

Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning

Seyed Mohamad Javidan ^a, Ahmad Banakar ^{a,*}, Keyvan Asefpour Vakilian ^b, Yiannis Ampatzidis ^c

^a Department of Biosystems Engineering, Tarbiat Modares University, Tehran, Iran

^b Department of Biosystems Engineering, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran

^c Agricultural and Biological Engineering Department, Southwest Florida Research and Education Center, University of Florida, 2685 FL-29, Immokalee, FL 34142, USA



ARTICLE INFO

Keywords:
Artificial intelligence
Deep learning
Disease classification
Grape diseases
Machine vision

ABSTRACT

Plant diseases often reduce crop yield and product quality; therefore, plant disease diagnosis plays a vital role in farmers' management decisions. Visual crop inspections by humans are time-consuming and challenging tasks and, practically, can only be performed in small areas at a given time, especially since many diseases have similar symptoms. An intelligent machine vision monitoring system for automatic inspection can be a great help for farmers in this regard. Although many algorithms have been introduced for plant disease diagnosis in recent years, a simple method relying on minimal information from the images is of interest for field conditions. In this study, a novel image processing algorithm and multi-class support vector machine (SVM) were used to diagnose and classify grape leaf diseases, i.e., black measles, black rot, and leaf blight. The area of disease symptoms was separated from the healthy parts of the leaf utilizing K-means clustering automatically, and then the features were extracted in three color models, namely RGB, HSV, and $l^*a^*b^*$. As an efficient classification method, SVM was used in this study, where principal component analysis (PCA) was performed for feature dimension reduction. Finally, the most important features were selected by the relief feature selection. Gray-level co-occurrence matrix (GLCM) features resulted in an accuracy of 98.71%, while feature dimension reduction using PCA resulted in an accuracy of 98.97%. The proposed method was compared with two deep learning methods, i.e., CNN and GoogleNet, which achieved classification accuracies of 86.82% and 94.05%, respectively, while the processing time for the proposed method was significantly shorter than those of these models.

1. Introduction

Agriculture is the primary source of national income for many countries, including Iran. Crop diseases are serious causes of reducing the quantity and quality of production; therefore, identifying plant diseases are of great importance. Disease symptoms can occur in different parts of the plant; however, plant leaves are commonly used to diagnose diseases [1,2]. Early and accurate diagnosis is a critical first step in mitigating losses caused by plant diseases. An incorrect diagnosis can lead to improper management decisions, such as selecting the unsuitable chemical application, potentially resulting in further health loss and yield reduction [3,4]. The unaided eye method is a traditional method of identifying diseases that requires enormous manpower and is prone to human error, time-consuming, and not applicable for large fields [5]. In addition, it is costly as it requires continuous monitoring by experts. Intelligent disease detection techniques can be beneficial in

detecting a plant disease at the initial growth stages [6].

As a reliable prediction methodology, machine learning can detect various fungal, bacterial, and viral diseases [7,8]. Intelligent detection of plant diseases by utilizing machine learning algorithms is an essential research topic as it may prove advantageous in monitoring large fields and automatically detecting diseases based on symptoms appearing on plant leaves. Advanced technologies can be used to reduce the adverse effects of plant diseases by diagnosing them in early development stages. The application of artificial intelligence and computer vision for the automatic diagnosis of plant diseases is now widely studied because human monitoring of plant diseases is tedious, time-consuming, and challenging [9].

In recent years, there has been a lot of research in the field of machine vision in agriculture, including fruit maturity classification and quality rating [10,11,12], fruit disease diagnosis [13,14], plant pest diagnosis [15–17], plant species classification [18,19], fruit

* Corresponding author.

E-mail address: ah_banakar@modares.ac.ir (A. Banakar).

| | Black measles | Black rot | Leaf blight | Healthy leaf |
|---------------------------------|--|---|---|--------------|
| Leaf image | | | | |
| Disease symptoms | The symptoms can be identified on the leaves as they take on a tiger-stripe pattern. | Spots on leaves appear as small, tan to reddish-brown circular lesions. | First, it appears as red spots on the upper leaf surface. These circular spots enlarge and become tan to light brown with distinct, dark borders. | No symptoms |
| Number of images in the dataset | 1309 | 1105 | 1058 | 413 |

Fig. 1. Dataset of healthy and diseased grape leaves and their symptoms.

identification in harvesting robots [20], weed control and recognition [21,22], and disease diagnosis and classification in plant organs [23]. Machine vision can include a variety of sensors, such as color, multispectral, and hyperspectral cameras. In typical machine vision applications, illumination used and captured by the sensor is in the visible spectral range.

Al Bashish et al. [24] proposed a framework for detecting and classifying RGB images of leaf and stem diseases for pepper. Their method was investigated for five diseases: early scorch, cottony mold, ashen mold, late scorch, and tiny whiteness. First, a color transformation was conducted, and then, *K*-means clustering was used for segmentation. The color co-occurrence method was applied for feature extraction, and finally, an artificial neural network (ANN) classifier was employed. Ratnasari et al. [25] proposed a system for sugarcane leaf disease detection utilizing RGB images. The proposed system has been verified only on three diseases: rust spot, ring spot, and yellow spot. A combination of color and texture features was used in feature extraction, while SVM was used for classification. In the SVM classifier, four kernel types were tested: linear, quadratic, radial basis function, and polynomial, among which linear kernel provided better results than the others. However, because of the limitations of the segmentation method, only an accuracy of 80% has been achieved. Rastogi et al. [26] developed a machine vision-based technology for maple and hydrangea leaf disease detection and grading. First, preprocessing was performed on leaf RGB images, then segmentation was applied using *K*-means clustering and the Euclidean distance technique. In feature extraction, the GLCM matrix is considered in which contrast, energy, homogeneity, and correlation have been calculated. In this study, ANN has been used for classification. For grading, using total leaf area (AT) and disease area (AD), the percentage of infection has been calculated. After calculating

the percentage of infection, grading has been applied by using fuzzy logic.

Recently, due to the advancement in computing, a large number of artificial neurons are stacked in a specific architecture that forms deep neural networks, and these are capable of learning the features automatically, contrary to the previous approach. These features are used for image classification (in different domains), and this is popularly known as deep learning. One of the deep learning approaches, namely convolutional neural network (CNN), is widely used for image classification [27]. Different CNN-based architectures, i.e., AlexNet, GoogLeNet, ResNet50, etc., have been developed to solve disease classification in various crops [28,29,30,31]. The training from scratch approach requires a large dataset for training as it is a data-driven technique that poses a significant challenge to the researchers. In several studies, authors have used their own developed datasets to classify various crop diseases [32,33]. Cruz et al. [34] presented a novel algorithm for framing a convolutional neural network (CNN) by fusing data at different levels of abstraction to improve the system's performance, which achieved a true positive rate of $98.60 \pm 1.47\%$. They utilized this method for detecting olive quick decline syndrome in olive leaves. Ferentinos [35] adopted a CNN model for disease diagnosis on an open database containing RGB images of 25 different plants. They achieved a 99.53% accuracy in identifying plant diseases. Through testing, Cruz et al. [36] evaluated six ANN architectures showing great potential for image analysis for RGB images of the grapevine yellow disease. They proved that deep learning offers 35.97% and 22.88% better predictive values for recognizing the disease from a baseline system without deep learning and trained humans, respectively.

Generally, it is difficult to compare studies that present emerging techniques for crop disease detection, mainly because they use different

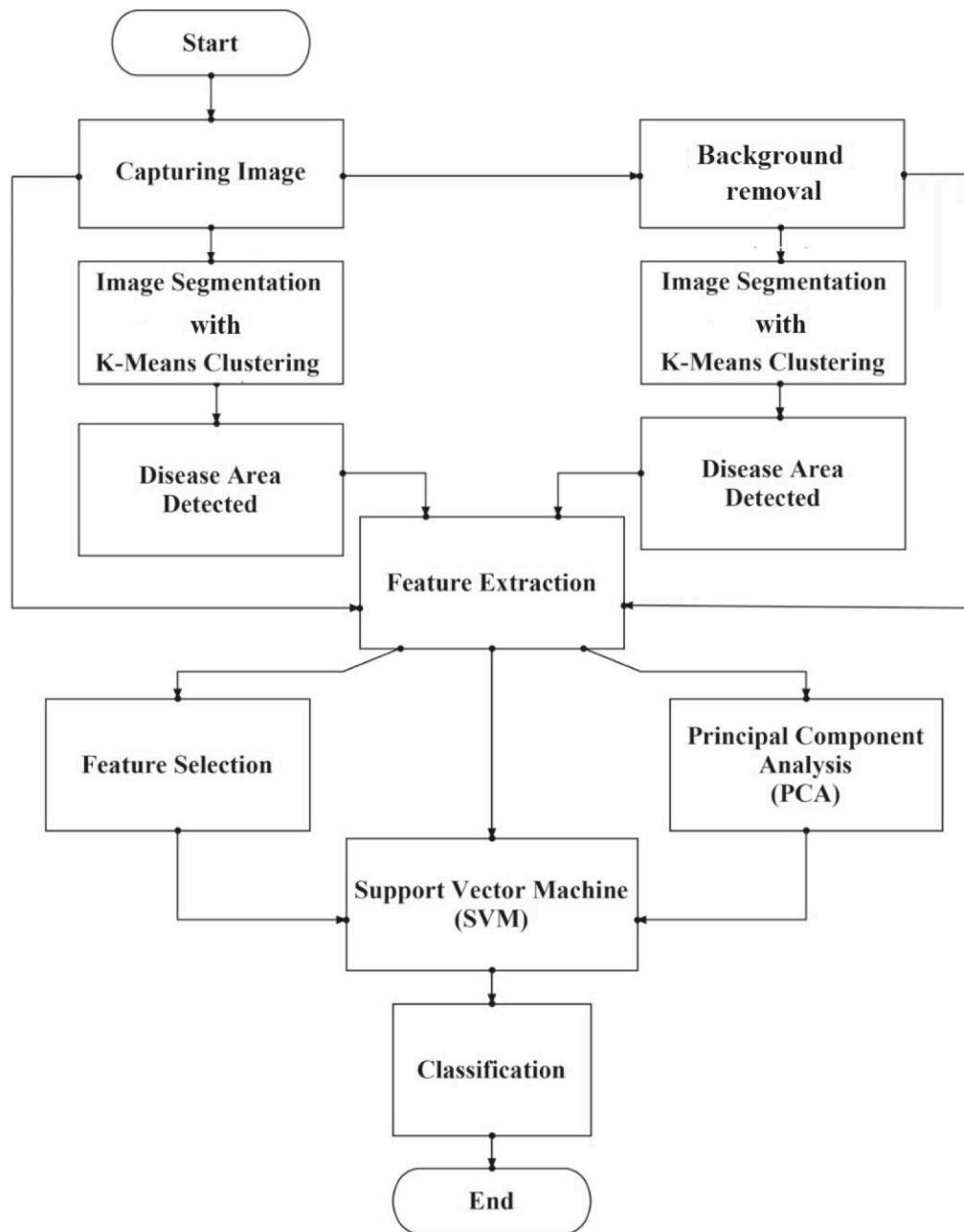


Fig. 2. Flowchart of the proposed methodology for disease detection and classification.

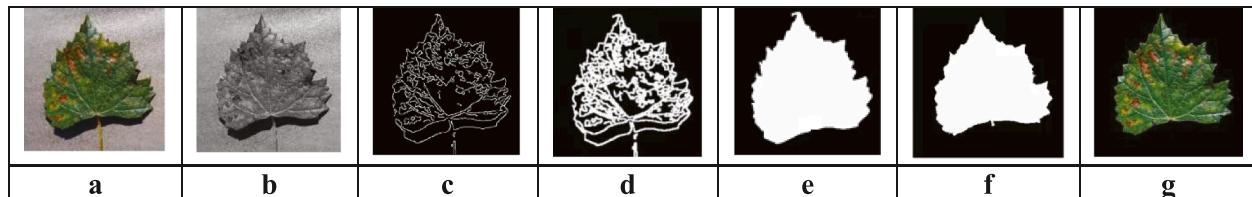


Fig. 3. (a) Image with shadow, (b) extracting green color band from the image, (c) edge detection, (d) morphological dilation to connect edges, (e) filling in the blanks in the edge image, (f) shadow removal based on thresholding, and (g) final image.

sensing systems, datasets, and evaluation metrics (Zhang et al., 2018). One of the limitations when using machine learning models for disease detection is the large and high-quality dataset required for model training and testing. Another challenge is the clustering sensitivity required to diagnose the diseased area(s) for accurate symptom-based

disease detection. This means detecting the region of interest (ROI) requires high precision and sensitivity to achieve the required application and accuracy. In previous research, to separate the diseased area from the leaf, it was necessary to determine the number of clusters before implementing the *K*-means clustering algorithm. Thus *K*-means divides

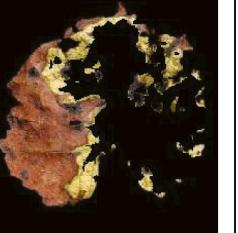
| Disease type | Captured image | <i>K</i> -means clustering of the captured image | Removed Background | <i>K</i> -means clustering of the DA with removed background |
|---------------|---|---|--|---|
| Black measles |  |  |  |  |
| Black rot |  |  |  |  |
| Leaf blight |  |  |  |  |
| Healthy |  |  |  |  |

Fig. 4. Preparing and separating the image from the background and detecting the DA.

the data set into a number of predefined clusters. Finding the optimal number of clusters is also a challenging task. Most current machine learning methods are not yet accurate, precise, and robust enough to be adopted for diagnosing plant diseases in the field and point-of-care situations. They also require a long time to diagnose leaf disease due to not separating either the diseased part or the background from the original image.

Furthermore, most machine learning and deep learning methods introduced cannot determine the most effective features extracted from images of diseased crops for disease classification. In fact, the machine learning and deep learning models are somehow black-box [37], and there is no clear explanation for better classification accuracies achieved when different methods/models are used [38,39]. One of the most critical challenges in disease detection is classifying a plant infected with multiple diseases (or disorders). Another challenge is identifying and classifying diseases with similar symptoms, especially in early disease development stages [40]. Therefore, it seems that a comprehensive

system should be developed to extract features for accurate disease diagnosis and classification, especially when comparing diseases and disorders with similar symptoms. This study aims to address this challenge by developing a methodology to classify multiple grape leaf diseases (black measles, black rot, leaf blight, and healthy leaves) utilizing RGB image processing-based feature extraction and multi-class SVM classification. The performance of the proposed method was compared with two well-known deep learning algorithms, i.e., CNN and GoogLeNet, to investigate the efficiency of the method.

2. Material and methods

In order to develop accurate image classifiers to diagnose plant diseases, a large verified data set of images of diseased and healthy plants is required. Until very recently, such a dataset did not exist, and even smaller datasets were not freely available. So, researchers had to create datasets by themselves [41]. In this study, the PlantVillage dataset, as a

Table 1

Description of the extracted features for disease detection and classification.

| Feature | Description | Formula | Reference |
|--|--|--|----------------------|
| Contrast | The measure of the difference between the brightness of the objects or regions and other objects within the same field of view | $\sum_{ij} i - j ^2 p(i, j)$ | Haralick et al. [46] |
| Correlation | The measure of degree and type of relationship between adjacent pixels | $\frac{\sum_{ij} (i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j}$ | Haralick et al. [46] |
| Energy | The sum of squared elements in the gray level co-occurrence matrix | $\sum_{ij} p(i, j)^2$ | Haralick et al. [46] |
| Homogeneity | the closeness of the distribution of elements in the GLCM | $\sum_{ij} \frac{p(i, j)}{1 + i - j }$ | Haralick et al. [46] |
| Mean | The measure of the average intensity value of the pixels present in the region | $\frac{1}{n} (\sum_{i=1}^n X_i)$ | Haralick et al. [46] |
| Standard Deviation | The measure of how much the gray levels differ from the mean | $\sqrt{\frac{1}{n} (\sum_{i=1}^n (X_i - \bar{X})^2)}$ | Haralick et al. [46] |
| Entropy | The measure of differences in gray levels | $E = \sum(p_i * \log_2(p_i))$ | Haralick et al. [46] |
| Root Mean Square (RMS) | The measure of root mean square value of an image | $X_{RMS} = \sqrt{\frac{1}{N} \sum_{n=1}^N X_n ^2}$ | Haralick et al. [46] |
| Variance | The measure of variance value of an image | $\frac{1}{n} (\sum_{i=1}^n (X_i - \bar{X})^2)$ | Haralick et al. [46] |
| Smoothness | A measure of relative smoothness of intensity in a region | - | Haralick et al. [46] |
| Kurtosis | A measure of peaks distribution related to the normal distribution | $K = \frac{E(x - \mu)^4}{\sigma^4}$ | Haralick et al. [46] |
| Skewness | A measure of asymmetry in a statistical distribution | $S = \frac{E(x - \mu)^3}{\sigma^3}$ | Haralick et al. [46] |
| Inventive Design Method (IDM) | The measure of the closeness of the distribution of GLCM elements to the GLCM diagonal | - | Haralick et al. [46] |
| Local binary patterns (LBP) | measure and extract the local texture information of images | $\sum_{i=0}^{p-1} s(n_i - G_C) 2^i$ $sx = \begin{cases} 1 & x > 0 \\ 0 & otherwise \end{cases}$ | Sairamya et al. [47] |
| Histogram of Oriented Gradients (HOG) | histograms of edge orientation | - | Zhou et al. [48] |
| Harris | An algorithm for extracting corners and inferring features of an image | - | Gong et al. [49] |
| Binary Robust Invariant Scalable Keypoints (BRISK) | An algorithm for feature point detection and description algorithm with scale invariance | - | Kashif et al. [50] |

Table 1 (continued)

| | | | |
|---------------------------------------|--|---|-------------------------|
| Oriented FAST and Rotated BRIEF (ORB) | and rotation invariance | - | Ma et al. [51] |
| MSER | An algorithm for extracting the number of co-variant regions in an image | - | Fouad et al. [52] |
| Hough | An algorithm to find imperfect instances of objects within a certain class of shapes by a voting procedure | - | Kazhagamani et al. [53] |

freely-available dataset, was used. The present study involved three types of plant diseases in addition to healthy leaves, which made four classes in total (Fig. 1).

2.1. Proposed algorithm

The proposed methodology applied in this work is depicted in Fig. 2. It includes (1) background removal, (2) image segmentation for detecting the disease symptoms (i.e., DA) using K -means clustering, (3) feature extraction, (4) feature selection, (5) feature dimension reduction, and finally, (6) multi-class SVM classification. The proposed methodology is described below in detail. This process was performed for two datasets, i.e., without background and with background, to check the classification accuracy and processing time. This operation was performed to determine the importance of the presence or absence of seismicity in image classification and reliable disease identification. For this purpose, the background will first be removed from the leaf to obtain two sets of data.

Background removal. This step involves removing the background to prevent any possible bias in the extracted properties and the trained framework. Color cast removal is performed by normalizing the gray values of three-color channels separately. The background removal task needs to be performed automatically, free from any human influence to increase its usability. It can be performed in two ways: pixel clustering and edge detection. In this study, a combination of both methods was used: gray-level thresholding for pixel clustering and the Canny method for edge detection. A critical challenge that occurs in images in real and natural conditions is the presence of shadows [42]. Fig. 3(a) shows a leaf with a shadow and a background. In the plant leaf images, the green color band is often dominant, so by extracting the green band, Fig. 3(b) can be obtained. Edge detection was then performed using the Canny algorithm, and then the pixel size of the edges was morphologically dilated to connect the borders. Figs. 3(c) and 3(d) show the edges extracted from the image. In the next step, the empty spaces in the image were filled. As shown in Fig. 3(e), the resulting image still had a shadow. In this step, by removing the shadow color pixels using gray-level thresholding, the shadow was omitted (Fig. 3f). The final image of a leaf without a background is shown in Fig. 3(g).

Image segmentation. Image segmentation is the process of dividing an image into several parts. There are various methods for image segmentation, ranging from simple thresholding methods to advanced color and frequency-domain image segmentation methods [6,43]. The Otsu algorithm and K -means method are among the reliable image segmentation processes for image thresholding due to their simple calculation process and methodology [44].

K -means clustering. The K -means clustering is used to classify objects based on a set of features into K number of classes. The classification of objects is done by minimizing the sum of the squares of the distance between the object and the corresponding cluster. The

Table 2

Grape disease classification results of the SVM method.

| | Input images | Features | Color bands | SVM Kernel Function | | | |
|----|---|--|-------------------|---------------------|------------|-----------------------|---------|
| | | | | Linear | Polynomial | Radial Basis Function | Sigmoid |
| 1 | Captured image | GLCM | RGB-HSV- L*a*b | 0.9472 | 0.5984 | 0.8553 | 0.779 |
| 2 | K-means clustering of the captured image | | | 0.9678 | 0.7082 | 0.9142 | 0.8848 |
| 3 | Images without background | | | 0.9781 | 0.7006 | 0.9168 | 0.8012 |
| 4 | K-means clustering of DA without background | | | 0.9871 | 0.7425 | 0.9197 | 0.8922 |
| 5 | Captured image | HOG | RGB-HSV- L*a*b | 0.7622 | 0.4677 | 0.6134 | 0.5367 |
| 6 | K-means clustering of the captured image | | | 0.7931 | 0.3501 | 0.6803 | 0.5776 |
| 7 | Images without background | | | 0.7815 | 0.5284 | 0.6862 | 0.5696 |
| 8 | K-means clustering of DA without background | | | 0.8860 | 0.3719 | 0.8422 | 0.7979 |
| 9 | Captured image | LBP | RGB-HSV- L*a*b | 0.8345 | 0.3640 | 0.6525 | 0.5683 |
| 10 | K-means clustering of the captured image | | | 0.6008 | 0.3467 | 0.4605 | 0.3964 |
| 11 | Images without background | | | 0.8595 | 0.4098 | 0.6891 | 0.6422 |
| 12 | K-means clustering of DA without background | | | 0.7107 | 0.3411 | 0.4618 | 0.4698 |
| 13 | Captured image | Harris, BRISK, ORB, MSER, Hough (line length and point number) and Histogram | RGB-HSV- L*a*b | 0.3254 | 0.3254 | 0.3254 | 0.3254 |
| 14 | K-means clustering of the captured image | | | 0.3045 | 0.3045 | 0.3045 | 0.3045 |
| 15 | Images without background | | | 0.2937 | 0.2937 | 0.2937 | 0.2937 |
| 16 | K-means clustering of DA without background | | | 0.2965 | 0.2965 | 0.2965 | 0.2965 |

algorithm for K -means clustering can be described by these steps: (1) picking the center of K -th cluster either randomly or based on some heuristics; (2) assigning each pixel to a cluster that minimizes the distance between the pixel and the cluster center; (3) computing the cluster centers by averaging all the pixels in the cluster; and finally, (4) repeating steps 2 and 3 until convergence is obtained [45]. Generally, in the mentioned steps, the selection of the value of K and the selection of ROI are performed manually, which depends on the user's skill, and sometimes ROI might not be selected correctly by the user. This means that for each number of images in the database, the ROI number must be manually selected to determine the desired area of the disease, so this is very time-consuming and error-prone. Therefore, automatic clustering can be useful for the automatic diagnosis of the disease area in the plant leaves. In this study, K -means clustering was used to automatically separate the disease symptoms from the "healthy" areas of a leaf. Therefore, the ROI in the K -means clustering, which had to be selected by the user in other methods, was determined automatically. This was done by thresholding between the color of the disease area (i.e., symptoms) and the color of the healthy leaf area. For this purpose, in the leaf image, the pixels in which the red color is less than the blue and green values were masked out. Therefore, in the rest of the image, only the diseased area of the leaf will remain. Fig. 4 shows the steps of preparing and separating the image from the background and detecting the DA of grape leaves.

Feature extraction. An image contains a remarkable amount of information, and usually, only some information is needed to distinguish objects in images, for example, texture, color, and shape [5]. In this study, several features such as gray-level co-occurrence matrix (GLCM), local binary patterns (LBP), histogram of oriented gradients (HOG), Harris corner detection, binary robust invariant scalable key-points (BRISK), oriented fast and rotated BRIEF (ORB), maximally stable external regions (MSER), Hough transform data, in three bands, i.e., RGB, HSV, and L*a*b were extracted. Color bands vary in color, brightness, and composition. The algorithms used in the field of machine vision are different in that they are based on the properties of color,

brightness, and composition. For example, an image-based algorithm distinguishes between different sections and thresholds based on color values; therefore, in this research, the three color bands of RGB, HSV, and L*a*b have been used. A description of the extracted features is given in Table 1.

The features were categorized into four groups: (1) GLCM; (2) HOG; (3) LBP; and (4) Harris, BRISK, ORB, MSER, and Hough (line length and point number). The features were extracted from four groups of images, including the captured image (Group 1), K -means clustering of the captured image (Group 2), images without background (Group 3), and K -means clustering of the DA without background (Group 4) (Fig. 4). Finally, each group of features that were extracted was examined more carefully and the most important and best features will be selected from that group.

Principal component analysis (PCA). PCA is a dimensionality reduction that is often used to reduce the dimension of the variables of a larger dataset that is compressed to the smaller one, which contains most of the information to build an efficient model. The justification behind the use of PCA is its versatility and simplicity in real-time implementations. PCA helps to efficiently remove correlated features in a dataset, resulting in the components being independent of one another, which further reduces training time when subjected to machine learning algorithms. Also, since the most significant features are selected efficiently, PCA reduces the chances of overfitting. The high dimensional dataset is converted into low dimensions when PCA is used, and hence, it becomes easy to visualize its nature in a 2D plot and derive inferences [54,55]. Therefore, the data were analyzed in two stages: one after feature extraction and the other after feature selection so that the results can be compared.

Feature selection. Feature selection aims to better classification results by listing significant features, which helps in reducing computational overload. The reason behind the feature selection is that classifiers trained on reduced feature space are more robust and reproducible than classifiers constructed on the original large feature space. In feature selection, the most important feature(s) for predicting a

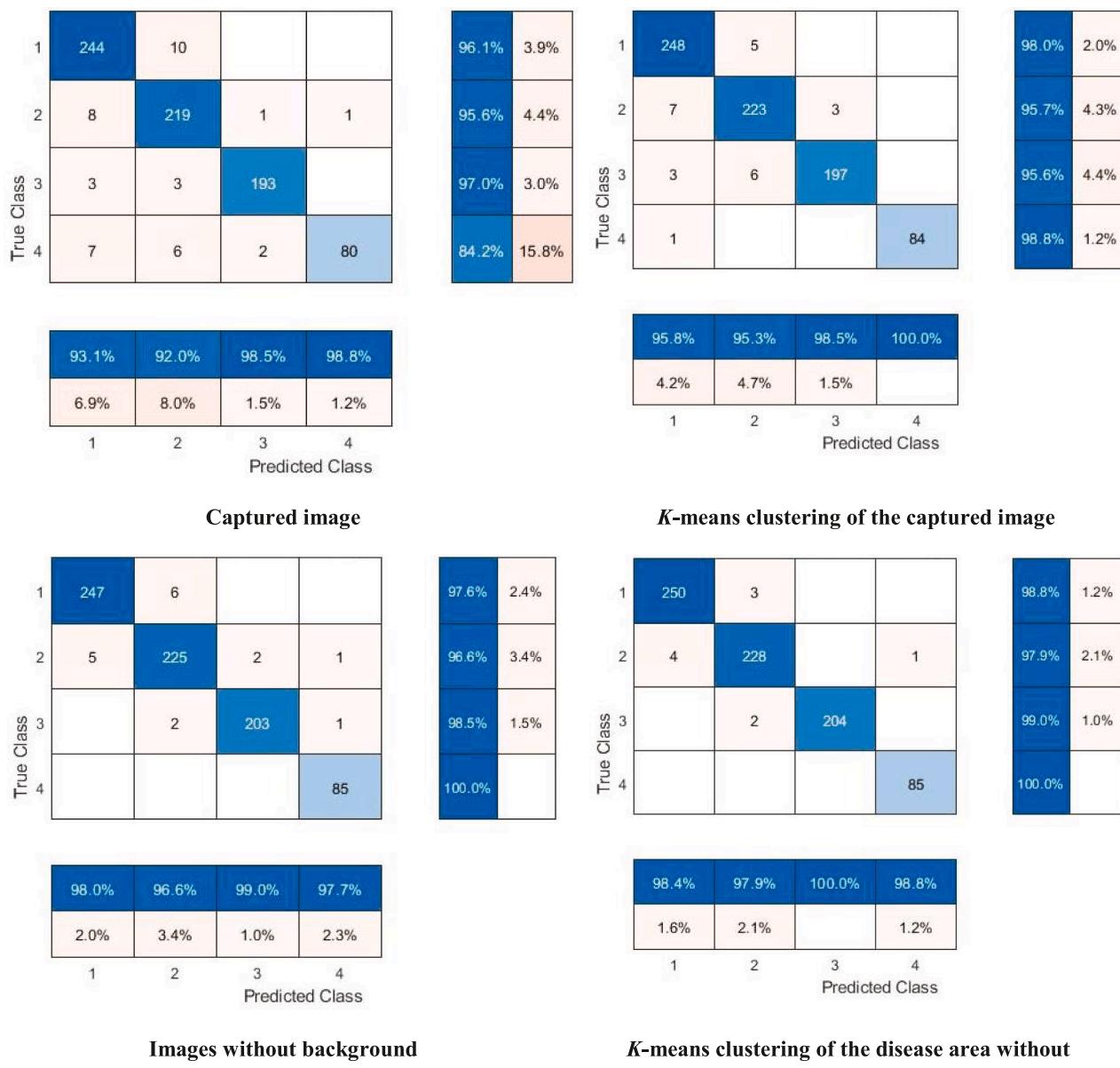


Fig. 5. Confusion matrix for grape leaf disease detection utilizing SVM linear kernel function.

variable (e.g., disease, yield, etc.) is identified [56]. The features which do not provide helpful information are called irrelevant features, and the features which do not provide more information than the currently selected features are called redundant features [57]. The features that are unrelated or uncorrelated to class variables are called noise, introducing bias in prediction and reducing classification performance. Hence, noise should be handled to improve prediction performance, and it can be made possible with dimensionality reduction. It can be achieved by either feature extraction or feature selection [58]. Many feature selection methodologies have been proposed, and research continues to support the claim that there is no universal “best” method for all tasks [54]. In this research, the Relief feature selection method was used to select the best and essential features for diagnosing grape leaf diseases.

Relief feature selection. In the Relief feature selection method, at each step and randomly, a sample is selected from the samples in the data set. Then, the degree of relevance of each feature is updated based on the difference between the selected sample and the two neighboring samples. If one of the selected sample properties differs from a similar

feature in the neighboring sample of the same class (hit sample), the score of this feature decreases. On the other hand, if the same feature in the selected instance differs from the similar feature in the neighboring instance of the opposite class (miss sample), the score of this feature increases.

Support Vector Machine (SVM). Various machine learning approaches, e.g., neuron-based and kernel-based methods, help predict and classify various diseases from the input plant leaf images of different plants [59]. SVM is a widely used supervised learning classification algorithm that utilizes various types of kernel functions. In this study, several kernel functions such as linear, polynomial, radial basis function (RBF), and sigmoid functions were investigated to diagnose and classify diseased and healthy leaves, and a confusion matrix was created for each result. Finally, the best algorithm was selected in terms of efficiency and accuracy for the diagnosis and classification of grape diseases.

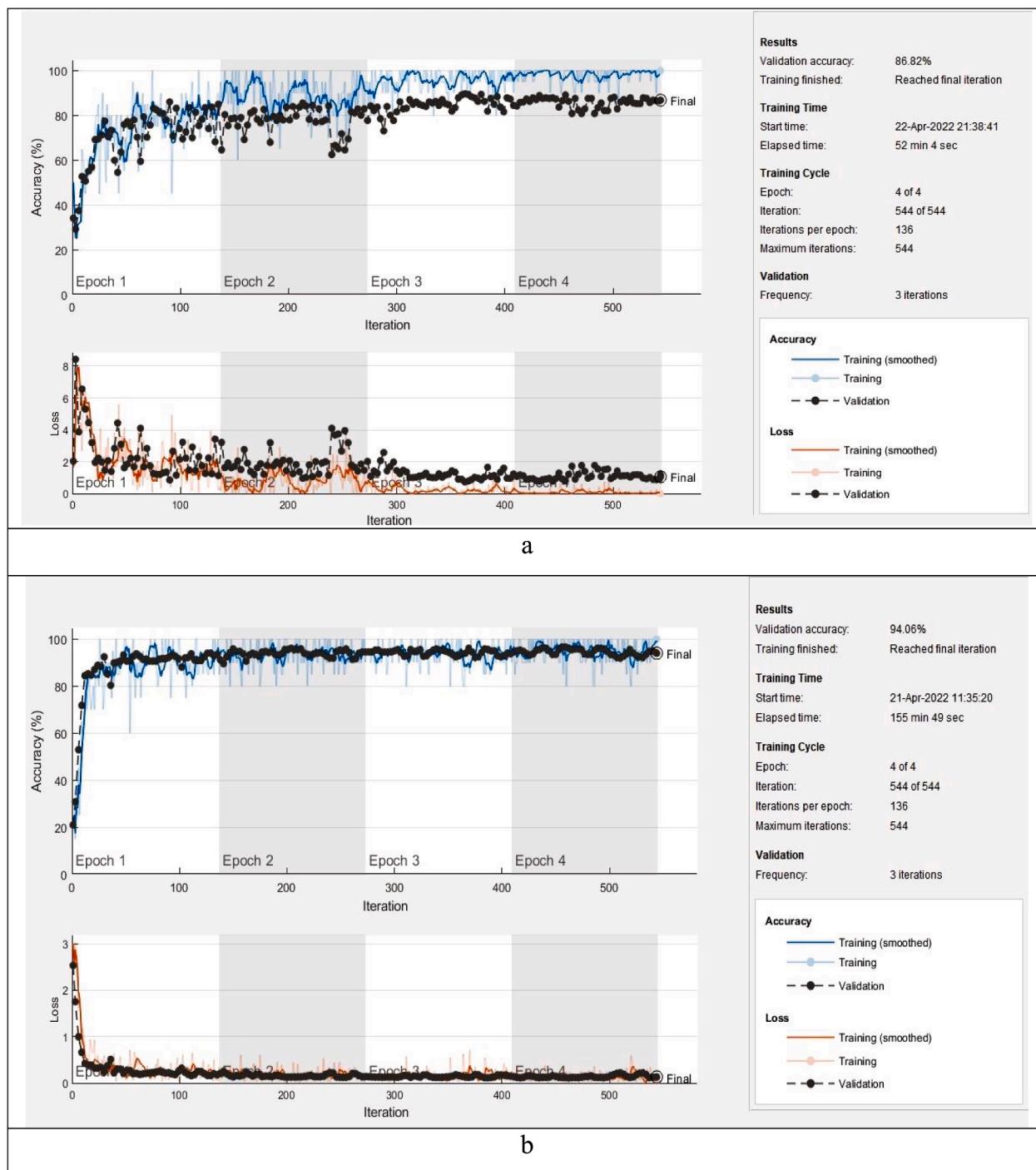


Fig. 6. Training plot for (a) CNN; (b) GoogleNet.

2.2. Deep learning methods

In this study, the proposed machine learning method was compared with two deep learning-based classifiers, i.e., CNN and GoogleNet, in terms of classification accuracy and duration of disease diagnosis. CNN and GoogleNet are freely accessible on the GitHub website (www.github.com). They implemented in MathWorks MATLAB R2010b programming environment considering the default values for the parameters of learner layers.

2.3. Performance evaluation indicators

Confusion matrices of the prediction process were used to evaluate the success and efficiency of disease diagnosis systems. A confusion matrix is a summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broken down by each class. In this study, accuracy was used to evaluate the performance of models. The accuracy of a classification method on test data is the percentage of observations of the test set that is correctly categorized by the model used (Eq. (1))

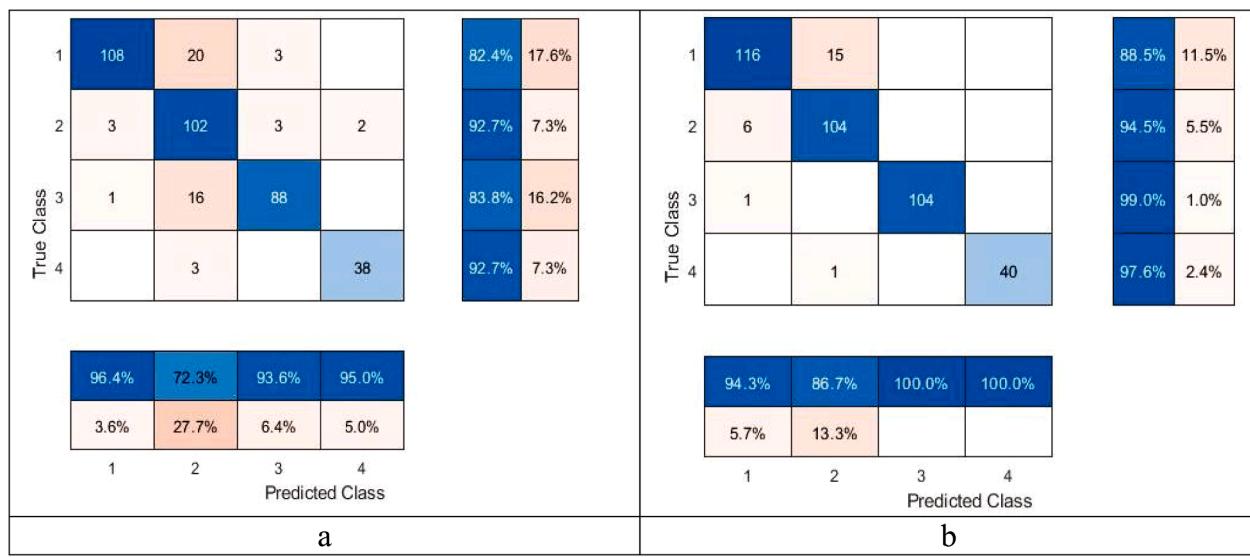


Fig. 7. Confusion matrix of deep learning methods (a) CNN, (b) GoogLeNet.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. The processing time reported in this study was based on the time consumed for five-fold cross-validation.

3. Results

3.1. Classification results using support vector machine

After extracting the features for four groups of images, including captured image, K -means clustering of the captured image, images without background, and K -means clustering of the disease area without background, the results of Table 2 were obtained using SVM. In this table, for the GLCM feature, a feature group (contrast, correlation, energy, homogeneity, mean, standard deviation, entropy, RMS, variance, smoothness, kurtosis, skewness, and IDM) was extracted from each group of images. Moreover, another group of features, including LBP and HOG features, and finally, Harris, BRISK, ORB, MSER, and Hough (line length and point number) feature groups were extracted from the images. In order to classify with the SVM, four kernel functions (linear, polynomial, radial basis function, and sigmoid) were used to obtain the classification accuracy for each of the feature groups.

According to Table 2, the highest accuracies were obtained for the GLCM features and the SVM with linear kernel function, which were 0.9472, 0.9678, 0.9781, and 0.9871 for the raw captured image, K -means clustering of the captured image, Images without background, and K -means clustering of DA without background, respectively (first to fourth rows of Table 2). The processing time for training of these algorithms were 365.22, 337.68, 267.78, and 187.8 s, respectively. The processing time obtained shows that the algorithm, after separating the background and the disease area from the original image, can achieve high leaf disease detection accuracy in a very short time. It can also be observed that removing the background from the raw image increases the algorithm's accuracy. This might be due to the fact that using K -means clustering increased the ability to diagnose the disease. The

highest accuracy was achieved by the “K-means clustering of DA without background” algorithm (Row 4 in Table 2), which removes both the background and the DA from the leaf. The processing time of disease diagnosis in this algorithm (187.8 s) was significantly less than the others (267-365 s).

Fig. 5 depicts the confusion matrices of this classification. In this matrix, the rows represent the true class of the samples, while each column represents the predicted classes. The diagonals show the number of samples which have been classified correctly. The class numbers for Black measles, Black rot, Leaf blight diseases were 1, 2 and, 3, respectively, and for Healthy leaves were 4. In this experiment, 2,861 images were used for training and 1024 images for testing.

3.2. Classification results with Convolutional Neural Network (CNN) and GoogleNet

To classify the three groups of disease and healthy leaf data, two deep learning classification method was used to be compared with the method introduced in this study. The prediction accuracy using these deep learning models, i.e., CNN and GoogleNet, were 0.8682 and 0.9405, respectively. The learning rate remained constant for the entire training of all the models. The training plot for the two models is shown in Fig. 6. The time taken for the training of the dataset using CNN and GoogleNet were approximately 53 and 156 min, respectively. Fig. 7 depicts the confusion matrices of the deep learning models in the classification task. According to the results, not only the prediction accuracy of deep learning methods was not favorable, but they are also highly time-consuming.

3.3. Result of feature dimension reduction

In this study, the PCA and SVM methods were used to classify grape leaf diseases based on the extracted color features, shape features, and texture features from disease images and their combined features. After reducing the dimensions of the data using PCA, acceptable prediction accuracies for image recognition could be obtained using the SVM as a classifier. Table 3 shows the results of the PCA of the classified output of grape leaves using SVM. The highest accuracies were obtained for the GLCM features, which were 0.9601, 0.9704, 0.9871, and 0.9897 (first to

Table 3

Grape disease classification result of the PCA and SVM combination.

| Number | Input images | Features | Color Bands | Kernel Function |
|--------|---|---|---------------|-----------------|
| 1 | Captured image | GLCM | RGB-HSV-1*a*b | Linear 0.9601 |
| 2 | K-means clustering of captured image | | | 0.9704 |
| 3 | Images without background | | | 0.9871 |
| 4 | K-means clustering of the disease area without background | | | 0.9897 |
| 5 | Captured image | HOG | RGB-HSV-1*a*b | 0.8077 |
| 6 | K-means clustering of the captured image | | | 0.8208 |
| 7 | Images without background | | | 0.8237 |
| 8 | K-means clustering of the disease area without background | | | 0.8891 |
| 9 | Captured image | LBP | RGB-HSV-1*a*b | 0.8486 |
| 10 | K-means clustering of the captured image | | | 0.6564 |
| 11 | Images without background | | | 0.8620 |
| 12 | K-means clustering of DA without background | | | 0.6577 |
| 13 | Captured image | Harris, BRISK, ORB, MSER, Hough (line length and point number), and Histogram | RGB-HSV-1*a*b | 0.4604 |
| 14 | K-means clustering of the captured image | | | 0.2942 |
| 15 | Images without background | | | 0.3318 |
| 16 | K-means clustering of the disease area without background | | | 0.3326 |

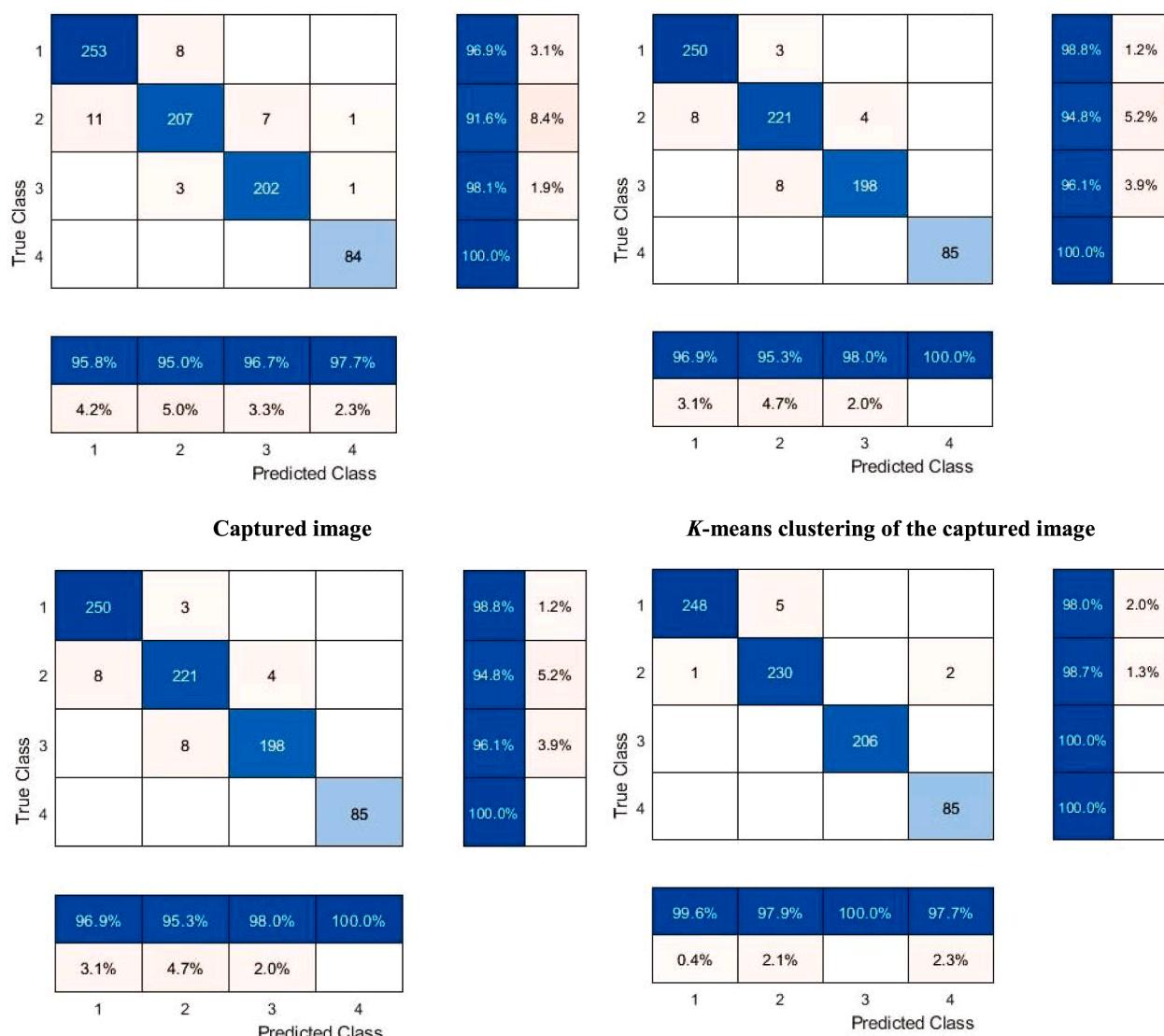


Fig. 8. Confusion matrix for grape leaf diseases detection utilizing PCA and SVM linear kernel function.

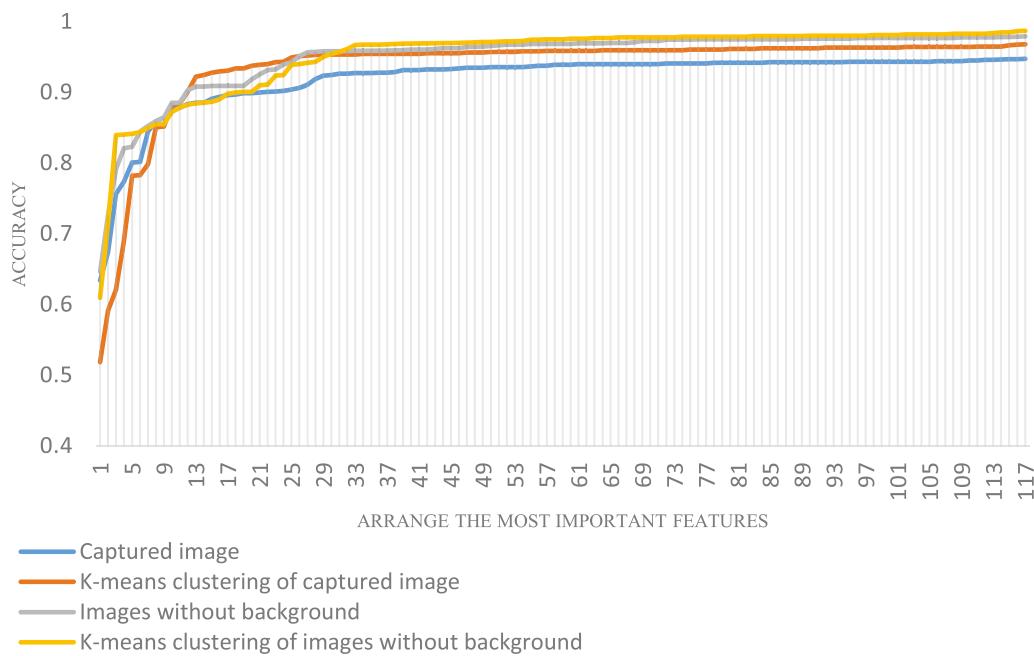


Fig. 9. The order of the selected features for various image groups for disease detection.

fourth rows of Table 3), with processing times of 351.02, 260.28, 120.69, and 96.38 s, respectively. Fig. 8 depicts the confusion matrices of the classification. As the results showed, reducing the dimensions of the feature data extracted from the images of grape leaf diseases could reduce the running time of the classification, and acceptable detection results could be obtained. In the practical application, PCA could be used to reduce the dimensions of the data extracted from the leaf disease images, and then SVM classification could be constructed for leaf disease detection. The method used in this study could also be used for the image recognition of other plant leaf diseases.

3.4. Results of the feature selection

As can be seen in Table 3, the highest accuracy for diagnosing and classifying grape leaf diseases belongs to the use of the GLCM features group. Therefore, in this section, the most effective/important features for disease classification are presented. Among the 117 features extracted from the images in four groups, the most effective/important features that played a role in diagnosing the disease were selected by the Relief method. Then, SVM was used again for disease detection using only the important features. Fig. 9 shows the order of the selected features. The accuracy of the algorithms that used *K*-means clustering is higher than the other algorithms. It was also found that the algorithm which removes the background of the leaf image and can separate the DA from the leaf was more accurate. In the next step, some of the most important features were selected. The number of these features was obtained based on reaching a prediction performance with a maximum of 2% difference from the highest accuracy. Table 4 lists the number of these features for each group, as well as the accuracy of diagnosing the disease for the same number separately. For groups 1, 3, 2 and 4 the accuracies were 0.9282, 0.9484, 0.9585, and 0.9673, respectively. The estimated time to reach the accuracy of 0.9673 using the most important features shown in Table 4 was 54.6 s, representing a 60% reduction in time compared to the use of all 117 extracted features. An essential advantage of this technique, compared to previous studies, is that by developing an automatic clustering algorithm and also by removing the

background from the original image, the algorithm was able to accurately diagnose the disease in less time. This can be seen in the yellow line compared to other lines (Fig. 9). This diagram shows that the proposed algorithm with a lower number of features has the ability to diagnose the disease with high accuracy in a faster time.

4. Discussion

Non-destructive methods, such as image processing, are emerging solutions for the intelligent diagnosis and classification of plant diseases. Although biochemical methods at the molecular level can accurately detect plant diseases, they are extremely labor- and time-consuming [60–62]. When evaluating disease classification techniques, the type and number of diseases that can be detected and the ability to diagnose diseases with similar symptoms should be considered. Since the goal of a disease detection algorithm is to identify a pattern for achieving the highest possible classification rate for disease diagnosis, factors such as the accuracy of diagnosis and classification of diseases, the shortest possible time to diagnose a disease (for real-time applications), and the determination of the important features, extracted from the collected images, for disease detection should be compared and discussed too. Table 5 compares several previous methods with this study for grape leaf disease detection, presenting detection accuracies and limitations. According to the table, limitations such as low accuracy and long diagnosis time, the number of diseases classified, uncommonness and similarity of diseases, the uncertainty of the extracted features, and selection of the most important features can be seen in previous research. Although deep learning techniques, which are black-box methods (the inputs and operations are not visible to the user), might result in promising performance, they require a high number of training samples, and the effective features for disease diagnosis are not specified [63]. Therefore, an accurate deterministic algorithm is considered in this study for diagnosing and classifying diseases with similar symptoms in grapes. The proposed novel algorithm was compared with two deep learning methods, CNN and GoogleNet, which are widely used methods in the classification of plant leaf diseases, and the results showed that the algorithm proposed

Table 4

The best and most important selected features for the four groups of algorithms.

| Number | Feature | Color Band | Feature | Captured image | | K-means clustering of the captured image | Image without background | K-means clustering of the disease area without background |
|--------|--------------------|---------------|--------------------|----------------|--------------------|--|--------------------------|---|
| | | | | Accuracy | 0.9282 | | | |
| 1 | Energy Homogeneity | HSV (H) | Skewness | l^*a^*b (a) | Energy | HSV (H) | Skewness | l^*a^*b (a) |
| 2 | | HSV (H) | Standard Deviation | l^*a^*b (a) | Entropy | HSV (H) | Kurtosis | l^*a^*b (a) |
| 3 | Kurtosis | l^*a^*b (a) | Entropy | l^*a^*b (a) | Homogeneity | HSV (H) | Standard Deviation | l^*a^*b (a) |
| 4 | Standard Deviation | l^*a^*b (a) | Mean | l^*a^*b (a) | Kurtosis | RGB (R) | Correlation | l^*a^*b (b) |
| 5 | Variance | l^*a^*b (a) | Variance | l^*a^*b (a) | Variance | l^*a^*b (a) | Correlation | RGB (B) |
| 6 | Contrast | l^*a^*b (b) | Skewness | l^*a^*b (b) | Kurtosis | RGB (B) | Variance | l^*a^*b (a) |
| 7 | Variance | HSV (H) | Energy | HSV (S) | Contrast | l^*a^*b (a) | Correlation | RGB (G) |
| 8 | Entropy | HSV (S) | Energy | HSV (H) | Skewness | RGB (B) | Correlation | RGB (R) |
| 9 | Mean | l^*a^*b (a) | Contrast | l^*a^*b (l) | Skewness | RGB (R) | Correlation | HSV (V) |
| 10 | Mean | HSV (H) | Contrast | HSV (H) | Kurtosis | l^*a^*b (a) | Contrast | l^*a^*b (a) |
| 11 | Contrast | HSV (H) | Contrast | l^*a^*b (a) | Standard Deviation | l^*a^*b (a) | Smoothness | l^*a^*b (a) |
| 12 | Energy | HSV (S) | Kurtosis | l^*a^*b (a) | Kurtosis | l^*a^*b (b) | Skewness | RGB (R) |
| 13 | Entropy | l^*a^*b (b) | Energy | l^*a^*b (b) | Kurtosis | RGB (G) | Skewness | HSV (V) |
| 14 | Contrast | l^*a^*b (a) | Homogeneity | l^*a^*b (l) | Kurtosis | l^*a^*b (l) | Mean | l^*a^*b (a) |
| 15 | Homogeneity | l^*a^*b (a) | Standard Deviation | l^*a^*b (b) | Kurtosis | HSV (V) | Smoothness | RGB (R) |
| 16 | Standard Deviation | HSV (H) | Entropy | HSV (H) | Correlation | RGB (R) | Smoothness | HSV (V) |
| 17 | Skewness | l^*a^*b (b) | Homogeneity | l^*a^*b (a) | Standard Deviation | HSV (H) | Homogeneity | l^*a^*b (b) |
| 18 | Kurtosis | l^*a^*b (b) | Skewness | HSV (S) | Skewness | l^*a^*b (b) | Skewness | RGB (B) |
| 19 | RMS | HSV (H) | Standard Deviation | HSV (S) | Variance | HSV (H) | Energy | HSV (H) |
| 20 | Entropy | l^*a^*b (a) | Entropy | l^*a^*b (B) | Standard Deviation | HSV (S) | Smoothness | RGB (B) |
| 21 | Energy | HSV (H) | Mean | HSV (H) | Correlation | l^*a^*b (a) | Skewness | RGB (G) |
| 22 | Energy | l^*a^*b (b) | Energy | l^*a^*b (a) | Mean | l^*a^*b (a) | Smoothness | HSV (H) |
| 23 | Correlation | HSV (S) | IDM | l^*a^*b (a) | Kurtosis | HSV (S) | Correlation | l^*a^*b (a) |
| 24 | Correlation | l^*a^*b (l) | Kurtosis | l^*a^*b (b) | Energy | l^*a^*b (a) | Kurtosis | HSV (S) |
| 25 | Kurtosis | RGB (B) | Correlation | HSV (S) | Contrast | l^*a^*b (b) | Homogeneity | l^*a^*b (l) |
| 26 | Skewness | RGB (B) | - | - | IDM | l^*a^*b (a) | Kurtosis | RGB (R) |
| 27 | Homogeneity | l^*a^*b (b) | - | - | Homogeneity | HSV (V) | Kurtosis | HSV (V) |
| 28 | Entropy | RGB (B) | - | - | Standard Deviation | RGB (B) | Energy | l^*a^*b (a) |
| 29 | IDM | l^*a^*b (a) | - | - | Homogeneity | RGB (R) | Entropy | l^*a^*b (b) |
| 30 | Correlation | l^*a^*b (a) | - | - | Homogeneity | RGB (G) | Skewness | HSV (H) |
| 31 | Kurtosis | HSV (S) | - | - | - | - | Standard Deviation | RGB (R) |
| 32 | IDM | HSV (H) | - | - | - | - | Smoothness | HSV (S) |
| 33 | Skewness | HSV (S) | - | - | - | - | Kurtosis | HSV (H) |
| 34 | - | - | - | - | - | - | Skewness | HSV (S) |

in this study had a higher detection accuracy with a less processing speed. The advantages of the proposed algorithm are as follows:

- using the automatic ROI selection to cluster the disease area, there is no need for user input at the time of segmentation. The proposed method is fully automated, while other methods require user input to select the best input image segmentation;
- increasing the classification accuracy using the proposed algorithm;

- having the ability to develop detection algorithms for more diseases as well as classifying diseases accurately with similar symptoms;
- identifying the most effective features in diagnosing and classifying diseases;
- reducing the time for algorithm training;
- achieving higher accuracy and reduced classification time compared to the CNN and GoogleNet deep learning methods;
- and reducing the diagnosis time using the proposed algorithm.

Table 5

A comparison of the methods, accuracy, and limitations of previous research and the present study for grape leaf diseases.

| Diseases | Methods | Accuracy | Limitation | Reference |
|--|--|-----------------------------------|--|-------------------------|
| Downy, Powdery | K-Means Clustering - SVM Classifier | 88.89% | Low number of diseases, low detection accuracy, does not extract all the features from the image, does not select the most effective features, common diseases in grape leaves have not been studied | Padol et al. (2016) |
| Black Rot, Black Measles, and Leaf Blight | SVM Classifier | 93% | It does not extract all the features from the images, does not select the most effective features | Jaisakthi et al. (2019) |
| Black Measles | SVM Classifier | 97. 2% | A low number of diseases does not extract all the features from the image and also does not specify the most effective features extracted | Singh et al. (2020) |
| Black Rot, Black Measles, and Isariopsis | AlexNet GoogLeNet ResNet-18 | 95.65% 92.29% | Using a black-box method, being time-consuming, does not select the most effective features | Lauguico et al. (2020) |
| Anthracnose, Brown Spot, Mites, Black Rot, Downy Mildew, and Leaf Blight | DICNN - CNN | 89.49% 97.22% | Using a black-box method, being time-consuming, does not select the most effective features | Liu et al. (2020) |
| Black rot, Esca measles, Leaf spot | GAN - CNN | 98.70% | Using a black-box method, being time-consuming, not select the most effective features | Liu et al. (2020) |
| Black Rot, Esca and Isariopsis Leaf Spot | UnitedModel - CNN | 99.17% | Using a black-box method, being time-consuming, not selecting the most effective features | Ji et al. (2020) |
| Black Rot, Black Measles, Leaf Blight, and Mites | DR-IACNN | 81.1% | Low detection accuracy, uses a black-box method, being time-consuming, does not select the most effective features | Xie et al. (2020) |
| Black Rot, Black Measles, and Leaf Blight | AlexNet MobileNet MobileNet ShuffleNet ShuffleNet | 99.01% 95.34% 96.69% 97.79% 94.8% | In some cases, it uses a black-box method, being time-consuming, does not select the most effective features | Tang et al. (2020) |
| Black Rot, Black Measles, and Leaf Blight | K-Means Clustering - GLCM Feature Extraction- SVM Classifier-PCA- Relief feature selection CNN GoogleNet | 98.97% 86.82% 94.05% | Compared to other methods used, the algorithm method used in this research includes the following advantages: Accurate diagnosis of the disease area Use automated clustering (ROI) Low operating time for diagnosis and classification of diseases Identify the best and most important features to diagnose the disease High accuracy for diagnosing and classifying diseases Higher accuracy for classification compared to CNN and GoogleNet deep learning methods | This Study |

5. Conclusion

In this study, to diagnose and classify grape leaf diseases, a new technique is proposed. Features of grape images, categorized in different groups, have been extracted and compared with each other for disease detection. The SVM classifier with linear kernel and using only the GLCM features resulted in the highest accuracy of 0.9472, 0.9678, 0.9781, and 0.9871 for the captured image, K-means clustering of the captured image, images without background, and K-means clustering of images without background, respectively. After using PCA for the feature dimension reduction of the same input images, the accuracy reached 0.9601, 0.9704, 0.9871, and 0.9897, respectively. In order to find the most effective features to diagnose the diseases, the Relief feature selection method showed that about thirty basic features are among the most important features. The proposed technique was compared to two deep learning methods (i.e., CNN and GoogLeNet), which achieved lower classification accuracies of 0.8682 and 0.9405, respectively. The results show that the algorithm developed in this study can detect plant leaf diseases in grapes with high accuracy in a short time.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] K. Asefpour Vakilian, J Massah, A farmer-assistant robot for nitrogen fertilizing management of greenhouse crops, *Comput. Electron. Agric.* 139 (2017) 153–163.
- [2] V.K. Vishnoi, K. Kumar, B. Kumar, Plant disease detection using computational intelligence and image processing, *J. Plant Dis. Prot.* 128 (2020) 19–53.
- [3] J. Abdulridha, Y. Ampatzidis, S.C. Kakarla, P. Roberts, Detection of target spot and bacterial spot diseases in tomato using UAV-based and benchtop-based hyperspectral imaging techniques, *Precis. Agric.* 21 (2020) 955–978.
- [4] J. Abdulridha, Y. Ampatzidis, J. Qureshi, P. Roberts, Laboratory and UAV-based identification and classification of tomato yellow leaf curl, bacterial spot, and target spot diseases in tomato utilizing hyperspectral imaging and machine learning, *Remote Sens.* 12 (17) (2020) 2732.
- [5] T. Ojala, M. Pietikinen, T. Maenp, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *Proc. IEEE Trans. Pattern Anal. Mach. Intell.* 24 (2002) 971–987.
- [6] V. Singh, A.K. Misra, Detection of plant leaf diseases using image segmentation and soft computing techniques, *Inf. Process. Agric.* 4 (1) (2017) 41–49.
- [7] H. Al Hiary, S. Bani Ahmad, M. Reyalat, M. Braik, Z ALRahamneh, Fast and accurate detection and classification of plant diseases, *Int. J. Comput. Appl.* 17 (1) (2011) 31–38.
- [8] L.S. Puspha Annabel, T. Annapoorani, P. Deepalakshmi, Machine learning for plant leaf disease detection and classification – a review, in: Proceedings of 2019 International Conference on Communication and Signal Processing (ICCSP), 2019.
- [9] S. Raut, A. Fulsunge, Plant disease detection in image processing using MATLAB, *Int. J. Innov. Res. Sci. Eng. Technol.* 6 (6) (2017).
- [10] L. Costa, Y. Ampatzidis, C. Rohla, N. Maness, B. Cheary, L. Zhang, Measuring pecan nut growth utilizing machine vision and deep learning for the better understanding of the fruit growth curve, *Comput. Electron. Agric.* 181 (2021), 105964.
- [11] J.J. Lopez, E. Aguilera, M. Cobos, Defect detection and classification in citrus using computer vision, *Lect. Notes Comput. Sci.* (2009) 11–18.
- [12] X. Zhou, W.S. Lee, Y. Ampatzidis, Y. Chen, N. Peres, C. Fraisse, Strawberry maturity classification from UAV and near-ground imaging using deep learning, *Smart Agric. Technol.* 1 (2021), 100001.
- [13] M. Bhange, H.A. Hingoliwala, Smart farming: pomegranate disease detection using image processing, *Procedia Comput. Sci.* 58 (2015) 280–288.
- [14] J. Hariharan, J. Fuller, Y. Ampatzidis, J. Abdulridha, A. Lerwill, Finite difference analysis and bivariate correlation of hyperspectral data for detecting Laurel wilt disease and nutritional deficiency in avocado, *Remote Sens.* 11 (15) (2019) 1748.
- [15] M.A. Ebrahimi, M.H. Khoshtaghaza, S. Minaei, B Jamshidi, Vision-based pest detection based on SVM classification method, *Comput. Electron. Agric.* 137 (2017) 52–58.
- [16] V. Partel, L. Nunes, P. Stansley, Y. Ampatzidis, Automated vision-based system for monitoring Asian citrus psyllid in orchards utilizing artificial intelligence, *Comput. Electron. Agric.* 162 (2019) 328–336.
- [17] K. Thenmozhi, U.S. Reddy, Image processing techniques for insect shape detection in field crops, in: Proceedings of 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, India, 2017, 23–24 November 2017.

- [18] G.L. Rinblat, L.C. Uzal, M.G. Larese, P.M. Granitto, Deep learning for plant identification using vein morphological patterns, *Comput. Electron. Agric.* 127 (2016) 418–424.
- [19] M. Seeland, M. Rzanny, D. Boho, J. Wälldchen, P. Mäder, Image-based classification of plant genus and family for trained and untrained plant species, *BMC Bioinf.* 20 (1) (2019).
- [20] Y. Onishi, T. Yoshida, H. Kurita, T. Fukao, H. Arihara, A. Iwai, An automated fruit harvesting robot by using deep learning, *ROBOMECH J.* 6 (1) (2019).
- [21] M. Dadashzadeh, Y. Abbaspour-Gilandeh, T. Mesri-Gundoshmian, S. Sabzi, J. L. Hernández-Hernández, M. Hernández-Hernández, J.I. Arribas, Weed classification for site-specific weed management using an automated stereo computer-vision machine-learning system in rice fields, *Plants* 9 (5) (2020) 559.
- [22] V. Partel, S.C. Kakarla, Y. Ampatzidis, Development and evaluation of a low-cost and smart technology for precision weed management utilizing artificial intelligence, *Comput. Electron. Agric.* 157 (2019) (2019) 339–350.
- [23] K.P. Panigrahi, H. Das, A.K. Sahoo, S.C. Moharana, Maize leaf disease detection and classification using machine learning algorithms, *Prog. Comput. Anal. Netw.* (2020) 659–669.
- [24] D. Al Bashish, M. Braik, S. Bani-Ahmad, A framework for detection and classification of plant leaf and stem diseases, in: Proceedings of 2010 International Conference on Signal and Image Processing (ICSIP), 2010.
- [25] E.K. Ratnasari, M. Mentari, R.K. Dewi, R.V. Hari Ginardi, Sugarcane leaf disease detection and severity estimation based on segmented spots image, in: Proceedings of International Conference on Information, Communication Technology and System (ICTS), 2014.
- [26] A. Rastogi, R. Arora, S. Sharma, Leaf disease detection and grading using computer vision technology & fuzzy logic, in: Proceedings of 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN), 2015.
- [27] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M.S. Lew, Deep learning for visual understanding: a review, *Neurocomputing* 187 (2016) 27–48.
- [28] M. Brahimi, K. Boukhalfa, A. Moussaoui, Deep learning for tomato diseases: classification and symptoms visualization, *Appl. Artif. Intell.* 31 (4) (2017) 299–315.
- [29] S.P. Mohanty, D.P. Hughes, M. Salathé, Using deep learning for image-based plant disease detection, *Front. Plant Sci.* (2016) 7.
- [30] J. Shijie, J. Peiyi, H. Siping, S.liu Haibo, Automatic detection of tomato diseases and pests based on leaf images, in: 2017 Chinese Automation Congress (CAC), 2017.
- [31] E.C. Too, L. Yujian, S. Njuki, L. Yingchun, A comparative study of fine-tuning deep learning models for plant disease identification, *Comput. Electron. Agric.* 161 (2019) 272–279.
- [32] J. Chen, Q. Liu, L. Gao, Visual tea leaf disease recognition using a convolutional neural network model, *Symmetry* 11 (3) (2019) 343.
- [33] Y. Lu, S. Yi, N. Zeng, Y. Liu, Y. Zhang, Identification of rice diseases using deep convolutional neural networks, *Neurocomputing* 267 (2017) 378–384.
- [34] A.C. Cruz, A. Luvisi, L. De Bellis, Y. Ampatzidis, X-FIDO: an effective application for detecting olive quick decline syndrome with deep learning and data fusion, *Front. Plant Sci.* (2017) 1741.
- [35] K.P. Ferentinos, Deep learning models for plant disease detection and diagnosis, *Comput. Electron. Agric.* 145 (2018) 311–318.
- [36] A.C. Cruz, Y. Ampatzidis, R. Pierro, A. Materazzi, A. Panattoni, L. De Bellis, A. Luvisi, Detection of grapevine yellows symptoms in *Vitis vinifera* L. with artificial intelligence, *Comput. Electron. Agric.* 157 (2019) 63–76.
- [37] S. Amanabadi, M. Vazirinia, H. Vereecken, K. Asefpour Vakilian, M. H. Mohammadi, Comparative study of statistical, numerical and machine learning-based pedotransfer functions of water retention curve with particle size distribution data, *Eurasian Soil Sci.* 52 (12) (2019) 1555–1571.
- [38] J.G.A. Barbedo, A review on the main challenges in automatic plant disease identification based on visible range images, *Biosystem. Eng.* 144 (2016) 52–60.
- [39] J.G.A. Barbedo, Factors influencing the use of deep learning for plant disease recognition, *Biosystem. Eng.* 172 (2018) 84–91.
- [40] J. Abdulridha, Y. Ampatzidis, P. Roberts, S.C. Kakarla, Detecting powdery mildew disease in squash at different stages using UAV-based hyperspectral imaging and artificial intelligence, *Biosystem. Eng.* 197 (2020) 135–148.
- [41] Hughes, D. P., and Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. ArXiv: 1511.08060.
- [42] H. Hejaziipoor, J. Massah, M. Soryani, K. Asefpour Vakilian, G. Chegini, An intelligent spraying robot based on plant bulk volume, *Comput. Electron. Agric.* 180 (2021), 105859.
- [43] B. Bhanu, Jing Peng, Adaptive integrated image segmentation and object recognition, *IEEE Trans. Syst. Man Cybern. Part C* 30 (4) (2000) 427–441.
- [44] A. Kumar, A. Tiwari, A comparative study of otsu thresholding and K-means algorithm of image segmentation, *Int. J. Eng. Tech. Res.* 9 (5) (2019).
- [45] A. Archana, G. Shrima, A.K. Modi, Improvement in K-means clustering using variant techniques, *SSRN Electron. J.* (2019).
- [46] R.M. Haralick, K. Shanmugam, I. Dinstein, On some quickly computable features for texture, in: Proc. Symp. Computer Image Processing and Recognition, Univ. Missouri, Columbia 2, 1972, p. 12, 2.
- [47] N.J. Sairamya, L. Susmitha, S. Thomas George, M.S.P. Subathra, Hybrid approach for classification of electroencephalographic signals using time-frequency images with wavelets and texture features, *Intell. Data Anal. Biomed. Appl.* (2019) 253–273.
- [48] W. Zhou, S. Gao, L. Zhang, X. Lou, Histogram of oriented gradients feature extraction from raw Bayer pattern images, *IEEE Trans. Circ. Syst. Express Briefs* 67 (5) (2020) 946–950.
- [49] L. Gong, J. Feng, R. Yang, Corner detection-based image feature extraction and description with application to target tracking, *Lect. Notes Electr. Eng.* (2015) 1069–1076.
- [50] M. Kashif, T.M. Deserno, D. Haak, S. Jonas, Feature description with SIFT, SURF, BRIEF, BRISK, or FREAK? A general question answered for bone age assessment, *Comput. Biol. Med.* 68 (2016) 67–75.
- [51] C. Ma, X. Hu, J. Xiao, H. Du, G. Zhang, Improved ORB algorithm using three-patch method and local gray difference, *Sensors* 20 (4) (2020) 975.
- [52] A. Fouad, H. Taha, Y. Elhalwagy, Automatic aerial target recognition using a robust surf-mser feature-based algorithm, *Int. Conf. Aerosp. Sci. Aviat. Technol.* 16 (2015) 1–14.
- [53] U. Kazhagamani, E. Murugasen, A hough transform based feature extraction algorithm for finger knuckle biometric recognition system, *Adv. Comput. Netw. Inform.* 1 (2014) 463–472.
- [54] V. Bolón-Canedo, N. Sánchez-Marcano, A. Alonso-Betanzos, A review of feature selection methods on synthetic data, *Knowl. Inf. Syst.* 34 (3) (2012) 483–519.
- [55] T.R. Gadekallu, D.S. Rajput, M.P.K. Reddy, K. Lakshmann, S. Bhattacharya, S. Singh, M. Alazab, A novel PCA-whale optimization-based deep neural network model for classification of tomato plant diseases using GPU, *J. Real-Time Image Process.* (2020).
- [56] L. Costa, J. McBreen, Y. Ampatzidis, J. Guo, M. Reisi Gahrooei, A. Babar, Using UAV-based hyperspectral imaging and functional regression to assist in predicting grain yield and related traits in wheat under heat-related stress environments for the purpose of stable yielding genotypes, *Precis. Agric.* 23 (2022) 622–642.
- [57] V. Kumar, Feature selection: a literature review, *Smart Comput. Rev.* 4 (3) (2014).
- [58] P. Drotár, J. Gazda, Z. Smékal, An experimental comparison of feature selection methods on two-class biomedical datasets, *Comput. Biol. Med.* 66 (2015) 1–10.
- [59] R. Vardasca, L. Vaz, J. Mendes, Classification and decision making of medical infrared thermal images, *Classif. BioApps* (2017) 79–104.
- [60] Y. Fang, R. Ramasamy, Current and prospective methods for plant disease detection, *Biosensors* 5 (3) (2015) 537–561.
- [61] N. Luchi, R. Ioos, A. Santini, Fast and reliable molecular methods to detect fungal pathogens in woody plants, *Appl. Microbiol. Biotechnol.* 104 (6) (2020) 2453–2468.
- [62] A. Luvisi, Y. Ampatzidis, L. De Bellis, Plant pathology and information technology: opportunity and uncertainty in pest management, *Sustainability* 8 (8) (2016) 831.
- [63] J. Massah, K. Asefpour Vakilian, M. Shabaniyan, S.M. Shariyatmadari, Design, development, and performance evaluation of a robot for yield estimation of kiwifruit, *Comput. Electron. Agric.* 185 (2021), 106132.