JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 7, No. 2, December 2024, pp. 212~223
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: 10.36378/jtos.v7i2.3972

Classification Of Rice Plant Diseases Using K-Nearest Neighbor Algorithm Based On Hue Saturation Value Color Extraction And Gray Level Co-Occurrence Matrix Features

Siti Saniah[1], Mhd Furqan[2]
[1,2]Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Indonesia

Article Info

Article history:

Received 12 01, 2024
Revised 12 17, 2024
Accepted 12 30, 2024

Keywords:

K-Nearest Neighbor
HSV
GLCM
Rice Disease
Image Processing

ABSTRACT

This research aims to classify diseases in rice plants using the K-Nearest Neighbor (K-NN) algorithm based on HSV color and GLCM texture feature extraction. The main problem is how to identify diseases in rice automatically using digital images. Diseases such as Blight, Tungro, and Crackle often attack rice, thus requiring an accurate early detection system. Lack of understanding in recognizing disease symptoms manually often leads to handling errors. This research develops an image processing-based classification system to detect rice diseases. The methods used include RGB to HSV color space conversion, texture feature extraction using GLCM, and classification using K-NN algorithm. The dataset consists of 240 images, divided into 192 training data and 48 testing data. Testing is done by calculating accuracy at parameter values K = 1, K = 3, and K = 5 to evaluate the model. The results showed that the combination of HSV and GLCM features produced the best accuracy at K=3 with 75% accuracy. This system can help farmers detect rice diseases quickly and effectively, minimize production losses, and support agricultural sustainability. This research is expected to provide practical solutions for farmers in detecting rice diseases, so that control can be carried out more accurately and efficiently.

Corresponding Author:

Siti Saniah
Department of Computer Science
Universitas Islam Negeri Sumatera Utara
Medan, Indonesia
Email: sitisaniahnasution@gmail.com

1. Introduction

Rice is a very important factor related to food security. According to the Central Bureau of Statistics (BPS), rice is currently the largest raw material, number one in 2023, with a harvest area of 10.4 million hectares and a production volume of 54.6 million tons, therefore rice productivity greatly affects food security and the economy. The level of rice productivity is affected by several inhibiting factors such as disease. This disease attack is the biggest challenge in rice cultivation [1]. Lack of understanding in recognizing the type of disease and changes in symptoms that occur both in color and shape on rice leaves, can lead to errors in handling, control and maintenance of plants and show that farmers use pesticides tend to be excessive and not in accordance with the prevalence of disease, as well as the selection of varieties that are not in accordance with the circumstances of the planted place. So that the quality and quantity of rice has decreased and will experience losses [2].

With the development of increasingly modern technology, image processing can be a solution to this problem. Image processing allows detecting rice plant diseases through images of infected leaves. As a first step, this research uses a widely used technique in image processing, namely feature extraction. Feature extraction is a stage to recognize the characteristics or object information in the image that will be the basis for classification [3]. Feature extraction that will be carried out on this rice plant disease uses the HSV (Hue Saturation Value) and GLCM (Gray Level Co-occurrence Matrix) methods. Then there are various algorithms to perform the classification process, one of which is the K-Nearest Neighbor Algorithm.

K-Nearest Neighbor is one of the algorithms that classifies and classifies by looking at the characteristics of its nearest neighbors [4]. It actually involves using the data closest to 'K' (its neighbors) to determine the class of new data based on its closeness/similarity to other data being classified. Of the many distances in K-Nearest Neighbor that are often used euclidean distance, Euclidean distance aims to test the size of the interpretation of the proximity of the distance of two objects [5].

In previous research conducted by Setyawan et al. there is an application of the K-NN Algorithm with the extraction of HSV Color and Texture Features to detect corn leaf disease by showing optimal results and at a value of k = 3 which is able to get results achieved with an accuracy of 84%, with correct prediction data identified as much as 168 data from the total training data [6].

Then research conducted by Irawan et al, where this research applies color extraction using the Mean Hue Saturation Value (HSV) method and feature extraction using Gray Level Co-Occurrence Matrix (GLCM) with the K-Nearest Neighbor (KNN) algorithm. Based on the comparison between KNN, GLCM-KNN and HSV-KNN, it is known that the value of KNN alone is the lowest. In KNN, the value of K = 1 with d = 1 is 80%, while the best results can be obtained in the HSV-GLCM combination in K = 1 and d = 1, namely 95% [7].

Based on the background described above, the researcher intends to create a research system that can classify rice plant diseases using the K-Nearest Neighbor algorithm and also use the completion method with HSV color features and texture extraction using GLCM. This system is expected to help farmers for classification, especially diseases in rice plants. In the future, farmers can find out what diseases attack the rice and the initial control that must be done.

2. Research Method

Rice is a cultivated annual crop, with the Latin name Oryza sativa L. It is a very important food commodity around the world that fulfills the carbohydrate needs of almost half of the world's population [8]. However, rice is also susceptible to bacterial diseases. Bacterial diseases in rice can cause significant yield losses. Therefore, it is very important to treat the disease properly [9]. The types of rice include paddy rice, upland rice, hybrid rice, organic rice, aromatic rice, brown rice, white rice, super rice, black rice, and wild rice [10].

The problem that often arises is that many rice plants are susceptible to disease attacks during the planting period. Generally, when rice plants are attacked by pests and diseases, farmers will immediately use pesticides or treatment methods that are sometimes not in accordance with the disease experienced. As a result, treatment is not optimal and can even cause new diseases [11].

Image processing is a process of changing pixels in a digital image for a specific purpose [14]. Image preprocessing is used to simplify the process of identifying images [15]. The stages carried out in this Preprocessing are cropping to extract the necessary parts of the image and reduce its size, and Resizing where 200 x 200 pixels are used for each image [16].

Feature extraction is the most important part of image analysis to find out and recognize the special characteristics of an image [17]. The feature extraction method used in this study is divided into 2 stages of analysis, namely HSV color and GLCM (Gray-Level Co-Occurrence Matrix) which aims to find the value of the characteristics for each leaf image that has a disease [18].

Hue Saturation Value (HSV) is a color extraction feature used for basic color classification. HSV also has tolerance to changes in light intensity [19]. This HSV stage aims to distinguish the color of the fund image to determine the redness, greenness, of light [20]. In the equation below before doing HSV extraction, the conversion must first be done from RGB to HSV.

$$H = tan \frac{3(G-B)}{(R-G)+(R-B)} \tag{1}$$

$$S = 1 - \frac{\min(R,G,B)}{V} \tag{2}$$

$$V = \frac{R+G+B}{3} \tag{3}$$

Conversion to HSV (Hue Saturation Value) values

$$R = \frac{R}{R+G+B} \qquad G = \frac{G}{R+G+B} \qquad B = \frac{B}{R+G+B} \tag{4}$$

$$V = \max(r, g, b) \tag{5}$$

$$S\begin{cases}0, & jika\ v = 0\\ 1 - \frac{\min(r,g,b)}{v}, & v > 0\end{cases} \quad (6)$$

GLCM is a second-order texture feature extraction method and its calculation is based on the proximity of two neighboring pixels with angular and distance directions [21]. At this stage, the matrix calculation describes the frequency. This GLCM has 4 proper headings with pixels specifically 0˚, 45˚, 90˚ and 135˚[22]. Gray Level Co-occurrence Matrix produces 4 feature extractions as follows [23].

Matrix Normalization :

$$P\{i,j|\Delta_x,\Delta_y\} = WQ\{i,j|\Delta_x,\Delta_y\}$$

Energy (ASM) :

$$Energy \sum_{j=1}^{N-1}\sum_{j=0}^{N-1} p_i j^2 \quad (7)$$

Contrast :

$$contrast \sum_{i,j}^{N-1}(K)^2 X(i,), |i-j| = k \quad (8)$$

Correlation :

$$Correlation \sum_{i,j}^{N-1} pi,j \left|\frac{(i-\mu_j)(j-\mu_j)}{\sqrt{\sigma i^2 \sigma j^2}}\right| \quad (9)$$

Homogenity :

$$Homogeneity \sum_{i,j}^{N-1} \frac{(pi,j)^2}{1+(i-j)} \quad (10)$$

K-Nearest Neighbors is a classification method that uses data algorithms from training data, and classification results are obtained based on the nearest neighbor distance [24]. In the context of classification, the KNN algorithm finds the k-nearest neighbors of an unknown data point in the feature space. The working principle is based on the Euclidean distance, usually the proximity or distance of the neighbor is calculated. Euclidean Distance is said to be good if the new data inputted by the user has a small enough distance and has a high similarity [25].

$$d(x,y)\sqrt{\sum_{i=1}^{p}(xi-yi)^2} \quad (11)$$

This research uses a quantitative approach [26] to evaluate the performance of the K-Nearest Neighbor (KNN) algorithm in classifying rice plant diseases with a total dataset of 240 data. Data in the form of rice leaf images are processed using HSV color extraction and GLCM (Gray Level Co-occurrence Matrix) texture features. The performance is measured by accuracy, precision, recall, and F1-score metrics.

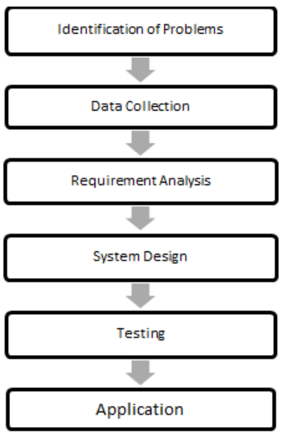


Figure 1. Research stages

Figure 1 shows the research framework that will be carried out starting from the problem identification stage to the application stage. This problem identification stage is a way to identify and analyze existing problems. At this stage will produce problem formulations, research objectives and also problem boundaries. The data collection techniques used in this research are literature studies and dataset downloads. The literature used is journals, websites, and previous research. And data collection is also done by downloading datasets from the Kaggle platform. The selected dataset focuses on rice plant diseases, which contains images of rice leaves infected with various diseases. This needs analysis is a breakdown of the requirements used in the implementation of the classification system software to be built to support it to run well. The requirements include hardware and software. This stage is the stage about the design of the system to be built. berdasarkan analisa yang dilakukan. Pada tahap pengujian bertujuan untuk mengimplementasikan system properly and correctly according to user needs and to find out whether the system made has met the requirements in accordance with the purpose of system design. This test is also to measure how much the percentage of accuracy of the system that has been made. The application of this system is to identify rice plant diseases based on HSV and GLCM extraction and classify rice plant leaves. By using images, the object of rice leaf disease can be processed to allow computers to recognize images like human vision.

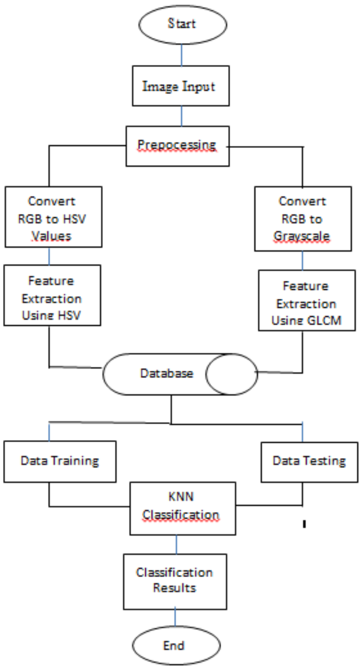


Figure 2. System Planning Diagram

The figure above is a system planning diagram of the classification of diseases in rice plants at this stage starting from acquisition which is the collection of image data then the processing stage, this stage is carried out cropping image cutting and resizing so that the image size becomes smaller. In addition, the RGB (Red Green Blue) color conversion stage is also carried out to HSV (Hue Saturation Value) and RGB conversion to grayscale. Characteristic extraction of HSV color features and GLCM (Gray Level Co-Occurrence Matrix) texture features whose results will be stored in the database. The last stage is the classification stage carried out using the K-Nearest Neighbor method which results in the classification of rice plant diseases. The system built on the classification of rice plant diseases is based on the Matlab GUI. Which only has one user only. Users can classify rice leaf disease.

3. Result and Discussion

Data that has been obtained as much as 240 data in the form of images will be processed. In this research, data analysis is carried out starting with inputting images of rice diseases that will go through the process of extracting rice plant image features using Hue Saturation Value (HSV) and Gray Level Co-Occurrence Matrix (GLCM). Furthermore, the classification process is carried out using the K-Nearest Neighbor (K-NN) algorithm and finally experiments will be carried out on the system that will be made based on the learning model that has been designed.

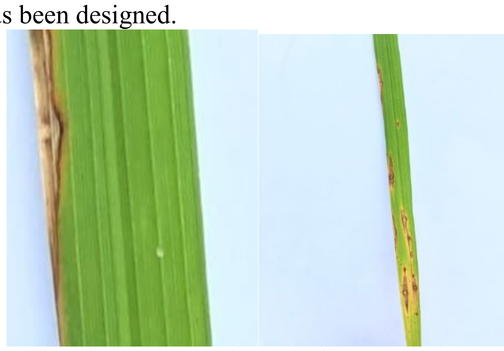


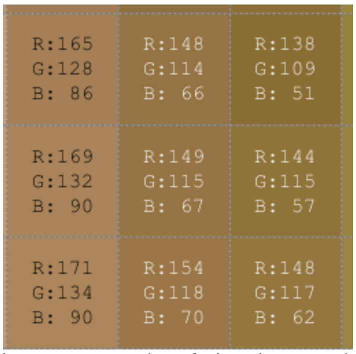
Figure 3. Rice Disease Image Before and After Preprocessing

The figure above is an image of rice disease before and after prepocessing the sample image of rice plants that have been prepocessing with Red Green Blue (RGB) values at 3×3 pixels which will be used to find the Hue Saturation Value (HSV) value..

5	5	4	3	2	2	2	1
5	5	4	4	3	2	2	1
5	5	4	4	3	2	2	1
5	5	4	4	4	3	2	1
5	5	4	4	4	3	2	1
5	5	4	4	4	4	3	2
5	5	4	4	4	4	4	3
4	5	4	4	4	4	5	3

Figure 4. Sample Image of Rice Plant with Grayscale Matrix

In the figure above is a sample image of rice plants with a grayscale matrix that has a value in each pixel and has gone through a preprocessing process and has 8 degrees of gray with a color range of 0-7. After the preprocessing stage, the rice plant image sample will be processed to the image feature extraction stage using Gray Level Co-Occurrence Matrix (GLCM). Then the image extraction process uses the Hue Saturation Value (HSV) method, where this method converts the RGB color model to HSV. The following is the RGB value of the 3×3 pixels in the rice plant image sample that has gone through the data pre-processing stage. for the calculation of RGB values taken at pixel 40 to pixel 43.



R:165 G:128 B: 86	R:148 G:114 B: 66	R:138 G:109 B: 51
R:169 G:132 B: 90	R:149 G:115 B: 67	R:144 G:115 B: 57
R:171 G:134 B: 90	R:154 G:118 B: 70	R:148 G:117 B: 62

Figure 5. RGB Value of Rice Plant Samples

The first step in converting the RGB color model to the HSV color model is to normalize the RGB value at each pixel of the rice sample in Figure 5. It is known that at pixel (1,1) the value of R = 165 G = 128 and B = 86, then the normalized RGB value at pixel (1,1) according to [14] is :

$$r = \frac{165}{165+128+86} = 0{,}435$$

$$g = \frac{128}{165+128+86} = 0{,}338$$

$$b = \frac{86}{165+128+86} = 0{,}227$$

The next step is to convert the rgb value that has been obtained to the HSV value. It is known from pixel (1,1) that the normalized value of r = 0.435 g = 0.338 and b = 0.227, then the HSV normalization value at pixel (1,1) is:

V = max (r, g, b)

V = 0,435

$$S = 1 - \frac{0{,}227}{0{,}435} = 1 - 0{,}521 = 0{,}479$$

Since the V value obtained is from the r value, the formula for finding the H value is as follows:

$$H = 60 \text{ X } \left(0 + \frac{0{,}338 - 0{,}227}{0{,}479 \; X \; 0{,}435}\right)$$

$$H = 60 \text{ X } \left(0 + \frac{0{,}111}{0{,}208}\right)$$

H = 60 X 0,533

H = 32,0

The calculation of finding the HSV value above as a whole is done with the same formula and method, so that the overall HSV value can be seen in the table below.

Table 1. HSV Conversion Values

X/Y		1	2	3
1	H	32,0	35,2	40,0
	S	0,479	0,552	0,631
	V	0,435	0,451	0,463
2	H	31,14	35,1	40,0
	S	0,466	0,551	0,605
	V	0,433	0,450	0,456
3	H	32,45	34,26	38,34
	S	0,473	0,545	0,581
	V	0,433	0,450	0,453

After the RGB value is converted to HSV form, the next step is to find the average (mean) of the three HSV features. The following is the average result of HSV feature extraction on rice plant image samples measuring 3×3 pixels.

Table 2. HSV Average Value

Mean H	35,39
Mean S	0,537
Mean V	0,491

Then, the Gray Level Co-Occurrence Matrix (GLCM) method extraction process uses an angle direction of 0. The matrix is transposed to obtain a symmetrical matrix then normalized and the search for the Contrast, Energy, Homogeneity, Correlation equation will also be carried out.

Table 3. Matrix Normalization Result 0^0

	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	0	0,0446	0	0	0	0	0
2	0	0,0446	0,0714	0,0536	0	0	0	0
…	…	…	…	…	…	…	…	…
7	0	0	0	0	0	0	0	0

Contrast:

$\text{Contrast}_{(1,2)} = (1-2)^2 \times 0{,}0446 = 0{,}0446$

$\text{Contrast}_{(2,1)} = (2-1)^2 \times 0{,}0446 = 0{,}0446$

$\text{Contrast}_{(2,2)} = (2-2)^2 \times 0{,}0714 = 0$

Contrast (total) =

(0,0446 + 0,0446 + 0 + 0,0536 + 0,0536 + 0,0625 + 0,0357 + 0,0625 + 0 + 0,0893 + 0,0357 + 0,0268 + 0)

Contrast(total) = 0,5714

Energy

$\text{Energy} = (0{,}0446)^2 + (0{,}0446)^2 + (0{,}0714)^2 + (0{,}0536)^2 + (0{,}0536)^2 + (0{,}0625)^2 + (0{,}0089)^2 + (0{,}0625)^2 + (0{,}2857)^2 + (0{,}0893)^2 + (0{,}0089)^2 + (0{,}0893)^2 + (0{,}1250)^2 = 0{,}1359$

Homogeneity

$$\text{Homogeneity}=\{\left(\frac{0{,}0446}{1+|1-2|}\right)+\left(\frac{0{,}0446}{1+|2-1|}\right)+\left(\frac{0{,}0714}{1+|2-2|}\right)+\left(\frac{0{,}0536}{1+|2-3|}\right)+\left(\frac{0{,}0536}{1+|3-2|}\right)+\left(\frac{0{,}0625}{1+|3-4|}\right)+\left(\frac{0{,}0089}{1+|3-5|}\right)+\left(\frac{0{,}0625}{1+|4-3|}\right)+\left(\frac{0{,}2857}{1+|4-4|}\right)+\left(\frac{0{,}0893}{1+|4-5|}\right)+\left(\frac{0{,}0089}{1+|5-3|}\right)+\left(\frac{0{,}0893}{1+|5-4|}\right)+\left(\frac{0{,}1250}{1+|5-5|}\right)\}$$

Homogeneity = 0,0223 + 0,0223 + 0,0714 + 0,0268 + 0,0268 + 0,0313 + 0,0018 + 0,0313 + 0,2857 + 0,0446 + 0,0029 + 0,0446 + 0,1250 = 0,6955

Correlation

$$\mu i = ((1x5) + (2x5) + (2x8) + (2x6) + (3x6) + (3x7) + (3x1) + (4x7) + (4x32) + (4x10) + (5x1) + (5x10) + (5x14))/112$$

$$\mu i = 3{,}6250$$

$$\mu j = ((1x5) + (2x5) + (2x8) + (2x6) + (3x6) + (3x7) + (3x1) + (4x7) + (4x32) + (4x10) + (5x1) + (5x10) + (5x14))/112$$

$$\mu j = 3{,}6250$$

$$\sigma i = \sqrt{\{(1-3{,}6250)^2 \times (0{,}0446)\} + \{(2-3{,}6250)^2 \times (0{,}0446)\}} + \{(2-3{,}6250)^2 \times (0{,}0714)\} + \{(2-3{,}6250)^2 \times (0{,}0536)\} + \{(3-3{,}6250)^2 \times (0{,}0536)\} + \{(3-3{,}6250)^2 \times (0{,}0625)\} + \{(3-3{,}6250)^2 \times (0{,}0089)\} + \{(4-3{,}6250)^2 \times (0{,}0625)\} + \{(4-3{,}6250)^2 \times (0{,}2857)\} + \{(4-3{,}6250)^2 \times (0{,}0893)\} + \{(5-3{,}6250)^2 \times (0{,}0089)\} + \{(5-3{,}6250)^2 \times (0{,}0893)\} + \{(5-3{,}6250)^2 \times (0{,}1250)\} = \sigma i = 1{,}1349$$

$$\sigma j = \sqrt{\{(1-3{,}6250)^2 \times (0{,}0446)\} + \{(2-3{,}6250)^2 \times (0{,}0446)\}} + \{(2-3{,}6250)^2 \times (0{,}0714) + \{(2-3{,}6250)^2 \times (0{,}0536)\} + \{(3-3{,}6250)^2 \times (0{,}0536)\} + \{(3-3{,}6250)^2 \times (0{,}0625)\} + \{(3-3{,}6250)^2 \times (0{,}0089)\} + \{(4-3{,}6250)^2 \times (0{,}0625)\} + \{(4-3{,}6250)^2 \times (0{,}2857)\} + \{(4-3{,}6250)^2 \times (0{,}0893)\} + \ + \{(5-3{,}6250)^2 \times (0{,}0893)\} + \{(5-3{,}6250)^2 \times (0{,}1250)\} = \sigma j = 1{,}1349$$

Correlation = {(0,1904 + 0,1904 + 0,1886 + 0,0544 + 0,0544 – 0,0146 – 0,0077 – 0,0146 + 0,0402 + 0,0460 – 0,0077 + 0,0460 + 0,2363)/ (1,1349)(1,1349)}

Correlation = 0,7782

This process is carried out until the results of the calculation of feature extraction with GLCM on the image of rice plants the author will use a distance of 1 pixel by using the four degrees of angular direction, namely 45^0, 90^0, 135^0.

Table 4. Calculation Results of Feature Extraction with GLCM

Features	Angel 0	Angel 45	Angel 90	Angel 135
Contrast	0,5714	0,856	0,2068	0,3468
Energy	0,1359	0,0978	0,1917	0,1634
Homogeneity	0,6655	0,6761	0,8928	0,6679
Correlation	0,7782	0,7210	0,1116	0,0844

After the calculation process of feature extraction with Gray Level Co-Occurrence Matrix (GLCM), the next step is to find the average value of all angles in order to obtain the single value of each feature so that it can facilitate the classification process. The average value of the four Gray Level Co-Occurrence matrix (GLCM) features can be seen in the table below.

Table 5. Average Value of Four GLCM Features

Contrast	Energy	Homogeneity	Correlation
0,5714	0,856	0,2068	0,3468

The next step is to analyze the data with the K-Nearest Neighbor model. K-Nearest Neihgbor (KNN) is a simple classification algorithm that works on the principle of similarity. K-Nearest Neihgbor classifies a new object based on the class of its K nearest neighbors in the feature space [27]. The following is the data of rice plant image extraction features using HSV and GLCM by using the system that has been designed:

Table 1. HSV Training Data Average Image Extraction

Photo	**Hue**	**Saturation**	**Value**	**Class Labels**

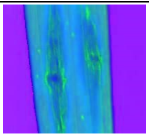
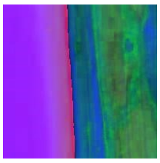
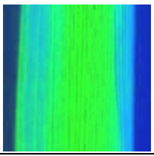
	0,32602	0,38299	0,88442	Blast
....	....	....	....	....
	0,34386	0,33315	0,62992	Blight
....	....	....	....	....
	0,14251	0,76144	0,50411	Tungro

Table 6 is the result of HSV feature image extraction, namely average Hue, average Saturation, and average Value on training data totaling 192 photo data of rice plants.

Table 7. Average HSV image extraction on training data

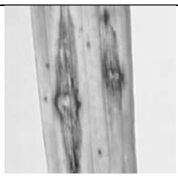
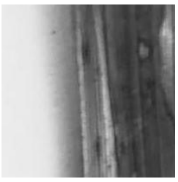
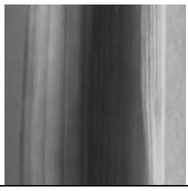
Photo	**Contrast**	**Energy**	**Homogeneity**	**Correlation**	**Class Labels**
	0,22852	0,28193	0,95894	0,97564	Blast
….	….	….	….	….	….
	0,14710	0,22070	0,96505	0,98703	Blight
....	....	....	....	....	....
	0,08775	0,23857	0,97757	0,98512	tungro

Table 7 is the result of extracting GLCM image features, namely average Contrast, average Energy, average Homogeneity, and average Correlation on training data totaling 192 photo data of rice plants. After the results of the GLCM and HSV image extraction of rice plants are known, then implement the test data samples using the K-Nearest Neighbor algorithm.

Tabel 8. Nilai Ekstraksi Fitur HSV dan GLCM pada Sampel Data Uji

Photo	**x1**	**x2**	**x3**	**x4**	**x5**	**x6**	**x7**	**Class Labels**

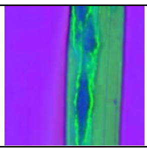
	0,21380	0,27170	0,96210	0,98531	0,42705	0,34240	0,78004	?

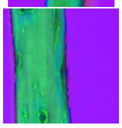
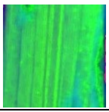
Table 8 will be used as test data to predict by looking at the training data in Table 6 and in Table 7 in order to find out which class the data belongs to. The variables that the author uses in Table 8 are the extraction of features from Gray Level Co-Occurance Matrix (GLCM) and features from Hue Saturation Value (HSV). With the following explanation. x1 = Contrast feature, x2 = Energy feature, x3 = Homogeneity feature, x4 = Correlation feature, x5 = Hue feature, x6 = Saturation feature, x7 = Value The distance measure used is Euclidean Distance. Where the K (neighborliness) value used is an odd value such as 1, 3, and 5.
Calculation of Euclidean Distance of test data against the 1st training data:

$$d1 = \sqrt{((0{,}22852 - 0{,}21380)^2 + (0{,}28193 - 0{,}27170)^2 + (0{,}95894 - 0{,}96210)^2 + (0{,}97564 - 0{,}98531)^2 + (0{,}326002 - 0{,}42705)^2 + (0{,}38299 - 0{,}34240)^2 + (0{,}88442 - 0{,}78004)^2}$$

$$d1 = 0{,}02317$$

For the calculation of Euclidean Distance on the 2nd test data up to 192 the process is the same as the calculation of Euclidean Distance on the test data against the first data. The results of the calculation of Euclidean Distance test data against all training data can be seen in the table below.

Table 9. Euclidean Distance Calculation Results of Test Data Against Training Data

Photo	x1	x2	x3	x4	x5	x6	x7	Euclidean Distance	Rank	Class Labels
	0,18773	0,30484	0,97256	0,98734	0,43360	0,30255	0,78177	0,00353	1	Tungro
	0,21561	0,32580	0,96313	0,98528	0,39647	0,35267	0,73881	0,00567	2	Blast
…	…	…	…	…	…	…	…	…	…	…
	0,07922	0,41281	0,97360	0,93879	0,18975	0,79355	0,36985	0,46841	192	Blast

The Matlab GUI is used to perform GLCM and HSV feature extraction on rice plant images.

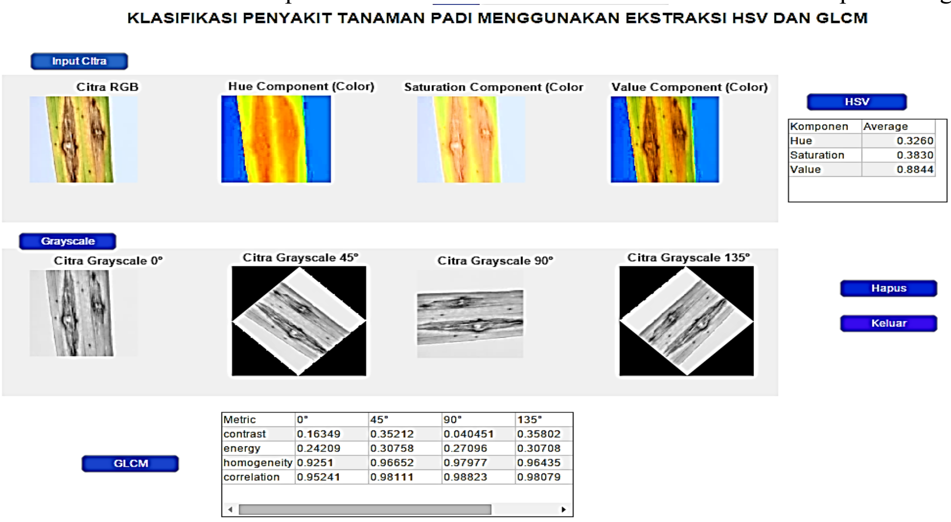


Figure 1. GLCM and HSV Extraction Process on Rice Plant Image

Then in this study testing the classification of diseases in rice plants using the K-NN algorithm with the Python programming language in Google Colab as an IDE. In this test, the author uses a neighbor value

(K) of 1, 3, and 5. Furthermore, the accuracy or success of the classification is evaluated using the confusion matrix as below.

```
Akurasi model KNN dengan K=1: 66.67%
Hasil Klasifikasi:
              precision    recall  f1-score   support

       Blast       0.53      0.62      0.57        16
      Blight       1.00      0.88      0.93        16
      Tungro       0.53      0.50      0.52        16

    accuracy                           0.67        48
   macro avg       0.69      0.67      0.67        48
weighted avg       0.69      0.67      0.67        48
```

Figure 7. K-NN Testing (K=1)

Testing the classification of rice plant images using values in the Gray Level Co-Occurrence Matrix (GLCM) and Hue Saturation Value (HSV) features using the K-Nearest Neighbor (K-NN) algorithm for a value of K = 1 gets a model accuracy of 66.67%.

```
Akurasi model KNN dengan K=3: 75.00%
Hasil Klasifikasi:
              precision    recall  f1-score   support

       Blast       0.63      0.75      0.69        16
      Blight       1.00      0.75      0.86        16
      Tungro       0.71      0.75      0.73        16

    accuracy                           0.75        48
   macro avg       0.78      0.75      0.76        48
weighted avg       0.78      0.75      0.76        48
```

Figure 2. Testing K-NN (K=3)

Testing the classification of rice plant images using values in the Gray Level Co-Occurrence Matrix (GLCM) and Hue Saturation Value (HSV) features using the K-Nearest Neighbor (K-NN) algorithm for a value of K = 3 gets a model accuracy of 75.00%.

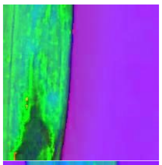
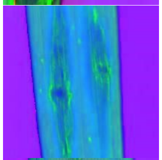
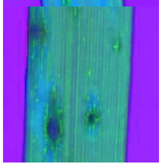
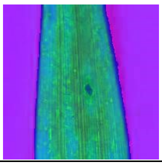
```
Akurasi model KNN dengan K=5: 70.83%
Hasil Klasifikasi:
              precision    recall  f1-score   support

       Blast       0.58      0.69      0.63        16
      Blight       1.00      0.75      0.86        16
      Tungro       0.65      0.69      0.67        16

    accuracy                           0.71        48
   macro avg       0.74      0.71      0.72        48
weighted avg       0.74      0.71      0.72        48
```

Figure 3. Testing K-NN (K=5)

Testing the classification of rice plant images by combining Gray Level Co-Occurrence Matrix (GLCM) and Hue Saturation Value (HSV) features was successfully carried out using the K-Nearest Neighbor (K-NN) algorithm with a value of (K = 5). Based on the test results, the model was able to achieve an accuracy of 70.83%, which shows a fairly good performance at an early stage. Details of the test data classification results can be seen in the following table.

Table 10. Classification Results of Rice Plant Image at Values K1, K3, and K5

K=1, Akurasi = 66,67%			K=3, Akurasi = 75%		K=5, Akurasi = 70,83%	
Photo	**Original Class Label**	**Prediction**	**Original Class Label**	**Prediction**	**Original Class Label**	**Prediction**

K=1, Akurasi = 66,67%			K=3, Akurasi = 75%		K=5, Akurasi = 70,83%	
Photo	**Original Class Label**	**Prediction**	**Original Class Label**	**Prediction**	**Original Class Label**	**Prediction**
	Blast	Blast	Blast	Blast	Blast	Blast
	Blast	Tungro	Blast	Tungro	Blast	Tungro
	Blast	Tungro	Blast	Tungro	Blast	Tungro
…	…	…	…	…	…	…
	Tungro	Blast	Tungro	Blast	Tungro	Tungro

4. Conclusion

Based on the research results, the rice plant image recognition system is carried out through a combination of two feature extraction methods, namely Hue Saturation Value (HSV) and Gray Level Co-Occurrence Matrix (GLCM), and the K-Nearest Neighbor (K-NN) algorithm for disease type classification. The HSV feature extraction process involves three main stages: converting the RGB color space to HSV, separating the image into Hue, Saturation, and Value components, and calculating the average of each component. Meanwhile, extraction using GLCM involves five steps, namely determining the probability of neighbor relationship between pixels at a certain distance, choosing an angular orientation (0°, 45°, 90°, and 135°), forming a cooccurrence matrix, normalizing it into probability form, and calculating the feature values of Contrast, Energy, Homogeneity, and Correlation.

In the classification stage using the K-NN algorithm, the Euclidean Distance is calculated to determine the K value with variations of K = 1, K = 3, and K = 5. The test results show that the highest accuracy is achieved at K = 3 with a value of 75.00%, followed by K = 5 with an accuracy of 70.83%, and K = 1 at 66.67%. These results show that the combination of HSV, GLCM, and K-NN algorithms has good potential to support disease classification in rice plants. The recommendation from this research is to explore additional algorithm parameters and features to improve the accuracy of future classification systems.

References

[1] M. K. Khamdani, N. Hidayat, and R. K. Dewi, "Implementasi Metode *K-Nearest Neighbor* Untuk Mendiagnosis Penyakit Tanaman Bawang Merah," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 1, pp. 11–16, 2021, [Online]. Available: http://j-ptiik.ub.ac.id

[2] A. Alfayanti, Y. Yesmawati, L. Harta, K. Dinata, and S. Yuliasari, "Persepsi petani terhadap teknologi pengendalian hama dan penyakit terpadu padi sawah dengan agensia hayati (studi kasus di Kelurahan Semarang Kota Bengkulu)," *Pros. Semin. Nas. Lahan Suboptimal*, vol. 9, pp. 233–241, 2021.

[3] W. S. Sari, C. A. Sari, F. I. Komputer, and U. D. Nuswantoro, "Klasifikasi Bunga Mawar Menggunakan KNN dan Ekstraksi Fitur GLCM dan HSV," vol. 5, pp. 145–156, 2022.

[4] W. Aliansa, H. N. Ifayatin, and R. A. Saputra, "Segmentasi Kematangan Pisang Raja Berbasis Fitur Warna HSV Menggunakan Metode KNN," *J. Sains Komput. Inform.*, vol. 7, no. 2, pp. 595–608, 2023.

[5] Z. Y. Lamasigi, S. -, H. -, and Y. Lasena, "Identifikasi Tingkat Kesegaran Ikan Tuna Menggunakan Metode GLCM dan KNN," *Jambura J. Electr. Electron. Eng.*, vol. 4, no. 1, pp. 70–76, 2022, doi: 10.37905/jjeee.v4i1.12045.

[6] M. A. Setyawan, P. Kasih, M. Ayu, and D. Widyadara, "Klasifikasi Penyakit Daun Jagung Berdasarkan Ruang Warna HSV dan Fitur Tekstur Dengan Algoritma K-NN," pp. 67–72, 2022.

[7] C. Irawan, E. H. Rachmawanto, C. A. Sari, and R. U. Nur, “Klasifikasi Citra Mengkudu Berdasarkan Perhitungan Jarak Piksel pada Algoritma K- Nearest Neighbour,” vol. 14, no. 02, pp. 200–207, 2023, doi: 10.35970/infotekmesin.v14i2.1827.

[8] I. F. Annur, J. Umami, M. N. Annafii, N. Trisnaningrum, and O. V. Putra, “Klasifikasi Tingkat Keparahan Penyakit Leafblast Tanaman Padi Menggunakan MobileNetv2,” *Fountain Informatics J.*, vol. 8, no. 1, pp. 7–14, 2023, doi: 10.21111/fij.v8i1.9419.

[9] H. P. Eryah, S. P. Telnoni, and Y. Da Costa, “Isolasi Dan Identifikasi Bakteri Penyakit Hawar Daun Di Desa Naibonat Kecamatan Kupang Timur, Kabupaten Kupang Nusa Tenggara Timur,” *Flobamora Biol.*, vol. 1, no. 2, pp. 9–18, 2022.

[10] M. D. Furqan, M. Sayuthi, and H. Hasnah, “Biodiversitas Arthropoda Predator pada Beberapa Varietas Padi Sawah,” *J. Ilm. Mhs. Pertan.*, vol. 8, no. 3, pp. 526–541, 2023.

[11] S. Katarina Sianturi, A. Syaefudin, and A. Nabila, “PKM Pengembangan Sistem Pakar Diagnosa Penyakit Padi pada Kelompok Tani Mekartani Desa Lebakwana,” *Sevana J. Pengabdi. Kpd. Masy.*, vol. 2, no. 1, pp. 43–48, 2023, doi: 10.47926/sjpkm.2023.2.1.43-48.

[12] I. Il Sanuriza *et al.*, “First report of Pyricularia oryzae, the cause of blast disease in upland rice, in Lombok, West Nusa Tenggara, Indonesia,” *Biodiversitas*, vol. 25, no. 2, pp. 683–689, 2024, doi: 10.13057/biodiv/d250227.

[13] F. Firmansyah, K. Khaerana, and E. A. Sidik, “Hubungan Skor Penyakit Tungro terhadap Kehilangan Komponen Hasil Padi,” *AGROSAINSTEK J. Ilmu dan Teknol. Pertan.*, vol. 7, no. 1, pp. 17–24, 2023, doi: 10.33019/agrosainstek.v7i1.315.

[14] A. Syarifah, A. A. Riadi, and A. Susanto, “Klasifikasi Tingkat Kematangan Jambu Bol Berbasis Pengolahan Citra Digital Menggunakan Metode K-Nearest Neighbor,” vol. 7, no. 1, pp. 27–35, 2022.

[15] J. Khatib, S. Dalam, M. K. Neighbor, and K. N. N. Berbasis, “Indonesian Journal of Computer Science,” vol. 12, no. 1, pp. 656–664, 2023.

[16] S. Ani, M. Furqan, and R. S. TP. Bolon, “Deteksi Tepi Pola Tulisan Arab Menggunakan Metode Canny pada Nisan Kuno di Sumatera Utara,” *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 1, p. 86, 2023, doi: 10.53513/jsk.v6i1.7385.

[17] M. Afriansyah, Joni Saputra, V. Y. P. Ardhana, and Yuan Sa’adati, “Algoritma Naive Bayes Yang Efisien Untuk Klasifikasi Buah Pisang Raja Berdasarkan Fitur Warna,” *J. Inf. Syst. Manag. Digit. Bus.*, vol. 1, no. 2, pp. 236–248, 2024, doi: 10.59407/jismdb.v1i2.438.

[18] V. Feriska Amalia and R. Rahma Dewi, “Penilaian Kesegaran Ikan Dengan Metode *K-Nearest Neighbor* Dan Pengolahan Citra Digital,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp. 7823–7829, 2024, doi: 10.36040/jati.v8i4.10441.

[19] A. I. Masruroh, . S., and A. Syauqi, “Klasifikasi Tingkat Kematangan Buah Pepaya California Dalam Ruang Warna Hsv (Hue Saturation Value) Dengan Algoritma K-Nearest Neighbors,” *J. Inform. dan Ris.*, vol. 1, no. 1, pp. 9–14, 2023, doi: 10.36308/iris.v1i1.470.

[20] Y. G. Lestari and H. Irsyad, “Penggunaan Metode SVM Dengan Fitur HSV HOG Dalam Mengklasifikasi Jenis Ikan Guppy,” *J. Algoritm.*, vol. 4, no. 1, pp. 21–30, 2023, doi: 10.35957/algoritme.xxxx.

[21] M. A. Lutfia, F. X. A. Setyawan, S. Alam, T. Yulianti, and H. Fitriawan, “Implementasi Ekstraksi Fitur Menggunakan Gray Level Co-Occurrence Matrices (GLCM) dan *K-Nearest Neighbor* (K-NN) Untuk Klasifikasi Jenis Kain Dasar,” *Semin. Nas. Tek. Elek*, pp. 3–8, 2023.

[22] Y. A. Prasaja, “Perbandingan Metode Glcm Dan Lbp Dalam Klasifikasi Jenis Kayu,” *Indexia*, vol. 4, no. 2, p. 61, 2022, doi: 10.30587/indexia.v4i2.4292.

[23] M. Furqon, S. Sriani, and L. S. Harahap, “Klasifikasi Daun Bugenvil Menggunakan Gray Level Co-Occurrence Matrix Dan K-Nearest Neighbor,” *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 6, no. 1, p. 22, 2020, doi: 10.24014/coreit.v6i1.9296.

[24] I. M. Ade Prayoga, G. Indrawan, and D. G. Hendra Divayana, “Pengelompokan Laras Suara Berdasarkan Pepatutan Atau Pathet Gamelan Bali Menggunakan Klasifikasi *K-Nearest Neighbor* Dan Support Vector Machine,” *Technomedia J.*, vol. 8, no. 2SP, pp. 151–161, 2023, doi: 10.33050/tmj.v8i2sp.2011.

[25] A. Rizal Efendi, I. N. Farida, M. A. Dusea, and W. Dara, “Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 954 Penerapan Metode KNN Untuk Mendeteksi Hama dan Penyakit Pada Tanaman Mangga,” *Agustus*, vol. 7, pp. 2549–7952, 2023.

[26] M. Furqan, Y. R. Nasution, and A. N. Siregar, “Penerapan Sistem Pakar Diagnosis Peradangan Pulpa Gigi Dengan Metode Certainty Factor,” *Technol. J. Ilm.*, vol. 14, no. 2, p. 152, 2023, doi: 10.31602/tji.v14i2.10448.

[27] Y. R. Nasution and A. Raja, “PENERAPAN METODE SIMPLE MULTI ATRIBUTE RATING TEHNIQUE DAN ALGORITMA K- NEAREST NEIGHBOR Prodi Ilmu Komputer , Fakultas Sains Dan Teknologi , Universitas Islam Negeri Sumatera Utara,” *J. Sci. Soc. Res.*, vol. 4307, no. February, pp. 61–65, 2021.