Indonesian Journal of Computer Science
ISSN 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam No. 1, Padang, Indonesia
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Machine Learning and Fuzzy C-Means Clustering for the Identification of Tomato Diseases

Amir Saleh[1], Achmad Ridwan[2], M Khalil Gibran[3]
amirsalehnst1990@gmail.com[1], achmadridwan@unprimdn.ac.id[2], khalilgibran1612@gmail.com[3]
[1,2]Department of Informatics Engineering, Universitas Prima Indonesia, Medan, Indonesia
[3]Department of Data Science, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

Article Information

Submitted : 5 Sep 2023
Reviewed: 18 Sep 2023
Accepted : 15 Oct 2023

Keywords

Identification of tomato plant diseases, Fuzzy c-means, Machine learning, Lab color space

Abstract

Diseases in tomato plants can cause economic losses in the agricultural industry. Identification of tomato plant diseases is important to choosing the right action to control their spread. In this research, we propose an approach to identify tomato plant diseases using a machine learning algorithm and lab colour space-based image segmentation using the fuzzy c-means (FCM) clustering algorithm. The segmentation method aims to separate the infected area, leaf image, and background in the tomato plant image. In the first step, the tomato image is represented in the Lab colour space, which allows for combining information on brightness (L), red-green colour components (a), and yellow-blue colour components (b). Then, the FCM algorithm is applied to segment the image. The segmentation results are then evaluated through an identification process using machine learning techniques such as k-Nearest Neighbors (kNN), Random Forest (RF), Support Vector Machine (SVM), and Naïve Bayes (NB) to measure the level of accuracy. The dataset used in this research is tomato images, which include various plant diseases obtained from the Kaggle dataset. The performance results of the proposed method show that the segmentation approach based on Lab colour space with the FCM clustering algorithm is able to identify infected areas well. The accuracy value of each machine learning method used is kNN of 85.40%, RF of 88.87%, SVM of 80.73%, and NB of 74.60%. The proposed method shows success in accurately identifying types of tomato plant diseases and obtains improvements compared to without using segmentation.

A. Introduction

Tomato plants (solanum lycopersicum) have many benefits and are important in various aspects of human life [1]. Maintaining the health of tomato plants is very important because it has a big impact on the productivity and quality of the crop. Healthy tomato plants tend to produce more and higher-quality fruit. By maintaining plant health, farmers can increase the potential for profitable yields. However, tomato plants are susceptible to various diseases and pest attacks, such as viruses, bacteria, fungi, fleas, caterpillars, and mealybugs [2]. By recognising diseased tomatoes, farmers can improve plant health, reduce the risk of losses, and support agricultural sustainability.

Detection of tomato plant diseases can be done by examining symptoms on the plant, using visual observation, or using advanced technology such as laboratory tests or plant sensors. However, this process takes a long time and costs a lot. Thus, an approach based on artificial intelligence is widely used in recognising tomato diseases [3]. One technique that is often used is computer vision, where by utilising machine learning approaches, artificial intelligence, and image processing, we can recognise an object using digital images [4].

The computer vision approach makes it possible to detect plant diseases automatically and quickly, allows farmers to take appropriate preventive and control measures before infections spread more widely, and reduces the risk of losses in crop yields [5]. Computer vision techniques for detecting tomato diseases are one example of the application of machine learning in agriculture that can increase the efficiency and sustainability of the agricultural sector.

Various identification techniques using machine learning have been widely used to detect diseases in tomato plants. The use of the deep learning neural network method obtained very good accuracy. This research uses the CNN method and various input variations to recognise images of tomato plants infected with plant diseases [6]. Other studies used SVM and KNN techniques to identify diseases in tomatoes, and both methods succeeded in identifying diseases very well [7]. From the research conducted, the machine learning method can effectively be used in identifying tomato plant diseases with very good accuracy.

In computer vision techniques, there is one process that is very important to do well to obtain accurate identification results, namely the segmentation process [8]. The image segmentation process is very important in detecting tomato diseases because it allows the separation of relevant parts of the tomato plant image. Segmentation is a method used to divide images into separate parts or objects that have certain characteristics or attributes. In the process of detecting tomato diseases, segmentation helps identify areas on the plant that are infected with the disease more accurately.

In this research, we will use the fuzzy c-means (FCM) clustering technique to carry out the segmentation process. The segmentation process using this method has been widely applied to assist in the object recognition process. The fuzzy c-means (FCM) is a methodology that is widely used in image grouping for segmentation. The fuzzy c-means clustering (FCM) algorithm starts by randomly selecting the initial membership degree value for each cluster and produces better clustering results when the membership degree value is used correctly. The final result of good clustering will be appropriate image segmentation and improved identification results compared to without carrying out the segmentation process. The FCM method has been used in several experiments to test segmentation, with positive outcomes. For the best outcomes in deciding clustering-based image segmentation, the fuzzy c-means (FCM) method is employed. The proposed method may be utilised to categorise mangoes with an accuracy rate of 87%, as demonstrated by the results of the experiments conducted for this study [9]. The FCM method is a type of image segmentation algorithm that is effective [10].

According to the justification provided, the segmentation procedure is essential to boosting accuracy when identifying tomato diseases. In this research, images of disease-infected tomato leaves will be identified using machine learning algorithms, namely kNN, RF, SVM, and NB, based on segmentation results using the proposed FCM algorithm to separate object parts from image noise. The RGB image will first be transformed into Lab colour space in order to extract the colour characteristics of the image. The lab colour space is virtually linear with human visual perception, which is why this alteration was done. It is anticipated that using the suggested strategy will provide a good segmentation process solution and result in high accuracy in the identification of tomato diseases.

B. Research Method

The major goal of this study is to increase the precision of tomato disease identification combining segmentation with fuzzy c-means (FCM) clustering and machine learning techniques. The details of the proposed research stages can be seen in Figure 1 below.

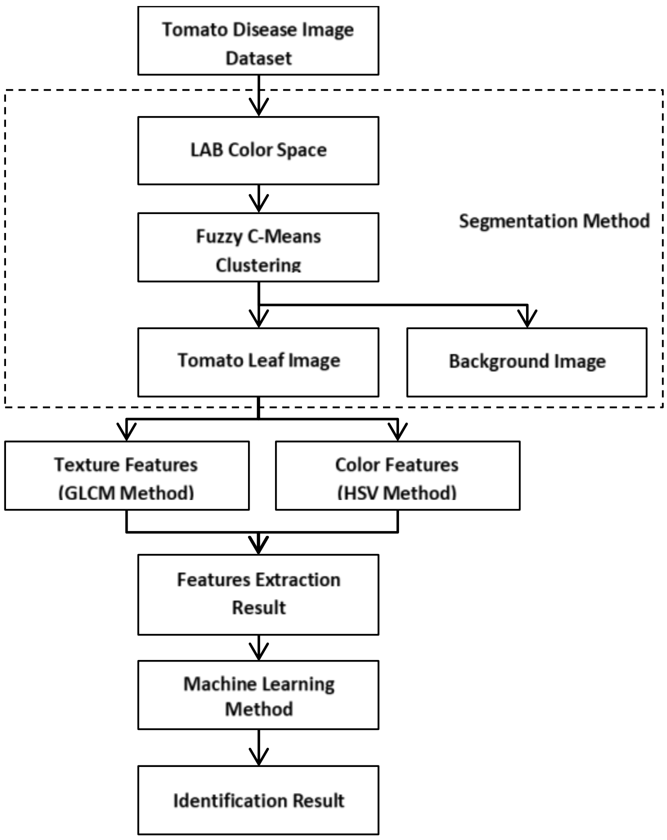

Figure1. The Proposed Method

The following steps might be used to describe the proposed approach for diagnosing tomato diseases:

1. Gather the acquired image data, namely images of infected tomato leaves.
2. Convert the tomato leaf image from RGB to Lab colour space to begin segmenting it.
3. Carry out calculations for the segmentation of an image into a total of three clusters. Clusters of leaf images, diseased leaf portions, and background will all be seen in segmentation with fuzzy c-means (FCM) clustering.
4. In tomato leaf images, do feature extraction calculations using the GLCM method for texture characteristics and HSV for colour attributes.
5. Create a data matrix by combining all image data using the same processing method. To diagnose diseases in tomatoes, this matrix will be used as input data and a data class in machine learning techniques.
6. Use the machine learning methods kNN, RF, SVM, and NB to identify. With the use of previously saved information (training data), this method calculates the identification results from a new image.
7. Calculating the overall identification results by comparing the identification results from machine learning algorithms with the actual results to test performance outcomes.

The Tomato Leaf Disease Dataset from the Kaggle Dataset provided the information needed in this investigation. There are 5000 images utilised as research items, with five different tomato diseases and different image orientations. Table 1 provides the following information about the dataset that was used in this study.

Tabel 1. The dataset used

Type of Diseases	Number of images
Bacterial Spot	1000
Spider Mites Two Spotted	1000
Target Spot	1000
Mosaic Virus	1000
Yellow Leaf Curl Virus	1000

The dataset will be divided into 2 parts, namely training data and testing data, with a division percentage of 70:30, where the entire image will be subjected to a disease identification process using the proposed method. Before carrying out the disease identification process on tomato plants, a segmentation process will first be carried out using the proposed method. In this research, the method used will be compared with methods in general to see to what extent the proposed method can perform segmentation well and influence identification results using machine learning techniques.

1. Image Segmentation with Fuzzy C-Means (FCM) Clustering

The fuzzy c-means (FCM) clustering technique is used in the segmentation procedure to remove portions of the leaf image from the background. The RGB image will be transformed into Lab colour space during the initial segmentation process in order to extract colour feature values from the image since Lab colour space is almost linear with human visual perception and has more colour variations [11]. To find the Lab colour value, you can use equations 1 and 2 below [12].

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

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

where,

- L : The colour brightness,
- a : The position between magenta and green,
- b : The position between yellow and blue, and
- X_w, Y_w dan Z_w : The values X = 95,04, Y = 100.00, and Z = 108,88 from the standard white reference point (D65).

The output of this conversion is feature value data from each image, after which the FCM method is used to group the image data. Three segmentations (clusters) of tomato leaf images, including the leaf image, the diseased portion of the leaf, and the background, make up the clustering findings of the suggested method. In the process of diagnosing tomato diseases, background clusters will be eliminated and turned into noise values. Equations 3 and 4 below illustrate how the FCM algorithm clusters images [10].

$$\mu_{ki} = \left[\sum_{j=1}^{m} \left(\frac{d_{ki}}{d_{ji}}\right)^{2/(w-1)}\right]^{-1} \quad (3)$$

where,

$$\begin{aligned} d_{ik} &= d(X_i - C_k) \\ &= \left[\sum_{j=1}^{m} (X_{ij} - C_{kj})^2\right]^{1/2} \end{aligned} \quad (4)$$

where,

μ_{ki} : The partition data (u matrix) at the k-th cluster centre and i-th data,
d_{ki} : The Euclidean distance function at the k-th cluster centre and i-th data,
d_{ji} : The Euclidean distance function at the centre of the j-th cluster and i-th data,
w : The weighting, and
X_{ij} : The data on the j-th attribute and i-th data.

2. Features Extraction

In order to extract discriminative feature values from image data, feature extraction is a crucial stage in the process of identifying pictures and pattern recognition. The choice of input features affects the precision of classification significantly and is a prerequisite for running training on the used algorithm [13][14]. The GLCM (Grey Level Co-occurrence Matrix), developed by Haralick et al. (1973), is one of the most well-known and commonly used texture-based feature extraction methods [14]. The GLCM method is used to represent image information in grayscale and obtain texture attributes from an image [15]. In this study, the contrast, correlation, energy, and homogeneity texture feature values were used as inputs for the feature extraction method with GLCM, which was carried out using the equations 5, 6, 7, and 8 below [16][17].

$$Contrast = \sum_i \sum_j (i-j)^2 P(i,j) \quad 5$$

$$Correlation = \frac{\sum_i \sum_j i,j \ P[i,j] - \mu_i \mu_j}{\sigma_i \sigma_j} \quad 6$$

$$Energy = \sum_i \sum_j P[i,j]^2 \quad 7$$

$$Homogeneity = \sum_i \sum_j \frac{P[i,j]}{1+[i-j]} \quad 8$$

The colour feature, which may be done using the HSV approach, is another characteristic that can be used to identify an image. Using this technique, RGB colours are defined and their hue, saturation, and value values are translated. The following equations 9, 10, and 11 can be used to calculate an image's colour characteristic value using the HSV method [17][18].

$$H\,(Hue) = \begin{array}{ll} 60^0 \text{ x } \left[\frac{g-b}{S \text{ x } V}\right] & if\ V = r \\ 60^0 \text{ x } \left[2 + \frac{b-r}{S \text{ x } V}\right] & if \max = g \\ 60^0 \text{ x } \left[4 + \frac{r-g}{S \text{ x } V}\right] & if\ max = b, \\ H + 360^0 & if\ H < 0 \end{array} \quad 9$$

$$S\,(Saturation) = \begin{cases} 0 & if\ V = 0 \\ V - \frac{\min(r,g,b)}{V} & if\ V > 0 \end{cases} \quad 10$$

$$V = \max(r,g,b) \quad 11$$

3. Machine Learning

Machine learning is a method that is widely applied in the process of recognising or identifying objects [19]. Several algorithms are popular and have good identification rates, such as kNN, random forest, SVM, and Naïve Bayes.

a. K-Nearest Neigbors (kNN)

K-Nearest Neighbors (KNN) is a machine learning algorithm that is simple yet effective in classifying. The basic theory of KNN is that an object will be classified based on the majority of the labels of its nearest neighbors in the feature space [20]. In kNN, the prediction of the class (label) of new data is based on the majority of the classes from its k nearest neighbors in the feature space. The mathematical equation for kNN in classification can be found equation 12 below [20].

$$d(x,y) = \sum_{i=1}^{n} \sqrt{x_i^2 - y_i^2} \quad 12$$

where,

N : The number of features such that,
x : $\{x_1, x_2, x_3, ..., x_N\}$, and
y : $\{y_1, y_2, y_3, ..., y_N\}$.

.

b. Random Forest

One of the most potent and well-liked machine learning classification methods is Random Forest. As an ensemble technique, random forest makes use of numerous decision trees, which are more basic machine learning models, to provide predictions that are more accurate [20]. The steps for finding results with the random forest method are as follows [20]:

1. Data Subset Selection,
2. Feature Subset Selection,
3. Making Decision Trees, and
4. Voting (classification).

c. Support Vector Machine

Support Vector Machine (SVM) is a classification machine learning technique [20]. SVM works by finding the best hyperplane (plane or line) that can separate two classes in the dataset. In the case of SVM with linear classification, the main goal is to find the best hyperplane that can separate two classes. Equation 13 below shows the employed formula [21].

$$f(x) = sign(\sum_{i=1}^{n} w_i x_i + b) \qquad 13$$

where,

f(x) : The decision function that produces class labels for input data x,
xi : The i-th attribute or feature of input data x,
wi : The weight corresponding to the ith attribute learned by the SVM model,
b : The bias that affects the hyperplane position, and
sign(·) : Signum function, which returns +1 if the argument is positive or 0, and -1 if the argument is negative.

d. Naïve Bayes

The Naïve Bayes is an effective strategy that can be used with all kinds of datasets and offers a higher level of accuracy for classification or identification outcomes utilising statistical approaches and probabilistic procedures [22]. Equation 14 below is used to find the Naïve Bayes classifier, a group of classification algorithms based on the Bayesian theorem [23].

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \qquad 14$$

where,

P(H|X) : The probability of Hypothesis H based on Condition X,
P(X|H) : The probability of X based on conditions in Hypothesis H,
P(H) : The probability of Hypothesis H, and
P(X) : The probability of X.

4. Performance Evaluation

Performance evaluation in the tomato disease identification process in this study used relevant confusion matrix techniques, such as accuracy, precision, recall, and f1-score [24]. To calculate the performance of the method used in this research, it can be calculated using equations 15, 16, 17, and 18 below [25][26].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (15)$$

$$Precision = \frac{TP}{TP+FP} \quad (16)$$

$$Recall = \frac{TP}{TP+FN} \quad (17)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \quad (18)$$

where,

TP : True Positive,
TN : True Negative,
FP : False Positive, and
FN : False Negative.

C. Result and Discussion

Before carrying out the process of identifying diseases in tomato plants, a segmentation process will be carried out using the proposed method. In this study, the method used will be compared with methods in general to see the extent to which the proposed method can perform segmentation properly so as to influence the proposed identification results.

1. Tomato Leaf Image Segmentation Results

The segmentation results from the fuzzy c-means (FCM) clustering method will be shown in this section. Figure 2 below, for instance, depicts a image of a tomato leaf that has a disease.

Figure 2. Plant Disease-Infected Tomato Leaves

A common technique for image recognition or identification is segmentation using the fuzzy c-means (FCM) clustering algorithm. In general, this strategy produces good segmentation results. As illustrated in Figure 3 below, segmentation using the fuzzy c-means (FCM) clustering algorithm can produce a variety of results.

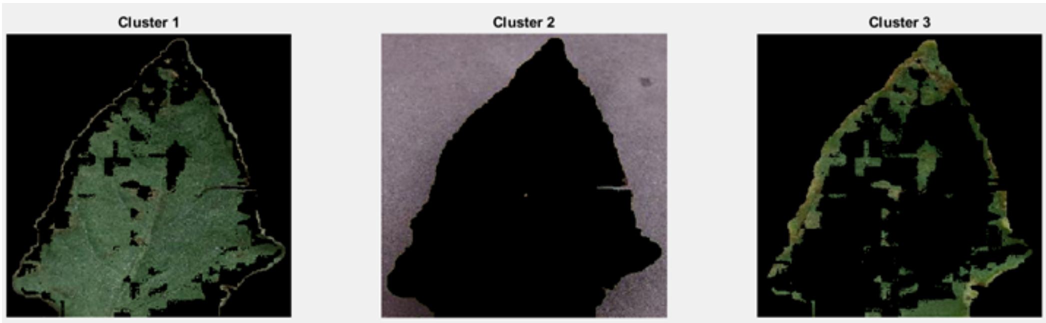

Figure 3. Image Clustering Results Using the Fuzzy C-Means Clustering Method

In Figure 3, Section 2 has the background cluster index, which needs to be deleted. Using an equation, each cluster with a small pixel area is added together, and the cluster with the biggest pixel area is discarded. The largest area, which is the area that is thought of as the background in this case, will be eliminated by the algorithm. In Figure 3, the area of the background that will be eliminated has the most pixels, which are located in cluster 2. As shown in Figure 4 below, the final segmentation results utilising the fuzzy c-means clustering technique.

Figure 4. Leaf Image Segmentation Results with the Fuzzy C-Means Clustering

2. Identification Results of Tomato Plant Diseases

After carrying out the segmentation process with the proposed method, the process continues with the identification of tomato plant diseases. The method used is machine learning, such as kNN, RF, SVM, and NB. This is done to see the level of accuracy in identifying the type of disease in tomato plants. The testing procedure is carried out either with or without the application of the FCM segmentation result method. The results of tests carried out without using FCM segmentation can be seen in Table 2 below.

Table 2. Identification Results Without Using the FCM Segmentation Method

Methods	Results of Identification			
	Accuracy	Precision	Recall	F1-Score
kNN	81.47%	0.8153	0.8147	0.8148
RF	84.27%	0.8426	0.8427	0.8422
SVM	72.87%	0.7497	0.7287	0.7165
NB	71.13%	0.7148	0.7113	0.7037

From the results of the tests carried out, the identification produced without using the FCM segmentation method has quite good results. The accuracy results obtained using the machine learning method, namely kNN of 81.47%, RF of 84.27%, SVM of 72.87%, and NB of 71.13%. Then, further testing will be carried out using FCM segmentation on the image of infected tomato plants. The test results using the proposed method can be seen in Table 3 below.

Table 3. Identification Results Using the FCM Segmentation Method

Methods	Results of Identification			
	Accuracy	Precision	Recall	F1-Score
kNN	85.40%	0.8534	0.8540	0.8536
RF	88.87%	0.8882	0.8887	0.8883
SVM	80.73%	0.8072	0.8073	0.8068
NB	74.60%	0.7445	0.7460	0.7434

From the results of the tests carried out, the identification produced using the FCM segmentation method produced better results than without using the FCM segmentation technique. The accuracy results obtained using the machine learning method can be described, namely kNN of 85.40%, RF of 88.87%, SVM of 80.73%, and NB of 74.60%. An increase in the accuracy of disease identification in tomato plants with FCM segmentation was seen in all machine learning algorithms tested (kNN, RF, SVM, and NB). Specifically, RF achieved the highest improvement in accuracy, rising from 84.27% to 88.87% after applying FCM segmentation.

This shows that the use of FCM segmentation has improved the model's ability to better recognise and classify types of disease on tomato plants. Thus, the results of this study indicate that the use of the FCM segmentation method can be an effective step in improving the quality of disease identification and diagnosis in tomato plants using machine learning techniques. This has the potential to support agriculture and monitor plant health more efficiently when implemented.

Image segmentation is very important in a variety of applications, including pattern recognition, object detection, computer vision, and image analysis. In pattern recognition, image segmentation helps identify and separate objects from the background so that further analysis can be carried out for shape and pattern recognition. In agriculture, image segmentation is very important for identifying and isolating areas showing disease. However, image segmentation is also a complex task and a challenge in itself, especially when the image has noise, high complexity, or when objects have varying sizes and shapes. The selection of the right segmentation method depends on the type of image and the desired analysis objectives.

Image segmentation FCM clustering has several advantages that make it an attractive option in image analysis. The FCM method is able to handle images contaminated by noise well. This method uses fuzzy membership values that allow pixels to belong to more than one group, so they are able to deal with uncertainties and variances that may exist in image data. In addition, the FCM method is relatively easy to implement and can be adapted for various image segmentation applications. Although the FCM method has many advantages, there are still some considerations that need to be made. As with other clustering methods, the results of segmentation are highly dependent on the initialization and the parameters used. To ensure good segmentation quality, proper parameter selection and repeated testing may be required. In addition, there are situations when the FCM method is not optimal for images with very heterogeneous regions or images with very large sizes because it requires significant computing resources.

In the disease identification experiment on tomato plants, machine learning methods such as kNN, RF, SVM, and NB were used to analyse plant images. Experimental results show that without FCM segmentation, the RF method provides the highest accuracy, which shows good ability to recognise diseases in plants. However, by applying FCM segmentation to images of disease-infected plants, all methods experienced significant improvements in identification accuracy. In this research, the RF method is the best method after FCM segmentation, but the increase in accuracy with FCM segmentation shows the importance of image processing in agricultural applications and disease detection in plants. The combination of image segmentation techniques such as FCM and machine learning can be a very effective tool in supporting modern agriculture with the early detection of diseases in plants.

D. Conclusion

In research on disease identification in tomato plants, machine learning methods such as k-Nearest Neighbors (kNN), Random Forest (RF), Support Vector Machine (SVM), and Naïve Bayes (NB), which were used to analyse plant images, obtained good results. Experimental results show that with fuzzy c-means (FCM) segmentation, the overall machine learning algorithm used provides better accuracy than without segmentation, which shows good capabilities in recognising diseases in plants. In this test, the RF method is the best method after FCM segmentation, with an identification accuracy of 88.87%. The combination of image segmentation techniques such as fuzzy c-means (FCM) clustering and machine learning can be a very effective tool in supporting agriculture with the early detection of diseases.

E. Acknowledgment

The Research Team would like to express its thanks to the Directorate General of Higher Education, Research, and Technology for funding this research under the Beginner Lecturer Research Grant Scheme (PDP) for the 2023 funding year. The Research Team would also like to thank FAST (Faculty of Science and Technology) at Universitas Prima Indonesia for their willingness and cooperation in carrying out research activities.

F. References

[1] M. M. Saffan *et al.*, "Sustainable Production of Tomato Plants (Solanum lycopersicum L.) under Low-Quality Irrigation Water as Affected by Bio-Nanofertilizers of Selenium and Copper," *Sustain.*, vol. 14, no. 6, pp. 1–17, 2022, doi: 10.3390/su14063236.

[2] V. K. Singh, A. K. Singh, and A. Kumar, "Disease management of tomato through PGPB: current trends and future perspective," *3 Biotech*, vol. 7, no. 4, pp. 1–10, 2017, doi: 10.1007/s13205-017-0896-1.

[3] A. Guerrero-Ibañez and A. Reyes-Muñoz, "Monitoring Tomato Leaf Disease through Convolutional Neural Networks," *Electron.*, vol. 12, no. 1, pp. 1–15, 2023, doi: 10.3390/electronics12010229.

[4] H. Orchi, M. Sadik, and M. Khaldoun, "On using artificial intelligence and the internet of things for crop disease detection: A contemporary survey," *Agric.*, vol. 12, no. 1, 2022, doi: 10.3390/agriculture12010009.

[5] N. Khasawneh, E. Faouri, and M. Fraiwan, "Automatic Detection of Tomato Diseases Using Deep Transfer Learning," *Appl. Sci.*, vol. 12, no. 17, 2022, doi: 10.3390/app12178467.

[6] N. K. Trivedi *et al.*, "Early detection and classification of tomato leaf disease using high-performance deep neural network," *Sensors*, vol. 21, no. 23, 2021, doi: 10.3390/s21237987.

[7] P. P. P. A. N. W. F. I. R. H. Zer, F. N. Tambunan, R. Rosnelly, and W. Wanayumini, "Comparison of Tomato Leaf Disease Classification Accuracy Using Support Vector Machine and K-Nearest Neighbor Methods," *SinkrOn*, vol. 8, no. 2, pp. 939–947, 2023, doi: 10.33395/sinkron.v8i2.12195.

[8] H. Mittal, A. C. Pandey, M. Saraswat, S. Kumar, R. Pal, and G. Modwel, "A comprehensive survey of image segmentation: clustering methods, performance parameters, and benchmark datasets," *Multimed. Tools Appl.*, vol. 81, no. 24, pp. 35001–35026, 2022, doi: 10.1007/s11042-021-10594-9.

[9] L. Marlinda, M. Fatchan, W. Widiyawati, F. Aziz, and W. Indrarti, "Segmentation of Mango Fruit Image Using Fuzzy C-Means," *SinkrOn*, vol. 5, no. 2, pp. 275–281, 2021, doi: 10.33395/sinkron.v5i2.10933.

[10] J. Chen, C. Yang, G. Xu, and L. Ning, "Image Segmentation Method Using Fuzzy C Mean Clustering Based on Multi-Objective Optimization," *J. Phys. Conf. Ser.*, vol. 1004, no. 1, 2018, doi: 10.1088/1742-6596/1004/1/012035.

[11] 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, doi: 10.1109/ICEEE.2013.6676048.

[12] 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, doi: 10.1109/CRV.2010.17.

[13] 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, doi: 10.1016/j.procs.2020.03.382.

[14] 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, doi: 10.1016/j.procs.2019.09.314.

[15] M. Yogeshwari and G. Thailambal, "Automatic feature extraction and detection of plant leaf disease using GLCM features and convolutional neural networks," *Mater. Today Proc.*, vol. 81, pp. 530–536, 2023, doi: https://doi.org/10.1016/j.matpr.2021.03.700.

[16] 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, doi: 10.1016/j.procs.2020.03.382.

[17] A. Saleh, D. Suryandy, and J. Nainggolan, “Face Image Retrieval System Using Combination Method of Self Organizing Map and Normalized Cross Correlation,” *J. Infokum*, vol. 9, no. 2, pp. 219–228, 2021.
[18] 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, doi: 10.1016/j.jmii.2020.08.003.
[19] J. Wäldchen and P. Mäder, “Machine learning for image based species identification,” *Methods Ecol. Evol.*, vol. 9, no. 11, pp. 2216–2225, 2018, doi: 10.1111/2041-210X.13075.
[20] E. Y. Boateng, J. Otoo, and D. A. Abaye, “Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review,” *J. Data Anal. Inf. Process.*, vol. 08, no. 04, pp. 341–357, 2020, doi: 10.4236/jdaip.2020.84020.
[21] Z. Jun, “The Development and Application of Support Vector Machine,” *J. Phys. Conf. Ser.*, vol. 1748, no. 5, 2021, doi: 10.1088/1742-6596/1748/5/052006.
[22] P. C. V, V. P. Pandian V, V. K. Kumar V, and S. M. Bharathi V, “Recommendation System Using Naive Bayes Classifier,” *Int. Res. J. Eng. Technol.*, pp. 5507–5510, 2020, [Online]. Available: www.irjet.net
[23] A. Saleh, N. Dharshinni, D. Perangin-Angin, F. Azmi, and M. I. Sarif, “Implementation of Recommendation Systems in Determining Learning Strategies Using the Naïve Bayes Classifier Algorithm,” *Sinkron*, vol. 8, no. 1, pp. 256–267, 2023, doi: 10.33395/sinkron.v8i1.11954.
[24] F. Azmi, M. K. Gibran, A. Ridwan, and A. Saleh, “Enhancing Water Potability Assessment Using Hybrid Fuzzy-Naïve Bayes,” *Indones. J. Comput. Sci.*, vol. 12, no. 1, pp. 1032–1043, 2023, doi: https://doi.org/10.33022/ijcs.v12i3.
[25] N. E. Ramli, Z. R. Yahya, and N. A. Said, “Confusion Matrix as Performance Measure for Corner Detectors,” *J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 29, no. 1, pp. 256–265, 2022, doi: 10.37934/araset.29.1.256265.
[26] A. Tasnim, M. Saiduzzaman, M. A. Rahman, J. Akhter, and A. S. M. M. Rahaman, “Performance Evaluation of Multiple Classifiers for Predicting Fake News,” *J. Comput. Commun.*, vol. 10, no. 09, pp. 1–21, 2022, doi: 10.4236/jcc.2022.109001.