ISSN 2277-2685
IJESR/Aug. 2018/ Vol-8/Issue-3/1-11
Mr. G. Venkateshwarlu./ International Journal of Engineering & Science Research

Crops Disease Detection using Machine Learning Techniques and CNN

Mr. G. Venkateshwarlu, MCA, MTech. *1, Mr. C. Santhosh Kumar Reddy, MCA *2, Mrs. K. Mary Leena, MBA *3

1. Faculty, Department of computer Science, Siva Sivani Degree college, kompally,sec-bad-100
2. Faculty, Department of computer Science, Siva Sivani Degree college, kompally,sec-bad-100
3. Faculty, Department of Management, Siva Sivani Degree college, kompally,sec-bad-100

Abstract: Crop disease detection is crucial for food security and agricultural productivity. Traditional methods are time-consuming and rely on human expertise. This research proposes an automated system using machine learning techniques to accurately identify diseases affecting crops based on plant leaves. The system uses Machine learning algorithms to analyze images and classify them into healthy or diseased categories. A comprehensive dataset is used to train and validate the model, and transfer learning is employed to enhance performance. Attention mechanisms are incorporated to improve interpretability. The trained model shows promising results in accuracy, sensitivity, and specificity. The system is scalable and adaptable to different crops and diseases, making it applicable across various agricultural settings. This machine learning-based approach could revolutionize agricultural practices, reducing environmental impact and improving crop yields.

Keywords: Plant diseases, Machine learning, Early Detection, Segmentation.

1. Introduction:

Crop diseases are a major problem for the world's economies, despite the importance of agriculture. Conventional approaches to disease diagnosis frequently call for human specialists, which can be costly and time-consuming. The diagnosis of crop diseases has found great support in machine learning, especially in deep learning methods. By employing Machine Learning algorithms, this research seeks to create a dependable and effective method for crop disease identification. CNNs are useful for processing visual data because they can automatically extract hierarchical features from images. The technology is capable of accurately classifying crop health status because it trained the model on a variety of datasets. To enhance generalization performance, transfer learning and attention mechanisms are also combined. This project uses machine learning and image processing to identify diseases of plants from photographs of the leaves. When fitting training data into decision-making models, machine learning aids in comprehension. In order to diagnose plant leaf diseases with the highest level of accuracy, the project analyzes picture parameters. In the past, big farms can incur high costs for disease identification due to the need for visual examination and ongoing monitoring. The suggested method, which makes use of statistical machine learning and image processing methods, is less computationally expensive and requires less prediction time than existing deep learning-based techniques.

There exists a close relationship between climate change and agriculture, as the former can modify the stages and rates of pathogenicity, resulting in physiological differences in the collaboration between the pathogen and the host[1]. Long-term infections may become resistant to pesticides due to inconsistent use, making it harder to fight them as shown in Table 1. Precision farming places a strong emphasis on accurately and quickly interpreting plant illnesses in order to combat persistent pathogenic resistance and lessen the negative consequences of climate change.

In this ever-changing environment, early and accurate illness detection is essential. Although there are a number of ways to identify plant diseases, the main option is through the use of skilled professionals[2]. Utilizing appearance and visual symptoms, automated systems can assist novices and seasoned experts in identifying plant diseases.

Advancements in computer vision present prospects for bolstering and expanding plant security measures, as well as expanding the market for accurate agricultural computer vision applications. The identification and categorization of diseases makes use of standard digital image processing technologies including color detection and threshold[3]. Recently, several domains, including computer vision, pharmacology, and bioinformatics, have shown positive outcomes with deep learning techniques like CNN. Deep learning makes use of the GPU's computational capacity and the ability to work with raw data directly, without the need for handcrafts. This enables the training of deep models and their application in computing parallelism.

Plant disease detection has been greatly enhanced by machine learning and deep learning approaches, allowing for automatic feature extraction and categorization. Because of their recognition and classification capabilities, convolution neural networks (CNNs) are the go-to option for automated plant disease identification. On paddy plants, CNNs have been employed by researchers for disease detection and segmentation, as well as classification and image processing. The greatest results are probably obtained by training deep-learning models with a big dataset, given there is little diversity in the datasets employed. PlantVillage, a plant disease dataset produced by Pennsylvania State University, has 54,305 RGB photos across 38 plant disease groups. Better results will only be possible with improvements to picture databases.

Table 1: Various segmentation techniques

S.No	Techniques	Advantages	Disadvantages
1	Regional based method[4]	This method helps to differentiate image segmentation among interactive and automated techniques and also gives an accurate result analysis.	Requires more computation power and memory.
2	Threshold method[5]	Requires very less computation cost, it is reliable and minimum knowledge of the plants is enough to use this method.	Easily distracts with noise data and choosing threshold value is important.
3	Clustering based method[6]	Based on the k-value this method works better and computation speed is fast.	Doesn't perform better when there are group of clusters.
4	Edge Detection[7]	Gives accurate results with higher contrast images.	Detection of object edges is difficult.

1.2 Role of machine learning techniques to detect diseases in Plants:

A industry that has been receiving a lot of attention lately because of technical breakthroughs like automation and data mining is agriculture, and machine learning techniques are being used there to monitor soil fertility. Unfortunately, the subject of study on data mining in agricultural soil databases is still in its infancy. Although a method for precisely identifying plant illnesses was described, the factors impacting disease detection were not fully understood. Early illness symptoms detection and less monitoring effort are two benefits of automated methods. Furthermore, a review is conducted on crop boundary algorithm prediction that is dependent on geographic information systems. To assist with identification, selection, fertilizer selection, and pest control, technical rice, cocoa, and coffee sustain software is being created. The various leaf images are shown in figure 2.

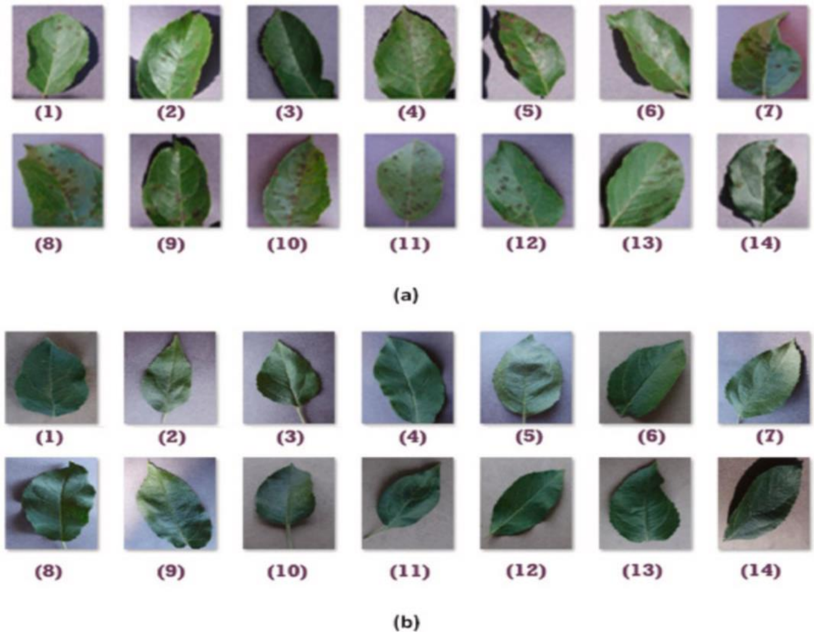


Fig. 1. (a) Diseased leaf images (b) healthy leaf images.

2. Literature survey:

Various techniques for recognizing and categorizing agricultural diseases are included in the literature study. Artificial neural networks for grading and detecting leaf diseases, back propagation neural networks for classifying pomegranate diseases, and pattern recognition algorithms for identifying cotton leaf diseases are a few examples[8]. Additional techniques include fuzzy logic for illness grading, GLCM for extracting textural features, and K-means clustering[9]. The diseases known as black sigatoka and banana bacterial wilt are diagnosed automatically using vision-based methods such as color histograms, SV classifiers, random forests, near neighbors, decision trees, random forests, and extremely randomized trees[10]. Wheat plant diseases are detected offline by SVM-based multiple classifier systems. Through these techniques, diseases in agriculture can be detected more precisely and adaptably[11].

By dividing images into smaller, more manageable chunks, image segmentation speeds up image retrieval[12]. Image classification classifies images into supervised and unsupervised groups, whereas feature extraction minimizes pixels by eliminating important features. Plant diseases can reduce output and result in financial losses, but agriculture is essential to the preservation of food supply[13]. Kulkarni et al. used image processing to detect plant illnesses, obtaining a 91% identification rate following extraction with an ANN classification classifier and Gabor filter[14].

Back propagation neural networks (BPNNs) and digital image processing techniques were used in a system for plant disease identification created in 2015 by S. Khirade et al[15]. The contaminated portions of the leaves were segmented using Otsu's thresholding, boundary detection, and spot detection techniques. For classification, characteristics such as color, texture, morphology, and edges were extracted. When utilizing all of the retrieved features, Shiroop Madiwalar and Medha Wyawahare's analysis of several image processing techniques for plant disease identification produced the maximum accuracy of 83.34%[16]. In order to show how hyperspectral imaging may be used to detect plant diseases, Peyman Moghadam et al. used vegetation indices in the VNIR spectral range to achieve 83% accuracy and entire spectrum to get 93% accuracy[17]. A Bacterial Blight detection system for Pomegranate plants was created by Sharath D. M. et al. based on parameters such as color, mean, homogeneity, SD, variance, correlation, entropy, and edges. A convolutional neural network was used by Garima Shrestha et al.[18]to classify 12 plant diseases with an accuracy of 88.80%.

3. Classification algorithms:

Automation systems that can recognize fruits based on their form, variety, ripeness, and intactness can be created using machine learning approaches and image processing concepts. A work on crop disease identification utilizing artificial intelligence logic and picture pre-processing methodology was proposed by Pushkara Sharma et al. in 2020. The objective is to safeguard plants and manage illnesses, showcasing the effectiveness and precision of the categorization strategy. This concept is efficient and flexible, making it possible to use it in the industrial sector. The usefulness of the model was validated by validation tests, and its 'unknown' function was designed to improve accuracy. A few procedures need to be followed in order to determine whether the leaf is infected or healthy. For example, preprocessing, feature extraction, classifier training, and classification. Reducing all of the image sizes to a single, uniform size is known as preprocessing and the block diagram of the proposed model is shown in figure 2. Green and brown masking images are concatenated using convolutional logic in the image masking concatenation procedure. The process of feature extraction is used to extract distinguishing characteristics—like color, texture, and shape—that can be utilized to identify particular plant diseases. Accurate classification is necessary for these systems to function well. Convolutional neural networks are employed to analyze brown color variation ranges; leaves classified as healthy or unwell have a threshold level of 200. There is also a graphical representation of the predicted processing time for deep machine learning techniques such as CNN, SVM, DT, KNN and RF.

3.1 KNN classifier:

KNN classifier is used for classification and regression issues, diagnosing distribution tree algorithms. Non-parametric, it calculates data attributes. Slow Learning (SL) algorithm, it doesn't require training data, allowing testing without hypothetical mathematical imaginations.

3.2 Decision Tree classifier:

A popular technique for classification and prediction is the Decision Tree Classifier (DT), which uses a tree-based structure with internal nodes for feature tests, branches for test results, and terminal nodes for data labels.

3.3 SVM classifier:

SVM computes hyperplanes with larger margins inside data classes to assist with regression and classification analysis. It can identify support vectors to lower classification errors by splitting hyperplanes into non-intersecting classes, albeit this may not always be applicable.

3.4 Random forest classifier:

RF is a supervised technique that uses a robust decision tree methodology for regression and classification. It is superior to a single DT classifier because it selects better outcomes through a defined procedure and uses prediction results from the entire tree.

3.5 CNN:

This technology takes a different approach from previous neural network approaches in order to identify photos with a certain perspective. In contrast to other picture arrangement calculations, CNNs require less pre-processing. Learning channels are started by categorizing certain highlights at spatial location data. Convolution, average computation, maximum pooling, and maximum pixel are all handled by symmetric and asymmetric blocks in the 48-layered deep CNN network known as Inception-V4. The Oxford University-developed VGG-16 uses softmax as the activation norm and concentrates on the max pool, fully connected layer, and padding convolution layer.

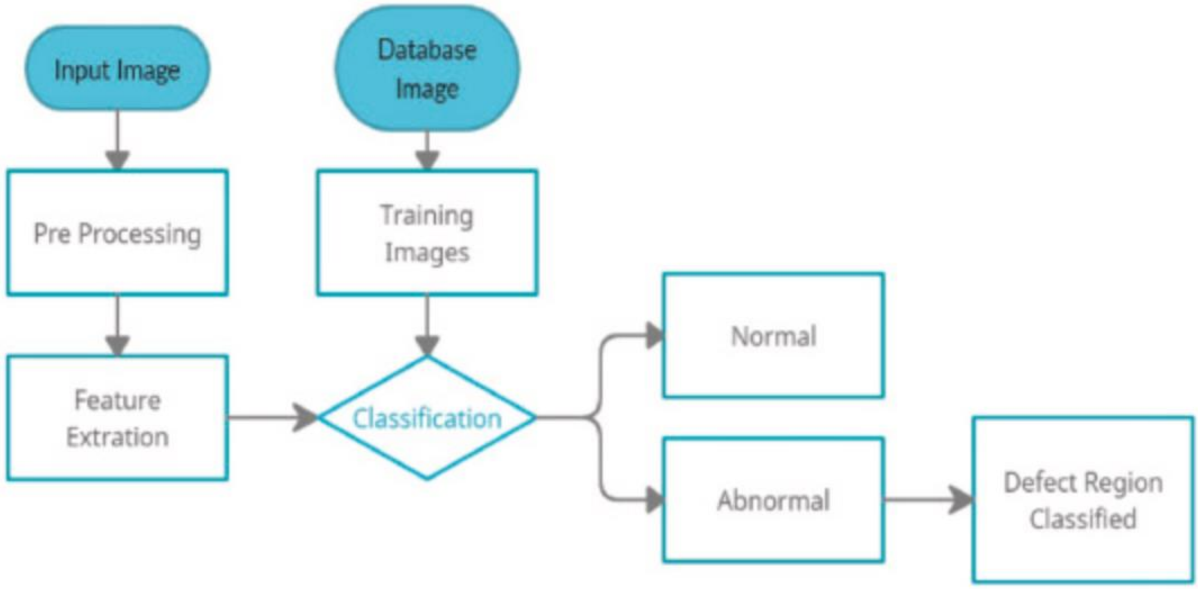


Fig 2: Block diagram of Proposed model

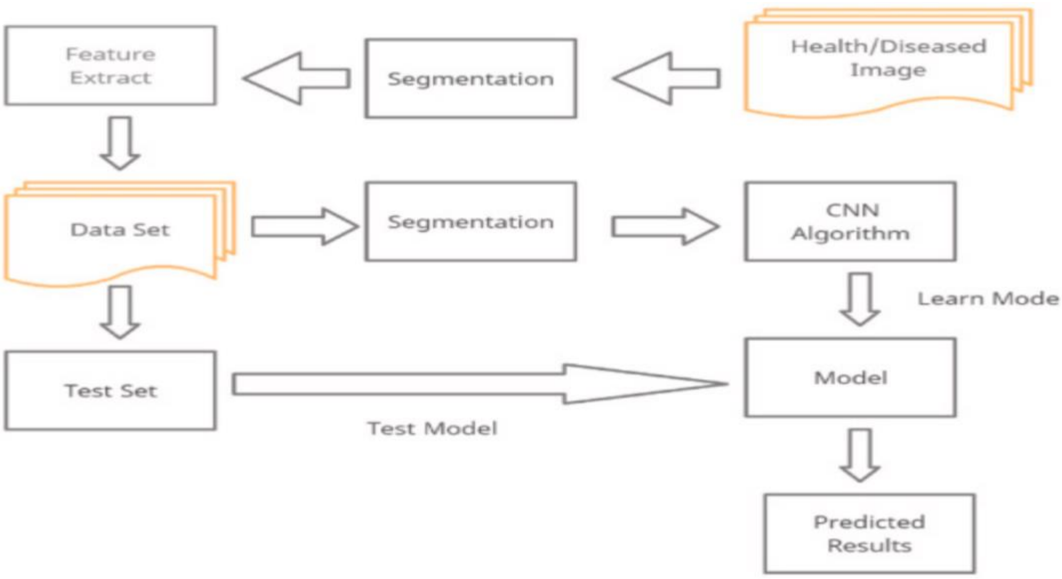


Fig 3: Diagram of Machine learning algorithms and CNN

4. Results and Discussion:

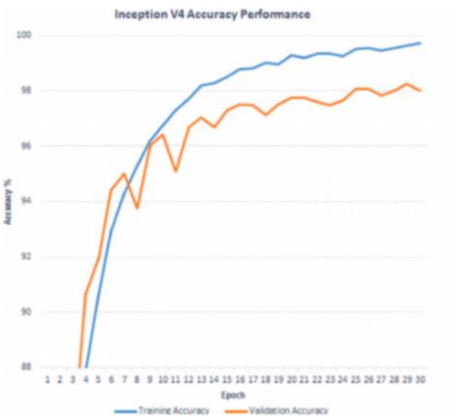
The work was done using a GPU NVIDIA GeForce GTX workstation equipped with a Core i5 9th generation processor, 8 GB RAM, GDDR5 graphics memory, and Windows 10. OpenCV, Numpy CuDNN, Anaconda3, Keras, and Theano libraries were among the software used in the implementation. Accuracy during testing and training was assessed, and losses were computed for every model. Based on the ImageNet dataset, pre-trained models included ResNet-50, Inception V4, VGG-16, and DenseNet-121.

4.1 Dataset: The PlantVillage dataset, consisting of 54,305 images, was used for experimental analysis. Pre-trained models were trained with 80% of the dataset, while 20% was used for validation and testing. The dataset split into training, testing, and validation sets, covering all 38 plant disease classes.

4.2 Data pre-processing: The PlantVillage dataset, which contains 54,000 photos of crop diseases, was utilized in the study to train transfer learning models. To accommodate varying input sizes for different models, images were downscaled to 256 × 256 pixels. Techniques for augmenting data, such as rotation, flipping, zoom intensity, and rescaling, were devised to avoid overfitting and model loss. By doing this, overfitting is avoided and the model's resilience is increased, improving its ability to accurately categorize photos of actual plant diseases.

4.3 Training a model: Pre-trained network models with varying filter widths for feature extraction were employed in the study to classify plant diseases. Accurate feature extraction from feature maps was ensured by using real pre-trained models with real convolution layers and filter sizes in the trials.

PlantVillage, a publicly accessible dataset, was used in the study to train deep CNN networks via transfer learning for illnesses of plants. Standardized parameters for the models included 38 output classes, a 0.5 dropout rate, and a learning rate of 0.01. 80% of the samples in the dataset were used for training, while the remaining samples were divided into test and validation samples. With a maximum recognition accuracy of 84.27%, the models converged after 10 epochs with high accuracy. The validation model's loss and training model's respective losses were 0.64% and 0.52%.



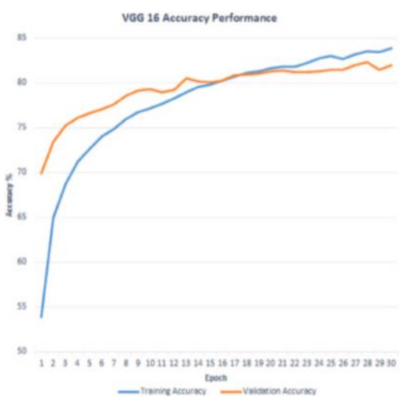


Fig 4: inception v4 performance.   Fig 5: VGG 16 performance

Performance estimates for inception v4 and VGG 16 accuracy are shown in Figs. 4 and 5, where results are graphically shown and training and testing accuracy are estimated. In agricultural production, early detection of crop illnesses is essential to achieving high yields. Models for deep learning and transfer learning are effective at classifying images and removing training complexity. Three pre-trained models were assessed in this study: VGG-16, VGG-19, and Inception V4. Compared to the other models, VGG-16 and 19 fared the best, attaining the highest validation accuracy of 0.97, or nearly an F1 score of 1. According to the report, agricultural industry should adopt the newest technologies for efficient early disease identification.

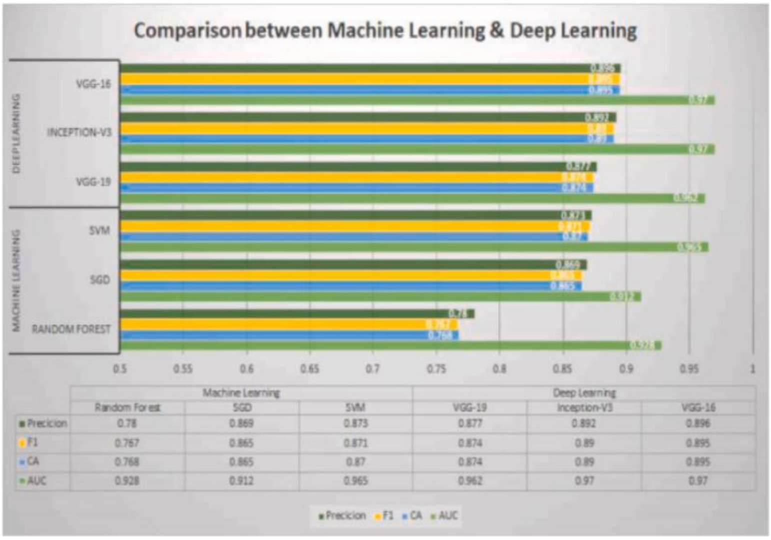


Fig 6: Comparison of proposed ML techniques and CNN

A comparison of deep learning and machine learning techniques is shown in Fig. 6, along with a graphical representation of the findings.

A refined model with the best accuracy of 97% is revealed when the study compares the PlantVillage dataset findings with those of cutting-edge transfer learning methods.

5. Conclusion and future scope:

This study investigates machine and deep learning methods for the identification and categorization of plant diseases. It covers a range of methods and strategies for identifying plant diseases, emphasizing the advancement of deep learning technologies for the diagnosis of leaf diseases. The paper contrasts deep learning and machine learning methods, emphasizing both advancements and areas in need of more research for efficient plant disease identification. The diagnosis and control of crop and plant pests have greatly improved with the application of deep learning and machine learning (DL) and deep learning (ML) techniques. But the majority of research has to include meteorological and plant health data, and it depends on images of pests and plant diseases that have been gathered. Training DL models and network learning can benefit from both unsupervised learning and human visual cognition. To reach maximum potential, cooperation between experts in plant protection and agriculture is essential. This study offers a thorough examination of current advancements and suggests ways to overcome obstacles and restrictions when applying DL and ML approaches for the identification and prevention of plant diseases.

References:

1. T. Bera, A. Das, J. Sil, A.K. Das A survey on rice plant disease identification using image processing and data mining techniques Emerging Technologies in Data Mining and Information Security, Springer (2019), pp. 365-376.
2. U. Shruthi, V. Nagaveni, B. Raghavendra A review on machine learning classification techniques for plant disease detection 2019 5th International Conference on Advanced Computing & Communication Systems, ICACCS) (2019), pp. 281-284
3. J. Shirahatti, R. Patil, P. Akulwar A survey paper on plant disease identification using machine learning approach 2018 3rd International Conference on Communication and Electronics Systems (ICCES) (2018), pp. 1171-1174
4. J. Singh, H. Kaur Plant disease detection based on region-based segmentation and KNN classifier International Conference on ISMAC in Computational Vision and Bio-Engineering (2018), pp. 1667-1675
5. V. Singh, A.K. Misra Detection of plant leaf diseases using image segmentation and soft computing techniques Information processing in Agriculture, 4 (2017), pp. 41-49
6. S. Zhang, Z. You, X. Wu Plant disease leaf image segmentation based on super pixel clustering and EM algorithm Neural Comput. Appl., 31 (2019), pp. 1225-1232
7. A. Badage Crop disease detection using machine learning: Indian agriculture Int. Res. J. Eng. Technol, 5 (2018)
8. S. S. Sannakki and V. S. Rajpurohit," Classification of Pomegranate Diseases Based on Back Propagation Neural Network," International Research Journal of Engineering and Technology (IRJET), Vol2 Issue:02 | May-2015
9. P. R. Rothe and R. V. Kshirsagar," Cotton Leaf Disease Identification using Pattern Recognition Techniques", International Conference on Pervasive Computing (ICPC),2015.
10. Abu Sarwar Zamani, L. Anand et.al, "performance of Machine Learning and Image Processing in plant leaf Disease Detection", Journal of Food Quality, 26 Apr 2022.
11. A. R. Bhagat Patil1, Lokesh Sharma, "A Literature Review on Detection of Plant Diseases", European Journal of Molecular & Clinical Medicine ISSN 2515-8260 Volume 7, Issue 07, 2020.
12. Kowshik B, Savitha V , Nimosh madhav M, "Plant Disease Detection Using Deep Learning", International Research Journal Volume 03 Issue 03 March 2021.
13. Aakanksha Rastogi, Ritika Arora and Shanu Sharma," Leaf Disease Detection and Grading using Computer Vision Technology &Fuzzy Logic" 2nd International Conference on Signal Processing and Integrated Networks (SPIN)2015.
14. Godliver Owomugisha, John A. Quinn, Ernest Mwebaze and James Lwasa," Automated Vision-Based Diagnosis of Banana Bacterial Wilt Disease and Black Sigatoka Disease ", Preceding of the 1'st international conference on the use of mobile ICT in Africa ,2014.
15. S. D. Khirade and A. B. Patil, "Plant Disease Detection Using Image Processing," 2015 International Conference on Computing Communication Control and Automation, 2015, pp. 768-771, doi: 10.1109/ICCUBEA.2015.153.

16. S. C. Madiwalar and M. V. Wyawahare, "Plant disease identification: A comparative study," 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI), 2017, pp. 13-18, doi: 10.1109/ICDMAI.2017.8073478.
17. P. Moghadam, D. Ward, E. Goan, S. Jayawardena, P. Sikka and E. Hernandez, "Plant Disease Detection Using Hyperspectral Imaging," 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2017, pp. 1-8, doi: 10.1109/DICTA.2017.8227476.
18. G. Shrestha, Deepsikha, M. Das and N. Dey, "Plant Disease Detection Using CNN," 2020 IEEE Applied Signal Processing Conference (ASPCON), 2020, pp. 109-113, doi: 10.1109/ASPCON49795.2020.9276722.