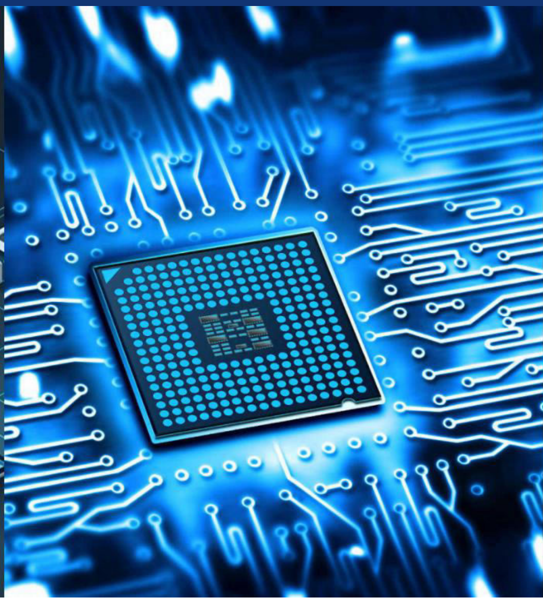
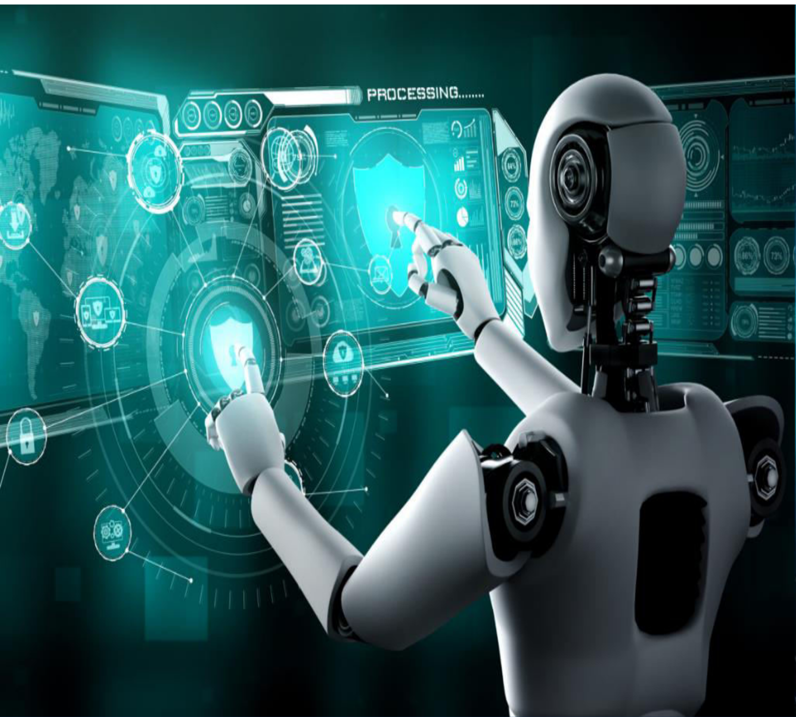
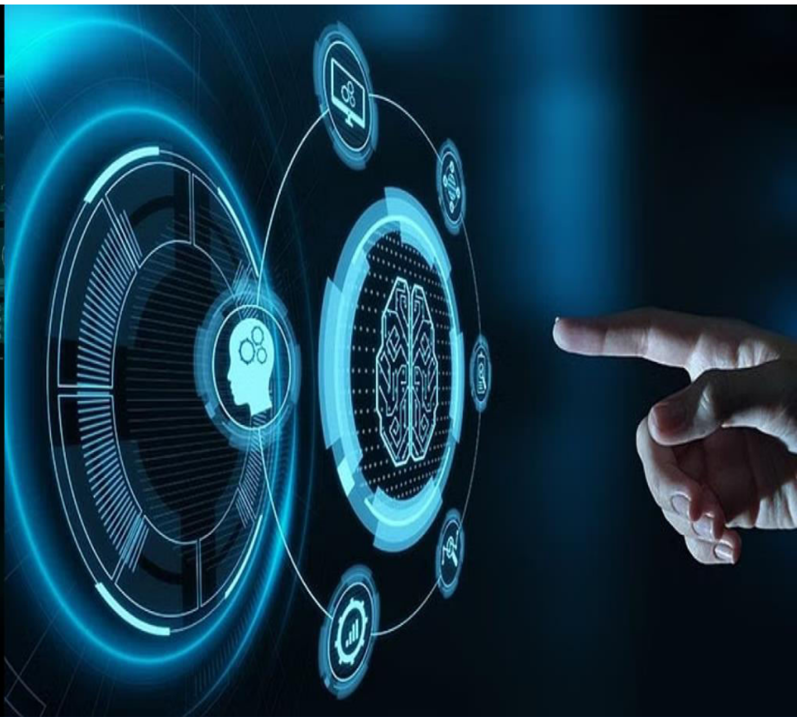
ISSN(O): 2320-9801
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771

Volume 13, Issue 3, March 2025

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Machine Learning Models for Plant Leaf Detection and Classification

Chilakala Vijay Kumar[1], Divya Erry[2]

Department of CSE (AI&ML), CMR Institute of Technology, Hyderabad, India[1,2]

ABSTRACT: Developing dedication to developing effective techniques for the early detection and classification of plant has been prompted by the need for environmentally friendly farming methods. Given their capacity to examine massive amounts of data and generate useful knowledge, machine learning (ML) techniques are growing as powerful tools in this field. The paper provides a summary of the most recent advances in machine learning frameworks designed particularly for identifying and classifying plant classification's. The study begins by exploring the importance of early detection of classification's in agriculture and the challenges related to conventional techniques. The fundamental concepts of machine learning (ML) and its applications in plant detection are then described, with an accent on the functions of feature extraction, feature selection, and classification algorithms. Machine learning has been widely used for classification and recognition tasks in numerous fields, primarily in the biological fields of study, as science and technology have developed. Researchers and laypeople may find simpler to quickly identify plant species with the use of an automated plant species identification system. While machine learning is better at providing deeper information about images, it is dependable for feature extraction. Pre-processing, segmentation, feature extraction, and classification are all related to the research field of plant species identification.

KEYWORDS: Machine Learning, Support Vector Machine (SVM), Random Forest

I. INTRODUCTION

One of the interesting regions when machine learning techniques are utilized for identifying between species is plant species classification. Automation in plant species classification occurs by extracting features from the plants. The most commonly employed feature in developing such automated plant classification systems is leaf shape, except other features such as texture, colour, and veins can also provide additional information that may help in the automation process. As science and technology evolved, machine learning has grown in prevalence for classification and recognition tasks in many domains, especially in the biological fields[1].

Conventional classification diagnostic techniques generally depend on observation by skilled agronomists, which can be a laborious, subjective, and human error-prone procedure.
Given this, there has been growing interest in employing machine learning (ML) approaches to enhance and automated the process of identifying and classifying classification's in plants. A particular component of artificial intelligence is machine learning, that includes creating algorithms that enable computers acquire knowledge and generate predictions and choices from data without requiring for specific programming. In a number of fields, including machine learning and computer vision, models have shown substantial potential by analysing images and identifying significant patterns. ML models may be trained on a large amount of photos showing both healthy and harmed plants for the purpose to identify plant classifications. They acquire the unique characteristics that differentiate various classifications through this process. This introduction seeks to provide a thorough overview of the way machine learning models are used in the identification and categorization of plant classifications. It will go over the disadvantages of traditional techniques, the possible advantages of ML-based methods, and the challenges involved with applying such models in actual agricultural environments[2].

Identifying and diagnosing classifications which impact plants is commonly referred to as plant classification detection. Early diagnosis is necessary for reducing the spread of classification's and reducing losses to agriculture. Visual examination is one of the most popular ways to identify plant classification's. Plant pathologists, agronomists, and farmers inspect plants for signs of classification, including discolouration, wilting, lesions, and aberrant growth. Certain signs

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

that can be used to determine the type of illness present are frequently seen in plant classification's. Spots on leaves, reduced growth and development, leaf yellowing, rotting, and additional symptoms are possible. Researchers and plant pathologists examine these signs to determine the root reasons.

The Flavia dataset has been used for this study. Data transformation is the method of data pre-processing that is used here. The image dataset has been transformed into the numerical dataset. A few classification techniques use features that have been obtained form the CSV file for training and learning. Support Vector Machine (SVM), Random Forest (RF), and k-nearest-neighbour (k-NN) are the classifiers used in this study. Because contours are memory-efficient through removing contain points, they have been employed in place of the conventional Sobel edge detection method. The machine learning model will be assessed using an additional validation dataset once it has been successfully trained in order to determine its efficacy and adjust parameters as needed[3]. The model is prepared for practical use after it exhibits good performance on an independent test dataset. This can entail incorporating it into programs, applications, or systems that evaluate images of plants and provide farmers with comments.

Figure 1: images of the 32 different species identified in the Flavia dataset.

II. RELATED WORK

With the objective to extract the vein structure, Bao Quoc Truong et al. [4] developed a developing vein classifier based on Ant Colony algorithms and genetic algorithms (GA). A plant identification technique based on a combination of color and textural features was proposed by Larese et al. an automated method for identifying legume leaves based only on vein morphology. A Random Forests technique was used to identify the vein anatomy once basic measurements made[5]. By adding new images to the dataset on a regular basis and retraining the model to account for changing environmental factors or classification trends, machine learning models for plant classification identification may be constantly enhanced. Researchers and agricultural professionals may create accurate and efficient plant classification detection systems by utilizing machine learning techniques. This will allow for early intervention and successful management strategies to protect crop health and maximize agricultural yield.

Since machine learning (ML) techniques are so effective at analysing large image collections and correctly classifying plant health issues, they are being used more and more to diagnose plant classification's. Gathering a sizable collection of photos of both healthy and classification plants is the initial stage in using machine learning for plant classification identification. Drones, cell phones, cameras, and other imaging equipment may all be used to take these pictures. Preprocessing techniques can be used to standardize image sizes, modify lighting, and eliminate noise or superfluous

© 2025 IJIRCCE | Volume 13, Issue 3, March 2025| DOI: 10.15680/IJIRCCE.2025.1303019

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

background information after the image collection has been gathered. This helps to keep machine learning algorithms consistent and improves their effectiveness.

Plant classification detection machine learning models can be developed and evaluated with the use of datasets that include Plant Village and Fungi DB. To increase model robustness and generalization, data preparation methods like picture augmentation and normalization are crucial Metrics like accuracy, precision, recall, and F1-score are frequently used to evaluate how well machine learning models perform in classification detection tasks. Comparative analyses are made easier by benchmarking studies, which also assist in determining the best methods and techniques. Environmental variability and the lack of well-labelled information are challenges in applying machine learning to plant classification diagnosis. In practical applications, models' inability to be transferred across many plant species and illnesses continues to be a major drawback [6].

III. METHODOLOGY

Plant species classification is seen to be essential to managing the ecological balance and enables botanists to identify large plant types with less human labour. The curves approach is a data pre-processing technique that transforms the plant picture data set into a numerical data set with the required characteristics as attributes. In addition, three machine learning algorithms Support Vector Machine (SVM), Random Forest (RF), and k-Nearest Neighbour (k-NN) are used for classification, and their accuracy is evaluated.

1. Support Vector Machine

Support Vector Machines (SVM), a supervised machine learning method, has been recognized as one of the most effective methods for classification because of its exceptional ability to handle high dimensional space and rapidly separated input points. Implementing linear SVM on feature-mapped data can increase classification performance and run quickly with less storage. Because this study uses a multi-class dataset, linear SVM with a "one versus all" method was implemented. In order to increase the characterization effectiveness and precision of SVM, the main goal of this study is to create an improved SVM model that can be used to cardiac classification data sorting[7].

The choice tree simulates the traditional method for creating archives by using a tree that represents both positive and negative inquiries. arrangement. The corresponding categorization of archives is represented by the leaves in the decision tree structure, while the connections of highlights that lead to those classes are represented by the branches. The tree expands until all of the material has been correctly or incorrectly sorted. Without much effort, the effective choice tree may classify a record by placing it in the root hub and letting it go through the inquiry structure until it reaches a particular leaf that addresses the goal of the archive's layout. Compared to other decision aids, the choice tree order approach offers a few advantages.

2. Random Forest

Random Forest is an adaptable, simple to operate machine learning method that typically yields fantastic results even when no hyper-parameter tuning is done. The fact that it is simple to use and can be applied to both grouping and relapse chores makes it unique among the most popular computations. Given this, it is possible to deduce the suitability and practicality of the random forest computations as well as several additional essential elements of it. Because it can be used for both regression and classification, this classifier is among the most used algorithms available today. Decision trees' primary flaw is their propensity to overfit the data. Making many decision trees on arbitrary subsamples of the data is one way to stop this. When a new data point has to be forecasted, all decision trees will calculate it and determine a majority vote. The data's class will be decided by this vote. As a result, the accuracy is higher than with a decision tree alone. Accuracy is closely correlated with the number of decision trees. This collaborative decision-making procedure, secured by multiple trees with their insights, provides an example that produces precise results[8]. Random forests are commonly used for classification and regression functions, which are noted for their capacity to handle complicated data, avoid overfitting, and produce reliable forecasts in varied situations.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

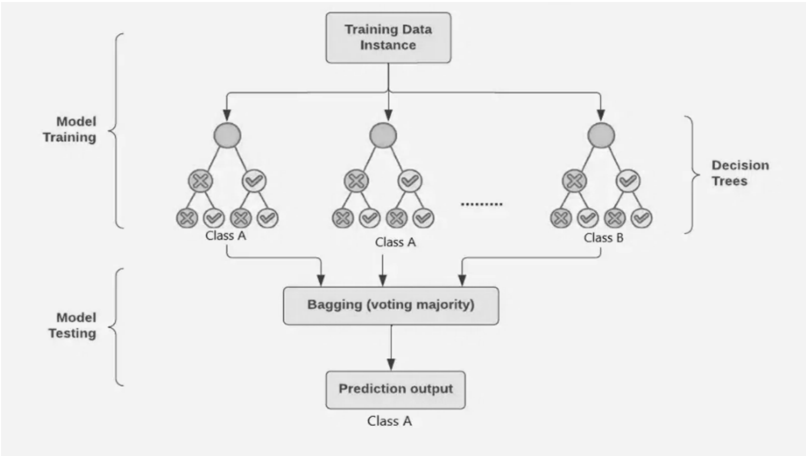


Figure 1: Implementation of Random Forest

3. k-Nearest Neighbour (k-NN)

A classification method termed a k-NN classifies data based on the majority vote of their neighbours. In the present research, the number of neighbours is set at one employing the city block distance metric. A K-Neighbours Classifier object is instantiated by the application with k set. When fitting the classification algorithm to the training data, the classifier's name is shown. It then uses the K-Neighbours Classifier with k=3 to do cross-validation. The software then makes predictions on the test data using the trained classifier, calculates the accuracy of these predictions, and shows the accuracy. Finally, the application uses the trained classifier to predict each class's probability, computes the log loss using the predicted probabilities log loss is printed. However, it is mostly used to production classification prediction complications. Finding the k-closest data points to a given test data point and using these nearest neighbour's to get a prediction is the fundamental notion fundamental KNN[9]. The number of neighbour's to take into account is represented by the hyperparameter k, which must be adjusted. The KNN approach allocates the test data point to the class that is most prevalent among the k-nearest neighbour's in classification tasks. Stated otherwise, the projected class is the one with the greatest number of neighbour's. The KNN approach uses the average of the values of the k-nearest neighbour's to choose the test data point for regression issues.

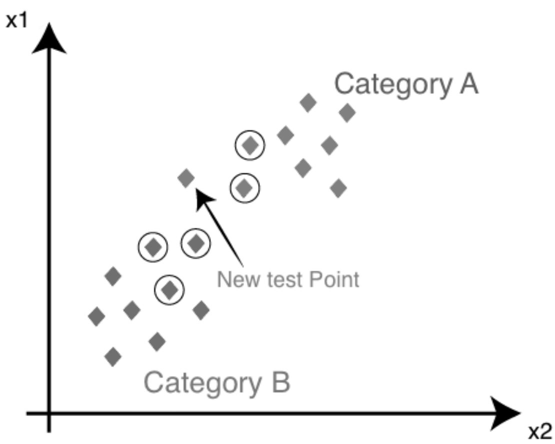


Figure 2: Implementation of k-Nearest Neighbour (k-NN)

© 2025 IJIRCCE | Volume 13, Issue 3, March 2025| DOI: 10.15680/IJIRCCE.2025.1303019

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. RESULTS

In the context of this work, we had conversations about the performance of the various models. Following data preprocessing, the data was separated into training and testing sets in a 70:30 ratio and fitted to three different classifiers. We subsequently extended our work by testing with some real-time plant species images. TABLE 1 discusses each classifier's performance. It was determined that the Random Forest classifier performed the best, with an accuracy of 85%. Nonetheless, the accuracy of the other classifiers was at least 80%. Therefore, we can say that RF is a very good fit for contours in this study's feature extraction model. On the other hand, k-NN and SVM performed worse with the contours. The traditional feature extraction approach, on the other hand, takes a long time and extracts each sort of feature independently[10].

Table 1: Classification precision of several classifiers

S.no	Classifiers	Accuracy
1	Random Forest	85%
2	Support Vector Machine	80%
3	k-Nearest Neighbours	82%

In machine learning, a real dataset is often split into training and testing sets to assess a model's performance. A training set and a testing set are created from the original dataset. The machine learning model is developed using the training set, and its performance on untested data is assessed using the testing set. The training set is used to train the machine learning model. The model learns the patterns and correlations in the data to make predictions or classify it. The input attributes (X_train) and the labels or goal values (Y_train) that correspond to them are used to train the model. To determine how well the model applies to fresh, untested data, we may examine how well it performs on the testing set. To identify how well the model applies to fresh, untested data, we may examine how well it performs on the testing set. The data-set will be gathered from Kaggle and will include 10 distinct types of plant leaves, each with one hundred photos, assuming the model performs well on the test set and correctly predicts new occurrences. Collectively, SVM classifiers can identify and categorize plant illnesses from photos of leaves. This system gives farmers an automated and effective way to find and diagnose crop illnesses, allowing for prompt treatments and raising the output of agriculture overall.

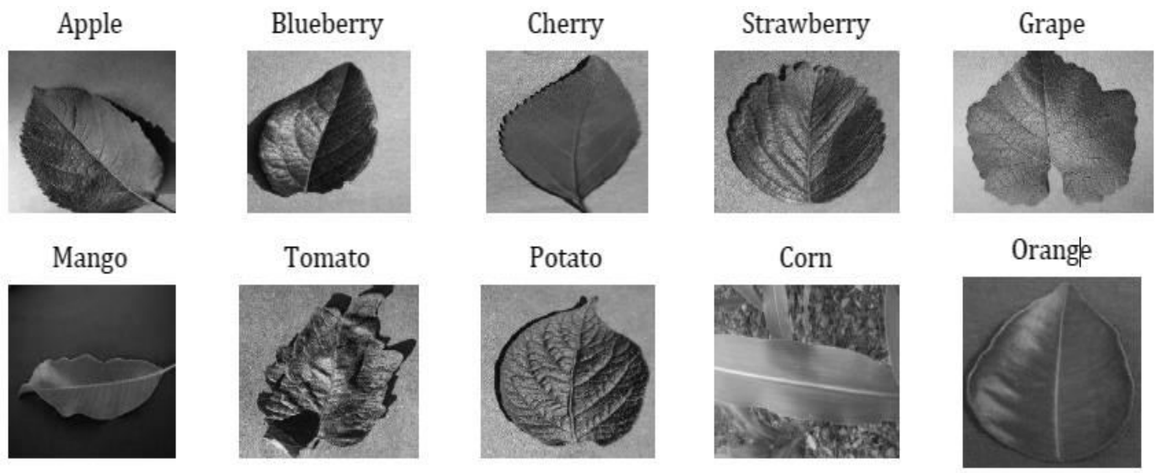


Figure 3: instances of leaf images from the collected dataset

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. CONCLUSION

Correctly determining plant species from their leaves can be performed with the application of machine learning techniques. Plant leaves have been identified using machine learning approaches. The models that performed the best in this regard were Random Forest and k-Nearest Neighbours Analysis, which showed the highest accuracy and the least amount of log loss. While K-Neighbours Classifier had a little lower accuracy, SVM functioned satisfactorily. This application may be useful in a number of fields, including botany, ecology, and agriculture, by utilizing machine learning techniques. Python is the best option for developing leaf identification systems since it offers a robust library and capabilities for image processing and machine learning approaches. We can accurately recognize and classify various kinds of leaf using these image processing techniques. Here, we place the photos from trained datasets for real-time detection. We developed the model for this project using suitable methodologies and implementation methods.

REFERENCES

1. Amiruzzaki Taslim, Sharifah Saon , Abd Kadir Mahamad, Muladi Muladi, Wahyu Nur Hidayat(Oct 2021) Plant leaf identification system using convolutional neural network. Bulletin of Electrical Engineering and Informatics (Vol. 10, pp. 3341-3352)
2. Rakibul Sk, Ankita Wadhawan,(Feb 2021) . Identification of Plants using Deep learning: A Review International Semantic Intelligence Conference(Vol 2786 ,paper 51).
3. Adnan Mohsin Abdulazeez1 , Diyar Qader Zeebaree2 , Dilovan Asaad Zebari2 , Thamer Hassan Hameed3 (05/2021). Leaf identification based on shape, color, texture, and vines using probabilistic neural network.
4. Hiep Xuan Huynh, Bao Quoc Truong, Kiet Tan Nguyen Thanh and Dinh Quoc Truong(4 March 2020), Vietnam Journal of Computer Science, Plant Identi¯cation Using New Architecture Convolutional Neural Networks Combine with Replacing the Red of Color Channel Image by Vein Morphology Leaf , (Vol. 7, No. 2 (2020) 197–208), https://dx.doi.org/10.1142/S2196888820500116.
5. Noon, S.K., Amjad, M., Qureshi, M.A., Mannan,(dec 2020) A., “Use of deep learning techniques for identification of plant leaf stresses: A review”, Sustainable Computing: Informatics and Systems, (Vol.28) http://dx.doi.org/10.1016/j.suscom.2020.100443.
6. S. Anubha Pearline, V. Sathiesh Kumar, S. Harini(2019). “A study on plant recognition using conventional image processing and deep learning approaches”, Journal of Intelligent & Fuzzy Systems (Vol. 36, no.3, pp.1997-2004).
7. Chen, Junde, et al. "Using deep transfer learning for image-based plant disease identification." Computers and Electronics in Agriculture 173 (2020): 105393.
8. Kolivand, Hoshang, et al. "A new leaf venation detection technique for plant species classification." Arabian Journal for Science and Engineering 44.4 (2019): 3315-3327.
9. Lee, Sue Han, Chee Seng Chan, and Paolo Remagnino. "Multi-organ plant classification based on convolutional and recurrent neural networks." IEEE Transactions on Image Processing 27.9 (2018): 4287-4301.
10. Singh, Vijai, and Ak K. Misra. "Detection of plant leaf diseases using image segmentation and soft computing techniques." Information processing in Agriculture 4.1 (2017): 41-49.