Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 7, No. 3, August, Pp. 299-308

KNN algorithm for identification of tomato disease based on image segmentation using enhanced k-means clustering

Amir Saleh*[1], Alvin[2], Ana Tince Siregar[3], Monica Sari Sinaga[4]
Universitas Prima Indonesia, Medan, Indonesia[1,2,3,4]

Article Info

Keywords:
Image segmentation, K-means clustering, Cosine similarity, Identifying tomato diseases, KNN

Article history:
Received: July 14, 2022
Accepted: August 20, 2022
Published: September 28, 2022

Cite:
A. S. Nasution, A. Alvin, A. T. Siregar, and M. S. Sinaga, "KNN Algorithm for Identification of Tomato Disease Based on Image Segmentation Using Enhanced K-Means Clustering", KINETIK, vol. 7, no. 3, Sep. 2022.
https://doi.org/10.22219/kinetik.v7i3.1486

*Corresponding author.
Amir Saleh
E-mail address:
amirsalehnst1990@gmail.com

Abstract

Image segmentation is an important process in identifying tomato diseases. The technique that is often used in this segmentation is k-means clustering. One of the main problems in this technique is the case of local minima, where the cluster that is formed is not suitable due to the incorrect selection of the initial centroid. In image data, this case will have an impact on poor segmentation results because it can erase parts that are actually important to be lost or there is still background in the recognition process, which has an impact on decreasing accuracy results. In this research, a method for image segmentation will be proposed using the k-means clustering algorithm, which has been added with the cosine similarity method as the proposed contribution. The use of the cosine method will determine the initial centroid by calculating the level of similarity of each image feature based on color and dividing them into several categories (low, medium, and high values). Based on the results obtained, the proposed algorithm is able to segment and distinguish between leaf and background images with good results, with the kNN reaching a value of 94.90% for accuracy, 99.50% for sensitivity, and 93.75% for specificity. The results obtained using the kNN method with k-means segmentation obtained a value of 92.46% for accuracy, 96.30% for sensitivity, and 91.50% for specificity. The results obtained using the kNN method without segmentation obtained a value of 90.22% for accuracy, 93.30% for sensitivity, and 89.45% for specificity.

1. Introduction

Indonesia is a country with two seasons that has abundant wealth in the fields of plantations and agriculture as a producer of high-quality horticultural products [1]. One of the products in this field is fruits and vegetables, including tomatoes. This plant can thrive in various places, which is one of the reasons for the choice of farmers in cultivating it. However, sometimes the yield of this fruit is still not as satisfying as expected. Some of the causes of crop failure include cultivation techniques, environmental conditions, and pest and disease disturbances. One of the main diseases in tomato plants is leaf spot [2].

Plant diseases can cause a decrease in agricultural production, so early detection and diagnosis of plant diseases is very important. Generally, tomato disease often appears on the leaves, with the characteristics of the affected leaves being varied and difficult to distinguish. In addition, these plant diseases are not only harmful to humans but also to animals [3]. Therefore, the detection and analysis of diseases in tomato plants to increase crop yields is very important because it is very difficult to detect and analyze tomato diseases manually [4]. Researchers have conducted many studies to obtain effectiveness in identifying plant diseases [5]. Image processing-based applications for the recognition and classification of plant diseases are a broad field of research today. This application is useful for precise identification or recognition of plant diseases, such as fungi, bacteria, and viruses that can damage plants [6].

Several studies have produced the right method for recognizing or identifying tomato diseases with various techniques, one of which is using the convolutional neural network (CNN) technique with good accuracy results above 90% [7][8]. Another technique uses a combination of random forest and k-nearest neighbor (kNN) with an accuracy of up to 96% [9]. In addition, the classification of diseases in tomatoes can be done using the classification tree technique, with an accuracy of recognition reaching 97.3% [6]. Basically, every identification or recognition technique that is carried out must be accompanied by image data, image processing techniques, and appropriate feature extraction to obtain the best results.

The image recognition process in computer science is known as computer vision, which can be applied to agriculture in identifying plant diseases [10][11]. Machine learning techniques that are commonly used in image identification [12] must begin with obtaining image data that does not contain a lot of noise (background) to maintain good identification results [13]. In this case, image segmentation is an important part of image processing techniques in several studies for image identification [14][15]. Image segmentation is the process of separating important objects from the background of an image [16] and is the basis of computer vision, where segmentation accuracy has a major impact on the identification results [17]. The output of the image segmentation process is a segmented image for several classes, where each class has its own attributes [18]. Various segmentation techniques can be used and generally use clustering techniques to separate objects from the background of an image. One of the most popular techniques in image segmentation is k-means clustering, which has been widely proposed to improve the accuracy of image-based object recognition.

The k-means clustering algorithm starts by selecting the k value (the number of clusters) and the initial centroid value randomly for each cluster. In each iteration, the data object is associated with the nearest centroid and is updated based on the distance value of the closest data object [19]. The k-means clustering algorithm in image segmentation is an unsupervised learning algorithm that is used to segment an image object with a background [16], so it is necessary to adjust the number of segmentation groupings and the initial centroid that have a good or bad influence on the segmentation quality [20]. A common problem that is often encountered is the determination of the initial centroid, which is done randomly, resulting in different clusters if the initialization is changed [13]. These changes can cause a properly segmented image to fail and affect the identification results [21]. The solution to this problem can be done by using an improved k-means clustering algorithm for precise separation between the studied image and the background [13][16].

Based on the previous explanation, the segmentation process in the identification of tomato diseases is very necessary in order to increase the accuracy of the results. In this study, identification of diseased tomato leaves will be carried out using the kNN (k-Nearest Neighbor) algorithm based on segmentation using an enhancement algorithm on k-means clustering to separate parts of objects from the background. The contribution to this research lies in the image segmentation process, which is carried out using the cosine similarity method on the k-means clustering algorithm in determining the initial centroid. Cosine similarity is a similarity method on data that produces a level of similarity with a smaller error value than the Euclidean and Jaccard similarity methods [22].

Cosine similarity will look for feature values with the highest and lowest similarity in each image data based on the color content of the image. This value will be used as the initial centroid in the k-means clustering algorithm in image segmentation. This feature separation is done so that data can be grouped according to the same feature value as it. Each clustering algorithm induces similarity between a given data point and the underlying clustering criteria [23]. Basically, the data grouping technique is done by testing the similarity between one feature and its central feature. The closer the size of the similarity of a feature to its central feature, the features will be grouped into the same cluster and vice versa.

In this study, the use of cosine will yield several characteristic features with different similarities between one another. The value of these features becomes the basis for the proposed segmentation process, and then the k-means clustering algorithm will look for the suitability of these features and partition them into several parts for image segmentation. The better the initial centroid value selected in the k-means clustering algorithm, the better the resulting clustering results [24]. The final result will be good clustering, which will find the appropriate image segmentation and improve the identification results.

2. Research Method

The main objective of this research is to improve the accuracy of tomato disease identification using the kNN method and segmentation using the proposed method. The stages carried out in this study can be seen in the block diagram as shown in Figure 1 below.

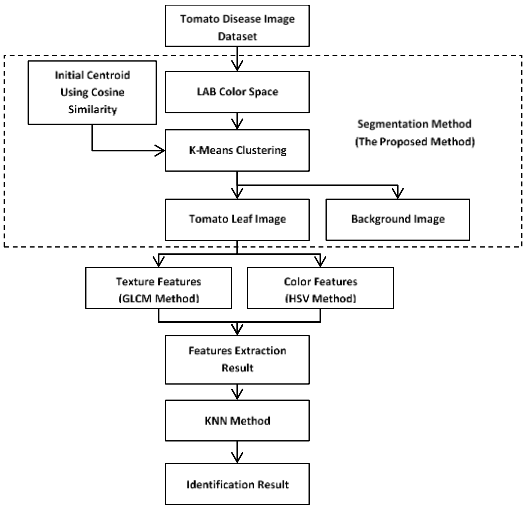

Figure 1. The Proposed Method

The steps in identifying tomato diseases using the proposed method can be described as follows.

1. Collecting image data obtained, namely images of tomato leaves infected with plant diseases.
2. Converting an RGB tomato image into a Lab color space to start segmenting the tomato leaf image. The result of changing this value is an image feature to segment the image at an early stage.
3. Perform the calculation of the initial centroid value search on the k-means clustering algorithm using the cosine similarity method. This method will look for the initial value of the centroid by comparing each image feature value to find the feature that has the highest similarity difference between one another. This process will produce the initial centroid point as many as the previously determined similarity value and is used as the initial centroid in the k-means clustering algorithm.
4. Perform calculations for the image segmentation process with the number of clusters (segmentation) of as many as 3 parts. Segmentation with k-means clustering will display clusters in the form of leaf images, the part of the leaf that is infected with the disease and the background. The cluster, which is part of the background, will be discarded and is the noise value in the process of identifying diseases in tomatoes.
5. Perform feature extraction calculations using the GLCM (Gray Level Co-occurance Matrix) method for texture characteristics and HSV for color features on tomato leaf images. The feature values used are 7, namely: contrast, correlation, energy, homogeneity, hue, saturation, and value.
6. Processing all image data with the same technique and combining them into a data matrix. This matrix is the input data and the data class that will be processed by the kNN method to carry out the process of identifying diseases in tomatoes.
7. Classification using the kNN method with a value of k = 3 (neighborhood value). This method works by entering a new image to calculate the distance or similarity with previously stored data (training data).
8. Conducting a voting process (class determination on the test image with the most similarities) as a result of identification. The next step is to calculate the overall identification results to get the accuracy of the test by matching the results of the identification using the kNN method with the actual results.

2.1 Dataset

The dataset used in this study was obtained from internet pages, namely Tomato leaf disease detection. The number of images used as research objects is 5000 images with 5 types of tomato disease with various image orientation angles. The description of the dataset used in this study can be seen in Table 1 as follows.

Table 1. Dataset

Class (Diseases Type)	No. of Images
Bacterial Spot	1000
Spider Mites Two Spotted	1000
Target Spot	1000
Mosaic Virus	1000
Yellow Leaf Curl Virus	1000

2.2 Image Segmentation Using Enhanced K-Means Clustering

After obtaining the image data, the next step will be a segmentation process to remove the background that is not needed in the tomato disease identification process. In this research, the process is carried out using the k-means clustering algorithm and cosine similarity, which is the proposed method. The purpose of image segmentation is to simplify or change the representation of an image into something more meaningful and easier to analyze [14]. The segmentation process is carried out to separate the leaf image from the background using the k-means clustering and cosine similarity algorithm. Cosine similarity in this study was carried out in selecting the initial centroid used in the k-means clustering algorithm in image segmentation. Next, the centroid value will be updated based on the object's proximity to the new centroid generated. This process will be carried out as many times as allowed during the initial determination of the clustering process with the k-means clustering algorithm. The final result is an image that is segmented into 3 parts, and the background will be removed.

At the initial stage of segmentation, the RGB image will be converted into the Lab color space to get the color feature value from the image. This change was made because the Lab color space is almost linear with visual perception and has more color differences for humans [25]. The equations used are in equation 1 and equation 2 below [26][27].

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4125 & 0.3576 & 0.1804 \\ 0.2127 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9502 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

$$L = 116 \ f\left(\frac{Y}{Y_w}\right) - 16, a = 500\left[f\left[\frac{X}{X_w}\right] - f\left[\frac{Y}{Y_w}\right]\right], b = 200\left[f\left[\frac{Y}{Y_w}\right] - f\left[\frac{Z}{Z_w}\right]\right] \quad (2)$$

Where: L is the brightness of the color, a is the position between magenta and green, and b is the position between yellow and blue. While Xw, Yw, and Zw are values, X = 95.04, Y = 100.00, and Z = 108.88 from the standard white reference point (D65).

The result of the conversion is in the form of feature value data from each image, and then the similarity search process for each feature will be searched using the cosine similarity equation. In this study, the number of similarities to be sought is 3 levels (taken from the number of clusters or the value of k = 3), namely: the highest similarity (max_similarity), medium similarity (med_similarity), and the lowest similarity (min_similarity). This process will produce 3 data features with different similarities and will be used as the initial centroid in the k-means clustering algorithm. This method is the method used in the clustering process, where each data feature of a different cluster will have a very small resemblance to the similarity of features in the cluster itself. The equation used can be seen in equation 3 below [28].

$$Cos\ \theta = \frac{a.b}{\|a\|.\|b\|} = \frac{\sum_{i=1}^{n} a_i * b_i}{\sqrt{\sum_{i=1}^{n}(a_i)^2} * \sqrt{\sum_{i=1}^{n}(b_i)^2}} \quad (3)$$

Where: Cos θ is a measure of the similarity of features (values in the range of 0–1), a is the comparison feature, and b is the compared feature.

By using the cosine equation, it can be made an equation that is used to find the level of similarity between features using equation 4, equation 5, and equation 6 below.

$$Centroid\ 1 = \ Max_{Similarity}(idx\ feature) = max(Cos\ \theta) \quad (4)$$

$$Centroid\ 2 = Med_{Similarity}(idx\ feature) = median(Cos\ \theta) \quad (5)$$

$$Centroid\ 3 = Min_{Similarity}(idx\ feature) = min(Cos\ \theta) \quad (6)$$

The next step is to enter the centroid value into the k-means clustering algorithm as the initial centroid initialization. The k-means clustering algorithm in image segmentation is an important step in separating leaf clusters, disease, and background in tomato leaf images. The success of this method is the result of a well-segmented image if the initial initialization is carried out properly. The equations used to separate image clusters can be seen in equation 7 and equation 8 below [16].

$$d = \|p(x, y) - ck\| \quad (7)$$

$$ck = \frac{1}{k}\sum_{y \in ck}\sum_{x \in ck} p(x, y) \quad (8)$$

Where: d is a measure of the similarity of the data with the Euclidean equation, p(x,y) is the data feature, k is the number of data clusters (number of image segmentations), and ck is the data centroid.

2.3 Feature Extraction

Feature extraction is an important step in the process of image classification and pattern recognition, with the aim of extracting discriminatory feature values from image data. The value of this feature will be fed as an input vector to the machine learning model from the visual description to retrieve information from an image and can be automatically distinguished by the machine [29]. The choice of input features has a significant impact on the classification accuracy and is a requirement for training the algorithm used [30][31].

One of the most popular and widely used texture-based feature extraction techniques is GLCM (Gray Level Co-occurrence Matrix), which was discovered by Haralick et al. (1973) [32].The GLCM feature is a feature that is taken from the relationship between two pixels that are given a certain distance in an image. The GLCM feature has a value that changes rapidly in the fine texture area and a value that changes slowly in the coarse texture area [33]. The GLCM method is used to retrieve the texture properties of an image and represent image information in grayscale [34]. The feature extraction method with GLCM in this study was carried out by taking the values of the texture characteristics, namely: contrast, correlation, energy, and homogeneity, using equations 9, 10, 11, and 12 below [30].

$$Contrast = \sum_i \sum_j (i-j)^2 P(i,j) \tag{9}$$

$$Correlation = \frac{\sum_i \sum_j i,j \ \ P[i,j] - \mu_i \mu_j}{\sigma_i \sigma_j} \tag{10}$$

$$Energy = \sum_i \sum_j P[i,j]^2 \tag{11}$$

$$Homogeneity = \sum_i \sum_j \frac{P[i,j]}{1+[i-j]} \tag{12}$$

Another feature that can be used to identify an image is the color feature that can be done using the HSV method. This method is used to define RGB colors that are converted into Hue, Saturation, and Value values. The RGB values contained in an image have a range of values between 0 – 255, which will be converted to HSV with values 0 – 360, 0 – 1, and 0 – 1. The equations that will be used to get the value of the color characteristics of an image using the HSV method are equations 13, 14, 15, and 16 below [35][36].

$$r(red) = \frac{R}{R+G+B}, g(green) = \frac{G}{R+G+B}, b(blue) = \frac{B}{R+G+B} \tag{13}$$

$$H(Hue) = \begin{cases} 60^0 \ x \left[\frac{g-b}{S \ x \ V}\right] & if \ V = r \\ 60^0 \ x \left[2 + \frac{b-r}{S \ x \ V}\right] & if \ max = g \\ 60^0 \ x \left[4 + \frac{r-g}{S \ x \ V}\right] & if \ max = b \\ H + 360^0 & if \ H < 0 \end{cases} \tag{14}$$

$$S(Saturation) = \begin{cases} 0 & if \ V = 0 \\ V - \frac{\min(r,g,b)}{V} & if \ V > 0 \end{cases} \tag{15}$$

$$V(Value) = max(r,g,b) \tag{16}$$

2.4 KNN Classifier

Research on the use of the kNN method has been a hot topic of research on data mining and machine learning techniques for data classification since this algorithm was proposed in 1967 [37]. The kNN method begins with selecting k (neighborhood values) samples from the closest training data sample for the test data sample, then predicts the test data sample with the highest class among the k closest training data samples. In this study, the use of the kNN method to identify diseases in tomatoes was tested. This stage is carried out after feature extraction using the GLCM and HSV methods is applied to each training image. At the next stage, the test image will be extracted with features in the same way and the classification or identification process will be carried out using the kNN method. The result of the identification is an index that shows the data is included in one type of disease in tomatoes. The general equation for identification matching can use the Euclidean distance with equation 17 below [38][39].

$$d(x,X) = \sqrt{\sum_{i=1}^{n}(x_i - X_i)^2} \tag{17}$$

Where: d is the minimum distance using the Euclidean distance equation, x is the test data, and X is the training data.

The final stage of this process is to calculate the accuracy, sensitivity, and specificity of the test classification results obtained with equations 18, 19, and 20 below [40].

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{18}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{19}$$

$$Specificity = \frac{TN}{TN + FP} \tag{20}$$

Where:
TP is the true positive, TN is the true negative, FP is the false positive, and FN is false negative.

3. Results and Discussion

The test begins by segmenting the image into several clusters to obtain important parts of the tomato leaf image. The result of segmentation is a leaf image without involving the background in the image. This segmentation is carried out using the k-means clustering method and the proposed method (cosine similarity and k-means clustering). The segmentation results will be continued by analyzing feature values based on leaf texture using the GLCM (Gray Level Co-occurance Matrix) and HSV methods for image color features. The final stage of this process is to identify the leaf using the kNN method to obtain accurate results for the method used. The identification stage is also used as a validation of the proposed method in image segmentation. This is because basic good segmentation will support good identification results and increase accuracy in the identification of the image under study.

3.1 Image Segmentation Results

In this section, we will compare the segmentation results obtained using the k-means clustering method and the proposed method. As an example, you can see the image of tomato leaves infected with the disease as shown in Figure 2.

Figure 2. Image of a Tomato Leaf

Segmentation using the k-means clustering algorithm is a popular method used in image recognition or identification. Segmentation results obtained using this method generally get good results. However, during the trial process of segmenting the image, there were still some images that were not segmented properly. This has an impact on the image that you want to research into an incomplete object. One of the causes is the difficulty in determining the cluster index, which is part of the background (the part that we want to eliminate). The results obtained show that the cluster index will vary according to the centroid specified at the beginning of this algorithm. An example of the results obtained by segmenting using the k-means clustering algorithm can be seen in Figure 3 below.

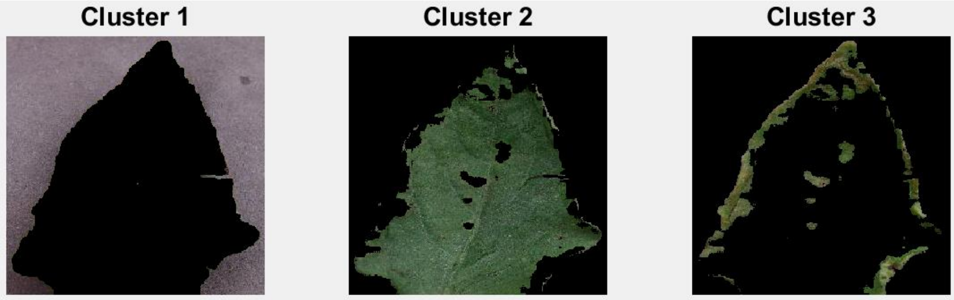

Figure 3. Clustering Results Using the K-Means Clustering Method (k = 3)

In Figure 3, the cluster index in the background is 1 and must be removed. This result may change when the centroid is initially changed. The equation used is to sum each cluster for a small pixel area and discard the cluster with the largest pixel area. In this case, the researcher will discard the maximum area, which is the area that is considered the background. In the case of Figure 3, the largest pixel area is in cluster 2 and is an important part of the image under study. This part will be removed, which makes the image object imperfect and only leaves cluster 1 and cluster 3. The final segmentation results using the k-means clustering method can be seen in Figure 4 below.


Figure 4. Results of Segmentation Using the K-Means Clustering Method

The segmentation in this section uses the cosine similarity method in determining the initial centroid value, and then the centroid will be corrected using the k-means clustering algorithm. The initial stage of segmentation with the proposed method is to find the 3 features with the highest, medium, and small similarity. The equations used are the maximum, median, and minimum similarity values according to the number of clusters that we want to use. In this study, 3 (k = 3) feature values will be used based on these similarities. Before the overall testing phase is carried out, several leaves will be tested first to find clusters containing the background. After obtaining the existence of a cluster background, the algorithm will calculate the entire image and remove the pixel area that contains the background. The clustering process obtained by the proposed method can be seen in Figure 5 below.

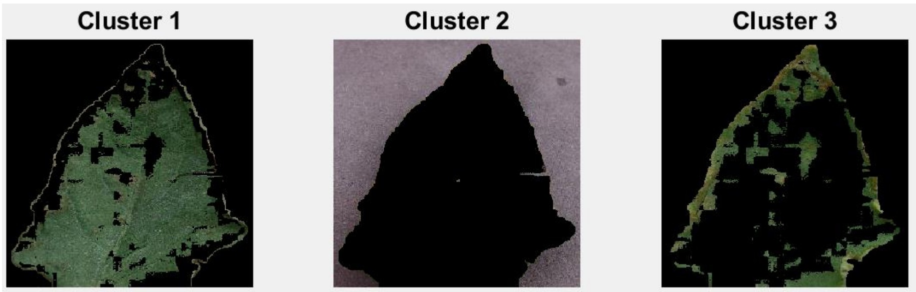

Figure 5. Clustering Results Using the Proposed Method (k = 3)

In Figure 5, the proposed algorithm is able to find the part of the cluster background that we want to remove. In this case, the omitted cluster is a cluster with an index value of 2, which is the background part of the image. The next stage is to unify the clusters containing the leaf images, namely clusters 1 and 3. The results obtained are better segmented images because the initial centroid value has been set and k-means clustering will help improve the results of the cluster and place clusters that are according to the given initial centroid value. The final segmentation results using the k-means clustering and cosine similarity method can be seen in Figure 6 below.

Figure 6. Results of Segmentation Using the Proposed Method

3.2 Identification Accuracy Results with the kNN Method

After the segmentation stage is completed using the k-means clustering method and the proposed method, the next step will be the feature extraction stage using the GLCM method for texture and the HSV method for color. The final stage is to identify the image using the kNN (k-Nearest Neighbor) algorithm. At this stage, the training image that has been feature extracted will be re-tested to determine the identification accuracy. The results obtained for each image test on tomato leaves infected with the disease can be seen in Figure 7 and Tabel 2 below.

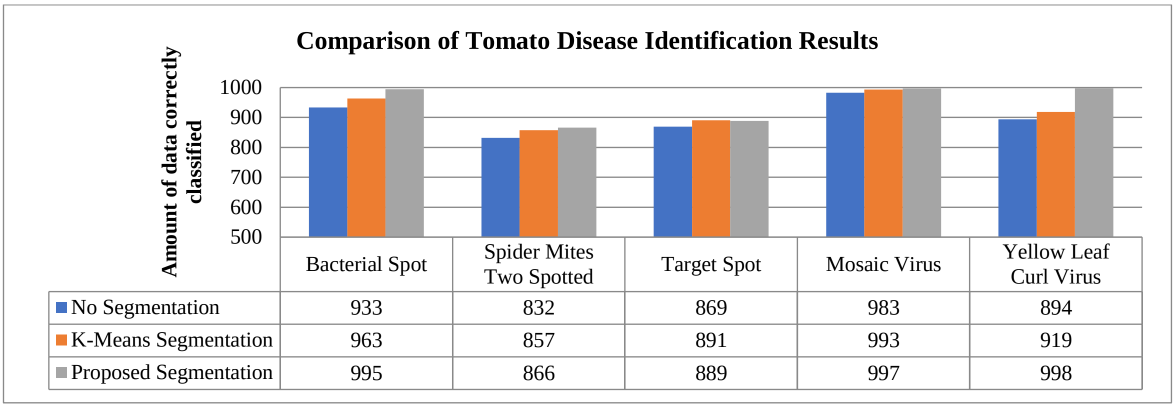

Figure 7. Comparison of Identification Results

In Figure 7, the results obtained with the proposed method obtain higher data accuracy than the methods without segmentation and with segmentation k-means clustering. These results show that in almost every type of image testing identified a disease, there is an increase using the proposed method.

Table 2. Results Comparison

Method	Results			
	No. Correct	Accuracy	Sensitivity	Specificity
KNN - No Segmentation	4511	90.22%	93.30%	89.45%
KNN - Kmeans Segmentation	4623	92.46%	96.30%	91.50%
KNN - Proposed Method	4745	94.90%	99.50%	93.75%

Based on Table 2 of the results of the tomato disease identification test, the results with the proposed method obtained an accuracy value of 94.90%, a sensitivity value of 99.50%, and a specificity value of 93.75%. The increase in identification results that occurred was 4.68% for accuracy, 6.20% for sensitivity, and 4.30% for specificity compared to without using segmentation. Meanwhile, the comparison of the identification results between the proposed method and the k-means clustering segmentation method is 2.44% for accuracy, 3.20% for sensitivity, and 2.25% for specificity. Basically, k-means clustering can have a higher yield when the selected centroid during the random centroid selection process is correctly selected. However, it takes a lot of testing to find the right combination of centroids. This is certainly not effective and efficient because it takes a long time because the process is done repeatedly.

The proposed method is an appropriate solution for selecting the initial centroid in the k-means clustering algorithm. The results obtained are the result of the process of analyzing the proximity of the features of the entered image. Features that have vastly different characteristics are less likely to be placed in the same cluster. The cosine similarity method used in this study is a method of measuring the similarity of image features that is able to find differences in features from one another and can be used as a basis for image segmentation. The results obtained from the tests carried out show that this process can determine the initial centroid and the location of the background in an image, making it easier to determine the clusters that you want to eliminate. Meanwhile, using the usual k-means clustering method, the results of each segmentation will vary according to the randomization process in the selection of the initial centroid. The method used is to assume that the background cluster is located in the largest part of the cluster area. This method can be done, but some images cannot be segmented properly and some leaf images are missing, which makes them more difficult to identify.

4. Conclusion

Based on the results of the tomato disease identification test using the kNN algorithm and segmentation enhanced K-Means Clustering, better results were obtained than without using segmentation or with the usual k-means clustering segmentation. This is obtained because the proposed method is able to provide a more precise centroid value on the k-means clustering algorithm in segmenting tomato leaf images. The image segmentation results obtained by the proposed method provide better results and are able to increase the accuracy of tomato disease identification. Results with the kNN method and the proposed method reached a value of 94.90% for accuracy, 99.50% for sensitivity, and 93.75% for specificity. The results obtained using the kNN method with k-means segmentation obtained an accuracy of 92.46%, a sensitivity of 96.30%, and a specificity of 91.50%. The results obtained using the kNN method without segmentation obtained an accuracy of 90.22%, a sensitivity of 93.30%, and a specificity of 89.45%.

Acknowledgement

The authors are very grateful to all those who have helped in the writing of this research, especially to the informatics engineering study program at Universitas Prima Indonesia.

References

[1] A. Hidayatuloh, M. Nursalman, and E. Nugraha, “Identification of Tomato Plant Diseases by Leaf Image Using Squeezenet Model,” *2018 Int. Conf. Inf. Technol. Syst. Innov. ICITSI 2018 - Proc.*, pp. 199–204, 2018. https://doi.org/10.1109/ICITSI.2018.8696087

[2] C. Usha Kumari, S. Jeevan Prasad, and G. Mounika, “Leaf disease detection: Feature extraction with k-means clustering and classification with ANN,” *Proc. 3rd Int. Conf. Comput. Methodol. Commun. ICCMC 2019*, no. Iccmc, pp. 1095–1098, 2019. https://doi.org/10.1109/ICCMC.2019.8819750

[3] H. Kibriya, R. Rafique, W. Ahmad, and S. M. Adnan, “Tomato Leaf Disease Detection Using Convolution Neural Network,” *Proc. 18th Int. Bhurban Conf. Appl. Sci. Technol. IBCAST 2021*, pp. 346–351, 2021. https://doi.org/10.1109/IBCAST51254.2021.9393311

[4] H. D. Gadade and D. K. Kirange, “Tomato leaf disease diagnosis and severity measurement,” *Proc. World Conf. Smart Trends Syst. Secur. Sustain. WS4 2020*, pp. 318–323, 2020. https://doi.org/10.1109/WorldS450073.2020.9210294

[5] C. Zhou, S. Zhou, J. Xing, and J. Song, “Tomato Leaf Disease Identification by Restructured Deep Residual Dense Network,” *IEEE Access*, vol. 9, pp. 28822–28831, 2021. https://doi.org/10.1109/ACCESS.2021.3058947

[6] H. Sabrol and K. Satish, “Tomato plant disease classification in digital images using classification tree,” *Int. Conf. Commun. Signal Process. ICCSP 2016*, pp. 1242–1246, 2016. https://doi.org/10.1109/ICCSP.2016.7754351

[7] M. M. Gunarathna and R. M. K. T. Rathnayaka, “Experimental determination of CNN hyper-parameters for tomato disease detection using leaf images,” *ICAC 2020 - 2nd Int. Conf. Adv. Comput. Proc.*, pp. 464–469, 2020. https://doi.org/10.1109/ICAC51239.2020.9357284

[8] Y. Altuntas and F. Kocamz, “Deep Feature Extraction for Detection of Tomato Plant Diseases and Pests based on Leaf Images,” *Celal Bayar Üniversitesi Fen Bilim. Derg.*, vol. 17, no. 2, pp. 145–152, 2021. https://doi.org/10.18466/cbayarfbe.812375

[9] A. Yousuf and U. Khan, “Ensemble Classifier for Plant Disease Detection,” *Int. J. Comput. Sci. Mob. Comput.*, vol. 10, no. 1, pp. 14–22, 2021. https://doi.org/10.47760/ijcsmc.2021.v10i01.003

[10] J. Shijie, J. Peiyi, H. Siping, and Sl. Haibo, “Automatic detection of tomato diseases and pests based on leaf images,” *Proc. - 2017 Chinese Autom. Congr. CAC 2017*, vol. 2017-Janua, pp. 3507–3510, 2017. https://doi.org/10.1109/CAC.2017.8243388

[11] M. Sardogan, A. Tuncer, and Y. Ozen, “Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm,” *UBMK 2018 - 3rd Int. Conf. Comput. Sci. Eng.*, pp. 382–385, 2018. https://doi.org/10.1109/UBMK.2018.8566635

[12] L. Liu, Y. Wang, and W. Chi, “Image Recognition Technology Based on Machine Learning,” *IEEE Access*, pp. 1–1, 2020. https://doi.org/10.1109/ACCESS.2020.3021590

[13] H. Yao, Q. Duan, D. Li, and J. Wang, “An improved K-means clustering algorithm for fish image segmentation,” *Math. Comput. Model.*, vol. 58, no. 3–4, pp. 790–798, 2013. https://doi.org/10.1016/j.mcm.2012.12.025

[14] E. Aghajari and G. D. Chandrashekhar, “Self-Organizing Map based Extended Fuzzy C-Means (SEEFC) algorithm for image segmentation,” *Appl. Soft Comput. J.*, vol. 54, pp. 347–363, 2017. https://doi.org/10.1016/j.asoc.2017.01.003

[15] C. Shanmugam and E. Chandira Sekaran, “IRT image segmentation and enhancement using FCM-MALO approach,” *Infrared Phys. Technol.*, vol. 97, no. December, pp. 187–196, 2019. https://doi.org/10.1016/j.infrared.2018.12.032

[16] N. Dhanachandra, K. Manglem, and Y. J. Chanu, “Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm,” *Procedia Comput. Sci.*, vol. 54, pp. 764–771, 2015. https://doi.org/10.1016/j.procs.2015.06.090

[17] B. Dong, G. Weng, and R. Jin, “Active contour model driven by Self Organizing Maps for image segmentation,” *Expert Syst. Appl.*, vol. 177, no. 178, p. 114948, 2021. https://doi.org/10.1016/j.eswa.2021.114948

[18] A. K. Helmy and G. S. El-Taweel, “Image segmentation scheme based on SOM-PCNN in frequency domain,” *Appl. Soft Comput. J.*, vol. 40, pp. 405–415, 2016. https://doi.org/10.1016/j.asoc.2015.11.042

[19] M. N. Qureshi and M. V. Ahamad, “An Improved Method for Image Segmentation Using K-Means Clustering with Neutrosophic Logic,” *Procedia Comput. Sci.*, vol. 132, pp. 534–540, 2018. https://doi.org/10.1016/j.procs.2018.05.006

[20] Y. Y. Sun, S. Chen, and L. Gao, “Feature extraction method based on improved linear LBP operator,” *Proc. 2019 IEEE 3rd Inf. Technol. Networking, Electron. Autom. Control Conf. ITNEC 2019*, no. Itnec, pp. 1536–1540, 2019. https://doi.org/10.1109/ITNEC.2019.8729320

[21] K. Tian, J. Li, J. Zeng, A. Evans, and L. Zhang, “Segmentation of tomato leaf images based on adaptive clustering number of K-means algorithm,” *Comput. Electron. Agric.*, vol. 165, no. March, p. 104962, 2019. https://doi.org/10.1016/j.compag.2019.104962

[22] M. Alobed, A. M. M. Altrad, and Z. B. A. Bakar, “A Comparative Analysis of Euclidean, Jaccard and Cosine Similarity Measure and Arabic

Wordnet for Automated Arabic Essay Scoring," *Proc. - CAMP 2021 2021 5th Int. Conf. Inf. Retr. Knowl. Manag. Digit. Technol. IR 4.0 Beyond*, pp. 70–74, 2021. https://doi.org/10.1109/CAMP51653.2021.9498119
[23] A. L. N. Fred and A. K. Jain, "Learning pairwise similarity for data clustering," *Proc. - Int. Conf. Pattern Recognit.*, vol. 1, pp. 925–928, 2006. https://doi.org/10.1109/ICPR.2006.754
[24] A. Ilham, D. Ibrahim, L. Assaffat, and A. Solichan, "Tackling Initial Centroid of K-Means with Distance Part (DP-KMeans)," *Proceeding - 2018 Int. Symp. Adv. Intell. Informatics Revolutionize Intell. Informatics Spectr. Humanit. SAIN 2018*, pp. 185–189, 2019. https://doi.org/10.1109/SAIN.2018.8673364
[25] J. Montenegro, W. Gomez, and P. Sanchez-Orellana, "A comparative study of color spaces in skin-based face segmentation," *2013 10th Int. Conf. Electr. Eng. Comput. Sci. Autom. Control. CCE 2013*, pp. 313–317, 2013. https://doi.org/10.1109/ICEEE.2013.6676048
[26] C. C. Tseng and S. L. Lee, "A Low-Light Color Image Enhancement Method on CIELAB Space," *2018 IEEE 7th Glob. Conf. Consum. Electron. GCCE 2018*, pp. 29–33, 2018. https://doi.org/10.1109/GCCE.2018.8574809
[27] M. Wirth and D. Nikitenko, "The effect of colour space on image sharpening algorithms," *CRV 2010 - 7th Can. Conf. Comput. Robot Vis.*, pp. 79–85, 2010.https://doi.org/10.1109/CRV.2010.17
[28] L. Sahu and B. R. Mohan, "An improved K-means algorithm using modified cosine distance measure for document clustering using Mahout with Hadoop," *9th Int. Conf. Ind. Inf. Syst. ICIIS 2014*, 2015. https://doi.org/10.1109/ICIINFS.2014.7036661
[29] S. Bakheet and A. Al-Hamadi, "Automatic detection of COVID-19 using pruned GLCM-Based texture features and LDCRF classification," *Comput. Biol. Med.*, vol. 137, no. August, p. 104781, 2021. https://doi.org/10.1016/j.compbiomed.2021.104781
[30] Priyanka and D. Kumar, "Feature Extraction and Selection of kidney Ultrasound Images Using GLCM and PCA," *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 1722–1731, 2020. https://doi.org/10.1016/j.procs.2020.03.382
[31] A. Zotin, Y. Hamad, K. Simonov, and M. Kurako, "Lung boundary detection for chest X-ray images classification based on GLCM and probabilistic neural networks," *Procedia Comput. Sci.*, vol. 159, pp. 1439–1448, 2019. https://doi.org/10.1016/j.procs.2019.09.314
[32] M. H. Abd Latif, H. Md Yusof, S. N. Sidek, and N. Rusli, "Implementation of GLCM Features in Thermal Imaging for Human Affective State Detection," *Procedia Comput. Sci.*, vol. 76, no. Iris, pp. 308–315, 2015. https://doi.org/10.1016/j.procs.2015.12.298
[33] M. Yogeshwari and G. Thailambal, "Automatic feature extraction and detection of plant leaf disease using GLCM features and convolutional neural networks," *Mater. Today Proc.*, no. xxxx, 2021. https://doi.org/10.1016/j.matpr.2021.03.700
[34] K. Prakash and S. Saradha, "Efficient prediction and classification for cirrhosis disease using LBP, GLCM and SVM from MRI images," *Mater. Today Proc.*, no. xxxx, pp. 2–7, 2021. https://doi.org/10.1016/j.matpr.2021.03.418
[35] L. S. Yu, S. Y. Chou, H. Y. Wu, Y. C. Chen, and Y. H. Chen, "Rapid and semi-quantitative colorimetric loop-mediated isothermal amplification detection of ASFV via HSV color model transformation," *J. Microbiol. Immunol. Infect.*, vol. 54, no. 5, pp. 963–970, 2021. https://doi.org/10.1016/j.jmii.2020.08.003
[36] G. Bargshady, X. Zhou, R. C. Deo, J. Soar, F. Whittaker, and H. Wang, "The modeling of human facial pain intensity based on Temporal Convolutional Networks trained with video frames in HSV color space," *Appl. Soft Comput. J.*, vol. 97, p. 106805, 2020. https://doi.org/10.1016/j.asoc.2020.106805
[37] Z. Deng, X. Zhu, D. Cheng, M. Zong, and S. Zhang, "Efficient kNN classification algorithm for big data," *Neurocomputing*, vol. 195, pp. 143–148, 2016. https://doi.org/10.1016/j.neucom.2015.08.112
[38] S. Alamuru and S. Jain, "Video event classification using KNN classifier with hybrid features," *Mater. Today Proc.*, no. xxxx, 2021. https://doi.org/10.1016/j.matpr.2021.03.154
[39] A. Almomany, W. R. Ayyad, and A. Jarrah, "Optimized implementation of an improved KNN classification algorithm using intel FPGA platform: Covid-19 case study," *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, 2022. https://doi.org/10.1016/j.jksuci.2022.04.006
[40] N. Binsaif, "Application of Machine Learning Models to the Detection of Breast Cancer," *Mob. Inf. Syst.*, vol. 2022, 2022. https://doi.org/10.1155/2022/7340689