AI Based Plant Disease Monitoring Using Maching Learning

Prof.Asha Gaikar[1], Prof.Pratiksha Shinde[2], Prof.Supriya Pawar[3], Prof.Namrata Wagh[4]

[1]*Department of Computer Engineering*
[2]*Department of Artificial intelligence and Machine Learning*
[3,4]*Department of Computer Engineering*
Bharat College of Engineering,Badlapur.

Abstract -Plants play a crucial role in human life as they are a major source of energy and help reduce the effects of global warming. However, plants are increasingly affected by diseases such as bacterial spot, late blight, and Septoria leaf spot, which significantly reduce crop productivity. Therefore, early detection of plant diseases is essential in agriculture to ensure effective disease management. Identifying diseases at an early stage helps prevent their spread and minimizes crop loss.

The main objective of this project is to detect plant leaf diseases using Machine Learning techniques applied to images and videos. For image-based detection, a Random Forest classifier, a machine learning algorithm, is used for classification. For video-based detection, the proposed method employs ResNet50, a deep learning algorithm. These techniques provide prediction results evaluated using performance metrics such as accuracy, precision, and efficiency. This system can be applied in agriculture fields, nurseries, college gardens, and similar environments

Key Words: Random forest classifier,Resnet-50.

1. INTRODUCTION

A plant disease is a disturbance in the normal condition of a plant that damages or disrupts its essential functions. All types of plants, whether wild or cultivated, can be affected by diseases. Each plant family is vulnerable to certain diseases, although many of these occur infrequently. The occurrence and spread of plant diseases vary from place to place, depending on factors such as the presence of pathogens, environmental conditions, and the types of crops and cultivars grown. Some plant varieties are highly susceptible to disease outbreaks, while others show greater resistance. Plants that are repeatedly exposed to stress or disturbance are more likely to develop diseases. These diseases arise when pathogens interfere with normal physiological functions, leading to disruptions in plant structure, growth, metabolism, or overall performance. Such interference alters essential biochemical and physiological processes, resulting in recognizable disease symptoms. Plant diseases are generally classified into two broad categories based on their cause: infectious and non-infectious. Infectious diseases are caused by living agents such as fungi, bacteria, mycoplasmas, viruses, viroids, nematodes, and parasitic flowering plants, which can grow within or on the host and spread to other susceptible plants.

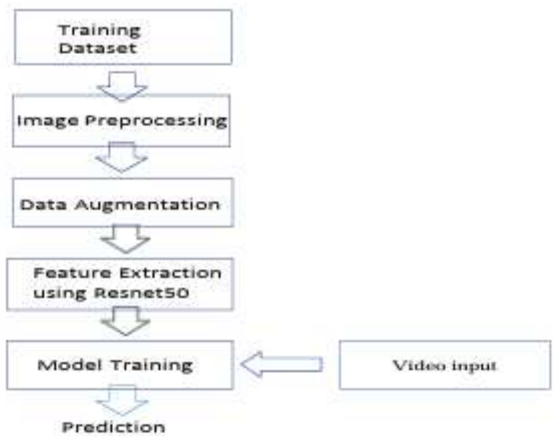

Fig.1.1 Basic Architecture

2. LITERATURE SURVEY

1. Unsupervised Image Translation Using GANs for Plant Disease Detection (Nazki et al.)

- **Dataset:** 2,789 tomato disease images
- **Method:** GAN with Deep CNN
- **Accuracy:** 86.1%
- **Advantages:** Produces clearer images and improves feature representation

- **Disadvantages:** Training is unstable, requires extensive tuning, and may suffer from mode collapse

1. **Cucumber Leaf Disease Identification Using GPDCNN (Zhang et al.)**

- **Dataset:** 600 infected cucumber leaf images (6 disease classes)
- **Method:** Global Pooling Dilated Convolutional Neural Network (GPDCNN)
- **Accuracy:** 94.65%
- **Advantages:** More robust compared to other methods
- **Disadvantages:** Large number of parameters in fully connected layers slows training and increases the risk of overfitting.

2. **Multilayered CNN for Mango Leaf Disease Classification (Singh Chouhan et al.)**

- **Dataset:** Mango leaf images captured at SMVDU, Katra
- **Method:** Multilayer Convolutional Neural Network (MCNN)
- **Accuracy:** 97.13%
- **Advantages:** Automatically extracts important features without human intervention
- **Disadvantages:** Training is time-consuming and requires high computational resources

3. **Sunflower Leaf Disease Detection Using PSO (Vijai Singh)**

- **Dataset:** Captured sunflower leaf images
- **Method:** Image segmentation using Particle Swarm Optimization (PSO)
- **Accuracy:** 98%
- **Advantages:** Easy to implement, requires minimal parameter tuning, and offers better computational efficiency than Genetic Algorithms
- **Disadvantages:** Limited applicability to complex problems due to its simple operational nature.

4. **Deep CNN-Based Real-Time Corn Plant Disease Detection (Mishra et al.)**

- **Dataset:** PlantVillage dataset
- **Method:** Deep Convolutional Neural Network (DCNN)
- **Accuracy:** 88.46%
- **Advantages:** Minimal preprocessing required, reduced human intervention, and self-learning capability
- **Disadvantages:** Requires a very large dataset for effective training

5. **Performance Analysis of CNN Models for Plant Disease Detection (Sharma et al.)**

- **Dataset:** Images of healthy and diseased tomato leaves
- **Method:** Convolutional Neural Network (CNN) with image segmentation
- **Accuracy:** 98.6%
- **Advantages:** Automatic feature extraction directly from raw images
- **Disadvantages:** CNNs lack structured representations similar to human visual perception

6. **Tomato Leaf Disease Detection Using CNN (Agarwal et al.)**

- **Dataset:** PlantVillage dataset
- **Method:** Convolutional Neural Network (CNN)
- **Accuracy:** 76%–100% (Average disease accuracy: 91.2%)
- **Advantages:** Very low storage requirement (~1.5 MB) compared to pre-trained models (~100 MB), demonstrating efficiency
- **Disadvantages:** CNNs can be slower due to operations such as pooling.

7. **Seasonal Crops Disease Prediction Using Deep Convolutional Encoder Network (Khamparia et al.)**

- **Dataset:** PlantVillage dataset
- **Method:** Deep Convolutional Encoder Network
- **Accuracy:** 97.50%
- **Advantages:** Uses a Softmax classifier at the output layer to provide class probabilities, enabling effective multi-class prediction
- **Disadvantages:** Lacks an effective mechanism to map deep feature representations back to the input dimensions.

8. **Deep Neural Network–Based Plant Disease Recognition (Sladojevic et al.)**

- **Dataset:** Leaf images captured by agricultural experts
- **Method:** Deep Convolutional Neural Network (DCNN)
- **Accuracy:** 96.3%
- **Advantages:** High accuracy due to multiple hidden layers; effective for image classification and segmentation tasks
- **Disadvantages:** Does not explicitly encode object position or orientation and lacks full spatial invariance

9. **Review of Machine Learning Approaches for Plant Leaf Disease Detection (Majji V. Appalanaidu & G. Kumaravelan)**

- **Dataset:** PlantVillage dataset
- **Techniques Used:** CCM, GLCM, MER, DWT, SIFT
- **Objective:** To compare state-of-the-art machine learning and deep learning methods for detecting and classifying plant leaf diseases
- **Outcome:** Provides a comprehensive comparative analysis of feature extraction and classification approaches used in plant disease diagnosis

10. **Research on a Machine Learning Framework Based on the Random Forest Algorithm
(Qiong Ren, Hui Cheng, and Hai Han)**

- **Technique Used:** Random Forest algorithm
- **Objective:** To study and analyze a machine learning framework built on the Random Forest algorithm with the aim of improving and overcoming its existing limitations

11. **Random Forest with Adaptive Local Templates for Pedestrian Detection (Xiang et al.)**

- **Dataset:** TUD Pedestrians, INRIA Pedestrians
- **Method:** Random Forest with Adaptive Local Templates
- **Accuracy:** 90.8%
- **Objective:** To detect pedestrians in complex and cluttered environments
- **Key Idea:** Multiple weak classifiers are generated using adaptive local templates and combined through a Random Forest framework. The forest is built iteratively in a layer-by-layer manner, where splitting functions are learned from adaptive local templates, and classifiers of equal depth collectively form a weak classifier.

12. **Improving the Random Forest Algorithm by Varying Bootstrap Sample Size (Md. Nasim Adnan)**

- **Technique Used:** Random Forest
- **Objective:** To enhance ensemble accuracy by increasing diversity among decision trees
- **Key Idea:** The method applies the random subspace technique with randomly varying bootstrap sample sizes for high-dimensional datasets, generating more diverse decision trees and improving overall classification performance

13. leaf and skin disease detection using image processing

- **Reference**: Manjunath Badiger ,Varuna kumara ,Sachin CN shetty,Sudhir poojary
- **Technique used**: K-means algorithm and SVM classifier
- **Output**: Accuracy = 96.3%
- **Advantages**: Easy to understand and implement. Can handle large datasets well.

- **Disadvantages:** It requires to specify the number of clusters (k) in advance.It cannot handle noisy data and outliers.It is not suitable to identify clusters with non-convex shapes.

3. TECHNIQUES USED FOR DISEASE DETECTION

Machine Learning Methods:

K-Nearest Neighbour (KNN) Algorithm:K-Nearest Neighbour is one of the most straightforward machine learning algorithms and operates under a supervised learning framework.The algorithm assumes that new data samples share similarities with existing labeled instances and assigns the new sample to the class most common among its nearest neighbours.KNN retains all available training data and classifies new inputs based on similarity measures, enabling efficient and accurate categorization of unseen data points.

Support Vector Machine: is a widely used supervised learning technique designed for both classification and regression tasks, although it is predominantly applied to classification problems. The primary goal of the SVM algorithm is to identify the optimal decision boundary that effectively separates data points within an *n*-dimensional feature space. This optimal separating boundary, known as a **hyperplane**, enables accurate classification of new and unseen data samples.

4. RESULTS

Confusion matrix Result:

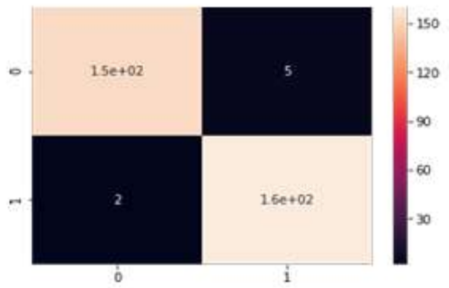

Fig. 4.1 Plot of Confusion matrix.

In the above confusion matrix, 5 images are predicted as false negative and 2 images are predicted as false positive

Comparison Table for Different Machine Learning Algorithms:

Parameters	Random forest	Logistic Regression	KNN	Naive Bayes	SVM
Precision	0.98	0.88	0.96	0.88	0.94
Recall	0.98	0.86	0.96	0.86	0.94
f1-score	0.98	0.86	0.96	0.86	0.94
Accuracy	0.9812	0.9265	0.9562	0.8578	0.94

Table 4.1 Comparison Table for Different Machine Learning Algorithms

Results of Plant Leaf Images:

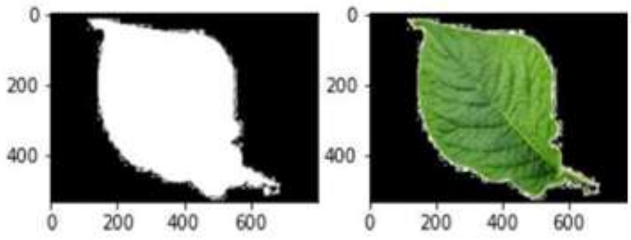

Fig. 4.2 Healthy Leaf

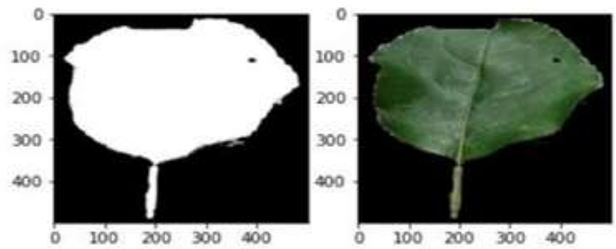

Fig.4.3 Diseased (Low)

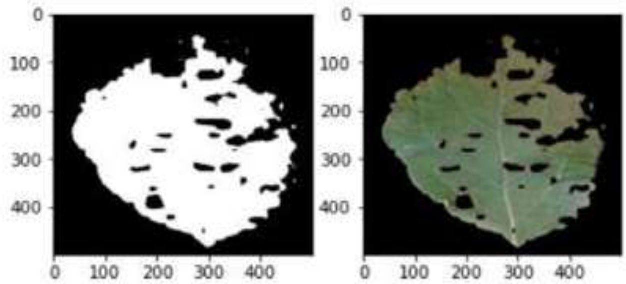

Fig.4.4 Diseased (High)

5. CONCLUSION AND FUTURE WORK

In recent years, machine learning–based approaches for plant leaf disease detection have demonstrated significant potential. These techniques enable accurate classification of leaves into healthy and diseased categories, supporting early disease identification and timely intervention to minimize crop damage and yield loss. This project investigates multiple machine learning algorithms, including Logistic Regression, K-Nearest Neighbour, Naïve Bayes, Support Vector Machines, and Random Forests, and evaluates their performance across different datasets using metrics such as accuracy, precision, recall, and F1-score.

With advances in technology, automated monitoring and management systems are increasingly adopted in agriculture. Crop yield losses are often caused by delayed disease detection, which typically occurs only at advanced stages. The proposed system addresses this issue by enabling early-stage disease detection directly from leaf images, thereby reducing yield loss and lowering reliance on agricultural experts.

REFERENCES

1. Haseeb Nazki, Sook Yoon, Alvaro Fuentes, Dong Sun Park "Unsupervised image translation using adversarial networks for improved plant disease recognition" Published by Elsevier B.V,(2020).

2. Shanwen Zhang, Subing Zhang, Chuanlei Zhang, Xianfeng Wang, Yun Shi "Cucumber leaf disease identification with global pooling dilated convolutional neural network" Published by Elsevier B.V, (2019).

3. Uday Pratap Singh, Siddharth Singh Chouhan, Sukirty Jain, And Sanjeev Jain "Multilayer Convolution Neural Network for the Classification of Mango Leaves Infected by Anthracnose Disease" (2019).

4. Vijai Singh "Sunflower leaf diseases detection using image segmentation based on particle swarm optimization" 2019 Published by Elsevier,(2019).

5. Sumita Mishra, Rishabh Sachan, Diksha Rajpal "Deep Convolutional Neural Network based Detection System for Real-time Corn Plant Disease Recognition" 2020 Published by Elsevier B.V, (2019).

6. Parul Sharma, Yash Paul Singh Berwal , Wiqas Ghai "Performance analysis of deep learning CNN models for disease detection in plants using image segmentation" open access 2019 Published by Elsevier B.V, (2019).

7. Mohit Agarwal, Abhishek Singh, Siddhartha Arjaria, Amit Sinha, Suneet Gupta "Tamato Leaf Disease Detection using Convolution Neural Network" 2019-2020 Published by Elsevier,(2019).

8. Aditya Khamparia, Gurinder Saini, Deepak Gupta, Ashish Khanna, Shrasti Tiwari, Victor Hugo C. de Albuquerque "Seasonal Crops Disease Prediction and Classification Using Deep Convolutional Encoder Network" ,(2019).

9. Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk and Darko Stefanovic "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification", Volume 2016 Hindawi Publishing Corporation,(2016). 10. Majji V Applalanaidu, G. Kumaravelan "A Review of Machine Learning Approaches in Plant Leaf Disease Detection and Classification" IEEE,(2021).

11. Qiong Ren , Hui Cheng and Hai Han "Research on machine learning framework based on random forest algorithm" : AIP Conference Proceedings,(2017).

12. Tao Xiang, Tao Li, Mao Ye, and Zijian Liu "Random Forest with Adaptive Local Template for Pedestrian Detection" Hindawi Publishing Corporation,(2015).

13. Md Nasim Adnan "Improving the Random Forest Algorithm by Randomly Varying the Size of the Bootstrap Samples"Adnan,(2014).

14. Ziming Wu, Weiwei Lin, Zilong Zhang and Angzhan Wen "An Ensemble Random Forest Algorithm for Insurance Big Data Analysis"IEEE,(2017).

15. Manjunath Badiger, Varuna kumara,Sachin CN shetty,Sudhir poojary "Leaf and skin disease detection using image processing" Global Transactions Proceedins,(2022).

16. Niveditha M, Pooja R, Prasad Bhat N, shashank N, "Plant disease detection using machine learning" IEEE (2021).

17. Nishant Shelar ,Suraj shinde ,Shubham sawant ,Shreyas dhumal "Plant disease detection using CNN " Turkish Journal of Computer and Mathematics Education, (2021).

18. Madhuri Devi Chodey, Dr.Noorilla Shariff C, Gauravi Shetty "Pest detection in crop using video and Image processing" IJRASET (2020).

19. Aryan Garg"Image Classification Using Resnet-50 Deep Learning Model"Analytics vidya,(2022).

20. Devvi Sarwinda , Radifa Hilya Paradisa , Alhadi Bustamam ,Pinkie Anggia "Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer" International Conference on Computer Science and Computational Intelligence,(2020).