Journal of Engineering Science and Technology
Special Issue on ICIT2022 Vol. 18, No. 6 (2023) 32 - 38

PLANT DISEASE DETECTION USING RANDOM FOREST CLASSIFIER WITH NOVEL SEGMENTATION AND FEATURE EXTRACTION STRATEGY

R. KARTHICKMANOJ[1,2,*], T. SASILATHA, NARINDERJIT SINGH SAWARAN SINGH[2]

[1]Department of EEE, Academy of Maritime Education and Training, Deemed to be University, Kanathur, Chennai, Tamil Nadu, India
[2]Faculty of Data Science and Information Technology, INTI International University, Nilai, Malaysia
*Corresponding author: karthickmanoj@ametuniv.ac.in

Abstract

In agricultural applications, plant disease identification is critical for increasing economic yield. To avoid yield loss, early identification of disease in leaves is critical. Machine learning algorithms can be used to classify diseases at an early stage, allowing farmers to take action to avoid further crop damage. The paper's key contribution is the development of an effective method for tracking plants in order to identify and classify diseases at an early stage. The camera sensor will be used by the machine to capture leaf images in the field. For extracting essential features for classification, a novel segmentation and feature extraction technique is proposed. The disease is classified using the random forest algorithm at the monitoring station. The system's efficiency is measured in terms of detection and classification accuracy.

Keywords: Accuracy, Agriculture, Classification, Feature extraction, Innovation, Process, RFC, Segmentation, Smallholder.

1. Introduction

Plant disease detection is critical in agriculture applications because diseases and pests are costly to farmers. Because visual tracking of plants for illnesses is a less accurate and time-consuming task, machine learning automates the procedure of diagnosing diseases based on features extracted from photos [1, 2]. At the field to be watched, image acquisition and feature extraction are carried out. Experts do the classification at the monitoring station. Finally, the farmers are given advice on how to solve the situation via text messages or a mobile app.

The farm's camera sensor is used to capture images of the plant leaves. The image is then further processed to identify and segment the affected leaf section. The agricultural specialist extracts feature from the segmented section and sends them to him for study. If the farmer had exact knowledge of where the disease had spread, he would be able to apply appropriate pesticide doses to the affected areas, resulting in both economic and environmental benefits.

Pre-processing and segmentation are critical components of a vision-based disease diagnosis system. The extraction of characteristics from the segmented image are critical for efficient categorization. The process of grouping or splitting an image into various portions is known as image segmentation [3-6].Different techniques, ranging from simple to complex segmentation procedures, can be used to segment an image. Colour, texture, and area are used to extract features. Experts use classifiers to classify information.

Singh and Misra [7] introduced a novel picture segmentation approach for detecting and classifying plant leaf diseases automatically. The genetic algorithm was employed to segment the images, and the SVM classifier was utilised to classify them. The accuracy of the suggested method was assessed to be 95.71 percent, with a classification accuracy of 97.6 percent.

Dhakate and Ingole [8] suggested a system for detecting and classifying illnesses in pomegranate plants that uses image processing and neural networks. Fruit Spot, Bacterial Blight, and Leaf Spot are the illnesses that will be demonstrated.

The proposed method produced satisfactory results with a 90% accuracy rate. Web enabled disease detection system (WEDDS) based on compressed sensing (CS) is proposed by Nandhini et al. [9] to identify and identify illnesses in leaves. For several plants, the system was tested using both simulation and experimental analyses. The classification was done using SVM, and the results indicated a 98.5 % accuracy.

In this work, a novel segmentation and feature extraction technique is proposed for classification of diseases. The goal is to extract characteristics based on the foreground of the image. The foreground of the leaf image is retrieved first, followed by the characteristics. The suggested approach's performance is measured in terms of detection and classification accuracy. There is also a comparison with K-means classification algorithm [10].

The remainder of the paper is laid out as follows. The proposed framework is discussed in Section 2.The performance evaluation is discussed in full in Section 3.Section 4 concludes with a conclusion and recommendations for future research.

2. Proposed Work

A novel and simple segmentation and feature extraction technique is proposed for the leaf disease detection system. The system is used to monitor plant leaves in the field from afar. This technique makes it easier to detect diseases at an early stage. The system will collect the image, pre-process it, and then segment it using a technique called foreground extraction. This method has the advantage of reducing manual monitoring in large farms and detecting diseases as soon as they appear on plant leaves. The system is depicted in Fig. 1 as a whole.

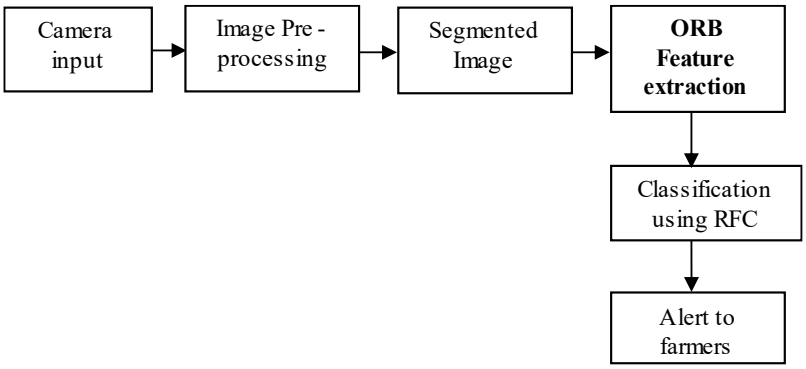


Fig. 1 Proposed system.

The photos of the leaves are captured by the camera sensor, which is then processed to segment the sick region. The image is first pre-processed before the foreground is extracted for segmentation. The camera is only activated when the colour of the leaf changes, which reduces the amount of power used. The photos are collected at regular intervals in some circumstances. Before the segmentation procedure, the image is pre-processed to improve its quality. To improve the quality, the image is enhanced before being transformed into the L*a*b* model. To isolate the foreground, which will be the affected section, background subtraction, and OTSU thresholding are utilized. For feature extraction, ORB descriptors are employed. ORB is essentially a hybrid of the FAST key point detector and the BRIEF descriptor, with numerous enhancements. ORB is a great replacement for SURF and SIFT detectors. The retrieved features are fed into the classification algorithm as training data.

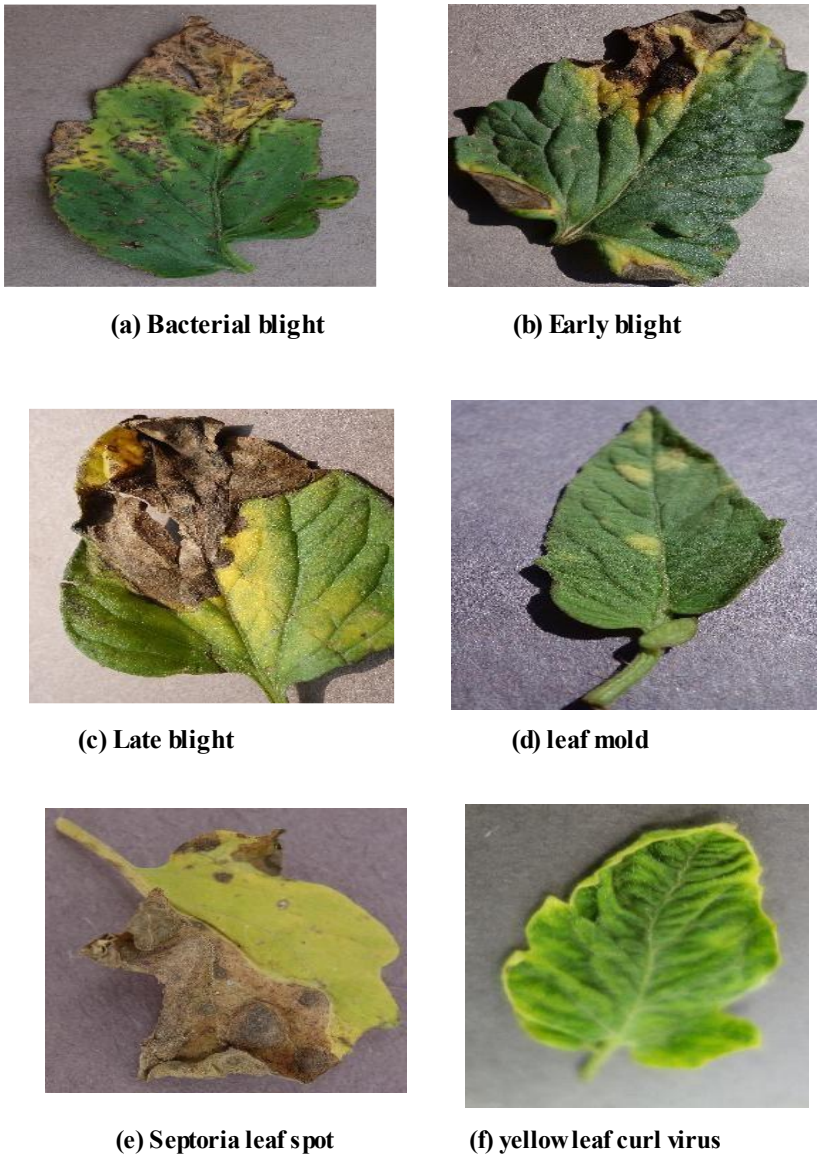
Random Forest Classifier (RFC) is a supervised learning algorithm commonly used for classification [6, 11]. The training dataset is used to predict the testing dataset in supervised learning. Random forest is a flexible, user-friendly machine learning methodology that, in most circumstances, produces great results even without hyper-parameter tweaking. It is also one of the most extensively utilized algorithms due to its simplicity and versatility. Following classification, the found disease is delivered to the farmers, together with the treatment, for further action to improve agricultural yield.

3. Performance Evaluation

Following classification, the found disease is delivered to the farmers, together with the treatment, for further action to improve agricultural yield. The suggested system

was tested using the software MATLAB 2019. Validation was done on tomato leaves, and evaluation was done on a database comprising 1000 photos for training and 400 images for testing.

Karthickmanoj et al. [12] was used to create the database. Bacterial spot, early blight, late blight, leaf mold, Septoria leaf spot, and tomato yellow leaf curl virus are among the diseases that will be demonstrated. The suggested system's performance is measured in terms of detection and classification accuracy. The training dataset's input photos for various disorders are shown in Fig. 2. Figure 3 shows the segmented diseased leaf using the proposed technique.

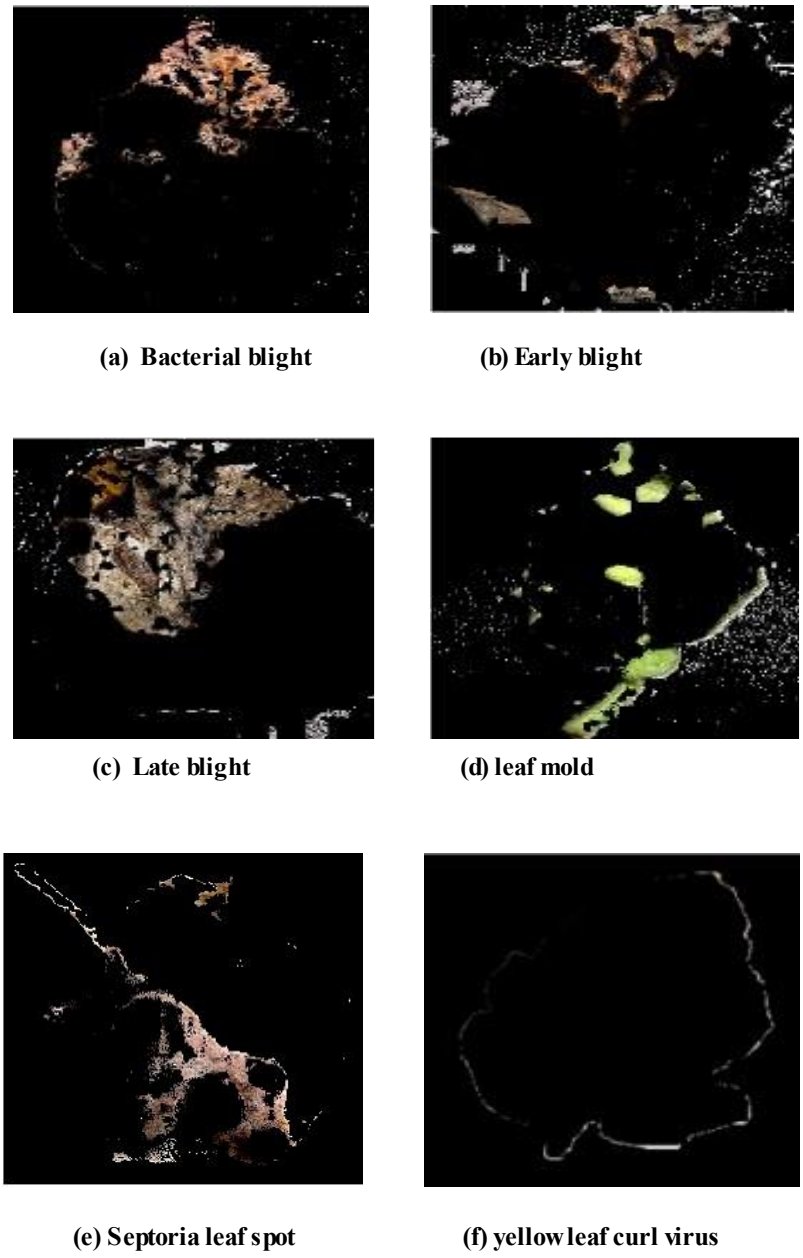


(a) Bacterial blight **(b) Early blight**

(c) Late blight **(d) leaf mold**

(e) Septoria leaf spot **(f) yellow leaf curl virus**

Fig. 2 Tomato leaves disease.



(a) Bacterial blight **(b) Early blight**

(c) Late blight **(d) leaf mold**

(e) Septoria leaf spot **(f) yellow leaf curl virus**

Fig. 3. Segmented tomato leaves.

The accuracy of the system (ACC) is calculated using Eq. (1).

$$ACC = \frac{\mathrm{p}}{\mathrm{T}} * 100 \quad (1)$$

where *P* signifies the number of positively classified results and signifies the total number of images considered for testing. Table 1 shows how accurate the system

is for six tomato leaf diseases. The suggested methodology is compared to segmentation based on the well-known K-means clustering algorithm.

Table 1 shows that the suggested approach achieves an average accuracy of 98.8 percent, which is higher than the existing work described in the literature. The six different tomato leaf diseases are considered for classification, and the accuracy of each disease's categorization is tabulated in Table 2.

Classification accuracy determines the rate at which specific diseases may be correctly classified from test photos. According to Table 2, the overall categorization accuracy is 99.75 percent. In comparison to previous work [6], the proposed system and RFC algorithm has a higher average classification accuracy.

Table 1. Comparison of the proposed work with the existing work.

Crop considered	Disease name	Detection accuracy (%)	
		Existing Work [2]	Proposed system
Tomato	Bacterial blight	98.38	98.50
	Early blight	97.67	97.38
	Late blight	96.54	98.20
	Leaf mold	97.46	99.6
	Septoria leaf spot	95.55	99.36
	Tomato yellow leaf curl virus	98.25	99.65
Overall Accuracy		97.30	98.78

Table 2. Classification accuracy of the proposed system.

Leaf Disease	Bacterial blight	Early blight	Late blight	Late blight	Septoria leaf spot	Yellow leaf curl	Accuracy
Bacterial blight	397	0	0	1	0	0	99.25
Early blight	0	398	0	1	0	1	99.5
Late blight	1	0	400	0	0	1	100
Leaf mold	0	0	0	400	0	0	100
Septoria leaf spot	0	1	0	0	399	0	99.75
Yellow leaf curl	0	0	0	0	0	400	100
Average							99.75

4. Discussion Conclusion and Scope for Future Work

One of the most important agricultural applications is plant disease detection, which allows farmers to detect diseases early on and enhance yield. At the field, the camera sensor can be utilized to capture photographs at regular intervals or when prompted. After a pre-processing stage and a foreground-based segmentation stage, the ORB features are recovered from the collected image. Tomato leaf photos with six distinct illnesses are used to train the RFC algorithm. The suggested system is simulated in MATLAB, and the findings show that the suggested system has an

overall detection accuracy of around 99 percent and a classification accuracy of about 99.75 percent, according to the results. Compared to the k-means clustering algorithm, the proposed method outperforms it.

References

1. Upadhyay, A.; Upadhyay, A.; and Maurya, S. (2017). Regression and artificial neural network based improved classification of LISS-III satellite image. *Proceedings of the 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*. p Mysore, India, 917-921.
2. Shanthi, H.J.; and Anita, M. (2014). Heuristic approach of supervised learning for intrusion detection. *Indian Journal of Science and Technology*, 7(is6), 11-14.
3. Khirade, S.D.; and Patil, A.B. (2015). Plant disease detection using image processing. *Proceedings of the* 2015 *International Conference on Computing Communication Control and Automation*. Pune, India, 768-771.
4. Linda, G.M.; Themozhi, G.; and Bandi, S.R. (2020). Colour-mapped contour gait image for cross-view gait recognition using deep convolutional neural network. *International Journal of Wavelets, Multiresolution and Information Processing*, 18(1), 1941012.
5. Themozhi, G.; and Ramareddy, S. (2012). Output voltage regulation in ZVS bidirectional DC to DC converter using neural network. *The IUP Journal of Electrical & Electronics Engineering*, V(3), 34-47.
6. Kaggle (2023). Random forest classifier. Retrieved October 5, 2023, from https://www.kaggle.com.
7. Singh, V.; and Misra, A.K. (2016). Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture*, 4(1), 41-49.
8. Dhakate, M.; and Ingole, A.B. (2015). Diagnosis of pomegranate plant diseases using neural network. *Proceedings of the 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*. Patna, India, 1-4.
9. Aasha Nandhini, S.; Hemalatha, R.; Radha, S.; and Indumathi, K. (2018). Web enabled plant disease detection system for agricultural applications using WMSN. *Wireless Personal Communications*, 102(2), 725-740.
10. Surendiran, J.; Saravanan, S.; and Kmanivannan, K. (2016). Detection of glaucoma based on color moments and SVM classifier using k mean clustering. *International Journal of Pharmacy and Technology*, 8(3), 16139-16148.
11. Bhange, M.; and Hingoliwala, H.A. (2015). Smart farming: Pomegranate disease detection using image processing. *Procedia Computer Science*, 58, 280-28.
12. Karthickmanoj, R.; Padmapriya, J.; and Sasilatha, T. (2021). A novel pixel replacement-based segmentation and double feature extraction techniques for efficient classification of plant leaf diseases. *Material Today Proceedings*, 47(Part 9), 2048-2052.