Identification of Bacterial Leaf Blight and Brown Spot Disease in Rice Plants with Image Processing Approach

Ihsanuddin[1], Eka Wahyu Hidayat[2], Alam Rahmatulloh[3]

[1,2,3]Department of Informatic, Siliwangi University, Tasikmalaya, 46115, Indonesia

ARTICLE INFO

ABSTRACT

In agriculture, technology can provide benefits to farmers. However, at present there are still very few farmers who use technology, especially computerization in their agricultural processes, such as the identification of diseases in rice plants, there are still many rice farmers who cannot recognize and distinguish the types of diseases in rice plants. Research on the identification of bacterial leaf blight and brown spots on rice plants has carried out before, but the accuracy rate is only 70%. This research developed a system to identify bacterial leaf blight and brown spot in rice plants through leaf images with an image processing approach. Image of affected rice leaves is segmented first using K-Means Clustering, then the texture features are extracted using the Gray Level Co-Occurrence Matrix (GLCM) with features extracted in the form of energy, contrast, correlation, homogeneity and shape pattern characteristics using metric and eccentricity features, then identified using Euclidean Distance. The training data used 40 images for each disease and 12 images for each disease. The test results show that the system has a better level of accuracy than previous studies that reached 100% with a Mean Squared Error (MSE) value of 0.007282214.

Corresponding Author:

Alam Rahmatulloh
Department of Informatic
Siliwangi University, Tasikmalaya, Indonesia
Email: alam@unsil.ac.id

1. INTRODUCTION

In agriculture, technology can provide benefits to farmers, from the initial process, harvesting, distribution to sales. However, at present only a few farmers use technology, especially computerization, in their agricultural processes, such as the identification of diseases in rice plants, there are still many rice farmers who cannot recognize and distinguish the types of diseases in rice plants. Many diseases in rice plants, make it difficult for farmers to identify and classify the types of diseases that attack the rice plants they plant. Farmers' knowledge about the identification of the types of diseases in rice plants and their classification can obtain from experience and counseling, from individuals and related agencies, and many are not computerized. Research on the identification of diseases in rice plants, in general, has been done before by [1] whereas more specific research on the identification of bacterial leaf blight and brown spots on rice plants was carried out by [2] using the Gabor Wavelet and K-Means Clustering algorithm, but the level of accuracy only reaches 70%.

From previous studies, accuracy is still low, it is necessary to make a clustering application with a different approach. For this reason, applications need to be made to identify leaf diseases in rice plants with better accuracy. Identification of diseases in rice plants based on symptoms that arise on rice leaves using an image processing approach, the diseases identified include rice crop diseases which cause the most significant crop loss, namely bacterial leaf blight and brown spots. From this problem, this research will create a system to identify diseases in rice plants through leaf images with an image processing approach, where the image of affected rice leaves is segmented first using K-Means Clustering, then extracting its texture features using the Gray Level Co-Occurrence Matrix with features extracted in the form of energy, contrast, correlation, and homogeneity. Where the research conducted on [3]–[6] all four of these features produce high accuracy, also the shape features extracted using metric features and eccentricity as in research [7] [8] which is able to extract

forms from the symptoms of bacterial leaf blight and brown spots on rice plants, then identified using Euclidean Distance. This system is expected to be more accurate so that it can help farmers to identify diseases in rice plants, notably bacterial leaf blight and brown spot diseases.

2. RESEARCH METHOD

The stages of the method in this study began from the Observation, Image Collection, Image Selection, Image Preprocessing, Image Acquisition, Developing System and Application, Testing and Result, Conclusion stages. The system design uses input images of rice leaves affected by bacterial leaf blight and brown spots obtained from observations. Before identification, bacterial leaf blight and brown spot images, first processed through the stages of image data collection, image data selection, the pre-processing stages and then acquired.

Image of training data on bacterial leaf blight and brown spot that has been acquired, segmented first using the K-Means Clustering algorithm, then extracted the texture characteristics using the Gray Level Co-Occurrence Matrix (GLCM) algorithm and the shape pattern characteristics using the Metric & Eccentricity algorithm, then entered into the training data database.

For the test data images that have been acquiring, are entered into the system one by one. Then the image that has been enter is segmented using the K-Means Clustering algorithm, then the image feature extraction is performed using the Gray Level Co-Occurrence Matrix (GLCM), Metric & Eccentricity algorithm, then the identification process is carried out using the Euclidean Distance algorithm by using data comparisons through a training data database, then the identification results are displayed. The design of the identification system for bacterial leaf blight and brown spot disease can see in Figure 1.
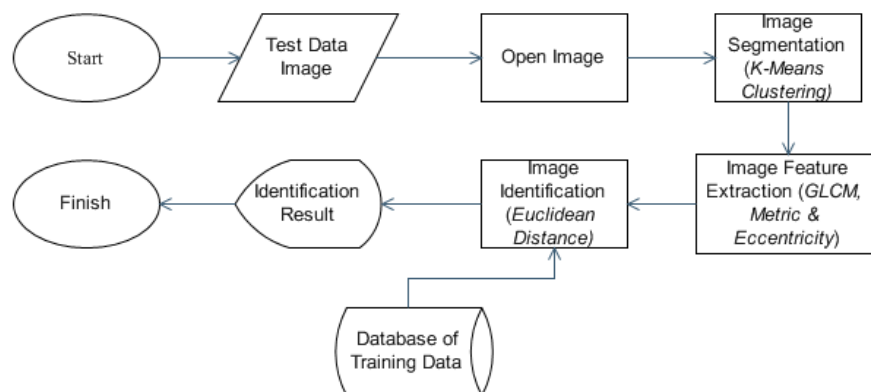

Fig. 1 Flowchart system

2.1. Observation

This process makes direct observations of rice leaves affected by bacterial leaf blight and brown spots, then take pictures using the camera of the Xiaomi Redmi 5A 13MP smartphone and serve as test data. For the training data image, image of bacterial leaf blight and brown spots on rice plants taken from previous research from the University of California, Irvine (UCI) Machine Learning Repository.

2.2. Image Data Collection

This stage is collecting training data images and test data from observations. The results of the observations obtained 40 training data images for each disease. For the test data, the images obtained totaled 40 images, with details of each type of disease totaling 20 images.

2.3. Image Data Selection

In this stage, image data of bacterial leaf blight and brown spot disease on rice plants that have been collected then selected. The image data that is unique and following needs then continued at the pre-processing stage and weak image data eliminated.

In the training data, there is no selection, besides because it is a dataset from previous research also because image data already has good quality. Whereas for the test data 8 images were eliminated for each disease, a part from poor quality, it is also due to the need for test data, which is only 30% of the training data, which is 12 images for each disease.

2.4. Image Data Pre-Processing

At this stage, the rice leaf image data contained symptoms of bacterial leaf blight and brown spot disease which had selected, then the image quality improvement process is carried out. Then the image quality improvement process is carried out and background removal. The background needs to remove so that the retrieval of character values is more focused on the image of rice leaves without interference from the background image. The process is carried out using image processing software, namely Adobe Photoshop CS6.

The training data not pre-processed because the data quality is excellent. Whereas all test data pre-processed in the form of background removal. The process and results of the pre-processing stages can see in Figure 2 and Figure 3.

Fig. 2 Before pre-processing   **Fig. 3** After pre-processing

2.5. Image Data Acquisition

This stage is the final stage of data collection, where the rice leaf image data contained symptoms of bacterial leaf blight and brown spot disease who have gone through stages of collection, selection, and pre-processing ready to be used in the research.

Image of training data acquired as many as 40 images for each disease, examples of training data images acquired for each disease can see in Figure 4. As for the test data acquired, there were 12 images for each disease. An example of the image of the acquired test data can see in Figure 5.
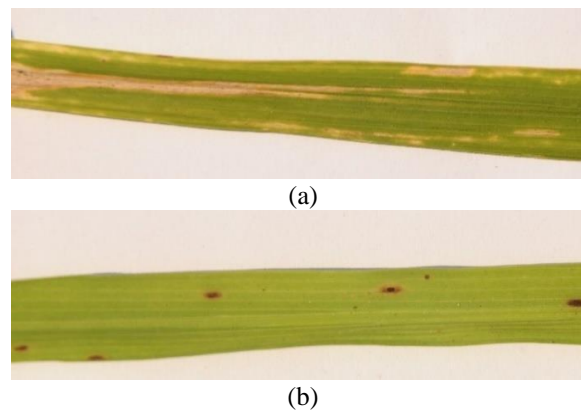

(a)

(b)

Fig. 4 (a) Image training data on bacterial leaf blight acquired (b) Image training data on brown spot acquired

(a)

(b)

Fig. 5 (a) Image test data on bacterial leaf blight acquired (b) Image test data on brown spot acquired

Image of training data and test data of bacterial leaf blight and brown spots on the rice plants obtained then changed first to the form of grayscale image using the Matlab R2015a application. The process and results of training data transformation changed to grayscale mode can be seen in Figure 6 and Figure 7.
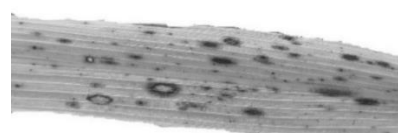
Fig. 6 RGB image

Fig. 7 Grayscale image

2.6. Database of Training Data

The database contains the value of texture features with features of contrast, correlation, energy, homogeneity, and the value of the shape pattern characteristics with the metric and eccentricity features of the training data image. The database consists of 80 rows with one row for one image and six columns for each feature value extracted from the image.

2.7. K-Mean Clustering Algorithm

The K-Mean Clustering algorithm used with the following stages: first, specify number of clusters K. Second, initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement. third, keep iterating until there is no change to the centroids. i.e. assignment of data points to clusters isn't changing. The algorithm is used to segment and divide the test data into 3 cluster symptoms, namely bacterial leaf blight disease, browns spot disease, and normal. Calculation of the distance between centroids and data is done using the Euclidean Distance algorithm.

2.8. Disease in Rice Plants

Rice plants attacked by various diseases with various causes. However, in this research, the diseases that will identify are only two diseases that cause the highest yield loss, namely bacterial leaf blight and brown spot disease.

a. Bacterial Leaf Blight

Bacterial Leaf Blight (*Xanthomonas Oryzae*) is a disease caused by widespread bacteria called *Xanthomonas Oryzae Pv. Oryzae*. This disease occurs in the rainy season or wet-dry season, especially on rice fields that always flooded.

Bacterial leaf blight disease produces two distinctive symptoms, namely crackle and blight. Crackle is a symptom that occurs in plants aged <30 days. The leaves are gray-green, folded, and curled. In severe conditions, all leaves roll, wither, and die while blight is the most common symptom found in plantations that have reached the stage of growth until the adult phase.

Symptoms begin with the onset of gray (yellowish) spots generally on the edge of the leaf. During its development, symptoms will spread, form blight, and finally the leaves dry. In moist conditions, groups of bacteria, in the form of golden yellow granules, can be easily found on leaves that show symptoms of blight [9], Symptoms of bacterial leaf blight on rice leaves can see in Figure 8.

Fig. 8 Symptoms of bacterial leaf blight

Loss of yield due to Bacterial Leaves Disease varies between 15 - 80%, depends on the stage of the plant when the disease arises [10].

b. Brown Spot

Brown spot disease caused by the fungus *Helmintosporium Oryzae* on plantations. Brown spots can cause the death of young plants and reduce grain quality. Like Cercospora, this disease is very damaging to rice cropping inland with weak drainage systems or land that is nutrient deficient.

The most common symptom of this disease is brown spots, oval to round, the size of a sesame seed, on the surface of the leaf, on the midrib or the grain, Pathogens are seed-borne. So that it is suitable, the disease can develop in very young plants [9]. This disease results in a loss of between 20% - 40%. Symptoms of brown spots on rice leaves can see in Figure 9.

Fig. 9 Symptoms of brown spot disease

2.9. Gray Level Co-Occurrence Matrix

Co-occurrence means a joint event, i.e., the number of occurrences of one level of neighboring pixel values with one level of other pixel values in the distance (*d*) and orientation of certain angles (*θ*) [11]–[14] Distance expressed in pixels and orientation expressed in degrees. The orientation formed in four angular directions with 45 ° angle intervals, i.e., 0°, 45°, 90°, and 135°. While the distance between pixels usually set at 1 pixel. These four directions represented in Figure 10.
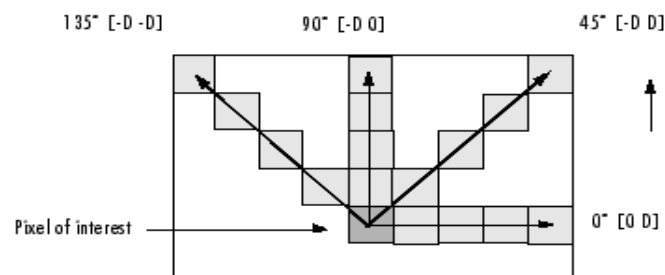

Fig. 10 GLCM angle

The co-occurrence matrix is a square matrix with as many elements as the square of the pixel intensity level in the image. Each point (*p*, *q*) on the orientation matrix is oriented (*θ*) contains the probability of occurrence of (*p*) value pixels with pixels worth (*q*) at a distance (*d*) and orientation (*θ*) and (180−*θ*) [15].

The co-occurrence matrix is formed from an image by looking at pairs of pixels that have a certain intensity. The use of this method based on a hypothesis that in a texture configuration recurrence will occur or gray level pair. For example, (*d*) defined as the distance between two-pixel positions, i.e. (*x1*, *y1*) and (*x2*, *y2*); and (*θ*) defined as the angle between the two. Then the co-occurrence matrix is defined as a matrix expressing the spatial distribution between two neighboring pixels which has intensities i and j, which has a distance d between the two, and the angle *θ* between the two. The co-occurrence matrix expressed as Pd, θ (*i*, *j*). A neighboring pixel that has a distance d between the two can be located in eight different directions [16]. Among the feature extraction features to be used, there are four features, as follows:

a. Contrast
The contrast shows the size of the spread (moment of inertia) elements of the image matrix. If it located far from the main diagonal, the contrast value is significant. Visually, the contrast value is a size of variation between the gray degrees of an image area. The contrast value obtained from (1).

$$Contrast = \sum_{i,j=1}^{N} P(i,j)(i-j)^2 \tag{1}$$

b. Correlation
Correlation shows the size of the linear dependence of the degree of a gray image so that it can give a hint of a linear structure in the image. Correlation values are obtained through (2).

$$Correlation = \sum_{i,j=1}^{N} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i \sigma_j} \tag{2}$$

c. Energy

Energy expresses the distribution of pixel intensity over the gray level. The energy value obtained from (3).

$$Energy = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{P(i,j)\}^2 \tag{3}$$

d. Homogeneity

Homogeneity is the value obtained from the similarity of variations in image intensity. Homogeneity value obtained from (4).

$$Homogeneity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i,j)}{1+|i-j|} \tag{4}$$

2.10. Metric and Eccentricity

In recognition of shape patterns, there are two components, namely eccentricity, and metric. *Eccentricity* (*e*) is the value of the ratio between the distance of the ellipse minor foci (*b*) *with the ellipse major foci* (a) of an area/shape on the object wrote in (5). Eccentricity values range from 0 to 1, an elongated area (approaching a straight line), then the eccentricity value will be close to 1, while the circular area, the eccentricity value will approach 0.

Metric (M) is the ratio between area (A) and circumference (C) of the area of an object [17]. Calculation of the metric values shown in (6).

$$e = \sqrt{1 - \frac{b^2}{a^2}} e = \sqrt{1 - \frac{b^2}{a^2}} e \tag{5}$$

$$M = \frac{4\pi \cdot A}{C^2} \tag{6}$$

3. RESULTS AND DISCUSSION

The results of this study are building applications by implementing the K-Mean Clustering algorithm to classify leaf diseases in rice planting. applications that are built can work well according to their functions.

3.1. System Implementation

Implementation starts with creating a graphical user interface, then provide functions to each component in the system according to the design. The results of system implementation can see in Figure 11.
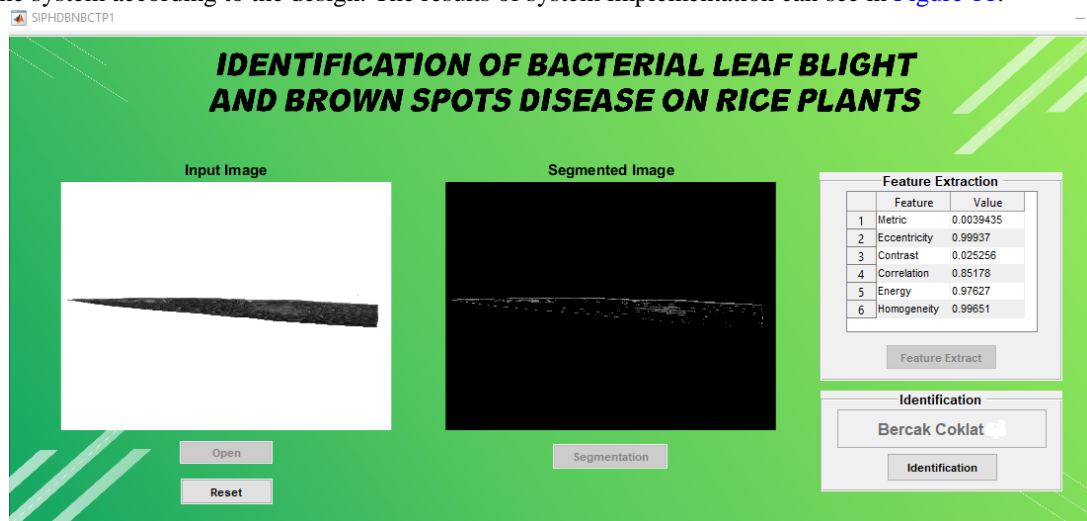

Fig. 11 System implementation

The identification process starts with inputting the image into the system, then the image that has inputted is segmented (K-Means Clustering), after segmenting the image then texture characteristics are taken (GLCM) and its shape patterns (Metric & Eccentricity) and finally identified (Euclidean distance) type of disease. The system must reset first, to start a new identification process.

3.2. System Testing

System Testing in this research using Black Box Testing. Then proceed with testing the accuracy of the system in identifying bacterial leaf blight and brown spot identification.

a. Black Box Testing

The black box testing is shown in Table 1. Based on table, the results of black-box testing of the system of identification of bacterial leaf blight and a brown spot on rice plants, that the system is working correctly and following the design that has made before.

Table 1. System black box testing

No.	Testing Scenarios	Expected results	System Response	Results
1.	Input the test image to the system.	- Display the loading bar of the image input process. - Display the test image on "axes1" on the system. - Activate the segmentation process button.	- Display the loading bar of the image input process. - Display the test image on 'axes1' on the system. - Segmentation process button active	Accepted
2.	Test image segmentation	- Display the loading bar of the image segmentation process. - Display test images that have segmented. - Activate the feature extraction button.	- Display the loading bar of the image segmentation process. - Display test images that have segmented. - Feature extraction button active	Accepted
3.	Feature extraction of test images	- Display the loading bar of the image feature extraction process. - Displays value of image characteristics in 'table1'. - Activate the identification button.	- Display the loading bar of the image feature extraction process. - Displays value of image characteristics in 'table1'. - Identification button active.	Accepted
4.	Test image identification	- Display the results of identifying the test image in the 'edit1' text box.	- Display the results of identifying the test image in the 'edit1' text box.	Accepted
5.	System reset	- Remove images from 'axes1' and 'axes2'. - Disable segmentation button, feature extraction button, and, identification button.	- Remove images from 'axes1' and 'axes2'. - Segmentation button, feature extraction button, and identification button are off.	Accepted

3.3.1. Accuracy Test

This stage of testing is carried out to see the accuracy of the system in identifying bacterial leaf blight and brown spot disease in rice plants. At this stage all the test images are processed into the system one by one, then the compatibility between the test image and the result of system identification is seen, then calculate the Mean Squared Error (MSE) value of each test. MSE is the mean square error value between the training data image and the test data image whose values are between 0-1, the more similar the two images. Then the MSE value is getting closer to zero. The equation used to calculate the MSE value can be seen in (7).

$$MSE = \frac{1}{m \, x \, n} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} [f(i,j) - g(i,j)]^2 \tag{7}$$

where m & n is a dimension of the image, i & j is coordinates of a point in the image, f is test data image, and g is the image of training data. The results of tests conducted on 24 test images are as in Table 2.

Table 2. Accuracy Test Results

No.	File Name	Identification Results	Information	MSE value
1.	HDB 01.JPG	Bacterial Leaf Blight	Correct	0.0335842
2.	HDB 02.JPG	Bacterial Leaf Blight	Correct	0.00302611
3.	HDB 04.JPG	Bacterial Leaf Blight	Correct	0.00255864
4.	HDB 05.JPG	Bacterial Leaf Blight	Correct	0.00255134
5.	HDB 07.JPG	Bacterial Leaf Blight	Correct	0.00314484
6.	HDB 09.JPG	Bacterial Leaf Blight	Correct	0.0026287
7.	HDB 11.JPG	Bacterial Leaf Blight	Correct	0.00305181
8.	HDB 12.JPG	Bacterial Leaf Blight	Correct	0.00298113
9.	HDB 16.JPG	Bacterial Leaf Blight	Correct	0.00172318
10.	HDB 17.JPG	Bacterial Leaf Blight	Correct	0.00487105
11.	HDB 18.JPG	Bacterial Leaf Blight	Correct	0.0032071
12.	HDB 20.JPG	Bacterial Leaf Blight	Correct	0.00256014
13.	BC 01.JPG	Brown Spots	Correct	0.00524459
14.	BC 02.JPG	Brown Spots	Correct	0.00530681
15.	BC 04.JPG	Brown Spots	Correct	0.00714653
16.	BC 05.JPG	Brown Spots	Correct	0.00644253
17.	BC 06.JPG	Brown Spots	Correct	0.00948043
18.	BC 08.JPG	Brown Spots	Correct	0.0112806
19.	BC 09.JPG	Brown Spots	Correct	0.00747053
20.	BC 11.JPG	Brown Spots	Correct	0.00561636
21.	BC 12.JPG	Brown Spots	Correct	0.00377095
22.	BC 13.JPG	Brown Spots	Correct	0.00660395
23.	BC 15.JPG	Brown Spots	Correct	0.00546869
24.	BC 18.JPG	Brown Spots	Correct	0.0652787
		Average		0.007282214

The Table 2 explains the results of MSE clustering testing accuracy with an average of 0.007282214. Mean Square Error (MSE) is a parameter used as an indicator to measure the similarity of two images from the training image and test image. These parameters are often used to compare the results of image processing with the initial image or original image. The more similar the two images, the closer the MSE value is to zero. The level of accuracy of the system in identifying bacterial leaf blight and brown spots on rice plants can see in (8).

$$Accuracy(\%) = \frac{\sum successful \, trial}{\sum trial} \, x \, 100\%$$
$$Accuracy \, (\%) = \frac{24}{24} x \, 100 = 100\% \tag{8}$$

4. CONCLUSION

Based on an accuracy test on 12 images of bacterial leaf blight test data and 12 images of brown spot test data on rice plants, 100% accuracy is obtained with an average value of Mean Squared Error (MSE) of

0.007282214. Combination of feature extraction algorithm, Gray Level Co-occurrence Matrix (GLCM), Metric and Eccentricity. K-Means Clustering algorithm for image segmentation, and the Euclidean Distance classification algorithm in identifying bacterial leaf blight and brown spots disease on rice plants. It has a better level of accuracy than in previous research with the same object.

In this study successfully made an application to cluster leaf diseases in rice plants into three cluster symptoms, namely bacterial leaf blight disease, browns spot disease, and normal. Applications made indicate the level of accuracy reaches a maximum value. Applications created are only limited to clustering. Detailed explanations need to be added as additional information on each leaf disease in rice plants such as symptoms, causes, how to prevent, and how to overcome leaf diseases in rice plants.

REFERENCES

[1] K. S. Archana and A. Sahayadhas, "Automatic rice leaf disease segmentation using image processing techniques," *International Journal of Engineering and Technology(UAE)*, 2018. DOI: 10.14419/ijet.v7i3.27.17756

[2] S. Santhosh Kumar and B. K. Raghavendra, "Diseases Detection of Various Plant Leaf Using Image Processing Techniques: A Review," in *2019 5th International Conference on Advanced Computing and Communication Systems, ICACCS 2019*, 2019. DOI: 10.1109/ICACCS.2019.8728325

[3] M. M. Santoni, D. I. Sensuse, A. M. Arymurthy, and M. I. Fanany, "Cattle Race Classification Using Gray Level Co-occurrence Matrix Convolutional Neural Networks," in *Procedia Computer Science*, 2015. DOI: 10.1016/j.procs.2015.07.525

[4] O. R. Indriani, E. J. Kusuma, C. A. Sari, E. H. Rachmawanto, and D. R. I. M. Setiadi, "Tomatoes classification using K-NN based on GLCM and HSV color space," in *Proceedings - 2017 International Conference on Innovative and Creative Information Technology: Computational Intelligence and IoT, ICITech 2017*, 2018. DOI: 10.1109/INNOCIT.2017.8319133

[5] J. M. G. Aglibut, L. L. Alonzo, M. F. B. Coching, J. L. Torres, and N. B. Linsangan, "Skin disease identification system using gray level co-occurrence matrix," in *ACM International Conference Proceeding Series*, 2017. DOI: 10.1145/3057039.3057044

[6] S. K. P.S and D. V.S, "Extraction of Texture Features using GLCM and Shape Features using Connected Regions," *International Journal of Engineering and Technology*, 2016. DOI: 10.21817/ijet/2016/v8i6/160806254

[7] N. Rishi and J. Singh Gill, "An Overview on Detection and Classification of Plant Diseases in Image Processing," *International Journal of Scientific Engineering and Research*, 2014. Online

[8] P. Revathi and M. Hemalatha, "Classification of cotton leaf spot diseases using image processing edge detection techniques," in *IEEE Proceedings of the INternational Conference On Emerging Trends in Science Engineering and Technology: Recent Advancements on Science and Engineering Innovation, INCOSET 2012*, 2014. DOI: 10.1109/incoset.2012.6513900

[9] B. S. Chauhan, K. Jabran, and G. Mahajan, *Rice Production Worldwide*. 2017. DOI: 10.1007/978-3-319-47516-5

[10] Y. Suryadi *et al.*, "Determination of Pathotypes from Indonesian Xanthomonas oryzae Pv. Oryzae Population causing Bacterial Leaf Blight and their Reactions on Differential Rice," *Makara Journal of Science*, 2016. DOI: 10.7454/mss.v20i3.6241

[11] M. M. Mokji and S. A. R. A. Bakar, "Gray level co-occurrence matrix computation based on haar wavelet," in *Computer Graphics, Imaging and Visualisation: New Advances, CGIV 2007*, 2007. DOI: 10.1109/CGIV.2007.45

[12] B. Pathak and D. Barooah, "Gray-Level Co-Occurrence Matrix Considering Possible," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2013. Online

[13] P. Mohanaiah, P. Sathyanarayana, and L. Gurukumar, "Image Texture Feature Extraction Using GLCM Approach," *International Journal of Scientific & Research Publication*, 2013. DOI: 10.1.1.414.96981

[14] a Gebejes, E. M. Master, and a Samples, "Texture Characterization based on Grey-Level Co-occurrence Matrix," *Conference of Informatics and Management Sciences*, 2013. Online

[15] M. Pratiwi, Alexander, J. Harefa, and S. Nanda, "Mammograms Classification Using Gray-level Co-occurrence Matrix and Radial Basis Function Neural Network," in *Procedia Computer Science*, 2015. DOI: 10.1016/j.procs.2015.07.340

[16] G. Capizzi, G. Lo Sciuto, C. Napoli, E. Tramontana, and M. Wozniak, "Automatic classification of fruit defects based on Co-occurrence matrix and neural networks," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015*, 2015. DOI: 10.15439/2015F258

[17] R. F. Cooper, M. A. Wilk, S. Tarima, and J. Carroll, "Evaluating Descriptive Metrics of the Human Cone Mosaic," *Investigative Opthalmology & Visual Science*, vol. 57, no. 7, p. 2992, Jun. 2016. DOI: 10.1167/iovs.16-19072