Plant Disease Recognition Using Machine Learning

Minor Project Report

Prepared by

Specialization	Name	SAP ID
B. Tech CSE BAO	Shivam Gupta	500083281
B. Tech CSE BAO	Shrey Gilotra	500083687
B. Tech CSE BAO	Hiya Chopra	500083441
B. Tech CSE AIML	Shivam Rathi	500084169

Department of Informatics

School Of Computer Science

UNIVERSITY OF PETROLEUM & ENERGY STUDIES,
DEHRADUN- 248007. Uttarakhand

Dr. TP Singh
Project Guide

Dr. TP Singh
Cluster Head

Project Title: - Plant Disease Recognition Using Machine Learning

Abstract

Plant diseases pose a significant threat to agricultural productivity and food security. Early detection and accurate identification of plant diseases are crucial for effective disease management and mitigation. In recent years, deep learning models have shown promising results in various computer vision tasks, including plant disease recognition. In this study, we explore the effectiveness of four widely used deep learning models: ResNet34, ResNet50, CNN, and VGG16, for plant disease recognition. We leverage a comprehensive dataset comprising images of healthy plants and plants affected by various diseases. Pretrained versions of ResNet34, ResNet50, CNN, and VGG16, which were trained on large-scale image datasets, are fine-tuned using transfer learning techniques. The fine-tuning process involves retraining the models on our plant disease dataset to adapt their learned features for disease recognition. We evaluate the performance of the models based on several metrics, including accuracy, precision, recall, and F1 score. Comparative analysis is conducted to identify the strengths and weaknesses of each model in terms of their ability to correctly classify plant diseases. Additionally, we investigate the effects of model architecture depth and complexity on the recognition performance. Our experimental results demonstrate that all four models achieve high accuracy in classifying plant diseases, with varying degrees of success. While ResNet34 and CNN exhibit good overall performance, ResNet50 and VGG16 demonstrate superior capabilities in capturing fine-grained details and subtle disease symptoms. However, the deeper architectures of ResNet50 and VGG16 require more computational resources and training time. The findings of this study contribute to the development of automated plant disease recognition systems. They provide insights into the selection and optimization of deep learning models for accurate and efficient disease diagnosis in plants. The proposed models can be integrated into real-time monitoring systems, empowering farmers and plant disease experts to detect and respond to diseases promptly, minimizing crop losses and ensuring food security.

Keywords:- Convulutional Neural Network, machine learning, Plant Disease, VGG16, ResNet34, ResNet50, Image Classification

Acknowledgment

We would like to convey our sincere gratitude to our mentor, Dr. TP Singh, for all the guidance, inspiration, and unwavering support he provided us with throughout the course of working on our project. Without his encouragement and helpful recommendations, this effort would not have been feasible.

We really appreciate Dr. TP Singh, the cluster chair, for all his help with our study at SOCS.

Additionally, we are grateful to our kind Dean of SOCS, UPES Dr. Ravi S. Iyer for providing us with the tools we needed to complete our project work effectively.

We appreciate the assistance and helpful feedback provided by all UPES faculty members during the course of our project work. Finally, we can only thank our parents in the deepest sense for showing us the world and for all of the support they have given us.

List of figures

Fig No.	Figure Description	Page no.
1	Model Flow Chart	10
2	UML diagram	11
3	Pert Chart	13
4	Web Application Result	14
5	Input Image	15
6	CNN Accuracy and Loss	15
7	ResNet 34 Accuracy and Loss	16
8	ResNet 50 Results	17
9	Comparison Between Different Models	18

Table of Contents

Topic		Page No
Table of Content		
1	Introduction	6
	1.1 Purpose of the Project	6
	1.2 Target Beneficiary	6
	1.3 Project Scope	6
2	Literature Review	7
3	Problem statement	7
4	Existing system issue	7
5	Project Description	8
	5.1 Reference Algorithm	9
	5.2 Characteristics of data	9
	5.3 SWOT analysis	10
	5.4 Project Features	11
	5.5 User Classes and Characteristics	11
6	PERT chart	13
7	Results	14
8	Conclusion	18
9	Future work	18
	References	19
	AppendixA: Glossary	20
	AppendixB: Analysis model	20
	Appendix C: Issues list	21

1. Introduction

Plant diseases can seriously harm the crops around us, and farmers who use inaccurate methods to detect diseases reduce crop yield and incur greater financial losses. Therefore, early disease detection is essential for efficient disease management. The need for effective and precise methods for detecting plant diseases has become more pressing due to the rising demand for agricultural production and the threat of climate change. In this study, we train machine learning models to recognize and detect patterns in plant images as well as to pinpoint their symptoms. We also have studied different algorithms to learn various image processing techniques. Here, it is a comparative study between the algorithms like ResNet 34, ResNet 50, and VGG 16 which gives us the answer to why we used the CNN approach to make a better use of Image processing technology. This involves extracting features from the images, training machine learning models to recognize patterns, and using the learned patterns to classify and predict new cures and diseases.

1.1 Purpose of the project

The purpose of this project is to use Machine Learning for plant disease recognition to provide farmers and agriculture experts with an efficient and accurate tool to detect the disease associated at an early stage. Also, Machine Learning algorithms can analyze large amounts of data and identify patterns that are difficult for humans to detect, making it possible to automate the detection and diagnosis. The use of machine learning algorithms is when we need to generalize and detect plant diseases. The technology can also help to mitigate the shortage of agricultural experts in certain parts of the world, making it possible to benefit more farmers in the world

1.2 Target Beneficiary

The target of detection of plant diseases is mainly farmers, agricultural experts, home farmers, and other people interested in plant cultivation and care. Farmers will benefit significantly from this technology as it provides them with an efficient and accurate way to detect plant diseases at an early stage. Agricultural professionals, including researchers and consultants, will also benefit from this technology, providing them with a tool to support their work in crop management and disease control. In addition to farmers and agricultural experts, other stakeholders involved in crop production and maintenance, such as decision-makers, input suppliers, and processors, will also benefit from this technology, reducing the economic losses and environmental impacts caused by plant diseases.

1.3 Project Scope

The scope of the machine learning detection of plant diseases project can be divided into the following phases: Data collection: The first step of the project is to collect an extensive database of images of plants with known diseases. This dataset must cover a wide range of plant species and diseases to ensure model accuracy and reliability. Data pre-processing: Collected data must be pre-processed to remove noise or irrelevant information that may affect the accuracy of the model. This includes image enhancement, feature extraction, and data normalization. Also, various image processing techniques are used to develop image classification models but we are using neural network technique. Model training and testing: The pre-processed data is then used to train a machine learning model using algorithms such as Convolutional Neural Networks (CNN), VGG16, ResNet 34, and ResNet 50. The trained model should be tested on a separate dataset of plant images with known diseases to ensure its accuracy and generalizability. Comparative study: All the four machine learning models used will compared on the basis on their accuracy and loss rate. Implementation of the model: Once the model is tested and validated, it should be implemented in a web-based platform or mobile application for use by farmers. The platform must be user-friendly and easily accessible to farmers in remote areas.

2. Literature review

P.Chaitanya Reddy et al [1] stated that improving the technical and mechanized support will increase agriculture productivity. They employed various Machine learning algorithms like SVM Classification for the detection of leaf diseases. Various Performance metrics like Root mean square Error (RMSE), Peak Signal Noise Ratio are compared to benefit the farmers with less time and resources.

Thakur et al. [2] used two pre-trained layers from VGG16 and Inception v& layers to train their CNN model. Their accuracy was 99.16%,93.66%,94.24%,91.36% and 96.67% respectively on five publicly available datasets.

In this work, a transfer learning-based CNN was constructed for detecting plant diseases. To train the model, we mainly used ResNet50 [3], a popular CNN architecture. The proposed model performed best with a training accuracy of 99.80%

3. Problem statement

The problem statement for plant disease recognition using machine learning is the need for a reliable and efficient method for early detection and diagnosis of plant diseases. Plant diseases can cause significant damage to crop, resulting in lower yields, economic losses, and food insecurity. Traditionally, farmers rely on visual inspection to detect diseases, which can be time-consuming and inaccurate. The goal is to develop a mobile application that can be used by farmers, researchers, and other people in the agriculture industry so that they can accurately identify and treat diseased plants.

4. Existing system issue

Based on the literature review done for this project the major issues in the existing projects are:-

• As we are considering a very large dataset, handling so many images is very difficult.

• Due to the size of the dataset, it is very time-consuming to run each and every epoch, which also adds load on the system.

5. Project Description

5.1 Reference Algorithm

The algorithms which are taken into account for this project are:-

Convolutional Neural Networks (CNNs) are particularly well-suited for plant disease recognition because they excel at processing images and extracting relevant features from them.

1. CNNs can be trained on large datasets of images of plants with and without diseases, allowing them to learn to identify subtle patterns and features that are indicative of different diseases. We are using 14 classes of 5 unique plants.

2. In this project we are implementing 10 layers in the Convolutional Neural network, 5 layers in Max Pooling, and 3 layers in the Flattening layer.

3. Along with this, the activation function used here is Rectified Linear Unit (RELU).

4. The Matrix used in MAX POOL is of 2*2 with a Kernel of 3, Stride of 1, and padding of 1.

5. It gives better accuracy than other models, which is up to 98.6% with a 7 % Validation Loss.

VGG 16:

VGG 16 is a type of CNN that is used for object detection and as a classification algorithm.

VGG is 16 layers deep and there are 13 convolutional layers, 5 max-pooling layers, and three dense layers.

It gives an accuracy of up to 97.9% with a 14 % Validation Loss.

In VGG, we are having Convolution layers of a 3x3 filter with stride 1 with the same padding.

ResNet 34 : • ResNet34 Architecture is designed to address the problem of vanishing gradients in Deep Neural networks.

• ResNet34 has 34 layers and has been used for various vision tasks like Object detection, Image classification, and semantic segmentation.

• In this, the residual blocks have a filter size of 3x3, and the number of filters is gradually increased in the deeper layers.

• Each residual block contains two convolutional layers with the same number of filters and a shortcut connection that adds the input to the output of the block. This allows the network to learn residual mappings that facilitate the propagation of gradients through the network.

• The network also includes global average pooling and fully connected layer at the end for classification.

5.2 Characteristics of data

- The dataset comprises 32516 images, comprising 25666 images for training and 6850 images for validation
-
- The plant images span the following 5 unique species of Plant: Apple, Grape, Cherry, Peach, Strawberry.
-
- The dataset contains a total of 38 classes, like: Apple Scab, Apple Black Rot, Apple Cedar Rust, Apple healthy, Cherry healthy, Cherry Powdery Mildew, Grape Black Rot, Grape Black Measles, Grape Leaf Blight, Grape healthy, Peach Bacterial Spot, Peach healthy, Strawberry Healthy, Strawberry Leaf Scorch.
-
- The dataset also consists of an additional class background to differentiate between leaves and its background features.

5.3 SWOT Analysis

Strengths -

● Accessibility: The technology can be implemented through smartphones, making it easily accessible to farmers in remote areas.

● Scalability: Once a machine learning model is trained, it can be easily scaled to analyze large amounts of data from multiple farms and regions.

● Efficiency: Detection of plant diseases based on machine learning is faster and more efficient than traditional manual diagnostic methods, reducing time and labor costs.

● Accuracy: Machine learning algorithms can accurately classify and diagnose plant diseases, resulting in better crop management and increased yields.

Weaknesses -

● Complexity: Developing a machine learning model for plant disease detection requires specialized knowledge and expertise in computer science and agricultural science.

● Data dependency: Machine learning algorithms require large amounts of data to train, and model accuracy depends on the quality and diversity of the data set.

Opportunities -

● Increased productivity: Accurate and timely disease diagnosis can increase yields and improve food safety.

● Sustainability: Technology can reduce pesticide use by enabling targeted and timely use, improving environmental sustainability.

● disease control, reduce disease spread and minimize crop losses.

Threats-

● Lack of awareness: Many farmers may not be aware of the technology and its benefits, which may limit its adoption and impact.

● Cost: The cost of developing and deploying machine learning-based plant disease detection systems can be a barrier, especially for small farmers.

5.4 Project features

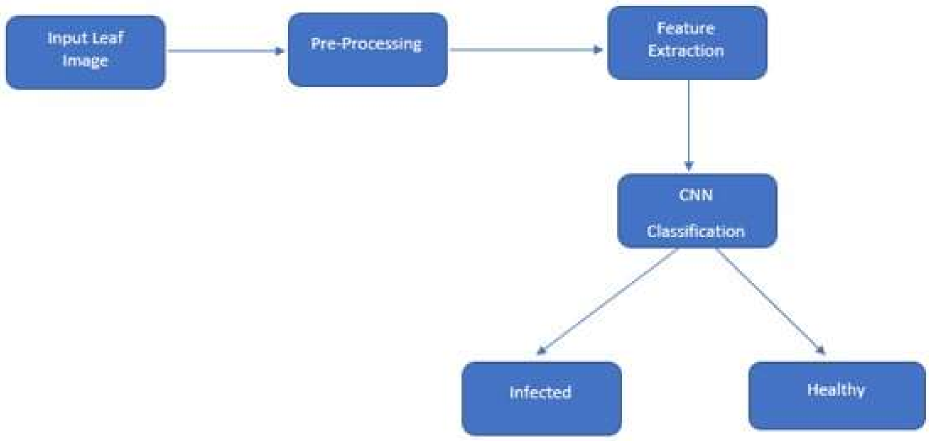


Fig 1. Model FlowChart

The project mainly focuses on features like Speed, Accessibility, Versatility, User-friendly, and Cost-effectiveness. Other than this, we have applied the approach as followsThe project shows a comparative study between 4 different machine learning models, being CNN, VGG16, ResNet 34, and ResNet 50.

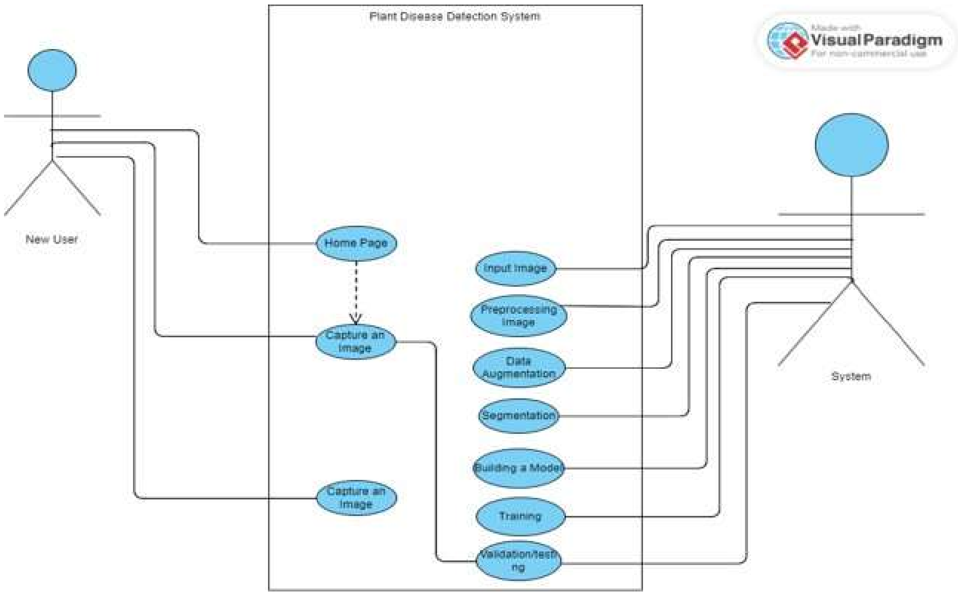


Fig 2. UML diagram

Step 1: Collecting the dataset The first step in our project is to collect image data of various plant leaves. We are using Plant Village Data which is available on Kaggle

Step 2: Preprocessing and Augmentation: Pre-Processing and Augmentation are done on the collected dataset by using Keras. We will improve the quality and quantity of images using Data Augmentation by rotating and changing the brightness of the image to help the model train faster. The total number of images in the dataset is 32516.

Step 3: Implementing machine learning models We are using 4 different models in order to find the most accurate one. We have used 25666 images for training and 6850 images for validation which is an 80:20 ratio

5.5 User Classes and Characteristics

Potential user classes that would benefit from this project will be -

1. Farmers:

- Domain knowledge: Farmers have practical knowledge about plant diseases and their symptoms.

- Limited technical expertise: Farmers may not have extensive knowledge of machine learning or computer vision techniques.

- Need for on-site solutions: Farmers require user-friendly systems that can be easily deployed and used in their fields.

2. Agricultural researchers:

- Technical expertise: Researchers have a deeper understanding of machine learning algorithms and computer vision techniques.

- Data annotation capabilities: Researchers are capable of annotating large datasets for training and evaluation.

- Need for accurate and interpretable results: Researchers often require detailed insights into the disease identification process for further analysis.

3. Agricultural extension officers:

- Communication skills: Extension officers act as intermediaries between farmers and researchers, and they need to effectively communicate disease-related information to farmers.

- Familiarity with field conditions: Extension officers are well-versed in the practical challenges faced by farmers in the field.

- Need for real-time updates: Extension officers' benefit from timely and accurate disease identification to provide relevant recommendations.

4. Plant disease experts:

- Deep knowledge of plant diseases: Experts possess extensive knowledge of various plant diseases, including their symptoms, causes, and treatments.

- Ability to validate and refine models: Experts can evaluate the performance of disease recognition models and provide insights to improve their accuracy.

- Need for advanced analysis: Plant disease experts may require additional information beyond basic disease recognition, such as disease severity estimation or pathogen identification.

5. Software developers:

- Programming skills: Developers possess the technical expertise to implement and optimize machine learning models.

- Integration capabilities: Developers can integrate disease recognition models into user-friendly software interfaces or mobile applications.

- Scalability and efficiency requirements: Developers focus on building systems that can handle large-scale datasets, process images efficiently, and provide real-time responses.

6. PERT Chart

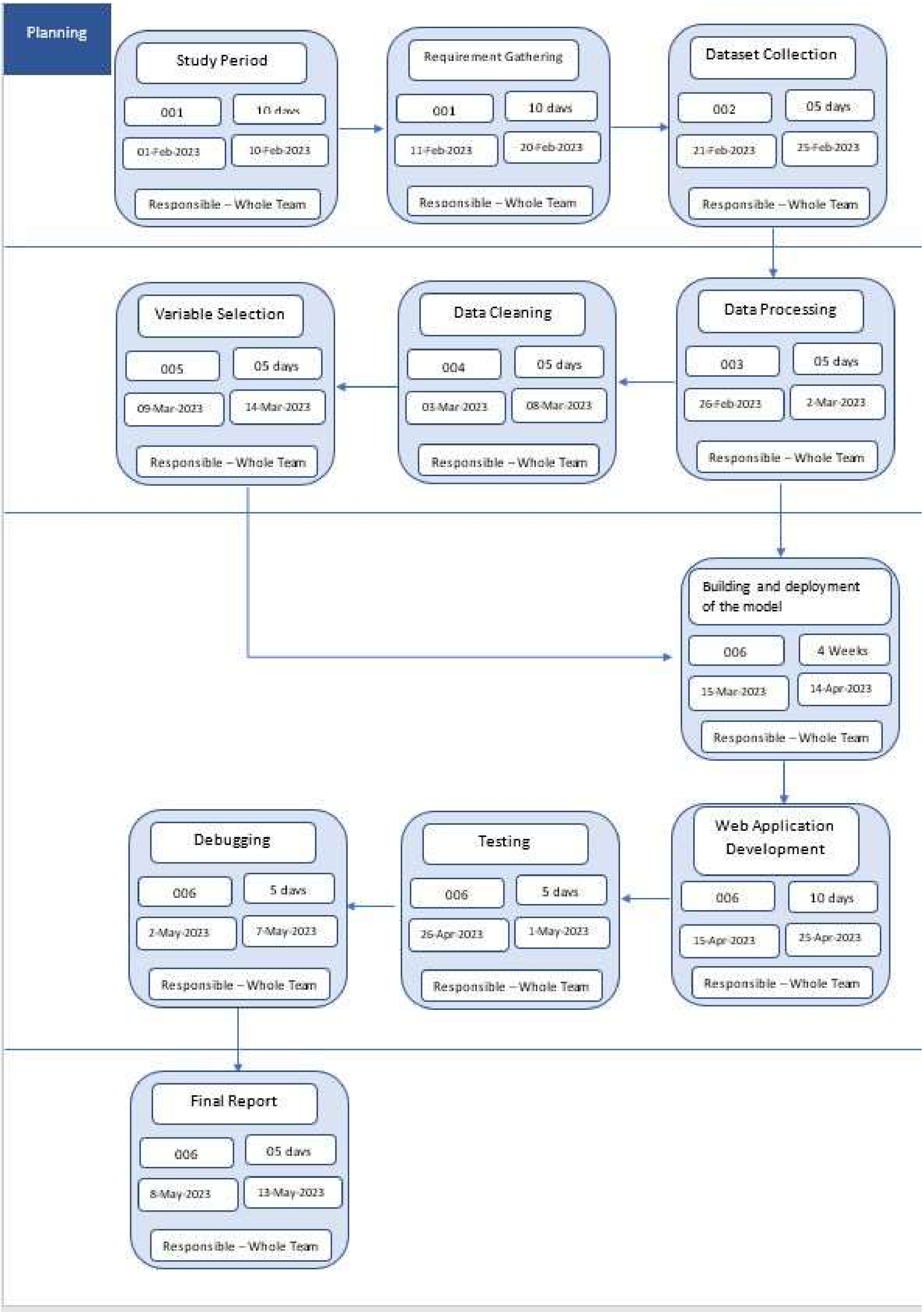


Fig 3. PERT chart

7. Results

The following disease prediction model gives detailed information about the cause of the disease the plant is having.

It provides you with the disease name, and cause of disease.

Also, it provides you with the cure which must be followed in order to get rid of the issue you are facing.

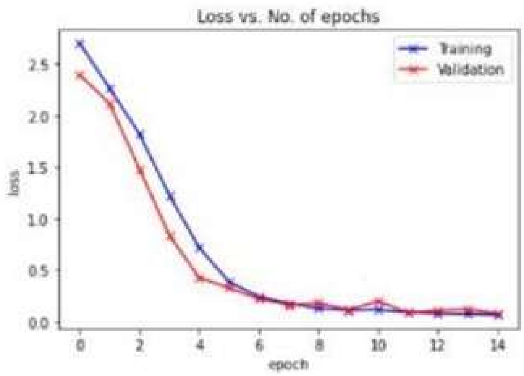
The following was our inference of the model :

- Convolutional Neural Network: 98.6% test accuracy
- VGG16: 97.9% test accuracy
- ResNet34: 98.4% test accuracy
- ResNet50: 99.3% test accuracy
- Vision Transformers: 98.7% test accuracy

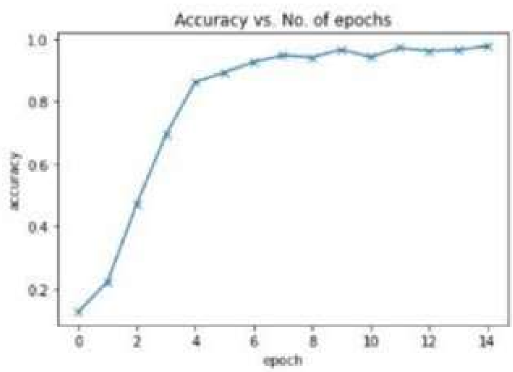
Fig 4.Web application result

Fig 5. Input of an image

Convolutional Neural Network (CNN)



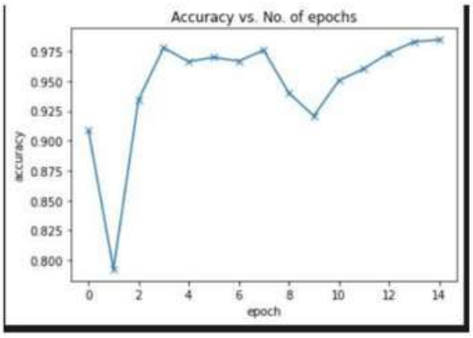
Loss V/S No. of Epoch



Accuracy V/S No. of Epoch

Fig 6. CNN Accuracy and Loss graph

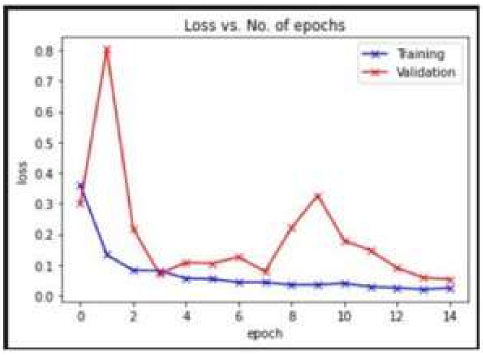
14



Accuracy V/S No. of Epoch

Fig 6. ResNet Accuracy

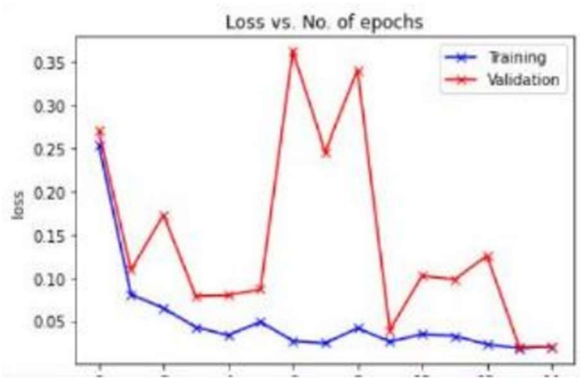
ResNet34



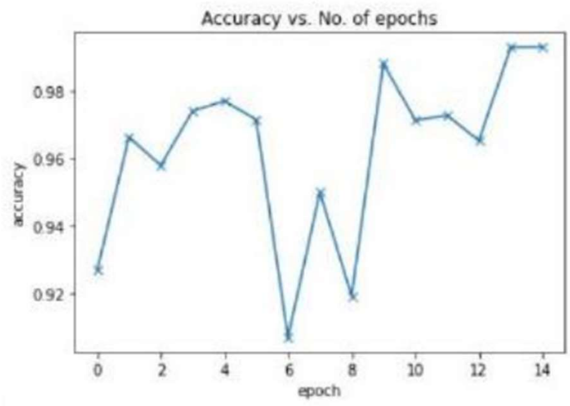
Loss V/S No. of Epoch

Fig 7. ResNet Loss graph

ResNet50



Loss V/S No. of Epoch



Accuracy V/S No. of Epoch

Fig 8: ResNet 50 Results

Models	Accuracy	Loss
Convolutional Neural Network	98.6%	7%
VGG16	97.9%	14%
ResNet34	98.4%	5%
ResNet50	99.3%	2%

Fig 7. Comparison between Different Models

8. Conclusion

• In this project we have integrated 4 Machine learning models CNN , VGG16 , RESNET34 , RESNET 50 to build a recommender system to control and identify the disease of a plant.

• We have also generated a simple web-based interface using Flask for the easy accessibility and cost-effective means of monitoring crops for diseases and their cures.

9. Future work

Artificial intelligence and machine learning: The effectiveness and efficiency of plant disease recognition systems can be increased by the application of artificial intelligence and machine learning algorithms.

With the aid of these technologies, the system may learn from a sizable dataset of plant diseases and their symptoms, simplifying the accurate diagnosis and treatment of diseases.

Internet of Things (IoT): By integrating IoT devices, such as cameras and sensors, it is possible to track the development of diseases and track the health of plants in real-time. This can assist growers and farmers in taking proactive steps to stop the spread of illnesses. Smartphone applications can give farmers and producers an easy-to-use interface for identifying and treating plant diseases. Based on the crop and area, these apps can also offer tailored advice.

References

[1]. P. Chaitanya Reddy, R. M. S. Chandra, P. Vadiraj, M. Ayyappa Reddy, T. R. Maheshand G. Sindhu Madhuri, "Detection of Plant Leaf-based Diseases Using Machine Learning Approach," 2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bangalore, India, 2021, pp. 1-4, doi: 10.1109/CSITSS54238.2021.9683020.

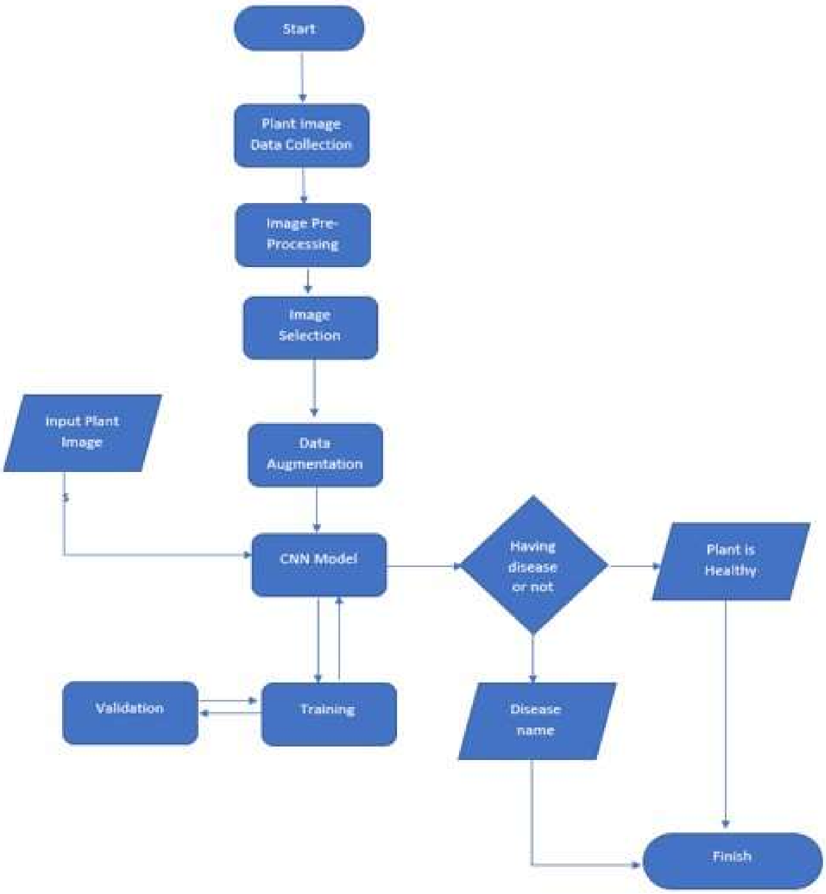
[2] Thakur, P. S., Sheorey, T., & Ojha, A. (2022). VGG-ICNN: A Lightweight CNN model for crop disease identification. Multimedia Tools and Applications, 1-24.

[3] I. Z. Mukti and D. Biswas, "Transfer Learning Based Plant Diseases Detection Using ResNet50," 2019 4th International Conference on Electrical Information and Communication Technology (EICT), 2019, pp. 1-6, doi: 10.1109/EICT48899.2019.9068805.

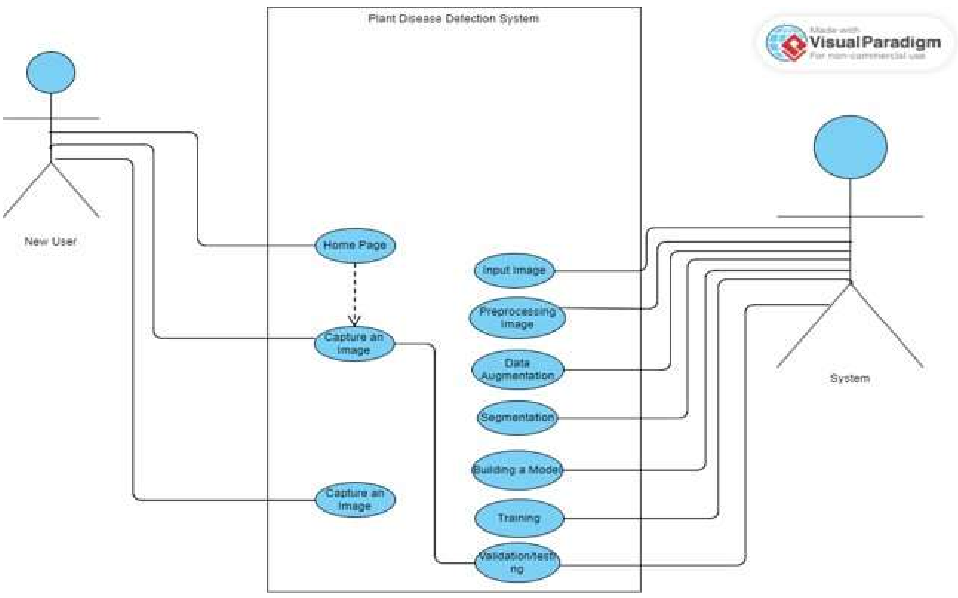
Appendix A: Glossary

Appendix B: Analysis Model

- Flow chart


Start
Plant image Data Collection
Image Pre-Processing
Image Selection
Input Plant Image
Data Augmentation
CNN Model
Having disease or not
Plant is Healthy
Validation
Training
Disease name
Finish

● UML


Plant Disease Detection System
Made with Visual Paradigm For non-commercial use
New User
Home Page
Capture an Image
Capture an Image
Input Image
Preprocessing Image
Data Augmentation
Segmentation
Building a Model
Training
Validation/testing
System

Appendix C: Issue List

- Data privacy
- API Limitations