32DETECTION OF CROP DISEASE USING MACHINE LEARNING

ENROLLMENT- 20103275, 20103281, 20103297

NAME OF STUDENTS- VRITI DAWRA, SRISHTI KASHYAP, KUSH KAPOOR

NAME OF THE SUPERVISOR- DR. BHARAT GUPTA

MINOR PROJECT 2022

PANEL MEMBERS: DR. ASHISH MISHRA

DR. APEKSHA AGGARWAL

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, NOIDA

TABLES OF CONTENTS

TABLE OF CONTENTS

INDEX

(II)

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Place: Signature:

Date: 02 December 2022

Name: Vriti Dawra
Srishti Kashyap
Kush Kapoor

Enrollment No: 20103275 (B10)
20103281 (B10)
20103297 (B10)

(III)

CERTIFICATE

This is to certify that the work titled “Detection Of Crop Disease Using Machine Learning ” submitted by “Vriti Dawra, Srishti Kashyap, Kush Kapoor” in partial fulfilment for the award of degree of B.Tech. of Jaypee Institute of Information Technology, Noida has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Signature of Supervisor	……………………....
Name of Supervisor	Dr. Bharat Gupta
Designation	Assistant Professor (SENIOR GRADE)
Date	02 December 2022

(IV)

ACKNOWLEDGEMENT

We deem it a pleasure to acknowledge our sense of gratitude to our project guide Dr. Bharat Gupta under whom we have carried out the project work. His incisive and objective guidance and timely advice encouraged us with constant flow of energy to continue the work.

We wish to reciprocate in full measure the kindness shown by Dr. Bharat Gupta Assistant Professor (SENIOR GRADE) who inspired us with his valuable suggestions in successfully completing the project work.

Finally, we must say that no height is ever achieved without some sacrifices made at some end and it is here where we owe our special debt to our parents and our friends for showing their generous love and care throughout the entire period of time.

Signature of Supervisor	………………………
Name of Supervisor	Dr. Bharat Gupta
Designation	Assistant Professor (SENIOR GRADE)
Date	02 December 2022

(V)

ABSTRACT

The major agricultural products in India are rice, wheat, pulses, and spices. As our population is increasing rapidly the demand for agriculture products also increasing alarmingly. A huge amount of data are incremented from various field of agriculture. Analysis of this data helps in predicting the crop yield, analysing soil quality, predicting disease in a plant, and how meteorological factor affects crop productivity. Crop protection plays a vital role in maintaining agriculture product. Pathogen, pest, weed, and animals are responsible for the productivity loss in agriculture product. Machine learning techniques like Random Forest, Bayesian Network,Decision Tree, Support Vector Machine etc. help in automatic detection of plant disease from visual symptoms in the plant. A survey of different existing machine learning techniques used for plant disease detection was presented in this paper. Automatic detection of disease in plant helps in early diagnosis and prevention of disease which leads to an increase in agriculture productivity.

(VI)

KEYWORDS

1. Crop Disease
2. Image Acquisition
3. Image Processing
4. Segmentation
5. Feature Extraction
6. Deep Learning
7. Classification
8. Deployment

CHAPTER 1
INTRODUCTION

1.1 General introduction

Agriculture is a very ancient methodology of food source. It is a very important source of economy all over the world . Without food no one can survive in this world. Plants are important not only humans, even animals rely on plants for their food, oxygen etc., To increase the production of crops, the government and researchers are taking immense steps which are successfully running in the real world. Once a plant gets infected with any kind of disease, then the entire living things will get affected in one or other form [9]. This plant disease may occur in any part of the plant like stem, leaf, branch and so on. And even the type of diseases affecting the plants may also vary like fungal, bacterial diseases etc.,
The disease that affects the crops will depend on factors like climatic conditions. There are a huge number of people who are starving for food. This happens because of insufficient production in food crops. Even drastic climatic changes will affect the growth of plants . This kind of natural disaster cannot be helped.

Identifying the plant disease in early stages helps in preventing huge loss of crops. Farmers are supposed to use proper pesticides for their crops [9]. Too much chemicals (pesticides) are dangerous for the crops as well as the farming land. Getting experts' advice will help from exceeding use of chemicals on plants. Huge research has been done on the plants to help farmers and also people interested in the agricultural field [10]. Disease detection is easy when it is visible to a human's naked eye [3]. Once the farmer has enough knowledge and continuous monitoring of the crops, then the disease may be detected and treated earlier. But this stage exists only when the disease is severe or the production of crops is less [6]. When the number of crops increased or the production is high then this type of detecting disease won't work. Then arises the new techniques. Automated techniques for diseases detection are introduced which makes the farmer's life easier [1]. The results produced in this kind of technique are efficient for both small and large production of crops. Importantly the results are accurate and take very less time to detect the diseases. Machine learning and neural networks play a vital role for these technologies [5]. In this paper, Deep Convolutional Neural Network (CNN) is used to classify the diseased and healthy leaves and to detect the disease in the affected leaves. The CNN model is built to fit both healthy and diseased leaves, the images are trained in the model and the output will be produced according to the input leaf. Following this section 2 will explain the dataset collection, and section 3 explains the image pre-processing, section 4 will explain about training and testing data and finally section 5 explains the Deep learning model which is used to detect the diseases in the plant leaf. The methodology and the results obtained are shown in the experimental results.

1.2 Problem Statement

Outbreak of plant life-form diseases significantly harm the quantity of production as well as reduce the quantity if agriculture products which leads to massive economic loss . Crop losses due to diseases and weeds across the world have increased from about 34.9% in 1965 to 42% in 2021. The conventional approach of disease detection and plant maintenance is human intervention in the majority of agriculture around the world. Agriculture loss forecasts show inefficiency of this approach due to variables such as human error. Thus we aim to establish a software solution for plant leaf disease detection using image processing which can both increase the accuracy of work and reduce the need for human workforce.

1.3 Significance of the problem

The overall agriculture productivity of a country plays a huge role in its economic infrastructure. The aim of our study is to design a system which will implement a machine learning based solution for automatic detection of plant leaf disease using image processing technology. We plan to develop a sophisticated system which will detect, identify and classify various categories of disease in plant life-form to a degree which will provide a realistic significance. Crop disease significantly decrease the qualitly and quantity of production of plants . Studies have shown that early detection of plant disease can significantly reduce the chances if wide-scale outbreak which in turn will minimize losses fir all parties involved in the production of a certain plant product. Detection of such diseases via automatic software instead of manual human supervision reduces a large portion of human work and drastically increase effectiveness

1.4 Proposed System

Deep Learning techniques are inspired by the architecture of neurons present in the human brain (Haykin, 1998). These techniques use Artificial Neural Networks (ANNs), and its other variants, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to identify the hidden structures in data. There are two prominent advantages of Deep Learning techniques over the Machine Learning techniques. First, they automatically extract various features from raw data, and hence there is no need for an extra feature extraction module. Second, Deep Learning techniques reduce the amount of time required to process large datasets of high dimensions. Therefore, the Deep Learning techniques are used to build the proposed hybrid model.

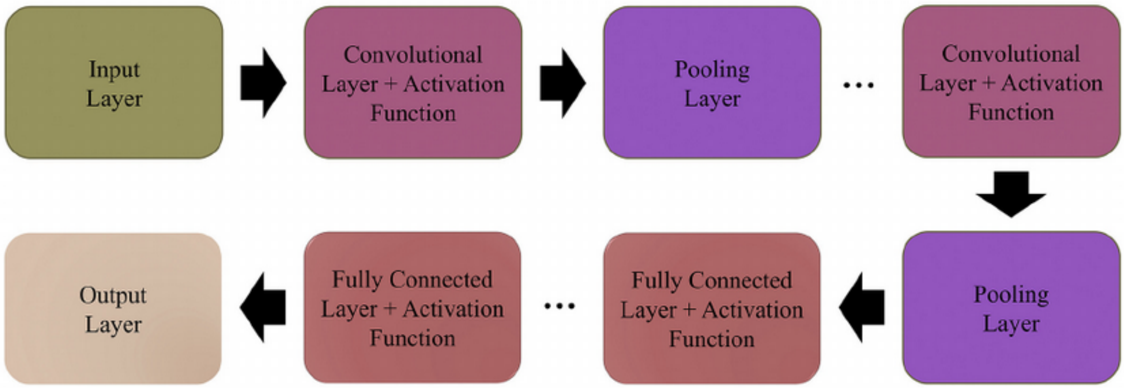


Fig. 1. The architecture of a typical CNN.

Convolutional Neural Networks (CNNs) and Convolutional Autoencoders (CAEs) are two Deep Learning techniques used in many computer vision applications due to their effectiveness on image data. Both these techniques use convolution operation to extract various spatial and temporal features from image data. CNNs are used to classify input images to their respective classes, whereas CAEs are used to reduce the dimensionality of an image efficiently.

1.5 Contribution of the Project

This initiative will capture and measure the global impacts of crop pests and disease, putting a much-needed spotlight on crop health and ensuring that money and goodwill are directed towards the real, evidence-based, causes of crop loss. With accurate and relevant information, decision-makers can allocate resources and systematically develop investment in, and capacity of, plant health systems. Overall, this project has the potential to transform global agriculture and serve as a cornerstone for agricultural policy decision-making.

CHAPTER 2
LITERATURE SURVEY

2.1 Summary of papers studied

In agriculture, due to climatic conditions, there are drastic changes and different types of diseases are affecting the crops. There are a lot more ways to detect the disease in the plant. To detect the plant leaf diseases, the considered dataset taken is the combination of both healthy and diseased leaves of different crops. The methods such as deep learning techniques and computer vision are used for the implementation in this paper. The neural network that is used to detect the disease and train the model is chosen as Convolution neural network (CNN).

In the paper **—"Deep learning for Image-Based Plant detection" [1]** the authors Prasanna Mohanty et al., have proposed an approach to detect disease in plants by training a convolutional neural network. The CNN model is trained to identify healthy and diseased plants of 14 species. The model achieved an accuracy of 99.35% on test set data. When using the model on images procured from trusted online sources, the model achieves an accuracy of 31.4%, while this is better than a simple model of random selection, a more diverse set of training data can aid to increase the accuracy. Also some other variations of model or neural network training may yield higher accuracy, thus paving path for making plant disease detection easily available to everyone.

Malvika Ranjan et al. in the paper **—"Detection and Classification of leaf disease using Artificial Neural Network"** proposed an approach to detect diseases in plant utilising the captured image of the diseased leaf. Artificial Neural Network (ANN) is trained by properly choosing feature values to distinguish diseased plants and healthy samples. The ANN model achieves an accuracy of 80%.

According to the paper **—" Detection of the unhealthy region of plant leaves and classification of plant leaf diseases using texture features" [3]** by S. Arivazhagan, the disease prediction process includes four main steps as follows: first, a color transformation structure taken for the input RGB image, and then by means of a specific threshold value, the green pixels are detected and uninvolved, which is followed by segmentation process, and for obtaining beneficial segments the texture statistics are computed. At last, classifier is used for the features that are extracted to classify the disease.

Kulkarni et al. in the paper **—"Applying image processing technique to detect plant diseases" [4]**, a methodology for early and accurately plant diseases detection, using artificial neural network (ANN) and diverse image processing techniques. As the proposed approach is based on ANN classifier for classification and Gabor filter for feature extraction, it gives better results with a recognition rate of up to 91%.

In paper **—"Plant disease detection using CNN and GAN" [5]**, by Emaneul Cortes, an approach to detect plant disease using Generative Adversarial networks has been proposed. Background segmentation is used for ensuring proper feature extraction and output mapping. It is seen that using Gans may hold promise to classify diseases in plants, however segmenting based on background did not improve accuracy.

In the paper **—"Convolutional Neural Network based Inception v3 Model for Animal Classification" [6]**, Jyotsna Bankar et al. have proposed use of inception v3 model in classifying animals in different species. Inception v3 can be used to classify objects as well as to categorize them, this capability of inception v3 makes it instrumental in various image classifiers.

2.2 Integrated summary of the literature studied

2.2.1 Image Segmentation

In the complex environment, the most crucial task is how to segment the images while localizing and detecting diseased plant leaves, since the major aim of image segmentation is to set the symptom information apart from the background. There are many researchers making a deep investigation on it. In 2017, Ali et al. applied the Delta E color difference algorithm to separate the disease-infected area [13]. In general, four major methods are used to perform the image segmentation which are discussed the detail in the following paragraph [14].

Some researchers integrate the region of interest (ROI) and other methods to segment images. For example, Kao et al. claimed that the convolutional autoencoder served as the background filter to determine the ROI in an image [15]. The second method only concerns region segmentation. In 2013, Pujari et al. claimed that images were divided into various regions which had a special meaning and extracted the images' feature [16]. Akram and other colleagues provided an image processing model with real-time synchronous processing. By dividing the image into three color spaces, it can carry out contrast stretching, feature vector, and salient region recognition [17]. In addition, other researchers chose deep learning techniques to segment and detect images. Marko et al. recommended a depth-based target detection algorithm and used the two-stage algorithm to optimize plant disease images detection [18]. At last, the thresholding is common in segmentation. In 2018, Li et al. applied multilevel thresholding techniques based on gray histogram for image segmentation [19]. Mohamed and Diego presented a new multiobjective metaheuristic on the basis of a multiverse optimization algorithm to segment grayscale images via multilevel thresholding [20].

However, there is a fact that cannot be ignored. Because of the complexity of color information in the complicated environment, the machine vision algorithm based on color, ROI, and threshold performs poorly in practice.

2.2.2 Feature Extraction

The feature extraction of plant disease faces many problems in identifying plant disease. The distinct image features include textures, shape, colour, and motion-related attributes, which are the essential conditions for disease feature extraction [21, 22]. Raza and his colleagues described a method that uses colour and texture features to extract disease spots [1]. Hu et al. proposed the Dempster–Shafer (D-S) evidence theory and multi feature fusion for extracting features as well as the results were processed by introducing variance to improve decision rules of D-S evidence theory [2]. In addition, Turkoglu depicted improved versions of the Local Binary Patterns (LBP) methodology, which uses the original LBP local quadratic value to transform the image into grayscale and processes the R and G channels of the image by considering overall and region [3]. Li et al. researched an IoT feature extraction for the intelligent city based on the deep migration learning model [4]. There was an application in music, which can extract meaningful audio features in order to enable the visualisations to be responsive to the music [5]. And in recent studies, lots of novel approaches have been put forward to implementing feature extraction. For example, concerning the challenging task that the extraction of relevant and distinct features from electroencephalogram (EEG), Meziani et al. proposed two new spectral estimators that were robust against non-Gaussian, nonlinear, and nonstationary signals [2]. What is more, as Liu et al. reported, the high-dimensional time–frequency spectrum features were extracted by using the residual neural network and the improved signal-to-clutter ratio (SCR) [1]. Xu et al. introduced a feature extraction method based on the Hilbert marginal spectrum to perform the wear of milling tools [3].

2.2.3 Disease Identification

As for the precise identification, so many techniques are developed and researched to get accurate results. The identification model focused on using class labels for training images and built a fine-grained image classification system [5]. Zhang et al. reported a recognition method for plant disease leaf images based on a hybrid clustering [6]. In 2017, Patil et al. described a content-based image retrieval (CBIR) system to extract texture features and means value to compute color features, and support vector machine classifier was used for classification [13]. Through above researches, the major goal was to design the classification schemes and image analysis for feature extraction and identification. Recently, other new approaches have been introduced to identify the disease more accurately and precisely.

A novel system based on the selection of pictures and short text descriptions helped nonexperts in identifying plant diseases that can be used remotely from a desktop as well as in a smart phone or personal digital assistants [3]. Pertot et al. presented a scheme that used mobile phones for real-time on-field imaging of diseased plants and used mobile devices for leaf image segmentation and spotting of disease patch with improved k-means clustering [35]. Yang et al. presented a microscopy image detection methodology based on the synergistic judgement of texture and shape features and the decision tree-confusion matrix [36].

Additionally, the convolutional neural network is numerously utilised in identifying diseases. Chad et al. established a system capable of automatically identifying plant disease in field-acquired images of maize plants [12]. Ni et al. used the deep convolution neural network to train 1632 images of corn kernels and designed an automatic corn detector [38]. Lu et al. proposed a rice diseases identification method based on deep convolutional neural networks (CNNs) techniques [13]. Zhang et al. designed an agricultural machinery image recognition network using the deep learning algorithm [14].

Zhang et al. improved deep convolution neural network to improve the accuracy of maize leaf disease identification [11]. Images were input into two deep learning-based architectures, namely, AlexNet and VGG-16 net, to perform detection [12].

Coulibaly et al. suggested a method using transfer learning for feature extraction to build an identification system [13]. However, due to the requirement for high hardware resources and traditional neural network models of high quality and quantity of data sets in the training process, the training wastes much time that is not conducive to the promotion and use of the model. In this paper, we recommend a transfer learning model for identification combined with the pretrained model, using the dataset of disease leaves to train the model.

From the above research findings, some achievements have been achieved in three aspects: leaf image segmentation, leaf lesion feature extraction, and leaf disease recognition. However, there are still many problems to be solved to realise plant disease identification in the complex environment.

CHAPTER -3
METHODOLOGY AND DATASET

3.1 Proposed Methodology

A. Dataset collection

Dataset used here is the New Plant Diseases Dataset Dataset [16]. It is an open-source dataset which consists of about 20,600 images of plant leaf. Three different species of plant leaves are used for this research process such as Tomato, Potato and Pepper. This disease detection is done by using both healthy and infected leaves. Nearly 12 diseases for 3 species of plant leaves are included in this dataset along with their respective healthy leaves. The original size of the leaf images in the dataset are 256x256. It will be segmented and resized further for processing of data. The sample dataset is shown in Fig 1

B. Image pre-processing

Pre-processing of images is important because we need a uniformity in images and to have a common background. It will help in increasing the accuracy of the neural network model. But here we are not using any image pre-processing technique before training the model. The dataset collected is directly trained by the CNN model which is built already. To reduce overfitting of the model we are rotating all the leaf images in the collected dataset from 0-360º i.e., in all different angles. It helps in enlarging the dataset. The architecture diagram of this model is shown in Fig 2. C. Deep learning model Convolution Neural Network (CNN) is the deep learning model used here for identifying the diseases in the plant leaves. Since there is no image pre-processing done, the images in the dataset will be converted into an array and be trained. The CNN model will run on Windows (any version above 7) or Ubuntu (64-bit) Operating Systems. On top of the TensorFlow GPU backend, a deep learning package called Keras is used to build the CNN model. TensorFlow is the framework used as a backend which provides both high and low-level APIs to support the deep learning model. Keras is an open-source neural network library which provides only high-level APIs (Application Programming Interface). Kerasis used to build and train deep learning models. So, here Keras is built for ease in training the model and making it little user-friendly.

The CNN model is built by using a sequential model. The table obtained as an output for the sequential model is illustrated in Fig 3. This model sequential stacks up sequential layers of Convolution Neural Network in the order of input to output. A function called ReLu activation function is used in all the layers. MaxPooling2D is used along with this Keras to reduce the computational cost. It reduces the number of parameters in the CNN layers. Dropout is used here to avoid overfitting in CNN layers. Flatten is used here to flatten the data into one dimensional. Since the data has to be converted before passing it into the next level, flatten helps to make the data one dimensional. Dense is used to collect the output from all the neurons of the previous layer and give them to the next layer.

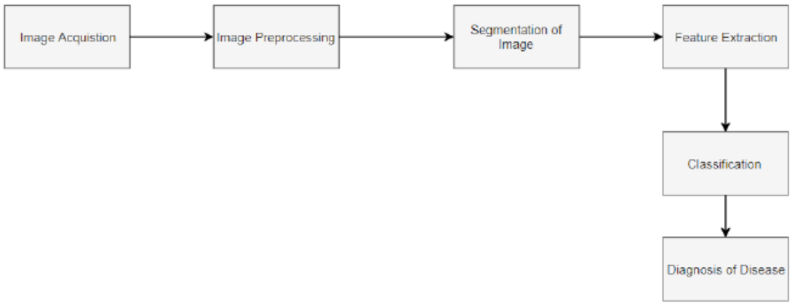


Figura 3: General workflow of plant disease prediction model

A generalised overview of the identification of crop diseases using VGG is presented in Figure 1. To implement our proposed work, firstly, the images of diseased leaves are collected from the New Plant Diseases Dataset [7] as well as from the field. Then the images are labelled according to the disease classes based on expert knowledge in case of field images. After that, pre-processing of images is performed which includes resizing of images, filtering of images, and different data augmentation techniques such as rotation, flipping, zca whitening, and shifting to increase the dataset size. The training and the testing images are fed into the VGGmodel and the features are extracted. At last, we have used two different machine learningclassifiers to classify the diseased leaf images. A detailed description of the implemented models is discussed in subsequent sub-sections.

3.1.1 Convolutional Neural Network

Convolutional Neural Network (CNN) is a deep learning technique used mainlyin recognition and classification purposes. In comparison with hand-crafted traditional approaches, CNN has the potential to perform better. It has the ability to learn the robust features directly from the input images. Whereas, to classify in the traditional hand-crafted feature-based approach, we need to extract the features separately. Different populardeep CNN models are used in identification of crop diseases such as AlexNet [40], VGGnet [20,41], InceptionV3 [14,42], ResNet50 [35], DenseNet [35], etc. From the research,it is shown that deep CNN-based models can achieve better performance accuracy in identification of crop diseases. Typically a CNN consist of convolutional layer, pooling layer, and fully connected layer [13].The convolutional layer is the main component in CNN, which extracts the features of the input images using different convolutional kernel. The output of the convolution operation is computed as:

$$x_j = I * W_j + b_j \quad where\ j = 1, 2, \ldots\ldots F \tag{1}$$

Here, xj is the output feature corresponds to j-th convolution filter,Wjis the corresponding weight, bjis the bias and Fis the number of filter . Poling layer down-samples the input vector and avoids over-fitting in the output.The pooling layer reduces the computational complexity of the model. The output of the pooling layer is evaluated as :

$$x_i^k = down(x_i^{k-1}, s) \tag{2}$$

Where down() represents the down sample, (xk−1i) represents the feature vector of previous layer and "s" represents the pool size. Max pooling and average pooling are two commonly used pooling operation.After the convolutional and pooling layer, several fully connected layers are there which transform the output of the previous layer to a single column vector. Usually,softmax function is used for multi-class prediction. Dropout regularisation is used to decrease the neuron size and to avoid over-fitting. The softmax function is written as:

$$softmax(z_j) = \frac{e^{z_j}}{\sum_k^K e^{z_k}} \quad for\ (j = 1, 2, \ldots\ldots K) \tag{3}$$

3.1.2 Visual Geometry Group (VGG19)

In recent times, several popular deep learning models such as AlexNet, VGGNet,ResNet, Inception, DenseNet, Xception, etc., have been used in the identification of plant diseases. Among this model, VGG is a relatively simple network developed by Simonyanand Zisserman [20]. VGG network consists of several convolution layer and pooling layer with different numbers of filter. VGGNet has two models VGG16 and VGG19. VGG16 consists of 16 convolutions and pooling layer with fully connected layer. VGG19 consist of19 convolutions and pooling layer with fully connected layer. The number of parameters generated in the VGG network is 140 million.

The VGG network is pre-trained on a large dataset (ImageNet) with 1000 categories.In our work, we have used VGG model which takes nine layers of the VGG19 model that includes 7 convolution layer and 2 max-pooling layer. The input size used in our implemented model is 256×256×3 and after performing the convolution and pooling operation the output size is 64×64×256. Instead of a fully connected layer, we have used global average pooling layer which reduces the number of parameters and dropout layer. The dropout layer plays an important role in reducing the overfittingproblem of the network. After extracting the features using VGG.

```
Model: "vgg19"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 256, 256, 3)]     0

 block1_conv1 (Conv2D)       (None, 256, 256, 64)      1792

 block1_conv2 (Conv2D)       (None, 256, 256, 64)      36928

 block1_pool (MaxPooling2D)  (None, 128, 128, 64)      0

 block2_conv1 (Conv2D)       (None, 128, 128, 128)     73856

 block2_conv2 (Conv2D)       (None, 128, 128, 128)     147584

 block2_pool (MaxPooling2D)  (None, 64, 64, 128)       0

 block3_conv1 (Conv2D)       (None, 64, 64, 256)       295168

 block3_conv2 (Conv2D)       (None, 64, 64, 256)       590080

 block3_conv3 (Conv2D)       (None, 64, 64, 256)       590080

 block3_conv4 (Conv2D)       (None, 64, 64, 256)       590080

 block3_pool (MaxPooling2D)  (None, 32, 32, 256)       0

 block4_conv1 (Conv2D)       (None, 32, 32, 512)       1180160

 block4_conv2 (Conv2D)       (None, 32, 32, 512)       2359808

 block4_conv3 (Conv2D)       (None, 32, 32, 512)       2359808

 block4_conv4 (Conv2D)       (None, 32, 32, 512)       2359808

 block4_pool (MaxPooling2D)  (None, 16, 16, 512)       0

 block5_conv1 (Conv2D)       (None, 16, 16, 512)       2359808

 block5_conv2 (Conv2D)       (None, 16, 16, 512)       2359808

 block5_conv3 (Conv2D)       (None, 16, 16, 512)       2359808

 block5_conv4 (Conv2D)       (None, 16, 16, 512)       2359808

 block5_pool (MaxPooling2D)  (None, 8, 8, 512)         0

=================================================================
Total params: 20,024,384
Trainable params: 0
Non-trainable params: 20,024,384
_________________________________________________________________
```

3.2 Dataset

This dataset is recreated using offline augmentation from the original dataset. The original dataset can be found on this github repo. This dataset consists of about 2.7GB images of crop leaves which is categorized into 38 different classes. The total dataset is divided into 80/20 ratio of training and validation set preserving the directory structure. A new directory containing 33 test images is created later for detection purpose.We analyze 54,306 images of crop leaves, which have a spread of 38 class labelled directories for training and 38 class labelled directories for validating are assigned to them. Inside each class label is a crop-disease pair, and we make an attempt to predict the crop-disease pair given just the image of the crop leaf. In all the approaches described in this paper, we resize the images to 256 x 256 pixels, and we perform both the model optimization and predictions on these downscaled images.

Dataset Name	**URL**
New Plant Diseases Dataset	https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset
Plant Village	https://www.plantvillage.org/
UCI Machine Learning Repository	https://www.archies.ics.uci.edu/ml/datasets/plants
Swedish Leaf Dataset	https://www.cvl.isy.liu.se
Plant Phenotyping dataset	https://plant-phenotyping.org/dataset

List of publicly available plant leaf dataset

To implement our proposed work, firstly, the images of diseased leaves are collected from **New Plant Diseases Dataset [7]** dataset as well as from the field. Then the images are labelled according to the disease classes based on expert knowledge in case of field images. After that, pre-processing of images is performed which includes resizing of images, filtering of images, and different data augmentation techniques such as rotation, and shifting to increase the dataset size. The training and the testing images are fed into the VGG model and the features are extracted. At last, we have used two different machine learning classifiers to classify the diseased leaf images. A detailed description of the implemented models is discussed in subsequent sub-sections.

Serial No.	**Crop**	**Crop Disease**
1.	Apple	1) Apple Scab 2) Balck Rot 3) Cedar Apple Rust 4) Healthy
2.	Blueberry	1) Healthy
3.	Cherry	1) Healthy

Serial No.	Crop	Crop Disease
		2) Powdery Mildew
4.	Corn	1) Northern Leaf Blight 2) Common Rust 3) Healthy 4) Gray Leaf Spot
5.	Grape	1) Black Rot 2) Esca 3) Healthy 4) Leaf Blight
6.	Orange	1) Huanglongbing
7.	Peach	1) Healthy 2) Bacterial Spot
8.	Pepper	1) Healthy 2) Bacterial Spot
9.	Potato	1) Healthy 2) Early Blight 3) Late Blight
10.	Raspberry	1) Healthy
11.	Soyabean	1) Healthy
12.	Squash	1) Powdery Mildew 2) Healthy
13.	Strawberry	1) Leaf Scroch 2) Healthy
14.	Tomato	1) Bacterial Spot 2) Early Blight 3) Healthy 4) Late Blight 5) Leaf Mold 6) Mosaic Virus 7) Yellow Leaf Curl Virus 8) Target Spot 9) Septoria Leaf Spot 10) Spyder Mites

Table : Crops and their respective diseases taken from the Dataset .

Reference link of dataset -
https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset

3.3 Image Processing

The various steps involved in Crop disease detection using Image Processing are image acquisition, segmentation, image analysis and classification.

3.3.1 Image Analysis for Disease Detection in Crop

The image acquisition process is the first stage in the process of image analysis. It is also known as digital image acquisition. It can be defined as the visual character of an object can be represented as a digitally encoded one. In simple terms, it can be defined as an image captured using a camera. Nowadays Digital image is extended to a mobile phone it makes the process of image acquisition a user-friendly. Photographs, printed paper, and photographic film are the media used for it. It mainly captures visualmoments.In the pre-processing of the image, there are two types in it namely digital image processing and analog image processing. Removal of unwanted features in the image is the main process involved in it.More algorithms are used for the removal of unwanted features in the image. The major steps involved in the image pre-processing are 1. Image Acquisition 2. Image Normalization 3. Image Enhancement 4.Segmentation 5. Morphology. Separating image into pixel and their similar attributes is image segmentation.

It mainly helps in the image interpretation process Oliver et al. [2018]. It transfers the low-level image into a high-level image. While analysing an image its success mainly depends on reliability in the segmentation process.Segmentation process involves both contextual and non-contextual. Several algorithms are used for the segmentation process.A copy of the original feature is kept during the feature selection process. In the feature extraction process, new sets of features are created and these two processes mainly deal with removing unwanted noise in the image and choosing the needed features only for the image analysis process.

During the classification process, the data are categorized into a number of classes. If new observation come into the process then identify whether the new observation belongs to which class Ferentinos[2018a]. There are several classification algorithms available for the process of classification which gives accurate classification result as well.

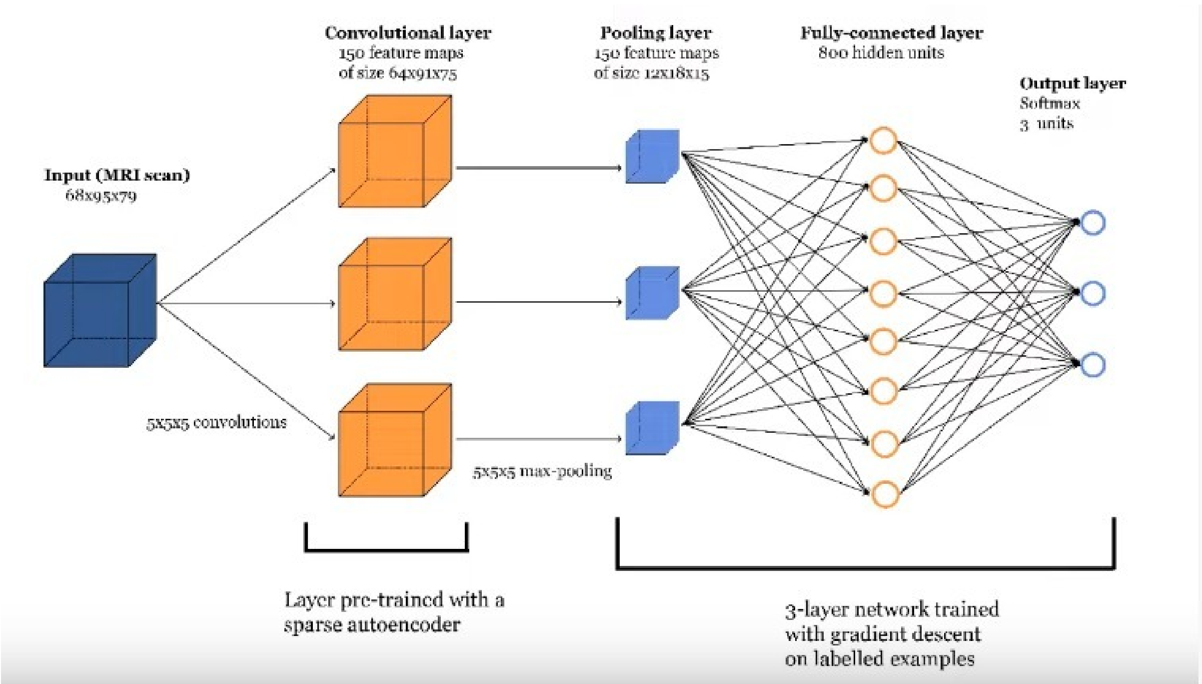
CHAPTER 4

MODELLING DETAILS

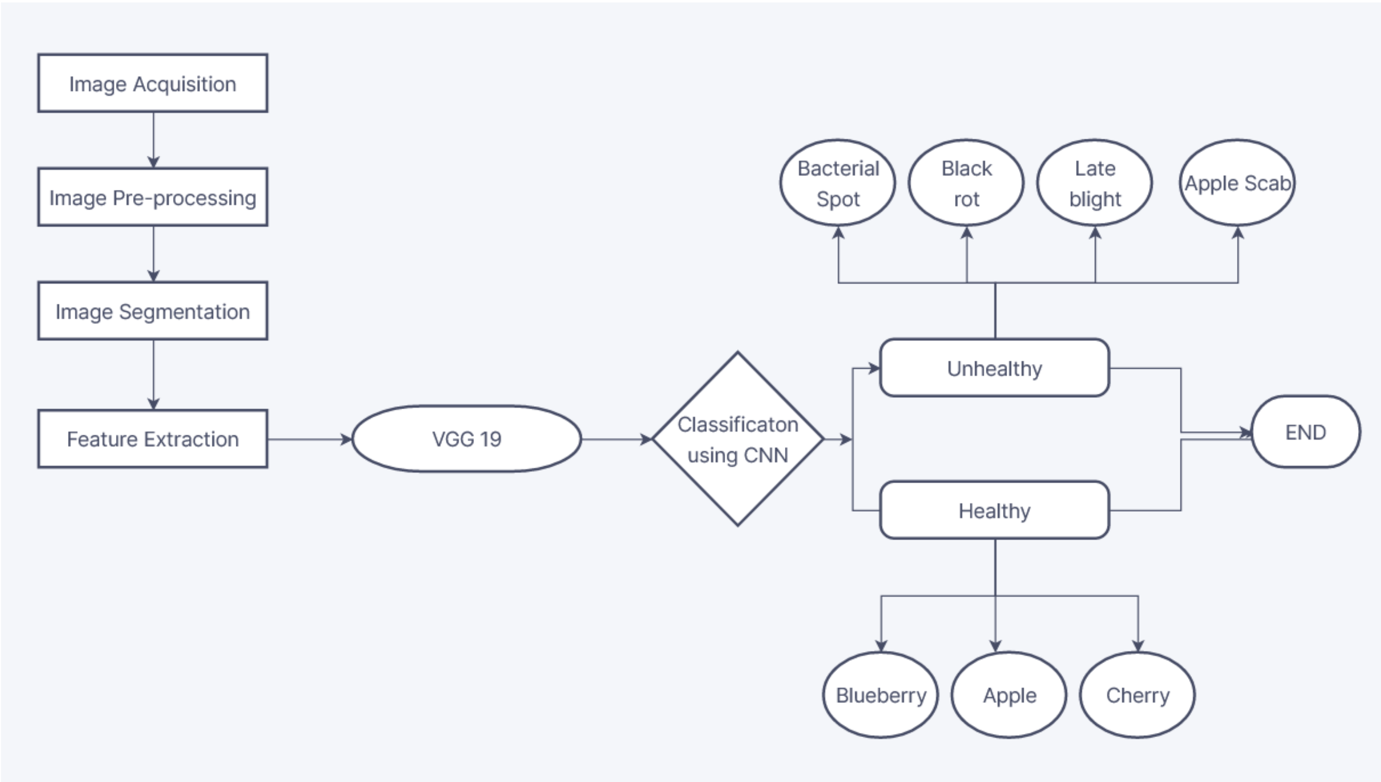
4.1 Experimental Setup

In our experiment, all the pre-processing, CNN based feature extraction and classification are performed in Google collab along with jupyter notebook environment. The model is implemented using keras, numpy, pandas, matplotlib, tensor flow, library, etc.
The experiment is conducted on Intel(R) Core (TM) i7-6700 CPU at 3.40 GHz with 8 GB memory.

4.2 Design Diagrams



4.3 Control Flow Diagram



CHAPTER 5

TESTING THE MODEL

5.1 Testing Plan

Here, we have almost 2.7GB of images and we have multiple sub folders. Each of the folders tells us the disease or the name of a disease that occurs in plants on their leaves. The agenda over here is to train all our images with respect to the classes, and then we are going to provide a new image from the internet and check whether our model is capable of detecting a particular image that belongs to that particular class of disease.

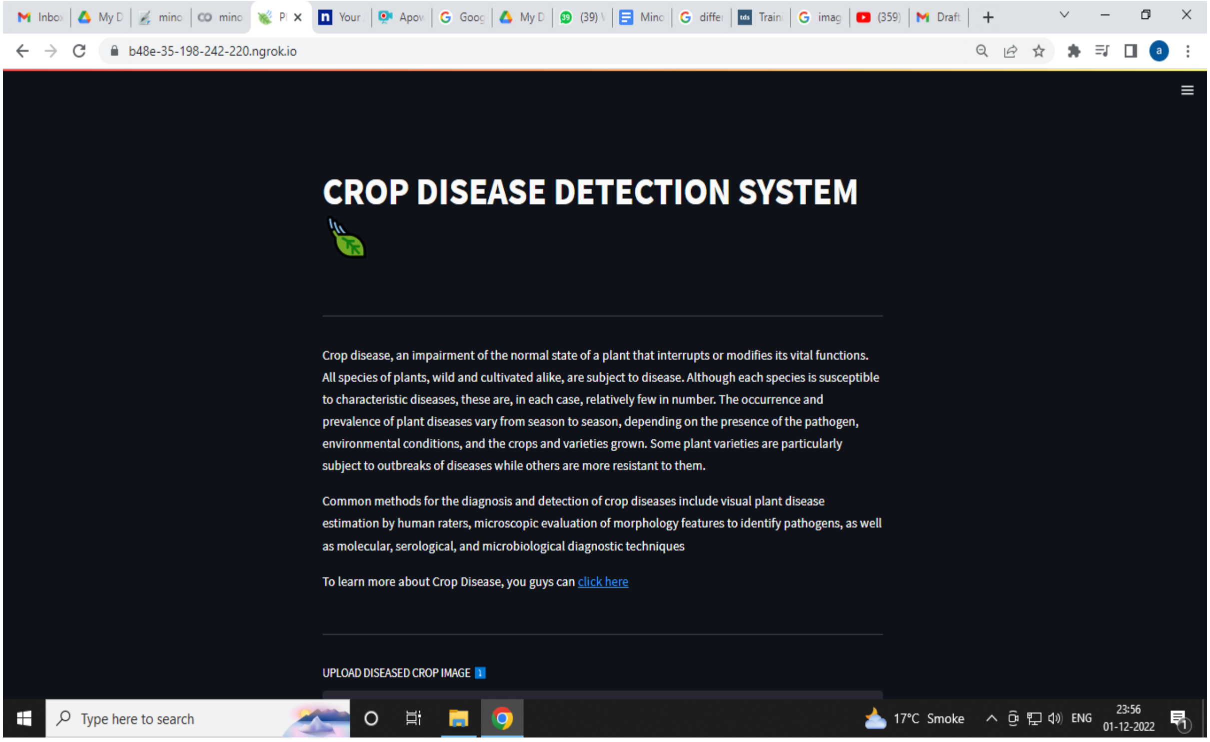
After importing the data from kaggle and after importing some main libraries, we have built our image data generator. For some exploratory data analysis, we've also calculated the number of classes present in our dataset. After loading our images to the data generator, a pre-processing function is added which has all the information. In order to pass the preprocessing function, we are importing vgg19. In order to create our own CNN model, libraries like numpy, matplotlib and keras are used. We've also imported functional API (usage of model).

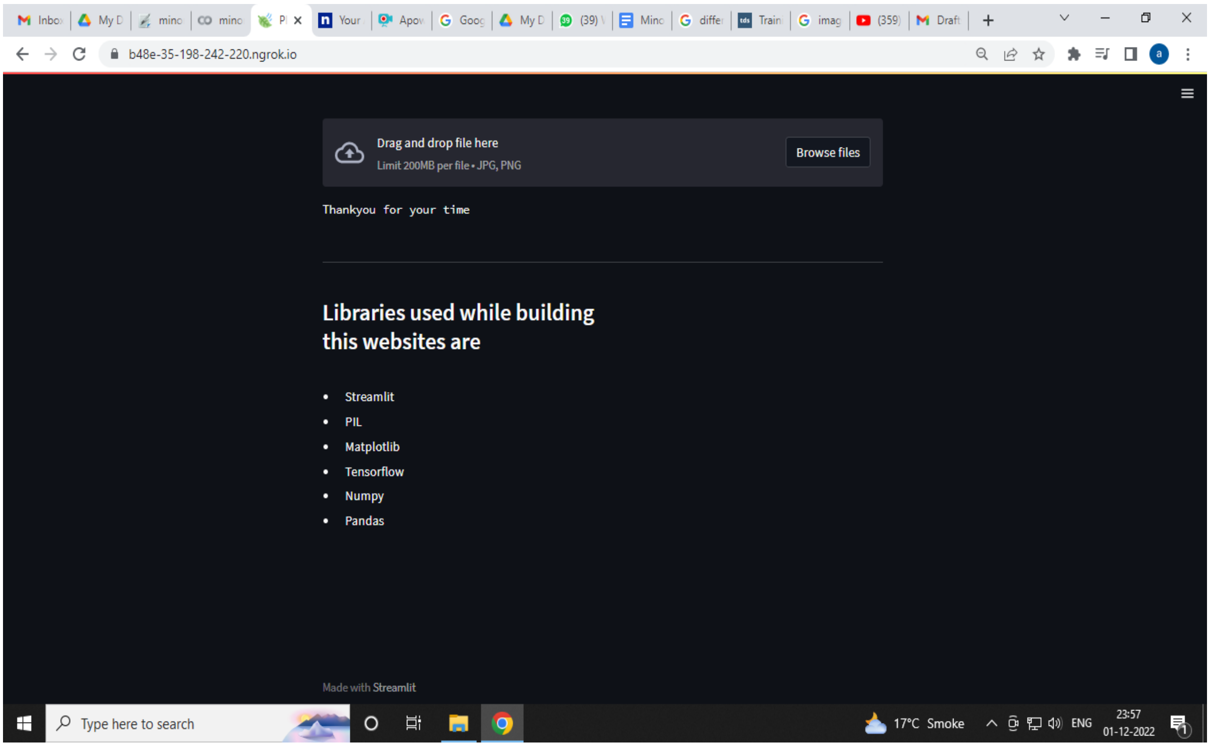
To train our model, we are doing early stopping and checkpoint using callbacks from keras. Early stopping basically monitors our validation. Due to early stopping, if there is no improvement in the val_accuracy, the model will stop training itself. Each step per epoch defines whether or not there is an increase in accuracy. Inbuilt detection of image is also happening in our code by copying the path.

5.2 Deployment

Further our model is deployed using Streamlit and Ngrok on web.
Ngrok is a globally distributed reverse proxy fronting your web services running in any cloud or private network, or your machine.
Streamlit is an open source app framework in Python language. It helps us create web apps for data science and machine learning in a short time. It is compatible with major Python libraries such as scikit-learn, Keras, PyTorch, SymPy(latex), NumPy, pandas, Matplotlib etc.
For loading our model, we have provided the directory where our model is stored (.h5 file).
When a user interacts with our website, he/she can upload the image of the crop he wants to be tested for disease. Then the website accepts ".jpg" and ".png" images as an input which further is sent to the trained model which applies techniques such as CNN and VGG19 on the test image and then extracts the required features. After extracting the required features , it forms a matrix of the features that are extracted, which is then compared with the matrix of

trained images and thus the model gives the instruction (which is based on the accuracy) to website. Website then prints the detected disease based on the accuracy of our model.





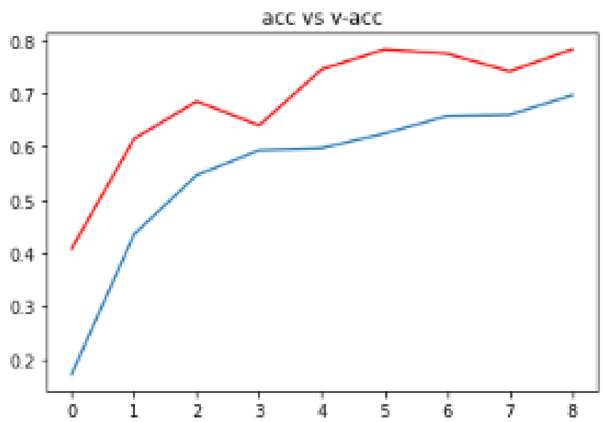
5.3 Performance Validation

The validation set is used to evaluate the models in order to perform model selection. On the other hand, the test set is used to evaluate whether the final model (that was selected in the previous step) can generalise well to new, unseen data.

For the same, graphs of ValAccuracy vs Accuracy and ValLoss vs Loss are attached:

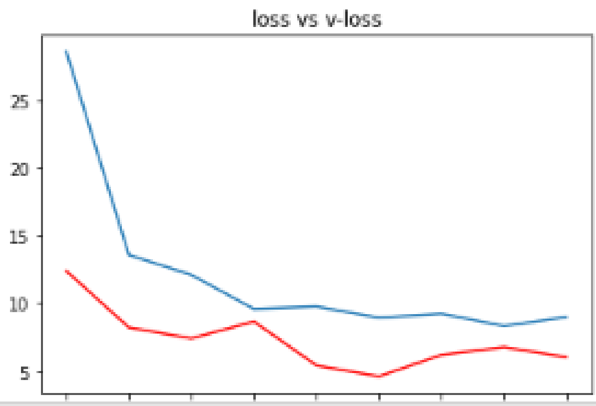
[21]

```
plt.plot(h['accuracy'])
plt.plot(h['val_accuracy'], c="red")
plt.title("acc vs v-acc")
plt.show()
```



[22]

```
plt.plot(h['loss'])
plt.plot(h['val_loss'], c="red")
plt.title("loss vs v-loss")
plt.show()
```



In order to evaluate our model and check the accuracy for the same the following code was run and an accuracy of 76.6% was generated.

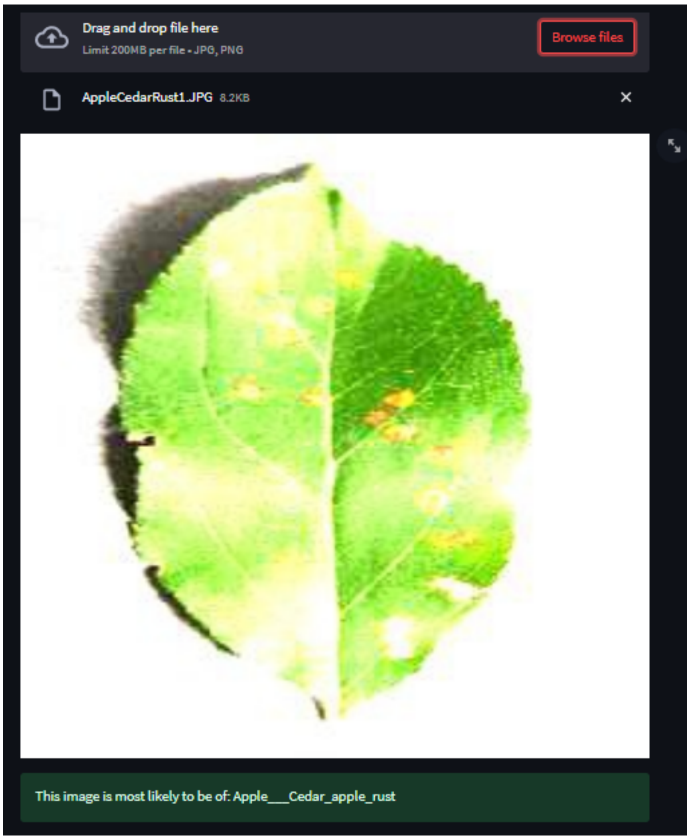
```
[24] #to evaluate our model

     acc= model.evaluate_generator(val)[1]

     print(f"The accuracy of your model is= {acc*100}%")

<ipython-input-24-9878eaec77c9>:3: UserWarning: `Model.evaluate_generator`
  acc= model.evaluate_generator(val)[1]
The accuracy of your model is= 76.68449878692627%
```

5.3 List Of All Test Cases

Here we have attached the test cases used on our deployed website along with the result that is the detected as the disease of the test image.

Here we have uploaded an image of corn -Common rust disease and thus the detected result was shown as output .

Here we have uploaded an image of Cedar-Apple-Rust disease and thus the detected result was shown as output .



Here we have uploaded an image of Cedar-Apple-Rust disease and thus the detected result was shown as output .

CHAPTER 6
CONCLUSION AND FUTURE WORK

6.1 Conclusion

Diseases in crops are one of the main reasons that directly affect in quality of agricultural crops. Therefore timely prediction of crop diseases is highly desired to protect the quality of the crops. Deep learning techniques, particularly CNN architectures show promising results in comparison with hand-crafted based approaches. In this paper, we proposed two models namely VGG to identify the diseases in plants. We have modified the pre-trained VGG19 model and used first nine layers and replaced the rest layers with a global average pooling layer. The average performance accuracy obtained using VGG From the result, we can conclude that VGG performs better in comparison with other deep learning models and hand-crafted based approaches. The computation time of our proposed approach is much lesser in compared to the other deep learning models. In future work, we will try to implement the proposed model on lightweight smart devices for automated detection of crop diseases. Similarly efforts will be made to identify the crops diseases, using other different parts of the crops such as flowers, stem etc.

6.2 Future Work

In today's scenario the government/ agricultural enthusiasts belonging to popular agricultural hubs of india can benefit from our project , but the concern arises when it comes to Isloated areas of our country .Thus to encounter this concern we can connect our project with Unmanned Aerial Vehicle(UAV) or better known as Drone . Drone technology is an emerging technology that is being used for many commercial and survey purposes. Drones can also be used in various agricultural applications to reduce human efforts and work time. Agricultural drones are highly efficient and their usage has been expanded to many areas in agriculture including fertiliser spraying, sowing seeds, growth monitoring, and mapping. This system includes the process of capturing images of the leaves automatically for image processing and analysis of leaf properties which would help to detect disease. Through which we would be able to capture images and obtain data from isolated areas of our country, which will then be available to the government/farmer which he could upload on our website and check for diseases in the yield. With the help of information regarding leaf disease can also be given to the farmer at an early stage and avoid loss in crop production. By this we would be able to increase the yield of the crops. Thus benefiting our country's growth in the agricultural sector.

REFERENCES

[1] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7(September), [1419].https://doi.org/10.3389/fpls.2016.01419

[2] Ser Serawork Wallelign, Mihai Polceanu, Cédric Buche. Soybean Plant Disease detection Using Convolutional Neural Network. FLAIRS-31, May 2018, Melbourne, United States. https://hal.archives-ouvertes.fr/hal-01807760

[3] S.Arivazhagan, R. Newlin Shebiah, S.Ananthi, S.Vishnu Varthini. 2013. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Agric Eng Int: CIGR Journal.

[4] Huu Quan Cap, Katsumasa Suwa, Erika Fujita, Satoshi Kagiwada, Hiroyuki Uga, and Hitoshi Iyatomi. A deep learning approach for on-site plant leaf detection. 2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA).

[5] Santhosh Kumar S, and B. K. Raghavendra. Diseases Detection of Various Plant Leaf Using Image Processing Techniques: A Review. 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)

[6] Jyotsna Bankar, and Nitin R Gavai. Convolutional Neural Network based Inception v3 Model for Animal Classification. International Journal of Advanced Research in Computer and communication Engineering, May 2018

[7] New Plant Diseases Dataset –"Image dataset containing different healthy and unhealthy crop leaves."- SAMIR BHATTARAI . https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset (as accessed on 22 November 2022)

[8] Abirami Devaraj, Karunya Rathan, Sarvepalli Jaahnavi and K Indira, " detection of Plant Disease using Image Processing Technique". International Conference on Communication and Signal Processing, April, pp (5386-7595), IEEE, April (2019)

[9] Shivani K. Tichkule, Prof. Dhanashri. H. Gawali, "Plant Diseases Detection Using Image Processing Techniques". 2016 Online International Conference on Green Engineering and Technologies (IC-GET). pp (5090-4556), IEEE (2016).

[10] Md. Selim Hossain, Rokeya Mumtahana Mou, Mohammed Mahedi Hasan, Sajib Chakraborty, M. Abdur Razzak, “Recognition and Detection of Tea Leaf’s Diseases Using Support Vector Machine ". IEEE 14th International Colloquium on Signal Processing & its Applications (CSPA 2018), pp (5386-0389), IEEE, March (2018).

[11] D. O. Shamkuwar, Gaurav Thakre, Amol R. More, Kishor S. Gajakosh, Muktanand O. Yewale, “An Expert System for Plant Disease Diagnosis by using Neural Network”. International Research Journal of Engineering and Technology (IRJET), pp (369-371), vol.5 issue.4, IRJET April (2018).

[12] Yong Zhong, Ming Zhao, “Research on deep learning in apple leaf disease recognition”. Journal in Computers and Electronics in Agriculture, pp (0168-1699), Elsevier, December (2019).

[13] Peng Zhang, Jun Meng, Yushi Luan, Chanjuan Liu, “Plant miRNA–lncRNA Interaction Prediction with the Ensemble of CNN and IndRNN”. Interdisciplinary Sciences: Computational Life Sciences (2020). vol.12 pp (82–89), Springer, December (2019).

[14] Qingxin Xiao, Weilu Li, Peng Chen, and Bing Wang, “Prediction of Crop Pests and Diseases in Cotton by Long Short-Term Memory Network”. ICIC 2018, LNCS 10955, pp (11–16), Springer (2018).

[15] Muammer Turkoglu1, Davut Hanbay, Abdulkadir Sengur, ‘Multi-model LSTM-based convolutional neural networks for detection of apple diseases and pests’ Journal of Ambient Intelligence and Humanized Computing, Springer (2019).