https://doi.org/ 10.48047/AFJBS.6.5.2024. 11442-11457

African Journal of Biological Sciences

Automated Early Detection of Plant Diseases a Machine Learning Approach

Ashish Nagila, Abhishek K Mishra

1. *Ashish Nagila Assistant Professor, Department of CSE, IFTM University, Moradabad , Uttar Pradesh, India Mail id: ashishnagila01@gmail.com
2. Abhishek K Mishra Associate Professor Department of CSE, IFTM University, Moradabad, Uttar Pradesh, India Mail id: abhimishra2@gmail.com

Article History
Volume 6, Issue 5, 2024
Received: 25 May 2024
Accepted: 02 Jun 2024
doi: ***10.48047/AFJBS.6.5.2024. 11442-11457***

Abstract

This study introduces an innovative strategy for the early detection of plant diseases leveraging machine learning algorithms, a crucial step towards bolstering global food security. Despite the importance of early disease detection for mitigating crop losses, inadequate infrastructure often hinders swift and accurate identification. This paper addresses these challenges by utilizing the Random Forest algorithm to distinguish between healthy and diseased plant leaves using a specially curated dataset. Our proposed method involves comprehensive phases of dataset creation, feature extraction, classifier training, and subsequent classification. For the extraction of image features, the Histogram of Oriented Gradient (HOG) technique is utilized, offering a robust method for image analysis. Embracing the capabilities of publicly accessible, large-scale datasets, our method provides a scalable solution to plant disease detection. The empirical results demonstrate promising levels of accuracy in disease monitoring, presenting transformative potential for disease management practices in agriculture. Particularly in resource-constrained settings, our machine learning approach paves the way for efficient and effective plant disease detection, carrying significant implications for future crop protection strategies.

keywords:
Plant diseases, early detection, machine learning, random forestry algorithms, directed gradient histograms (HOG), image classification, feature extraction, agriculture, food security.

1. Introduction

The agriculture sector in the world confronts several difficulties. Plant diseases significantly reduce food output each year [EU, Erke (2006)]. If we wish to reduce the harm that plant diseases do to food security, farmer earnings, and companies in general [Savary, S., Fike, A., Oberto, JN, & Hollier, K. (2012)], early diagnosis of plant diseases is crucial.

Traditional techniques for identifying plant diseases rely significantly on specialised expertise and need ongoing manual monitoring and performance assessment. However, these procedures are frequently arduous, drawn-out, and prone to mistakes [Pessybridge, SC, Nelson, ME (2015)]. Consequently, a method for monitoring plant diseases that is automated, quick, and accurate is required.

It may be challenging for farmers in rural areas to discern between several illnesses that might destroy their crops. To find out what the infection is, you don't need to go to the farm office. Our major objective is to identify the illnesses that are spread from the plant form using image processing and machine learning.

Food poverty is made worse by pests and diseases that decimate portions of crops and fields, making it challenging to grow adequate food. Furthermore, in many underdeveloped nations, little is known about how to manage and control illnesses and pests. Harmful microorganisms, inadequate disease control, and extreme climate change are a few of the key causes of the fall in food supply.

A number of innovative techniques have been created to lessen the amount of labour required following harvest, improving farming's sustainability and yields. The polymerase chain reaction, gas chromatography, mass spectrometry, thermography, and hyperspectral techniques have all been employed in laboratories to identify illnesses. These techniques take a lot of time and don't save much money.

Disease detection has long been done using servers and mobile devices. High-resolution cameras, high-performance computing, and numerous integrated devices enable automatic illness detection.

In order to increase the accuracy of the findings, contemporary tools like machine learning and deep learning techniques have been applied. Numerous research have examined the detection and diagnosis of plant illnesses using conventional machine learning techniques such random forests, artificial neural networks, support vector machines (SVM), fuzzy logic, k-means, and convolutional neural networks.

A group of learning techniques called random forests are used for classification, regression, and other applications. A forest of decision trees is created during training. Both category and numerical data may be processed using random forests. The training dataset does not need to be too comparable to the actual world, unlike decision trees.

In computer vision and image processing, a directional gradient histogram (HOG) is a component description used to recognise objects. Three component descriptions are utilised in this instance:

1. Hu times
2. Fabric from Haralick
3. A green bar chart

Leaf form is mostly determined by hu time. The colour histogram demonstrates how the colours are dispersed in the image, and the Haralik technique is used to determine the leaf structure.

Recent advancements in machine learning software have shown that there could be a solution to this issue. For the detection and classification of plant diseases in particular, image classification techniques are well established [4].

This study shows how machine learning may be used to automatically and early detect plant illnesses. To categorise plant illnesses based on images of leaves, our technique employs a random forest algorithm and directed gradient histogram (HOG) feature extraction. By giving farmers access to real-time data on the condition of their crops and enabling them to react more swiftly to illness, these automated technologies have the potential to revolutionise farming.[5]

This research extends previous work by giving a detailed examination of the operation of our suggested machine learning model and its implications for the future of agriculture and food security.

2. Literature review

SS Sannakki and VS Rajpurohit have studied the categorization of pomegranate illness using backpropagation neural networks [6]. This method largely focuses on identifying intact from damaged parts, with additional clues coming from colour and structure. In this illustration, categorization was carried out via neural network techniques. The key benefit is that, after extracting the chrominance layer from the image by converting it to L*a*b, the marking is calculated with an accuracy of 97.30%. The fact that it can only be applied to a small number of plant species is its main flaw.

[7] Moment Hu is a method that PR Rothe and RV Kshirsagar explain in their work "Identification of Cotton Diseases Using Pattern Recognition Techniques" for using to differentiate between various illnesses. BPNN classifiers can handle a broad range of class issues by limiting the amount of energy that may reach an infection site using an active loop model. The average level of the distribution by groups is 85.52%.

In their research, authors Aakanksha Rastogi, Hrithik Arora, and Shanu Sharma employed fuzzy logic and computer vision to identify and evaluate leaf illnesses. Damage severity is calculated using fuzzy logic, texture information is retrieved using GLCM, and damaged regions are segmented using K-means clustering. To gauge the severity of the leaf lesions, researchers employed artificial neural network (ANN) predictions.

Godleaver Ovomgisha, John A. Quinn, Ernest Mwebaze, and James Lwasa invented the idea of automated visual diagnosis of banana wilt and black Sigatoka [9]. For colour histograms, the RGB colour space is transformed into the L*a. *b, HSV, and HSV colour spaces. For classification, area under the curve analysis was employed, and maximum trees with five morphological characteristics were built from maximal components. These methods included Naive Bayes, Decision Tree, Random Forest, Extremely Random Tree, Nearest Neighbours, and SV. Randomised trees provide several benefits, including the excellent performance of seven classifiers, real-time data access, and programming flexibility.

[10] Support vector machines are used by Wang Tian, Chunjiang Zhao, Shenglian Lu, and Xinyu Guo in "SVM-Based Multiple Classification System for Wheat Leaf Disease Recognition" to categorise wheat leaves and explain how to recognise the illness. of HIS describes colour qualities as formal parameters using the GLCM and seven stable moments. displays a color's RGB values. The MCS and SVM classifier were used to identify wheat illnesses during the dormant season.

3. Methodology

A variety of tests should be carried out to see if the leaves are healthy. Preprocessing, feature extraction, classification, and classifier training are the four primary phases. Preprocessing results in a reduction in all picture sizes, which is maintained. The next step is to use HOG to extract features from the validated picture. HoG [11] is a description of the functionality utilised for object detection. This feature's gradients provide details about the look of objects and the general structure of an image. The benefit of HoG feature extraction is the availability of extra cells. This does not affect any changes. In this example, three function descriptors were used.

Images with prominent image pixel features, or "Hu moments", are illustrative. Here, the Hu moments help describe the particular shape of the leaf. Hu events are measured on a channel. Hu moments cannot be calculated until the RGB data is converted to grayscale. Here, the forms are described using different terms.

Halal Leaf Texture: Healthy and diseased leaves usually have a contrasting texture. Here, the halal texture attribute is used to distinguish between healthy and diseased leaf patterns. This approach is based on an adjacency matrix representing the coordinates (I, J). The frequency with which the I pixel occupies the adjacent J pixel is used in texture calculations [12]. Before defining a halal pattern, the image must be converted to grayscale.

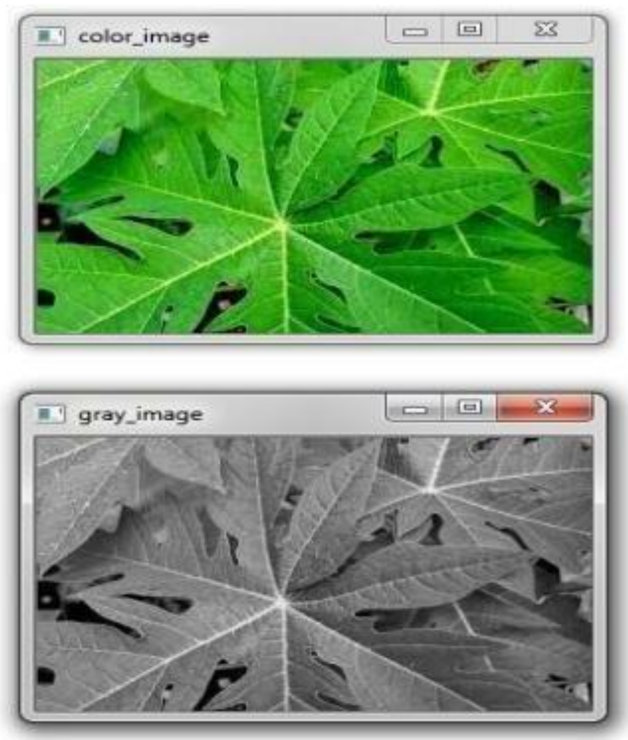


Figure 1: Converting sheets from RGB to grayscale.

Color histogram: A color histogram shows the distribution of colors in an image. The color set is converted from RGB to HSV before creating graphics. RGB images should be converted to HSV because the HSV model more closely matches the color perception of the human eye.

The number of pixels belonging to each color category is indicated in the histogram [13].

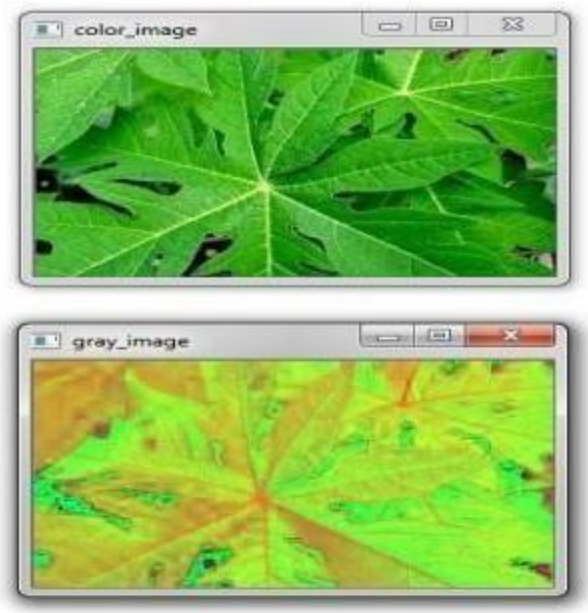


Figure 2. Converting RGB Sheets to HSV

4. Description of the algorithm

Here we implement an approach using a random forest classifier. They are adaptive and have two purposes: regression and classification. Compared to other machine learning methods such as support vector machines, naive Gaussian arrays, logistic regression, and linear discriminant analysis, random forest performed better on less data. Take a look at the attached image to see how our proposed approach works.

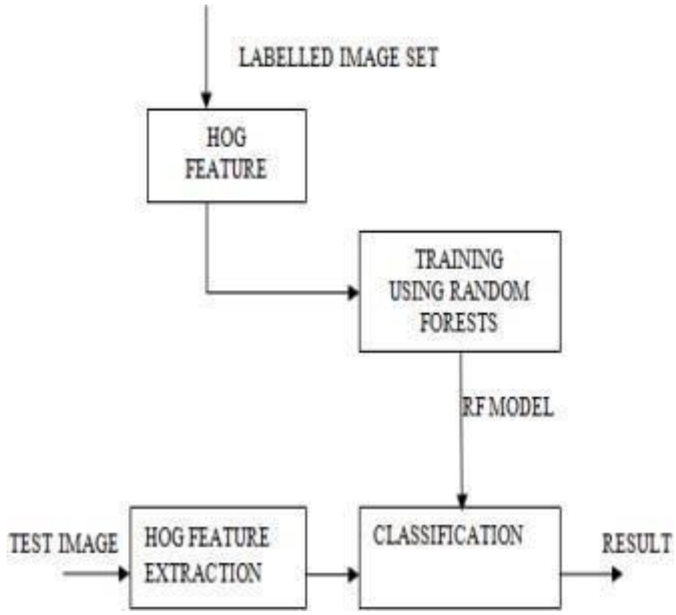


Figure 3. Architecture of the proposed model

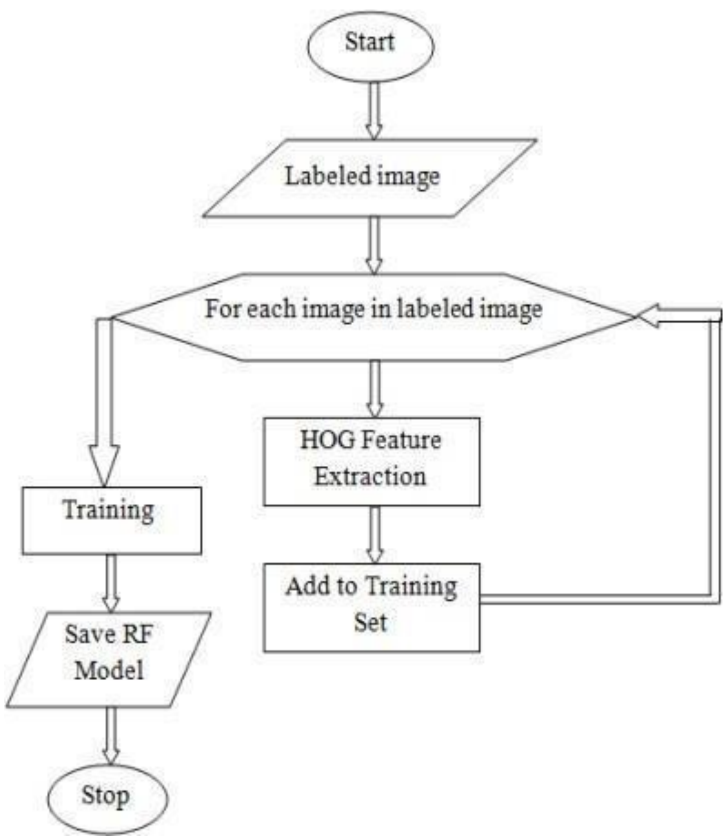


Figure 4 - Training flowchart

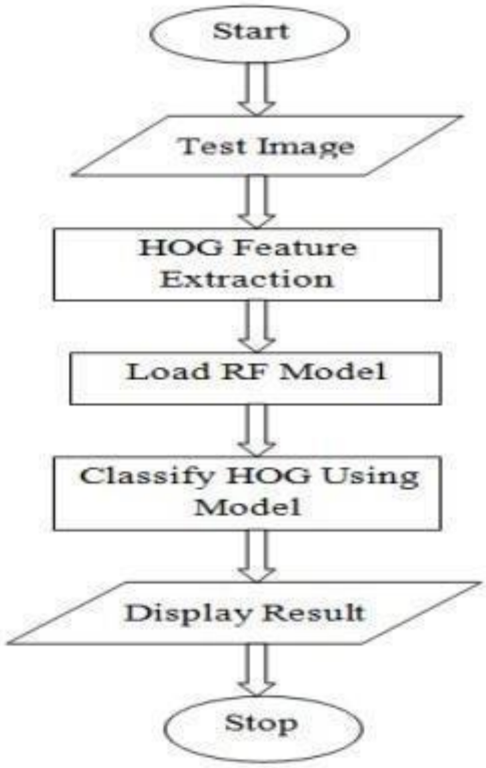


Figure 5. Classification block diagram

A given data set contains "training data" and "test data". The feature vector for the training dataset is created using HoG feature extraction. The random forest classifier is trained using an input feature vector. In Figure 3, we can see that the trained classifier also comes with test data feature vectors created using HoG feature extraction for prediction.

HoG feature extraction transforms the labeled training samples into perfectly matched feature vectors, as shown in Figure 4. The training samples are updated to contain the reconstructed feature vectors. The learned feature vectors are trained using a random forest classifier [14, 15]. The generation of feature vectors for test images by HoG feature extraction is shown in Fig. 5. Feature vectors generated by the constructed model are used to predict future outcomes by the trained and stored classifier.

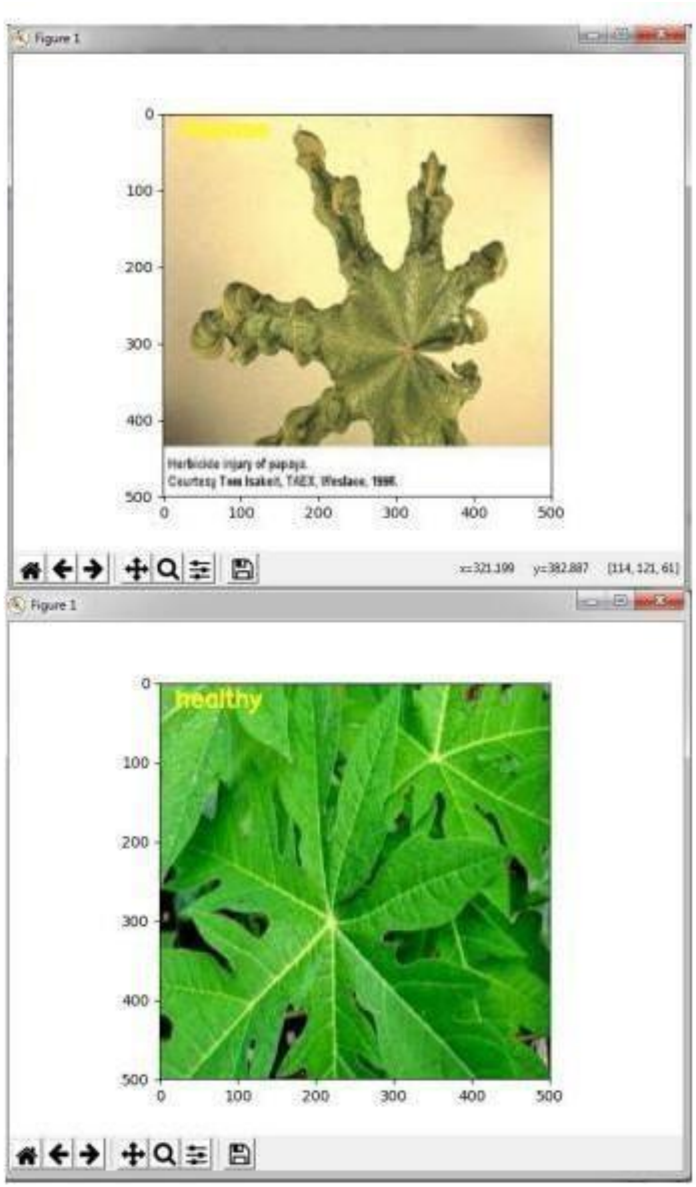
Random Forest algorithm:

- ➢ Select k random features from a total of 'm' features where k << m.
- ➢ Among the 'k' features, calculate the node 'd' using the best split point.
- ➢ Split the node into daughter nodes using the best split.
- ➢ Repeat the steps 1 to 3 until 'l' number of nodes has been reached.
- ➢ Build the forest by repeating steps 1 to 4 for 'n' number of times to create 'n' number of trees.

The forest, as a whole, will be used for making the final prediction. For a new input vector, the Random Forest algorithm will make 'n' predictions (one prediction from each tree). The final prediction will be decided by the majority vote.

5. Results

The first step in image processing is to convert the original RGB file to a grayscale version. Indeed, Hu shape descriptors, moments and halal features can only be computed in a single channel. Therefore, Hu moments and halal properties cannot be calculated without first converting RGB to grayscale. As an example, consider Figure 3. To calculate a histogram, we first need to convert an RGB image to an HSV (Hue, Saturation, Value) image as shown in Figure 4. In Figure 6, you can see: In other words, I'm trying to use a random forest classifier to determine if a leaf is diseased or healthy.



In fig. 7 shows the final result of the classifier.

Compare machine learning algorithms in the table and graph below.

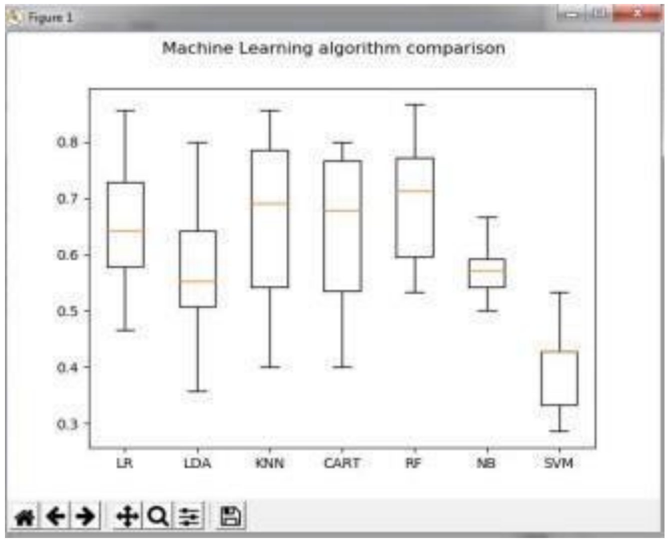


Figure 8. Evaluation of some machine learning models

Figure 9 - Comparative table

Machine learning model	**Accuracy (%)**
logistic regression	65.33
Support vector machine	40.33
K nearest neighbor	66.76
Cart	64.66
random forest	70.14
naive bayes	57.6

Table showing a comparison of the accuracy of various machine learning models in detecting plant diseases.

It can be seen from the table that the random forest algorithm outperforms other models with an accuracy of 70.14%, which highlights its effectiveness in early detection of plant diseases. In contrast, the support vector machine has the lowest performance with an accuracy of 40.33%.

This comparison provides valuable insights into selecting the best machine-learning models for plant disease detection.

6. Conclusion

The system aims to detect abnormal growth patterns of crops, whether in natural or artificial environments. Objects in photographs stand out best against a neutral background. It was compared to various machine learning models to assess the accuracy of the strategy. The model was developed using 160 photos of papaya leaves and a random forest classifier. This model offers an astonishing classification accuracy of nearly 70%. SIFT (scale invariant feature transformations), SURF (fast reliable features), and DENSE (dense networks using random features) using the BOVW (Bag Of Visual Word) framework can be integrated with local features during training. An example of a global function.

Future Work

While our results demonstrate a promising direction for the application of machine learning in plant disease detection, there are several ways that this work could be extended to provide more comprehensive and practical solutions.

1. Expanding Datasets: The availability of large, diverse, and high-quality datasets is crucial for the effective training of machine learning models. One possible direction is to expand our current dataset by adding more classes of plant diseases or including other types of plants. Data augmentation techniques, such as rotation, flipping, and scaling, could be utilized to artificially increase the size of the dataset.

2. Experimenting with Different Models: While the Random Forest algorithm was used in this study, there are many other machine learning and deep learning models that might perform better. Future work could involve testing different models, such as Convolutional Neural Networks (CNNs), Support Vector Machines (SVMs), or other ensemble methods.

Ashish Nagila / Afr.J.Bio.Sc. 6(5) (2024). 11442-11457

3. Real-time Detection: Building a real-time plant disease detection system would be beneficial for farmers and agricultural workers, allowing them to detect diseases as soon as they appear. This would involve optimizing the model for real-time performance, possibly through the use of edge computing or more efficient algorithms.

4. Multi-Label Classification: In the real world, a plant may be affected by more than one disease at the same time, or by a disease and a pest simultaneously. In future work, we could adapt our model to predict multiple classes for each image, known as multi-label classification.

5. User-Friendly Applications: In addition to improving the accuracy and performance of the model, a key direction for future work would be to develop user-friendly applications for non-technical users. This could involve developing a mobile app with an intuitive user interface, allowing farmers to easily take pictures of their crops and receive immediate feedback.

6. Incorporation of Additional Information: Plant diseases can be influenced by various environmental factors, including temperature, humidity, and soil conditions. The inclusion of such contextual information could potentially enhance the performance of our model and is an interesting avenue for future exploration.

By building on the work presented in this paper, we hope to bring more sophisticated, accurate, and practical tools for early plant disease detection into the hands of those who need them most.

- Competing Interests: Not Applicable

- Funding Information: Not Applicable

- Author contribution: Both Author

- Data Availability Statement: Specially Curated Dataset

- Research Involving Human and /or Animals: Not Applicable

- Informed Consent: Not Applicable

References

[1] EU, Erke (2006). Yield loss due to pests. Journal of Agricultural Sciences, 144(1), 31-43.

[2] Savary, S., Fike, A., Oberto, JN, & Hollier, K. (2012). Crop losses due to disease and their impact on global losses in food production and food security. Food Safety, 4(4), 519-537.

[3] Pessybridge, SC, Nelson, ME (2015). Leaf Doctor: A new wearable app to quantify the severity of plant diseases. Plant Diseases, 99(9), 1310-1316.

[4] Bulan, J., Fouche, S., Theo, J., St. Charles, PL (2019). Overview of spectral methods (bands and indices) based on high-resolution satellite imagery for monitoring and detecting pressures on vegetation. Remote Sensing, 11(9), 1121.

[5] Camilaris, A., & Prenafeta-Boldo, FX (2018). Deep learning in agriculture: research. Computers and electronics in agriculture, 147, 70-90.

[6] Sannakki, SS, Rajpurohit, VS (2015). Pomegranate disease classification based on backpropagation neural network. International Journal of Research in Engineering and Technology (IRJET), Volume 2: 02 | May 2015

[7] PR Rote and RV Xirsagar (2015). Identification of cotton leaf disease using pattern recognition technology. International Conference on Comprehensive Computing (ICPC).

[8] Rastogi A., Arora R., Sharma S. (2015). Leaf disease detection and classification using computer vision technology and fuzzy logic. Second International Conference on Signal Processing and Integrated Networks (SPIN).

[9] Owomgisha, G., Quinn, JA, Mwebaze, E., & Lwasa, J. (2014). Automated visual diagnosis of banana wilt and black Sigatoka. Before the first international conference on the use of mobile ICTs in Africa.

[10] Tian J., Zhao K., Lu S., Guo X. (2010). SVM-based multiple classification system for wheat leaf disease recognition. Proceedings of the Trusted Computing Conference 2010 (CDC'2010).

[11] Yun S., Xianfeng W., Shangwen Z. and Chuanlei Z. (2015). PNN-based crop disease recognition using leaf image characteristics and weather data. International Journal of Agricultural Biotechnology, volume 8, no. 4, p.60.

[12] Valedo, JGA (2013). Digital imaging techniques for the detection, quantification and classification of plant diseases. Springer Plus, Volume 2, No. 660, pages 1-12.

[13] Chaglayan A., Guklu O., Dzhan AB. (2013). An approach to plant recognition using the shape and color characteristics of leaf images. At the International Conference on Image Analysis and Processing (pp. 161-170). Springer, Berlin, Heidelberg.

[14] Zhen S., Wang Z., Islam A., Chang Y. and Li S. (2014). Direct assessment of biventricular cardiac volumes using a regression forest. In: Medical Imaging Computing and Computer Intervention - adopted for MICCAI 2014.

[15] Wang P, Chen K, Yao L, Hu B, Wu S, Zhang J et al (2016). Multimodal classification of mild cognitive impairment based on the partial least squares method.