A PROJECT BASED LEARNING REPORT
ON
“PLANT LEAF DISEASE PREDICTION SYSTEM”

Submitted to
SAVITRIBAI PHULE PUNE UNIVERSITY

In Partial Fulfilment of the Requirement for the Award of

MASTER OF COMPUTER APPLICATION(Under Engineering)

BY

Aniket Pravin Sarjine 0703

UNDER THE GUIDANCE OF
Prof.Monika Shinde

DEPARTMENT OF MASTER OF COMPUTER APPLICATION
TRINITY ACADEMY OF ENGINEERING
Kondhwa Annex, Pune - 411048
2024-2025

TRINITY ACADEMY OF ENGINEERING

Department of Master of Computer Application

CERTIFICATE

This is certify that the Project Based Learning entitled

"PLANT LEAF DISEASE PREDICTION SYSTEM"

submitted by

Aniket Sarjine

This is to certify that **Aniket Sarjine** has successfully submitted Project Based Learning entitled " **"PLANT LEAF DISEASE PREDICTION SYSTEM""** under the guidance of " **Prof.Monika Shinde**" in the Academic Year 2024-25 at Master of Computer Application of Trinity Academy of Engineering , under the Savitribai Phule Pune University. This Project Based Learning work is duly completed.

Date: / / 2025
Place: Pune

Prof.Monika Shinde **PBL Guide**	**Dr. A. A. Bhusari** **HOD MCA**	**Dr. R. J. Patil** **Principal**

Internal Examiner **External Examiner**

Acknowledgement

I would like to acknowledge all the teacher and friends who ever helped and assisted me throughout my Project Based Learning work.

First of all I would like to thank my respected guide **Prof.Monika Shinde**, Introducing me throughout features needed. The time-to-time guidance, encouragement and valuable suggestion received from him are unforgettable in my life. This work would not have been possible without the enthusiastic response, insight and new idea from him.

Furthermore, I would like to thank respected **Dr. R. J. Patil**, Principal and **Dr. A. A. Bhusari**, Head of Department of Master of Computer Application for the provided by him during my **Project Based Learning** work. I am also grateful to all the faculty members of Trinity Academy of Engineering, Pune for their support and cooperation. I would like to thank my lovely parent for time-to-time support and encouragement and valuable suggestion, and I would specify like to thank all my friends for their valuable suggestion and support. The acknowledgement world be incomplete without mention of the blessing of the almighty, which helped me in keeping high moral during difficult period.

Aniket Pravin Sarjine
Department of MCA

Declaration by the candidate

I hereby declare that this project report titled **"PLANT LEAF DISEASE PREDICTION SYSTEM"** submitted towards partial fulfillment of requirements for the degree of MCA is an authentic record of my work carried out under the guidance of **Prof.Monika Shinde.**

I further declare that the material obtained from other resources is duly acknowledged in this report.

Date: / /2025
Place: Pune

Aniket Pravin Sarjine
Department of MCA

Abstract

Plants are very essential in our life as they provide source of energy and overcome the issue of global warming. Plants now a days are affected by diseases like bacterial spot, late blight, Septoria leaf spot. These diseases effect the efficiency of crop yield. So ,the early detection of diseases is important in agriculture. Detection of diseases as soon as they appear is vital step for effective disease management. Aim of the project is to detect plant leaf disease by Machine Learning using image and videos. For Image, the proposed algorithm is Random forest classifier-Machine learning Algorithm used for classification andfor video the proposed technique is Resnet50- Deep Learning Algorithm. These techniques will obtain prediction results using various metrics like accuracy, precision and efficiency. This project can be implemented in agriculture, nursery, college gardens etc.

Plant diseases are challenging to monitor manually as it requires a great deal of work, expertise on plant diseases, and excessive processing time. Hence, this can be achieved by utilizing image processing techniques for plant disease detection. These techniques include image acquisition, image filtering, segmentation, feature extraction, and classification. Convolutional Neural Network's(CNN) are the state of the art in image recognition and have the ability to give prompt and definitive diagnoses. We trained a deep convolutional neural network using 70295 images on 38 folders of diseased and healthy plant leaves. This project aims to develop an optimal and more accurate method for detecting diseases of plants by analysing leaf images.

Keywords: - *Machine Learning,CNN,Classification.Random Forest Algorithm*

Contents

List of Figures

List of Abbreviation

CNN - Convolutional Neural Network

SVM - Support Vector Machine

KNN - K-Nearest Neighbhor

NLP - Natural Language Processing

API - Application Programming Interface

UI - User Interface

HTTPS - Hypertext Transfer Protocol Secure

DCNN - Deep Convolution Neural Network

ML - Machine Learning.

AI - Artificial Intelligence

DBSCAN - Density-based spatial clustering of applications with noise

RF - Random Forest

LR - Logistic Regression

JPEG - Joint Photographic Experts Group

bmp - Bitmap

WSGI - Web Server Gateway Interface

1 About Project

1.1 Title

The title of our project is **"PLANT LEAF DISEASE PREDICTION SYSTEM"**

1.2 Domain

Agriculture

1.3 Aim

when plants are consistently disturbed, they might catch illness. pathogens that result in abnormal physiological processes that disrupt the normal structure, growth, function, or other activities of the plant. The essential physiological or biochemical processes of the plant are disrupted, which results in the typical diseased states or symptoms. Depending on whether the primary cause of the disease is infectious or noninfectious, plant diseases may be broadly categorised. Infectious plant diseases are caused by pathogens such as fungi, bacteria, mycoplasma, viruses, viroids, nematodes, or parasitic flowering plants. Within or on a host, infectious organisms can grow and spread from one vulnerable host to another. Unfavorable growth circumstances, including as excessive temperatures, unfavorable moisture-oxygen ratios, soil and air pollutants, and an abundance or shortage of vital minerals, are the root causes of non-infectious plant illnesses.

1.4 Objective

The main objective of this study is to develop an automated system for detecting and classifying plant leaf diseases using CNN. Specifically, this study aims to:

- Develop a CNN model that can accurately detect and classify common plant leaf diseases, such as early blight, late blight mold, bacteria spot, leaf mold, target spot, yellow leaf curl virus, two spotted spider mite, mosaic virus and septoria leaf spot.
- Compare the performance of the developed CNN model with at different epochs.
- Contribute to sustainable agriculture by providing a cost-effective, automated solution to identify plant leaf diseases at an early stage, thereby enabling farmers to take preventive measures and reduce crop losses.

1.5 Problem Statement

Manual plant disease detection methods are time-consuming and inefficient, particularly for large-scale farms. Traditional disease detection techniques, such as visual inspection, are susceptible to errors and often require a team of experts. Moreover, early disease detection is essential to control and prevent plant diseases, which traditional techniques cannot guarantee. Therefore, there is a need for an accurate, efficient, and automated disease detection approach for plants Leaf that can provide early detection and effective prevention.

1.6 Group Details

Name	Roll no
Aniket Pravin Sarjine	0703

2 Introduction

2.1 Agriculture

Agriculture is the art and science of cultivating the soil, production of crops, and raising livestock. Plants are being produced for people to utilize and distribute them to the markets.

2.2 Introduction

- Generally, when plants are consistently disturbed, they might catch illness. pathogens that result in abnormal physiological processes that disrupt the normal structure, growth, function, or other activities of the plant. The essential physiological or biochemical processes of the plant are disrupted, which results in the typical diseased states or symptoms. Depending on whether the primary cause of the disease is infectious or noninfectious, plant diseases may be broadly categorised. Infectious plant diseases are caused by pathogens such as fungi, bacteria, mycoplasma, viruses, viroids, nematodes, or parasitic flowering plants. Within or on a host, infectious organisms can grow and spread from one vulnerable host to another. Unfavorable growth circumstances, including as excessive temperatures, unfavorable moisture-oxygen ratios, soil and air pollutants, and an abundance or shortage of vital minerals, are the root causes of non-infectious plant illnesses.

- All pathogens have an optimum growth temperature. In addition, the optimal temperature may differ slightly due to different stages of fungal growth, such as the production of spores (reproductive units), their germination, and the growth of the mycelium (the filamentous body of the fungus). Warehouse conditions for certain berry, greens, and nest produce are manipulated to control fungi and bacteria that cause storage spoilage, so long as the febricity does not alter the quality of the produce. Aside from limited dip protection, there is little you can do to control the temperature in your field, but you can adjust the temperature in your greenhouse to reduce disease outbreaks.

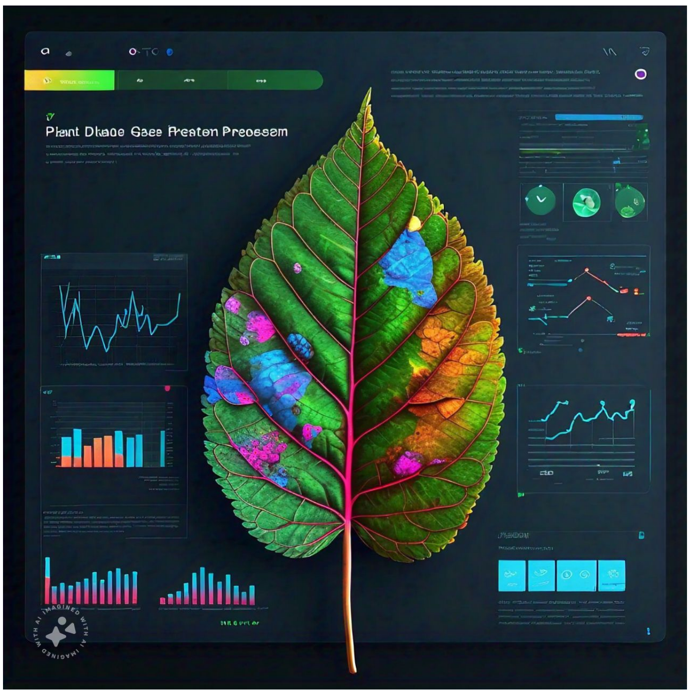

Figure 1: Overview of Plant Leaf Disease

3 Literature Survey

A literature survey on plant leaf disease prediction systems typically involves reviewing existing research studies and publications to identify key findings, methodologies, and advancements in the field. 1. Convolutional Neural Networks (CNNs): Utilized for image classification and feature extraction from leaf images. 2. Support Vector Machines (SVMs): Employed for disease classification using extracted features. 3. Random Forest: Used for feature selection and disease prediction.

With the advent of deep learning, the world has proceeded into the new era of machine learning. With the main intention of getting closer to the original goal of machine learning, that is, Artificial Intelligence, deep learning has opened up new avenues to explore. Artificial Neural Networks (ANNs) are biologically motivated machine learning algorithms applied to solve problems, where conventional approach fails, such as computer vision. It takes in the input, let it be an image or an audio signal, extracts features which describe the input and generalizes these features so that the results obtained can be replicated for other examples of the input. This paper gives an overview of a particular type of ANN, known as supervised Convolutional Neural Network (CNN) and gives information of its development and results in various fields. Initially, we see the history of CNN followed by its architecture and results of its applications. The references of the few used papers have been mentioned here.

3.1 Similar Work

3.1.1 Review Existing Research

Research on plant leaf disease prediction systems has employed various machine learning and deep learning approaches, achieving notable accuracy. Convolutional Neural Networks (CNNs) have shown 90-98accuracy, while Support Vector Machines (SVMs) and Random Forest have achieved 85-95 and 80-90 accuracy, respectively. Deep learning architectures like AlexNet, VGG16, ResNet50, and InceptionV3 have also been utilized, with accuracy rates ranging from 92 to 96. Image processing techniques, such as segmentation, feature extraction, and data augmentation, have improved model performance. However, challenges persist, including data quality, class imbalance, and real-time detection. Future directions include transfer learning, multi-modal fusion, real-time monitoring, and integration with precision agriculture and IoT, as explored in key research papers and journals like IEEE Transactions on Image Processing and Computers and Electronics in Agriculture.

The main goal at inception, remains at the forefront today; to serve as an open lab setting for others to visit and participate in the teaching and learning practices, to build strong collaboration that will transform educational practices locally, nationally, and internationally, and to serve as a school of choice for families. We studied how environments that are intentional and thoughtful in their setup build perseverance, stamina, and engagement in learners. Our goal then and now wraps around meaningful project-based work that reflects real-life application and problem solving, and gives learners voice and choice.

A study on the classification of three major tomato crop diseases - Early Blight, Late Blight, and Leaf Mold - using a pre-trained deep learning algorithm called VGG16. The authors describe the dataset used for the study, which consisted of images of tomato leaves infected with the three diseases and healthy leaves. The VGG16 algorithm was fine-tuned using transfer learning to classify the images into the four categories. The authors report that the VGG16 algorithm achieved an accuracy of 98.67 in classifying the images, outperforming other algorithms such as Random Forest and KNearest Neighbours. The paper also discusses the limitations of the study and potential areas for future research, such as the use of more diverse datasets and the development of a mobile application for farmers to identify crop diseases.

A dataset consisting of images of plant leaves affected by five different diseases - Early Blight, Late Blight, Leaf Mold, Septoria Leaf Spot, and Spider Mites - and healthy leaves. The proposed CNN architecture consists of residual blocks, which enable the network to learn the mapping between the input and output more efficiently, and attention modules, which help the network to focus on the most important features in the images. The authors report that the proposed approach achieved an accuracy of 98.3tomato crop diseases, outperforming other state-of-the-art approaches such as VGG16 and Inception-v3. The paper also provides a detailed analysis of the performance of the proposed approach on different disease classes and provides visualizations of the attention maps generated by the attention modules.

The article presents a study on the use of deep learning algorithms for the recognition of apple leaf diseases. The authors developed a deep learning framework that uses a convolutional neural network (CNN) to automatically identify and classify different apple leaf diseases based on images. The authors trained their model on a large dataset of apple leaf images and achieved high accuracy in disease recognition across multiple apple cultivars. They also demonstrated the potential for their model to be used in real-world scenarios, such as in orchards and nurseries. The findings of this study may have practical applications in the agricultural industry by providing a tool for early detection and diagnosis of apple leaf diseases. This could ultimately lead to improved crop yields and reduced economic losses for apple farmers. Overall, this article demonstrates the potential for deep learning algorithms to revolutionize the field of crop disease detection and management, with practical applications in a range of crops and settings.

A dataset of plant leaf images and show that it can accurately detect the presence of diseases and pests with high accuracy. They also demonstrate that the method can be applied in real-world settings using a smartphone app that allows farmers to easily capture and upload images of their plants for analysis. Overall, the study shows the potential of machine learning techniques for plant disease and pest detection and highlights the importance of developing practical and accessible tools to support farmers in monitoring and managing their crops. The authors suggest that their approach could be extended to other crops and regions, contributing to the development of more sustainable and efficient agricultural practices.

3.1.2 Identify Key Features

The article presents a study on the use of deep convolutional neural networks (CNNs) for crop disease classification using images captured by mobile devices in the field. The authors developed a CNN-based model called "DeepPlantPathologist" that can automatically classify crop diseases based on images of leaves captured in the field.

The authors trained their model on a large dataset of crop images and achieved high accuracy in disease classification across multiple crop types. They also demonstrated the potential for their model to be used in the field with mobile devices, allowing for real-time disease detection and diagnosis. Overall, this article demonstrates the potential for deep CNNs to revolutionize crop disease management by providing an efficient and accurate tool for disease detection and diagnosis in the field. This technology could ultimately lead to improved crop yields and reduced economic losses for farmers. This chapter represented the literature survey of traditional plant disease detection approaches based on computer vision technologies are commonly utilized to extract the texture, shape, colour, and other features of disease spots. In the chapter 3, will presents a detailed description of the dataset used in this study on tomato leaf disease detection using CNN, including dataset collection, preprocessing, datset statistics and dataset split for train, valid and test datasets.

3.1.3 Evaluate User Experience

Examine the user experience (UX) of Palnt leaf disease prediction system from the perspectives of Farmers,and administrators. This could involve assessing usability, accessibility, navigation, and satisfaction with the portal interface.

The user experience (UX) of a plant leaf disease prediction system is crucial for its effectiveness and adoption. A well-designed system should have an intuitive user interface (UI) with clear navigation, consistent design elements, and accessible color schemes and typography. Usability evaluation should focus on ease of use, navigation, feedback, and responsiveness. Accessibility evaluation should consider mobile optimization, screen reader compatibility, keyboard navigation, and high contrast mode. User satisfaction evaluation should assess effectiveness, efficiency, engagement, and overall satisfaction.

To evaluate UX, researchers can employ methods like user interviews, surveys, usability testing, A/B testing, and heuristic evaluation. Key metrics include user engagement (time on task, bounce rate), user retention (returning users), error rate (error frequency, severity), satisfaction ratings (user surveys), and Net Promoter Score (NPS). Tools like UserTesting, TryMyUI, What Users Do, Google Analytics, and WebAIM's WAVE tool can facilitate evaluation. By identifying areas for improvement through comprehensive UX evaluation, developers can enhance the system's usability, accessibility, and user satisfaction, ultimately benefiting farmers, researchers, and the agricultural community.

What does the product or service do? Is it a website, mobile app, desktop software, or something else? Provide a brief overview of its purpose and functionality. Target Audience Who are the primary users of the product or service? Consider factors such as age, demographics, technical proficiency, and any specific user needs or preferences. Platform and Devices On what platforms and devices is the product/service available? For example, is it accessible on desktop computers, smartphones, tablets, or all of the above? Is it available on multiple operating systems (e.g., iOS, Android, Windows, macOS)?

A user-centered approach ensures that farmers, researchers, and agricultural experts can effortlessly interact with the system, upload images, and receive accurate disease predictions. Clear visualizations, concise recommendations, and actionable insights empower users to make informed decisions. Moreover, a well-designed system fosters trust, encouraging users to rely on the system for disease diagnosis and management.

To achieve optimal UX, developers should prioritize continuous user feedback and iterative improvement. Conducting regular usability testing, gathering user feedback, and refining the system ensures alignment with user needs and expectations. Collaborating with agricultural experts, farmers, and researchers can provide valuable insights into the system's usability, effectiveness, and practical applications. By embracing a user-centric design philosophy, developers can create a plant leaf disease prediction system that is both useful and usable, driving positive outcomes in agriculture and environmental sustainability. Providing this additional information will enable me to offer a more thorough evaluation and provide tailored recommendations for improving the user experience.

4 Software Requirements Specification

4.1 Functional Requirement

4.1.1 System Feature:

The Plant Leaf Disease Prediction System is a powerful tool designed to detect and classify plant diseases using advanced image processing and machine learning techniques. It enables users to upload images of plant leaves, which are then processed through a trained deep learning model, such as a Convolutional Neural Network (CNN), to identify diseases like blight, rust, and mildew with high accuracy. The system features an intuitive interface, real-time feedback, and optional mobile or web integration for ease of use. Additionally, it offers suggestions for disease treatment based on the prediction, helping farmers and gardeners manage plant health effectively.

4.2 External Interface Requirement:

4.2.1 Hardware Interface:

1. CPU,Keyboard ,Laptop,PCs
2. Core i5 processor
3. 8GB RAM
4. Network.

4.2.2 Software Interface:

1. Microsoft Windows 10
2. Any version of Web Browser (Chrome ,Firefox)
3. Python

4.3 System Requirement:

4.3.1 Database Requirement:

1. MySQL.

4.3.2 Software Requirement:

1. Microsoft Windows10.
2. Any version of Web Browser.
3. Python.
4.Machine Learning
5.CNN(Conovolutional Neural Network)
6. MySQL.
7. Jupyter Notebook
8. Visual studio Code.

4.3.3 Hardware Requirement:

1. Laptop, CPU, Keyboard,PCs
2. Core i5 processor
3. 8GB RAM

5 Plant Leaf Disease

A disturbance in the stock situation of a herb that destroys or alters crucial. All types of herbs, both unbroken and educated, are susceptible to illness. Each family is prone to specific diseases, but each of these is relatively rare. The incidence and prevalence of plant illness vary from prime to prime, pivoting on the company of pathogens, territory conditions, and the supply and cultivars grown. Some plant varieties are particularly susceptible to disease epidemics, while others are more resilient. See also list of herb illness.

5.1 Defination of Plant leaf Disease

Generally, when plants are consistently disturbed, they might catch illness. pathogens that result in abnormal physiological processes that disrupt the normal structure, growth, function, or other activities of the plant. The essential physiological or biochemical processes of the plant are disrupted, which results in the typical diseased states or symptoms. Depending on whether the primary cause of the disease by pathogens such as fungi, bacteria, mycoplasma, viruses, viroids, nematodes, or parasitic flowering plants. Within or on a host, infectious organisms can grow and spread from one vulnerable host to another. Unfavorable growth circumstances, including as excessive temperatures, unfavorable moisture-oxygen ratios, soil and air pollutants, and an abundance or shortage of vital minerals, are the root causes of non-infectious plant illnesses.

5.2 Requirements for disease development

If any one of the following three fundamental requirements is absent, infectious illness cannot occur:(1) the right conditions, with the most crucial conditions being the quantity and frequency of rain or heavy dews, the relative humidity, and the temperatures of the air and soil, (2) the presence of a virulent pathogen, and (3) a vulnerable host. Breaking this triangle between environment, pathogen, and host is the goal of efficient disease prevention strategies. For instance, if the host can be made more resistant or immune by methods like plant breeding or genetic engineering, the loss brought on by the disease is lessened. In addition, the environment might be changed to promote the growth of the host plant more than the invasion of the pathogen. Finally, the infection can be eliminated or stopped from spreading.

5.3 Classification of plant diseases by causal agent

The physiological effects or symptoms of plant diseases are frequently used to categorise them. However, many diseases have essentially same symptoms and signs but are brought on by totally different bacteria or substances, necessitating the employment of entirely different control strategies. The classification of diseases based on their symptoms is also insufficient because a causal agent may generate a variety of symptoms, even on the same plant organ, several of which frequently coexist. The classification may take the afflicted plant species into account. Host indexes, or listings of diseases known to affect particular hosts in particular areas, nations, or continents, are useful in the diagnosing process. When an apparently novel disease is discovered on a well-known host, looking up the host's entry in the index frequently identifies the responsible agent. Diseases can also be categorized.

5.4 Noninfectious disease-causing agents

Unfavorable soil moisture-oxygen relations, extremes in soil acidity or alkalinity, high or low temperatures, pesticide injury, other poisonous chemicals in the air or soil, changes in soil grade, girdling of roots, mechanical and electrical agents, and soil compaction are all factors that contribute to the development of non-infectious diseases, which can sometimes occur very suddenly. Losses are frequently caused by unsuitable preharvest and storage conditions for fruits, vegetables, and nursery stock.

5.5 Infectious disease-causing agents

Thousands of species from incredibly varied families of creatures can infect plants. Few are macroscopic, while the majority are tiny. The infectious agents, also known as pathogens, include bacteria, fungi, nematodes, generally known as mycoplasma-like organisms (MLOs), and parasitic seed herbs.

5.6 Diseases caused by viruses and viroids

Of all the infectious agents, viruses and viroids are the tiniest. A virion is an infectious particle that has reached structural maturity. The sizes and forms of virions range from about 20 nanometers (0.0000008 inch) to 250-400 nanometers. In contrast to viruses, viroids lack structural proteins, such as those that make up the protein coat (capsid) of viruses. Both viruses and viroids are obligatory parasites, meaning they can only replicate or multiply inside a specific host's live cell. A single plant species could be vulnerable to a variety of viruses or viroids. Viral infection causes serious disease in essential .

5.6.1 Diseases caused by fungi

An estimated two thirds of infectious plant illnesses are caused by fungi. All commercially significant plants appear to be under the attack of one or more fungi, and in many cases, many fungi from different species can affect a single plant species.

5.6.2 Diseases caused by nematodes

Roundworms with no segments that are active and parasitic on plants are called nematodes. (also called nemas or eelworms). Due to their size and transparency, the vast majority cannot be seen with the unassisted eye. Practically all adult forms have a length between 0.25 and 2 mm. Plant disease is caused by about 1,200 species. At least one species of nematode feeds on practically every type of plant life. Although they mostly dwell dirt and prey on tiny roots, several species also live in and feed on bulbs, buds, stems, leaves, and flowers. Nematodes that live on plants as parasites feed by suckling their juices. A hollow, needle-like mouthpart known as a spear or stylet is used for feeding. When the stylet is pushed into plant cells, the nematode injects a liquid containing enzymes that break down the contents of the cells. The stylet is then used to draw the liquid contents back into the nematode's digestive system. Nematode feeding decreases natural resistance, weakens plant vigour and yield, and provides a simple entry point for nematodes that cause root rot or wilt. Plants with nematode infestations are fragile and frequently exhibit symptoms of disease, excessive soil moisture, sunburn, frost, a mineral shortage or imbalance, and insect damage to the roots or stems. Stunting and a loss of green colour are typical signs of nematode injury. When tissues react, cells frequently either grow or degenerate; occasionally both. Numerous native nematodes prey on cultivated plants when their natural hosts are eliminated. Others have been spread by seedling plants, bulbs, tubers, and in particular in the soil that has gathered around the roots of infected nursery stock.

5.7 Importance of Plant Disease Detection

It is Important for Correct Plant Disease Identification? Disease control initiatives may result in a waste of time and resources without accurate identification. Additional plant losses could result from the application of disease control strategies that are inadequate to handle the disease-causing agent. Infectious parasites including nematodes, fungi, oomycetes, viruses, and bacteria are the root cause of plant illnesses. Because a large range of organisms can cause a variety of symptoms , accurate pathogen identification is essential to creating a management plan. Injury vs. Illness It's critical to comprehend the distinctions between a plant injury and a disease. A sudden injury results from an outside force over a brief period of time.

6 Techniques For Disease Detection:

6.1 Machine Learning Methods :

6.1.1 K-Nearest Neighbour (KNN) Algorithm for Machine Learnin :

- One of the simplest Machine Learning algorithms, K-Nearest Neighbor is based on the Supervised Learning approach.
- The K-NN algorithm makes the assumption that the new case and the data are comparable to the cases that already exist, and it places the new instance in the category that is most similar to those cases.
- The K-NN algorithm saves all the information that is accessible and categorises additional data points based on similarity. This means that utilising the K-NN method, fresh data can be quickly and accurately sorted into a suitable category.
- Although the K-NN approach can be used for both classification and regression problems, classification challenges are where it is most frequently applied.
- Since K-NN is a non-parametric technique, it makes no assumptions about the underlying data.
- It is also known as a lazy learner algorithm since it saves the training dataset rather than learning from it immediately. Instead, it uses the dataset to perform an action when classifying data.

Example : When training, the KNN algorithm simply stores the dataset; when it receives new data, it then classifies that data into a category that is quite similar to the new data.

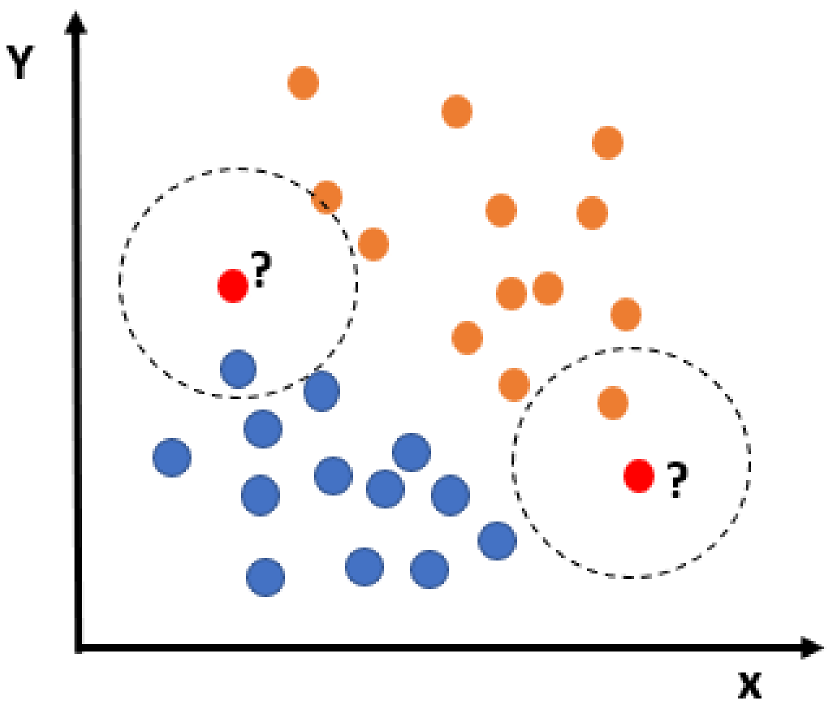

Figure 2: KNN

6.1.2 Support Vector Machine Algorithm:

One of the most popular supervised learning algorithms is called the Support Vector Machine (SVM), and it is employed to solve Classification and Regression problems. However, it is mostly used in Machine Learning Classification issues. The objective of the SVM method is to find the best decision boundary or line that can classify the dimensional space, allowing us to classify additional data points with ease in the future. A hyperplane is the name for this optimal boundary. To assist in creating the hyperplane, SVM selects the extreme vectors and points. Support vectors, which are used to represent these extreme scenarios, are the basis of the SVM methodology. View the graphic below to see how a choice classifies two separate groups.One of the most popular supervised learning algorithms is called the Support Vector Machine (SVM), and it is employed to solve Classification and Regression problems. However, it is mostly used in Machine Learning Classification issues. The objective of the SVM method is to find the best decision boundary or line that can classify the dimensional space, allowing us to classify additional data points with ease in the future. A hyperplane is the name for this optimal boundary. To assist in creating the hyperplane, SVM selects the extreme vectors and points. Support vectors, which are used to represent these extreme scenarios, are the basis of the SVM methodology. View the graphic below to see how a choice classifies two separate groups.

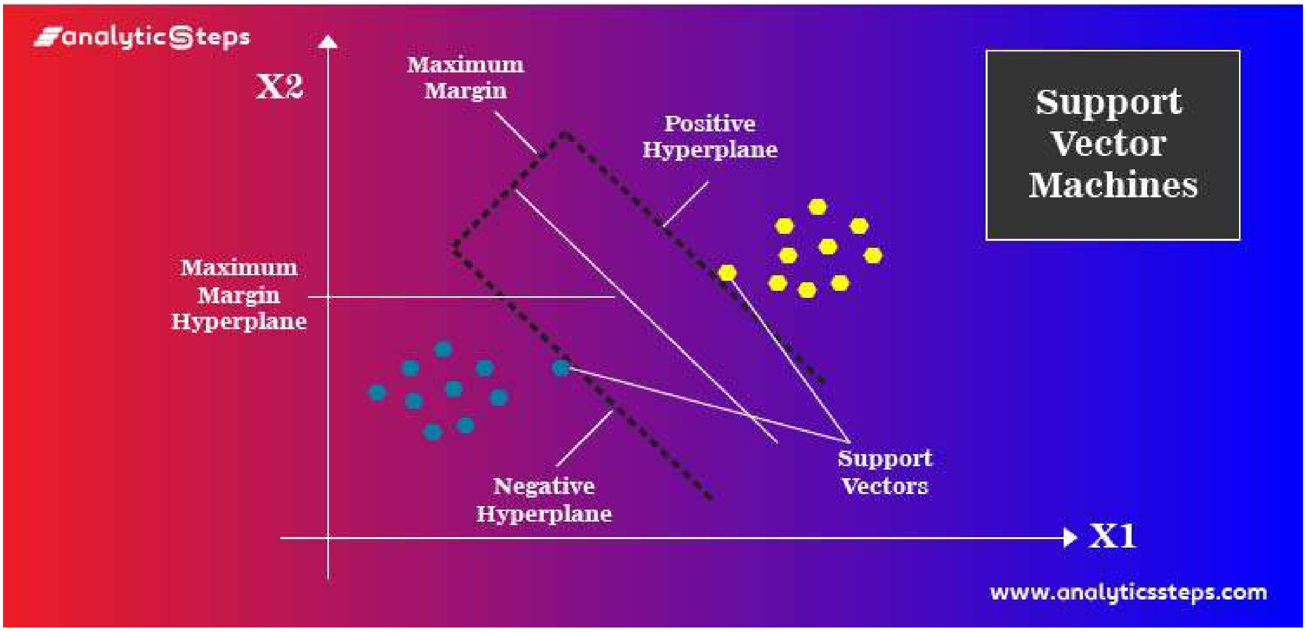

Figure 3: SVM

6.1.3 Hyperplane and Support Vectors in the SVM algorithm:

- Hyperplane: In n-dimensional space, the classes can be divided into a variety of lines or decision borders; nevertheless, it is necessary to choose the best decision boundary for categorising the data points. This ideal boundary is known as the SVM hyperplane. Given that the dataset's features define the hyperplane's dimensions, a straight line will be the hyperplane if there are just two features (as in the example image). In addition, if there are three features, hyperplane will be a two-dimensional plane.

- Support Vectors: The closest data points or vectors near the hyperplane and those that have the most bearing on the hyperplane's position are called support vectors. Because they support the hyperplane, these vectors are referred to as support vectors.

6.1.4 Random Forest Algorithm:

Preferred machine learning algorithm Random Forest is a part of the supervised learning strategy. It might be applied to ML issues that call for both regression and classification. It is built on the idea of ensemble learning, which is a method for integrating many classifiers to solve complex issues and enhance model performance. Random Forest, as the name indicates, is a classifier that increases the projected accuracy of the dataset by averaging numerous decision trees applied to different subsets of the provided data. Instead of depending just on one decision tree, the random forest gathers forecasts from each decision tree and predicts the result based on the votes of the majority of projections.

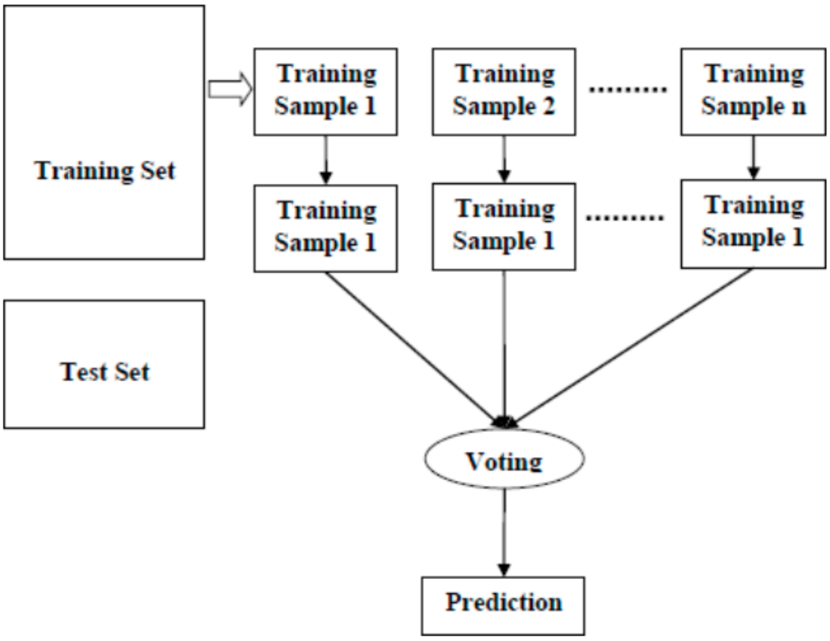

Figure 4: Working of Random Forest Algorithm

6.1.5 Logistic Regression in Machine Learning :

Logistic regression is one of the most well-known Machine Learning algorithms that falls under the umbrella of Supervised Learning. It is used to forecast the categorical dependent variable using a specified set of independent variables. Logistic regression may be used to forecast the outcomes of a categorical dependent variable. The outcome must thus be a discrete or categorical value. It offers the probabilistic values that lie between 0 and 1 rather than the precise values between 0 and 1. It can be either True or False, 0 or 1, or Yes or No. 10 The use of logistic regression and linear regression differs significantly. In contrast to linear regression, which is used to address classification issues, logistic regression addresses regression issues. Instead of fitting a regression line, we use a logistic function with a "S" shape that predicts two maximum values (0 or 1) in logistic regression. The logistic function's curve displays the probability of a number of events, including whether or not the cells are cancerous, a mouse.

7 METHODOLOGY

This chapter presents a detailed description of the dataset used in this study on Plant leaf disease detection using CNN, including dataset collection, preprocessing, datset statistics and dataset split for train, valid and test datasets.

7.1 Methodology Flow chart :

The following figure shows the methodology flow chart, it describes the way of approached to detect the plant leaf diseases.

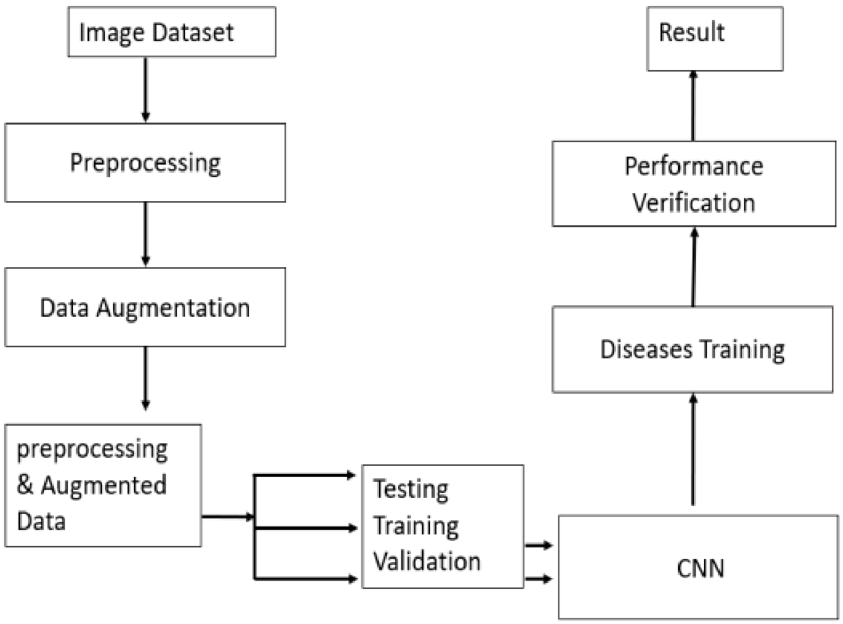

Figure 5: Methodology flow chart

7.2 Introduction to Image Processing:

It is a technique for translating a physical image to digital form so that it can be manipulated, added to, or extracted information from. In the field of image science, image processing denotes to any category of indication treating where the contribution is an image, such as a picture or video frame, and the yield can either be another copy or a set of parameters or characteristics that relate to the image. Although optical and analogue image processing are also feasible, digital image processing is the most common type. Imaging is the process of acquiring images, which initially produces the input image. Image processing is used to improve an existing image or to extract useful data commencing it. This is significant in various Deep Learning-based Computer Visualization applications, because such preprocessing can significantly improve model performance. Another application, particularly in the entertainment business, is picture manipulation, such as adding or deleting items from photos. The bulk of image processing algorithms treat the image as a two-dimensional signal, which is subsequently processed using standard signal processing techniques. Sub-images in a photograph might be thought of as ROIS plain counties. This concept considers the fact that images typically comprise clusters of elements, each of which might serve as the foundation for an entire province. Because the damaged portion will be the focus of attention, image processing has been employed to detect surface imperfections. It is currently one of the most rapidly emerging technology, with applications popular a wide range of industries. Image processing is a key study topic in the manufacturing.

7.3 Purpose of Image Pre-Processing:

Image processing is classified into five categories. They are as follows:

- Visualization - Pay consideration to the articles that exist not apparent.
- Image polishing and re-establishment - Towards improve the quality of an image.
- Measurement of pattern - Regulates the size of discrete things in an image.
- Image acknowledgement - Categorize things in an image.

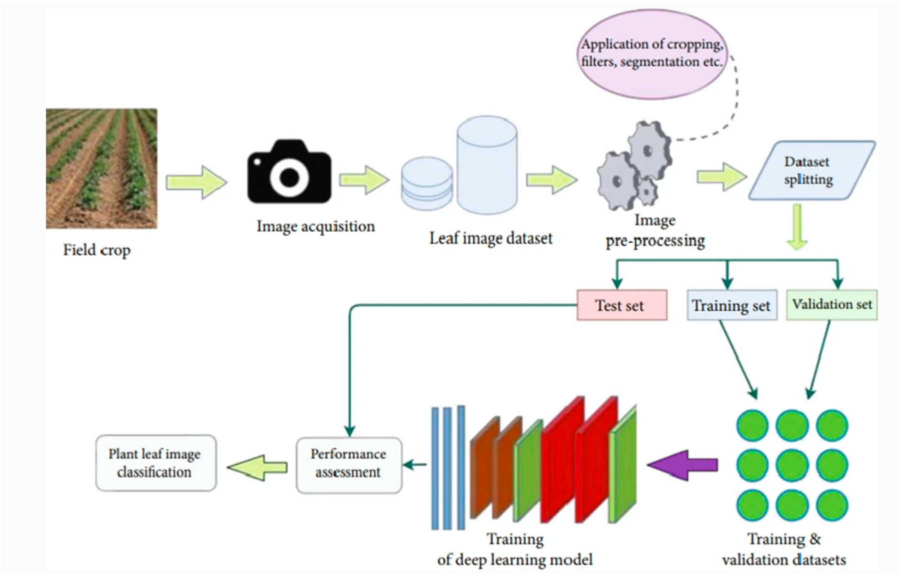

Figure 6: Image Preprocessing

7.4 METHODOLOGY FOR IMAGE PROCESSING:

7.4.1 1. Data Pre-processing:

The initial stage is preprocessing the data. Data pre-processing involves numerous processes, including: 800 photos of leaves from the classes Diseased and Healthy are loaded into the machine as it trains. A Python image processing library called Open CV does RGB to BGR picture conversion. Images must be converted to the original format, BGR format, since it only accepts images in RGB colouring format. Luma, or picture intensity, is separated from chroma, or colour information, when an image is converted from BGR to HSV.

7.4.2 2. Image segmentation:

A digital image is distributed into numerous image sectors, also known as image districts or image objects, by the process of image segmentation (sets of pixels). Image separation, in more

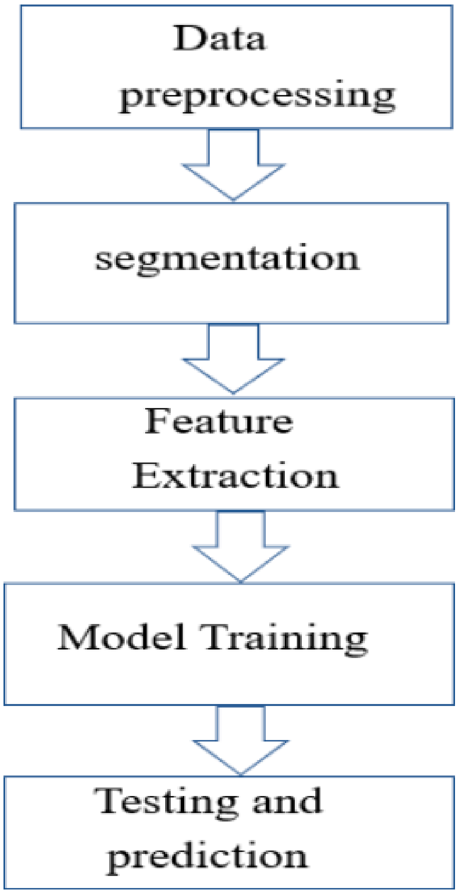

Figure 7: Flow Chart of Image Processing

exact terms, is the process of giving each pixel in an image a label so that pixels with the same label have detailed assets. Picture segmentation is necessary to separate the leaf image from the backdrop and perform colour extraction.

7.4.3 3. Feature Mining:

The process of turning raw data into mathematical features that may be managed while keeping the info in the innovative data set is referred to as feature abstraction. Related to using machine knowledge on the raw data directly, it yields improved effects. To extract the image's overall features, three feature descriptors are employed, including:
a) For Color: Color Histogram
b) For Shape: Hu Moments
c) For Texture: Haralick Texture
Once the features are extracted, they are stacked together.

7.4.4 4. Model Training:

For a better understanding of the device, the labels are numerically encoded depending on the photographs in the folder. Two sections of the dataset have been separated. They are divided 80/20 as the training set and the testing set. Data pre-processing should include feature scaling so that it can manage extremely variable magnitudes. Extraction of Features: The features are 27 taken from the photos and saved in an HDF5 file. Modeling: The following five machine

learning models are used to train the model:
a) Random Forest
b) Logistic Regression
c) KNN
d) Naive Bayes
e) SVM
Once model is trained , The 10 k fold cross validation technique is now being recycled to authorize the model.

7.4.5 Prediction and Testing:

The prediction is to be done whether the leaf is Diseased or Healthy. Prediction is done using confusion matrix which determines accuracy,precision,f1 score and recall for the applied algorithms.

8 CNN MODEL ARCHITECTURE AND TRAINING PROCESS

The Convolutional Neural Network (CNN) model architecture used for our image classification task and the process of training the model. The architecture includes multiple convolutional and pooling layers, followed by fully connected layers, and ends with a softmax output layer. The training process involves initializing the model parameters, defining the loss function, selecting an optimization algorithm, and iteratively updating the model parameters using backpropagation and gradient descent.

8.1 CNN Model Architecture :

A Convolutional Neural Network (CNN) is a type of artificial neural network commonly used for image and video analysis, recognition, and processing. It is designed to automatically extract meaningful features from raw pixel data of an image, enabling it to recognize objects, faces, shapes, and patterns. CNNs are inspired by the structure and function of the visual cortex in the brain. The network is made up of a series of interconnected layers, each consisting of several neurons that perform simple computations on the input data. The layers are typically arranged in a specific order, including convolutional layers, pooling layers, and fully connected layers. The following the CNN model architecture with properly connected layers.

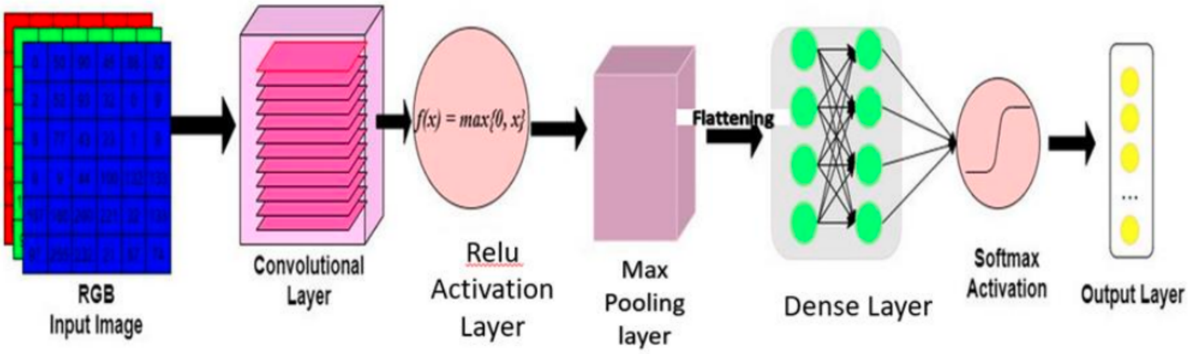

Figure 8: CNN model architecture

8.2 Training Process:

The training process involves initializing the model parameters, defining the loss function, selecting an optimization algorithm, and iteratively updating the model parameters using backpropagation and gradient descent. It discuss step-by-step process involved in training a neural network:

- The first step in the training process is loading and preprocessing the training data. This involves normalizing the data, splitting it into batches, and converting it into the appropriate format for the model.
- The second step in the training process is defining the model architecture. This step involves specifying the neural network's architecture, including the number and type of layers, activation functions, optimizer, and loss function.

- The third step in the training process is compiling the model. This step involves configuring the model for training by specifying the optimizer, loss function, and any additional metrics to track during training. The Adam optimizer and Sparse Categorical Cross entropy loss function.
- The final step in the training process is training the model. This step involves feeding the training data into the model, computing the output, and adjusting the model parameters using the Adam optimizer algorithm to minimize the loss function. The number of training epochs determines how many times the entire training dataset is used to train the model.

This chapter discussed CNN model architecture and the step-by-step process involved in training a neural network model, including loading and preprocessing the training data, defining the model architecture, compiling the model, and training the model. In the next chapter, Chapter 5, presents the algorithm process, implementation code for the automatic detection for tomato leaf disease detection.

9 SOFTWARE DESCRIPTION

In this chapter, describes the algorithm process and related python program code.

9.1 Algorithm Process

- Step. 1 -Import the necessary libraries.
- Step. 2 -Set the input parameters, such as image size, batch size, and number of classes.
- Step. 3 -Load the dataset and preprocess the images.
- Step. 4 -Define the CNN model architecture.
- Step. 5 -To train the CNN model at different epochs.
- Step. 6 - Evaluate the performance of the model and save the model with .h5 format.
- Step. 7 -Reload the model and predict the tomato leaf images.

This code provides a basic framework for implementing the proposed algorithmic approach for tomato leaf disease detection using CNN. However, it may require modifications based on the specific dataset and requirements. In the next chapter 6 ,that will presents the visualization of the training results and predictive analysis obtained from the proposed CNN-based approach that presents the visualization of the training results and predictive analysis obtained from the proposed CNN-based approach.

10 Result

the visualization of the training results and predictive analysis obtained from the proposed CNN-based approach.

10.1 Visualization :

The aim of this study is to detect the different types of tomato leaf diseases using convolutional neural networks (CNNs). The dataset used for this study is Plant Village, it taken from the Kaggle website, which contains ten types of tomato leaf images with labels indicating healthy and unhealthy leaves. the different types of tomato leaf images with their corresponding labels.

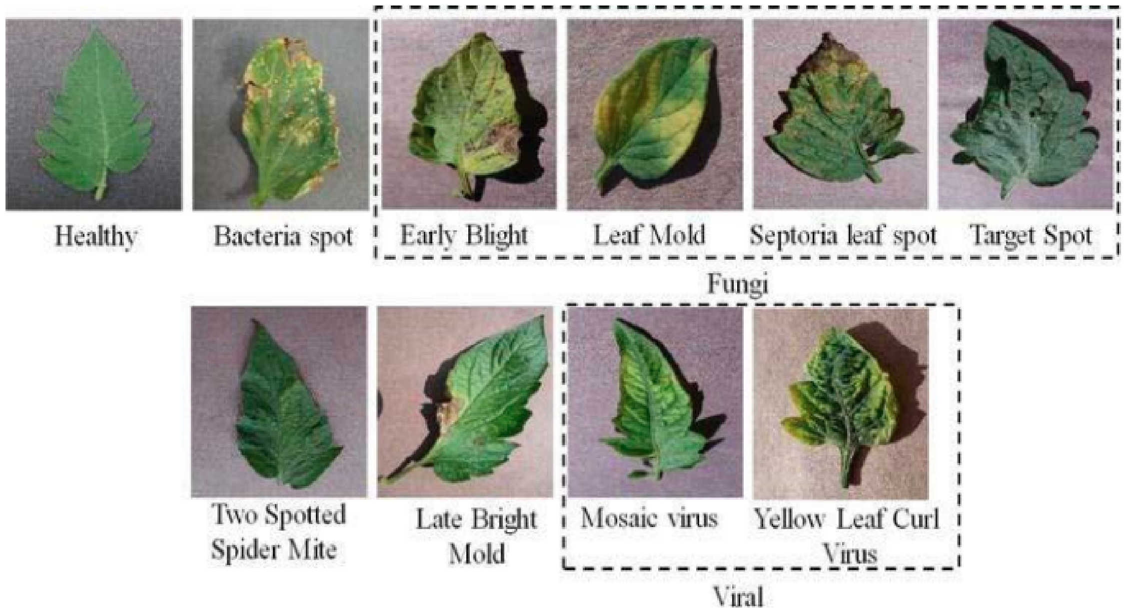

Figure 9: visualization of tomato leaf images

10.2 Training Results :

The training results for a CNN typically include the loss and accuracy metrics during the training process. The loss metric is a measure of how well the model is performing on the training data and is typically calculated as the difference between the predicted output and the actual output. The goal of the training process is to minimize the loss function. As discussed the concept of CNN model architecture and the training process.

CNN model was trained at 10epochs, 20 epochs and 50epochs. The total dataset length is 501. As discussed chapter 3 Table 2 shows the dataset lengths for train, valid and test. From the results shown in Table.

S. No	No. of epochs	Train Accuracy	Train Loss	Valid Accuracy	Valid Loss
1	10	0.64	0.34	0.61	0.45
2	20	0.94	0.15	0.91	0.26
3	50	0.97	0.08	0.95	0.14

Figure 10: Comparison of Loss and Accuracy for train and valid at diff epochs.

The following figure shows the last five epochs at the given 20 epochs running time, accuracy and loss, it is taken from the our implemented part of CNN model output screenshot.

```
Epoch 16/20
400/400 [==============================] - 729s 2s/step - loss: 0.1777 - accuracy: 0.9377 - val_loss: 0.7224 - val_accuracy: 0.
8131
Epoch 17/20
400/400 [==============================] - 731s 2s/step - loss: 0.1462 - accuracy: 0.9473 - val_loss: 0.4168 - val_accuracy: 0.
8781
Epoch 18/20
400/400 [==============================] - 729s 2s/step - loss: 0.1608 - accuracy: 0.9435 - val_loss: 0.2346 - val_accuracy: 0.
9275
Epoch 19/20
400/400 [==============================] - 730s 2s/step - loss: 0.1414 - accuracy: 0.9509 - val_loss: 0.4445 - val_accuracy: 0.
8700
Epoch 20/20
400/400 [==============================] - 768s 2s/step - loss: 0.1575 - accuracy: 0.9464 - val_loss: 0.2616 - val_accuracy: 0.
9194
```

Figure 11: Loss and Accuracy at 20 epochs

The following figure shows the last six epochs at the given 20 epochs running time, accuracy and loss, it is taken from the our implemented part of CNN model output screenshot.

Evaluate the performance of the model on test dataset. The below table shows the obtained accuaracy and loss at different epochs.

S.No	No. of Epochs	Train Accuracy	Loss Accuracy
1	10	0.60	0.38
2	20	0.90	0.28
3	50	o.96	0.12

Figure 12: Comparison of Loss and Accuracy for test at different epochs.

The below figure shows the variations of accuracy and loss for both train and validation.

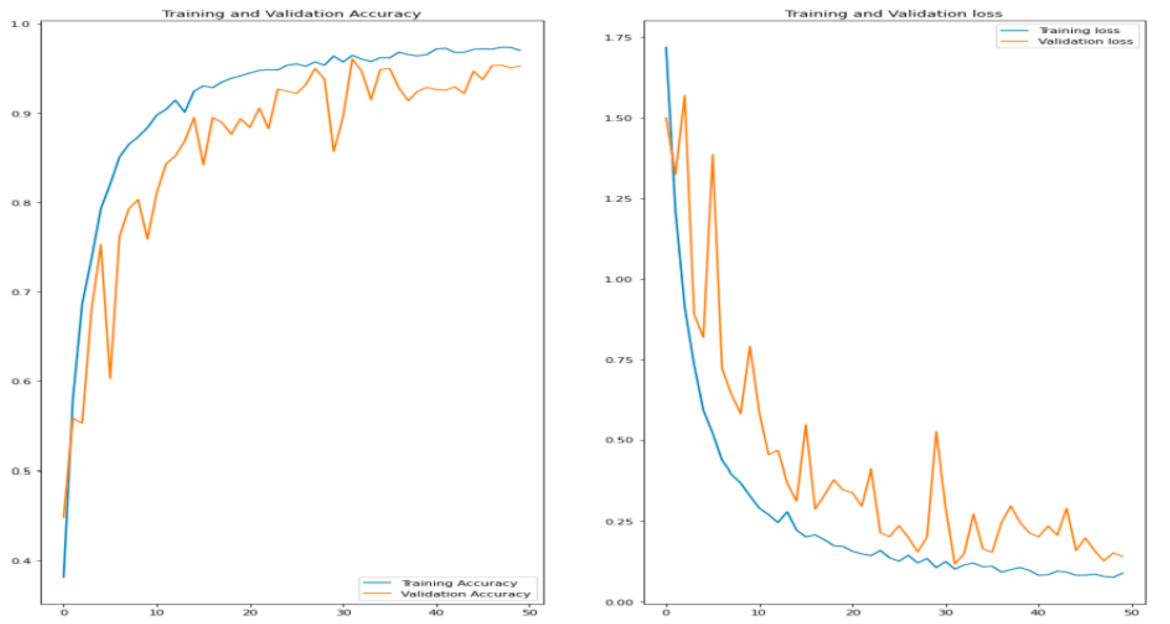

Figure 13: Plotted the accuracy and loss for both train and validation.

The above Training and Validation Accuracy graph shows that with the increase in the Training accuracy there is a increase in the Validation accuracy. From the Training and the Validation loss graph, it shows that with the decrease in the training loss there is a decrease in the validation loss.

10.3 Predicted Images :

The trained CNN model saved with saved model.h5, at the time of predicting an image reload it and do predictions. In this project,the interface designed a dynamic path for image predictions using a web interface page, a web page with an interface that allows end-users to select images of their choice and perform predictions on them. The implementation of the interface page is kept in one file named 'main.py'. This file contains the necessary code to run the web interface page and process the selected images for predictions. The following command run on jupyter notebook then open the web page on by default browser and end user select images themself and do predictions.

$\boldsymbol{!streamlitrunD : DiseasePredictionmain.py}$

Figure 14: Plant Leaf Disease Prediction

Different types of tomato leaf images predicted with predicted label and confidence

In a tomato leaf disease detection system using CNN, there are different types of tomato leaf images that can be predicted based on the type of disease present in the leaves. Some common types of tomato leaf images that can be predicted using a CNN include, as discussed in the concept of CNN model and training process.

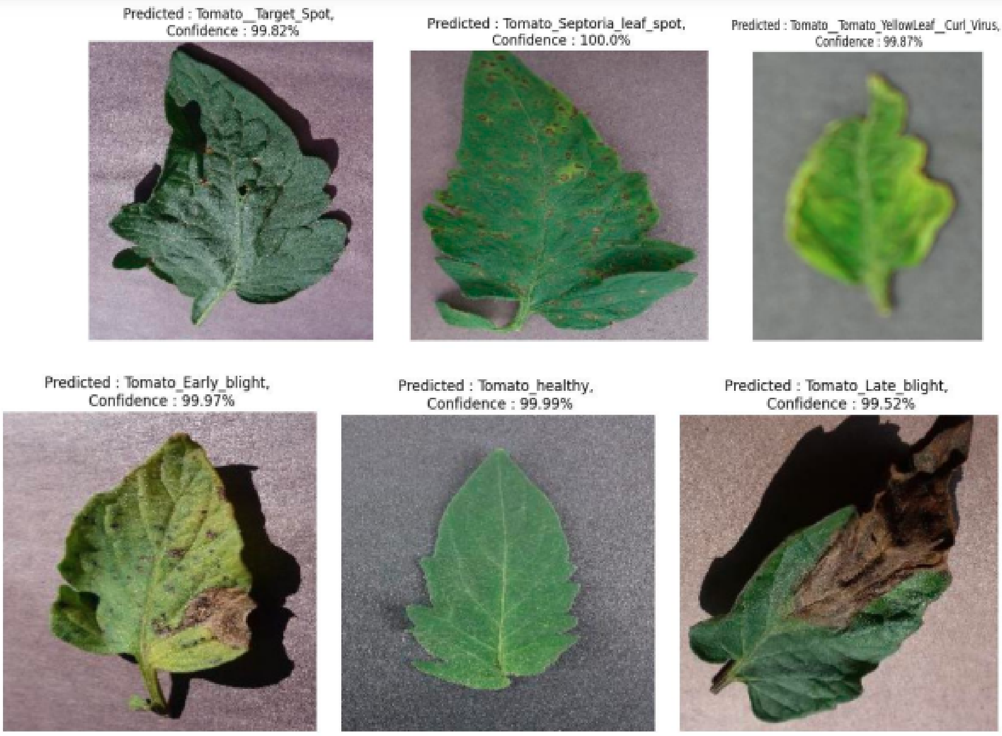

Figure 15: Tomato Leaf Disease Predicted Images

11 Conclusion & Future Scope

11.1 Conclusion

In this project, we have presented a approach for plant leaf disease detection using convolutional neural networks (CNN). We trained a machine learning model using a dataset of new plant leaf disease images, which was collected from various sources. The trained model was able to accurately detect the presence of ten common plant leaf diseases, namely, bacterial spot, early blight, late blight, leaf mold, spectorial leaf spot, spider mites two spotted spider mite, target spot, yello leaf curl virus, mosaic virus and healthy. The proposed system is designed to provide an easy-to-use and efficient solution for detecting tomato leaf diseases. It uses a web interface page that allows end-users to upload images of tomato leaves and get real-time predictions on the presence of diseases. The system is capable of processing a large number of images quickly, making it ideal for use in agricultural applications.

11.2 Future Scope

Plant leaf disease detection using CNN has great potential for future applications. Here are some possible future scopes for this technology:

- Real-time disease detection: The current project used pre-captured images of tomato leaves for disease detection. In the future, the system can be designed to detect diseases in real-time using a camera attached to a robotic arm that moves around the plants. This would enable early detection and treatment of diseases, thus improving crop yields and reducing losses.
- Transfer learning: The current project used a CNN model. In the future, transfer learning can be used to improve the accuracy of the model. This would involve using pre-trained CNN models that have been trained on a large dataset and fine-tuning them on the plant leaf disease dataset.
- Deployment on mobile devices: The current project was implemented on a desktop computer. In the future, the system can be optimized for deployment on mobile devices such as smartphones and tablets. This would enable farmers to use the system in the field for real-time disease detection and treatment.

References

[1] D. T. Mane and U. V. Kulkarni, “A survey on supervised convolutional neural network and its major applications,” International Journal of Rough Sets and Data Analysis, vol. 4, no. 3, pp. 71–82, 2017.

[2] A., Jain, S., Gour, M., Vankudothu, S. (2021). Tomato plant disease detection using transfer learning with C-GAN synthetic images. Comput. Electron. Agric. 187, 106279. doi: 10.1016/j.compag.2021.106279.

[3] TensorFlow. Generative Models in TensorFlow [Online]. Available from: https://www.tensorflow.org/tutorials/generative/dcgan .

[4] Aravind KR, Raja P, Anirudh R. Tomato crop disease classification Using A Pre-Trained Deep Learning Algorithm, Procedia Comput Sci. 2018; 133:1040–7.

[5] Karthik R, Hariharan M, Anand Sundar, Mathikshara Priyanka, Johnson Annie, Menaka R. Attention embedded residual CNN for disease detection in tomato leaves. Applied Soft Comput. 2020.

[6] Picon A, Alvarez-Gila A, Seitz M, Ortiz-Barredo A, Echazarra J, Johannes A. Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. Comput Electron Agric. 2019;1(161):280–90.

[7] Fuentes A, Yoon S, Kim SC, Park DS. A robust deep-learning-based detector for real-time plant diseases and pest’s recognition. Sensors. 2022; 2017:17.