6	27	0.086	0.319	0.089	1
7	28	0.122	0.331	0.040	3
8	29	0.035	0.404	0.150	1
9	30	0.186	0.272	0.120	3
10	31	0.041	0.361	0.171	1
11	32	0.281	0.267	0.189	2
12	33	0.096	0.305	0.103	1
13	34	0.139	0.306	0.066	3
14	35	0.118	0.369	0.025	3
15	36	0.112	0.328	0.046	3
16	37	0.217	0.401	0.082	3
17	38	0.251	0.436	0.121	3
18	40	0.857	0.530	0.876	2

As seen in the results of the K-Means calculation, the test data with sample number #40 have cluster status 2, namely the cluster for the category of *anthracnose* leaf spot disease, which can be seen in Figure 8.

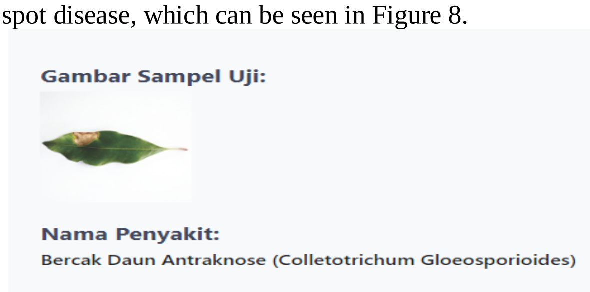


Figure 8. Clustering Results of K-Means

The results of the accuracy test with energy, entropy, homogeneity, and contrast show that testing for all *Coniella* diseases obtains the values of 0.637 for energy, 0.191 for entropy, 0.948 for homogeneity, and for 0.663 contrast. Meanwhile, testing for all *Anthracnose* diseases gains the values of 0.583 for energy, 0.175 for entropy, 0.939 for homogeneity, and 0.175 for contrast. To identify clove leaf diseases, it is necessary to look at the value of variations in color levels in the co-occurrence matrix (contrast), the similarity of variations in the color intensity of the image (homogeneity), the level of texture uniformity (energy), and the degree of randomness of the color of an image on the matrix (entropy). Concerning the effectiveness of the algorithm used with the GLCM method, it can detect clove plant diseases based on leaf images which are then grouped with leaf extraction features with the K-means algorithm.

Testing the level of accuracy of sample images to obtain diagnostic results for the type of disease or pest is carried out by examining the three levels of the image and diagnosing the sample images as shown in figures in Table 10.

Table 10. The Results of Testing Sample Images

No.	Figures	Type of Image Capture	Success Rate
1		Focused on the object	90%
2		Medium Distance	75%
3		Long Distance	30%

Based on the test results of the samples, it is concluded that the recommended image capture for obtaining a more accurate diagnosis is by taking images focusing directly on the sample objects.

3. Conclusion

This system can provide clustering results of disease types using the K-Means algorithm based on the results of GLCM image processing from images of affected clove plant leaves. Furthermore, this system can be used to help clove farmers in diagnosing what diseases are infecting their plants by only uploading photos from the leaves of the affected clove plant. Concerning the results of the sample image test, the highest percentage of success accuracy is 90% by capturing images focusing

DOI: https://doi.org/10.29207/resti.v6i4.4033

on the object, followed by 75% due to capturing images at a medium distance and 30% because of the long distance. To test for clove leaf disease, it is necessary to look at the values of contrast, homogeneity, energy, and entropy. Further studies are expected to compare other feature extraction algorithms, such as *Haar Wavelet* and *Local Binary* which focus on disease-contaminated leaves.

References

[1] D. A. Bimantoro dan S. Uyun, "Pengaruh Penggunaan Information Gain untuk Seleksi Fitur Citra Tanah dalam Rangka Menilai Kesesuaian Lahan pada Tanaman Cengkeh," *JISKA J. Inform. Sunan Kalijaga*, vol. 2, no. 1, hlm. 42–52, Agu 2017, doi: 10.14421/jiska.2017.21-06.

[2] Y. Rasud, "In Vitro Callus Induction from Clove (Syzigium aromaticum L.) Leaves on Medium Containing Various Auxin Concentrations," *J. Ilmu Pertan. Indones.*, vol. 25, no. 1, hlm. 67–72, Jan 2020, doi: 10.18343/jipi.25.1.67.

[3] Badan Pusat Statistik Kabupaten Kepulauan Selayar, "Statistik Pertanian Non Tanaman Pangan Kabupaten Kepulauan Selayar Tahun 2018," *Badan Pusat Statistik Kabupaten Kepulauan Selayar*, 2019. https://selayarkab.bps.go.id/publication/2019/10/25/e1af1d6d0563d921cc41bfdd/statistik-pertanian-non-tanaman-pangan-kabupaten-kepulauan-selayar-tahun-2018.html

[4] M. Sofwan, A. H. Hamid, dan I. A. Kadir, "Motivasi Petani, Faktor Pendorong dan Faktor Penghambat dalam Budidaya Tanaman Cengkeh Di Mukim Lampuuk Kecamatan Lhoknga Kabupaten Aceh Besar," *J. Ilm. Mhs. Pertan.*, vol. 3, no. 4, hlm. 355–367, Nov 2018, doi: 10.17969/jimfp.v3i4.8765.

[5] M. K. Amrulloh, H. S. Addy, dan W. S. Wahyuni, "Karakterisasi fisiologis dan biokimia penyebab penyakit bakteri pembuluh kayu pada tanaman cengkeh (Syzygium aromaticum L.) di PT Tirta Harapan," vol. 2, no. 1, hlm. 1–7, Feb 2021, doi: 10.19184/jptt.v2i1.17919.

[6] A. V. Efrilla *dkk*., "Klasifikasi Penyakit Pada Daun Stroberi Menggunakan K-Means Clustering dan Jaringan Syaraf Tiruan," *J. Keteknikan Pertan. Trop. Dan Biosist.*, vol. 8, no. 2, hlm. 161–170, Agu 2020, doi: 10.21776/ub.jkptb.2020.008.02.06.

[7] N. Sivi Anisa dan T. Herdian Andika, "Sistem Identifikasi Citra Daun Berbasis Segmentasi Dengan Menggunakan Metode K-Means Clustering," *Aisyah J. Inform. Electr. Eng. AJIEE*, vol. 2, no. 1, hlm. 9–17, Feb 2020, doi: 10.30604/jti.v2i1.22.

[8] R. H. Ariesdianto, Z. E. Fitri, A. Madjid, dan A. M. N. Imron, "Identifikasi Penyakit Daun Jeruk Siam Menggunakan K-Nearest Neighbor," *J. Ilmu Komput. Dan Inform.*, vol. 1, no. 2, hlm. 133–140, Nov 2021, doi: 10.54082/jiki.14.

[9] N. A. Haris, "Kombinasi Ciri Bentuk dan Ciri Tekstur Untuk Identifikasi Penyakit Pada Tanaman Padi," *JATISI J. Tek. Inform. Dan Sist. Inf.*, vol. 7, no. 2, hlm. 237–250, Agu 2020, doi: 10.35957/jatisi.v7i2.239.

[10] Y. I. Nurhasanah, "Sistem Pengenalan Jenis Kanker Melanoma pada Citra Menggunakan Gray Level Cooccurrence Matrices (GLCM) dan K-Nearest Neighbor (KNN) Classifier," *Mindjournal*, vol. 5, no. 1, hlm. 66–80, 2020, doi: 10.26760/mindjournal.v5i1.66-80.

[11] S. A. Hantoush Alrubaie dan A. H. Hameed, "Dynamic Weights Equations for Converting Grayscale Image to RGB Image," *J. Univ. BABYLON Pure Appl. Sci.*, vol. 26, no. 8, hlm. 122–129, Okt 2018, doi: 10.29196/jubpas.v26i8.1677.

[12] A. Septiarini, Rizqi Saputra, Andi Tejawati, dan Masna Wati, "Deteksi Sarung Samarinda Menggunakan Metode Naive Bayes Berbasis Pengolahan Citra," *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 5, no. 5, hlm. 927–935, Okt 2021, doi: 10.29207/resti.v5i5.3435.

[13] K. Mahesa, B. Sugiantoro, dan Y. Prayudi, "Pemanfaatan Metode DNA Kriptografi Dalam Meningkatkan Keamanan Citra Digital," vol. 07, no. 02, hlm. 108–113, Sep 2019, doi: 10.33884/jif.v7i02.1356.

[14] P. N. Andono dan E. H. Rachmawanto, "Evaluasi Ekstraksi Fitur GLCM dan LBP Menggunakan Multikernel SVM untuk Klasifikasi Batik," *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 5, no. 1, hlm. 1–9, Feb 2021, doi: 10.29207/resti.v5i1.2615.

[15] W. Wiliawanto, M. Bernard, P. Akbar, dan A. I. Sugandi, "Penerapan Strategi Pembelajaran Aktif Question Student Have Untuk Meningkatkan Kemampuan Berpikir Kritis Matematik Siswa SMK," *J. Cendekia J. Pendidik. Mat.*, vol. 3, no. 1, hlm. 139–148, Mei 2019, doi: 10.31004/cendekia.v3i1.86.

[16] Z. Y. Lamasigi, "DCT Untuk Ekstraksi Fitur Berbasis GLCM Pada Identifikasi Batik Menggunakan K-NN," *Jambura J. Electr. Electron. Eng.*, vol. 3, no. 1, hlm. 1–6, Jan 2021, doi: 10.37905/jjeee.v3i1.7113.

[17] L. Maulida, "Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. DKI Jakarta Dengan K-Means," *JISKA J. Inform. Sunan Kalijaga*, vol. 2, no. 3, hlm. 167, Mar 2018, doi: 10.14421/jiska.2018.23-06.

[18] S. Rustam, H. A. Santoso, dan C. Supriyanto, "Optimasi K-Means Clustering Untuk Identifikasi Daerah Endemik Penyakit Menular Dengan Algoritma Particle Swarm Optimization Di Kota Semarang," *Ilk. J. Ilm.*, vol. 10, no. 3, hlm. 251–259, Des 2018, doi: 10.33096/ilkom.v10i3.342.251-259.

DOI: https://doi.org/10.29207/resti.v6i4.4033