International Journal of Research in Computer & Information Technology (IJRCIT),
Vol.7, Issue 2, March 2022
E-ISSN: 2455-3743
Available online at www.ijrcit.co.in

# Comparative Study on Plant Leaf Disease Detection and Classification, Based on Machine Learning Techniques

[1]Prasad W. Bhombe, [2]Dr. Shirish V. Pattalwar

**Abstract**— Plants are a major source of food for the world population. Plant diseases contribute to production loss, which can be tackled with continuous monitoring. Manual plant disease monitoring is both laborious and error-prone. Early detection of plant diseases using computer vision and artificial intelligence (AI) can help reduce the adverse effects of diseases and overcome the shortcomings of continuous human monitoring. To identify the recent advancements in the development of plant disease detection and classification system based on Machine Learning (ML) and Deep Learning (DL) models [5]. An organized way of analysis of various plant disease classification models has been shown in well-formed tables. In this paper, we have conducted a systematic literature study on the applications of the state-of-the-art ML and DL algorithms such as Support Vector Machine (SVM), Convolutional Neural Network (CNN), K-Nearest Neighbor (KNN), Naïve Bayes (NB), other few popular ML algorithms and AlexNet, GoogLeNet, VGGNet, and other few popular DL algorithms respectively for plant disease categorization. Each stated algorithm is characterized through the corresponding processing methods such as image segmentation, and feature extraction, along with the standardized experimental-setup metrics such as total number of training/testing datasets employed, number of diseases under consideration, type of classifier utilized, and the percentage of classification accuracy.

**Keywords**— Plant Disease, Machine Learning, Deep Learning, Support Vector Machine, K-Nearest Neighbor, Convolutional Neural Network.

## I. INTRODUCTION

One of the important sectors of the Indian Economy is Agriculture. Employment for almost 50% of the country's workforce is provided by the Indian agriculture sector. India is known to be the world's largest producer of pulses, rice, wheat, spices, and spice products [7]. Farmer's economic growth depends on the quality of the products they produce, which relies on the plant's growth and the yield they get. Therefore, in the field of agriculture, the detection of disease in the early stage of plants plays an instrumental role [10]. Plants are highly prone to diseases that affect the growth of the plant which in turn affects the ecology of the farmer.

To detect a plant disease at the very initial stage, the use of an automatic disease detection technique is advantageous. The symptoms of plant diseases are conspicuous in different parts of a plant such as leaves, stems, branches, etc. Manual detection of plant disease using leaf images is a tedious job. Hence, it is required to develop computational methods that will make disease detection and classification using leaf images automatic &time-consuming.

Despite the challenges in plant disease detection, it is still an active area of research. Numerous approaches have been proposed over the years. In traditional systems approach for detection and differentiation of plant, diseases can be achieved using deep learning with CNN & ALDD (orange Leaf disease dataset) for real-time detection of apple leaf diseases using deep learning approach based on improved convolution neural networks, which has got an accuracy of 78.80% & 23.13FPS. Another approach is Support Vector Machine (SVM) algorithms [2] like LSVM, QSVM, CKNN, ESD (Ensemble Subspace Discriminative) are used. This technique was implemented for grape leaf diseases and depending on the type and stage of disease, the classification accuracy was between 90% and 92%. Another approach based on leaf images and using CNNs as a technique for automatic detection and classification of plant diseases was used with K-means is a clustering procedure. CNN consisted of 10 hidden layers. The number of outputs was 6 which was the number of classes representing five diseases along with the case of a healthy leaf.

Having diseases is quite natural in crops due to changing climatic and environmental conditions. Diseases affect the growth and production of crops and are often difficult to control. To ensure good quality and high production, it is necessary to have accurate disease diagnosis and control actions to prevent them in time. In India, we have a wide variety of crops and they may be affected by different types of diseases on leaf, stem, and fruit. Leaf diseases are the early symptoms caused due to fungi, bacteria, and viruses. So, there is a need to have an automatic system that can detect the type of diseases and take appropriate actions.

## II. COMPARISON OF TECHNIQUES FOR PLANT DISEASE CATEGORIZATION

This section summarizes 15 papers that are compared for plant disease categorization with respect to methodology,

**Manuscript Received Feb. 12, 2022; Revised 28 Feb. 2022 and Published on March 28, 2022**

***Prasad W. Bhombe,*** *Department of Electronics & Telecommunication, Prof. Ram Meghe Institute of Technology &Research, Badnera, Amravati, India. Mail Id: prasadbhombe123@gmail.com*
***Dr. Shirish V. Pattalwar,*** *Department of Electronics & Telecommunication, Prof. Ram Meghe Institute of Technology &Research, Badnera, Amravati, India. Mail Id: shirishpattaalwar@rediffmail.com*

parameter considerations, advantages, and drawbacks for various plants leaf diseases.

A summary of the comparison is given in the following table.

Table 1 shows the related previous research work on plant leaf disease detection. It is concluded based on the literature survey carried out on the machine learning and deep learning techniques, that machine learning technique suffers from slow convergence problem. It is essential to improve the convergence rate and generalization ability of machine learning.

Table 1: Comparative study of various techniques for detection of plant leaf diseases

| Ref. No. | Authors | Methodology | Parameter considered | Advantages | Drawbacks |
|---|---|---|---|---|---|
| [1] | Peng Jiang et. al. .2019 IEEE journal | Deep learning with CNN & ALDD (Apple Leaf disease dataset) | GoogLeNet Inception & Rainbow concatenation are used to detect apple leaf diseases, received accuracy of 78.80% & 23.13FPS | The real-time monitoring of apple leaf diseases, which is of high practical value for agricultural applications. | Low recognition accuracy for Alternaria leaf spot & Grey spot |
| [2] | AlishbaAdeela et. al. 2019 | Multiclassifiers like LSVM, QSVM, CKNN, ESD (Ensemble Subspace Discriminative) are used | Detection of grape leaf diseases namely Black rot, black Measles, Leaf Blight. Achieved segmentation accuracy of 90% & Classification accuracy of 92% | Using multilevel k means and SVM algorithm efficiency of performance will be calculated. | 1. I/P images are low contrast & noisy Whichaffect segmentation. 2. Degrade accuracy for complex images |
| [3] | Ali Khadivi et. al. 2018 | A statistical method like PCA & Cluster analysis is used | Morphological Char's are identified for pomegranate by classifying Germplasm | The performance parameters, namely accuracy and sensitivity of the extreme learning machine, is 95% | PCA shows high variation among genotypes of the total variance. |
| [4] | Zhang Chuanlei et. al. 2017 | Image processing techniques & pattern recognition | GA-CFS is implemented with an SVM classifier which gives a 90% identification rate | Extracting the features can be done statistically using different image processing algorithms depending on what features best describe the disease symptoms. | Apple leaf images are captured in a controlled environment |
| [5] | Siddharth Singh Chouhan et. al. 2018 | BRBFNN (Bacterial Foraging optimization based radial basis function Neural n/W) is used for identification & Classification | Received high accuracy in identified fungal diseases like common rust, cedar apple rust, late blight, leaf curl, leaf spot, early blight, etc. Vpc- 0.8621/0.8307 Vpe- 0.1118/ 0.1527 | The algorithm was contrasted with other machine learning models for accuracy | Worked only on fungal diseases, not on bacteria/viruses |
| [6] | Jayme G.A. Barbedo 2018 | Plant disease recognition is applied to plant pathology using Deep Learning with CNN & GoogLeNet | Corn leaf samples are used. Achieved Accuracy:76% for original images 79% for background removed images | Regression analysis we can find new trends and data by location of user and using crowdsourcing results will be influenced | Overfitting problem appears for random noise or errors |
| [7] | Jayme Garcia 2019 | Nutrient detection, quantification & classification using digital images & ML | Nutritious deficiencies in plants, plant color & morphology can be detected. | U-net segmentation models to investigate the best segmentation | The annotation process needs to be improved to avoid unreliable labels. |
| [8] | Sukhvir Kaur et. al. 2019 | Summarizes the pros & cons for plant disease detection with the CV) approach | The efficiency of the system depends greatly on the quality of training data. | The developed real-time hardware with ELM classifier is highly capable | Problems of overfitting or overtraining occurred with NN, SVM & GA |
| [9] | Juncheng Maa et. al. 2018 | Deep CNN with conventional classifiers like Random Forest, SVM, AlexNet is used | Cucumber diseases are identified using leaf symptom images with DCNN, which shows an accuracy of 93.41% | Extracting the features is the most tasking process for the SVM model implementation | Degradation inaccuracy for unbalanced data. |
| [10] | ShanwenZhanga et. al 2019 | GPDCNN (Global Pooling Dilated) along with AlexNet is used | Found higher recognition rate & learning rate for cucumber leaf diseases. | CNNs refers to their ability to learn rich mid-level image representations as opposed to hand-designed low-level features used in other image classification | Shows lesion detection failure. |

International Journal of Research in Computer & Information Technology (IJRCIT),
Vol.7, Issue 2, March 2022
E-ISSN: 2455-3743
Available online at www.ijrcit.co.in

| | | | | | |
|---|---|---|---|---|---|
| | | | | methods | |
| [11] | Jorge Parraga-Alava et. al. 2019 | ML with annotation techniques like object segmentation & classification is used | A robusta coffee leaf Rust infection level (1-4) identification is carried out using OIRSA (Organismo International Regional Sanded Agropecuaria) method. | The farmer will be able to accurately detect the type of disease a particular plant is having using the image of the plant. | No. of samples & accuracy is less |
| [12] | Lin yua et. al. 2019 | Unsupervised classification & adaptive 2D thresholding is developed for hyper spectralngg imaging | Tea leaf disease (anthracnose) is detected.<br>Accuracy:<br>Pixel level- 94%<br>Leaf level-98% | By using foliar hyperspectral images of anthracnose-infected tea leaves, we demonstrated the feasibility of using a comprehensive disease-scab-detecting strategy to identify the disease. | This model is unstable for different leaves |
| [13] | Hu Genshenga et. al. 2019 | SVM is used along with C-DCGAN (Conditional deep convolution generative adversarial n/w) | VGG16 DL model is used which gives identification accuracy of 90% | To solve the over-fit problem using deep learning | Identification accuracy of images is affected with VGG16. |
| [14] | Alessandro dos Santos Ferreiraa et. al. 2019 | Unsupervised deep clustering algorithm & image clustering | Weed (grass) & broadleaf labeling with 87% of accuracy. | To reduce the number of manual annotations time. | Labeling is possible only where the cluster has noticeably more elements of a specific class |
| [15] | VinitBodhwania et. al. 2019 | Convolutional Neural Network | DL n/w with 5 stages is designed which gives a recognition rate of 93.09% | To solve the Vanishing Gradient problem in plant leaf disease | The stochastic approach is used not focused on leaf diseases. |

Following are other challenges in detecting plant leaf diseases based on Machine learning and deep learning.

1. Quality of plant leaves images.

2. Needs a large amount of data set for Training & Testing.

3. Acquired images may affect by background data and noises.

4. Identification of the exact diseased spot in a leaf is difficult.

5. Depending on environmental conditions; the color size & texture of the leaf may vary.

6. Regular observation is needed for a particular category of plants leaf.

Hence a comprehensive methodology is proposed to overcome these challenges as discussed in the next section.

## III. PROPOSED METHODOLOGY

A methodology proposed here tries to minimize the challenges listed above. It is elaborated with the following diagram.

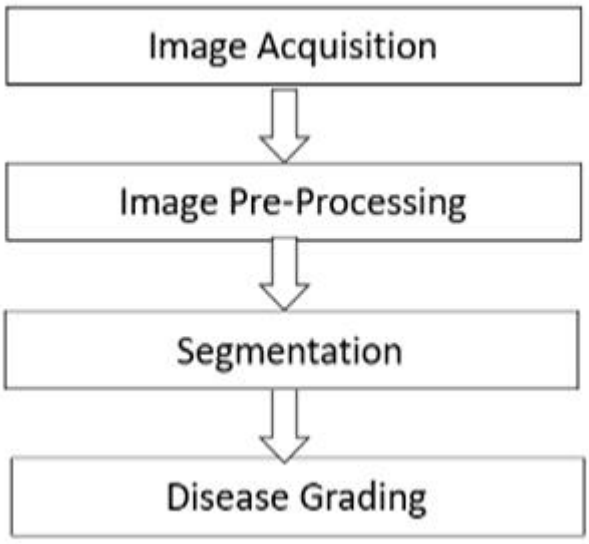


Figure 1: Architecture of Proposed Model for plant leaf disease detection

### A. *Datasets*

In all the experimental setups of the researchers, ML algorithms have experimented with a lesser number of inputs in their dataset in turn in comparison with the DL algorithms to detect and classify plant diseases. Most of the authors used the plant village dataset, in which the images were taken in a controlled environment. However, the trained model in the controlled environment fails when the model is tested in an uncontrolled environment. For this reason, to improve the accuracy of these models, experimentation using real-time images is needed. So, the researchers must focus on the real-time image dataset. Recently, the hyperspectral dataset has been used by some of the researchers in this domain, which shows better performance.

### B. *Image Acquisition*

In this step, the infected disease image is captured using computer vision. Computer vision is the ability of a computer to see, which consists of a camera, Analog-Digital converter, and Digital Signal Processing system.

### C. *Image Preprocessing*

In Image preprocessing, the quality of the image is enhanced and the noise is removed from the captured image by using various image processing techniques.

### D. *Image Segmentation*

In the leaf disease diagnosis system, the spots that appear on images are important. Hence the whole image is segmented into parts and only infected parts of the image are extracted using a segmentation algorithm.

### E. *Feature Extraction*

From the infected image, its features are extracted using different feature extraction techniques. Color, texture, and shapes are the features that are used in many leaf disease diagnosis systems.

### F. *Classifier*

There are many supervised machine learning algorithms such as Naive Bayes, SVM, and BP. We used one of the classifiers in the leaf disease diagnosis system.

## CONCLUSION

A comparative analysis of various state-of-the-art ML and DL algorithms to identify and categorize plant leaf diseases shows that the process of plant leaf disease diagnosis becomes slower with a big database in Machine learning algorithms. On the other hand, the researcher has to compromise on efficiency if the database is compressed. The following potential areas for further investigations are listed as follows: 1) It would be of great importance to diagnose a specific stage of a plant leaf disease. 2) The analysis would be useful to control the number of chemicals to be applied in an accurate quantification. 3) An online system should be needed for identifying and classifying all plant diseases. 4) To develop a system to detect an infected leaf due to loss of nutrient deficiency. 5) More research investigations must also be needed to carry out the analysis on the backside of the leaf. 6) Inclusion of real-time images should be used for accurate detection of diseases of the plant leaf. 7) Image collection under mixed lighting conditions can be addressed. 8) Automatic severity estimation of detected diseases can be measured. 9) To develop a system to extract the desired affected area from the image's complex background. 10) More research attempts are needed to detect the diseases of the other parts of the plants like stems. A proposed system is an effort in the direction of accurate plant leaf disease diagnosis with considerable speed.

## REFERENCES

[1] Peng Jiang, Yuehan Chen, Bin Liu, Dongjian he, Chunquan Liang, “Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks", IEEE journal Vol. 7, Access. 2914929, 2019.

[2] Alishba Adeela, Muhammad Attique Khan, Muhammad Sharifa, Faisal Azama ,Jamal Hussain Shaha, Tariq Umera, Shaohua Wanc, “Diagnosis and recognition of grape leaf diseases: An automated system based on a novel saliency approach and canonical correlation analysis based multiple features fusion”, Sustainable Computing: Informatics and Systems 2210-5379/© 2019 Elsevier Inc.

[3] Ali Khadivi, DaneyalAyenehkar, MiladKazemi, Alireza,“Phenotypic and pomological characterization of a pomegranate (Punicagranatum L.) germplasm collection and identification of the promising selections”, Elsevier ScientiaHorticulturae 238 (2018) PP 234–245.

[4] Zhang Chuanlei, Zhang Shanwen, Yang Jucheng, Shi Yancui, Chen Jia, “Apple leaf disease identification using genetic algorithm and correlation-based feature selection method”, International Journal of Agriculture &BiolEng Vol. 10 No.2, March 2017.

[5] Siddharth Singh Chouhan, Ajay Kaul, Uday Pratap Singh, Sanjeev Jain,“Bacterial Foraging Optimization Based Radial Basis Function Neural Network (BRBFNN) for Identification and Classification of Plant Leaf Diseases: An Automatic Approach Towards Plant Pathology”, 2169-3536, 2018, IEEE Translations.

[6] Jayme G.A. Barbedo, “Factors influencing the use of deep learning for plant disease recognition” Science Direct BiosystemEngg,, 2018, Published by Elsevier.

[7] Jayme Garcia ArnalBarbedo,“Detection of nutrition deficiencies in plants using proximal images and machine learning: A review”, Computers and Electronics in Agriculture 162 (2019) 482–492, Published by Elsevier.

[8] SukhvirKaur,Shreelekha Pandey, Shivani Goel, “Plants Disease Identification and Classification Through Leaf Images:A Survey”, Published by Springer Archives of Computational Methods in Engineering (2019)26: PP507–530.

[9] JunchengMaa, KemingDua, FeixiangZheng, LingxianZhang, ZhihongGong, ZhongfuSuna,“A recognition method for cucumber diseases using leaf symptom images based on the deep convolutional neural network”, Published by Elsevier Computers and Electronics in Agriculture 154(2018), PP.18–24

[10] ShanwenZhanga, SubingZhang, ChuanleiZhang, XianfengWanga, Yun Shia, “Cucumber leaf disease identification with global pooling dilated convolutional neural network”, Published by Elsevier Computers and Electronics in Agriculture 162 (2019) PP. 422–430

[11] Jorge Parraga-Alava, Kevin Cosme, Ang_elicaLoor, Esneider Santander, “RoCoLe: A robusta coffee leaf images dataset for

evaluation of machine learning-based methods in plant diseases recognition", Published by Elsevier Data in brief 25 (2019)

[12] Lin yuana, PengYang, WenyanHanb, YanboHuang, Bin Wangd, JingchengZhang, HaiboZhanga, ZhiyanBaoa, "Detection of anthracnose in tea plants based on hyperspectral imaging", Published by Elsevier, Computers, and Electronics in Agriculture 167 (2019)

[13] Hu Genshenga, Wu Haoyua, Zhang Yana,Wan Mingzhub,"A low shot learning method for tea leaf's disease identification", Published by Elsevier, Computers, and Electronics in Agriculture 163 (2019)

[14] Alessandro dos Santos Ferreira, Daniel Matte Freitas, Garcinia Gonçalves da Silva, Hemers on Pistorib, Marcelo TheophiloFolhesc, "Unsupervised deep learning and semi-automatic data labeling in weed Discrimination", Published by Elsevier, Computers, and Electronics in Agriculture 165 (2019)

[15] VinitBodhwania, D. P. Acharjyaa, UmeshBodhwania,"Deep Residual Networks for Plant Identification", Published by Elsevier, Procedia Computer Science 152 (2019) 186-194