**Crop Disease Diagnosis using Convolutional Neural Network**

# Solving for India –

# College of Engineering,Pune

**Team Members**

1)Apurva Deshpande

2)Divya Gaygol

3)Pooja Gulhane

4)Sakshi Mehetre

## Abstract

In India two third of the population i.e approximately 70% relies on agriculture for their livelihood. Agriculture sector occupies about 20.5% of total GDP (Gross Domestic Product) in India. Crop Diseases cause yield loss up to 70% causing major problems for the farmers impacting their income. Potential reason behind this is farmers unable to identify disease on the crops. Traditionally, crop disease detection has been done through visual inspec- tion by trained experts. These traditional methods include macroscopic and microscopic examination, serological tests, etc. But these methods require specific knowledge, skills, and equipment and can be time-consuming and labor-intensive.

With advancement in technology, there are new methods like machine learning and com- puter vision to detect the disease in crops which are faster and more efficient. In recent years there have been models like ANN (Artificial Neural Network), KNN (K-Nearest Neighbors) developed for crop disease detection. ANN requires a lot of computational resources to train and can be prone to overfitting. Whereas KNN can have difficulty in handling highly non- linear data as well as it can be computationally expensive while handling large datasets. The results of these models can be affected by presence of outliers and noise in the data. To overcome these flaws, we propose a model of ResNet (Residual Network) which uses CNN (Convolutional Neural Network). ResNet is a model that achieves higher accuracy compared to shallower models like ANN and KNN. It is able to handle large datasets using a residual learning strategy. Also ResNet is able to handle noise and outliers as well as overfitting by using skip connections and residual blocks. All these benefits make ResNet a more powerful model for crop disease detection especially when data is scarce, noisy and high dimensional.

Our aim is to develop an application using the model that we have implemented. This app will take leaf image as an input and it will predict crop disease based on that image.

**Introduction**

Agriculture plays a crucial role in India, where two-thirds of the popula- tion, approximately 70%, rely on it for their livelihood [10]. The agriculture sector contributes approximately 20.5% to the total Gross Domestic Product (GDP) of the country. However, crop diseases pose a major challenge for farmers, causing yield losses of up to 70% [10]. This has a significant impact on the income of farmers and is often a result of the inability to identify crop diseases. Traditionally, crop disease detection was performed through visual inspection by trained experts, using methods such as macroscopic and mi- croscopic examination, serological tests, and others. These methods can be time-consuming and labor-intensive, requiring specific knowledge and skills, as well as specific equipment.

With advancements in technology, new methods for crop disease detection have emerged, including machine learning and computer vision. Artificial Neural Network (ANN) and K-Nearest Neighbors (KNN) are examples of models that have been developed for crop disease detection. However, ANN requires a lot of computational resources to train and is prone to overfitting [9][14]. KNN, on the other hand, can have difficulty handling highly non- linear data, and can also be computationally expensive when working with large datasets [7]. The results of these models can be affected by the presence

of outliers and noise in the data, making it difficult to achieve high accuracy. To overcome these limitations, we propose a Residual Network (ResNet) model that uses Convolutional Neural Network (CNN). ResNet is a more powerful model for crop disease detection compared to shallower models like ANN and KNN, as it is able to handle large datasets and overcome overfitting through residual learning and skip connections [10]. Additionally, it can handle noise and outliers effectively. Thus, ResNet is the preferred model for crop disease detection due to its high accuracy and ability to handle large

datasets, noise and outliers, and to prevent overfitting.

Our goal is to develop an application that utilizes the ResNet model to de- tect crop diseases. The application will take leaf images as input and predict the disease based on that image. Our aim is to make the application user- friendly. This application will be of great benefit to farmers, as it will help them to improve their crop yields, reduce crop loss, and increase their prof- its. Additionally, the application could be used by agricultural organizations and government bodies to monitor crop health on a larger scale, providing insights into disease trends and helping to prevent widespread outbreaks. By leveraging technology, we can help to make agriculture more sustainable and secure, benefiting both farmers and the wider community.

**Research Gaps and Problem Statement**

## Research Gap

Based on existing reports we found out that there is lack of a compre- hensive and generalized model that can accurately detect and classify crop diseases across different types of crops using a large and diverse dataset. Most of the models listed in our study have limitations in terms of the crop types they can detect diseases for, the less size of the dataset used, or their abil- ity to only classify crops as healthy or unhealthy without predicting specific diseases. A potential research focus could be on developing a more versatile and robust model that can be applied to various crop species and trained on a larger and more diverse dataset to improve accuracy and generalization.

## Problem Statement

To explore and make a comprehensive and generalized model with the help of CNN that can effectively detect and classify diseases across a specific range of local crop species using large and diverse dataset.

**Proposed Methodology**

## Dataset

We used the publicly available PlantVillage dataset for our crop disease diagnosis model. The dataset was curated by S. P. Mohanty and it consists of 87,000 RGB images of healthy and unhealthy crop leaves with 38 different classes. We narrowed down our experimentation to 25 classes. The selected 25 classes, which are detailed in Table 4.1 covers a range of specific local crop species and diseases. To ensure an accurate representation of various types of crop diseases, a broad set of selected classes provides a dataset for training and evaluation of our model.

Table 4.1: Crop Disease Dataset

|  |  |  |
| --- | --- | --- |
| **Crop** | **Disease** | **No. of Images** |
| Tomato | Healthy | 1926 |
|  | Bacterial Spot | 1702 |
|  | Early Blight | 1920 |
|  | Late Blight | 1851 |
|  | Leaf Mold | 1882 |
|  | Septoria Leaf Spot | 1745 |
|  | Two-Spotted Spider Mite | 1741 |
|  | Target Spot | 1827 |
|  | Yellow Leaf Curl Virus | 1961 |
|  | Mosaic Virus | 1790 |
| Potato | Healthy | 1824 |
|  | Early Blight | 1939 |
|  | Late Blight | 1939 |
| Apple | Healthy | 2008 |
|  | Scab | 2016 |
|  | Black Rot | 1987 |
|  | Cedar Apple Rust | 1760 |
| Grape | Healthy | 1692 |
|  | Black Rot | 1888 |
|  | Esca | 1920 |
|  | Leaf Blight | 1722 |
| Corn | Healthy | 1859 |
|  | Gray Leaf Spot | 1642 |
|  | Common Rust | 1907 |
|  | Northern Leaf Blight | 1908 |

## Data preprocessing and Feature Extraction

ResNet model requires preprocessed data because the input data needs to be in a specific format before it can be used by the model. This typically involves resizing, cropping, and normalizing the images. The significance of feature extraction in ResNet is that it allows the model to learn more abstract features from the input images. By extracting high-level features, the model can better understand the content of the images and make more accurate predictions. This can be especially important for complex tasks where image dataset is used.

Accordingly,preprocessing is especially important in crop disease diagnosis from leaf images using ResNet model. To achieve accurate results, the input images first need to be preprocessed. Initially, the images are resized to a fixed size of 256 x 256 pixels, to maintain consistency in the dataset. Then, the images are normalized to improve the performance of the model. After normalization, data augmentation techniques such as flipping, rotating, and cropping are applied to increase the size of the dataset and to introduce varia- tion in the dataset. This helps in improving the performance of the model and prevents overfitting. The preprocessed images are then fed into the ResNet model for feature extraction. The ResNet model is a deep learning model that can extract complex features from the images. The extracted features are then passed through a fully connected layer, which is used for classifi- cation. The classification layer consists of softmax activation that outputs the probability of the image belonging to a particular class. After training the model, it is tested on the test dataset to evaluate the accuracy of the model.By performing these preprocessing steps, the ResNet model can learn to accurately classify images of plant diseases.

## ResNet Model

ResNet is a machine learning architecture designed to tackle the prob- lem of vanishing gradients that often arises in deep neural networks. ResNet introduces a unique structure called a ”residual block”, which enables the network to learn an identity mapping between layers. A residual block con- tains two convolutional layers followed by a batch normalization layer and a ReLU activation function. The output of the second convolutional layer is added to the input of the block, creating a shortcut connection that allows the gradients to flow directly from the output of one layer to the input of an- other. The ResNet architecture can be extremely deep and typically contains multiple stages of residual blocks, each stage containing a varying number of blocks with increasing numbers of filters. The final stage is followed by a global average pooling layer, a fully connected layer, and a softmax activation function for classification. ResNet’s architecture has been shown to achieve state-of-the-art performance on a variety of image classification tasks.

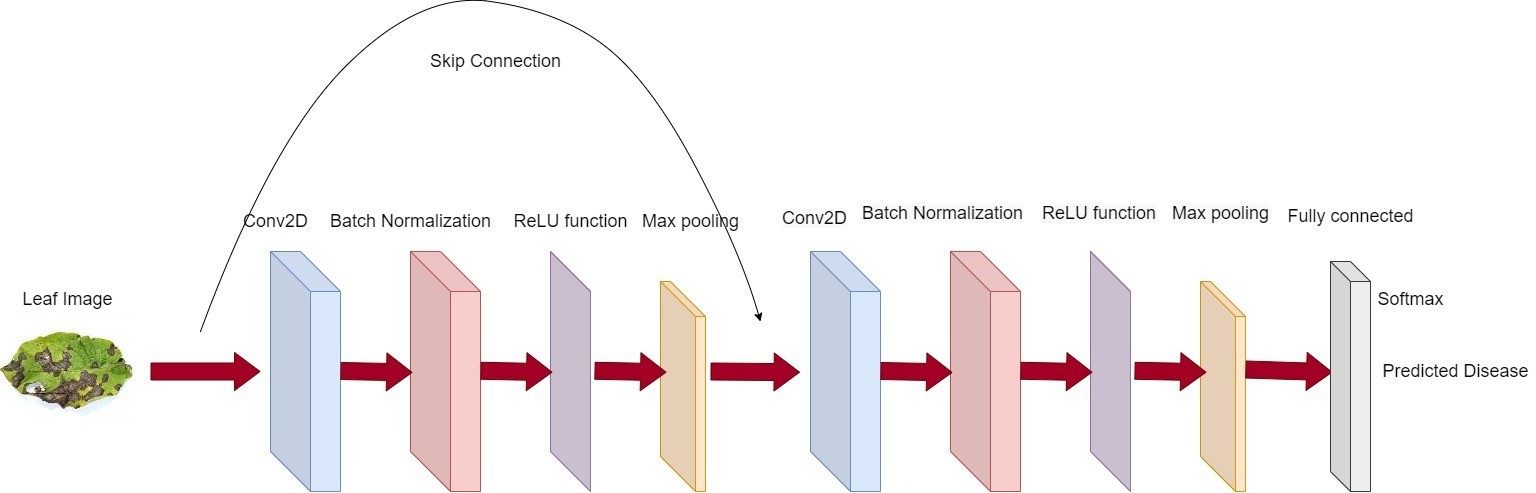


Figure 4.1: ResNet Block Diagram

## Project Plan

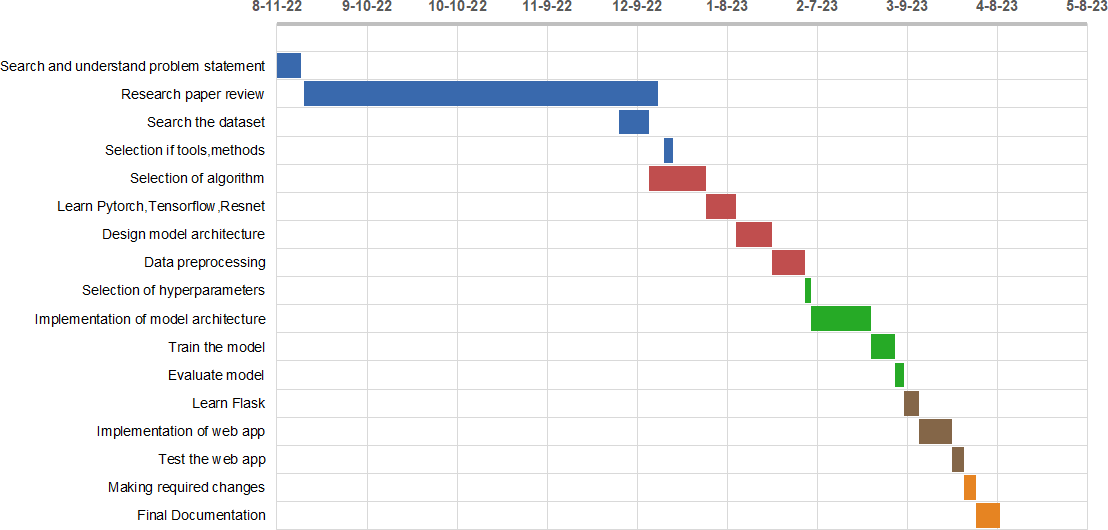


Figure 4.2: Project Plant Chart

# Experimental Setup

## System Architechture

**Architecture**

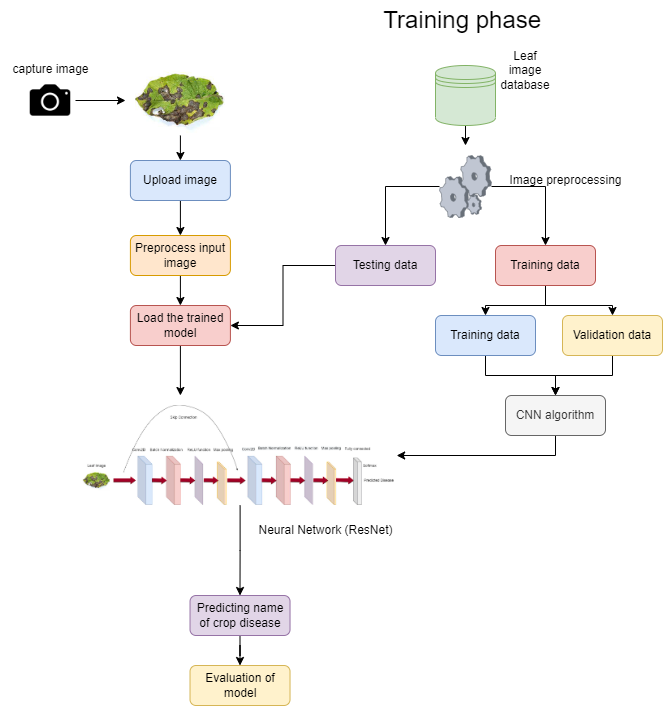


Figure 5.1: System Architechture

**Usecase Diagram**

In this usecase diagram, there are two actors - the user and the system. The user initiates the process by capturing an image of the leaf, which is then uploaded to the system. The system then preprocesses the input image and performs image segmentation to isolate the affected leaf. Next, the system detects the affected leaf and extracts its features. Based on the extracted features, the system classifies the disease and displays the name of the disease to the user.

In this system, the user is involved in the first two use cases - capturing and uploading the image. The remaining six use cases - preprocessing, seg- mentation, detection, feature extraction, disease classification, and displaying the name of the disease to the user are performed by the system.

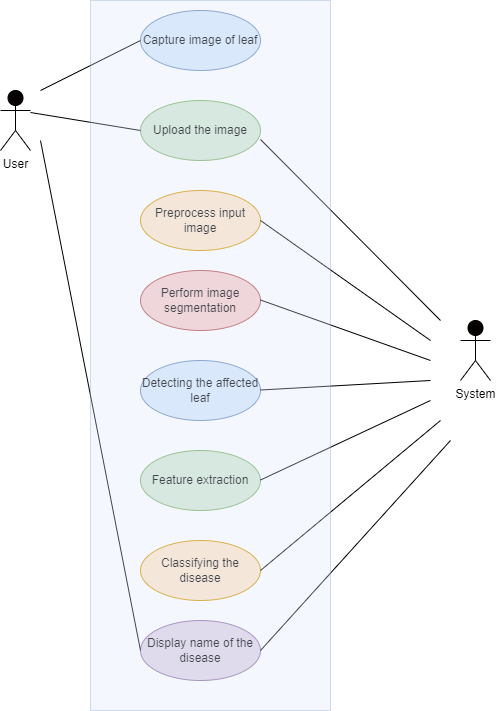


Figure 5.2: Usecase Diagram for the system

**Sequence Diagram**

In this sequence diagram, the user initiates the process by providing an input image to the system. The system then performs the methods of the system in sequence:

*Preprocess the image → Segment the image → Extract Feature →*

*Train model*

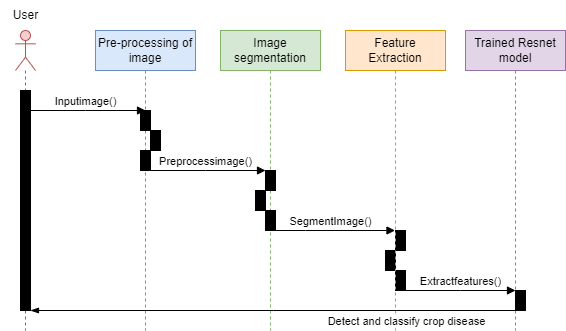


Figure 5.3: Sequence Diagram for the system

In Figure 5.3 it is shown that messages like Inputimage(), Preprocessim- age(), SegmentImage() and Extractfeatures() are used to define communica- tion between user and objects in the system.

## System Requirement Specifications

### Description

**Product Perspective**

The objective of this product is to develop a Crop Disease Diagnosis System that can discover and extract hidden information associated with crop diseases from a dataset. This system aims to exploit machine learning techniques, specifically the ResNet model, on agricultural data sets to assist in predicting crop diseases.

There are large volumes of records in the agricultural data domain, even- tually making it necessary to use machine learning techniques to help in decision support and prediction in agriculture. Therefore, it contributes to business intelligence and helps farmers diagnose crop diseases early on.

Some specific objectives of the system are:

* + - * Enable early detection of crop diseases to prevent widespread crop dam- age.
      * Reduce the cost of labor and resources associated with manual detection methods.
      * Help farmers save time and increase productivity.
      * Improve the accuracy of crop disease diagnosis, leading to more effective treatment plans.

**Product Function**

1. Data collection: The dataset for this system is collected from Kaggle or other crop pathology databases. The dataset consist of images of healthy and diseased crop. The dataset can be augmented by rotating, flipping, and scaling the images to increase the variety of data.
2. Data cleaning: Data cleaning involves removing irrelevant or duplicate data, resizing images to a standard size, and converting the data to a format that can be used by the ResNet model. This step also involves splitting the dataset into training and testing sets to assess the perfor- mance of the model.
3. Model development: The ResNet model can be used for crop disease diagnosis by training the model on the cleaned dataset. The model can identify patterns and features in the images that are indicative of healthy or diseased crops. The model can be fine-tuned by adjusting the hyperparameters and adding more layers to improve the accuracy of the predictions.
4. Predictions: Once the model is trained, it can be used to predict whether a crop is healthy or diseased along with the disease name based on its image. The model can classify images with high accuracy, allowing for early detection of crop diseases and preventing further spread of the dis- ease. This can help farmers and researchers identify crop diseases more accurately and efficiently, leading to better crop yields and healthier crops.

### Operating Conditions

For a project on crop disease diagnosis using ResNet model, the following technocal requirements may be needed:

**Software Requirements**

* + - * Operating System: Windows (10 and 11), Linux
      * Programming Language: Python 3
      * Platform: Google Colab
      * Libraries: Python libraries such as Numpy, pandas, pytorch, matplotlib

**Hardware Requirements**

* + - * Processor: 2.5+ GHz processor with a minimum of dual core.
      * RAM: 8 GB minimum

### Non-Functional Requirements

**Reliability**: For a crop disease diagnosis application using the ResNet model, it is important to have a reliable and strong structure. Changes made by the programmer should be reviewed. Bugs or errors discovered by users should be resolved promptly.

**Maintenance**: To ensure efficient performance of the application, the sys- tem monitoring and maintenance should be easy and straightforward. There should be no excess jobs running on different machines, and the application should be designed to make updates and maintenance easier.

**Results and Discussion**

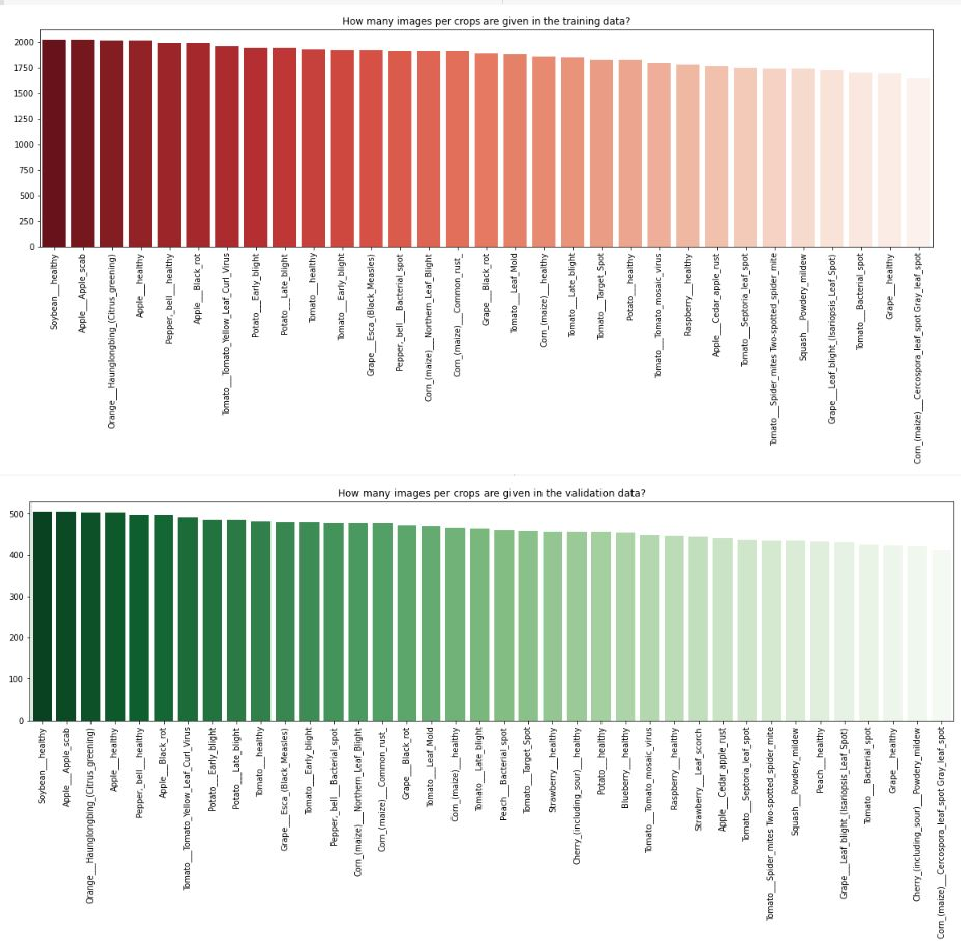
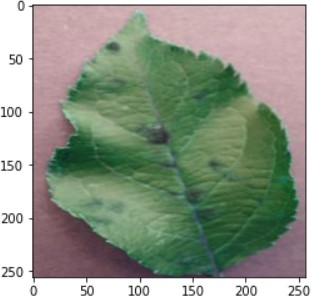
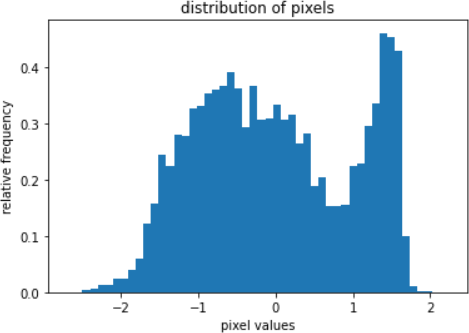
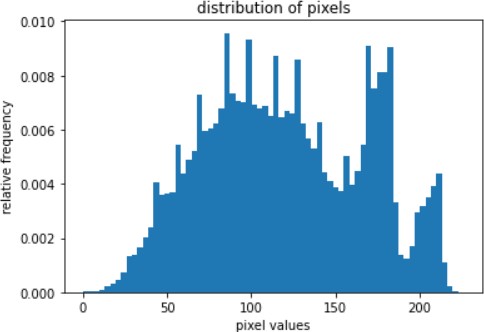


Figure 6.1: Training and Validation Data

(a) Original Image (b) Normalized Image

Figure 6.2: Apple Leaf



(a) Original Image (b) Normalized Image

Figure 6.3: Pixel Distribution

# Conclusion

This project provides valuable insights into the use of machine learning techniques for crop disease diagnosis. The classifier plays a critical role in the agriculture industry, enabling accurate and efficient prediction of crop diseases. Various machine learning techniques are studied and compared to identify the most effective systems. By improving the accuracy of crop disease prediction, these techniques can help farmers identify diseased crops in their early stages and implement preventive measures.

The model’s predictions can be used as a basis for developing measures to prevent crop disease. Similar models can be developed to study other crops and help farmers better serve their communities with analytical and classification techniques.

In the future, this project could be enhanced by integrating other com- puter vision and machine learning techniques to improve model accuracy and versatility. Additionally, the application could be scaled to include a larger database of crop diseases and to be made available to farmers in various re- gions in their regional languages to improve agriculture in those areas. By leveraging these techniques, more innovative solutions could be developed for farmers, leading to a more sustainable and profitable agriculture industry.