# Plant Disease Detection using CNN

## 1. Introduction

Plant diseases pose a significant threat to agricultural productivity, quality, and sustainability. Early and accurate detection of plant diseases is crucial to ensure timely intervention and to minimize crop losses. With the advancement in artificial intelligence and computer vision, deep learning techniques, especially Convolutional Neural Networks (CNNs), have shown great promise in the field of plant disease detection.

## 2. Goal

The goal is to develop a CNN-based model that can accurately detect and classify plant diseases from leaf images of various crops such as apple, cherry, grape, and corn. The model should be capable of distinguishing between healthy and diseased leaves and should identify the specific type of disease present, if any. This will enable early diagnosis and precise disease management strategies, thereby supporting precision agriculture practices.

## 3. Objectives

The primary objectives of this project are:

* - To collect and preprocess a comprehensive dataset of healthy and diseased leaf images from different crops.
* - To design and implement a CNN-based deep learning model for image classification.
* - To train the model using labeled data and optimize it for high accuracy and efficiency.
* - To evaluate the model's performance using standard metrics such as accuracy, precision, recall, and F1-score.
* - To deploy the trained model for real-time prediction of plant diseases.

## 4. Significance

This system will play a crucial role in enhancing agricultural productivity by enabling farmers to identify diseases at an early stage. By automating the disease detection process, it reduces the dependency on manual inspection and expert consultation, which can be time-consuming and costly. Moreover, it supports sustainable agriculture by facilitating targeted pesticide application, thereby reducing environmental impact.

## 5. Scope

The scope of this project includes building a model that works with leaf images of selected crops (apple, cherry, grape, corn). It is limited to disease detection based on visual symptoms observable on the leaf surface. Future enhancements may include expanding the system to other crops, integrating with mobile applications for field use, and incorporating real-time image acquisition through drone or camera feeds.

**PipeLine:**

1. **Data Collection and data loading**:

The success of a CNN-based plant disease detection model heavily relies on the quality and diversity of the dataset used for training and evaluation. In this project, data collection involves sourcing high-resolution images of plant leaves exhibiting both healthy and diseased conditions across various crops, including apple, cherry, grape, and corn. Publicly available datasets such as the PlantVillage dataset can serve as a primary source, as they offer thousands of labeled images categorized by crop type and disease class. Additionally, to enhance the model’s generalizability, images can be collected from agricultural fields using digital cameras or smartphones under different lighting conditions and backgrounds.

1. **Dataset:**

**Train:**The training dataset forms the backbone of the CNN model’s learning process. It consists of a large and diverse set of labeled images of plant leaves from crops such as apple, cherry, grape, and corn, each annotated with its corresponding health status and disease type. To ensure the model learns robust features, the dataset includes a wide variety of images with different lighting conditions, angles, resolutions, and background settings. The dataset is typically split into training, validation, and testing subsets, with around 70–80% of the data allocated for training.

**Valid:**The validation dataset is a crucial component used to evaluate the CNN model’s performance during training. It typically comprises 10–15% of the entire dataset and includes images that the model has not seen during the training phase. These images are used to monitor the model's generalization ability, tune hyperparameters, and prevent overfitting.

**Test:**The test dataset plays a vital role in objectively evaluating the performance of the final CNN model after training is complete. It typically comprises 10–15% of the total dataset and contains images that the model has never encountered during either training or validation.

1. **Zip File:**

Here we will have a zip file on the lms portal and download it and again upload into google drive.

1. **Image processing and Image Augmentation:**

Image processing is a crucial step in preparing raw leaf images for training a CNN model. It involves cleaning and transforming images into a consistent format suitable for model input. Typical preprocessing steps include resizing all images to a standard resolution (e.g., 224x224 pixels), converting color formats (e.g., RGB), and normalizing pixel values to improve training stability and convergence. Additionally, noise removal and contrast enhancement may be applied to highlight disease-related features.  
To improve model generalization and prevent overfitting, image augmentation techniques are employed during training. Augmentation artificially increases the size and variability of the dataset by applying random transformations to the original images.

1. **CNN Model:**

The Convolutional Neural Network (CNN) is a deep learning architecture specifically designed for image recognition and classification tasks. In the context of plant disease detection, a CNN model is highly effective due to its ability to automatically learn spatial hierarchies of features from input leaf images. The CNN typically consists of several layers: convolutional layers that apply filters to detect features like edges, spots, and textures; pooling layers that reduce spatial dimensions and computation; and fully connected layers that interpret the extracted features to make predictions.

**6.Test/evaluate**:

In AI and Machine Learning, **test and evaluate** refer to the process of assessing a model's performance after it has been trained on data. Testing involves applying the model to a separate dataset that it hasn't seen during training to check how well it generalizes to new data. Evaluation uses various metrics—such as accuracy, precision, recall, F1 score for classification tasks, or Mean Squared Error for regression—to measure how well the model performs. This step helps in identifying areas for improvement, guiding adjustments to the model, its parameters, or the data used. The process is iterative, ensuring that the model is optimized for better real-world application.