Identification of Diseases in Rice Plants Using the Gray Level Co-Occurrence Matrix Method

1Muhammad Nasir*, 2Mursyidah, 3Muchsin Fajri
Department of Information Technology and Computers, Politeknik Negeri Lhokseumawe, Aceh, Indonesia

*Correspondent Author: muhnasir.tmj@pnl.ac.id

Abstract. Rice plants (Oryza sativa, sp) are a staple food raw material that is vital for the people of Indonesia. Rice plants are included as a year / annual plant group. The management of rice plants is an innovative and dynamic approach in increasing the production and opinion of farmers through assembling technological components in a participatory manner with farmers. For this reason, appropriate technology facilities are needed that can be used by farmers in determining the type of leaf disease that burns during the growing period of rice so that it can increase production. This study aims to apply the Gray level Co-Occurrence Matrix method to produce feature values on rice leaves. Differences in the patterns of rice leaves are used by identifying each rice leaf pattern, which results in the type of leaf disease burning through the rice leaf pattern. In this study, the accuracy of this system was obtained at 72.5%.

Keywords - Identification of Disease in Rice Plants, Gray Level Co-Occurrence Matrix.

1. Introduction
In this day and age, rice fields are rarely found especially in urban areas, many rice fields that have become housing. We can only find rice fields in remote rural areas, such as rice, corn, sugar cane, soybeans and so on. Reduced rice field area makes production yields decrease, not to mention the many types of diseases that exist in rice plants, such as leaf disease on fire. For this reason, appropriate technology facilities are needed that can be used by farmers to find out the types of diseases that attack rice plants by knowing the factors or symptoms that arise as early as possible, so as to increase crop productivity. [2] [5]. Make a system to identify diseases of rice leaves from existing symptoms and provide solutions based on the type of disease caused after 4 MST (Sunday After Planting) until harvest. [4] [6]. In this study, the gray-level co-occurrence matrix (GLCM) method was used where in statistical calculations using gray degree distribution (histogram) by measuring the level of contrast, granularity, and roughness of an area from neighboring relationships between pixels in the image. This statistical paradigm is of unlimited use, so it is suitable for natural textures that are not structured from sup patterns and rules (microstructures). This research is expected to be a reference for farmers or people who want to learn to farm to know how to identify the type of rice disease early so that it can be anticipated quickly and precisely.

2. Research Methods
The Gray Level Co-Occurrence Matrix (GLCM) was first proposed by Haralick in 1979 with 28 features to explain spatial patterns. The first step to calculating GLCM features is to convert the RGB image into a grayscale image [5]. The second step is to create a co-occurrence matrix and proceed by determining the spatial relationship between reference pixels and neighboring pixels based on angles $\theta$ and distanced. The next step is to create a symmetric matrix by adding a co-occurrence matrix with the transposing matrix. Then the normalization of the symmetric matrix is done by calculating the
probability of each matrix element. The final step is to calculate the GLCM feature. Each feature is calculated with one-pixel distance in four directions, namely 00, 450, 900 and 1350 to detect co-occurrence [1] [9].

There are 4 GLCM features used in this study.

1. Energy
   Energy states the measure of the concentration of a pair with a certain gray intensity on the matrix.

   \[
   \text{ASM} = \sum_{i,j=0}^{N-1} P_{ij}^2
   \]

2. Correlation
   Express the size of the pixel-dependent relationship to neighboring pixels in the image.

   \[
   \text{correlation} = \sum_{i,j=0}^{N-1} P_{ij} \frac{1-(\mu_i)(\mu_j)}{(\sigma_i^2)(\sigma_j^2)}
   \]

3. Homogeneity
   Mathematically, the homogeneity of GLCM is the inverse of the GLCM contact. That is the uniformity of gray intensity in the image.

   \[
   \text{homogeneity} = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2}
   \]

4. Entropy
   Entropy is used to measure the traceability of the intensity distribution.

   \[
   \text{entropy} = \sum_{i,j=0}^{N-1} P_{ij} (-\ln P_{ij})
   \]

According to Putra, Dharma. (2010), the last step in the identification application is to look for similarities between training images and features of the test image. In the disease identification application program in the rice plant used in this study was Euclidean Distance. Euclidean Distance is the simplest technique for calculating distances between two vectors. Suppose that two vector features P and Q are given, then the distance between two vector features P and Q.

Euclidean Distance is the most commonly used metric to calculate the similarity of two vectors. The level of similarity in the form of a score and based on the score can be said to be similar or not. The Euclidean Distance formula is the root of the difference 2 square vector (root of square difference between 2 vectors).

The formulas of Euclidean Distance are as follows:

\[
D_{ij} = \sqrt{(P_{i1}-Q_{j1})^2 + (P_{i2}-Q_{j2})^2 + (P_{in}-Q_{jn})^2}
\]

In identifying the type of burnt rice leaf disease the method of data collection was carried out by taking the object of the type of rice disease by using a 15 cm camera taking an object. Data on the type of burnt rice disease used 40 images of rice leaf disease on fire, Figure 1 as follows.
In general, the study of the identification of this type of burnt rice leaf disease through several stages of the process, as shown in Figure 2 below.

The image processing process in this study consists of the image input process, RGB process, and matching feature values using value closeness. Data on rice input will be separated into two parts, namely training data and test data. Training data is the data used for training to get the required features to be included in the database, while the test data is data that is used as extracted input, then matched with the database to get the ranking results of similarities between the database and the input data. The test data used as many as 40 types of images of rice leaf disease types burned with different classes, used to extract information contained in the image of the fruit. The first step in the GLCM extraction process is extracting values from GLCM contained in the image. Each training data and test uses feature extraction from GLCM, used to obtain the overall results of the GLCM feature which will then be compared to find out the interrelationship between the closest GLCM values of the test data. The result of the matching process is recognizing the image of the type of burning rice leaf disease that is similar to one another.

Euclidean distance is the distance between two points in an euclidean space. Both designate P and Q in two dimensional Euclidean spaces and P with coordinates (p1, p2), Q with coordinates (q1, q2). The segment line with the end points of P and Q will form the slant of the right elbow triangle. The distance between two points p and q is defined as the square root of the sum of the squares of the sum of the differences between the corresponding coordinates of the points. Then it will be shown in the training data flowchart in Figure 2.3 below.
From Figure 3 above after Snort is run. The system of obtaining input in the form of data on leaf burnt rice disease. Then specify the RGB value. Feature extraction value here is to find the average value of the angle value, then the data will be stored in the database. The test data flowchart can be seen in Figure 4 below.

Figure 3. Flowchart training

Figure 4. Test flowchart
From Figure 4 above after snort is run. The system gets input in the form of rice disease data to be tested. Then determine the value of the extraction process. The extraction value here is to find the value of homogeneity, energy, entropy, correlation, then calculate the distance between the test data to each data label in the training data. Then specify k labeled data that has the closest distance, then identify new data to the majority data label. Then it will produce output in the form of a predetermined type of rice disease.

3. Results And Discussion

An identification form is a form for testing the system in recognizing the type of rice disease that is present in the image that has been entered. In this form, there is a picture capture button that functions to take the image you want to recognize or want to test. After inputting the image, then clicking the glcm extraction process is to find the value of homogeneity, energy, entropy, the correlation in the tested image. Then the next step is to click on the identification button to process the identification of the type of burnt leaf disease by comparing the test image with all the data contained in the system database. The system will check feature values in each image that are the same and not the same. The test results obtained are images that have the closest distance to the image contained in the database, to measure the proximity of the distance to the identification process using the Euclidean distance method, the identification process will produce the name of the type of disease that has the closest distance value to the input image. Sample output from the process of identifying images of burnt rice leaf disease, namely the image of a burnt rice leaf disease in accordance with the image of a burnt rice leaf disease. The results of the identification of burnt leaf disease images can be seen in Figure 5 below.

![Figure 5. Display](image)

**Burned Rice Leaf Image Identification Form**

| Id  | Nilai     |
|-----|-----------|
| DT52| 0.002089  |
| DT40| 0.015237  |
| DT57| 0.020607  |
| DT48| 0.022494  |
| DT12| 0.023614  |
| DT14| 0.025656  |
| DT31| 0.025765  |
|   |   |
|---|---|
| DT15 | 0.026435 |
| DT49 | 0.026931 |
| DT34 | 0.028887 |
| DT55 | 0.030548 |
| DT32 | 0.030839 |
| DT51 | 0.032308 |
| DT33 | 0.032439 |
| DT19 | 0.033552 |
| DT47 | 0.033977 |
| DT13 | 0.034389 |
| DT03 | 0.035238 |
| DT06 | 0.035249 |
| DT50 | 0.035951 |
| DT29 | 0.036152 |
| DT07 | 0.036842 |
| DT35 | 0.037042 |
| DT30 | 0.037286 |
| DT38 | 0.037766 |
| DT18 | 0.038007 |
| DT24 | 0.039125 |
| DT25 | 0.040668 |
| DT05 | 0.041050 |
| DT17 | 0.041093 |
| DT23 | 0.041785 |
| DT59 | 0.043428 |
| DT46 | 0.043932 |
| DT20 | 0.045393 |
| DT16 | 0.046103 |
| DT44 | 0.046906 |
| DT42 | 0.047219 |
| DT54 | 0.047892 |
| DT01 | 0.048028 |
| DT21 | 0.048248 |
| DT02 | 0.048419 |
| DT10 | 0.048777 |
| DT45 | 0.049528 |
| DT53 | 0.049810 |
| DT41 | 0.050758 |
| DT36 | 0.051108 |
| DT11 | 0.051232 |
| DT60 | 0.051269 |
| DT26 | 0.051488 |
The distance test results can be analyzed, namely comparing the extraction value of test image features at an angle of 0º compared to all values of training image feature extraction at an angle of 0º using the Euclidean distance method. From comparing the entire distance between the test image and the distance training that approaches the test image is id DT52 with a distance value of 0.002089 so that the result is that the leaves burn. The application to identify rice leaf disease is burnt using the Euclidean Distance method to recognize the type of rice disease based on the feature extraction of previously trained and stored matrix kookuresi method. The result of measuring the distance of the value taken is the smallest distance value from the results of the test image value with the value in the database. For the type of paddy disease, the total amount of data is 40, the number of test data that has been identified using the Euclidean distance is 29 types of burning disease, the total accuracy of success using the Euclidean distance method reaches 72.5%. Then the total accuracy of success recognized reaches 72.5% using the following equation 3.1. The percentage of the success rate of the system testing of all types of burning leaf disease can be calculated using the following equation:

\[
Accurate = \frac{Correct\ Data}{Overall\ Data} \times 100\%
\]

\[
Accurate = \frac{29}{40} \times 100\% = 72.5\%
\]

### 4. Conclusion

After conducting research and experiment objects regarding the disease identification system in rice plants using the Gray-Level Co-occurrence Matrix (GLCM) and Euclidean Distance methods in the previous chapter, the following conclusions can be drawn:

1. Based on the trials that have been carried out using the distance between the test data and the reference data, this disease identification system in rice plants can identify the types of burning leaf disease.
2. The results of identification of disease in these rice plants have identified results that are not in accordance with the data tested, this is caused by light factors and the retrieval of inappropriate image objects when retrieving inappropriate image objects will produce inappropriate output.
3. The more training carried out, the better the ability of the system in carrying out the introduction of the accuracy of this application system in recognizing the type of burnt leaf disease, which reached 72.5%.
Acknowledgment

The authors acknowledged the support of fund from Politeknik Negeri Lhokseumawe

5. Reference

[1] Angkoso, C. V., 2011. Introduction to Texture-Based Gender in Digital Photo Face Images. Journal of the National Conference on Innovation in Design and Technology.

[2] Irwan, N., 2016. Application of Expert System for Identifying Diseases in Rice Plants using the Certainty Factor method (Doctoral dissertation, Maulana Malik Ibrahim State Islamic University).

[3] Munir, Rinaldi. Digital Image Processing. Cet. 1. Bandung: Informatics, 2004.

[4] Nasir, M., 2013. Detection of Age of Rice Plants Based on Color Index. Journal of the National Information and Communication Technology Seminar (SNASTIKOM).

[5] Nasir, M., 2015. Texture Analysis of Rice Plants Using Extraction of Gray-Level Co-Occurrence Matrix Features. LITEK Journal

[6] Nasir, M., 2013. Detection of Age of Rice Plants Based on Color Index. National Seminar on Information and Communication Technology (SNASTIKOM)

[7] Son, Dharma. 2010. Digital Image Processing. ANDI. Yogyakarta.

[8] Rahman, M. And Wasista, s., 2010. Face Recognition System uses Webcem for Attendance using the TemplateMatching method. EEPIS Final Project.

[9] Wibawanto, H., Susanto, A., Widodo, T.S and Tjokronegoro, S.M., 2009. Identification of Mass Kistik Image Based on the Graylevel Co Occurrence Matrix Feature. In the National Seminar on Information Technology Applications (SNATI).