# PLANT LEAF DISEASE CLASSIFICATION AND PEST DETECTION USING DEEP LEARNING

B.TECH PROJECT REPORT

submitted by

**ANGELYN BEENA SHAW (LBT17CS018)**

**ANJALI M (LBT17CS019)**

**JANEEFA J JUSTIN (LBT17CS056)**

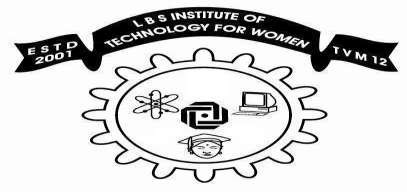
**JINCY JOSE A J (LBT17CS058)**

*to*

*the APJ Abdul Kalam Technological University*

*in partial fulfillment of the requirements for the award of the Degree of*

*Bachelor of Technology in Computer Science and Engineering*



## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

**LBS INSTITUTE OF TECHNOLOGY FOR WOMEN POOJAPPURA, THIRUVANANTHAPURAM, KERALA**

JUNE 2021

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

**LBS INSTITUTE OF TECHNOLOGY FOR WOMEN THIRUVANANTHAPURAM-695012**



## CERTIFICATE

This is to certify that the report entitled **‘PLANT LEAF DISEASE CLASSIFICATION AND PEST DETECTION USING DEEP LEARNING’** submitted by **ANGELYN BEENA SHAW, ANJALI M, JANEEFA J JUSTIN and JINCY**

**JOSE A J** to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering is a bonafide record of the project work carried out by them under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

#### Prof. JAYALEKSHMI S

Associate professor

Department of Computer Science and Engineering

(Project Guide)

#### Dr. SMITHA VAS P

Assistant professor

Department of Computer Science and Engineering

(Project Coordinator)

#### Prof. JAYALEKSHMI S

Head of the Department

Department of Computer Science and Engineering

# ACKNOWLEDGEMENT

First and foremost, we wish to express our deep and sincere gratitude to our principal **Dr. JAYAMOHAN J** for providing us with all facilities and infrastructure for the completion of this work. We record our sincere gratitude and thanks to **Prof. JAYALEKSHMI S**, Head of the Department, Department of Computer Science and Engineering for providing us with all facilities for the completion of this work. We would like to express our sincere gratitude to our project coordinator **Dr. SMITHA VAS P**, Department of Computer Science and Engineering for her valuable assistance provided during the project. We would like to express our sincere gratitude to our project guide **Prof. JAYALEKSHMI S,** Associate Professor, Department of Computer Science and Engineering, for providing us with an opportunity and guidance to do our final year project. We also express our sincere gratitude to all the staff members in the Department for the valuable support and encouragement they rendered. Last but not least we thank all my friends and well-wishers for their kind cooperation, immense support, and useful suggestions we are sincerely grateful to them for sharing their truthful and illuminating views on many issues related to the project.

ANGELYN BEENA SHAW

ANJALI M JANEEFA J JUSTIN JINCY JOSE A J

# ABSTRACT

In a developing country like India, the primary source of food and income is agriculture. Healthy and resourceful individuals are the result of healthy food. So agriculture plays the main part in the life of human beings. As the population increases, the need for food supplies increases. Without adequate food, there will be famine which leads to poverty in the country. So a satisfactory growth in the food production should be ensured. This can be made fruitful by successful disease control. Diseases are a major threat to the agriculture industry since they reduce the quantity and quality of crops. Infection of diseases in plants gives rise to huge losses in the agriculture industry. Farmers are struggling a lot to control plant diseases since they are unaware of the exact name or reason behind the disease. Many types of diseases can adversely affect plant health. So, quick and reliable detection of plant diseases is very important. To help the farmers who are struggling in disease detection and to improve the productivity in agriculture, research papers in different fields of disease prediction using machine learning are studied. The advantages and disadvantages of each method are analyzed and the matrices used for evaluating the prediction accuracy were also discussed in brief. In this work, we are presenting an application that helps farmers to identify the diseases and pests that affected their plants.

# CONTENTS

1. [INTRODUCTION 1](#_TOC_250018)
2. [LITERATURE SURVEY 2](#_TOC_250017)
   1. Machine Learning-based for Automatic Detection of Corn Plant Diseases Using Image Processing 2
   2. A Review on Machine Learning Techniques for Rice Plant Disease Detection in Agricultural Research 2
   3. Application of Machine Learning In Detection of Blast Disease in South India

Rice Crop 3

* 1. Diagnosis of Tomato Plant Diseases using Random Forest 3
  2. Identification of Plant Leaf Diseases Using a Nine-layer Deep Convolutional

Neural Network 4

* 1. Plant Diseases Detection and Classification Using Machine Learning Models 4
  2. A Detection and Severity Estimation System for Generic Diseases of Tomato Greenhouse Plants 5
  3. Detection of Banana Plants and Their Major Diseases through Aerial Images and Machine Learning Methods: A Case Study in DR Congo and Republic Of Benin 5
  4. Identification of Plant Leaf Disease Using Machine learning techniques 6
  5. Identification of Plant Diseases Using Convolutional Neural Networks 6
  6. Monitoring Of Rice Plants for Disease Detection Using Machine Learning 6
  7. Plant Disease Detection Using Computational Intelligence and Image Processing 7
  8. Research on Recognition Model of Crop Diseases and Insect Pests Based on

Deep Learning in Harsh Environments 7

* 1. Uncertainty Quantification for Plant Disease Detection Using Bayesian Deep Learning 8
  2. A Generic Approach for Wheat Disease Classification and Verification Using

Expert Opinion for Knowledge-Based Decisions 8

* 1. Automated Tomato Leaf Disease Classification Using Transfer Learning-based

Deep Convolution Neural Network 9

* 1. Identifying and Classifying Plant Disease Using Resilient LF-CNN 9
  2. Performance of Deep Learning Vs Machine Learning in Plant Leaf Disease

Detection 9

* 1. Plant Leaf Disease Classification Using Efficient Net Deep Learning Model 10
  2. Potato Plant Leaves Disease Detection and Classification Using Machine

Learning Methodology 10

1. [REQUIREMENT 13](#_TOC_250016)
   1. [Front-end development 13](#_TOC_250015)
   2. [Back-end development 13](#_TOC_250014)
2. EXISTING SYSTEM 16
3. [PROPOSED SYSTEM 17](#_TOC_250013)
   1. [System Design 17](#_TOC_250012)
   2. [Implementation 17](#_TOC_250011)
4. [RESULTS AND DISCUSSIONS 22](#_TOC_250010)
   1. [Experimental setup 23](#_TOC_250009)
   2. [Dataset 23](#_TOC_250008)
   3. [Performance evaluation metrics 25](#_TOC_250007)
   4. [Performance evaluation 26](#_TOC_250006)
      1. [Training and testing phase 26](#_TOC_250005)
      2. [Analysis of various leaf disease detection 28](#_TOC_250004)
      3. [Summary of disease detection 32](#_TOC_250003)
      4. [Pest detection analysis 33](#_TOC_250002)
   5. Screenshots of web application 34
5. [CONCLUSION 37](#_TOC_250001)
6. [FUTURE SCOPE 38](#_TOC_250000)

REFERENCE 39

4.1 Classification of machine learning algorithms based on plant disease 18

detection approaches.

* 1. System Architecture Diagram 19
  2. Process Flow Diagram 19
  3. ResNet 152 architecture 22
  4. MobileNet V2 23
  5. Epoch Vs Accuracy graph on training 25
  6. Epoch Vs Loss graph on training 26
  7. Epochs Vs Accuracy values in validation 27
  8. Epochs Vs Loss values in validation 28

2.1 Comparison of existing ML-based plant disease detection methods 10

* 1. The accuracy and loss values for each epochs during training 25
  2. The accuracy and loss values for each epoch during validation 27
  3. The performance analysis on the dataset 30
  4. The confusion matrix for pest detection 34
  5. Name of the plants with their total count in the dataset 35

|  |  |
| --- | --- |
| ANN | Artificial Neural Network |
| CNN | Convolutional Neural Network |
| FN | False Negative |
| FP | False Positive |
| GLCM | Grey Level Co-occurrence Matrix |
| GPU | Graphics Processing Unit |
| KNN | K-Nearest Neighbor |
| MC | Monte Carlo |
| ML | Machine Learning |
| ORB | Oriented FAST and rotated BRIEF |
| RGB | Red Green Blue |
| RMS | Root Mean Square |
| SGD | Stochastic Gradient Descent |
| SGLD | Stochastic Gradient Langevin Dynamics |
| SIFT | Scale Invarient Feature Transformation |
| SURF | Speeded Up Robust Features |
| SVM | Support Vector Machine |
| TN | True Negative |
| TP | True Positive |

# INTRODUCTION

Agriculture possesses a title role in India. India has a variety of climates and soil in different parts of the country. The crops are enduring a large number of insects, pests, diseases, nematodes, weeds, and even nutritional deficiencies [1]. Annual crop loss in India due to pest and disease infestation amounts to a whopping 50,000 to 90,000 crores. For ensuring good crop productivity; pests, diseases, and nutrients have to be managed efficiently. Indiscriminate use of chemical pesticides and fertilizers has resulted in several problems including harm to humans, animal health, and the environment in general. So, the judicious use of chemical pesticides and their timely recommendations are very important in crop production [22]. In this scenario, a detailed study is carried out to analyze the existing plant disease detection techniques. Accurate knowledge about the pests and diseases prevailing in their crops helps to eradicate economic loss. Plant diseases are examined by the farmers through their professional experience or by an agricultural proficient through their partially accurate observation, both are time-consuming processes. Nowadays, most nutrient deficiency symptoms are misinterpreted as diseases [17]. Early detection of symptoms is crucial in the agriculture field to increase crop productivity [6]. No serious attempts were made to address crop health. Plant disease detection using deep learning ensures timely and accurate crop management practices.

The accurate diagnosis of the disease is a tedious task. If the disease is diagnosed correctly, half of the problem is completed. To ensure successful agricultural activities, effective disease control is necessary. To predict the diseases accurately and user friendly, an interface is developed. Pests are also responsible for crop reduction. The pest prediction is included in this project to successfully identify the pests in the plants.

# LITERATURE SURVEY

All the papers studied for this project work uses machine learning or deep learning-based technologies for disease classification. The two major learning methods in machine learning are supervised learning and unsupervised learning. Supervised learning techniques such as Random Forest [9] [13] [16] [19], K-Nearest Neighbors [15], and Support Vector Machine [24] and Unsupervised learning techniques like K-Means [5] [20] are being used by different models. Some major deep learning techniques such as transfer learning [23], pretrained AlexNet [12], resident LF [8], EfficientNet [4], ResNet [10] [2] [25], Artificial Neural Network [14], Deep Convolutional Neural Network [7] [18], VGG16 [21], and Bayesian Convolutional Neural Network [11] techniques are also used in these papers.

## Machine Learning-based for Automatic Detection of Corn Plant Diseases Using Image Processing [13]

In 2018, Holman et al. proposed image processing-based, automatic disease identification in corn plants using a Machine Learning algorithm. Features like RGB, scale-invariant feature transform (SIFT), speeded up robust features (SURF), and Oriented FAST and rotated BRIEF (ORB) are used for the identification of diseases on the corn plants. These features are examined using Machine Learning algorithms like Support Vector Machines (SVM), Decision Trees, Random Forest, and Naive Bayes. The outcome of this model specifies that the color properties like RGB attained better performance with SVM. The accuracy of this model is about 95.6%. In this approach, the characteristics are identified rapidly and are categorized to the appropriate infections. The disadvantage of this approach is that the assumptions do not hold on to real life conditions.

## A Review on Machine Learning Techniques for Rice Plant Disease Detection in Agricultural Research [5]

Daniya et al. (2019) proposed a framework for detecting diseases in paddy crops using Machine Learning and image processing techniques to diagnose diseases like false smut, blast disease, rice bacterial blight, and sheath rot. Probabilistic Neural Networks, Genetic Algorithms, K- Nearest Neighbor classifier, and SVM were the classification techniques that are used for detecting diseases in various agricultural researches. Images are collected using cameras with

high resolution. Neural Network (NN) based classifiers were used for classifying the diseases. The NN classifiers have the top most accuracy level, the Radial basis function network models have an accuracy of about 95.5% and the SVM classifiers have the least accuracy level. This approach has higher segmentation efficiency and accuracy. Its drawbacks include isolated noise and certain holes in the image even after segmentation.

## Application of Machine Learning in Detection of Blast Disease in South India Rice Crop [18]

Ramesh et al. proposed a methodology to detect rice blast disease using Machine Learning algorithms in 2019. Images are collected from the paddy field using cameras. In the image pre- processing stage, the input images with RGB color are changed to HSV (Hue Saturation Value). Here, segmentation is done using K-means clustering. The Grey level co-occurrence matrix (GLCM) is used for feature extraction. K-Nearest Neighbor (KNN) and Artificial Neural Network (ANN) are utilized to classify the images. The confusion matrix is applied to compare the performance of these classifiers. ANN classifier performed better than the KNN classifier in this approach. This model attained a high accuracy of 95.48%.

## Diagnosis of Tomato Plant Diseases using Random Forest [9]

In 2019, Govardhan et al. utilized the Random Forest algorithm and created a model to identify and classify various diseases in tomato plants. Dataset consists of pathological as well as healthy leaves. Mainly three features are extracted, they are, texture using Haralick Texture, color using Color Histogram, and shape using Hu Moments. The project is split into two phases. Initially, attributes are derived and stored in an array. In the subsequent stage, the Random Forest algorithm is used to train the model and to predict the diseases correctly. This model has 95% accuracy in disease forecasting. When trained with a huge variety of pictures, results can be obtained at a quick pace. This project cannot identify the diseases caused by nutritional deficiencies.

## Identification of Plant Leaf Diseases Using a Nine-layer Deep Convolutional Neural Network [7]

Convolutional neural networks are one of the most used technologies for plant disease detection. In 2019, Geetharamani et al. proposed a novel plant leaf disease identification model based on a Deep Convolutional Neural Network. They used augmentation methods such as image flipping, scaling, noise injection, gamma correction, and rotation. The data augmentation can increase the performance of the system considerably. When this model is compared with other existing models, the proposed model gives a higher performance in the validation data. The accuracy of this work is quite larger (96.46%) than the existing machine learning approaches. The main disadvantage of this system is that it does not use the enhanced dataset (an enhanced dataset contains images from more vast sources like different plant types, leaf growths, geographic areas, and cultivation conditions).

## Plant Diseases Detection and Classification Using Machine Learning Models [16]

In 2019, Poojan et al. detected plant diseases quickly and efficiently using machine learning models. Image acquisition, image processing, image segmentation, feature extraction, and classification are done in this method. To improve the accuracy of plant disease detection, the Random Forest algorithm is applied, since it has minimum computational complexity for the accurate detection of diseases. The disadvantage of this approach is the need for a highly adaptable and generalized system for disease detection. Feature extraction is done by applying segmented images. Features are extracted using Gray level co-occurrence matrix (GLCM) and they are stored in a data set. Images of leaves with late blight, early blight, and bacterial spots were used for training the model. The efficiency of this classification is compared by the types of diseases predicted correctly to the original dataset. Random Forest, Decision Tree, K-Nearest Neighbor, and Support Vector Machine are used as classifiers. Random Forest yielded an accuracy of 98%.

## A Detection and Severity Estimation System for Generic Diseases of Tomato Greenhouse Plants [25]

This paper proposed by Patrick Wspanialy and Medhat Moussa in 2020, designed a computer vision system that can detect plant diseases. The plant village dataset is used for training. There were nine kinds of diseased tomato leaves in the dataset. To decrease the chance of over fitting, the variations in the image dataset are increased through data augmentation. A 50 layers deep, convolutional neural network called ResNet-50 is the main attraction of this approach. ImageNet, which is a large dataset that contains millions of images, is used to pre-train this network. By training the model with the ImageNet dataset, the model can identify features such as lines and corners which are common to all objects. The training phase of this model is divided into epochs. After training, the model is tested on various datasets. The outcome of this approach was found using the Jaccard index.

## Detection of Banana Plants and Their Major Diseases through Aerial Images and Machine Learning Methods: A Case Study in DR Congo and Republic Of Benin [19]

Detection of diseases in banana plants is very important because of the large variety of diseases that can attack the banana plants. In 2020, Selvaraj et al. used aerial images and Machine Learning methods to design a banana plant leaf disease detection system. The dataset is collected from the Kabare district. Here two main technologies are used: remote sensing and Machine Learning. Remote sensing in the disease detection systems helps to provide cost-efficient and precise information. Most disease detection systems only focus on the single sensor-based solution. But combining different information sources is very important. Therefore, the machine learning model is integrated with high-resolution satellite data. The advantage of this method is that it is possible to integrate the output of this method with other disease detection systems. The major disadvantage of this approach is that accurate disease detection with an open-source medium solution is still a challenging task.

## Identification of Plant Leaf Disease Using Machine learning techniques [14]

Shabari et al. in 2020 suggested a framework for the identification and classification of leaf diseases along with the percentage of infected parts on the leaf. The input sample consists of leaf diseases like Black measles and Black rot on grape leaves and Blast diseases of rice. Image pre- processing is done by using a median filter to remove distortion on the collected images. K- means clustering is used for segmentation to extract a desired part from the leaf image. The Grey-Level Co-Occurrence Matrix is used for feature extraction. Artificial Neural Network (ANN), Support Vector Machine (SVM), and Naive Bayes classifier are used for classifying the diseases. ANN performed better than the SVM and Naive Bayes classifier. This work gives an accuracy of 93.4%. The advantage of this approach is that the pests were detected at every surrounding in the picture. The disadvantage of this approach is that the pigment on the plant leaf is not properly captured.

## Identification of Plant Diseases Using Convolutional Neural Networks [12]

In 2020, Sachin et al. found an accurate and reliable plant disease detection system using a transfer learning approach in the soya bean plant. A pretrained AlexNet and GoogLeNet Convolutional neural network recognize the soya bean diseases using the transfer learning method. 649 images are utilized to train AlexNet, and GoogLeNet CNN used 550 image samples of infected and healthy soya bean leaves to strongly predict soya bean diseases. Here a five-fold cross-validation strategy is used to evaluate the performance. The classification was carried out with the help of AlexNet and GoogleNet models by improving different hyper parameters. This model achieved an accuracy of 98.75% and 96.25%, respectively. Advantages are it is more accurate than conventional pattern recognition techniques and has higher efficiency. The disadvantage is the low performance rate.

## Monitoring Of Rice Plants for Disease Detection Using Machine Learning [15]

Swetha et al. (2020) recommended a framework for diagnosing and categorizing diseases on rice crops like the bacterial blight of rice, rice blast, tungro of rice, and false smut. Images are

collected using cameras from paddy fields. During image processing, the size of gathered images is converted to a standard size to minimize the effect of lightning on images. Morphological features like the shape and color of leaves are considered for feature extraction. Support Vector Machines and K Nearest Neighbor are used as classifiers to classify the disease. The training process of Convolutional Neural Network includes convolution, pooling, and training. Forward propagation is done on images to minimize the error during training and back propagation is performed to correct the filter values on the images. This model has an accuracy of 91.23%.

## Plant Disease Detection Using Computational Intelligence and Image Processing [24]

In 2020, Vishnoi et al. detected plant diseases using computer vision. Support Vector Machine (SVM) and Artificial Neural Network (ANN) are used in this project. During preprocessing, the infected area is segmented, followed by final feature extraction and selection. The purpose of this work is to classify plant images in the order of their symptoms and infections of the diseases. Soft computing and machine learning techniques are applied to complete this initiative successfully. SVM, ANN, and K-Nearest Neighbor algorithms are used for this work. Color histogram, which helps to know the color contents, is used to extract color features. So an accurate and reliable detection of plant diseases is obtained. In this approach, multiple diseases can be classified easily and efficiently. The drawback of this method is its high computational complexity.

## Research on Recognition Model of Crop Diseases and Insect Pests Based on Deep Learning in Harsh Environments [2]

Yong Ai et al. (2020) designed a system to recognize leaf abnormalities and pests. Data pre- processing techniques like color dithering, contrast changes, and Gaussian noises are added to augment the image. The deep learning network model, Inception-ResNet-v2 is used to predict diseases and pests. Wechat applet is made to help normal people to recognize crop disorders and pests. The unhealthy crops are identified and proper prevention guidelines are given. The hybrid network model gives better results with a precision of 86%. The use of an elaborate dataset can increase the efficiency of the model.

## Uncertainty Quantification for Plant Disease Detection Using Bayesian Deep Learning [11]

Uncertainty quantification for plant disease detection using Bayesian deep learning is proposed by S Hernandez and Juan L Lopez in 2020. The Plant Village dataset is used in this approach. The dataset contains images of 26 varieties of diseases and 14 varieties of crops. It contains 54,306 images. Machine learning is used by a lot of experts in different areas. One of these is the plant's disease detection. A lot of data should be handled in such areas. It will make the system more complicated. Deep learning, which came under machine learning, is very well suited for handling signals and large volumes of images. Deep learning models learn from raw data. Using Convolutional Neural Network (CNN) in deep learning we can avoid the tedious process of manual creation of features from data. Here we use the Bayesian technique to achieve a well calibrated uncertainty estimate. The use of the Bayesian technique is the major feature of this approach. In this approach, we use three kinds of optimization techniques like Stochastic gradient Langevin dynamics (SGLD), Stochastic Gradient Descent (SGD), and Monte Carlo (MC) dropout. A technique among these three is hard to choose since each one has its advantages and disadvantages.

## A Generic Approach for Wheat Disease Classification and Verification Using Expert Opinion for Knowledge-Based Decisions [10]

A generic classification approach for leaf disease is proposed by Waleej Haider et al. (2021) to make knowledge-based decisions. The unreliable solutions due to the recurrent changes like diseases and the absence of an agricultural expert's opinion are rectified in this journal. The image dataset necessary for infection detection in wheat is directly collected from different fields. After data pre-processing, the size of training data is enhanced by augmentation techniques. The image dataset is classified using the Convolutional Neural Network. Another text dataset consisting of different symptoms of diseases that can occur in wheat is collected and that data is used to identify the infection. The Decision Tree algorithm is used to sort the dataset. The main advantage of this approach is the classification accuracy of 97.2%.

## Automated Tomato Leaf Disease Classification Using Transfer Learning- based Deep Convolution Neural Network [23]

In 2021, Thangaraj et al. detected plant diseases with minimal image data. AI-based deep learning technology is used in this approach. In deep learning technology, the knowledge obtained when a problem is solved can be reused for a different problem and this is called transfer learning. Here TL-based Deep CNN model is used for detecting different types of diseases in tomatoes like leaf mold, yellow leaf curl virus, mosaic virus, early blight, late blight, and target spot. Adaptive moment estimation, stochastic gradient descent (SGD) optimizers, and Root Mean Square (RMS) prop optimizers are used for estimating performance. The accuracy of this approach is about 99%.

## Identifying and Classifying Plant Disease Using Resilient LF-CNN [8]

Gokulnath et al. detected plant diseases with high accuracy using resilient LF-CNN in 2021. Plants get infected when the virus affects the host during a favorable environment. Early detection helps to avoid the severe destruction of crops. Here Convolutional neural networks are used. Images are collected from the plant village data set and then the images are augmented. The loss function is calculated and at the final stage, the software function is employed for classification. This model has an accuracy of 98.93%.

## Performance of Deep Learning Vs Machine Learning in Plant Leaf Disease Detection [21]

Sujatha et al. compared the performance of deep learning and machine learning in plant leaf disease detection through this paper published in 2021. Plant disease identification is an important and challenging task. There are many models and technologies available now to find plant diseases. But each one has its advantages and disadvantages. Different models use different technologies for finding diseases. Machine learning and deep learning are the trending technologies used in most systems. There are a lot of models in machine learning and deep learning. By examining those existing works, this paper compares the performance of deep learning and machine learning models in plant disease detection. The areas under the curve, precision, accuracy and F1 score are considered to get better results.

## Plant Leaf Disease Classification Using Efficient Net Deep Learning Model [4]

In 2021, Ümit et al. detected plant diseases with great accuracy, sensitivity, specificity, and precision. The ordinary diagnosis of plant diseases is time-consuming and the accurate disease prediction depends on the pathologist's capabilities. For plant disease detection, computer-aided diagnostic systems were used previously. In this model, there is no need for pre-processing. Here plant leaf disease classification is done using EfficientNet deep learning architecture. To train the model, the Plant Village dataset is used. The model is trained with the original dataset having 55,448 images and an augmented dataset having 61,486 images. This model uses the transfer learning approach. The EfficientNet B4 approach has an accuracy of 99.97% in the augmented dataset and the B5 approach achieved 99.91% efficiency in the original dataset.

## Potato Plant Leaves Disease Detection and Classification Using Machine Learning Methodology [20]

Aditi et al. in 2021 proposed a model to detect and classify the diseases on potato plants such as late blight and early blight disease using Machine Learning methodology. Segmentation is done using the k-means algorithm. GLCM is used for the extraction of features. A multi-class Support Vector Machine (SVM) is used for classifying the diseases. For the dataset, images are collected from the 'plant village dataset' that consists of both diseased and healthy leaf images of 14 crop species. Image acquisition, pre-processing, segmentation, extraction of features, and classification of disease are the major phases in the detection of plant diseases. This model achieved an accuracy of about 96%, 96.5% of precision, 96.25% of recall, and 96.2% of F1 score. The overall accuracy of this model is based on the precision, recall, and F1-score.

**Table 2.1 Comparison of existing ML-based plant disease detection methods**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ref.,**  **Crop** | **Diseases Found** | **Technique Used** | **Observations** | **Disadvantages** |
| [13]: Corn | Grey leaf spot, common rust, northern leaf blight | Random Forest | Easy and quick classification. | Expectations do not  meet original conditions. |
| [5]: Rice | Rice blast, leaf brown spot, heath rot, and Bacterial blight (BB). | K-Means | Segmentation efficiency and accuracy was high. | Isolated noise and holes exist in the picture even after splitting up. |
| [18]: Rice | Blast | Artificial Neural  Network(ANN) | Perfection is enhanced to  99% for ANN-based classifying. | Training time is high. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| [9]: Tomato | Late blight, septoria leaf spot, spider mite, target spot, early blight, mosaic  virus, yellow leaf curl, etc. | Random Forest | Prediction accuracy is 95%, adaptable, and sharp results. | Hard to identify diseases caused by nutritional  deficiencies. |
| [7]: Apple, grape, etc. | Black rot, leaf blight, leaf mold, powdery mildew, early blight, bacterial spots, etc. | Deep CNN | Reliability of 96.46%, b e t t e r , and unambiguous classification. | Absence of an enhanced dataset. |
| [16]: Pepper  and tomato | Bacterial spot, early blight, and late blight. | Random Forest | Maximum a c c u r a c y and m i n i m u m  computational complexity. | Lacks a highly adaptable and generalized system. |
| [25]:  Tomato | Bacterial Spot, Early Blight, Late Blight, Leaf Mold, Mosaic Virus, Septoria Leaf Spot, Spider Mites, Yellow Leaf Curl  Virus | Residual  network with 50 layers(ResNet- 50) | Automated, economical, and high precision disease identification system. | Improved datasets are required. |
| [19]: Banana | BXW (Xanthomonas Wilt of Banana) and BBTD (Banana Bunchy Top  disease) | Random Forest | An enhanced digital disease analysis system. | Perfection in banana mapping is still challenging. |
| [14]: Grape | Black rot and black measles | ANN | The pests are detected from the picture. | The pigments in the  plants are not properly extracted. |
| [12]: Soya bean | Bacterial blight, frogeye leaf spot, and brown spot. | AlexNet | Better result and higher accuracy. | Low-performance rate. |
| [15]: Rice | Bacterial blight of rice (RBB), Rice blast disease (RB), Rice tungro, and  false smut | K-Nearest Neighbors (k- NN) algorithm | Ranking precision for  infe c t i on detection is high. | Irrelevant inputs can cause expensive testing time. |
| [24]:  Tomato, pepper, apple | Bacterial spot, early blight, black rot, and common rust | SVM algorithm | Efficient, easier, and multiple disease  classification. | High computational complexity. |
| [2]: Tomato, potato, corn | Rust diseases, Puccinia polysra, peach scab,  etc. | Inception- ResNet-v2 model | Hybrid network models are efficient. | Lacks extended data set. |
| [11]:  Blueberry, cherry, etc. | Apple scab, Common rust, Black rot, Early blight, Late blight, Bacterial spot, Leaf scorch, etc. | Bayesian deep learning  techniques | Fine-tuned evaluation for uncertain data. | Computational cost for the  Bayesian model is high. |
| [10]: Wheat | Common bunt, fusarium head blight, a n d sooty head mold | Decision Tree and ResNet 50 | Improvements in the classification accuracy from 93.4% to 97.2%. | Lacks optimization. |
| [23]:  Tomato | Early blight and late blight | Deep CNN | For unbalanced data, it has a better recognition date. | It cannot identify  tomato leaf disease severity. |
| [8]: Tomato  and potato | Early blight, Late blight,  Yellow leaf, and Leaf mold. | Resilient LF-  CNN model | It gives 98% accuracy  a nd improves the lifecycle of plants. | It doesn’t get evaluated  under distinct conditions. |
| [21]: Citrus | Black spot, canker, greening, Melanose | VGG-16 | F1 score, precision , accuracy , and the area below the curve is contemplated to get a  better outcome. | It doesn’t make use of the fuzzy logic and bio- inspiredways. |
| [4]: Apple, pepper, tomato, grape,  blueberry, potato | mosaic virus, black rot, leaf mold, early blight, late blight, rust, andbacterial spot | EfficientNet | It gives average accuracy, sensitive, and precise output. | It can’t give accurate predictions in difficult environments. |
| [20]: Potato | Late blight and early blight | K-Means | The prediction accuracy is high and robust working is ensured. | Require more training time. |

Various methods employed for the detection of diseases in a variety of crops like rice, tomato, apple, pepper, etc using machine learning and deep learning-based approaches are summarized inthis table. Some models only focused on a single crop for better performance while some other models included a variety of crops. Each technique has its benefits. Performance of different machine learning methodologies and classification techniques are evaluated concerning accuracy. Efficient- Net deep learning model in [4] attains the topmost accuracy of 99.9% but it can’t predict the images with less clarity. The Random forest algorithm used in [16] achieved an accuracy of 98% which is the second most accurate model. Moreover, the Inception-ResNet-v2 model employed in [2] attains the least accuracy rate when compared with all other techniques used in plant disease detection.

# REQUIREMENT

The front-end of this project is done with the help of pyqt5 and back-end is developed with flask and python.

## FRONT-END DEVELOPMENT

#### PyQt5

Pyqt5 is a GUI widgets toolkit. It is a Python interface for Qt, one of the most powerful, and popular cross-platform GUI libraries. It gives programmers the freedom to create GUIs of their own choice while also providing a lot of good pre-built designs.

PyQt API is a set of modules that contains a large number of classes and functions. While the QtCore module contains non-GUI functionality for working with files and directories. The qtGui module contains all the graphical controls. In addition, there are modules for working with XML (QtXml), SVG (QtSvg), and SQL (QtSql), etc.

Some of the modules are;

QtCore − Core non-GUI classes used by other modules QtGui − Graphical user interface components

QtMultimedia − Classes for low-level multimedia programming QtOpenGL − OpenGL support classes

QtScript − Classes for evaluating Qt Scripts

QtWidgets − Classes for creating classic desktop-style UIs QtDesigner − Classes for extending Qt Designer

PyQt is compatible with all the popular operating systems including Windows, Linux, and Mac OS.

## BACK-END DEVELOPMENT

#### Flask

Flask is a web framework. This means flask provides you with tools, libraries, and technologies that allow you to build a web application. This web application can be some web pages, a blog, a wiki, or go as big as a web-based calendar application or a commercial website.

Flask is part of the categories of the micro-framework. Micro-framework is normally a framework with little to no dependencies on external libraries. This has pros and cons. Pros would be that the framework is light, there is little dependency to update and watch for security bugs, cons is that sometimes you will have to do more work by yourself or increase yourself the list of dependencies by adding plug-ins.

#### Python

In technical terms, Python is an object-oriented, high-level programming language with integrated dynamic semantics primarily for web and app development. It is extremely attractive in the field of Rapid Application Development because it offers dynamic typing and dynamic binding options. Python is relatively simple, so it's easy to learn since it requires a unique syntax that focuses on readability. Developers can read and translate Python code much easier than other languages. In turn, this reduces the cost of program maintenance and development because it allows teams to work collaboratively without significant language and experience barriers. Additionally, Python supports the use of modules and a package, which means that programs can be designed in a modular style and code can be reused across a variety of projects. Once you've developed a module or package you need, it can be scaled for use in other projects, and it's easy to import or export these modules. One of the most promising benefits of Python are that both the standard library and the interpreter are available free of charge, in both binary and source form. There is no exclusivity either, as Python and all the necessary tools are available on all major platforms. Therefore, it is an enticing option for developers who don't want to worry about paying high development costs. If this description of Python is over your head, don't worry. You'll understand it soon enough. What you need to take away from this section is that Python is a programming language used to develop software on the web and in app form, including mobile. It's relatively easy to learn, and the necessary tools are available to all free of charge. That makes Python accessible to almost anyone. If you have the time to learn, you can create some amazing things with the language.

#### Python libraries OpenCV

OpenCV stands for open-source computer vision. It contains machine learning libraries. These libraries contain more than 2000 optimized algorithms like machine learning algorithms and computer vision algorithms. These algorithms are used for identifying an object, track the

moving of an object, face recognition. Opencv is mainly used in image processing. It is written in C++ and also its interface is C++. It supports deep learning frameworks like TensorFlow, torch.

#### Tensorflow

Tensorflow is an open-source artificial intelligence library. It gives better functionality and service when compared with other machine learning frameworks. It is written in C++. Tensorflow is used for computing tensor operation which indicates the representation of data in N-dimensional matrices. It is an extension of NumPy. TensorFlow contains a math library that can be used in machine learning applications. It is very flexible because TensorFlow is a low- level library. Tensorflow helps the user to create a large-scale neural network with many layers. It is also used classification, perception, prediction

#### Keras

Keras is a free software library that gives a python interface for neural networks. It is used in machine learning. Keras is the interface of the TensorFlow library. It can run on top of TensorFlow, a Microsoft cognitive toolkit. Keras is used for implementing the concepts like regression, clustering, etc.

#### Numpy

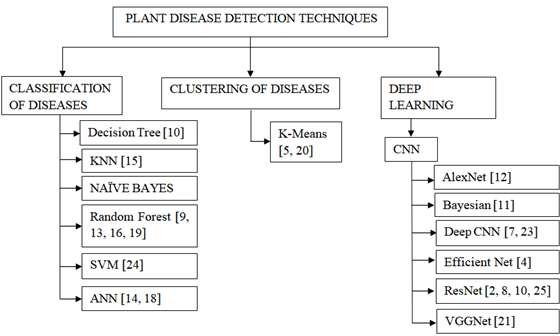
Numpy stands for Numerical Python is a library for the Python programming language. It is used as an internal library. Numpy is used for array representation. Numpy reduces loops and prevents them from getting tangled up in iteration indices.

#### Scikit-learn

Scikit-learn is a free software library for the Python programming language. It is used in machine learning. It is a very useful library for machine learning in a python programming language. The sklearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering, and dimensionality reduction.

# EXISTING SYSTEMS

Machine learning and deep learning techniques have been extensively used for the identification of plant diseases. For disease classification the machine learning algorithms such as Random Forest, K- Nearest Neighbors, Support Vector Machine, Decision tree, Logistic Regression, Naive Bayes, and Stochastic Gradient Descent are widely used. In deep learning, the CNN models such as AlexNet, VGG-16, VGG-19, InceptionNet, GoogLeNet, ResNet, Squeeze Net, DenseNet, and Bayesian are used for plant leaf disease detection and classification.



**Fig.4.1. Classification of machine learning algorithms based on plant disease detection approaches.**

Some of the disadvantages of existing systems are that they have specific background requirements, selective applicability to plant/disease types, noisy images.

In the case of pest detection, there are not many existing systems in this area. There is quite a lot of research going on in this area but the unavailability of a dataset makes the implementation more difficult.

# PROPOSED SYSTEM

Our proposed system is used to collect images of leaves of different plants and then classify them as healthy or unhealthy and also to detect the pest present in them. In data collection, we can collect the image of the data from the plant village database. Through preprocessing, the quality of the input image is improved and the undesired distortion is removed from the image. Then collect the train set or test the set of data and then compare with the output of the desired data it will classify that the given leaf is healthy or not. To detect the plant disease, we used the ResNet152 technique and to detect pests we used the MobileNetV2 technique.

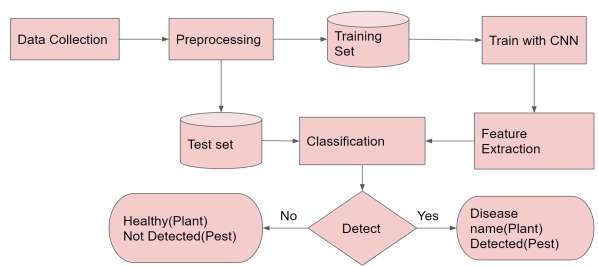
## SYSTEM DESIGN

### In this project, the client selects the image of diseased leaf or pest they need to recognize and sends it to the server. Server accepts the image from the client User Interface. The image is classified to its most appropriate category with the help of the ResNet 152 model. The detail of the classified image is sent to the client User Interface as a response from the server. The disease of the leaf or the pest detected is displayed on the client Interface.

**Fig.5.1. System Architecture Diagram**

## IMPLEMENTATION

### The implementation of the project from data collection to the classification of the disease or pest in a plant is discussed.



**Fig.5.2. Process Flow Diagram**

#### Data collection

Data collection is the process of gathering information from different sources. Collecting data for training the particular CNN model is the basic step in any deep learning project. Here we collected plant leaf dataset and pest dataset. Plant village dataset is used for plant leaf disease classification. Both the datasets are taken from Kaggle. In testing phase, users capture the image of a leaf that they want to check if that particular plant has a disease or net and that image is given as input to the system. We have captured images of the healthy and unhealthy leaves of the plant and those were the input to the system.

#### Image preprocessing

The aim of preprocessing is an improvement of the image data that suppresses unwanted distortion or enhances some image features important for further processing. Normalization and image augmentation are the two preprocessing steps we do.

#### Normalization

Normalization is a process that changes the range of pixel intensity values. The purpose of normalization is to transform data in a way that they are having similar distributions. In image processing, normalization is a process that changes the range of pixel intensity values. Here the pixel intensity values are transformed within the range (0, 1). Normalization dramatically improves model accuracy

#### Data augmentation

Data augmentation is used to expand the training dataset to improve the performance and ability of the model to generalize. Image data augmentation is supported within the Keras deep learning library via the ImageDataGeneratorclass. The image is shift, flip, brighten, and enlarge using image data augmentation. We mainly do two kinds of data augmentation, Center crop and scaling.

1. Center crop

The center crop scale type is used to Scale the image uniformly. It means to take care of the image's ratio as by doing that both dimensions width and height of the image are going to be adequate to or larger than the corresponding dimension of the image view.

1. Scaling

When scaling a raster graphics image, a new image with a higher or lower number of pixels must be generated. In the case of decreasing the pixel number this usually results in a visible quality loss.

#### Training with CNN

Before training the CNN models prepare the training and testing data by dividing the dataset in the ratio 70:30.Build the CNN layers using the TensorFlow library. Select the Optimizer. Then, train the network and save the checkpoints.

#### Feature extraction

Feature extraction may be a part of the dimensionality reduction process, during which an initial set of the data is split and reduced to more manageable groups. The most important characteristic of those large data sets is that they need an outsized number of variables. These variables require a lot of computing resources to process. Feature extraction helps to get the best feature from those big data sets by selecting and combining variables into features, thus, effectively reducing the amount of data. These features are easy to process, but still ready to describe the particular data set with accuracy and originality.

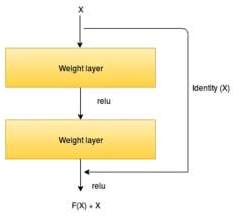
#### Classification

Classification is done by respective CNN model. Plant disease classification is done by ResNet 152 and for pest detection classification is done by MobileNet V2.

#### (i)ResNet 152

ResNet 152 can have a very deep network of up to 152 layers by learning the residual representation function instead of learning the signal representation directly. ResNet is different from other CNN models because of its no. of layers. Here we are using Resnet which has 152 layers.

Resnet was released in 2015 by Microsoft Research and it's the winner of ImageNet large scale visual recognition challenge 2015.Resnet is considered as very deep model in terms of no. of layers. Theoretically, when we go deeper and deeper training and testing error will improve. But practically it doesn't work in that way. If the number of layers is very high vanishing gradient problem will arise. To get over this problem Resnet introduce skip connections to provide alternative path for gradient to flow.



**Fig.5.3. ResNet 152 architecture [28]**

From this figure,

The input here is X . Let the output of the whole system is H(x). The output of second layer alone is F(x). But here the input is also copied to output. So, the net output H(x) = H(x) + x

It can be also written as, F(x) = F(x)-x

In the worst case if it’s not learning from anything or the weights are not updated or all the values are very small or zero then, it will at least learn x. Thus, gradient flow happens.

Here 2 layers are skipped at a time. That's why it's called as skip connection.

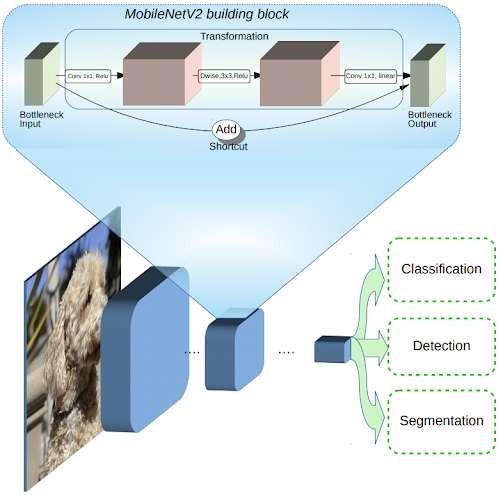
Advantages

1. It can be trained easily without increasing the training error percentage
2. Helps in tackling the vanishing gradient problem using identity mapping.
3. The error rate of Resnet 152 is very low. It's around 4% which is even lesser than human error.

#### (ii)MobileNet V2

MobileNet V2 is a pretrained, single stage, object detection model. It is pretrained with MS-COCO data which has about three lakh images and more than eight object classes. It doesn't have any intermediate stage. It supports classification, detection and segmentation. The main feature of mobile net is that it has the ability to run deep networks on personal mobile devices. MobileNet V2 is realized as a part of tensor flow slim image classification library in 2019. Compared to MobileNet V1 MobileNet V2 introduced two new features to its architecture.

1. Linear bottle neck between the layers
2. Shortcut connections between the bottle necks



**Fig.5.4. MobileNetV2 [27]**

The linear bottle neck between the layers helps to transform lower-level concepts to higher level descriptors. Shortcut connections between the bottle neck helps with flow of gradient through the network. So, there is a direct connection from input to the output. By doing this, gradient can never be zero. Architecture of MobileNet V2 is inverted residual architecture. MobileNet V2 uses depth wise separate convolutions as building blocks.

Advantages

1. Faster for an equivalent accuracy across the whole latency spectrum.
2. Light weight
3. Compatible

# RESULTS AND DISCUSSIONS

## EXPERIMENTAL SETUP

The training and testing of the model is done on Python 3 in Google Colab with Tesla T4 GPU. Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz Processor with Ram capacity of 8.00GB and 64 bit operating system device is used to complete this work. The proposed system is evaluated on “Plant village” dataset. It consist of 39(Apple scab, Black rot, Cedar apple rust, Apple healthy, Background without leaves, Blueberry healthy, Cherry Powdery mildew, Cherry healthy, Corn Cercospora leaf spot Gray leaf spot, Corn Common rust, Corn Northern Leaf Blight, Corn healthy, Grape Black rot, Grape Esca (Black Measles), Grape Leaf blight(Isariopsis Leaf Spot), Grape healthy, Orange Haunglongbing (Citrus greening), Peach Bacterial spot, Peach healthy, Bell pepper Bacterial spot, Bell pepper healthy, Potato Early blight, Potato Late blight, Potato healthy, Raspberry healthy, Soybean healthy, Squash Powdery mildew, Strawberry Leaf scorch, Strawberry healthy, Tomato Bacterial spot, Tomato Early blight, Tomato Late blight, Tomato Leaf Mold, Tomato Septoria leaf spot, Tomato Spider mites, Tomato Target Spot, Tomato Yellow Leaf Curl Virus, Tomato mosaic virus, Tomato healthy) categories of various plant diseases. Training: testing ratio is taken as 70:30 to help the model achieve more accuracy.

## DATASET

Dataset is taken from plant village dataset from Kaggle. 55,448 images of plant leaf diseases with 39 categories are obtained.

**Table 6.1 Name of the plants with their total count in the dataset**

|  |  |
| --- | --- |
| **Name of the plant** | **Total** |
| Apple scab | 630 |
| Black rot | 621 |
| Cedar apple rust | 275 |
| Apple healthy | 1645 |
| Background without leaves | 1143 |
| Blueberry healthy | 1502 |
| Cherry Powdery mildew | 854 |
| Cherry healthy | 1052 |

|  |  |
| --- | --- |
| Corn Cercospora leaf spot Gray leaf spot | 513 |
| Corn Common rust | 1192 |
| Corn Northern Leaf Blight | 1162 |
| Corn healthy | 985 |
| Grape Black rot | 1180 |
| Grape Esca (Black Measles) | 1383 |
| Grape Leaf blight(Isariopsis Leaf Spot) | 423 |
| Grape healthy | 1076 |
| Orange Haunglongbing (Citrus greening) | 5507 |
| Peach Bacterial spot | 2297 |
| Peach healthy | 360 |
| Bell pepper Bacterial spot | 997 |
| Bell pepper healthy | 1478 |
| Potato Early blight | 1000 |
| Potato Late blight | 152 |
| Potato healthy | 1000 |
| Raspberry healthy | 371 |
| Soybean healthy | 5090 |
| Squash Powdery mildew | 1835 |
| Strawberry Leaf scorch | 456 |
| Strawberry healthy | 1109 |
| Tomato Bacterial spot | 2127 |
| Tomato Early blight | 1000 |
| Tomato Late blight | 1591 |
| Tomato Leaf Mold | 1909 |
| Tomato Septoria leaf spot | 952 |
| Tomato Spider mites | 1771 |
| Tomato Target Spot | 1676 |
| Tomato Yellow Leaf Curl Virus | 1404 |
| Tomato mosaic virus | 373 |
| Tomato healthy | 5357 |

## PERFORMANCE EVALUATION METRICS

It's very important to find the efficiency of a system after its completion. To measure the performance of a machine learning algorithm, various performance metrics are used. A metric that is relevant to the algorithm should be chosen as its performance metric. The performance metrics have a large influence on the evaluation and the comparison of various machine learning algorithms. Some of the examples for the performance metrics are Precision and Recall.

#### Accuracy

Accuracy is the percentage of diseases correctly predicted in a given set of data. The number of correctly predicted outcomes is divided by the total number of predictions to obtain accuracy.

𝐴𝑐𝑐𝑢𝑟𝑎𝑐𝑦 = 𝑐𝑜𝑟𝑟𝑒𝑐𝑡 𝑝𝑟𝑒𝑑𝑖𝑐𝑡𝑖𝑜𝑛𝑠 × 100%

𝑡𝑜𝑡𝑎𝑙 𝑝𝑟𝑒𝑑𝑖𝑐𝑡𝑖𝑜𝑛𝑠

#### Precision

Precision is the fraction of diseases predicted correctly true positives from the total number of diseases that are predicted as correct (true positives +false positives).

𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛 = 𝑇𝑟𝑢𝑒 𝑃𝑜𝑠𝑖𝑡𝑖𝑣𝑒𝑠 × 100%

𝑇𝑟𝑢𝑒 𝑃𝑜𝑠𝑖𝑡𝑖𝑣𝑒𝑠 + 𝐹𝑎𝑙𝑠𝑒 𝑃𝑜𝑠𝑖𝑡𝑖𝑣𝑒𝑠

#### Recall

The recall is the fraction of diseases that are predicted correctly (true positive) from the total number of disease (true positives + false negatives) samples.

𝑅𝑒𝑐𝑎𝑙𝑙 = 𝑇𝑟𝑢𝑒 𝑝𝑜𝑠𝑖𝑡𝑖𝑣𝑒𝑠

𝑇𝑟𝑢𝑒 𝑝𝑜𝑠𝑖𝑡𝑖𝑣𝑒𝑠 + 𝐹𝑎𝑙𝑠𝑒 𝑛𝑒𝑔𝑎𝑡𝑖𝑣𝑒𝑠

× 100%

#### Confusion matrix

A confusion matrix is applied to illustrate the performance of a classifier in a table format. It is easy to understand and can be used on a classification model where the true values are known. There are four confusing terms used in the confusion matrix.

True Positive (TP): A leaf that has the disease and it's correctly predicted by the classifier is called True Positive TP).

True Negative (TN): A leaf that doesn't have the disease (healthy) and it's correctly predicted by the classifier is called True Negative (TN).

False Positive (FP): A leaf that doesn't have the disease (healthy) but it's incorrectly predicted as diseased by the classifier is called False Positive (FP).

False Negative (FN): A leaf that has the disease and it's incorrectly predicted as healthy by the classifier is called False Negative (FN).

## PERFORMANCE EVALUATION

#### TRAINING AND TESTING PHASE

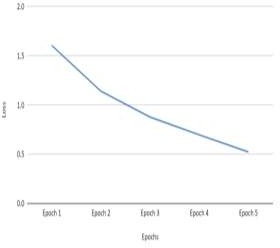
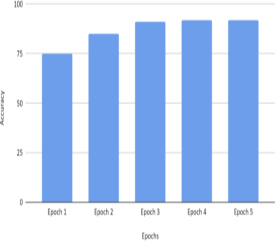
Training is the stage of machine learning. The model learns about the dataset it needs to classify through training. Training helps in optimizing the model. In this work, about 55,448 images that’s split in the ratio of 70:30(training: validation) is used for training. Accuracy is the number of images predicted correctly by the total number of images. Loss is the mean of squared differences between actual and predicted values. Testing helps us to find the accuracy by giving unseen images. 30% of images in our dataset are used for the testing or validation purpose. This helps us to identify how many new images the model can predict correctly. Performance of a fully trained model is obtained by giving the testing data collected. 39 disease categories are tested using various images to ensure the correctness of the model generated. Testing gives the model a clear confidence on what they are able to predict and the model can be trained again if they lack the desirable accuracy with more epochs. When the epoch increases, the model accuracy increases. If the accuracy is remaining constant then that’s the optimum accuracy that can be gained in this model from training.

**Table 6.2 The accuracy and loss values for each Table 6.3 The accuracy and loss values for epoch during training each epoch during validation**

|  |  |  |
| --- | --- | --- |
| Epoch | Accuracy | Loss values |
| Epoch 1 | 70 | 3.1 |
| Epoch 2 | 72 | 2.5 |
| Epoch 3 | 80 | 2.3 |
| Epoch 4 | 87 | 1.9 |
| Epoch 5 | 89 | 0.92 |

|  |  |  |
| --- | --- | --- |
| Epoch | Accuracy | Loss values |
| Epoch 1 | 75 | 1.605 |
| Epoch 2 | 85 | 1.14 |
| Epoch 3 | 91 | 0.877 |
| Epoch 4 | 92 | 0.697 |
| Epoch 5 | 92 | 0.523 |

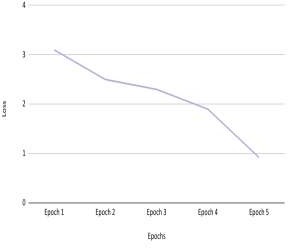
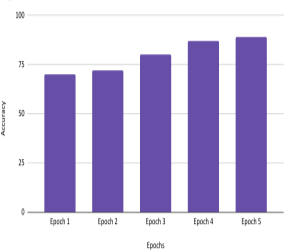
#### Epoch Vs accuracy and loss graph on training

During the training phase, the model is trained with 70% of the images in our dataset. The accuracy of the model is increasing with the number of iterations or epochs. This model attains an accuracy of 92% from 75% while training. Loss is the measure to indicate how badly the prediction happens in the model. In this project it’s found to be decreasing from 1.6 to 0.5.This indicates the increasing capability of the model to predict the disease correctly

**Fig 6.1 Epochs Vs Accuracy graph on training Fig 6.2 Epochs Vs Loss graph on training**

#### Epochs Vs accuracy and loss values in validation

During testing, the test images are given to the model and the results are analyzed to obtain the overall accuracy of the model. Here the accuracy increases from 70% to 89% during validation. The loss values are remainders of how far our model is growing by decreasing the amount of loss values.



**Fig 6.3 Epochs Vs Accuracy values in validation Fig 6.4 Epochs Vs Loss values in validation**

#### ANALYSIS OF VARIOUS LEAF DISEASE DETECTION

1. **APPLE**



#### Apple scab b. Black rot c. Cedar apple rust

In the apple plant, 3 main diseases like apple scab, black rot, and cedar apple rust is considered in our project. In apple scab, there are 630 images out of which 623 images are predicted correctly and 7 images are predicted incorrectly. An accuracy of 91% is obtained while predicting this disease. In black rot out of 621 images, 589 images are true positive and 32 images false positives. This gives an accuracy of 94%. In cedar apple rust, dataset contained 275 images. 258 images are classified correctly while 17 images are classified wrongly as healthy leaves. An accuracy of 92% is obtained while classifying cedar apple rust. When comparing the 3 diseases of apple tree high accuracy is obtained by black rot i.e., 94%.

#### CORN



* 1. **Corn Cercospora**

#### leaf spot

* 1. **Corn Common**

#### rust

* 1. **Corn Northern Leaf Blight**

In corn plant, main diseases taken are corn cercospora leaf spot, corn common rust, and corn northern leaf blight. Out of 513 images, 482 images are true positive and 31 images are false positive in corn cercospora leaf spot with an accuracy of 91%. In corn common rust, 94% accuracy is obtained with 1144 correct outcomes and 48 incorrect outcomes on 1192 images. Corn northern leaf blight has 1092 true positive values and 70 false positives. It provides an accuracy of 92%. So corn common rust has higher accuracy than the rest of the diseases predicted.

#### GRAPES



* 1. **Grape Black rot b. Grape Esca (Black Measles)**

#### c. Grape Leaf blight(Isariopsis Leaf Spot)

The diseases considered in grapes are black rot, esca and leaf blight. 1180 images are used to train the black rot and 1355 images are classified correctly and 36 images are incorrectly classified. An accuracy of 91% is found for black rot. 94% accuracy is obtained for the esca which has 1383images in which 1355 are true positive images and 28 false positive images. In grape leaf blight, out of 423 images, 397 are correct and 26 are incorrect. It gives an accuracy of 92%. Grape esca gives the highest accuracy.

#### POTATO



* 1. **Potato Early blight b. Potato Late blight**

In potato, the dataset contained early and late blight diseases. In potato early blight and late blight there are 1000 and 152 images respectively. Out of 1000 images in early blight, 910 images are true positive and 90 images are false positive giving an accuracy of 91%. In potato late blight, 141 images are correctly identified and 11 images are incorrect in 152 images. Potato late blight has higher accuracy of 92%.

#### TOMATO



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **a. Tomato** | **Bacterial** | **b.** | **Tomato** | **Early** | **c.** | **Tomato** | **Late** |
| **spot** |  |  | **blight** |  |  | **blight** |  |



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **d. Tomato Leaf Mold** | **e.** | **Tomato** | **Septoria** | **f.** | **Tomato** | **Spider** |
|  |  | **leaf spot** |  |  | **mites** |  |

#### Tomato Target Spot

1. **Tomato Yellow Leaf Curl Virus**

#### Tomato mosaic virus

In the tomato plant, 9 types of diseases like bacterial spot, early blight, late blight, leaf mold, septoria leaf spot, spider mites, target spot, yellow leaf curl virus and mosaic virus which have 2127, 1000, 1591, 1909, 952, 1771, 1676, 1404, 373 leaves in the dataset respectively. Tomato bacterial spot has 94% accuracy with 1914 true positive values and 213 false positive values. Early blight and spider mites have 92% accuracy. Tomato late blight has lowest accuracy of 87%. Target spot disease is predicted with 93% which is more accuracy than [25].

#### SUMMARY OF DISEASE DETECTION

**Table 6.4 The performance analysis on the dataset**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name of the plant** | **Total** | **TP** | **TN** | **FP** | **FN** | **Accuracy** | **Precision** | **Recall** |
| Apple scab | 630 | 623 | 510 | 7 | 106 | 91 | 0.99 | 0.85 |
| Black rot | 621 | 589 | 518 | 32 | 39 | 94 | 0.95 | 0.94 |
| Cedar apple rust | 275 | 258 | 216 | 17 | 25 | 92 | 0.94 | 0.91 |
| Apple healthy | 1645 | 1562 | 1343 | 83 | 136 | 93 | 0.95 | 0.92 |
| Background without  leaves | 1143 | 1051 | 861 | 92 | 98 | 91 | 0.92 | 0.91 |
| Blueberry healthy | 1502 | 1441 | 1123 | 61 | 257 | 89 | 0.96 | 0.85 |
| Cherry powdery  mildew | 854 | 811 | 681 | 43 | 87 | 92 | 0.95 | 0.90 |
| Cherry healthy | 1052 | 1009 | 867 | 43 | 99 | 93 | 0.96 | 0.91 |
| Corn cercospora leaf  spot | 513 | 48 | 395 | 31 | 56 | 91 | 0.94 | 0.90 |
| Corn common rust | 1192 | 1144 | 1006 | 48 | 90 | 94 | 0.96 | 0.93 |
| Corn northern leaf  blight | 1162 | 1092 | 917 | 70 | 105 | 92 | 0.94 | 0.91 |
| Corn healthy | 985 | 965 | 829 | 20 | 116 | 93 | 0.98 | 0.89 |
| Grape black rot | 1180 | 1144 | 938 | 36 | 170 | 91 | 0.97 | 0.87 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grape esca (black measles) | 1383 | 1355 | 1192 | 28 | 135 | 94 | 0.98 | 0.91 |
| Grape leaf  blight(Isariopsis leaf spot) | 423 | 397 | 333 | 26 | 38 | 92 | 0.94 | 0.91 |
| Grape healthy | 1076 | 1022 | 878 | 54 | 90 | 93 | 0.95 | 0.92 |
| Orange  haunglongbing(citrus greening) | 5507 | 5176 | 4244 | 331 | 601 | 91 | 0.94 | 0.90 |
| Peach bacterial spot | 2297 | 2113 | 1859 | 184 | 70 | 94 | 0.92 | 0.97 |
| Peach healthy | 360 | 338 | 283 | 22 | 33 | 92 | 0.94 | 0.91 |
| Bell pepper bacterial  spot | 997 | 947 | 814 | 50 | 83 | 93 | 0.95 | 0.92 |
| Bell pepper healthy | 1478 | 1359 | 1114 | 119 | 126 | 91 | 0.92 | .92 |
| Potato early blight | 1000 | 910 | 746 | 90 | 74 | 91 | 0.91 | 0.92 |
| Potato late blight | 152 | 141 | 118 | 11 | 12 | 92 | 0.93 | 0.92 |
| Potato healthy | 1000 | 940 | 808 | 60 | 72 | 93 | 0.94 | 0.93 |
| Raspberry healthy | 371 | 330 | 270 | 41 | 19 | 91 | 0.89 | 0.95 |
| Soybean healthy | 5090 | 4682 | 4120 | 408 | 154 | 94 | 0.92 | 0.97 |
| Squash powdery  mildew | 1835 | 1724 | 1448 | 111 | 165 | 92 | 0.94 | 0.91 |
| Strawberry leaf  scorch | 456 | 437 | 375 | 19 | 43 | 93 | 0.96 | 0.91 |
| Strawberry healthy | 1109 | 987 | 809 | 122 | 56 | 91 | 0.89 | 0.95 |
| Tomato bacterial  spot | 2127 | 1914 | 1684 | 213 | 17 | 94 | 0.9 | 0.99 |
| Tomato early blight | 1000 | 960 | 806 | 40 | 114 | 92 | 0.96 | 0.89 |
| Tomato late blight | 1591 | 1495 | 1106 | 96 | 293 | 87 | 0.94 | 0.84 |
| Tomato leaf mold | 1909 | 1813 | 1486 | 96 | 231 | 91 | 0.95 | 0.89 |
| Tomato septoria leaf spot | 952 | 923 | 812 | 29 | 82 | 94 | 0.97 | 0.92 |
| Tomato spider mites | 1771 | 1682 | 1412 | 89 | 181 | 92 | 0.95 | 0.90 |
| Tomato target spot | 1676 | 1541 | 1325 | 135 | 81 | 93 | 0.95 | 0.90 |
| Tomato yellow leaf  curl | 1404 | 1361 | 1116 | 43 | 202 | 91 | 0.97 | 0.87 |
| Tomato mosaic virus | 373 | 339 | 271 | 34 | 34 | 90 | 0.91 | 0.91 |
| Tomato healthy | 5357 | 5035 | 4229 | 322 | 484 | 92 | 0.94 | 0.91 |

#### PEST DETECTION ANALYSIS

Pest detection is done on a dataset with 1,070 images. The images are trained 50 epochs with batch size 16. It gains an accuracy of 82% on the pest dataset. It classifies the images on the basis of pests.

**Table 6.5 The confusion matrix for pest detection**

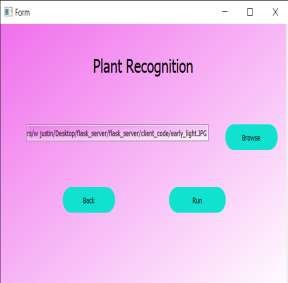
|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Predicted** | |
| **Total outcomes** | **Pest** | **Healthy** |
| **Actual** | **Pest** | 560(TP) | 70(FN) |
| **Healthy** | 436(FP) | 54(TN) |

## SCREENSHOTS OF USER INTERFACE

A user interface is constructed for the effective disease prediction. Both plant disease recognition and pest detection takes place in this interface. The name of the plant disease found is displayed on the screen while the pest will be caught in a box labeled pest.



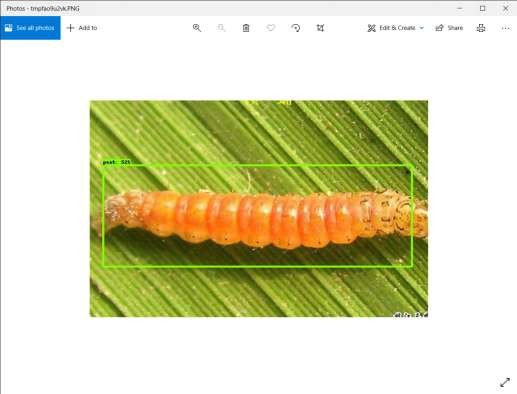
#### Welcome page of the web interface Web interface with plant recognition selected

#### Testing an image selected from the device The disease is predicted in the interface and run is clicked to view the result



**Pest recognition The image to search for the pest is selected from the device**



#### The pest is successfully found from the image

The images of 39 classes of diseases are successfully predicted. An accuracy of 92% is obtained by the ResNet 152 model for plant leaf disease detection. The parameters used to evaluate the accuracy are accuracy, precision, and recall. The accuracy can be increased by increasing the number of epochs. Pest detection is performed by the model MobileNet V2. An accuracy of 89% is obtained in the model. The confusion matrix is used to represent the accuracy of pest detection.

# CONCLUSION

In this project, we presented an interface for identifying plant diseases and pest detection using Deep learning techniques, ResNet 152 for farmers who have trouble in finding the plant diseases. Thus, it is very important in the agricultural sector. Plants are an important resource that is unavoidable in human life. It's quite natural for a plant to have a disease. Early detection of the diseases helps the farmers to have quality products. Initially, symptoms appear at the leaf. Automatic disease detection can be done with the help of symptoms from the leaves, which is easier and cheaper. The model is trained with an adequate amount of similar images so the chance of diagnosing the diseases increases. When the dataset increases, the authenticity of the model improves. Diseases occurring to a plant can be sorted based on the pathogen, the host, lifespan of the host, the name of the disease, plant part, indications, environment, causative agent, and topography. To estimate the disease; systems must be authentic, easy to understand, economical, and facilitate the prediction of a wide range of infections.

Along with plant diseases, pests are also a threat to farmers. There are a lot of pests attacking the plants without giving any visible symptoms. Pest detection is also as important as leaf disease detection. Both plant leaf disease detection and pest detection are, incorporated in a web interface.

# FUTURE SCOPE

Plant disease detection has incredible opportunities for future work. Forthcoming works with modern machine learning technologies like EfficientNet can increase efficiency. In the upcoming research, the labeled pest dataset can be collected from the affected fields. This helps to identify the pests attacking the plants efficiently. The nutritional deficiencies in the soil can lead to plant disorders. A dataset has to be collected accurately to predict the nutritional deficiencies in plants. The nutritional deficiencies along with the solutions can be a great help to the farmers. Along with the prediction, the instructions can be given to farmers to improve their crop productivity.

Global warming, pollution, and ozone layer depletion cause a temperature rise on Earth. Climate change has a great impact on plants and pests. The dataset containing the weather pattern can be used to plan the watering and nutritional supply to the plants. Overwatering and under watering are not good for crops. Overwatering leads to the rotting of roots while under watering can kill a plant. So plants have to be properly watered for productivity. Training data plays a significant role in disease prediction. Training the model with more pictures can increase the accuracy.

The following issues can be addressed for future works.

* Device to detect diseases efficiently in real-time.
* Efficient detection of nutritional deficiency-based symptoms.
* Weather recognition to water the plants smartly.
* Training with a large dataset for accurate prediction.

# REFERENCES

1. Ahasan Ullah Khan, M., Choudhury, A.R., Abdul, M., Maleque, C.K.D., Talucder, M.S.A., Maukeeb, A.R.M., Ema, I.J., Adnan, M.: Management of insect pests and diseases of jackfruit (artocarpus heterophyllus l.) in agroforestry system: A review. Acta Entomology and Zoology 2(1), 3 -4 (2021)
2. Ai, Y., Sun, C., Tie, J., Cai, X.: Research on recognition model of crop diseases and insect pests based on deep learning in harsh environments. IEEE Access 8, 1 1 8 -1 1 93 (2020)
3. Alpaydin, E.: Introduction to Machine Learning. MIT Press, US (2004)
4. Atila, Ü., Uçar, M., Akyol, K., Uçar, E.: Plant leaf disease classification using efficient net deep learning model. Ecological Informatics 61, 101182 (2021)
5. Daniya, T., Vigneshwari, S.: A review on machine learning techniques for rice plant disease detection in agricultural research. system 28(13), 49- 2 (2019)
6. (Food, F., of the United Nations), A.O.: Plant pest surveillance: A guide to understand the principal requirements of surveillance programmes for national plant protection organizations. FAO, Rome, Italy (2016)
7. Geetharamani, G., Pandian, A.: Identification of plant leaf diseases using a nine-layer deep convolutional neural network.Computers& Electrical Engineering 76, 323-338 (2019)
8. Gokulnath, B., et al.: Identifying and classifying plant disease using resilient lf-CNN. Ecological Informatics 63, 101283 (2021)
9. Govardhan, M., Veena, M.: Diagnosis of tomato plant diseases using random forest (2019)
10. Haider, W., Rehman, A.U., Durrani, N.M., Rehman, S.U.: A generic approach for wheat disease classification and verification using expert opinion for knowledge-based decisions. IEEE Access 9, 31104-31129 (2021)
11. Hernandez, S., Lopez, J.L.: Uncertainty quantification for plant disease detection using bayesian deep learning. Applied Soft Computing 96, 10 59 (2020)
12. Jadhav, S.B., Udupi, V.R., Patil, S.B.: Identification of plant diseases using convolutional neural networks. International Journal of Information Technology pp. 1- 10 (2020)
13. Kusumo, B.S., Heryana, A., Mahendra, O., Pardede, H.F.: Machine learning-based for automatic detection of corn- plant diseases using image processing. In: 2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA). pp. 93-9. IEEE (2018)
14. Monalisa Saha, E.S.: Identification of plant leaf diseases using machine learning algorithms. International Journal of Advanced Science and Technology 29, 2900 - 2910 (2020)
15. Naga Swetha R, V.S.: Monitoring of rice plants for disease detection using machine learning. International Journal of Engineering and Advanced Technology 9 (2020)
16. Panchal, P., Raman, V.C., Mantri, S.: Plant diseases detection and classification using machine learning models. vol. 4, pp. 1- (2019)
17. Prajapati, J., et al.: Balanced plant nutrition manual for crop. International Journal Of All Research Writings 1(1), 1- (2019)
18. Ramesh, S., Vydeki, D.: Application of machine learning in detection of blast disease in south indian rice crops. J. Phytol 11(1), 31-3 (2019)
19. Selvaraj, M.G., Vergara, A., Montenegro, F., Ruiz, H.A., Safari, N., Raymaekers, D., Ocimati, W., Ntamwira, J., Tits, L., Omondi, A.B., et al.: Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in drcongo and republic of benin. ISPRS Journal of Photogrammetry and Remote Sensing 169, 110-124 (2020)
20. Singh, A., Kaur, H.: Potato plant leaves disease detection and classification using machine learning methodologies. In: IOP Conference Series: Materials Science and Engineering. vol. 1022, p. 012121. IOP Publishing (2021)
21. Sujatha, R., Chatterjee, J.M., Jhanjhi, N., Brohi, S.N.: Performance of deep learning vs machine learning in plant leaf disease detection. Microprocessors and Microsystems 80, 103 15 (2021)
22. Tambo, J.A., Romney, D., Mugambi, I., Mbugua, F., Bundi, M., Uzayisenga, B., Matimelo, M., Ndhlovu, M.: Can plant clinics enhance judicious use of pesticides? evidence from Rwanda and zambia. Food Policy p. 1020 3 (2021)
23. Thangaraj, R., Anandamurugan, S., Kaliappan, V.K.: Automated tomato leaf disease classification using transfer learning-based deep convolution neural network. Journal of Plant Diseases and Protection 128(1), 3-8 (2021)
24. Vishnoi, V.K., Kumar, K., Kumar, B.: Plant disease detection using computational intelligence and image processing. Journal of Plant Diseases and Protection pp. 1-35 (2020)
25. Wspanialy, P., Moussa, M.: A detection and severity estimation system for generic diseases of tomato greenhouse plants. Computers and Electronics in Agriculture 178, 105 01 (2020)
26. Liu, Jun, and Xuewei Wang. "Plant diseases and pests detection based on deep learning: a review." *Plant Methods* 17.1 (2021): 1-18.
27. <https://ai.googleblog.com/2018/04/mobilenetv2-next-generation-of-on.html?m=1>
28. <https://neurohive.io/en/popular-networks/resnet/>
29. <https://colab.research.google.com/>
30. https://towardsdatascience.com/review-resnet-winner-of-ilsvrc-2015-image- classification-localization-detection-e39402bfa5d8