Journal of Agricultural Machinery

Homepage: https://jame.um.ac.ir

Research Article
Vol. 15, No. 1, Spring 2025, p. 65-79

Diagnosis and Classification of Two Common Potato Leaf Diseases (Early Blight and Late Blight) Using Image Processing and Machine Learning

H. Koroshi Talab[1], D. Mohammadzamani[1*], M. Gholami Parashkoohi[2]

1- Department of Biosystems Engineering, Takestan Branch, Islamic Azad University, Takestan, Iran
2- Department of Mechanical Engineering, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran
(*- Corresponding Author Email: dr.dmzamani@iau.ac.ir)

Received: 27 January 2024 Revised: 15 March 2024 Accepted: 06 April 2024 Available Online: 15 February 2025	**How to cite this article:** Koroshi Talab, H., Mohammadzamani, D., & Gholami Parashkoohi, M. (2025). Diagnosis and Classification of Two Common Potato Leaf Diseases (Early Blight and Late Blight) Using Image Processing and Machine Learning. *Journal of Agricultural Machinery, 15*(1), 65-79. https://doi.org/10.22067/jam.2024.86571.1226

Abstract

Diagnosing plant diseases is an important part of crop management and can significantly affect the quantity and quality of production. Traditional methods of visual assessment by human observers are time-consuming, costly, and error-prone, making accurate diagnosis and differentiation between various diseases difficult. Advances in agriculture have enabled the use of non-destructive machine vision systems for plant disease detection, and color imaging sensors have demonstrated great potential in this field. This study presents a framework for diagnosing early blight and late blight diseases in potatoes using a combination of Relief feature selection algorithms and Random Forest classification, along with color, texture, and shape features in three color spaces: RGB, HSV, and CIELAB (Lab*). The results indicated that the diagnostic accuracy for the early blight and late blight disease groups, as well as the healthy leaf group, were 94.71%, 95%, and 95.2%, respectively. The overall accuracy for disease classification was 95.99%. Additionally, the diagnostic accuracy for the early blight and late blight disease groups, along with the healthy leaf group, was 91.07%, 98.36%, and 98.93%, respectively, with an overall classification accuracy of 96.12%. After separating the diseased area from the healthy part of the leaf, a total of 150 features were extracted, including 45 color, 99 textural, and 6 shape features. The most effective features for disease detection were identified using a combination of all three feature sets. This study demonstrated that integrating these three sets of features can lead to a more accurate classification of potato leaves and provide valuable insights into the diagnosis and classification of potato diseases. This approach can help farmers and other plant disease specialists to accurately diagnose and manage potato diseases, and ultimately lead to an increase in product quality and yield.

Keywords: Artificial intelligence, Disease diagnosis, Feature extraction, Feature selection, Potato diseases

Introduction

Potatoes are one of the products that occupy a wide area of production in Iran. As a staple food, potato has its priority in cultivation all over the world. However, several diseases such as late blight and early potato blight affect the production of this product and destroy agricultural development. Therefore, diagnosing the disease in the early stages can be a better solution for successful crop cultivation (Abdulridha, Ampatzidis, Kakarla, & Roberts, 2019). The early blight and late blight, as the most destructive foliage diseases of potato crops, could cause major losses in most potato-growing areas in the world. On potato leaves, late blight appears as light green or olive green areas that rapidly turn brownish-

https://doi.org/10.22067/jam.2024.86571.1226

black, water-soaked, and oily. Likewise, early blight is round or irregular, which shows dark brown or black spots. Overall, early blight and late blight can occur in all stages of potato growth (Da Silva Silveira Duarte, Zambolim, Machado, Pereira Porto, & Rodrigues, 2019). To control and prevent diseases effectively and timely, it is of great significance to identify and detect the diseases of potato leaves.

In general, the traditional diagnosis of crop diseases is performed by experienced experts, but manual diagnosis is inefficient, subjective, and unsuitable for large regional scenarios. Traditional diagnostic techniques of crop diseases tend to include polymerase chain reaction (PCR), fluorescence in situ hybridization (FISH), enzyme-linked immunosorbent assay (ELISA), thermal imaging, and hyperspectral imaging (Fang & Ramasamy, 2015; Madufor, Perold, & Opara, 2018; Xie, Shao, Li, & He, 2015). In real-life production, farmers need simple, rapid, and accurate ways to identify potato diseases. Therefore, it is crucial to develop a fast, low-cost, time-saving, and labor-saving automatic identification system for potato diseases. In terms of image analysis, the application of machine learning techniques has shown great potential in the effective monitoring and identification of plant diseases. The visible patterns on the leaves of plants can be captured and processed through various image processing techniques to obtain specific patterns related to a particular disease. By comparing the obtained features or patterns with historical data, it is possible to classify different diseases. Machine learning techniques such as the combination of image processing and machine learning are highly effective in identifying and classifying diseases.

One of the potential advantages of machine learning is its ability to identify various diseases such as bacterial, fungal, and viral diseases. Intelligent diagnostic systems based on machine learning algorithms can automatically detect diseases in plants based on symptoms observed in their leaves. This technique can assist in monitoring large farms and diagnosing diseases earlier to minimize the adverse effects caused by such diseases (Barbedo, 2018).

Machine learning models, however, have a limitation in that a large and high-quality data set is required for model training and testing. Additionally, most of the existing machine learning and deep learning methods are unable to determine the most effective features extracted from images of plants for disease classification. Deep learning models, in particular, are complex and their functioning is not fully understood, which makes it difficult to determine why they perform better using different methods or models. Another challenge in the diagnosis of plant diseases is identifying and classifying diseases with similar symptoms, especially in the early stages of development. To overcome these challenges, it's crucial to develop a comprehensive system that can effectively extract features for accurate diagnosis and classification of diseases, particularly when comparing diseases and disorders with similar symptoms.

In recent years, significant research has been conducted in the field of machine vision in agriculture, particularly for fruit ripeness classification and quality ratings (Lopez, Aguilera, & Cobos, 2009; Mohamadzamani, Sajadian, & Javidan, 2020). This research has extended to include a range of sensors, such as color, multispectral, and hyperspectral cameras. Typically, the sensor used for these applications records information in the visible spectral range. This indicates the extent to which machine vision technology has progressed in the agriculture industry, with a focus on improving the effectiveness of farmers' decision-making processes.

Xiao and Liu, (2017) presented an adaptive feature combination and rapid diagnosis method for potato diseases. As proved by the detection test of three potato diseases, the average detection rate of the modified adaptive feature fusion method is at least 1.8% higher than the traditional adaptive method. Meanwhile, the average detection rate of the detection method was 92.5%. Fan and Li,

(2019) proposed a keypoint-based detection method that can quickly detect disease in target regions by combining the color and Textural Features. Although this method detects 10 types of potato diseases with high speed and accuracy, it does not perform well for diagnosis in complex environments. The detection accuracy in this method was 80%. Yang, Feng, Zhang, Sun, and Wang (2020) proposed a potato leaf disease detection method based on the combination of deep CNN and composite feature dictionary, adopted the faster R-CNN to detect disease areas, and built a composite feature dictionary through image feature extraction. The disease diagnosis model was trained by support vector machine and its average diagnosis accuracy can reach 84.16%. However, the background of the image was relatively simple. In research by Singh and Kaur (2021), an automatic system for potato leaf image disease detection was developed, and it was investigated using SVM algorithm on two well-known potato leaf diseases such as late blight and early blight. The proposed model achieved 91% accuracy.

Automatic plant disease detection system uses various techniques including: image acquisition, processing, segmentation, feature extraction, and machine learning to achieve quick and accurate diagnosis of plant diseases for farmers. The implementation of this automatic system eliminates the need for manual inspection, reducing the time and effort required for crop disease detection. As a result, farmers can focus on more critical tasks, such as ensuring optimal plant health, yields, and productivity.

This study presents a framework for potato leaf disease diagnosis. For this purpose, a dataset was compiled from farms in Qazvin province, Iran, consisting of one class of healthy potato leaves and two classes of diseased leaves, specifically late blight and early blight. A total of 300 images were prepared, with 100 images for each group. Subsequently, a combination of the Relief feature selection algorithm and the Random Forest classification algorithm was developed to diagnose and classify these diseases while also identifying the best and most effective features for recognizing common potato leaf diseases.

Materials and Methods

Machine learning has played a significant role in automating various systems, particularly in the field of agriculture, specifically in the classification and detection of plant diseases based on images. The proposed framework includes various steps, including system requirements, data collection, image segmentation, feature extraction, and classification. Each step is designed to achieve the ultimate goal of diagnosing and classifying images into different categories of diseases using the advantages of machine learning methods. The discussion will be organized to cover each of these steps, including their details and importance in achieving the desired results. Figure 1 shows the research process.

Creating the database of images

The creation of accurate and reliable image classifiers for plant disease detection is a crucial and challenging task. A large validated dataset of diseased and healthy plant images is highly necessary to develop effective classifiers. However, the lack of such a dataset has been one of the key barriers to achieving this goal. Previously, researchers had to create their own image datasets, which was a time-consuming and labor-intensive process. In this study, a data set of images of infected leaves including late blight and early blight will be prepared and analyzed by a Samsung mobile phone model (model A32) equipped with a 64 megapixel camera and in an environment with natural light. The images of the studied diseases are shown in Figure 2. The obtained dataset includes 300 leaves of potato plants, which were classified into 3 categories. The image database contains 100 healthy leaves and 200 diseased leaves. The image database was divided into two training databases and test databases. The training database contains 80% of the image database, i.e. 240 images, and the test database contains the remaining 20% of the image database, i.e. 60 images. The proposed framework was implemented on Windows 10 operating system with Intel®Core^{TM} i3-8130U processor @ 2.20 GHz -2.21 GHz with 8 GB RAM. The proposed algorithm has been implemented in order to diagnose and classify potato diseases using Matlab2018 software.

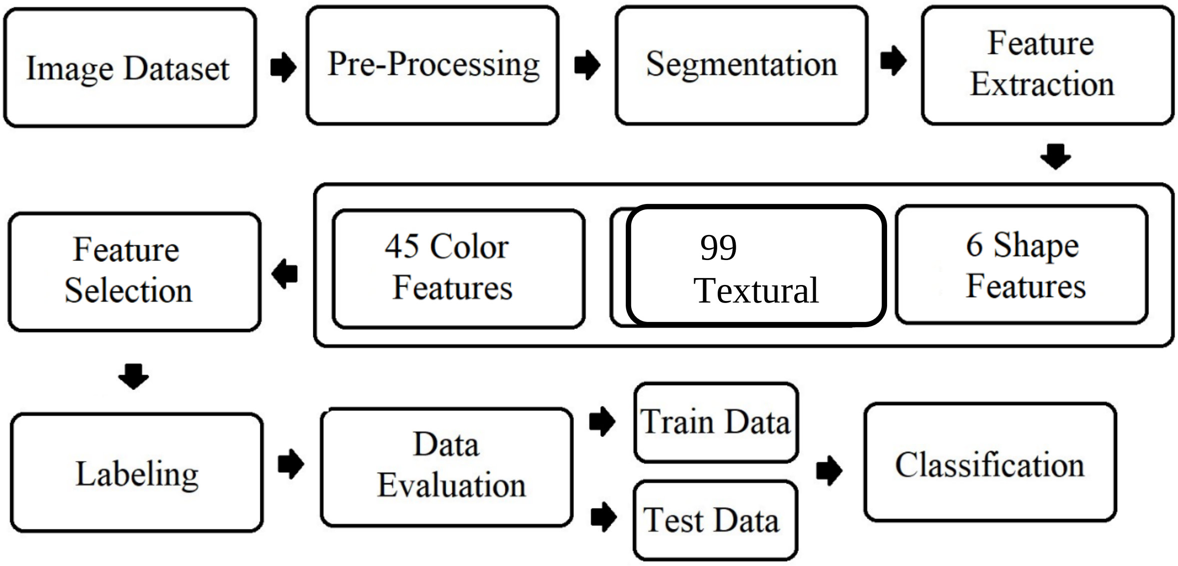

Fig. 1. The process of the proposed method to diagnose and classify potato diseases

Disease Name	Late Blight	Early Blight	Healthy
Leaf Image			
Symptoms	Early symptoms of late blight are small, light to dark green, circular to irregular spots soaked in water.	Symptoms of premature burn occur in the fruit, stem and foliage of potato and the stem, foliage and tubers of potato.	-

Fig. 2. Pictures of diseases along with their symptoms

Preprocessing

K-means clustering is used to classify objects based on a set of features into a total of *K* classes. The objects are classified by minimizing the sum of squares of the distance between the object and the corresponding cluster. The *K*-means clustering algorithm can be described by four steps (Mohammadi & Asefpour Vakilian, 2023): (1) picking the center of the *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.

Generally, the selection of the value of *K* and also the selection of Region of Interest (ROI) are made manually, depending on the skill of the user. Sometimes, ROI may not be selected correctly by the user. This means that for each number of images in the database, the ROI number should be manually selected to determine the desired area of the disease. This is very time-consuming and error-prone. Therefore, automatic clustering can be useful for the reliable 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 (cluster number indicating the disease), which had to be selected by the user in other methods, was done automatically 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. After that step, in the rest of the image, only the diseased area of the leaf will remain. Figure 3 shows the steps of preparing and separating the image from the background and detecting the disease area of potato leaves.

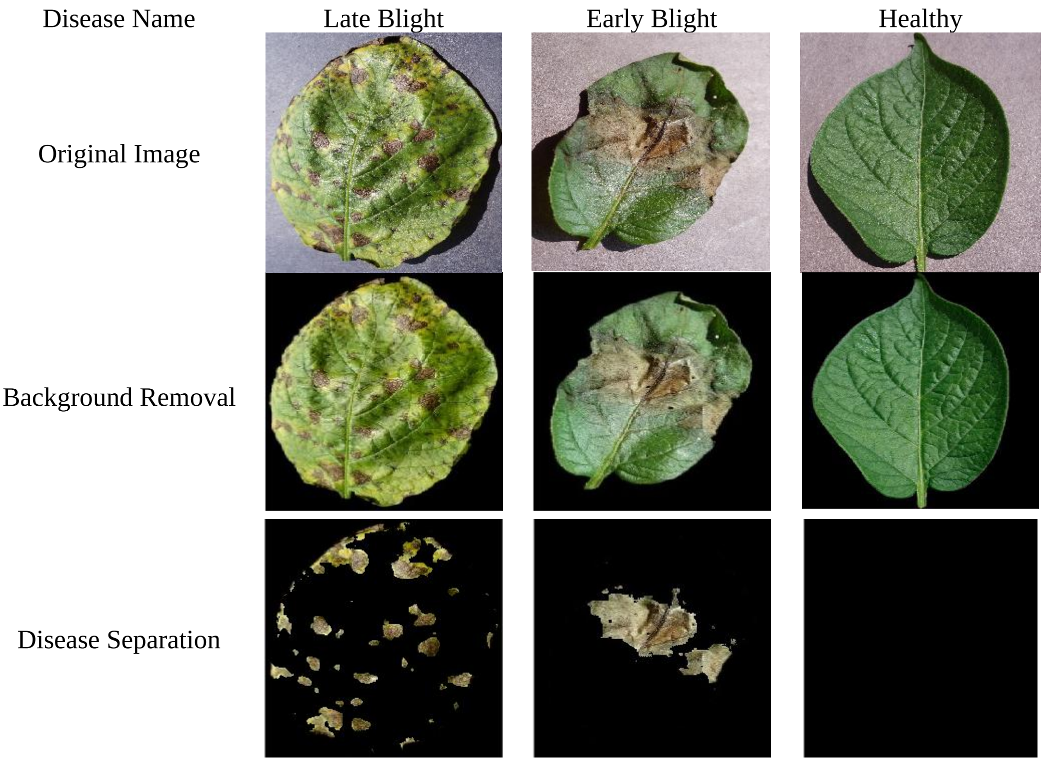

Fig. 3. Separation of the diseased area from the healthy leaf image

Feature extraction

The process of feature extraction involves transforming raw data into numerical features that can be understood and analyzed by machine learning models. Compared to directly using raw data, feature extraction can greatly improve the performance of machine learning models. In this case, after segmenting the disease area and extracting the damaged part of the leaf from input image, a variety of features such as colors, texture, and shapes were extracted (Ampatzidis, Partel, & Costa, 2020). In this study, after dividing the disease area, which means extracting the damaged part of the leaf from the input image, features such as color, shape, and texture were extracted. Color features (mean, maximum, minimum, standard deviation, and median) and Textural Features from GLCM (contrast, correlation, energy, homogeneity, mean, standard deviation, entropy, variance, smoothness, kurtosis, and skewness) in each band of three color spaces RGB, Lab*, and HSV, 45 color features and 99 Textural Features.

Furthermore, 6 shape features including: area, perimeter, number of spots, length of major and minor axis of points, and eccentricity index were extracted from images. Table 1 shows a description of the extracted features.

Table 1- Extracted features to diagnose and classify the diseases

Feature	Description	Equation	Reference
		Color Features	
Mean	The average value is computed as the sum of all the observed outcomes from the sample, divided by the total number of events	$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$	Ashfaq *et al.* (2019)
Max	Assign the maximum amount of neighborhood grays to the center pixel	-	Ashfaq *et al.* (2019)
Min	Assign the minimum amount of neighborhood grays to the center pixel	-	Ashfaq *et al.* (2019)
Standard Deviation	The measure of how far the data values lie from the mean	$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x-\bar{x})^2}$	Ashfaq *et al.* (2019)
Median	The middle value of a series of values is equal to the middle value of a series of numbers	-	Ashfaq *et al.* (2019)
		Textural Features	
Contrast	The measure of the difference between the brightness of the objects or regions and other objects within the same field of view	$\sum_{i,j}\lvert i-j\rvert^2 p(j,j)$	Haralick, Shanmugam, and Dinstein (1972)
Correlation	The measure of degree and type of relationship between adjacent pixels	$\sum_{i,j}\frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i\sigma_j}$	Haralick *et al.* (1972)
Energy	The sum of squared elements in the gray level co-occurrence matrix	$\sum_{i,j}p(i,j)^2$	Haralick *et al.* (1972)
Homogeneity	The closeness of the distribution of elements in the GLCM	$\sum_{i,j}\frac{p(i,j)}{1+\lvert i-j\rvert}$	Haralick *et al.* (1972)
Mean	The measure of the average intensity value of the pixels present in the region	$\frac{1}{n}\left(\sum_{i=1}^{n}X_i\right)$	Haralick *et al.* (1972)
Standard Deviation	The measure of how much the gray levels differ from the mean	$\sqrt{\frac{1}{n}\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)}$	Haralick *et al.* (1972)

Entropy	The measure of differences in gray levels	E=sum(p.*log2(p))	Haralick *et al.* (1972)
Variance	The measure of variance value of an image	$\frac{1}{n}\left(\sum_{i=1}^{n}(X_i-\bar{X})\right)^2$	Haralick *et al.* (1972)
Smoothness	A measure of relative smoothness of intensity in a region	-	Haralick *et al.* (1972)
Kurtosis	A measure of peaks distribution related to the normal distribution	$K=\frac{E(x-\mu)^4}{\sigma^4}$	Haralick *et al.* (1972)
Skewness	A measure of asymmetry in a statistical distribution	$S=\frac{E(x-\mu)^3}{\sigma^3}$	Haralick *et al.* (1972)
	Shape Features		
Area	Mean number of pixels in each segmented region in the image	$a=\sum_{u=1}^{M}\sum_{v=1}^{N}A[u,v]$	Vishnoi, Kumar, and Kumar (2021)
Perimeter	Mean of number of boundary pixels on each segmented region in the image	2 (length + width)	Vishnoi *et al.* (2021)
Number of objects	Number of disease spots in the leaf image	-	-
Major axis length	Length of the major axis of the ellipse that has the same normalized second central moments as the region	$M=x_1+x_2$	Vishnoi *et al.* (2021)
Minor axis length	Length of the minor axis of the ellipse that has the same normalized second central moments as the region	$m=\sqrt{((x_1+x_2)^2-d)}$	Vishnoi *et al.* (2021)
eccentricity index	The eccentricity of the ellipse that has the same second-moments as the region	$\frac{\text{Major axis length}}{\text{Minor axis length}}$	Vishnoi *et al.* (2021)

The Textural Features include energy, standard deviation, entropy, mean, and median, which are usually used to describe the texture of an image. Also, the selected band for each feature is specified in Table 1. The energy feature measures the amount of energy at a given pixel location based on the contrast of adjacent pixels, and higher values indicate more contrast. The standard deviation feature measures the deviation of the intensity values of neighboring pixels from the mean, with higher values indicating greater variability in the texture. The entropy property measures the amount of information at a given pixel location, with higher values indicating more random patterns (Ashfaq *et al.*, 2019). The average property represents the average intensity of all pixels in an image, while the mean property represents the average gray level of a given pixel.

The color features include median in two RGB bands, the maximum in Lab* band, and the average in two RGB bands, which are usually used to describe the color of an image. The median feature takes the middle value of the intensity values of the three color bands, and higher values indicate darker colors. The maximum feature in the Lab* band takes the highest intensity value of the Lab* color space and higher values indicate brighter colors. The average attribute represents the average intensity of all colors in an image, while the Lab* attribute captures the hue, saturation, and brightness values of a specific color (Zhou, Gao, Zhang, & Lou, 2020). Finally, color properties indicate the degree of red, green or blueness of a color.

The shape features include perimeter, area, and number of spots in the image, minor axis length and main axis length, which are usually used to describe the shape of an object in an image. The perimeter attribute indicates the distance around the perimeter of a given object, with higher values indicating larger objects. The area property represents the area value of an object, with higher values representing larger objects. The number of points features indicates the amount of variation in a given object, with higher values indicating more complex shapes. The minor axis length and major axis length indicate the length of the shortest and longest lines that can be drawn within an object, with higher values indicating longer or irregular shapes.

Feature selection

Feature selection is crucial for developing effective and reliable machine learning models for plant disease detection. Research has shown that selecting the right features can significantly impact the performance and accuracy of these models. Choosing the optimal set of features can also reduce the computational time required for training the system, particularly in cases where the number of features is large. One way to identify the most effective features is by using the Relief feature selection method, which reduces the redundancy of data features. This method randomly selects samples from the dataset and updates the score of each feature based on its correlation with the selected samples. Features with high scores are considered to be effective in the diagnosis and classification of plant diseases, making them ideal candidates for inclusion in machine learning models. This research applies the Relief method to extract the most important features for the detection and classification of potato diseases, making it a valuable resource for researchers in this field who are seeking to improve the accuracy and efficiency of their models.

Using Random Forest for classification

The Random Forest method is one of the comprehensive methods of machine learning that has competitive predictive application in various fields such as biological sciences, earth sciences, finance, chemical engineering, etc. The Random Forest can be introduced as a flexible and easy-to-use machine learning algorithm for different prediction and classification, which produces very good results most of the time even without setting too many parameters (Mohammadzamani, Javidan, Zand, & Rasouli, 2023). This new and powerful method has provided significant improvements in data mining technology. It is also one of the most widely used algorithms, because it can be used for classification and regression due to its simplicity and variety. Random Forest technique is a developed model of tree regression and classification method.

Performance evaluation in the classification of potato diseases

One of the most important criteria for evaluating the performance of a classification model is the confusion matrix. This matrix shows the relationship between the predicted and actual classes, where TP stands for the number of true positives, TN stands for the number of true negatives, FP stands for the number of false positives, and FN stands for the number of false negatives. In this research, the performance of the model was evaluated by classification accuracy, as well as other metrics such as Precision, Sensitivity, Specificity, and F-Measure (Equations 1 to 5).

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

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (3)$$

$$Specificity = \frac{TN}{TN+FP} \quad (4)$$

$$F\text{-}Measure = \frac{2\times(Precision\times Sensitivity)}{(Precision+Sensitivity)} \quad (5)$$

Results and Discussion

Image classification results

The proposed method for potato leaf diseases classification consists of several stages: segmentation of diseased region, feature extraction, feature selection (Relief algorithm), and classification (Random Forest algorithm). Finally, classification of diseases and healthy leaves before using feature selection algorithm and using random forest classification algorithm with 95.99% accuracy, 96.12% precision, 96.25% recall, and 96.16% F1 score. The classification criteria for diseases are presented in Table 2. The proposed method provides a promising approach for the precise and accurate diagnosis of potato diseases.

According to Figure 4, the late blight class achieved a lower accuracy than the other two classes. Additionally, the healthy leaves category obtained a higher accuracy compared to the other classes. The proposed model showed a satisfactory overall performance in terms of precision, recall, and F1 score. It was evident that the healthy leaves class performed better than the other two classes. This suggests that more attention should be paid to increasing the accuracy of the late blight class in future research. The proposed framework provides a promising and efficient approach for the accurate diagnosis of plant diseases.

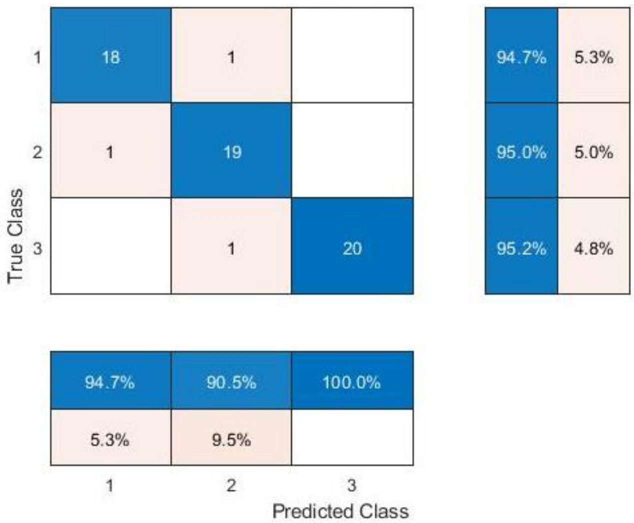

Fig. 4. Confusion matrix for classification (1- early blight group, 2- late blight group, and 3- healthy leaf group)

Table 2- Evaluation criteria for the classification of potato diseases

Category Name	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Late Blight	91.07	95.41	93.29	94.71
Early Blight	98.36	94.71	96.43	95
Healthy	98.93	98.62	98.76	95.2
Overall	96.12	96.25	96.16	95.99

According to Figure 5, the proposed model has performed well in identifying late blight and early blight in potato plants with accuracy rates of 94.71 and 95% and precision of 91.07 and 98.36%, respectively. In addition, it was able to accurately identify healthy leaves with an accuracy rate of 95.2%. The overall accuracy rate of the model was 95.99%, which shows that it had a relatively low error and was able to accurately identify plant diseases in most cases. The high model accuracy and good overall accuracy are promising indicators of its potential in plant disease diagnosis. However, it is important to note that the performance of the model is affected by the size and quality of the training dataset, and the scope of the specific application may also affect its performance (Zhai, Qiu, Weckler, He, & Jabran, 2023). In addition, there may be cases where the model misclassifies healthy or diseased plants, which can lead to false positives or negatives. These mistakes can potentially lead to incorrect treatment or lack of intervention, which can lead to financial losses for farmers or reduce the quality and quantity of crops (Jaisakthi, Mirunalini, & Thenmozhi, 2019). Therefore, it is important to continue validating the model performance in a larger dataset with different plant types. The results of these validation experiments provide valuable insights into model strengths and limitations and help guide the development of more accurate and robust models for plant disease diagnosis.

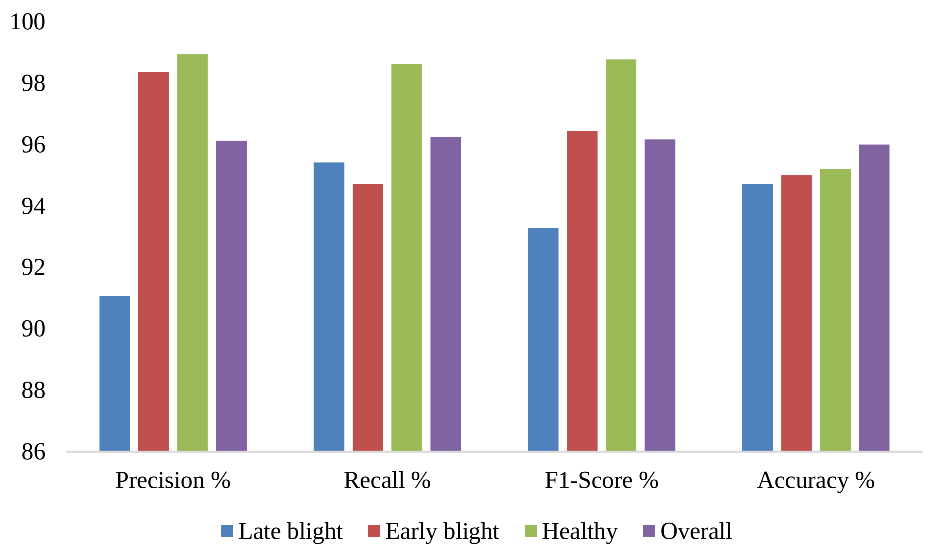

Fig. 5. A representation of evaluation criteria for the classification of potato diseases

Results of feature selection for potato disease diagnosis

After separating the diseased area from the healthy part of the leaf, a total of 150 features were extracted. These features included 45 color features, 99 Textural Features, and 6 shape features. To identify and classify each group, the most effective features were selected. This selection process included a combination of all three feature sets. The analysis of each feature set showed the features that had the greatest impact in the classification process. The final results showed that the combination of these three sets of features led to the most accurate classification.

The classification accuracy of individual feature sets was 70.45% for color features, 82.50% for Textural Features, and 70% for shape features. On the other hand, the classification accuracy of the combined feature set of the Random Forest classifier was

95.99%. In Table 3, five effective features for recognizing and classifying potato leaves are listed along with their descriptions, weights, and scores. In general, this study shows that the combination of three sets of features can lead to a more accurate classification of potato leaves. The selected features can also provide valuable insights into the diagnosis and classification of potato diseases. This approach can help farmers and other plant disease specialists to accurately diagnose and manage potato diseases, and ultimately lead to an increase in product quality and yield.

According to Table 3, the most important features selected by the Relief algorithm for diagnosing and classifying potato diseases based on texture, color and shape can be summarized as follows:

In general, these selected features represent the most important features that can be extracted from the leaves affected by potato-potato disease and can help in the diagnosis and classification of potato diseases. Textural features describe the structure and texture of the leaf, while color features describe the color and brightness of the leaf. Shape features describe leaf shape, including perimeter, area, number of spots in the image, minor axis length, and major axis length. These features (texture, color, and shape), when combined, can provide valuable information about the presence and nature of potato diseases and help farmers and plant disease specialists to identify and manage potato diseases more effectively.

Table 3- The results of the best and most important features selected for the diagnosis of potato leaf diseases

Row	1	2	3	4	5	Accuracy%
Color	Median (RGB) (G)	Median (RGB) (B)	Max (Lab*) (b)	mean (RGB) (G)	mean (RGB) (B)	70.45
Texture	Energy (Lab*) (l)	Standard Deviation (RGB) (B)	Entropy (Lab*) (b)	Mean (HSV) (H)	Median (Lab*) (l)	82.50
Shape	Area	Perimeter	Number of objects	Minor axis length	Major axis length	70
All Features	Skewness (HSV) (S)	Homogeneity (Lab*) (b)	Correlation (Lab*) (b)	Energy (RGB) (R)	Entropy (RGB) (R)	95.99

Compared to the research done by other researchers, what is more visible is the difference in classification accuracy, the number of extracted features, the classification method, and the number of data available in the diagnosis of diseases, which is one of the most important things that make the classification different (Padol & Yadav, 2016). The most important challenge in past research is the limited availability of diverse image datasets in various deep learning methods and convolutional neural networks, which weakens the performance of classifiers (Liu, Tan, Li, He, & Wang, 2020; Xie *et al.*, 2020). In such cases, using machine learning and correctly and effectively extracting the main features such as texture, color, and shape from images of diseased leaves can be a more effective approach (Jaisakthi *et al.*, 2019; Singh & Kaur, 2021). The weakness of fewer data can be overcome by using classification in machine learning, but the condition of high classification accuracy is the extraction of features that also increase classification accuracy (Padol & Yadav, 2016). Comparing the results with research that used machine learning shows that the use of different groups of features has a great impact on classification accuracy (Padol & Yadav, 2016). In this research, it was shown that an algorithm can be used to diagnose and classify the common diseases of potato plants, which determined the exact location of the disease by processing color images. In addition, researchers were introduced to features that are highly accurate in diagnosis. They are most effective in the classification of diseases with almost identical symptoms in the leaves of the potato plant. Also, the results obtained by a feature selection algorithm showed that different feature groups such as texture, and color in different color and shape spaces have almost the same importance in classification, so removing or ignoring each of the feature groups in the algorithm's performance will reduce the classification of the disease. On the other hand, there are cases such as not diagnosing the disease in the early days within two or three days after contracting the disease. Also, the limited spectrum in the color camera can be solved by using hyperspectral cameras, but using them and taking pictures cost a lot, and it will be difficult for farmers to carry out, as well as the small number of diseases, which will be increased in future research or can be suggested to be carried out by other researchers. Finally, the limited number of images used in deep learning algorithms can be considered as the limitation of this research.

A practical application of the proposed method can help plant breeders and farmers identify and diagnose potato diseases early in the crop cycle. This can help farmers to take timely measures and take measures to reduce disease outbreaks that can reduce yield and profitability. The proposed method can also be used to diagnose potato diseases alongside the knowledge of plant pathologists.

One of the challenges facing breeders and farmers in diagnosing potato diseases is the large number of diseases that can affect potato crops. Potato disease can be caused by viruses, bacteria, fungi, and other pathogens, and each disease can cause unique symptoms that can make identification and diagnosis difficult. Another challenge is the lack of reliable diagnostic tools such as laboratories and experienced plant pathologists in rural areas. This can make it difficult for farmers to access accurate diagnoses and effective treatments, which can lead to reduced yields and profitability.

Conclusion

This study presents a framework for detecting late blight and early blight diseases using a combination of Relief feature selection algorithms and Random Forest classification, along with color features (in three color spaces: RGB, HSV, and CIELAB), Textural Features, and shape features. After separating the diseased area from the healthy part of the leaf, a total of 150 features were extracted, including 45 color features, 99 textural features, and 6 shape features. The most effective features for disease detection were identified using a combination of all three feature sets. This study demonstrated that integrating these three sets of features can lead to more accurate classification of potato diseases than using each group of features separately and can provide valuable insights into the diagnosis and classification of potato diseases. The results indicated that the diagnostic accuracy for the early blight and late blight disease groups, as well as the healthy leaf group, was 94.71%, 95%, and 95.2%, respectively, with an overall accuracy for disease classification of 95.99%. Additionally, the diagnostic accuracy for the two disease groups (early blight and late blight) and the healthy leaf group was 91.07%, 98.36%, and 98.93%, respectively, with an overall classification accuracy of 96.12%.

This approach can help farmers and plant disease specialists accurately diagnose two important potato diseases: early blight and late blight. It assists farmers and plant pathologists in making informed decisions for managing these diseases. By providing a quick, accurate, and convenient method for diagnosing early blight and late blight using digital images of leaves, this automated method reduces the time and effort required for disease diagnosis and can be utilized by farmers and plant breeders with minimal training. Furthermore, this method can be integrated with existing digital technologies, such as drones and remote sensing, to monitor large areas of potato crops for disease symptoms, ultimately saving time and money.

Competing Interests

The authors declare that they have no conflict of interest in the publication of this article. There are no copyrights relevant to this work. There are no relationships or activities

to disclose that could be perceived to have influenced the submitted work.

Authors Contribution

H. Koroshi Talab: Conceptualization, Study design, Data collection, Methodology and analysis.

D. Mohammad Zamani: Supervision, Writing the manuscript.

M. Gholami Parashkoohi: Thesis advisor.

References

1. Abdulridha, J., Ampatzidis, Y., Kakarla, S. C., & Roberts, P. (2019). Detection of target spot and bacterial spot diseases in tomato using UAV-based and benchtop-based hyperspectral imaging techniques. *Precision Agriculture, 21*(5), 955-978. Springer Science and Business Media LLC. https://doi.org/10.1007/s11119-019-09703-4
2. Ampatzidis, Y., Partel, V., & Costa, L. (2020). Agroview: Cloud-based application to process, analyze and visualize UAV-collected data for precision agriculture applications utilizing artificial intelligence. *Computers and Electronics in Agriculture, 174*, 105457. Elsevier BV. https://doi.org/10.1016/j.compag.2020.105457
3. Ashfaq, M., Minallah, N., Ullah, Z., Ahmad, A. M., Saeed, A., & Hafeez, A. (2019). Performance Analysis of Low-Level and High-Level Intuitive Features for Melanoma Detection. *Electronics*, *8*(6), 672. https://doi.org/10.3390/electronics8060672
4. Barbedo, J. G. A. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosystems Engineering*, *172*, 84-91. https://doi.org/10.1016/j.biosystemseng.2018.05.013
5. Da Silva Silveira Duarte, H., Zambolim, L., Machado, F. J., Pereira Porto, H. R., & Rodrigues, F. A. (2019). Comparative epidemiology of late blight and early blight of potato under different environmental conditions and fungicide application programs. *Semina: Ciências Agrárias, Londrina*, *40*, 1805-1818. https://doi.org/10.5433/1679-0359.2019v40n5p1805
6. Fan, Z., & Li, X. (2019). Recognition of potato diseases based on fast detection and fusion features of ROI. *Southwest China Journal of Agricultural Science*. 544-550. https://doi.org/10.16213/j.cnki.scjas.2019.3.015
7. Fang, Y., & Ramasamy, R. P. (2015). Current and prospective methods for plant disease detection. *Biosensors*, 5, 537-561. https://doi.org/10.3390/bios5030537
8. Haralick, R. M., Shanmugam, K., & Dinstein, I. (1972). On some quickly computable features for texture. *In Symposium of Computer Image Processing and Recognition*, University of Missouri, Columbia, *2*(2), 12-2.
9. Jaisakthi, S. M., Mirunalini, P., & Thenmozhi, D. (2019). Grape leaf disease identification using machine learning techniques, *in Proceedings of the 2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, Chennai, 1-6. https://doi.org/10.1109/ICCIDS.2019.8862084
10. Liu, B., Tan, C., Li, S., He, J., & Wang, H. (2020). A Data Augmentation Method Based on Generative Adversarial Networks for Grape Leaf Disease Identification. *IEEE Access*, *8*, 102188-102198. Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/access.2020.2998839
11. Lopez, J. J., Aguilera, E., & Cobos, M. (2009). Defect Detection and Classification in Citrus Using Computer Vision. *In Neural Information Processing* (pp. 11-18). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-10684-2_2
12. Madufor, N. J. K., Perold, W. J., & Opara, U. L. (2018). detection of plant diseases using biosensors: a review. *Acta Horticulturae*, *1201*, 83-90. https://doi.org/10.17660/ActaHortic.2018.1201.12
13. Mohamadzamani, D., Sajadian, S., & Javidan, S. M. (2020). DDetection of *Callosobruchus*

maculatus F. with image processing and artificial neural network. *Applied Entomology and Phytopathology*, *88*(1), 103-112. https://doi.org/10.22092/jaep.2020.341684.1324

14. Mohammadi, P., & Asefpour Vakilian, K. (2023). Machine learning provides specific detection of salt and drought stresses in cucumber based on miRNA characteristics. *In Plant Methods, 19*(1). Springer Science and Business Media LLC. https://doi.org/10.1186/s13007-023-01095-x
15. Mohammadzamani, D., Javidan, S. M., Zand, M., & Rasouli, M. (2023). Detection of Cucumber Fruit on Plant Image Using Artificial Neural Network. *Journal of Agricultural Machinery*, *13*(1), 27. https://doi.org/10.22067/jam.2022.73827.1077
16. Padol, P. B., & Yadav, A. A. (2016). SVM classifier based grape leaf disease detection, in *Proceedings of the 2016 Conference on Advances in Signal Processing (CASP)*, April 12-15. Lisbon,
17. Singh, A., & Kaur, H. (2021). Potato Plant Leaves Disease Detection and Classification using Machine Learning Methodologies. *In IOP Conference Series: Materials Science and Engineering*, *1022*(1), 012121. IOP Publishing. https://doi.org/10.1088/1757-899x/1022/1/012121
18. Vishnoi, V. K., Kumar, K., & Kumar, B. (2021). A comprehensive study of feature extraction techniques for plant leaf disease detection. *Multimedia Tools and Applications*, *81*, 367-419. https://doi.org/10.1007/s11042-021-11375-0
19. Xiao, Z., & Liu, H. (2017). Adaptive features fusion and fast recognition of potato typical disease images. *Transactions of the Chinese Society for Agricultural Machinery*, *48*(12), 26-32. https://doi.org/10.6041/j.issn.1000-1298.2017.12.003
20. Xie, C., Shao, Y., Li, X., & He, Y. (2015). Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging. *Scientific Reports*, 5, 1-11. https://doi.org/10.1038/srep16564
21. Xie, X., Ma, Y., Liu, B., He, J., Li, S., & Wang, H. (2020). A Deep-Learning-Based Real-Time Detector for Grape Leaf Diseases Using Improved Convolutional Neural Networks. *Frontiers in Plant Science*, *11*(2). Frontiers Media SA. https://doi.org/10.3389/fpls.2020.00751
22. Yang, S., Feng, Q., Zhang, J., Sun, W., & Wang, G. (2020). Identification method for potato disease based on deep learning and composite dictionary. *Transactions of the Chinese Society for Agricultural Machinery*, *51*(7), 22-29. https://doi.org/10.6041/j.issn.1000-1298.2020.07.003
23. Zhai, C., Qiu, W., Weckler, P., He, X., & Jabran, K. (2023). Editorial: Advanced application technology for plant protection: Sensing, modelling, spraying system and equipment. *Frontiers in Plant Science*, *14*. Frontiers Media SA. https://doi.org/10.3389/fpls.2023.1113359
24. Zhou, W., Gao, S., Zhang, L., & Lou, X. (2020). Histogram of Oriented Gradients Feature Extraction from Raw Bayer Pattern Images. *In IEEE Transactions on Circuits and Systems II: Express Briefs*, *67*(5), 946-950). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/tcsii.2020.2980557

نشریه ماشین‌های کشاورزی
https://jame.um.ac.ir

مقاله پژوهشی
جلد ۱۵، شماره ۱، بهار ۱٤۰٤، ص ۷۹-٦٥

تشخیص و طبقه‌بندی دو بیماری شایع برگ سیب‌زمینی (لکه‌موجی زودرس و سفیدک داخلی) با استفاده از پردازش تصویر و یادگیری ماشینی

حسن کوروشی طلب[۱]، داود محمدزمانی[۱*]، محمد غلامی پرشکوهی[۲]

تاریخ دریافت: ۱۴۰۲/۱۱/۰۷
تاریخ پذیرش: ۱۴۰۳/۰۱/۱۸

چکیده

تشخیص بیماری‌های گیاهی بخش مهمی از فرآیند مدیریت مزرعه است و می‌تواند تاثیر قابل‌توجهی بر کمیت و کیفیت تولید داشته باشد. روش‌های سنتی ارزیابی چشمی توسط ناظران انسانی زمان‌بر، پر هزینه و مستعد خطا هستند و تشخیص دقیق و تمایز بین بیماری‌های مختلف را دشوار می‌سازند. پیشرفت‌های کشاورزی امکان استفاده از سامانه‌های بینایی ماشین غیرمخرب را برای تشخیص بیماری‌های گیاهی فراهم کرده است و حسگرهای تصویربرداری رنگی توانایی بالایی در این زمینه از خود بروز داده‌اند. این مطالعه چارچوبی را برای تشخیص بیماری لکه موجی زودرس و سفیدک داخلی سیب‌زمینی با استفاده از ترکیبی از الگوریتم‌های انتخاب ویژگی Relief و طبقه‌بندی تصادفی جنگل و ویژگی‌های رنگ، بافت و شکل در سه فضای رنگی RGB، HSV و Lab* توصیف کرد. نتایج این بررسی نشان داد که دقت تشخیص برای گروه بیماری لکه موجی زودرس و سفیدک داخلی و گروه برگ سالم به‌ترتیب ۹۴/۷۱، ۹۵ و ۹۵/۲ درصد و دقت کلی برای طبقه‌بندی بیماری ۹۵/۹۹ درصد بود. همچنین دقت تشخیص برای دو گروه بیماری لکه موجی زودرس و سفیدک داخلی و گروه برگ سالم به‌ترتیب ۹۱/۰۷، ۹۸/۳۶ و ۹۸/۹۳ درصد و دقت کلی برای طبقه‌بندی بیماری‌ها ۹۶/۱۲ درصد بود. پس از جداسازی ناحیه بیمار از قسمت سالم برگ، در مجموع ۱۵۰ ویژگی شامل ۴۵ ویژگی رنگی، ۹۹ ویژگی بافتی و شش ویژگی شکلی استخراج شد. مؤثرترین ویژگی‌ها برای تشخیص بیماری با استفاده از ترکیبی از هر سه مجموعه ویژگی شناسایی شدند. این مطالعه نشان داد که ترکیب این سه مجموعه از ویژگی‌ها می‌تواند منجر به طبقه‌بندی دقیق‌تر برگ‌های سیب‌زمینی شود و بینش ارزشمندی در تشخیص و طبقه‌بندی بیماری‌های سیب‌زمینی ارائه دهد. این رویکرد می‌تواند به کشاورزان و سایر متخصصان بیماری‌های گیاهی کمک کند تا بیماری‌های سیب‌زمینی را به‌طور دقیق تشخیص داده و مدیریت کنند و در نهایت منجر به افزایش کیفیت و عملکرد محصول شود.

واژه‌های کلیدی: انتخاب ویژگی، استخراج ویژگی، بیماری‌های سیب زمینی، تشخیص بیماری، هوش مصنوعی

۱- گروه مهندسی بیوسیستم، واحد تاکستان، دانشگاه آزاد اسلامی، تاکستان، ایران
۲- گروه مهندسی مکانیک، واحد شهر قدس، دانشگاه آزاد اسلامی، تهران، ایران
(*- نویسنده مسئول: Email: dr.dmzamani@iau.ac.ir)

https://doi.org/10.22067/jam.2024.86571.1226