Robusta Coffee Leaf Disease Classifications Using SVM Method and GLCM Feature Extraction

Agus Supriyanto[1], R. Rizal Isnanto[2], Oky Dwi Nurhayati[3]

[1] Department of Electrical Engineering, Faculty of Engineering, Universitas Diponegoro, Jl. Prof. Soedarto, Tembalang, Kec. Tembalang, Semarang 50275 INDONESIA, (tel.: 024-7460057; email: agussupriyanto.elektro@gmail.com)
[2,3] Department of Computer Engineering, Faculty of Engineering, Universitas Diponegoro, Jl. Prof. Soedarto, Tembalang, Kec. Tembalang, Semarang 50275 INDONESIA (tel.: 024-76480609; email: [2]rizal_isnanto@yahoo.com, [3]okydwin@gmail.com)

[Received: 20 July 2023, Revised: 12 September 2023]
Corresponding Author: Agus Supriyanto

ABSTRACT — Many farmers in Indonesia derive their income from coffee plants, which also play a crucial role in the country's foreign exchange earnings. However, coffee plant production may decrease due to pests and disease attacks. Leaf diseases, such as leaf spot (*Cercospora coffeicola*) and leaf rust (*Hemileia vastatrix*), are among the most common diseases to occur in coffee plants. This research seeks to identify leaf diseases in robusta coffee leaves and determine the classification. The application of machine learning-based image processing using the support vector machine (SVM) classification method based on the gray-level co-occurrence matrix (GLCM) feature extraction can be the proposed solution. The preprocessing must precede the processing stage for easier analysis of the image's quality. Then, the k-means clustering segmentation process was conducted to distinguish leaf parts affected by leaf spot and rust from those unaffected. The GLCM method was employed as the feature extraction based on the angular second moment (ASM) or energy features, contrasts, correlations, inverse different moment (IDM) or homogeneities, and entropy with angles of 0°, 45°, 90°, and 135°, as well as inter-pixel distances of 1 until 3. The classification was done with the SVM method using the linear, polynomial, and radial basis function (RBF) Gaussian kernels. This research used leaf spot and rust images, with training and test data of 320 and 80 images, respectively. The RBF Gaussian achieved the best test results with the best accuracy of 97.5%, precision of 95.24%, recall of 100%, and $F1$-score of 97.56%.

KEYWORDS — Robusta Coffee Leave, Leaf Rust Diseases, Leaf Spot Diseases, SVM, GLCM.

I. INTRODUCTION

Coffee is a tree-shaped plant belonging to the *Rubiaceae* family and the genus *Coffea* [1], [2]. The genus *Coffea* has about a hundred types, but only two species have high commercial value, especially robusta and arabica coffees. Other types of coffee, such as excelsa and liberica coffee beans, are only used as a mixture to enhance the aroma [3]. Coffee is one of the most widely consumed drinks in the world, so it is a foodstuff that is quite relevant from the economic perspective [4].

Coffee is a leading commodity that contributes to foreign exchange earnings, provides income for farmers, produces industrial commodities, stimulates job creation, and drives regional development. Indonesia is the world's sixth-largest coffee producer after Brazil, Vietnam, Colombia, Honduras, and India. It is also the second-largest coffee producer in Southeast Asia. The six countries export 73.7% of the world's coffee, with Brazil accounting for 29.1%, Vietnam 20.5%, Colombia 10.5%, Honduras 5.3%, India 4.7%, and Indonesia 3.6% [5], [6]. Coffee production in Indonesia has declined due to farmers' limited knowledge of various diseases and pests attacking coffee plants.

Plant disease is a condition in which symptoms appear when plant tissues and cells stop functioning normally as a result of constant pathogens or the environmental interferences [7]. The diseases that attack coffee plants include leaf spot, leaf rust, coffee rot, and fungal upas. Leaf spot (*Cercospora coffeicola*) and leaf rust (*Hemileia vastatrix*) diseases are two plant diseases that attack coffee leaves. These diseases may reduce coffee productivity and cause crop failures and plant death. In addition, the farmers' limited knowledge of the coffee plant disease impacts leads to crop failures, which is detrimental and unsettling to coffee farmers. Coffee leaf spot disease caused by the fungus *Cercospora coffeicola* is known as brown-eye spot and is widespread not only in Indonesia, but also worldwide. Round, concentric, reddish-brown, or dark brown spot indicate this diseases attack on the leaves. Humid weather can aggravate leaf spot, which then cause leaf drop [6], [8]. Leaf rust disease is caused by the fungus *Hemileia vastatrix*, which infects the genus *Coffea* and is more severe in arabica and robusta coffee [9]. Symptoms of leaf rust disease can be seen by the presence of orange-colored spots on both sides of the leaf. Affected leaves exhibit brown spots that then turn yellow [6].

The advancement of technology can affect all aspects of life, including farming. The employment of technology in farming, such as image processing to detect diseases in coffee leaves, is necessitated. Image processing involves numerous fields, including mathematics, physics, electronics, photography, arts, and computer technology. Hence, it plays an essential role in this research. Computer vision and image processing are interrelated. The main objectives of computer vision are object detection, segmentation, and classification [10].

Farmers can manually identify and classify diseases in coffee leaves. However, this practice is not that effective since morphological characteristics of the leaf diseases, such as shape, texture, and color, cannot be distinguished. Research employing image processing has been conducted to resolve this problem. In earlier research, a fuzzy k-nearest neighbor (FK-NN) method was used to diagnose diseases in the arabica coffee plants and resulted in an accuracy level of 80% [11]. Other research applied the web-based breadth-first search (BFS) method to identify pests and diseases in coffee plants, with an accuracy of 83.39% [12]. Using the Euclidean distance and Hough transform to identify brown eye spot diseases in coffee leaves resulted in an accuracy of 55% and 50% for the arabica

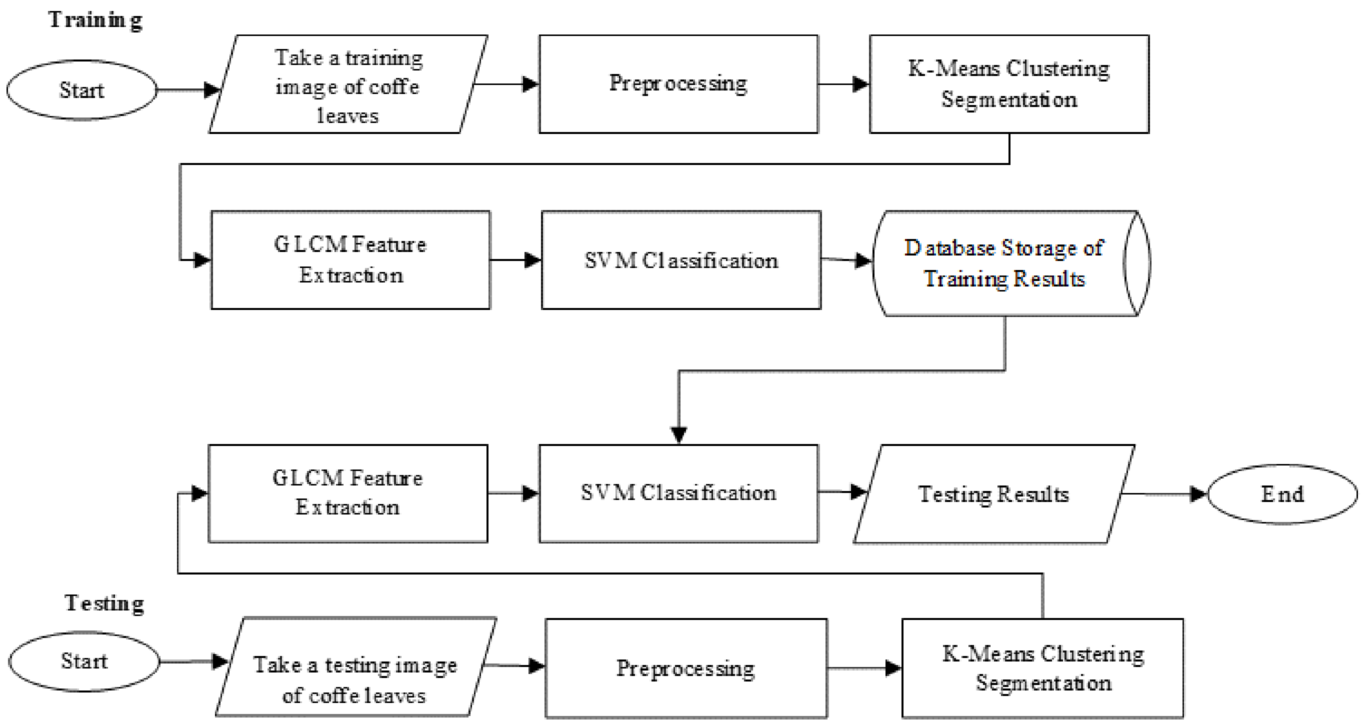


Figure 1. System design flowchart.

and the robusta coffee leaves, respectively [2]. In coffee plant leaf images, edge detection using the Laplacian of Gaussian method produced an average mean square error value of 237.629 pixels [13]. The application of expert systems and location-based services for disease detection in coffee plants using decision tree classification has also been studied. This method produced an accuracy rate of 85% [14]. Then, the study of potato leaf disease detection using the support vector machine (SVM) method based on texture features and color features yielded an average accuracy of 80% [15]. Classification of clove leaves using particle swarm optimization-support vector machine (PSO-SVM) and gray-level co-occurrence matrix (GLCM) to determine the leaf surface produced an accuracy rate of 90.5% [16]. An accuracy value of 96.8% was obtained in a study that used the SVM method for classifying and a convolutional neural network (CNN) for extracting disease characteristics in rice leaves [17].

The classification accuracy achieved using the GLCM and SVM methods in previous studies was greater than 80%. This research employed the SVM method for classification and the GLCM method for feature extraction. The research began with the acquisition of image data to obtain digital images in the form of robusta coffee leaves. Preprocessing was used to increase image contrast to get new and better RGB values. Segmentation with k-means clustering was used to distinguish parts of leaves affected by the disease from those unaffected. Texture feature extraction was performed using the GLCM process, yielding angular second moment (ASM) or energy, contrast, correlation, inverse different moment (IDM) or homogeneity, and entropy values. The SVM was used in the final stage of classification to determine robusta coffee leaf disease. This process was computer-processed using MATLAB software.

II. METHODOLOGY

The proposed methods presented in Figure 1 include image acquisition/taking images, preprocessing using contrast stretch, segmentation of k-means clustering, extraction of the GLCM feature, and classification of the SVM. The system design, both training and testing, shares similar process flows. The flow diagram of the system design in the form of training and testing is outlined in Figure 1, describing the flow of the conducted research.

A. IMAGE ACQUISITION

The image acquisition process is the initial step in capturing or obtaining digital images using devices or certain additional devices, for which this research used a digital scanner. Images of robusta coffee leaves used in this study had 300 dpi resolution and were in JPEG format (*.jpg extension). The image samples were collected from robusta coffee plantation in the Plaosan Village, Cluwak Subdistrict, Pati Regency, Central Java. The original images taken were leaf spot and rust. The obtained data were divided into two parts, namely training data and test data, using the splitting method with a comparison of 80:20 [18]. A total of 320 training data and 80 test data were obtained. The sample data of coffee leaves with leaf spot and rust amounted to 200 leaves, in which each leaf disease type consisting of 160 training data and 40 test data. Figure 2 shows two coffee leaf samples.

B. PREPROCESSING

Preprocessing was employed to improve the image quality for easier process and analysis. The process of enhancing contrast to expand image differences was performed to obtain another RGB value with better differentiation. Significant image differences can expand the variation of the objects' sharpness in the images, and clearly visible images can help the image segmentation process. The difference in image pixels with the highest and the lowest intensity values can be used to determine contrast. This research utilized contrast stretching by increasing the intensity value to obtain clearer images [19], [20]. The contrast enhancement process was performed using the MATLAB program. The original images were extracted in each RGB component, then contrast stretching was done to get

(a) (b)

Figure 2. Images of robusta coffee leaves, (a) leaf spot disease, (b) rust disease leaf.

(a) (b)

Figure 3. Preprocessed results, (a) RGB image of leaf rust, (b) image of leaf rust after increasing the contrast.

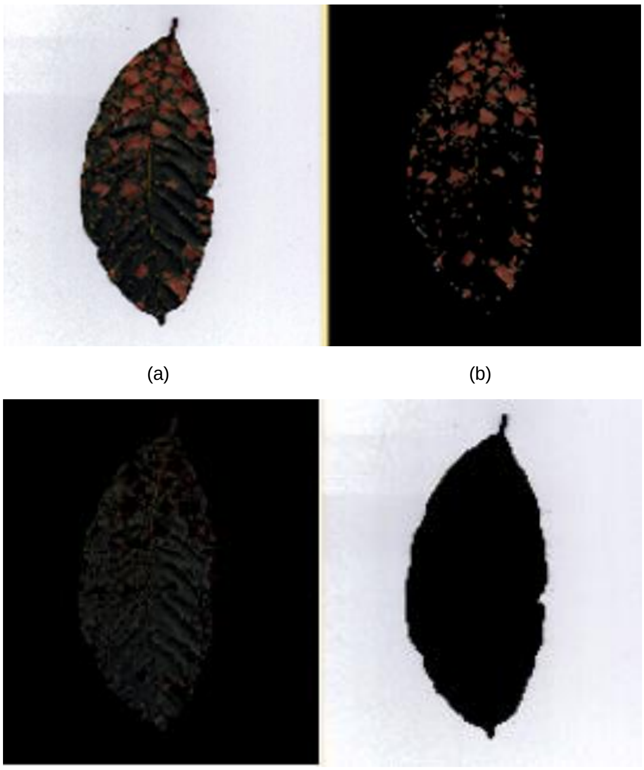
better image quality using the imadjust function. The results of stretching the contrast of the leaf spot image are depicted in Figure 3.

C. K-MEANS CLUSTERING SEGMENTATION

Segmentation partitions a region into several segments to make it easier to analyze. Images are divided into three clusters using k-means clustering, with images located in the main area of the region affected in at least one of the clusters [21]. Figure 1 shows the segmentation using k-means clustering to select which of the three clusters has more apparent disease. Figure 4(a) is the original images after the preprocessing stage, while Figure 4(b) until Figure 4(d) depict the results of clusters 1 to 3 from the k-means clustering segmentation process.

D. FEATURE EXTRACTION OF THE GRAY-LEVEL CO-OCCURRENCE MATRIX

In the subsequent process, the segmentation results were extracted to obtain information on area affected by the disease. The GLCM method was utilized for the feature extraction. In the texture analysis, the GLCM is a statistical method to extract features. The GLCM calculates the pixel frequency with grayscale intensity values horizontally adjacent to the pixel with a j value [22]. The number of times a pixel value level is adjacent to another at a given distance (d) and angular direction (θ) is known as co-occurrence. Pixels represent distance, while degrees represent orientation. With an interval of 45°, the orientation is formed in four angular directions, including θ = 0°, θ = 45°, θ = 90°, and θ = 135° [23], [24]. Texture features



(a) (b)

(a) (b)

Figure 4. K-Means clustering segmentation results, (a) leaf rust image, (b) cluster 1, (c) cluster 2, (d) cluster 3

in this research were calculated from the resulting GLCM, such as ASM or energy, contrast, correlation, IDM or homogeneity, and entropy. The extraction features of texture characteristics are as follows [25]–[27].

1) *ANGULAR SECOND MOMENT (ENERGY/UNIFORMITY)*

The ASM or energy is useful for measuring the gray intensity of an image in the GLCM matrix or texture uniformity. When the intensity variance of images decreases, the ASM value increases. Equation (1) is used to calculate the ASM value.

$$f_1 = \sum_i \sum_j \{p(i,j)\}^2. \quad (1)$$

2) *CONTRAST/INERTIA*

Contrast represents an image matrix spread measure or moment of inertia. The further away the contrast is from the main diagonal, the greater the contrast values. The contrast value is a visual indicator of the difference in gray levels between image areas. Equation (2) is used to determine the contrast value.

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \begin{matrix} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \\ |i-j| = n \end{matrix} \right\}. \quad (2)$$

3) *CORRELATION*

Correlation represents the measure of the linear dependence between the degree value of gray images. Equation (3) is used to determine the correlation value.

$$f_3 = \frac{\sum_i \sum_j (i.j) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}. \quad (3)$$

4) *INVERSE DIFFERENCE MOMENT/HOMOGENEITY*

The IDM is the feature indicating image homogeneity in the co-occurrence matrix with the same gray degree. In multiple coordinates, if a pair of pixels meets the requirements of the co-occurrence probability matrix, the energy value will increase. In contrast, if they are dispersed, the energy value will decrease. The result is a homogeneous image with a high IDM value. The equation used to determine the IDM value is shown in (4).

$$f_4 = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j). \quad (4)$$

5) *ENTROPY*

The intensity distribution irregularity of the images' gray level in the co-occurrence matrix can be measured using the entropy. The display will be good if the relative values of the GLCM elements are the same. In contrast, the display will be poor when the values of the GLCM elements are close to 0 or 1. It indicates that the gray transition is small, so is the change. The entropy value is calculated using (5).

$$f_5 = -\sum_i \sum_j p(i,j) \log(p(i,j)). \quad (5)$$

In (1) until (5), $p(i,j)$ and (i,j) is the input in the spatial dependence matrix of the gray level being normalized, $P(i,j)/R$. In $P_x(i)$ value, (i) is inputted to the low probability matrix obtained by adding up rows from $P(i,j) = \sum_{j=1}^{N_g} P(i,j)$. Value of n is the number of gray levels in the images, while N_g is the number of different gray levels in each quantified image Σ_i, Σ_j, $\Sigma_{i=1}^{N_g}$, and $\Sigma_{j=1}^{N_g}$. Value of $\mu_x \mu_y$ is the average of the column elements in the image matrix, while the value of $\sigma_x \sigma_y$ is the standard deviation of the matrix column.

The feature extraction in this research was calculated from the resulting GLCM, such as the ASM or energy, contrast, correlation, IDM or homogeneity, and entropy. Four directions of the texture feature formation are 0°, 45°, 90°, and 135°, with each having an interval of 45°. The GLCM was determined by measuring the inter-pixel distance (d) = 1, 2, and 3.

E. SUPPORT VECTOR MACHINE CLASSIFICATION

Following the feature extraction, the SVM method was used for classification. The SVM is supervised learning using algorithms which is able to analyze data and identify patterns to provide high-quality support for the hyperplane in the dimensional space [28]. This method was used in the regression analysis and classification to identify diseases in the coffee plant. For the nonlinearity, the notion of the kernel trick in the high-dimensional workspace can be included in the future development of the SVM. The fundamental notion of the SVM is linear classification. Several kernel functions can be utilized for the nonlinearity. The SVM learning can be easier when using the kernel trick. The SVM classification has several kernel functions that are often used, including the following [29]–[31].

1) *LINEAR KERNEL*

Of all the kernel functions, the linear kernel is the simplest. In the case of text classification, this kernel is often used. To determine the linear kernel, (6) is used.

$$K(x_i, x_j) = x_i . x_j. \quad (6)$$

2) POLYNOMIAL *KERNEL*

The kernel that is often used to classify images is the polynomial kernel. Equation (7) is used to determine the polynomial kernel.

$$K(x_i, x_j) = (x_i . x_j + c)^d. \quad (7)$$

TABLE I
RESEARCH SCENARIO

Scenario	K-Means Clustering	GLCM
	Cluster	***Distance (Pixels)***
1, 10, 19	1	1
2, 11, 20	1	2
3, 12, 21	1	3
4, 13, 22	2	1
5, 14, 23	2	2
6, 15, 24	2	3
7, 16, 25	3	1
8, 17, 26	3	2
9, 18, 27	3	3

3) *RADIAL BASIS FUNCTION* (RBF) GAUSSIAN KERNEL

The RBF kernel, which is the standard kernel for valid (available) data, is the one used as the SVM tool by default. Equation (8) specifies the RBF kernel.

$$K(x_i, x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \quad (8)$$

In (6) until (8), $K(x_i, x_j)$ is the function kernel, while x_i, x_j value is a pair of two data from the entire training dataset. Value of c, d, σ is the constant and $\|x_i - x_j\|$ is the square of the distance between the vectors x_i and x_j.

This research utilized kernel in the SVM method for the classification system. Kernels used included linear, polynomial, and RBF Gaussian kernels. Leaf spot and rust in the coffee leaves are the classification results.

F. EVALUATION OF CLASSIFICATION SUCCESS LEVEL

The classification success rate of a machine learning algorithm can be determined using a confusion matrix containing information on actual and predicted classification results. Accuracy represents the number of correctly classified cases divided by the total amount of data. Accuracy is calculated using (9). The higher the classification accuracy, the better the performance of the classification technique. Precision and recall are used as a measure of how precise and complete the classification results are; they are calculated using (10) and (11). $F1$-score is the harmonic mean of precision and recall, calculated using (12) [32], [33].

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \quad (9)$$

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision+Recall}. \quad (12)$$

Evaluation results depends on true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values.

III. RESULT AND DISCUSSION

System testing was performed by testing images that had been obtained previously. These images were collected using a digital scanner to obtain the exact distance between one image and another. Testing was done 27 times with three SVM classification kernels. Table I displays three k-means clustering segmentation clusters and three inter-pixel distances on the GLCM feature extraction. Scenarios 1 to 9 test SVM classification using a linear kernel with three k-means

TABLE II
EVALUATION OF LINEAR SVM TESTING RESULTS

Scenario	Linear SVM Testing Results			
	Accuracy	*Precision*	*Recall*	*F1-Score*
1	52.50%	52.94%	45.00%	48.65%
2	53.75%	56.00%	35.00%	43.08%
3	65.00%	71.43%	50.00%	58.82%
4	70.00%	66.67%	80.00%	72.73%
5	41.25%	40.54%	37.50%	38.96%
6	52.50%	52.38%	55.00%	53.66%
7	68.75%	63.64%	87.50%	73.68%
8	68.75%	62.30%	95.00%	75.25%
9	58.75%	100.00%	17.50%	29.79%

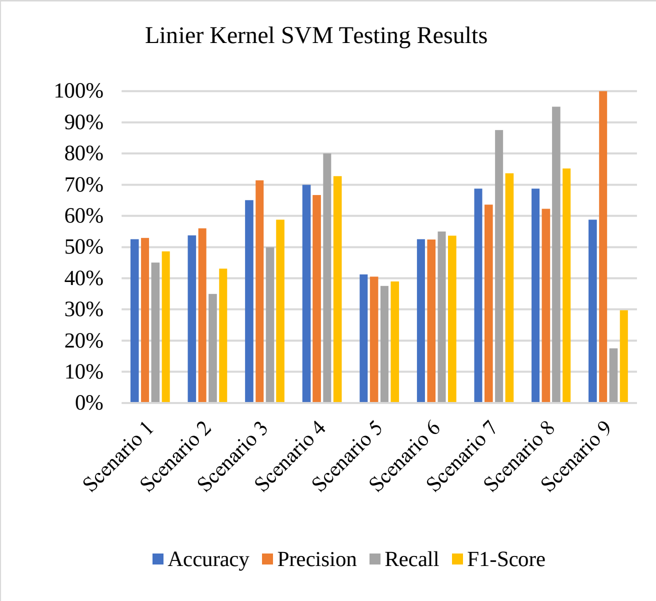


Figure 5. Graph of the evaluation results of the linier kernel SVM testing.

TABLE III
EVALUATION OF POLYNOMIAL SVM TESTING RESULTS

Scenario	Polynomial SVM Testing Results			
	Accuracy	*Precision*	*Recall*	*F1-Score*
10	72.50%	100.00%	45.00%	62.07%
11	76.25%	95.65%	55.00%	69.84%
12	82.50%	93.33%	70.00%	80.00%
13	46.25%	0.00%	0.00%	0.00%
14	50.00%	0.00%	0.00%	0.00%
15	86.25%	96.77%	75.00%	84.51%
16	50.00%	0.00%	0.00%	0.00%
17	50.00%	0.00%	0.00%	0.00%
18	50.00%	0.00%	0.00%	0.00%

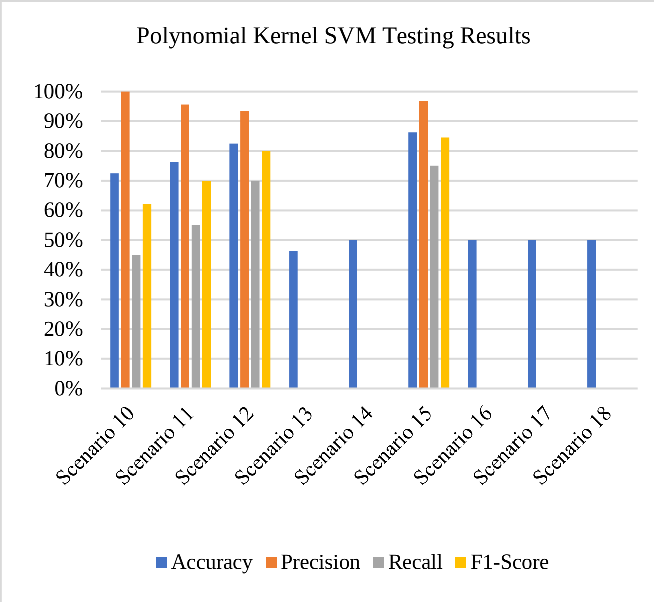


Figure 6. Graph of the evaluation results of the polynomial kernel SVM testing.

clustering segmentation clusters and three inter-pixel distances on the GLCM feature extraction. Scenarios 10 to 18 test SVM classification using a polynomial kernel with three k-means clustering segmentation clusters and three GLCM feature extraction pixel distances. Scenarios 19 to 27 test SVM classification using the RBF Gaussian kernel with three k-means clustering segmentation clusters and three GLCM feature extraction pixel distances.

A. DISCUSSION AND EVALUATION OF THE SVM TESTING RESULTS WITH THE LINEAR SVM KERNEL

There are multiple factors affecting accuracy, precision, recall, and $F1$-score from the SVM classification testing using the linear kernel. These influencing factors include the clusters in the k-means clustering segmentation process and the inter-pixel distances in the GLCM feature extraction. In this research, three clusters and three inter-pixel distances were used. The test results were obtained from the resulting GLCM feature extraction values, such as ASM or energy, contrast, correlation, IDM or homogeneity, and entropy with angles of 0°, 45°, 90°, 135°, and average angle. The test findings obtained from the evaluation of two types of leaf diseases, specifically leaf spot and leaf rust, utilizing the linear kernel, are presented in Table II.

The SVM classification testing using the linear kernel in scenarios 1 up to 9 is shown in Table II and Figure 5. Scenario 4 achieved the highest accuracy of 70%, scenario 9 achieved the highest precision of 100%, and scenario 8 achieved the highest recall of 95%, as well as the highest $F1$-score of 75.25%. The testing diagram in Figure 5 shows that the accuracy value increases as the inter-pixel distance increases, which occurred in cluster 1. However, the accuracy values in clusters 2 and 3 were not stable when the inter-pixel distance was increased. The unstable accuracy value is due to the random segmentation results, so the three image clusters with detected and undetected disease areas appear randomly. The precision and recall values were dependent on the initial classification result, which was the leaf spot. The more accurate the classifications, the higher the values. The precision and recall results determine the $F1$-score value. The higher the precision and recall values, the greater the $F1$-score value.

B. DISCUSSION AND EVALUATION OF THE SVM TESTING WITH THE POLYNOMIAL KERNEL

There are multiple factors affecting accuracy, precision, recall, and $F1$-score from the SVM classification testing using the polynomial kernel. These influencing factors include the clusters in the k-means clustering segmentation process and the inter-pixel distances in the GLCM feature extraction. This research used three clusters and three inter-pixel distances. The test results were obtained from the resulting GLCM feature extraction values, such as ASM or energy, contrast, correlation, IDM or homogeneity, and entropy with angles of 0°, 45°, 90°, 135°, and average angle. The test findings obtained from the evaluation of two types of leaf diseases, specifically leaf spot

TABLE IV
EVALUATION OF GAUSSIAN RBF SVM RESULTS

Scenario	Gaussian RBF SVM Testing Results			
	Accuracy	*Precision*	*Recall*	*F1-Score*
19	65.00%	75.00%	45.00%	56.25%
20	75.00%	85.71%	60.00%	70.59%
21	77.50%	82.35%	70.00%	75.68%
22	87.50%	91.67%	82.50%	86.84%
23	92.50%	92.50%	92.50%	92.50%
24	97.50%	95.24%	100.00%	97.56%
25	81.25%	82.05%	80.00%	81.01%
26	81.25%	80.49%	82.50%	81.48%
27	77.50%	76.19%	80.00%	78.05%

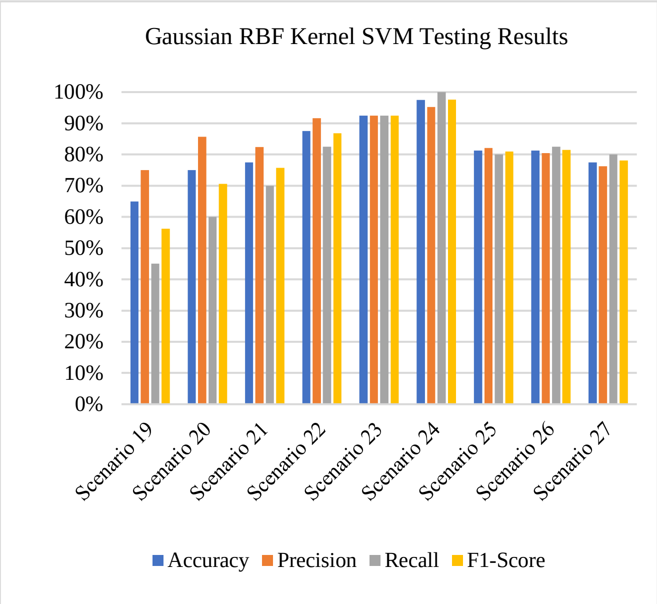


Figure 7. Graph of the evaluation results of the Gaussian RBF kernel SVM testing.

and leaf rust, utilizing the polynomial kernel, are presented in Table III.

The SVM classification testing using the polynomial kernel in scenarios 10 up to 18 is shown in Table III and Figure 6. Scenario 15 achieved the highest accuracy of 86.25%, scenario 10 achieved the highest precision of 100%, and scenario 15 achieved the highest recall of 75%, as well as the highest $F1$-score of 84.51%. The testing diagram in Figure 6 shows that the accuracy value increases as the inter-pixel distance increases, which occurred in clusters 1 and 2. However, the accuracy value in cluster 3 was stable when the inter-pixel distance was increased. The unstable accuracy value is due to the random segmentation results, so the three image clusters with detected and undetected disease areas appear randomly. The precision and recall values were dependent on the initial classification result, which was the leaf spot. Cluster 3 resulted in a value of 0 since the classification results of the leaf spot were all incorrect. The higher the precision and recall values, the better the $F1$-score, and vice versa.

C. DISCUSSION AND EVALUATION OF THE SVM TESTING WITH THE RGB GAUSSIAN

There are multiple factors affecting accuracy, precision, recall, and $F1$-score from the SVM classification testing using the RBF Gaussian kernel. These influencing factors include the clusters in the k-means clustering segmentation process and the

TABLE V
BEST SVM PERFORMANCE RESULTS

No	Kernel	Accuracy	Precision	Recall	F1-Score
1	Linear	70.00%	66.67%	80.00%	72.73%
2	Polynomial	86.25%	96.77%	75.00%	84.51%
3	Gaussian RBF	97.50%	95.24%	100.00%	97.56%

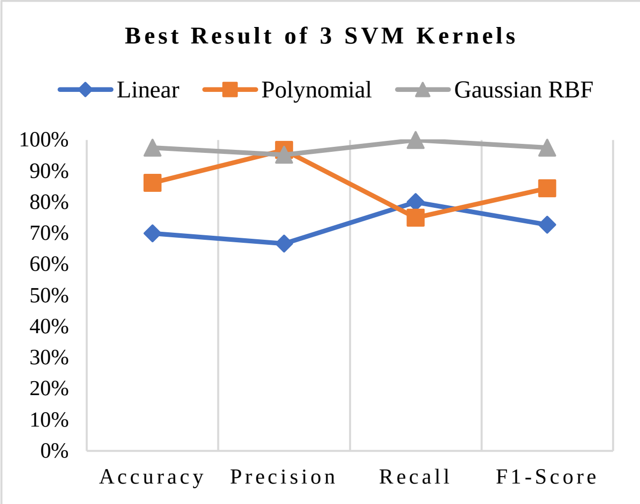


Figure 8. Graph of the best evaluation results of the SVM testing.

inter-pixel distances in the GLCM feature extraction. This research used three clusters and three inter-pixel distances. The test results were obtained from the resulting GLCM feature extraction values, such as ASM or energy, contrast, correlation, IDM or homogeneity, and entropy with angles of 0°, 45°, 90°, 135°, and average angle. The test findings obtained from the evaluation of two types of leaf diseases, specifically leaf spot and leaf rust, utilizing the RBF Gaussian kernel, are presented in Table IV.

The SVM classification testing using the RBF Gaussian kernel in scenarios 19 up to 27 is shown in Table IV and Figure 7. Scenario 24 achieved the highest accuracy, precision, recall, and $F1$-score of 95.24%, 100%, 100%, and 97.56%, respectively. The testing diagram in Figure 7 shows that the accuracy value increases as the inter-pixel distance increases, which occurred in clusters 1 and 2. However, the accuracy values in cluster 3 decreased when the inter-pixel distance was increased. The unstable accuracy value is due to the random segmentation results, so the three image clusters with detected and undetected disease areas appear randomly. The precision and recall values were dependent on the initial classification result, which was the leaf spot. The more accurate the classifications, the higher the precision and recall values. The precision and recall results affect the $F1$-score value. The higher the precision and recall values, the greater the $F1$-score value, and vice versa.

D. PERFORMANCE RESULTS OF THE KERNELS

The results of the classification performance of robusta coffee leaf disease types in the form of leaf spot and leaf rust using k-means clustering segmentation were divided into three clusters, which could detect leaf parts getting leaf spot and leaf rust diseases. ASM or energy, contrast, correlation, IDM or homogeneity, and entropy are the GLCM parameters used in this research. The four angles of 0°, 45°, 90°, and 135°, with inter-pixel distances of 1, 2, and 3, were used to form those parameters. The SVM method using the linear, polynomial, and

RBF Gaussian kernels was used for system classification. The test results for the three kernels exhibiting the highest performance are presented in Table V.

Figure 8 shows the results of testing accuracy values with a combination of clusters from k-means clustering, GLCM parameters, and the proposed SVM kernels. The RBF Gaussian kernel yielded the best performance among the other kernels. The polynomial kernel obtained the best precision value test based on accuracy, while the RBF Gaussian kernel obtained the best recall and $F1$-score. Kernels in SVM classification also affect the level of accuracy obtained during research.

IV. CONCLUSION

SVM classification of robusta coffee leaf disease based on GLCM feature extraction has been conducted. The segmentation process was done using k-means clustering with three clusters. GLCM feature extraction used ASM or energy, contrast, correlation, IDM or homogeneity, and entropy features with angles of 0°, 45°, 90°, 135°, and average angle, and inter-pixel distances of 1 to 3. Linear, polynomial, and RBF Gaussian kernels were used as the SVM classification method. The best test results of leaf spots and rust classification on robusta coffee were obtained with the RBF Gaussian kernel. The highest accuracy was 97.5%, precision was 95.24%, recall was 100%, and $F1$-score was 97.56%. The use of kernels in the SVM method is very influential in the classification process. Of the three kernels used during the research, namely linear, polynomial, and RBF Gaussian, the highest accuracy value was obtained in testing using the RBF Gaussian kernel.

However, this research still has shortcomings, one of which is in image preprocessing. Further research is required to acquire a preprocessing model and recognize the specific characteristics of leaf spot and leaf rust more precisely. Therefore, the classification results have higher accuracy. It is necessary to test robusta coffee leaf disease with other methods, such as deep learning. Then, the research is compared with this research to get the best method.

CONFLICT OF INTEREST

Authors declare no conflict interest.

AUTHOR CONTRIBUTION

Conceptualization, Agus Supriyanto; methodology, Agus Supriyanto; software, Agus Supriyanto; validation, Agus Supriyanto; formal analysis, R. Rizal Isnanto and Oky Dwi Nurhayati; resources, Agus Supriyanto; data curation, R. Rizal Isnanto and Oky Dwi Nurhayati; writing, Agus Supriyanto; funding acquisition, R. Rizal Isnanto and Oky Dwi Nurhayati.

REFERENCES

[1] M. Rizwan, *Budidaya Kopi*. West Pasaman, West Sumatra: CV. Azka Pustaka, 2022.

[2] Y. Defitri, "Pengamatan Beberapa Penyakit yang Menyerang Tanaman Kopi (Coffea Sp) di Desa Mekar Jaya Kecamatan Betara Kabupaten Tanjung Jabung Barat," *J. Media Pertan.*, Vol. 1, No. 2, pp. 78–84, Oct. 2016, doi: 10.33087/jagro.v1i2.19.

[3] I. Fibriani, Widjonarko, C.S. Sarwono, and F. Dwika, "Deteksi Penyakit Brown Eye Spot pada Daun Kopi Menggunakan Metode Euclidean Distance dan Hough Transform," *J. JEETech*, Vol. 1, No. 1, pp. 44–49, May 2020, doi: 10.48056/jeetech.v1i2.120.

[4] A.S. Franca and L.S. Oliveira, "Coffee," in *Integrated Processing Technologies for Food and Agricultural By-Products*, Z. Pan, R. Zhang, dan S. Zicari, Eds., Cambridge, MA, USA: Academic Press, 2019, pp. 413–438, doi: 10.1016/B978-0-12-814138-0.00017-4

[5] International Coffee Organization, "Total Production by All Exporting Countries." Distributed by International Coffee Organization, https://www.ico.org/historical/1990%20onwards/PDF/1a-total-production.pdf

[6] R. Harni *et al*., *Teknologi Pengendalian Hama dan Penyakit Tanaman Kopi*, ed. 2. Jakarta, Indonesia: IAARD Press, 2018.

[7] U.D. Rosiani, C. Rahmad, M.A. Rahmawati, and F. Tupamahu, "Segmentasi Berbasis K-Means pada Deteksi Citra Penyakit Daun Tanaman Jagung," *J. Inform. Polinema*, Vol. 6, No. 3, pp. 37–42, May 2020, doi: 10.33795/jip.v6i3.331.

[8] E.P. Ramdan *et al*., *Penyakit Tanaman dan Pengendaliannya*. Medan, Sumatera Utara: Yayasan Kita Menulis, 2021.

[9] N.E.T. Castillo *et al*., "Impact of Climate Change and Early Development of Coffee Rust – An Overview of Control Strategies to Preserve Organic Cultivars in Mexico," *Sci. Total Environ.*, Vol. 738, pp. 1–14, Oct. 2020, doi: 10.1016/j.scitotenv.2020.140225.

[10] W. Li *et al*., "Intelligent Metasurface System for Automatic Tracking of Moving Targets and Wireless Communications Based on Computer Vision," *Nat. Commun.*, Vol. 14, pp. 1–10, Feb. 2023, doi: 10.1038/s41467-023-36645-3.

[11] A.Y.P. Putri, "Pemodelan Sistem Pakar Diagnosa Penyakit Tanaman Kopi Arabika Dengan Metode Fuzzy K-Nearest Neighbor (FK-NN)," Skripsi, Universitas Brawijaya, Malang, Indonesia, 2015.

[12] F.R. Lumbanraja, S. Rosdiana, H. Sudarsono, and A. Junaidi, "Sistem Pakar Diagnosis Hama dan Penyakit Tanaman Kopi Menggunkan Metode Breadth First Search (Bfs) Berbasis Web," *Explore J. Sist. Inf., Telemat.*, Vol. 11, No. 1, pp. 1–9, Jun. 2020, doi: 10.36448/jsit.v11i1.1452.

[13] T.S. Prihartini and P.N. Andono, "Deteksi Tepi dengan Metode Laplacian of Gaussian pada Citra Daun Tanaman Kopi," Skripsi, Universitas Dian Nuswantoro, Semarang, Indonesia, 2015.

[14] W.A. Nugraha, M. Lestari, M. Yasin, and D. Suhartono, "Perancangan Sistem Pakar Pendeteksi Penyakit pada Tanaman Kopi dengan Layanan Berbasis Lokasi," Access date: 20-Jun-2023, [Online], https://socs.binus.ac.id/2014/07/18/perancangan-sistem-pakar-pendeteksi-penyakit-pada-tanaman-kopi-dengan-layanan-berbasis-lokasi/

[15] P.U. Rakhmawati, Y.M. Pranoto, and E. Setyati, "Klasifikasi Penyakit Daun Kentang Berdasarkan Fitur Tekstur dan Fitur Warna Menggunakan Support Vector Machine," *Seminar Nas. Teknol., Rekayasa (SENTRA) 2018*, 2018, pp. 1–8, doi: 10.22219/sentra.v0i4.2127.

[16] S.I. Novichasari and Y.S. Sipayung, "PSO-SVM untuk Klasifikasi Daun Cengkeh Berdasarkan Morfologi Bentuk Ciri, Warna dan Tekstur GLCM Permukaan Daun," *J. Multimatrix*, Vol. 1, No. 1, pp. 18–21, Dec. 2018.

[17] F. Jiang *et al*., "Image Recognition of Four Rice Leaf Diseases Based on Deep Learning and Support Vector Machine," *Comput., Electron. Agriculture*, Vol. 179, pp. 1–9, Dec. 2020, doi: 10.1016/j.compag.2020.105824.

[18] Trivusi (2022) "Data Splitting: Pengertian, Metode, dan Kegunaannya," [Online], https://www.trivusi.web.id/2022/08/data-splitting.html, access date: 20-Jun-2023.

[19] L. Hussain *et al*., "Lung Cancer Prediction Using Robust Machine Learning and Image Enhancement Methods on Extracted Gray-Level Co-occurrence Matrix Features," *Appl. Sci.*, Vol. 12, No. 13, pp. 1–20, Jun. 2022, doi: 10.3390/app12136517.

[20] I.M.O. Widyantara, N.M.A.E.D Wirastuti, and I.B.P. Adnyana, "Metode Contrast Stretching untuk Perbaikan Kualitas Citra pada Proses Segmentasi Video," *Maj. Ilm. Teknol. Elekt.*, Vol. 16, No. 2, pp. 1–6, May–Aug. 2017, doi: 10.24843/MITE.2017.vl6i02p01.

[21] H. Armagan, "K-Means Kümeleme Algoritması ile Renk Tabanlı Segmantasyon ve Renk Uzaylarının Görüntü Niceliklerine Etkisinin Sayısal Analizi," *El-Cezerî J. Sci., Eng.*, Vol. 9, No. 4, pp. 1506–1517, Dec. 2022, doi: 10.31202/ecjse.1141148.

[22] N. Mourya, Vidyashanakara, and G.H. Kumar, "Leaf Classification Based on GLCM Texture and SVM," *Int. J. Comput. Appl.*, Vol. 4, No. 3, pp. 156–159, Mar. 2018, doi: 10.5120/ijca2020919846.

[23] E. Alvansga, "Pengenalan Tekstur Menggunakan Metode GLCM serta Modul Nirkabel," Undergraduate thesis, Universitas Sanata Dharma, Yogyakarta, Indonesia, 2019.

[24] M. Furqan, S. Sriani, and L.S. Harahap, "Klasifikasi Daun Bugenvil Menggunakan Gray Level Co-Occurrence Matrix dan K-Nearest Neighbor," *J. CoreIT*, Vol. 6, No. 1, pp. 22–29, Jun. 2020, doi: 10.24014/coreit.v6i1.9296.

[25] R. Suganya, S. Rajaram, and A.S. Abdullah, *Big Data in Medical Image Processing*, ed. 1. Florida, AS: CRC Press, 2018, doi: 10.1201/b22456.

[26] M.F.T. Putra, "Penerapan Gray Level Co-Occurrence Matrix (GLCM) dan Learning Vector Quantization (LVQ) untuk Klasifikasi Penyakit Retina Mata," Final Project, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia, 2021.

[27] J. Webel, J. Gola, D. Britz, and F. Mücklich, "A New Analysis Approach Based on Haralick Texture Features for the Characterization of Microstructure on the Example of Low-Alloy Steels," *Mater. Charact.*, Vol. 144, pp. 584–596, Oct. 2018, doi: 10.1016/j.matchar.2018.08.009.

[28] Y.M. Oo and N.C. Htun, "Plant Leaf Disease Detection and Classification Using Image Processing," *Int. J. Res., Eng.*, Vol. 5, No. 9, pp. 516–523, Sep.–Oct. 2018, doi: 10.21276/ijre.2018.5.9.4.

[29] E. Prasetyo, *Data Minning: Mengolah Data Menjadi Informasi Menggunakan Matlab*, ed. 1. Yogyakarta, Indonesia: Andi, 2014.

[30] F. Hilmiyah, "Prediksi Kinerja Mahasiswa Menggunakan Support Vector Machine untuk Pengelola Program Studi di Perguruan Tinggi (Studi Kasus: Program Studi Magister Statistika ITS)," Master's thesis, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, 2017.

[31] Y.F. Khan, B. Kaushik, C.L. Chowdhary, and G. Srivastava, "Ensemble Model for Diagnostic Classification of Alzheimer's Disease Based on Brain Anatomical Magnetic Resonance Imaging," *Diagnostics*, Vol. 12, No. 12, pp. 1–27, Dec. 2022, doi: 10.3390/diagnostics12123193.

[32] S. Adinugroho and Y.A. Sari, *Implementasi Data Mining Menggunakan Weka*, ed. 1. Malang, Indonesia: UB Press, 2018.

[33] A.N. Rais, W. Warjiono, W. Kurniawan, and R. Ardianto "Analisa Akurasi dan F1 Score pada Algoritma Smote dan Naïve Bayes pada Dataset Bank Direct Marketing," *Speed-Sentra Penelit. Eng., Edukasi*, Vol. 11, No. 4, pp. 1–7, Oct. 2019, doi: 10.55181/speed.v11i4.620.