International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

# Crop Leaf Disease Prediction Using Machine Learning

Mr. Rushikesh Pawar , Mr. Rushikesh Awatade , Mr. Omkar Shedge ,Mr. Saif Shaikh
Under The Guidance Of Prof. Vikram Popat Deokate
**Department Of Computer Enguneering**

**Al Ameen Educatonal & Medical Foundation's College of Engineering Koregoan Bhima, Pune-412216**

## ABSTRACT

- In India, crop area is largest in the world and produces major crops like wheat, pulses, fruits, rice and vegetables Despite of using modem taming techniques along with traditional, infectious plant diseases is major problem which can be caused by different viruses, fungus and bacteria. This mainly affects crop production as well as crop quality. It is very important to identify diseases at early stage Nowadays, automatic crop de detection has become a important research domain. It helps in detecting the symptoms of the disease when they are found on the e In this paper we will focus on finding the diseases in order to increase crop quality and production effectively. Here, we will focus on r diseases by observing leaves of plants at initial stage using machine learning.

- In this paper, we designed a Deep Convolutional Neural Network based on LeNet to perform soybean leaf spot disease recognition and classification using affected areas of disease spots. The affected areas of disease spots were segmented from the leaves images using the Unsupervised fuzzy clustering algorithm. The proposed Deep Convolutional Neural Network model achieved a testing accuracy of 89.84%, and poor per class recognition results in 1378 images misclassified, and 1271 images correct classified. TheVGG16 achieved the best performance reaching a 93.54% success rate, and better per class recognition results in 1245 images misclassified, and 1404 images correct classified

- In order to address the challenges related to the classification and recognition of soybean disease and healthy leaf identification, it is essential to have access to high-quality images. A meticulously curated dataset named "SoyNet" has been created to provide a clean and comprehensive dataset for research purposes. The dataset comprises over 9000 highquality soybean images, encompassing healthy and diseased leaves. These images have been captured from various angles and directly sourced from soybean agriculture fields; The soybean leaves images are organized into two sub-folders: SoyNet Raw Data and SoyNet Pre-processing Data.The SoyNet Pre-processing Data folder comprises resized images of 256∗256 pixels and the grayscale versions of disease and healthy images, following a similar organizational structure. We captured the images using the Nikon digital camera and the Motorola mobile phone camera, utilizing different angles, lighting conditions, and backgrounds. They were taken in different lighting conditions and backgrounds at soybean cultivation fields to represent the real-world scenario accurately. The proposed dataset is valuable for testing, training, and validating soybean leaf disease classification.

# 1. INTRODUCTION

## 1.1 GENERAL

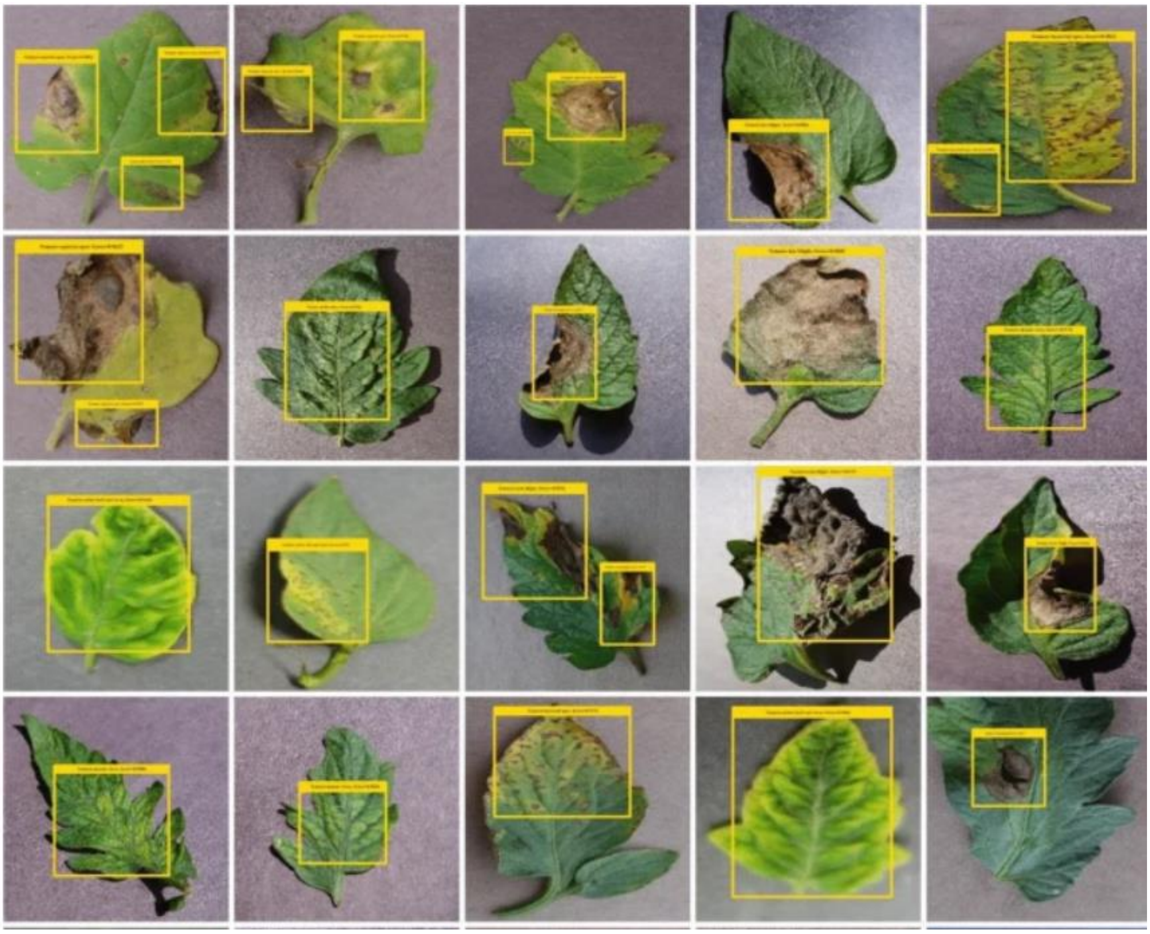
In India, about 70% of the population depends on agriculture for their livelihood. Identifying errors is important to prevent outages. The study of plant diseases is very problematic. It requires a lot of studies and a lot of sessions, especially in plant diseases. Therefore, image processing and machine learning models can be used to detect plant diseases. This project describes methods for detecting plant diseases with the help of leaf images. Image processing is the branch of the configuration function that extracts image objects or useful information from images. Machine learning is a subset of artificial intelligence that performs tasks or provides instructions to perform certain tasks. The main purpose of machine learning is to understand the data presented and fit it into the models that should be useful to humans. Thus, it helps to make good decisions and predict outcomes using a lot of information Color, leaf damage, leaf shape, and texture of leaves were used for classification. In this project, we analyzed different images or features to identify different diseases of leaves to obtain the highest accuracy. Previously, the diagnosis of plant diseases was made by experts by visual inspection of leaves or by some chemical treatment This requires a large group of experts and continuous analysis of plants, which is very expensive when we are a large farm. Based on these conditions, the proposed method was found to be effective in monitoring crop fields. By looking at the symptoms on the leaves of plants, automatic diagnosis of diseases has been made easier and cheaper. Compared to the other deep learning methods, the proposed solution for plant diseases is less expensive and takes less time to guess because it uses machine research and image processing algorithms.

## 1.2 MOTIVATION:

Crop Yield and Food Security: Plant diseases can have a significant impact on crop yield and quality. By detecting and diagnosing leaf diseases early, farmers and agricultural experts can take timely action to prevent or mitigate the spread of diseases, ensuring higher crop productivity and better food security. Economic Losses: Crop diseases lead to

Substantial economic losses for farmers and the agricultural industry as a whole. Detecting leaf diseases early can help minimize these losses by enabling targeted interventions such as disease-specific treatments, optimized resource allocation, and timely crop management practices. Environmental Sustainability: Disease detection plays a crucial role in promoting sustainable agricultural practices. By identifying plant diseases accurately, farmers can minimize the unnecessary use of pesticides, fungicides, and other agrochemicals, thereby reducing the environmental impact associated with their usage.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930



**1.3 PROBLEM STATEMENT**

The leaf shows the symptoms by changing their color, showing the spots on the leaf like yellow spots or Black spots or chocolate brown spots. Some leaves are having mildew such as Powdery Mildew, Downey Mildew. This identification of the disease is done by manual observation and pathogen detection which is more time consuming and somehow costly with a lower accuracy. So there is a better option which is fast and accurate detection by using image processing techniques using CNN.

**1.4 OBJECTIVE:**

The objective of plant disease detection using machine learning (ML) is to develop an automated and efficient system that can accurately identify and classify diseases in plants. By leveraging ML algorithms and techniques, this approach aims to enhance the speed and accuracy of disease diagnosis, leading to timely interventions and improved crop management.

## 2. LITERATURE REVIEW:

**1.Geeta Gunasekar ,S .Samundeswari,Saranya Gangadhara Moorty "**Plant Leaf Disease Classification and Detection System Using Machine Learning" This paper is presenting
Agricultural intervention in the livelihood of rural India indulges by about 58%. Among the agricultural products, tomato is one of the most used crops. Thus, preventing significant loss in quantity and yield of tomatoes is majorly dependent on the recognition and classification of diseases a tomato plant might possess. Latest and fostering technologies like Image processing is used to rectify such issues using different types of techniques and algorithms. Initially, the leaves of a tomato plant get affected, when the plant develops a particular type of disease.2020

2. **Mr .V Suresh ,D Gopinath,M Hemavarti ,K Jayathan and Mohana Krishnan** "Plant Disease Detection using Image Processing" this respective project presents the –Identification of plant diseases is the key to prevent the losses in the yield and quantity of the agricultural product. The studies of the plant diseases mean the studies of visually observable patterns seen on the plant. Health monitoring and disease detection on plant is very critical for sustainable agriculture. It is very difficult to monitor the plant diseases manually2020

**3.Shima Ramesh ,Mr.Ramchandra Hebbr,Nivedita M,Pooja R,Prasad Bhat and Shashank** "Plant Disease Detection Using Machine Learning "This paper makes use of Random Forest in identifying between healthy and diseased leaf from the data sets created. Our proposed paper includes various phases of implementation namely dataset creation, feature extraction, training the classifier and classification. The created datasets of diseased and healthy leaves are collectively trained under Random Forest to classify the diseased and healthy images. For extracting features of an image we use Histogram of an Oriented Gradient (HOG). Overall, using machine learning to train the large data sets available publicly gives us a clear way to detect the disease present in plants in a colossal scale.2018

**4. Vijai Singh and A.K. Misra** "Detection of plant leaf diseases using image segmentation and soft computing techniques" This paper presents an algorithm for image segmentation technique which is used for automatic detection and classification of plant leaf diseases. It also covers survey on different diseases classification techniques that can be used for plant leaf disease detection. Image segmentation, which is an important aspect for disease detection in plant leaf disease, is done by using genetic algorithm.2018

**5.B.Sai .Reddy and S.Neeraja** "Plant leaf disease classification and damage detection system using deep learning models" This paper presents an automatic plant leaf damage e detection and disease identification system. The first stage of the proposed method identifies the type of the disease based on the plant leaf image using DenseNet. The DenseNet model is trained on images categorized according to their nature, i.e., healthy and the type of the disease. This model is then used for testing new leaf images. The proposed DenseNet model produced a classification accuracy of 100%, with fewer images used during the training stage.2022

**6.Amrita S Tulshan And Nataasha Raul** "Plant Leaf Disease using Machine Learning "his process involved steps like image pre-processing, image segmentation, feature extraction. Furthur K Nearest Neighbor (KNN) classification is applied on the outcome of these three stages. Proposed implementation has shown 98.56% of accuracy in predicting plant leaf diseases. It also presents other information regarding a plant leaf disease that is Affected Area, Disease Name, Total Accuracy, Sensitivity and Elapsed Time.2019

7. **H.Al-Hiary, S. Bani-Ahmad, M.Reyalat, M.Braik and Z.AlRahamneh** Discussed various techniques to segment the diseased part of the plant. This paper also discussed some Feature extraction and classification techniques to extract the features of infected leaf and the classification of plant diseases. The use of ANN methods for classification of disease in plants such as self-organizing feature map, back propagation algorithm, SVMs, etc. can be efficiently used. From these methods, we can accurately identify and classify various

plant diseases using image processing techniques.

**8.Dae Gwan Kim, Thomas F. Burks, Jianwei Qin, Duke M.Bulanon**

An approach based on image processing is used for automated plant diseases classification based on leaf image processing the research work is concerned with the discrimination between diseased and healthy soybean leaves using SVM classifier. They have tested our algorithm over the database of 120 images taken directly from different farms using different mobile cameras. The SIFT algorithm enables to correctly recognize the plant species based on the leaf shape. The SVM classifier can help in recognizing normal and diseased soybean leaves with an average accuracy as high as 93.79%. The main aim of the proposed work is to provide inputs to an autonomous DSS which will provide necessary help to the farmers as and when required over the mobile. This system will provide help to the farmer with minimal efforts. The farmer only needs to capture the image of the plant leaf using a mobile camera and send it to the DSS, without any additional inputs.

## 3. METHODOLOGY

### 3.1 METHODOLOGY USED BY PREVIOUS RESEARCHER

The current approach for detecting plant disease is simple naked-eye observation by plant experts, which can be used to detect and identify plant diseases. In these circumstances, the suggested technique is useful for tracking vast fields of crops. Furthermore, in some nations, farmers lack adequate facilities or are unaware that they can contact experts. As a result, consulting experts is not only more expensive but also more time-consuming. In those circumstances, the suggested technique for tracking a large number of plants would be useful.

**DISADVANTAGES OF THE EXISTING SYSTEM**

- Only humans are capable of predicting diseases.
- The procedure is extremely slow.
- Consumption of time and space is also very high.
- The price is also high

### 3.2 METHODOLOGY USED IN THE PROPOSED SYSTEM

Digital camera or similar devices are use to take images of leafs of different types, and then those are used to identify the affected area in leafs. Then different types of image processing techniques are applied on them, to process those mages, to get different and useful features needed for the purpose of analyzing later.

Algorithm written below illustrated the step by step approach for the proposed image recognition and segmentation processes:

(1) Image acquisition is the very first step that requires capturing an image with the help of a digital camera.

(2) Preprocessing of input image to improve the quality of image and to remove the undesired distortion from the image. Clipping of the leaf image is performed to
get the interested image region and then image smoothing is done using the smoothing filter. To increase the contrast Image enhancement is also done.

(3) Mostly green colored pixels, in this step, are masked. In this, we computed threshold value that is used for these pixels. Then in the following way mostly green

pixels are masked: if pixel intensity of the green component is less than the pre-computed threshold value,then zero value is assigned to the red, green and blue
components of the this pixel.
(4) In the infected clusters, inside the boundaries, remove the masked cells.
(5) Obtain the useful segments to classify the leaf diseases.Segment the components using genetic algorithm For doing clustering appropriately, the search capability of
GAs can be used, to set of unlabeled points in N-dimension into K clusters. On image data, we have applied the same idea in our proposed scheme. We have taken a color image of size m n and every pixel has Red, Green and Blue components.Every chromosome shows a solution, which is a sequence of K cluster centers. Population is initialized in various rounds randomly and from existing chromosome best chromosome survives in each round for the next round processing. In the first step of fitness computation the dataset of pixel is clustered according to nearest respective cluster centers such that each pixel xi of color image is put into the respective cluster with cluster center zj for j = 1, 2, ..., K by the following
Equations If

$$\text{If} \quad ||x_i - z_j|| < ||x_i - z_l||,$$

$$i = 1, 2, \ldots, m \times n, \quad l = 1, 2, \ldots, K, \text{ and } p \neq j.$$

In the further step new cluster centers are obtained by calculating the mean of each pixel of the assigned clusters. The new center of cluster Zi is given by for the cluster Ci as:

$$Z_i(r, g, b) = \frac{1}{n_i} \sum_{x_j \in c_i} (x_j(r, g, b)) \quad i = 1, 2, \ldots k \tag{1}$$

Now the fitness function is computed by calculating Euclidean distance between the pixels and their respective cluster by using following equations

$$M = \sum M_i \tag{2}$$

$$M_i = \sum_{x_j \in c_i} |(x_j(r, g, b) - z_i(r, g, b))| \tag{3}$$

(6) Computing the features using color co-occurrence methodology
For feature extraction the method used is color cooccurrence method. It is the methodology in which both the texture and color of an image are considered, to come to the unique features, which shows that image. Over the traditional gray-scale representation, in the visible light spectrum, the use of color image features provides
an additional feature for image characteristic. There are three major mathematical processes in the color co-occurrence method. First, conversion of the RGB images of leaves is done into HIS color space representation. After completion of this process, to generate a color co-occurrence matrix, each pixel map is used, which results into three color co-occurrence matrices, one for each of H, S, I.Features called as texture features, which include Local homogeneity, contrast, cluster shade, Energy, and cluster prominence are computed for the H image as given in Eqs. (4)–(7).

$$\text{CONTRAST} = \sum_{i,j=0}^{N-1} (i,j)^2 C(i,j) \qquad (4)$$

$$\text{Energy} = \sum_{i,j=0}^{N-1} C(i.j)^2 \qquad (5)$$

$$\text{Local Homogeneity} = \sum_{i,j=0}^{N-1} C(i,j)/(1+(i-j)^2) \qquad (6)$$

## 4. SOFTWARE REQUIREMENT AND SPECIFICATION

### 4.1 Functional Requirements

**1.** A large amount of images(dataset)

2. CNN Algorithm

3. Tkinter

### 4.2 Software Quality Attribute

- Reliability

  system can continue to keep operating over time

- Portability

A software application can run on numerous platforms such as data portability, hosting, viewing, etc.,

- Testability

It shows how well the system or component facilitates performing tests to determine whether the predefined test criteria have been met.

- Scalability

A system can handle the demand for stress caused by increased usage without decreasing performance.

- Flexibility

A system can adapt to future changes

- Maintainability

A software application can maintain easily and support changes cost-effectively

- Supportability

It is the ability of a system that satisfies requirements and needs to identify and solve problems.

- Interoperability

Two or more systems can communicate or exchange data easily and use the data that has been exchanged.

- Performance

It is the ability of a system in the form of responsiveness to various actions within a certain period.

## 4.3 SYSTEM REQUIREMENTS

### 4.3.1 Software Requirements

1. Python
2. Python libraries like Python tensor flow
3. C library
4. **Python**

Python is a computer programming language often used to build websites and software, automate tasks, and conduct data analysis. Python is a general-purpose language, meaning it can be used to create a variety of different programs and isn't specialized for any specific problems. This versatility, along with its beginner-friendliness, has made it one of the most-used programming languages today. A survey conducted by industry analyst firm Red Monk found that it was the second-most popular programming language among developers in 2021

- What is Python used for?

Python is commonly used for developing websites and software, task automation, data analysis, and data visualization. Since it's relatively easy to learn, Python has been adopted by many non-programmers such as accountants and scientists, for a variety of everyday tasks, like organizing finances.

"Writing programs is a very creative and rewarding activity," says the University of Michigan and Coursera instructor Charles R Severance in his book *Python for Everybody*. "You can write programs for many reasons, ranging from making your living to solving a difficult data analysis problem to having fun to helping someone else solve a problem."

➢ **What can you do with Python?**

Some things include:

- Data analysis and machine learning
- Web development
- Automation or scripting

- Software testing and prototyping
- Everyday tasks

**4.3.2 Data analysis and machine learning**

Python has become a staple in data science, allowing data analysts and other professionals to use the language to conduct complex statistical calculations, create data visualizations, build machine learning algorithms, manipulate and analyze data, and complete other data-related tasks.

Python can build a wide range of different data visualizations, like line and bar graphs, pie charts, histograms, and 3D plots. Python also has a number of libraries that enable coders to write programs for data analysis and machine learning more quickly and efficiently, like Tensor Flow and Kera's.

**4.3.3 Using Tensor Flow in Python**

Tensor Flow provides all of this for the programmer by way of the Python language. Python is easy to learn and work with, and it provides convenient ways to express how high-level abstractions can be coupled together. Tensor Flow is supported on Python versions 3.7 through 3.10, and while it may work on earlier versions of Python it's not guaranteed to do so.

Nodes and tensors in Tensor Flow are Python objects, and Tensor Flow applications are themselves Python applications. The actual math operations, however, are not performed in Python. The libraries of transformations that are available through Tensor Flow are written as high-performance C++ binaries. Python just directs traffic between the pieces and provides high-level programming abstractions to hook them together.

High-level work in Tensor Flow—creating nodes and layers and linking them together—uses the Keras library. The Keras API is outwardly simple; a basic model with three layers can be defined in less than 10 lines of code, and the training code for the same takes just a few more lines of code. But if you want to "lift the hood" and do more fine-grained work, such as writing your own training loop, you can do that.

**4.2.4 Hardware Requirements And Specifications**

**4.2.4.1. Hardware Requirements**

1) Processer: core i3

2) RAM: 4GB

3) HDD: 256GB

**4.4 ANALYSIS MODELS**

The SDLC model presents a set of phases where each phase depends on the results of the preceding stage. Various SDLC models are considered for different projects based on their suitability to project conditions such as user requirements, project risks, cost, and development timeframe. A particular SDLC model may suit a particular project while at the same time other models may appear suitable for the elicited requirements but is it essential to consider trade-offs when choosing an SDLC model. A formal SDLC model theoretically consists of the following phases for developing and implementing computer software.

i. Planning
ii. Analysis
iii. Design
iv. Implementation
v. Testing
vi. Deployment and Evolution

**4.3.1 Planning and Requirement Analysis**: It's the first basic phase in which the need for a software product is analysed. The user requirements are gathered for development. The information gathered is used in preparation for an ideal project approach as well as feasibility examination from the cost-effective, functional and technical perspectives.

**4.3.2 Design**: The next phase of SDLC is Design. It is concerned with designing a basic structural framework that identifies the significant component of the product and the communication between these components.

**4.3.3 Implementation**: During this stage, the actual process of product development is executed. The product meets the SRS produced.

**4.3.4 Deployment**: In this phase, we verify and validate our system.

**4.3.5 Evolution**: It is the last phase of SDLC. Once the software is ready after the completion of testing, it is deployed into the runtime environment. Based on the feedback the product may enhance further development which is called maintenance.

**Need of SDLC**

The development team must determine a suitable life cycle model for a particular plan and then observe to it.

Without using an exact life cycle model, the development of a software product would not be in a systematic and disciplined manner. When a team is developing a software product, there must be a clear understanding among team representative about when and what to do. Otherwise, it would point to chaos and project failure. This problem can be defined by using an example. Suppose a software development issue is divided into various parts and the parts are assigned to the team members. From then on, suppose the team representative is allowed the freedom to develop the roles assigned to them in whatever way they like. It is possible that one representative might start writing the code for his part, another might choose to prepare the test documents first, and some other engineer might begin with the design phase of the roles assigned to him. This would be one of the perfect methods for project failure.

A software life cycle model describes entry and exit criteria for each phase. A phase can begin only if its stage-entry criteria have been fulfilled. So without a software life cycle model, the entry and exit criteria for a stage cannot be recognized. Without software life cycle models, it becomes tough for software project managers to monitor the progress of the project.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

## 5. PROJECT DESIGN-MODULE WISE

### 5.1 BLOCK DIAGRAM:

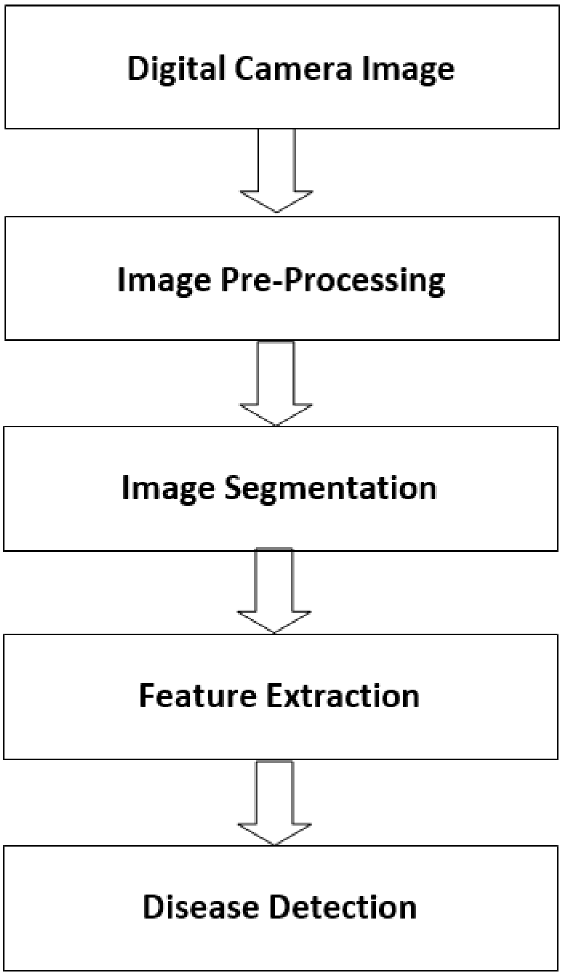



Fig 5.1 Block Diagram

**WORKING:**

**Digital Camera:**

The plant disease detection process includes four steps, as illustrated in Figure 1. The first step is the acquisition of images from cameras and mobile phones or websites. The second step divides the image into multiple groups and different techniques can be applied. The next step consists of the extraction process and the final step deals with the classification of the virus.

**Image Acquisition**

In the meantime, images of sheets are collected using digital media (camera, mobile phone, etc.) at the desired resolution and size Creating image databases is entirely up to application system developers The database image is responsible for further processing of the class in the final stage of detection.

**Image Segmentation**

This step aims to simplify the representation of the image so that it is useful and is easy to identify This step is the basis of the feature extraction and image processing process. Images can be segmented using various methods such as k-means clustering, Otsu's algorithm, and thresholding. k-Means Clustering classifies objects or pixels into K classes based on criteria. Classification is done by reducing to the sum of the squares of the distances between objects and their corresponding clusters.

**Feature Extraction**

Hence, in this step, the features from this area of interest need to be extracted. These features are needed to determine the meaning of a sample image. Features can be based on color, shape, and texture. Recently, most researchers are intending to use texture features for the detection of plant diseases. There are various methods of feature extraction that can be employed in developing the system such as gray-levels-occurrence matrix I (GLCM), color co-occurrence method, spatial grey-level independence matrix, and histogram-based feature extraction. The GLCM method is a statistical method of texture classification.

**Classification**

The classification phase implies determining if the input image is healthy or diseased. If the image is found to be diseased, some existing works have further classified it into a number of diseases. For classification, a software

routine is required to be written in MATLAB, also referred to as a classifier. A number of classifiers have been used in the past few years by researchers such as k-nearest neighbor (KNN), support vector machines (SVM), artificial neural network(ANN), back propagation neural network (BPNN), Naïve Bayes and Decision tree classifiers. The most commonly used classifier is found to be SVM. Every classifier has its advantages and disadvantages, SVM is simple to use and a robust technique.

**Disease Classifications:**

It is the method of using our qualified deep learning model to recognize plant disease. A digital camera or equivalent system should be used to take an image of the contaminated plant's leaf. Opencv was used to scan the image. Then it determines what kind of plant it is. It determines what kind of disease the plant has after finding it.

Gain (%) = number of correct classification/ Total no of test images *100

**Disease detection:**
In this process the Disease of the Leaf is detected and also we can classify the Disease by its Percentage in slight Disease, moderate Disease, and heavy Disease.

**5.2 PROCEDURE**

Plant disease detection using machine learning, specifically Convolutional Neural Networks (CNN), is a popular and effective approach for the automated identification of plant diseases. CNNs are well-suited for image-based tasks, making them suitable for analyzing plant images and detecting diseases based on visual patterns.

Here's a high-level overview of the process:

1. Dataset Collection: Gather a large dataset of plant images, including healthy plants and plants affected by various diseases. These images should cover different plant species and disease types. It's essential to ensure the dataset is diverse and representative of the target plant population.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

2. Data Preprocessing: Preprocess the collected images to enhance their quality and prepare them for training the CNN. Common preprocessing steps include resizing the images to a consistent size, normalizing pixel values, and augmenting the dataset by applying transformations like rotation, flipping, and cropping. Data augmentation helps to increase the model's robustness and generalization capability.

3. Model Training: Build and train a CNN model using the preprocessed dataset. The CNN consists of multiple layers, including convolutional layers, pooling layers, and fully connected layers. The convolutional layers are responsible for learning local patterns and features from the input images, while the fully connected layers combine the extracted features and make predictions. During training, the model learns to minimize the difference between its predicted outputs and the ground truth labels using a loss function and an optimization algorithm like stochastic gradient descent.

4. Model Evaluation: Evaluate the trained model's performance on a separate validation or test set to assess its accuracy and generalization ability. Metrics such as accuracy, precision, recall, and F1-score can be used to quantify the model's performance.

5. Deployment and Inference: Once the model is trained and evaluated, it can be deployed to detect plant diseases in real-world scenarios. New plant images can be fed into the trained model, and it will output predictions indicating whether the plant is healthy or affected by a specific disease.

It's worth noting that the success of the CNN model for plant disease detection relies on the availability of a comprehensive and accurately labeled dataset. Additionally, continuous updates and improvements can be made by incorporating new data and periodically retraining the model.

This overview provides a general idea of how plant disease detection using CNNs works. Implementing a complete system requires further details on the CNN architecture, hyperparameter

Tuning, and fine-tuning for specific plant species and diseases.

**5.2 .1 FLOW CHART**

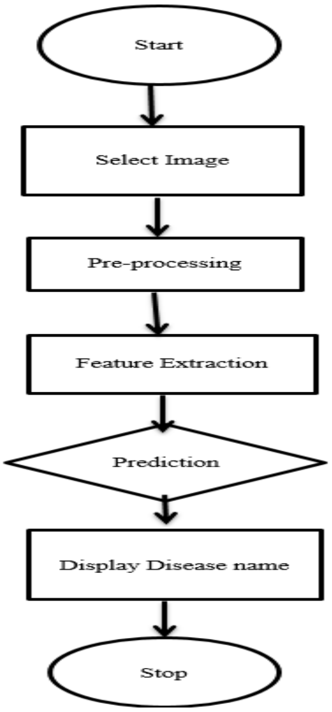


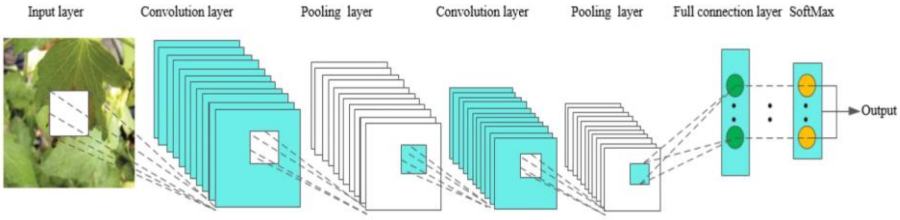

Fig 4.2.1 Flow Chart

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

### 5.2.2 CNN

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other

A convolutional neural network is a feed-forward neural network that is generally used to analyze visual images by processing data with grid-like topology. It's also known as a ConvNet. A convolutional neural network is used to detect and classify objects in an image.




- **Layers in a Convolutional Neural Network**

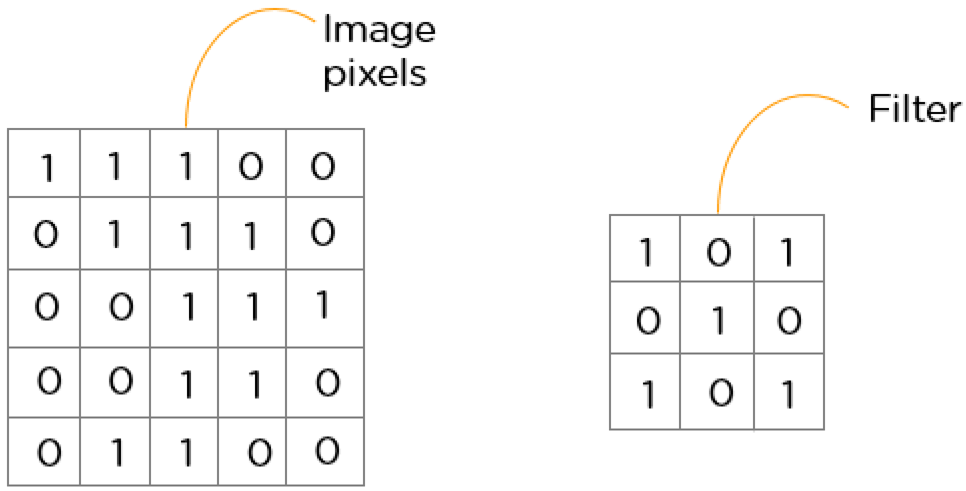
A convolution neural network has multiple hidden layers that help in extracting information from an image. The four important layers in CNN are:

1. Convolution layer
2. ReLU layer
3. Pooling layer
4. Fully connected layer

- Convolution Layer

This is the first step in the process of extracting valuable features from an image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values.

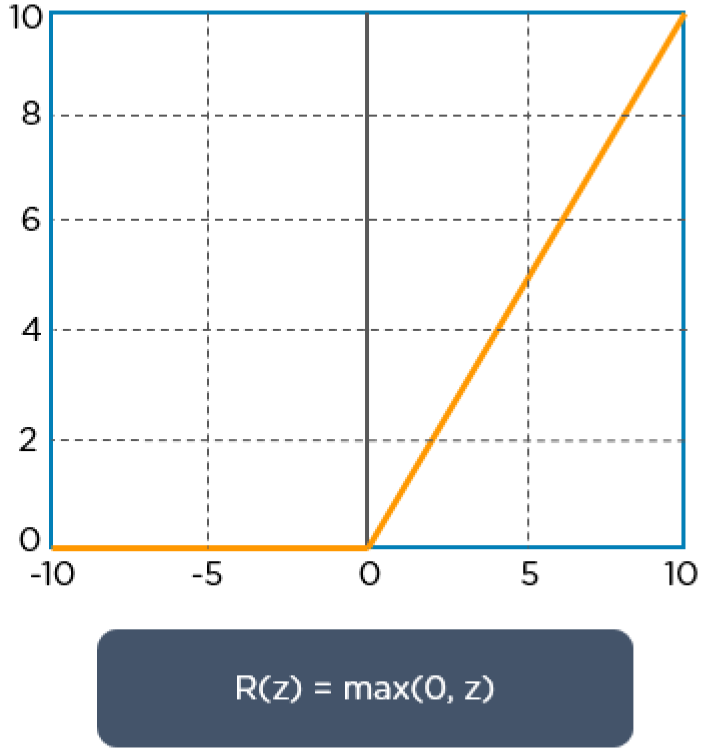
Consider the following 5x5 image whose pixel values are either 0 or 1. There's also a filter matrix with a dimension of 3x3. Slide the filter matrix over the image and compute the dot product to get the convolved feature matrix.




ReLU layer

ReLU stands for the rectified linear unit. Once the feature maps are extracted, the next step is to move them to a ReLU layer.

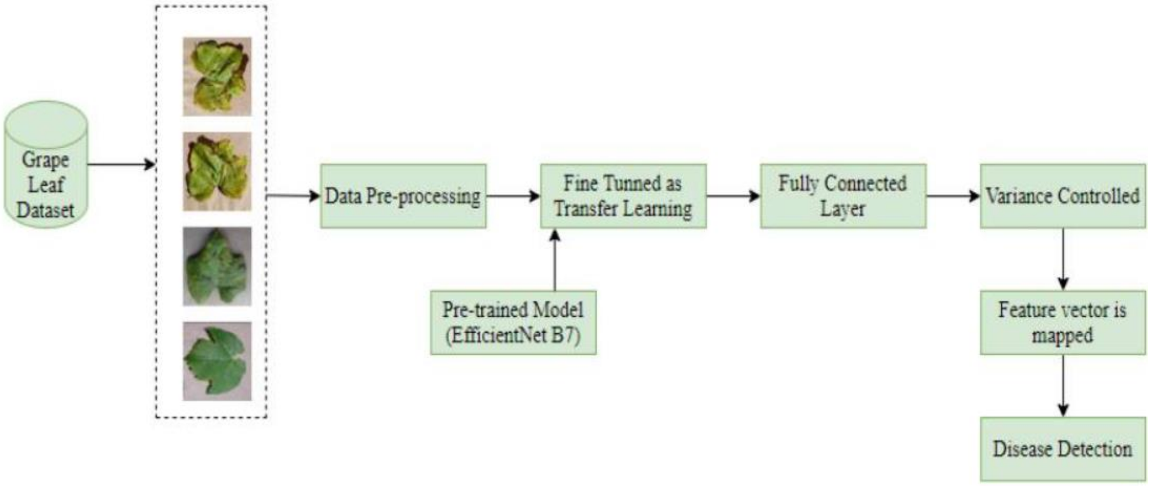
ReLU performs an element-wise operation and sets all the negative pixels to 0. It introduces non-linearity to the network, and the generated output is a rectified feature map. Below is the graph of a ReLU function:




The original image is scanned with multiple convolutions and ReLU layers for locating the features.

Here's how exactly CNN recognizes a leaf disease :

- The pixels from the image are fed to the convolutional layer that performs the convolution operation
- It results in a convolved map
- The convolved map is applied to a ReLU function to generate a rectified feature map
- The image is processed with multiple convolutions and ReLU layers for locating the features
- Different pooling layers with various filters are used to identify specific parts of the image
- The pooled feature map is flattened and fed to a fully connected layer to get the final output




**5.2.3 DataSet:**

For training and testing purposes, we used the standard open-access PlantVillage dataset which consists of 54,305 numbers of healthy- and infected-plant leaves. Detailed database information, the number of classes and images in each class, their common and scientific names, and the disease-causing viruses are shown in and . The database contains 38 different classes of 14 different plant species with healthy- and disease-affected-leaf images. All images were captured in laboratory conditions. **Figure** shows some sample leaf images from the PlantVillage datasets

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

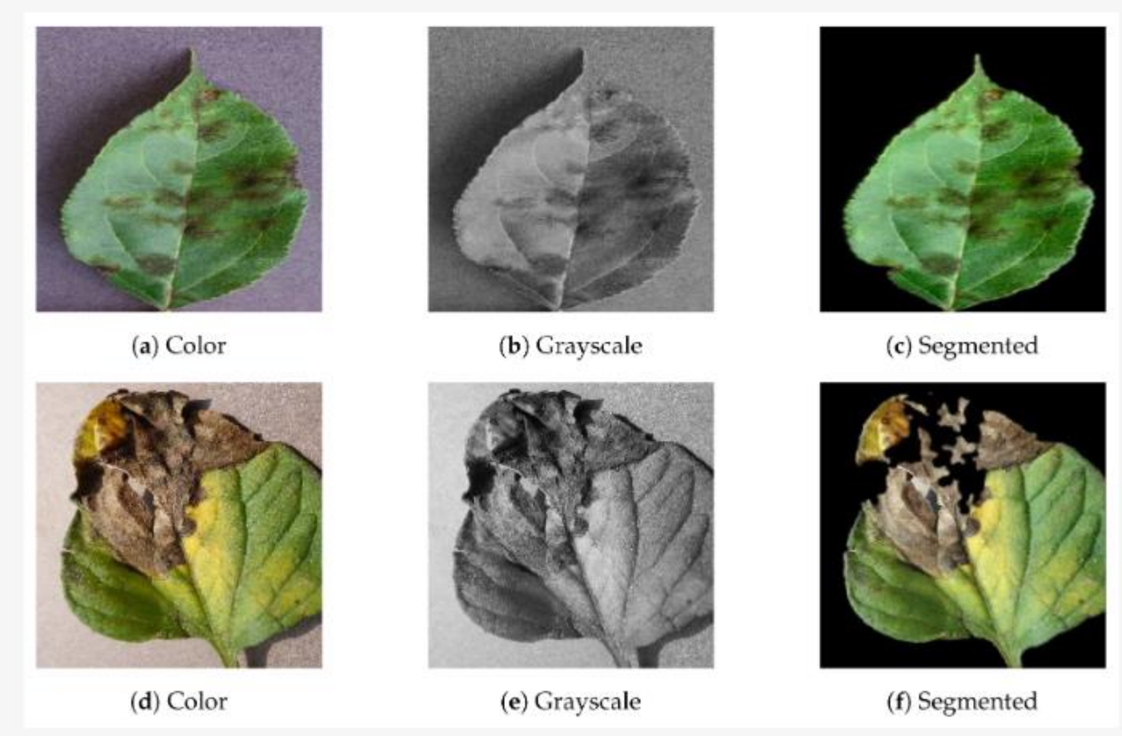



**Table .** Detailed description of PlantVillage dataset with relative information.

| Class | Plant Name | Disease Name | Causes Virus Name | Type of Disease | No of Images |
|---|---|---|---|---|---|
| C1 | Tomato | Healthy | | | 1591 |
| C2 | Tomato | Bacterial spot | Xanthomonas perforans | Bacterial | 2127 |
| C3 | Tomato | Early Blight | Alternaris sp | fungal | 1000 |
| C4 | Tomato | Late blight | Phytophthora infestans | fungal | 1909 |
| C5 | Tomato | Leaf Mold | Lycopersicon | fungal | 952 |
| C6 | Tomato | Septoria leaf spot | Septoria Lycoprsici | fungal | 1771 |
| C7 | Tomato | Spider mites | Tetranychus spp | pest | 1676 |
| C8 | Tomato | Target Spot | Corynespora Cassicola | fungal | 1404 |
| C9 | Tomato | Tomato mosaic virus | Tomato mosaic | viral | 373 |
| C10 | Tomato | Tomato Yellow Leaf | Begomovirus | viral | 5357 |

In our experiment, we used three different formats of PlantVillage datasets. First, we ran the experiment with colored leaf images, and then with segmented leaf images of the same dataset. In the segmented images, the background was smoothed, so that it could provide more meaningful information that would be easier to analyze. Lastly, we used grayscale images of the same dataset to evaluate the performance of the implemented methods. All leaf images were divided into two sets, a training set and the testing set. To evaluate performance, we split leaf images into three different sets, namely 80–20 (80% training images and 20% testing images), 70–30 (70% training images and 30% testing images), and 60–40 (60% training images and 40% testing images).

## 6.PROJECT IMPLEMENTATION

Project implementation (or project execution) is the phase where visions and plans become reality. This is the logical conclusion, after evaluating, deciding, visioning, planning, applying for funds and finding the financial resources of a project.

### 6.1 ALGORITHM DETAILS

#### 6.1.1 CNN Algorithm

A Convolutional Neural Network (CNN) is a type of Deep Learning neural network architecture commonly used in Computer Vision. Computer vision is a field of Artificial Intelligence that enables a computer to understand and interpret the image or visual data.

When it comes to Machine Learning, Artificial Neural Networks perform really well. Neural Networks are used in various datasets like images, audio, and text. Different types of Neural Networks are used for different purposes, for example for predicting the sequence of words we use Recurrent Neural Networks more precisely an LSTM, similarly for image classification we use Convolution Neural networks. In this blog, we are going to build a basic building block for CNN.

In a regular Neural Network there are three types of layers:

i**nput Layers:** It's the layer in which we give input to our model. The number of neurons in this layer is equal to the total number of features in our data (number of pixels in the case of an image).

**Hidden Layer:** The input from the Input layer is then feed into the hidden layer. There can be many hidden layers depending upon our model and data size. Each hidden layer can have different numbers of neurons which are generally greater than the number of features. The output from each layer is computed by matrix multiplication of output of the previous layer with learnable weights of that layer and then by the addition of learnable biases followed by activation function which makes the network nonlinear.

**Output Layer:** The output from the hidden layer is then fed into a logistic function like sigmoid or softmax which converts the output of each class into the probability score of each class.

The data is fed into the model and output from each layer is obtained from the above step is called feedforward, we then calculate the error using an error function, some common error functions are cross-entropy, square loss error, etc. The error function measures how well the network is performing. After that, we backpropagate into the model by calculating the derivatives. This step is called Backpropagation which basically is used to minimize the loss.

### 6.2.ADVANTAGES:

Plant disease detection using machine learning (ML) offers several advantages compared to traditional methods. Here are some key advantages:

1. Early Detection: ML algorithms can analyze large volumes of plant data and detect disease symptoms at an early stage. This allows for prompt intervention and treatment, minimizing the spread of diseases and reducing crop losses.

2. Accuracy: ML models can be trained to accurately identify and classify different plant diseases. They can recognize subtle patterns and variations that may be difficult for human observers to detect. This leads to more reliable and consistent diagnoses.

3. Speed and Efficiency: ML algorithms can process plant images or sensor data quickly, enabling rapid detection and diagnosis. This efficiency is particularly crucial for large-scale agricultural operations, where timely disease identification can prevent extensive damage.

4. Scalability: ML-based disease detection systems can be scaled up to handle large datasets and monitor vast areas of crops. This scalability makes it suitable for both small-scale farming and large commercial agricultural operations.

5. Cost-Effective: Implementing ML-based disease detection can potentially reduce the costs associated with manual labor, extensive field surveys, and expert consultations. It streamlines the process by automating the detection and diagnosis, resulting in cost savings for farmers.

6. Real-Time Monitoring: ML algorithms can continuously monitor plants in real-time, using sensors or image processing techniques. This allows for early detection of disease outbreaks, enabling farmers to take immediate action.

**6.3 DISADVANTAGES**

While there are several advantages to using machine learning (ML) for plant disease detection, there are also some potential disadvantages to consider:

1. Data Requirements: ML algorithms require large amounts of accurately labeled training data to perform effectively. Obtaining such data can be challenging, especially for rare or newly emerging plant diseases. Creating and maintaining a comprehensive dataset can be time-consuming and costly.

2. Limited Generalization: ML models trained on specific datasets may struggle to generalize well to new or unseen plant diseases. If a new disease or variant emerges, the model may not have the necessary information to accurately detect or classify it. Regular updates and retraining are necessary to ensure the model's effectiveness.

3. Dependency on Data Quality: The accuracy and reliability of ML-based disease detection heavily depend on the quality and diversity of the training data. If the training dataset is biased, incomplete, or contains mislabeled samples, it can negatively impact the model's performance and lead to false positives or false negatives.

4. Interpretability and Explainability: ML models, especially complex ones like deep neural networks, can be challenging to interpret and explain. The lack of transparency in decision-making can make it difficult for farmers or agricultural experts to understand why a particular diagnosis or prediction was made. This can hinder trust and adoption of ML-based systems.

5. Hardware and Infrastructure Requirements: Implementing ML-based disease detection may require substantial computational resources, including powerful hardware and storage capacity. This can pose a challenge for farmers or organizations with limited access to such resources, particularly in remote or resource-constrained areas.

6. Maintenance and Upkeep: ML models require regular maintenance, updates, and retraining to ensure their performance remains optimal over time. Additionally, as new diseases or variants emerge, the models may need to be adjusted or retrained to accommodate these changes. This ongoing maintenance can add complexity and cost to the system.

**6.4 FUTURE SCOPE:**

Plant disease detection using machine learning (ML) techniques has a promising future scope in agriculture and plant science. ML algorithms have the potential to revolutionize the way we detect and diagnose plant diseases, leading to improved crop management and increased agricultural productivity. Here are some of the future prospects for plant disease detection using ML:

1. Increased accuracy and speed: ML algorithms can be trained on large datasets of plant images, including healthy and diseased plants. As these algorithms continue to improve, they are likely to provide more accurate and rapid detection of plant diseases. This will enable farmers to take timely actions to prevent the spread of diseases and minimize crop losses.

2. Real-time disease monitoring: With advancements in sensor technologies and the Internet of Things (IoT), it will be possible to integrate ML-based disease detection models with smart farming systems. This integration will enable real-time monitoring of plant health, allowing farmers to receive immediate alerts and take proactive measures against diseases.

3. Automation and robotics: ML can be combined with robotics and automation to develop intelligent systems capable of autonomously detecting and treating plant diseases. Robotic systems equipped with cameras and ML algorithms can scan large agricultural fields, identify diseased plants, and apply targeted treatments. This can significantly reduce the need for manual labor and improve efficiency in crop management.

4. Disease prediction and forecasting: ML models can analyze various environmental and biological factors to predict the occurrence and spread of plant diseases. By utilizing historical data, weather patterns, and crop conditions, these models can generate disease risk assessments and provide early warnings to farmers. This proactive approach will help farmers make informed decisions regarding disease prevention and crop protection.

5. Mobile applications for farmers: Mobile applications powered by ML algorithms can be developed to provide on-the-spot disease diagnosis and treatment recommendations to farmers. Farmers can capture images of affected plant parts using their smartphones and upload them to the application. The ML model running in the backend can analyze the images and provide instant insights, including disease identification, severity assessment, and suggested management practices.

6. Transfer learning and data sharing: ML models trained on diverse datasets can be utilized to detect multiple diseases across different plant species. By applying transfer learning techniques, ML models can leverage knowledge gained from one plant disease to detect similar patterns in other plants. Moreover, data sharing platforms can be established to encourage collaboration among researchers and farmers, allow in for the creation of more comprehensive and robust disease detection models.

7. Integration with precision agriculture: ML-based disease detection can be seamlessly integrated with precision agriculture techniques, such as variable rate applications of fertilizers and pesticides. This integration will enable targeted and precise treatments, minimizing the use of agrochemicals and reducing environmental impact.

Overall, the future of plant disease detection using ML looks promising. As technology continues to advance, and more data becomes available, ML algorithms will become increasingly accurate, efficient, and accessible to farmers worldwide. This will lead to improved plant health management, increased crop yields, and sustainable agricultural practices.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

## 7. RESULT AND CONCLUSION

### 7.1RESULT:

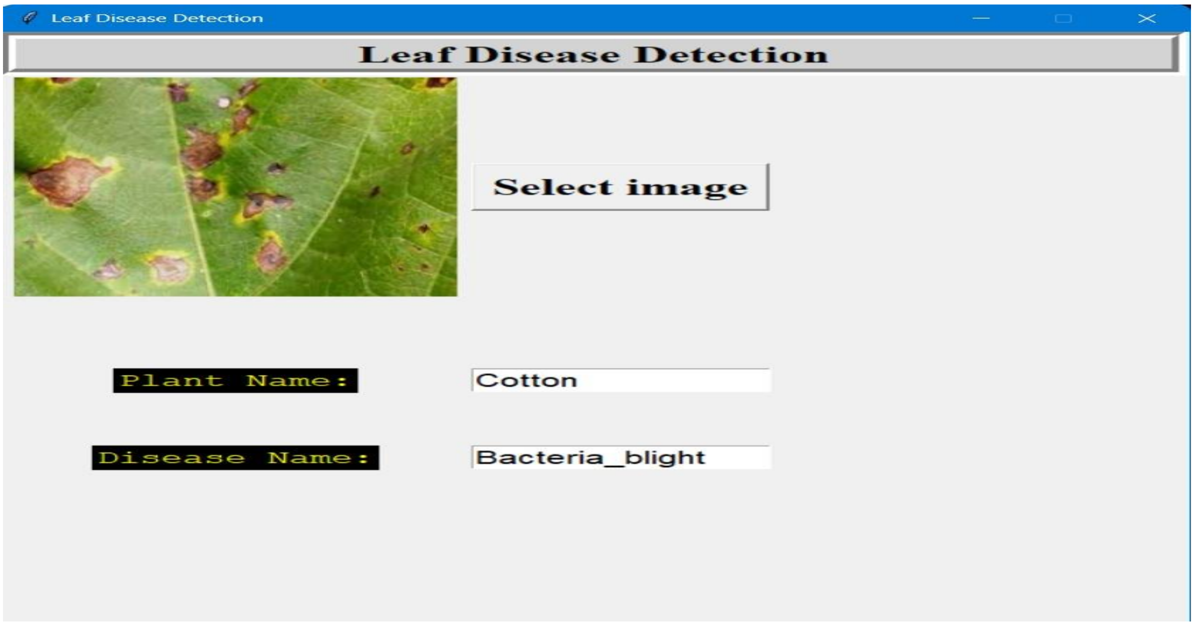



Fig 7.1 .1


ffffff

ffff Fig 7.1 .2

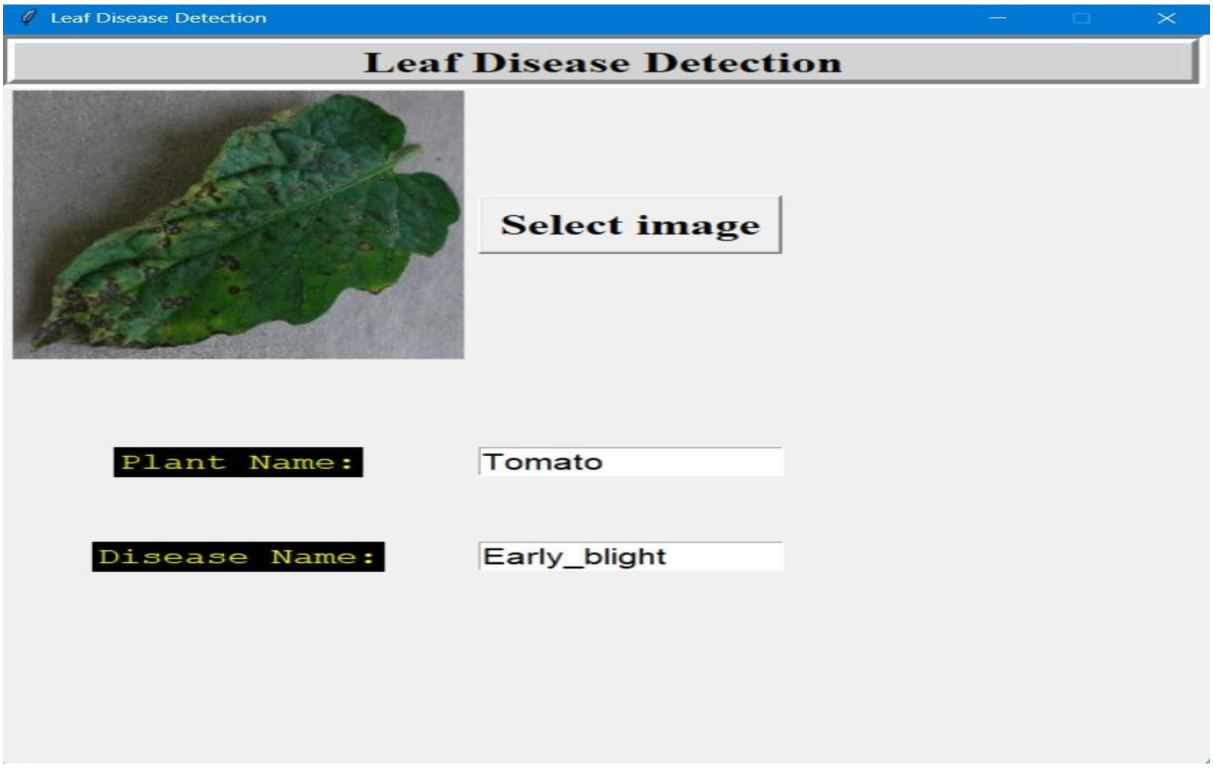



Fig 7.1.3

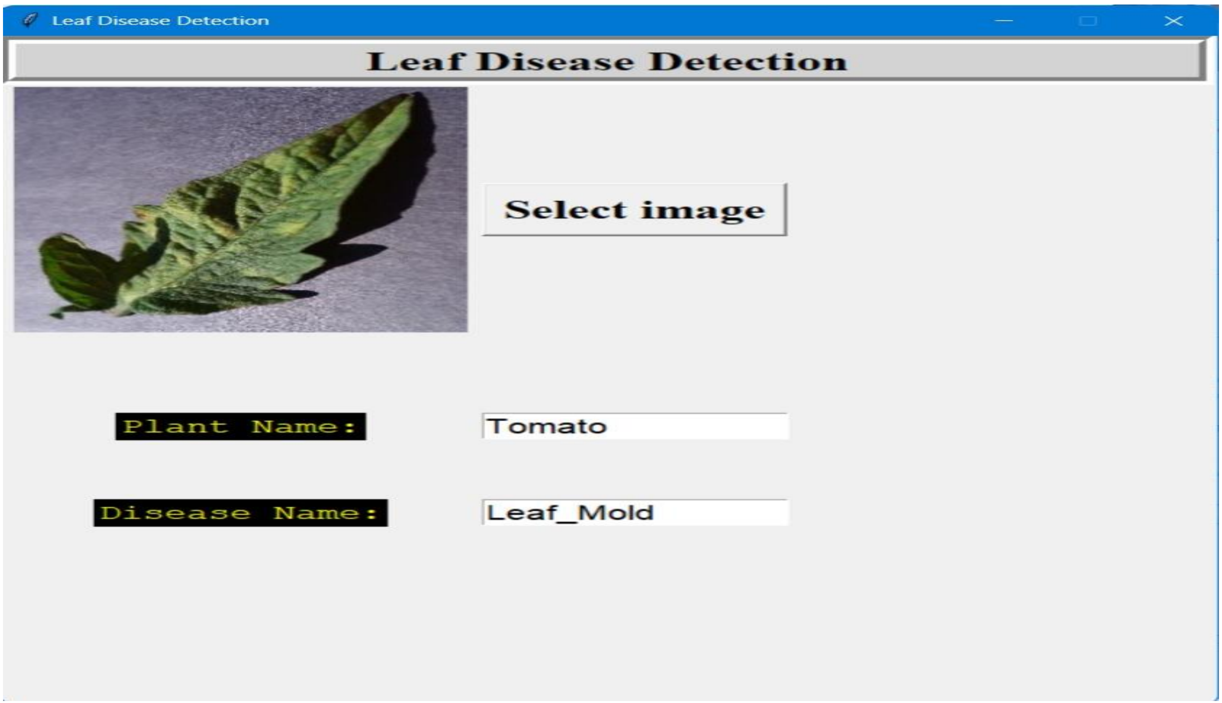



Fig 7.1.4

### 7.2 CONCLUSION:

The proposed system tracks the cultivated field on a regular basis. The CNN and DNN algorithms are used to identify crop diseases at an early stage. Machine learning methods are used to train the model, which aids in making appropriate disease decisions. To contain infected diseases, the farmer is advised to use pesticides as a cure. In the future, the proposed scheme could be expanded to provide additional facilities such as nearby government markets, pesticide price lists, and a nearby open market, among others..This paper presents a review of various disease classification strategies for crop disease detection, as well as an algorithm for image segmentation that can be used for automated detection and classification of plant leaf diseases in the future. Some of the organisms on which the proposed algorithm is evaluated include banana, beans, jackfruit, lemon, mango, potato, tomato, and sapota. As a result, similar diseases for these plants were investigated. The best results were obtained with very little computational effort, demonstrating the efficacy of the proposed algorithm in recognising and classifying crop diseases. Another benefit of this approach is that plant diseases can be detected at an early stage, or even at the beginning. Convolution neural network and Deep neural network algorithms may be used to increase recognition rates in the classification process.

### 8.REFERENCES


- [1] S. S. Sannakki and V. S. Rajpurohit," Classification of Pomegranate Diseases Based on Back Propagation Neural Network," International Research Journal of Engineering and Technology (IRJET), Vol2 Issue: 02 | May-2015
- [2] P. R. Rothe and R. V. Kshirsagar," Cotton Leaf Disease Identification using Pattern Recognition Techniques", International Conference on Pervasive Computing (ICPC),2015.
- [3] Aakanksha Rastogi, Ritika Arora and Shanu Sharma," Leaf Disease Detection and Grading using Computer Vision Technology &Fuzzy Logic" 2nd International Conference on Signal Processing and Integrated Networks (SPIN)2015.
- [4] Godliver Owomugisha, John A. Quinn, Ernest Mwebaze and James Lwasa," Automated Vision-Based Diagnosis of Banana Bacterial Wilt Disease and Black Sigatoka Disease ", Preceding of the 1'st international conference on the use of mobile ICT in Africa ,2014.
- [5] uan Tian, Chunjiang Zhao, Shenglian Lu and Xinyu Guo," SVM-based Multiple Classifier System for Recognition of Wheat Leaf Diseases," Proceedings of 2010 Conference on Dependable Computing (CDC'2010), November 20-22, 2010.
- [6] S. Yun, W. Xianfeng, Z. Shanwen, and Z. Chuanlei, "Pnn based crop disease recognition with leaf image features and meteorological data," International Journal of Agricultural and Biological Engineering, vol. 8, no. 4, p. 60, 2015.
- [7] J. G. A. Barbedo, "Digital image processing techniques for detecting, quantifying and classifying plant diseases," Springer Plus, vol. 2, no.660, pp. 1–12, 2013.
- [8] Caglayan, A., Guclu, O., & Can, A. B. (2013, September). "A plant recognition approach using shape and color features in leaf images." In International Conference on Image Analysis and Processing (pp. 161-170). Springer, Berlin, Heidelberg.
- [9] Zhen, X., Wang, Z., Islam, A., Chan, I., Li, S., 2014d. "Direct estimation of cardiac bi-ventricular volumes with regression forests." In: Accepted by Medical Image Com- puting and Computer-Assisted Intervention– MICCAI 2014.
- [10] Wang P., Chen K., Yao L., Hu B., Wu X., Zhang J., et al. (2016)." Multimodal classification of mild cognitive impairment based on partial least squares Sankaran, S.; Mishra, A.; Ehsani, R. "A review of advanced techniques for detecting plant diseases". Comput.Electron. Agric. 2010, 72, 1–13. [CrossRef]