Plant Leaf Disease Classification and Prediction Using Customized Deep Transfer Learning Model

Table of Contents

List of Figures

Abstract

There is a significant productivity, financial damage due to plant diseases, and diminished overall quality of farm products. Detecting plant diseases has become more important in the surveillance of vast fields of crops in the modern-day. When it comes to disease management, farmers have difficulty transitioning from one strategy to another. The conventional method for detecting and identifying plant diseases is professional naked-eye inspection. This research examines the necessity for a simple technique for detecting plant leaf disease that would aid agricultural innovations. Early knowledge of crop health and disease detection may facilitate effective monitoring tactics. Crop yields will rise as a result of this method. In addition, the advantages and drawbacks of each of these prospective approaches are discussed in this study. Image capture, image analysis, extraction of features, and categorization based on neural networks are all part of the process. We get the best result to help the farmers through the processed methodology by implementing this model. The resulted accuracy of the implemented model is 81.09492659568787%. The proposed work enhances the farming culture to predict certain diseases and get a good yield of crops.

Keywords: Plant leaf Diseases, Deep Learning, VGG-19 Model, Crop Yield.

CHAPTER 1
INTRODUCTION

INTRODUCTION

Plants get affected by various diseases due to pathogens and environmental conditions. The disease occurrence varies from season to season. To get out of these crop diseases, farmers use different types of pesticides to control the disruption caused by the pathogen and provide security for the plant to get much yield as per their farming investment. If any plant of a particular crop is subjected to a disease outbreak, then the leaves of the plant change drastically with specific features. It is due to the infection caused by that disease, resulting in a decrease in crop yield and loss of crops. Incredibly less developed countries where admittance to disease control methods is minimum should face hunger and starvation. In this proposed algorithm, we use a transfer learning model to process the image input. Transfer learning is a machine learning technique where we use the knowledge obtained from the pre-trained model as the entry point for the model of a new task. It results in optimization that allows rapid progress when modelling the new task. It ensures higher performance even for a large amount of data. The target model obtained from the transfer learning model is highly efficient with reasonable accuracy. Out of various transfer learning models, we used VGG19 to recognize the disease by taking the plant leaves and processing those leaves images through the pre-trained model.

Fig. 1 Infected leaf of Apple plant

Background:

What is Deep Learning?

Deep learning can be considered as a subset of machine learning. It is a field that is based on learning and improving on its own by examining computer algorithms. While machine learning uses simpler concepts, deep learning works with artificial neural networks, which are designed to imitate how humans think and learn. Until recently, neural networks were limited by computing power and thus were limited in complexity. However, advancements in Big Data analytics have permitted larger, sophisticated neural networks, allowing computers to observe, learn, and react to complex situations faster than humans. Deep learning has aided image classification, language translation, speech recognition. It can be used to solve any pattern recognition problem and without human intervention.

Artificial neural networks, comprising many layers, drive deep learning. Deep Neural Networks (DNNs) are such types of networks where each layer can perform complex operations such as representation and abstraction that make sense of images, sound, and text. Considered the fastest-growing field in machine learning, deep learning represents a truly disruptive digital technology, and it is being used by increasingly more companies to create new business models.

How does Deep Learning work?

Neural networks are layers of nodes, much like the human brain is made up of neurons. Nodes within individual layers are connected to adjacent layers. The network is said to be deeper based on the number of layers it has. A single neuron in the human brain receives thousands of signals from other neurons. In an artificial neural network, signals travel between nodes and assign corresponding weights.

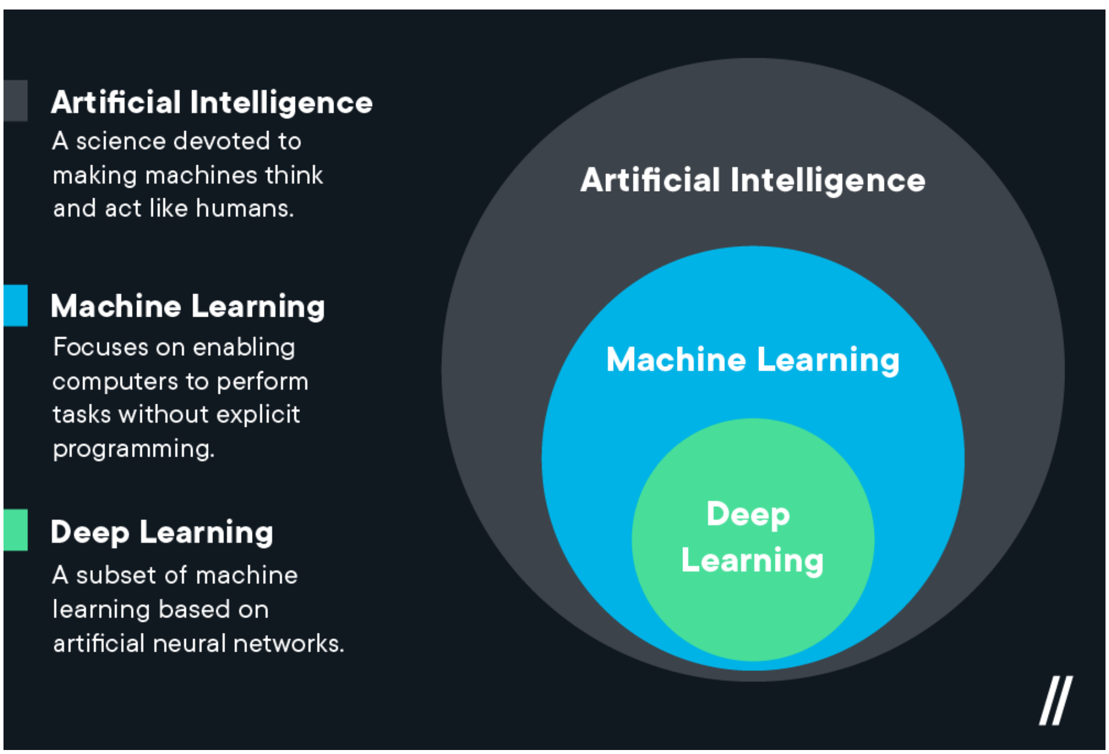

Fig. 2. Infographic:Get AI Development

A heavier weighted node will exert more effect on the next layer of nodes. The final layer compiles the weighted inputs to produce an output. Deep learning systems require powerful hardware because they have large amount of data being processed and involves several complex mathematical calculations. Even with such advanced hardware, however, training a neural network can take weeks.

Plant diseases and pests' detection is a very important research content in the field of machine vision. It is a technology that uses machine vision equipment to acquire images to judge whether there are diseases and pests in the collected plant images. At present, machine vision-based plant diseases and pests' detection equipment has been initially applied in agriculture and has replaced the traditional naked eye identification to some extent.

For traditional machine vision-based plant diseases and pests' detection method, conventional image processing algorithms or manual design of features plus classifiers are often used. This kind of method usually makes use of the different properties of plant diseases and pests to design the imaging scheme and chooses appropriate light source and shooting angle, which is helpful to obtain images with uniform illumination. Although carefully constructed imaging schemes can greatly reduce the difficulty of classical algorithm design, but also increase the application cost. At the same time, under natural environment, it is often unrealistic to expect the classical algorithms designed to completely eliminate the impact of scene changes on the recognition results [3]. In real complex natural environment, plant diseases and pests' detection is faced with many challenges, such as small difference between the lesion area and the background, low contrast, large variations in the scale of the lesion area and various types, and a lot of noise in the lesion image. Also, there are a lot of disturbances when collecting plant diseases and pests' images under natural light conditions. At this time, the traditional classical methods often appear helpless, and it is difficult to achieve better detection results.

In recent years, with the successful application of deep learning model represented by convolutional neural network (CNN) in many fields of computer vision (CV, computer-vision), for example, traffic detection, medical Image Recognition [5], Scenario text detection, expression recognition, face Recognition, etc. Several plant diseases and pests' detection methods based on deep learning are applied in real agricultural practice, and some domestic and foreign companies have developed a variety of deep learning-based plant diseases and pests detection Wechat applet and photo recognition APP software. Therefore, plant diseases and pests' detection method based on deep learning not only has important academic research value, but also has a very broad market application prospect.In view of the lack of comprehensive and detailed discussion on plant diseases and pests detection methods based on deep learning, this study summarizes and combs the relevant literatures from 2014 to 2020, aiming to help researchers quickly and systematically understand the relevant methods and technologies in this field. The content of this study is arranged as follows: “Definition of plant diseases and pests detection problem” section gives the definition of plant diseases and pests detection problem; “Image recognition technology based on deep learning” section focuses on the detailed introduction of image recognition technology based on deep learning; “Plant diseases and pests detection methods based on deep learning” section analyses the three kinds of plant diseases and pests detection methods based on deep learning according to network structure, including classification, detection and segmentation network; plant diseases and pests detection and compares the performance of the existing studies; “Challenges” section puts forward the challenges of plant diseases and pests detection based on deep learning; “Conclusions and future directions” section prospects the possible research focus and development direction in the future.

a) Definition of Plant Diseases and Pests:

Plant diseases and pests is one kind of natural disasters that affect the normal growth of plants and even cause plant death during the whole growth process of plants from seed development to seedling and to seedling growth. In machine vision tasks, plant diseases and pests tend to be the concepts of human experience rather than a purely mathematical definition.

b) Definition of Plant Diseases and Pests Detection Problem:

Compared with the definite classification, detection and segmentation tasks in computer vision, the requirements of plant diseases and pests’ detection is very general. In fact, its requirements can be divided into three different levels: what, where and how. In the first stage, “what” corresponds to the classification task in computer vision. As shown in Fig. 2, the label of the category to which it belongs is given. The task in this stage can be called classification and only gives the category information of the image. In the second stage, “where” corresponds to the location task in computer vision, and the positioning of this stage is the rigorous sense of detection. This stage not only acquires what types of diseases and pests exist in the image, but also gives their specific locations. As shown in Fig. 2, the plaque area of gray mold is marked with a rectangular box. In the third stage, “how” corresponds to the segmentation task in computer vision. As shown in Fig. 2, the lesions of gray mold are separated from the background pixel by pixel, and a series of information such as the length, area, location of the lesions of gray mold can be further obtained, which can assist the higher-level severity level evaluation of plant diseases and pests. Classification describes the image globally through feature expression, and then determines whether there is a certain kind of object in the image by means of classification operation; while object detection focuses on local description, that is, answering what object exists in what position in an image, so in addition to feature expression, object structure is the most obvious feature that object detection differs from object classification. That is, feature expression is the main research line of object classification, while structure learning is the research focus of object detection. Although the function requirements and objectives of the three stages of plant diseases and pests detection are different, yet in fact, the three stages are mutually inclusive and can be converted. For example, the “where” in the second stage contains the process of “what” in the first stage, and the “how” in the third stage can finish the task of “where” in the second stage. Also, the “what” in the first stage can achieve the goal of the second and the third stages through some methods. Therefore, the problem in this study is collectively referred to as plant diseases and pests detection as conventions in the following text, and the terminology differentiates only when different network structures and functions are adopted.

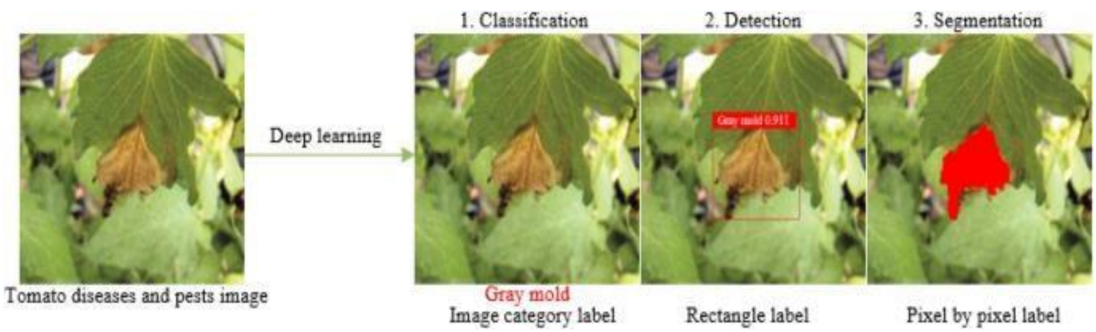

Fig. 3 Definition of Plant Diseases and Pests Detection Problem

Image Recognition technology based on Deep Learning:

Compared with other image recognition methods, the image recognition technology based on deep learning does not need to extract specific features, and only through iterative learning can find appropriate features, which can acquire global and contextual features of images, and has strong robustness and higher recognition accuracy.

Deep learning Theory:

The concept of Deep Learning (DL) originated from a paper published in Science by Hinton et al. in 2006. The basic idea of deep learning is: using neural network for data analysis and feature learning, data features are extracted by multiple hidden layers, each hidden layer can be regarded as a perceptron, the perceptron is used to extract low-level features, and then combine low-level features to obtain abstract high-level features, which can significantly alleviate the problem of local minimum. Deep learning overcomes the disadvantage that traditional algorithms rely on artificially designed features and has attracted more and more researchers' attention. It has now been successfully applied in computer vision, pattern recognition, speech recognition, natural language processing and recommendation systems.

Traditional image classification and recognition methods of manual design features can only extract the underlying features, and it is difficult to extract the deep and complex image feature information. And deep learning method can solve this bottleneck. It can directly conduct unsupervised learning from the original image to obtain multi-level image feature information such as low-level features, intermediate features and high-level semantic features. Traditional plant diseases and pests' detection algorithms mainly adopt the image recognition method of manual designed features, which is difficult and depends on experience and luck, and cannot automatically learn and extract features from the original image. On the contrary, deep learning can automatically learn features from large data without manual manipulation. The model is composed of multiple layers, which has good autonomous learning ability and feature expression ability, and can automatically extract image features for image classification and recognition. Therefore, deep learning can play a great role in the field of plant diseases and pests image recognition. At present, deep learning methods have developed many well-known deep neural network models, including deep belief network (DBN), deep Boltzmann machine (DBM), stack de-noising autoencoder (SDAE) and deep convolutional neural network (CNN). In the area of image recognition, the use of these deep neural network models to realize automate feature extraction from high-dimensional feature space offers significant advantages over traditional manual design feature extraction methods. In addition, as the number of training samples grows and the computational power increases, the characterization power of deep neural networks is being further improved. Nowadays, the boom of deep learning is sweeping both industry and academia, and the performance of deep neural network models are all significantly ahead of traditional models. In recent years, the most popular deep learning framework is deep convolutional neural network.

Convolutional neural network:

Convolutional Neural Networks, abbreviated as CNN, has a complex network structure and can perform convolution operations. As shown in Fig. 2, the convolutional neural network model is composed of input layer, convolution layer, pooling layer, full connection layer and output layer.

In one model, the convolution layer and the pooling layer alternate several times, and when the neurons of the convolution layer are connected to the neurons of the pooling layer, no full connection is required. CNN is a popular model in the field of deep learning. The reason lies in the huge model capacity and complex information brought about by the basic structural characteristics of CNN, which enables CNN to play an advantage in image recognition. At the same time, the successes of CNN in computer vision tasks have boosted the growing popularity of deep learning.

Plant diseases and pests' detection methods based on deep learning:

This section gives a summary overview of plant diseases and pests detection methods based on deep learning. Since the goal achieved is completely consistent with the computer vision task, plant diseases and pests' detection methods based on deep learning can be seen as an application of relevant classical networks in the field of agriculture. The network can be further subdivided into classification network, detection network and segmentation network according to the different network structures. This paper is subdivided into several different sub-methods according to the processing characteristics of each type of methods.

Detection Network:

Object positioning is one of the most basic tasks in the field of computer vision. It is also the closest task to plant diseases and pests' detections in the traditional sense. Its purpose is to obtain accurate location and category information of the object. At present, object detection methods based on deep learning emerge endlessly. Generally speaking, plant diseases and pests detection network based on deep learning can be divided into: two stage network represented by Faster R-CNN [54]; one stage network represented by SSD [55] and YOLO [56,57,58]. The main difference between the two networks is that the two-stage network needs to first generate a candidate box (proposal) that may contain the lesions, and then further execute the object detection process. In contrast, the one-stage network directly uses the features extracted in the network to predict the location and class of the lesions.

Classification Network:

In real natural environment, the great differences in shape, size, texture, color, background, layout and imaging illumination of plant diseases and pests make the recognition a difficult task. Due to the strong feature extraction capability of CNN, the adoption of CNN-based classification network has become the most commonly used pattern in plant diseases and pests' classification. Generally, the feature extraction part of CNN classification network consists of cascaded convolution layer + pooling layer, followed by full connection layer (or average pooling layer) + softmax structure for classification. Existing plant diseases and pests' classification network mostly use the muture network structures in computer vision, including AlexNet, GoogleLeNet, VGGNet, ResNet, Inception V4, DenseNets, MobileNet and SqueezeNet. There are also some studies which have designed network structures based on practical problems. By inputting a test image into the classification network, the network analyses the input image and returns a label that classifies the image. According to the difference of tasks achieved by the classification network method, it can be subdivided into three subcategories: using the network as a feature extractor, using the network for classification directly and using the network for lesions location.

Segmentation Network:

Segmentation network converts the plant diseases and pests' detection task to semantic and even instance segmentation of lesions and normal areas. It not only finely divides the lesion area, but also obtains the location, category and corresponding geometric properties (including length, width, area, outline, center, etc.). It can be roughly divided into: Fully Convolutional Networks (FCN) and Mask R-CNN.

CHAPTER 2
LITERATURE SURVEY

LITERATURE SURVEY

Several studies describe how to identify and treat the disorders. Techniques illustrating how to put them into practice, as seen and being the subject of this discussion, the possibility of finding diseased cotton leaves.

Earlier papers are describing to diagnosis the cotton leaves using various approaches suggesting the various implementation ways as illustrated and discussed below. [1] Diseases Control has been developed in a BP neural network as a decision-making system. [2] Cotton foliar diseases presented a method for automatic classification of cotton diseases used Wavelet transform energy has been used for feature extraction while Support Vector Machine has been used for classification. [3] Existing the research work described in the features could be extracted using a self-organizing feature map with a back-propagation neural network is used to recognize the colour of the image. [4] Earlier paper the fuzzy feature selection approach fuzzy curves (FC) and surfaces (FS) - is proposed to select features of cotton disease leaf the image. [5] Presented work carried out RPM and Dis Bin and compared with the classical PCA based technique. [6] The cotton leaf disease segmentation is performed using modified self-organizing feature map with genetic algorithms for optimization and support vector machines for classification [7] proposed use these techniques to extract Eigenfeature from cotton leaf.

Presently, in the recent agricultural system, advance computation techniques have been developed to help farmers (or) agricultures to monitor the proper development of their crops. In our early agricultural system, during the harvesting process of the crops, the exposed eye observation of farmers or experts is the main approach adopted in practice for the

detection and identification of crop diseases under microscopic conditions in the laboratory. However, this requires continuous monitoring of experts which might be prohibitively expensive in large farms. Further, in some developing countries, farmers may have to go long distances to contact experts, this makes consulting experts too expensive and time consuming. The basic problems regarding with crop is on the field, a fast and accurate recognition and classification of the diseases is required by inspecting the infected leaf spot images also identify the severity of the diseases. There are two main characteristics of plant-disease detection machine-learning methods that must be achieved, they are: performance and accuracy.

Proposed Research work will describe the process of Advance computing techniques for recognition of leaf spot diseases as this can give much benefit in monitoring large fields of crops and discover the symptoms of diseases. In this work we have to find out the computer systems which analyze the input images using the RGB pixel counting values feature used and identify (each and every disease) wise and next using homogenization techniques Sobel and canny using edge detection to identify the affected parts of the leaf spot to recognize the diseases boundary is white lightning and the result (recognition of the diseases and pest recommended) is given as output to the farmers.

(Anand. H. Kulkarni, 2012). A method in which the color and spot characteristics may be retrieved backward propagation of a self-organizing feature map a network of neurons detection of a diseased area on the leaves of a plant and texture traits are used to classify plant leaf diseases.

An algorithm was devised by (S. Arivazhagan, R. Newlin Shebiah, S. Ananthi & Varthi, 2013) to process the data, transforming the supplied color. The red, green, and blue components of an RGB picture are green pixels disguised after being formed. Shape and form texturing characteristics are considered from the data. In order to classify objects, the Minimum Distance is used. Vector machines for criterion evaluation and support (SVMs).

(Al Bashish et al., 2010), they have developed a framework for the Identification and Classification of Plant Diseases of the Leaf and Stem. The K-Means algorithm uses to segment and transforms RGB pictures into HIS Tonal system. Color and texture characteristics are then calculated. A statistical neural network classifier. Classification is used for categorization purposes. Image segmentation has been hypothesized by (Nandini Kakran & Pratik Singh, 2019) for crop disease detection.

The three-part technique is as follows: A device-independent color format, RGB, was used to store the leaf pictures at first. To separate the photos, they had to be resized and converted to the CIELAB color space independent of hardware. Region-based extraction of the contaminated area is the second method. The use of segmentation was made. An essential part of K-means clustering is A method used during the segmentation stage. Twelfth extract characteristics are based on color, shape, and texture. Typically used to describe a geographic area. As a result, The co-occurrence matrix of characteristics, wavelet, and grey levels. It was necessary to make use of certain strategies.

Classifying apples using a support vector regression (SVR) algorithm disease of the leaves. Agricultural plant disease has been hypothesized by (Sanjay B. Dhaygude, 2013); a prototype software approach for detecting rice disease based on diseased photos of diverse rice plants is described in Identification Using Pattern Recognition Techniques. Photos of diseased rice plants are taken using a digital camera, and sick sections of the plants are identified using image growth and image segmentation algorithms. After that, a neural network is used to classify the affected leaf tissue. Cotton Leaf Spot Diseases Detection Using Feature Selection with Skew Divergence Approach has been suggested by (Dheeb AI Bashish, Malik Braik, 2010). In this study, the upgraded PSO feature selection method uses the skew divergence method and extracts Edge, CYMK color, GA, Color, and Texture variations features. The collected features were analyzed using SVM, BPN, and Fuzzy with Edge Selection & Classification.

(Yan-Cheng Zhang, Han-Ping Mao, Bo hu, 2007), A panel of experts analyzed the images on the remote server. Computer vision methods are used to identify and classify damaged areas in a picture. The disease-affected lesions are segmented using a straightforward color difference technique. With the use of a mobile phone notification, the expert may assess the study's findings and offer comments to the farmers. An image recognition system capable of identifying crop diseases is the purpose of this study. The scanned color picture of a diseased leaf is the first step in image processing. These photos are segmented using a mathematical morphology technique. These attributes were then utilized in conjunction with a classification approach called the membership function to separate the three categories of illnesses.

(Haiguang Wang, Guanlin Li, Zhanhong Ma, 2012a) cotton leaf disease was identified and classified using image processing and machine learning approaches. The poll results on background removal and segmentation approaches were also brought up throughout the discussion. A translation from RGB to HSV color space for background removal was discovered during this research. In addition, we discovered that thresholding produces superior results compared to other methods of removing background noise. We used thresholding to the masked image after performing color segmentation by masking green pixels in the background removed image to produce a binary picture. This segmentation process is important in extracting disease-specific characteristics. For illness categorization, we observed that SVM performed well in accuracy. A total of three of our recommended processes have been implemented: Image Acquisition, Image Preprocessing, and Image Segmentation. Our proposed work has five key parts.

(H. AI-Hiary, S. Bani-Ahmad, M. Reyalat, 2011) After taking a picture, the collected pictures are initially enhanced using Image Edge detection and segmentation algorithms. Image segmentation is conducted out

using the R, G, and B Color Feature images (disease spots). An image feature extraction process is then used to identify disease areas and implement a pest management strategy based on these attributes. Cotton leaf spot, cotton leaf color segmentation, and image segmentation based on edge detection are included in this study's analysis and categorization of illness.

(M. Meunkaewjinda, P. Kumsawat, K. Attakitmongcol, 2008) We propose and test a new software method for detecting and classifying plant illnesses using Image Processing. Few agricultural professionals in rural India can examine crop photographs and provide advice. Farmers often get expert solutions to their inquiries too late because of the delay in expert responses. Image processing algorithms based on color, texture, and shape will be developed in this article to identify crop illnesses or other issues that may impact crops from photos and provide farmers with the most up-to-date information through SMS. As a result, more efficient use of chemicals will lead to increased production and a better-quality output due to these technologies.

(Lili li, Shujuan zhang, 2021), Using deep learning and advanced imaging techniques, the disadvantages in plant disease recognition get removed, and the efficiency of the proposed algorithm is also increased by using a particular methodology in processing the application. Certain visualization techniques are used to organize deep learning architectures. A roll of metrics is used for training and validation. Mean Average Precision (MAP) is an indicator for evaluating the framework.

(Haiguang Wang, Guanlin Li, Zhanhong Ma, 2012b) In this proposed technique, Principal component analysis (PCA) and Backpropagation networks (BP) are used to predict two kinds of wheat diseases. Generalized Regression Networks GRNNs and Probabilistic Neural Networks PNNs are used for classification to identify the grape diseases. The fitting and prediction accuracy of the class labels are identified via these processing algorithms.

(Song kai, Liu zhikun, Su hang, 2011) Based on image processing and analysis for maize crop disease prediction, image recognition of corn leaf. Texture features of corn disease and segmentation of disease spots are used to recognize the disease. The backpropagation neural network (BP) mechanism is also implemented to coordinate disease detection input/output relations. A spatial gray matrix is used for fast identification of the maize disease.

(Mokhled S, 2013) image analysis and classification techniques for spotted disease detection of Olive leaves are used for the research. Various images of leaves were collected from the field and taken as samples. C- means clustering is used to detect the defect. The severity of the disease is determined via the classification of leaf areas. As a result, 86% accuracy is obtained through the processed algorithm.

(Argenti et al., 1990) An algorithm for fast analysis is proposed for calculating the parameters of matrices. A supervised learning mechanism is used for reading the input data. Statistical features get extracted through classification and segmentation.

The paper by Amiya Halder and Nilavra Pathak titled, "An Evolutionary Dynamic Clustering Based Colour Image Segmentation" present that each pixel in color images have three components that are RED, GREEN, BLUE and these components are more complex than gray scale images which have single intensity value for pixels. RGB segmentation can be used in medical imaging, mining and mineral imaging, bioinformatics, and material sciences.

(Mehmet Sezgin et al) have done an extensive survey over "Image Thresholding Techniques". Besides image segmentation these techniques are also used in applications such as, "Document image analysis, Map processing, Scene processing, Inspection of materials for quality, Cell images, Knowledge representation, Non-destructive testing, Ultrasonic, eddy current and thermal images, X-ray computed tomography and Endoscopic images etc."

Lot of work has been devoted to the detection of leaf diseases using image processing in the history and it continues to attract research to carry out their research work in this field. Automatic crop disease detection using image processing and machine learning has been gaining prominence in recent years.

P. Krithika et al., pre-processed by image resizing, contrast enhancement and color-space conversion. The K-Means clustering for segmentation and feature extraction using GLCM is performed. Classification was made using multiclass SVM. R. Meena et al., performed color space conversion followed by enhancement process. The primary colors of leaves are converted into L*A*B*. The K-Mean clustering algorithm is used for segmentation. The GLCM and SVM are used for feature extraction and classification respectively. Bharat et al., acquired images using digital camera and median filter is used for image enhancement. K-Mean clustering is used for segmentation. SVM is used for classification. Pooja et al., segmentation is done to get the areas of interest that is the infected region. It is done using k-Mean clustering algorithm, Otsu's detection converting RGB to HSI later segmentation is done using boundary and spot detection algorithm. Rukaiyya et al., performed pre-processing by contrast adjustment and normalization. The conversion of color transform into YCBCR and Bi-level thresholding is performed. The GLCM, and HMM are used for features extraction and classification.Chaitali et al., segmentation of image is applied for background subtraction. The classification approach is carried out by KNN, ANN and SVM method. In KNN, it classifies samples using nearest distance between trained and testing subjects. Varun et al., has developed model for extraction thresholding technique and morphological operation. Then multiclass SVM is used as classifier. For segmentation, based on a set of marks generated by analysis of the color and luminosity components of different regions of image is L*A*B* color spaces. The GLCM is used for feature extraction. Vijai Singh et al., [19] considered samples of plant leaves like rose/beans (bacterial disorder), lemon (sun burn disorder), banana (early scorch) and beans (fungal) that are captured using a digital camera. The green regions as background using thresholding algorithm. Finally, the genetic algorithm is used to get the segmented image. The color co-occurrence is adapted for useful extraction of features from the segmented images. The Minimum Distance Criterion and then SVM classifier is used for classification purpose. The average accuracy of 97.6% has been recorded.

Sa'ed Abed et al., performed scaling and stretching (min-max linear) process for the input samples to improve the quality. The creation of HIS model is completed and the same is segmented later. The techniques of combined Euclidean distance and K-mean clustering is performed for segmentation of the samples. The GLCM and SVM are used for feature extraction and classification respectively. Arya et al., takes input RGB image and creates color transformation then conversion of the input samples to HIS format. Finally, segment the components using Otsu's method. Nema et al., images of 81 were included in the database and analysis was performed in L*a*b color space. Segmentation of the leaf disease was carried using k-means clustering and the classification of the disease was performed using SVM. Statistical information such as mean, median, mode, standard deviation was used by authors to record their findings. Vidyashree Kanbur et al., developed the model for leaf detection disease using multiple descriptors. The model was tested on local leaf database and the performance of the model was superior., but it can be tested on publicly available dataset.

Pushpa et al., Indices Based Histogram technique is used to segment unhealthy region of the leaf. The authors have surpassed other segmentation techniques such as slice segmentation, polygon approximation, and mean-shift segmentation. Kaleem et al., considered pre-processed to resize them into 300*300 sized images, remove background noise, enhance brightness, and adjust the contrast. The K-means clustering for segmentation and the useful features are extracted using Statistical GLCM and SVM classifier is used for classification of leaf disorders.

CHAPTER 3

PROBLEM STATEMENT & OBJECTIVES

PROBLEM STATEMENT

Crop diseases are a major threat to food security, but their rapid identification remains difficult in many parts of the world due to the lack of the necessary infrastructure. The combination of increasing global smartphone penetration and recent advances in computer vision made possible by deep learning has paved the way for smartphone-assisted disease diagnosis. Using a public dataset of 54,306 images of diseased and healthy plant leaves collected under controlled conditions, we train a deep convolutional neural network to identify 14 crop species and 26 diseases (or absence thereof). The trained model achieves an accuracy of 99.35% on a held-out test set, demonstrating the feasibility of this approach. Overall, the approach of training deep learning models on increasingly large and publicly available image datasets presents a clear path toward smartphone-assisted crop disease diagnosis on a massive global scale.

OBJECTIVES

The primary contributions of this research are:

1. We understand the features of infected and uninfected leaves of different fruits and vegetable crops.
2. We explore the existing machine learning and deep learning models used in plant leaf disease detection.
3. We are developing deep transfer learning models to detect plant leaf diseases and customizing the model to work with different class labels.
4. We are testing the model with randomly selected plant leaf images to predict whether the given leaf is infected or uninfected, finding the disease name.

CHAPTER 4
PROPOSED METHODOLOGY

METHODOLOGY

The four major diseases affected by the plant selected here are Alternaria Alternata, Anthracnose, Bacterial Blight and Cercospora Leaf Spot and also addition with the healthy leaves. Plant diseases control accomplished a desired aim only with welltimed detection of diseases and perfect identification of causative agents. In that majorly diseases seen on the leaves of a plant, then it will spread to different part of the plant. So here first need to identify the diseases and then need to feed the proper dosage of medicine treatment to cure the diseases. The image processing method helps to identify and classify the diseases automatically, efficiently, fast and accurately

The basic steps involved in plant leaf disease detection are given below.

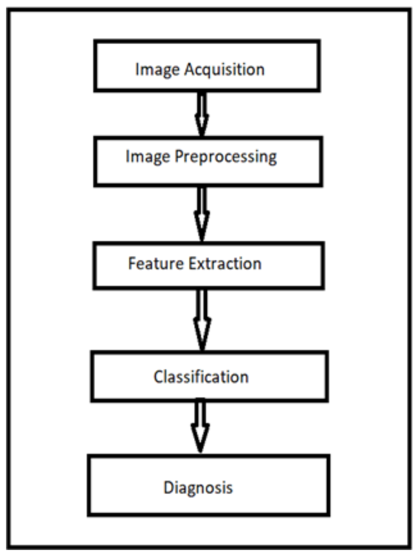

Fig. 4 Basic Steps for Disease Detection Algorithm

Acquisition of images:

High-quality RGB color photographs are taken using a digital camera with the necessary resolution. The project explicitly dictates how an image database is built. The classifier, which determines the algorithm's resilience, is made more efficient by the picture database.

Processing and Segmentation of images:

During the pre-processing stage, background and noise are removed from the picture data, and unwanted distortions are suppressed. It improves picture processing and analysis capabilities.

HIS and CIELAB color spaces are used to transform the RGB color picture. When HSI and CIELAB were developed as color-independent space models, they were also based on human perception.

Segmentation is the first step in locating the area that is contaminated. K-means clustering and edge detection algorithms are often used for segmentation.

A variety of characteristics are used to define the contaminated area after segmentation. For describing a location, color, texture, and form are often employed.

Color characteristics are essential for determining the visual surroundings, identifying objects, and conveying information.

An object's texture is one of its most critical identifying characteristics. Image retrieval benefits from its use as a robust regional descriptor.

The texture is described, by contrast, as homogeneity, dissimilarity, energy, and entropy. An image's shape is one of the most fundamental properties that may be used to describe its content.

To cut down the large monitoring work in farms of crops, symptoms at initial stage is taken care of. Naked eye observation by experts is the most prominent and existing method of plant disease identification and detection. But this method is convenient only when farms are smaller in size. As for the larger farms this method is quite cumbersome. A big team of experts and their continuous monitoring is needed for doing so which results in high cost in terms of both time and labour. In most of the developing nations the farmers lack the knowledge about the plant diseases and methods to prevent them or deal with them. They are not able to contact the experts in time for their advice and if they do so it is again time consuming and expensive process. For such situations, crop monitoring in large farms, propounded technique is profitable. On the basis of symptoms, automatic detection of the diseases is easier and cheaper as identification is done simply by checking the symptoms on the leaves of plant. This also supports machine vision to provide image based automatic process control, inspection, and robot guidance.

Image segmentation is a process wherein different parts of an image are either grouped together or separated based on certain attributes. There are many approaches prevailing now days for image segmentation. They include thresholding methods from the simple to advanced segmentation for the colour image. Those parts which generally humans can easily identify and can be viewed as individual objects, corresponding to these methods. As variety of methods have been developed for image segmentation, therefore for computers recognizing objects intelligently have no meaning.

Segmentation process is based upon on the various features available in the image. This information might be boundaries, colour or segment of an image.

Image Segmentation is performed using these techniques

1. Edge Detection Technique:

Image Segmentation by using Edge Detection technique determine the presence of edge or line in an image. "To define the edge, the boundary pixels that connect two separate regions with changing image amplitude attributes such as different constant luminance and tri-stimulus values in an image are considered".

Edge Detection means the boundary between two homogeneous regions. The technique is used to find background of the image and outlines of an object within an image. The technique refers to the process of identifying and locating sharp discontinuities in an image.

Edge Detection is done in following ways;

a) Filtering

Noise, which is "random variations in intensity values", is of following types:

i) Salt & pepper noise,

ii) Impulse noise and

iii) Gaussian noise

Salt and pepper noise contain random occurrences of both black and white intensity values. However, there is a trade-off between edge strength and noise reduction. More filtering to reduce noise will results in the loss of edge strength.

a) Enhancement

Edge detection is facilitated by "determine changes in intensity in the neighbourhood of a point". "Enhancement emphasizes pixels where there is a significant change in local intensity values and is usually performed by computing the gradient magnitude".

b) Detection

"Many points in an image have a nonzero value for the gradient, and not all of these points are edges for a particular application. Therefore, some method should be used to determine which points are edge points. Frequently, thresholding provides the criterion used for detection".

2. Thresholding Techniques:

When grey levels of pixels of objects are quite different from those of background, then objects can be easily separated from the background using thresholding techniques. Thresholding technique converts the grey scale image into a binary image of which “one state will indicate the foreground objects, while the other state will indicate the background.

Depending on the application, the foreground of the image can be represented by grey-level 0, that is, black as for text, and the background by the highest luminance for document paper that is 255 in 8-bit images, or conversely the foreground by white and the background by black”. Besides image segmentation these techniques are also used in applications such as, “Document image analysis, Map processing, Scene processing, Inspection of materials for quality, Cell images, Knowledge representation, Non-destructive testing, Ultrasonic, eddy current and thermal images, X-ray computed tomography and Endoscopic images etc.”

The thresholding techniques can be divided in to following six categories:

1. Histogram shape-based methods,
2. Clustering-based methods,
3. Entropy-based methods,
4. Object attribute-based methods,
5. Spatial methods,
6. Local methods

3. Clustering:

A set of objects when grouped into classes of similar objects based on some similarity measure, forms a cluster. The data objects within a cluster are similar to each other while dissimilar to those in another cluster. “By clustering, one can identify dense and sparse regions and therefore, discover overall distribution patterns and interesting correlations among data attributes”. Cluster analysis serves as a pre-processing step for other algorithms, such as classification, which would then operate on detected clusters. Hierarchical agglomerative clustering techniques start with as many clusters as there are unique values. Then pairs of clusters are successively merged till the optimal number of clusters is reached, depending on the termination condition. Termination condition is to be chosen carefully; else the hierarchical agglomerative clustering technique will ultimately yield one cluster containing all the values.

4. Classification:

The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data. In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups. The main goal of the Classification algorithm is to identify the category of a given dataset, and these algorithms are mainly used to predict the output for the categorical data.

Classification algorithms can be better understood using the below diagram. In the below diagram, there are two classes, class A and Class B. These classes have features that are similar to each other and dissimilar to other classes.

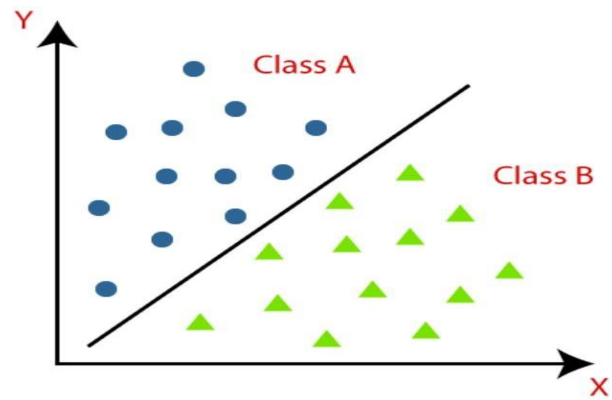

Fig.5 Fundamentals of Machine Learning

A classifier is the algorithm itself – the rules used by machines to classify data. A classification model, on the other hand, is the end result of your classifier's machine learning. The model is trained using the classifier, so that the model, ultimately, classifies your data.

There are both supervised and unsupervised classifiers. Unsupervised machine learning classifiers are fed only unlabeled datasets, which they classify according to pattern recognition or structures and anomalies in the data. Supervised and semi-supervised classifiers are fed training datasets, from which they learn to classify data according to predetermined categories.

At this point, no other diseases may be detected. Classifiers such as the Support Vector Machine and the Artificial Neural Network are often employed.

According to specified traits and allocating each illness to one of the preset groups, it is identifying a rule.

The overall architecture of our research work is represented in the following flowchart.

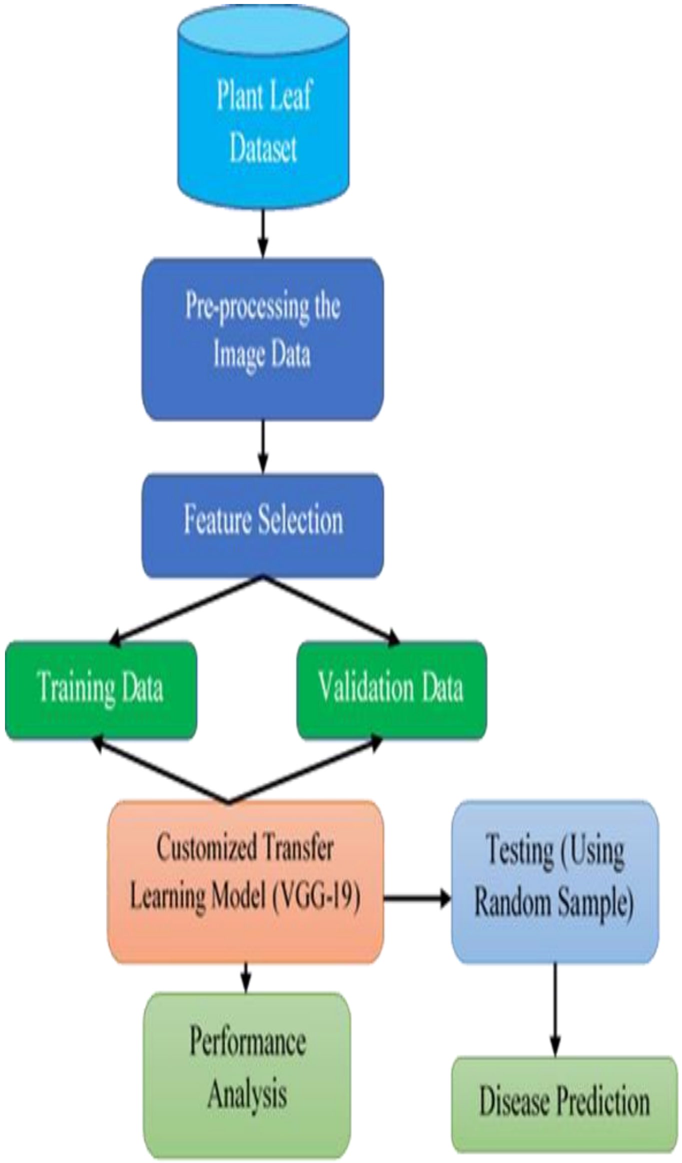

Fig.6 Overall Architecture of Proposed Model

Analysis of Plant disease and their symptoms:

1. Disease in plants:

It can occur on leaf, stem or any other part of the plant. The detection of disease in early stage is important task. Farmers need to have continuous monitoring which might be expensive. The farmers basically need fast and less expensive methodology that automatically detect the disease from symptoms that are on plant leaf. In most of the cases, symptoms of the disease in plants are seen on the stem, fruit and leaves. In the present work, symptoms of plant leaf have been considered for the detection of disease. In plants leaf, brown and yellow spots are common symptoms for general diseases. Early and late scorch, viral, bacterial and other fungal diseases are also generally found in plants.

2. Symptoms of plant disease:

Viral Disease:

Viral diseases are most difficult to diagnose. Viral disease does not show tell-tale signs so that they can easily observed and often confused with nutrient deficiencies. Some common viruses are leaf hoppers, whiteflies, aphids and cucumber beetle insects etc. with the carrier of disease mosaic virus, look for yellow or green stripes.

Bacterial Disease:

Bacterial diseases are spread out by rain, wind, birds or insects. The disease having symptoms on the leaf with tiny pole green spots on foliage, sometimes with yellow halo.

Fungal disease symptoms:

The fungal leaf diseases are late blight caused by fungus phytophthora infesters. This disease appears on older leaves like water soaked, grey green spots, in its early stages. When it gets matures, the spots get darken and white fungal growth form on the undersides.

VGG-19 Model:

VGG stands for Visual Geometry Group. As the name VGG19 indicates, it consists of 19 layers. Out of those 19 layers, 16 are convolution layers, 3 fully connected layers, 5 Max pool layers, and 1 Soft max layer.

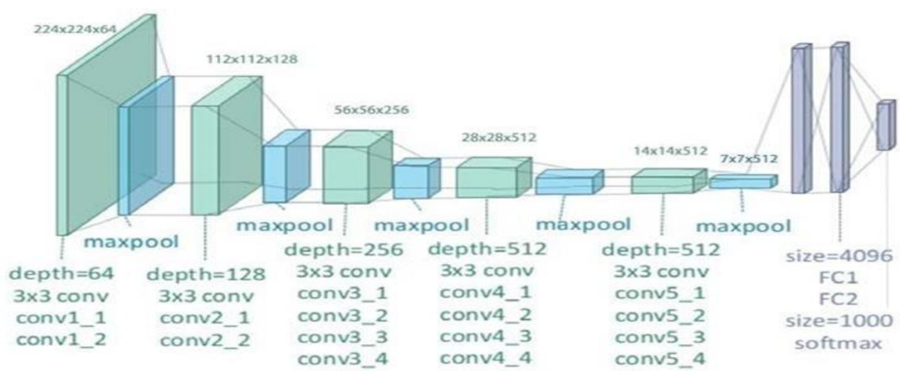

Fig. 7 VGG19 Architecture Diagram

Source: https://www.researchgate.net/figure/llustration-of-the-network-architecture-of-VGG-19-model-conv-means-convolution-FC-means_fig2_325137356

It is mostly used for transfer learning mechanisms. It is also named a good classification architecture for processing the input data.

The transfer learning model VGG-19 was customized and used in plant leaf disease prediction. The following diagram depicts the basic architecture of the VGG-19 model.

Architecture:

- A fixed size of (224 * 224) RGB image was given as input to this network which means that the matrix was of shape (224,224,3).
- The only pre-processing that was done is that they subtracted the mean RGB value from each pixel, computed over the whole training set.
- Used kernels of (3 * 3) size with a stride size of 1 pixel, this enabled them to cover the whole notion of the image.
- spatial padding was used to preserve the spatial resolution of the image.
- Max pooling was performed over a 2 * 2pixel windows with sride 2.
- This was followed by Rectified linear unit (ReLu) to introduce non-linearity to make the model classify better and to improve computational time as the previous models used tanh or sigmoid functions this proved much better than those.
- Implemented three fully connected layers from which first two were of size 4096 and after that a layer with 1000 channels for 1000-way ILSVRC classification and the final layer is a soft max function.

CHAPTER 5
EXPERIMENTATION & RESULTS

EXPERIMENTATIONS RESULTS

This section presents the dataset description and experimental results obtained from our implementation.

Dataset Description:

We analysed various images of the leaves of the plant. We rescaled all the images of leaves by using image processing and segmentation. The dataset consists of 38 class labels categorized into trained and validation data; as per the features and texture of the image input sent through the class labels, the plant's disease gets detected. We run all our experiments across a range of train-test set splits, namely 80–20 (80% of the whole dataset used for training and 20% for testing).

The following are the names of various class labels in the data set.

1. Apple Scab
2. Black Rot
3. Cedar Apple Rust
4. Apple Healthy
5. Blueberry healthy
6. Cherry healthy
7. Cherry Powdery Mildew
8. Corn Gray Leaf Spot
9. Corn Common Rust
10. Corn Healthy
11. Corn Northern Leaf Blight
12. Grape Black Rot
13. Grape Measles
14. Grape Healthy
15. Grape Leaf Blight
16. Orange Citrus Greening
17. Peach Bacterial Spot
18. Peach Healthy
19. Pepper Bell Bacterial Spot
20. Pepper Bell Healthy
21. Potato Early Blight
22. Potato Healthy
23. Potato Late Blight
24. Raspberry
25. Soyabean Healthy
26. Squash Powdery Mildew
27. Strawberry Healthy
28. Strawberry Leaf Scorch
29. Tomato bacterial spot
30. Tomato healthy
31. Tomato late blight
32. Tomato leaf mold
33. Tomato Septoria leaf spot
34. Tomato early blight
35. Tomato Spider mite
36. Tomato target spot
37. Tomato Mosaic Virus
38. Tomato Yellow Leaf Curl Virus

The following diagram shows the different class labels considered for experimentation.

Fig. 8 Samples of Different Class Labels in the Dataset

Experimental setup:

The system is configured with Intel(R) Core (TM) i7-8550U CPU @ 1.80GHz-1.99GHz and 16GB Random Access Memory (RAM). It has a 64bit operating system with X64 based processor. Here we used the Google Collaboratory platform to execute the code.

Results and Discussions:

The results obtained from the implementation of the VGG-19 model are represented in this section. We obtained 81.1% accuracy in plant leaf disease classification and prediction. The following snapshot gives the code implemented to print the model's accuracy.

```
acc = model.evaluate_generator(val)[1]

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

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: UserWarning: `Model.evaluate_generator` is deprecated and will be removed in a future
  """Entry point for launching an IPython kernel.
The accuracy of your model is = 81.09492659568787 %
```

Fig. Accuracy of the Model

The following figures show the comparison of training accuracy and validation accuracy.

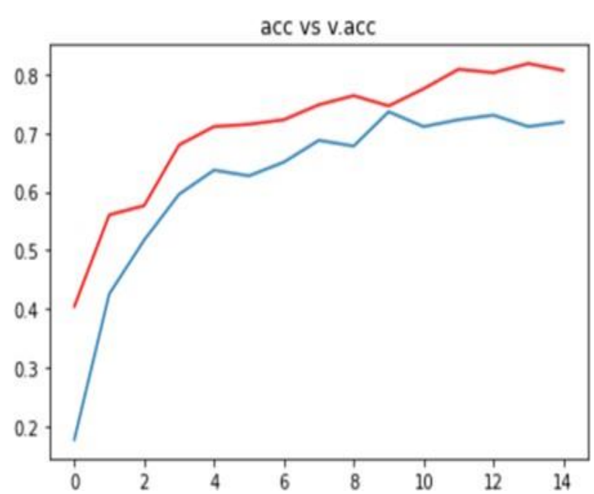

Fig. 9 Comparison of Training Accuracy and Validation Accuracy

The following figures show the comparison of training loss and validation loss

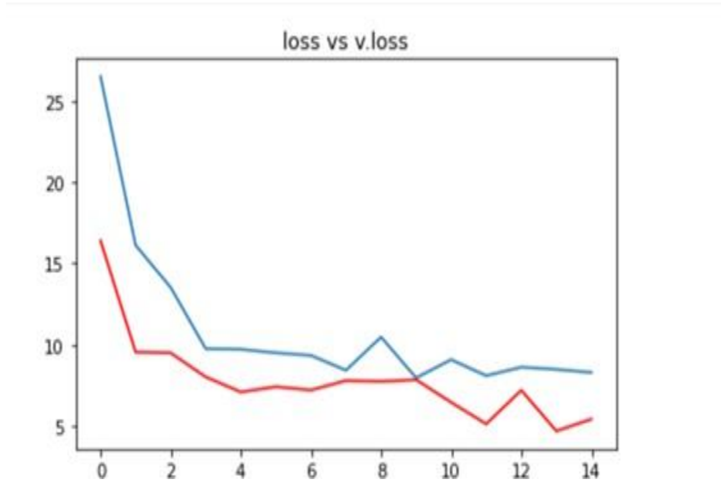

Fig. 10 Comparision of Training Loss and Validation loss

We tested the customized VGG-19 model by giving a random plant leaf image to predict whether it is infected or not, if it is infected, and displaying the name of the disease.

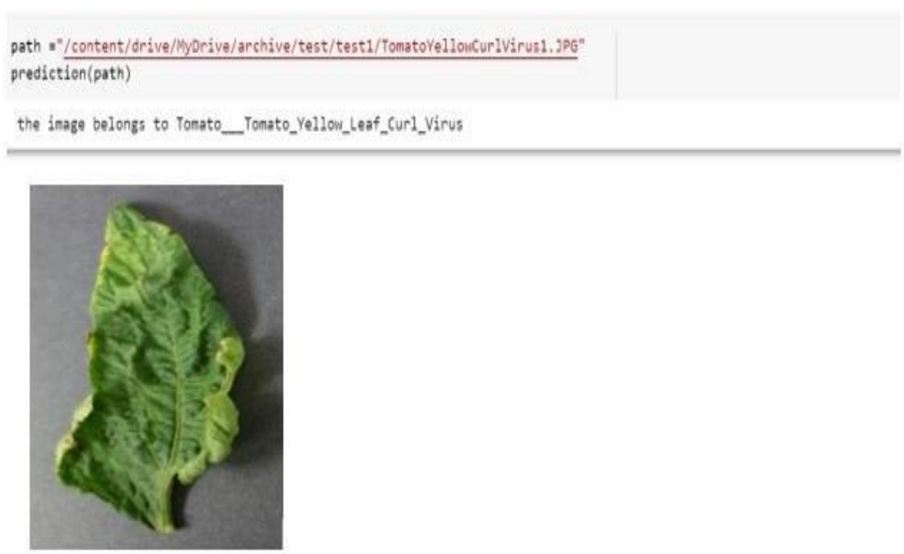

Fig. 11 Testing with a Random Image (Tomato_Yellow_Leaf_Curl_Virus)

The customized VGG-19 model efficiently predicts different types of plant leaf diseases. The model can predict 38 types of diseases from the plant leaves of different types of fruits and vegetables. The model accuracy is 81.1%. In the future, we will improve the accuracy of the model.

SOURCE CODE

```
from google.colab import drive
drive.mount('/content/drive')
root_path='drive/MyDrive/archive/New_Plant_Diseases_Dataset/New_Plant_Diseases_Dataset1'
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import os
import keras
from keras.preprocessing.image  import ImageDataGenerator, img_to_array, load_im
train_datagen= ImageDataGenerator(zoom_range=0.5, shear_range=0.3, rescale=1/255, horizontal_flip= True)
val_datagen = ImageDataGenerator(rescale= 1/255)
train = train_datagen.flow_from_directory(directory='/content/drive/MyDrive/archive/New_Plant_Diseases_Da
taset/New_Plant_Diseases_Dataset1/train',target_size=(256,256), batch_size=32)
val = val_datagen.flow_from_directory(directory='/content/drive/MyDrive/archive/New_Plant_Diseases_Datas
et/New_Plant_Diseases_Dataset1/valid', target_size=(256,256), batch_size=32)
Found 70411 images belonging to 38 classes.
Found 17572 images belonging to 38 classes.
print(type(train))
print(type(val))
<class 'keras.preprocessing.image.DirectoryIterator'>
<class 'keras.preprocessing.image.DirectoryIterator'>
t_img, label = train.next()
print(t_img)
print(label)
def plotImage(img_arr, label):
for im, l in zip(img_arr , label):
plt.figure(figsize=(5,5))
plt.imshow(im)
plt.show()
plotImage(t_img[1:4],  label[1:4])
from keras.layers import Dense, Flatten
from keras.models import Model
from keras.applications.vgg19 import VGG19
import keras
base_model = VGG19(input_shape=(256,256,3),
```

```
include_top= False)
for layer in base_model.layers:
layer.trainable = False
base_model.summary()
X = Flatten()(base_model.output)
X = Dense(units= 38, activation='softmax')(X)
#Creating our model
model = Model(base_model.input, X)
model.summary()
from keras.callbacks import ModelCheckpoint, Ear      lyStopping
# early stopping
es = EarlyStopping(monitor = 'val_accuracy', min_delta=0.01, patience=3, verbose=1)
# model check point
mc = ModelCheckpoint(filepath="best_model.h5",
            monitor = 'val_accuracy',
            min_delta= 0.01,
            patience = 3,
            verbose = 1 ,
            save_best_only = True)
cb = [es, mc]
 his = model.fit_generator(train,
              steps_per_epoch =16,
              epochs= 50,
              verbose=1,
              callbacks = cb,
              validation_data= val,
              validation_steps=16)
h.keys()
plt.plot(h['accuracy'])
plt.plot(h['val_accuracy'], c="red")
plt.title("acc vs v.acc")
plt.show()
plt.plot(h['loss'])
plt.plot(h['val_loss'], c="red")
plt.title("loss vs v.loss")
```

```
plt.show()
# load best model
from keras.models import load_model
model = load_model("/content/best_model.h5")
acc = model.evaluate_generator(val)[1]
print(f"The accuracy of your model is = {acc*100} %")
ref=dict(zip(list(train.class_indices.values()),
list(train.class_indices.keys())))
ref=dict(zip(list(train.class_indices.values()),
list(train.class_indices.keys())))
prediction(path):
img = load_img(path, target_size=(256,256))
i = img_to_array(img)
  im = preprocess_input(i)
  img = np.expand_dims(im , axis=0)
  pred =np.argmax (model. predict(img))
  print(pred)
path ="/content/drive/MyDrive/archive/test/test1/TomatoYellowCurlVirus1.JPG"
prediction(path)
```

CHAPTER 6
CONCLUSION

CONCLUSION

The central theme of this research is to predict the disease of plant leaves caused by different types of pathogens. In this approach, we used a transfer learning mechanism that initiates the usage of knowledge obtained from the previously implemented model. We took the dataset as input which consists of various images of leaves of certain fruits and vegetable plants. We take 38 class labels indicating various diseases of the train and validation datasets. Under pre-processing, the images of leaves are normalized and rescaled, and their size is customized. The transfer learning model VGG19 is implemented in 38 classes. The model's accuracy on the given dataset is 81.1%. In the future, we will improve the accuracy of the model.

REFERENCES

1. Anand.H.Kulkarni, A. P. R. K. (2012). Applying Image Processing technique to detect plant diseases. International Journal of Modern Engineering Research(IJMER), 28(1–2), 54–69. https://doi.org/10.1177/0958305X16685471
2. S.Arivazhagan, R. Newlin Shebiah, S. Ananthi, & Vishnu Varthini, S. (2013). Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Agricultural Engineering International: CIGR Journal, 15(1), 211–217. https://doi.org/10.1049/ip-f-2.1990.0064
3. Al Bashish, D., Braik, M., & Bani-Ahmad, S. (2010). A framework for detection and classification of plant leaf and stem diseases. Proceedings of the 2010 International Conference on Signal and Image Processing, ICSIP 2010, 113–118. https://doi.org/10.1109/ICSIP.2010.5697452
4. Nandini Kakran, Pratik Singh, K. P. J. (2019). Crop disease detection using image segmentation. International Journal of Computer Applications, 2(4), 190–194.
5. Sanjay B. Dhaygude, N. P. K. (2013). Agricultural plant Leaf Disease Detection Using Image Processing. 2(1), 599–602.
6. Revathi, P., & Hemalatha, M. (2012). Classification of cotton leaf spot diseases using image processing edge detection techniques. International Conference on Emerging Trends in Science, Engineering and Technology, 169–173. https://doi.org/10.1109/incoset.2012.6513900
7. Yan-Cheng Zhang, Han-Ping Mao, Bo hu, M.-X. L. (2007). Features selection of cotton disease leaves image based on fuzzy feature selection techniques. Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR '07, 1, 124–129. https://doi.org/10.1109/ICWAPR.2007.4420649
8. Haiguang Wang, Guanlin Li, Zhanhong Ma, X. L. (2012). Image recognition of plant diseases based on backpropagation networks. 2012 5th International Congress on Image and Signal Processing, CISP 2012, Cisp, 894–900. https://doi.org/10.1109/CISP.2012.6469998
9. H. AI-Hiary, S. Bani-Ahmad, M. Reyalat, M. B. and Z. Alr. (2011). Fast and Accurate Detection and Classification of Plant Diseases. International Journal of Computer Applications, 17(1), 31–38. https://doi.org/10.5120/2183-2754

10. M. Meunkaewjinda, P. Kumara, K. Attakitmongcol, A. S. (2008). Grape leaf disease detection from color imagery using hybrid intelligent system. Proceedings of ECTI-CON, 1, 513–516. https://doi.org/10.1109/ECTICON.2008.4600483
11. Lili li, Shujuan zhang, B. wang. (2021). Plant Disease Detection and Classification by Deep Learning - A Review. IEEE Access, 9(Ccv), 56683–56698. https://doi.org/10.1109/ACCESS.2021.3069646
12. Haiguang Wang, Guanlin Li, Zhanhong Ma, X. L. (2012b). Image recognition of plant diseases based on principal component analysis and neural networks. International Conference on Natural Computation, Icnc, 246–251. https://doi.org/10.1109/ICNC.2012.6234701
13. Song kai, Liu zhikun, Su hang, G. chunhong. (2011). A research of maize disease image recognition of corn based on BP networks. Third International Conference on Measuring Technology and Mechatronics Automation, ICMTMA, 1(2009921090), 246–249. https://doi.org/10.1109/ICMTMA.2011.66
14. Mokhled S, A.-T. (2013). An Emperical Investigation of Olive Leave Spot Disease Using Auto-Cropping Segmentation and Fuzzy C-Means Classification. World Applied Science Journal, 23(9), 1207–1211. https://doi.org/10.5829/idosi.wasj.2013.23.09.1000
15. Argenti, F., Alparone, L., & Benelli, G. (1990). Fast algorithms for texture analysis using co-occurrence matrices. IEEE Proceedings, 137(6), 443–448. https://doi.org/10.1049/ip-f-2.1990.006

PUBLICATION

Plant Leaf Disease Classification and Prediction Using Transfer Learning Model

Chandrika Bhargavi Achanta[1[0000-0002-1934-3893]], Kavuru Devi Keerthi[2[0000-0002-1607-1677]] and Sujatha Kamepalli[3 [1111-2222-3333-4444]]

[1] Graduate Student, Department of IT,
VFSTR Deemed to be University, Guntur, A.P., India.
[2] Graduate Student, Department of IT,
VFSTR Deemed to be University, Guntur, A.P., India.
[3]Associate Professor, Department of IT,
VFSTR Deemed to be University, Guntur, A.P., India.

Abstract. There is a significant productivity, financial damage due to plant diseases, and diminished overall quality of farm products. Detecting plant diseases has become more important in the surveillance of vast fields of crops in the modern-day. When it comes to disease management, farmers have difficulty transitioning from one strategy to another. The conventional method for detecting and identifying plant diseases is professional naked-eye inspection. This research examines the necessity for a simple technique for detecting plant leaf disease that would aid agricultural innovations. Early knowledge of crop health and disease detection may facilitate effective monitoring tactics. Crop yields will rise as a result of this method. In addition, the advantages and drawbacks of each of these prospective approaches are discussed in this study. Image capture, image analysis, extraction of features, and categorization based on neural networks are all part of the process. We get the best result to help the farmers through the processed methodology by implementing this model. The resulted accuracy of the implemented model is 81.09492659568787%. The proposed work enhances the farming culture to predict certain diseases and get a good yield of crops.

Keywords: Plant leaf Diseases, Deep Learning, VGG-19 Model, Crop Yield.

1 Introduction

Plants get affected by various diseases due to pathogens and environmental conditions. The disease occurrence varies from season to season. To get out of these crop diseases, farmers use different types of pesticides to control the disruption caused by the pathogen and provide security for the plant to get much yield as per their farming investment. If any plant of a particular crop is subjected to a disease outbreak, then the leaves of the plant change drastically with specific features. It is due to the infection caused by that disease, resulting in a decrease in crop yield and loss of crops. Incredibly less developed countries where admittance to disease control methods is minimum should face hunger and starvation. In this proposed algorithm, we use a transfer learning model to process the image input. Transfer learning is a machine learning technique where we use the knowledge obtained from the pre-trained model as the entry point for the model of a new task. It results in optimization that allows rapid progress when modeling the new task. It ensures higher performance even for a large amount of data. The target model obtained from the transfer learning model is highly efficient with reasonable accuracy. Out of various transfer learning models, we used VGG19 to recognize the disease by taking the plant leaves and processing those leaves images through the pre-trained model.

Fig. 1 Infected leaf of Apple plant

2 Literature Review

Several studies describe how to identify and treat the disorders. Techniques illustrating how to put them into practice, as seen and being the subject of this discussion, the possibility of finding diseased cotton leaves. (Anand.H.Kulkarni, 2012). A method in which the color and spot characteristics may be retrieved backward propagation of a self-organizing feature map a network of neurons detection of a diseased area on the leaves of a plant and texture traits are used to classify plant leaf diseases. An algorithm was devised by (S. Arivazhagan, R. Newlin Shebiah, S. Ananthi & Varthi, 2013) to process the data, transforming the supplied color. The red, green, and blue components of an RGB picture are green pixels disguised after being formed. Shape and form texturing characteristics are considered from the data. In order to classify objects, the Minimum Distance is used. Vector machines for criterion evaluation and support (SVMs). (Al Bashish et al., 2010), they have developed a framework for the Identification and Classification of Plant Diseases of the Leaf and Stem. The K-Means algorithm uses to segment and transforms RGB pictures into HIS Tonal system. Color and texture characteristics are then calculated. A statistical neural network classifier. Classification is used for categorization purposes. Image segmentation has been hypothesized by (Nandini Kakran & Pratik Singh, 2019) for crop disease detection. The three-part technique is as follows: A device-independent color format, RGB, was used to store the leaf pictures at first. To separate the photos, they had to be resized and converted to the CIELAB color space independent of hardware. Region-based extraction of the contaminated area is the second method. The use of segmentation was made. An essential part of K-means clustering is A method used during the segmentation stage. Twelfth extract characteristics are based on color, shape, and texture. Typically used to describe a geographic area. As a result, The co-occurrence matrix of characteristics, wavelet, and grey levels. It was necessary to make use of certain strategies. Classifying apples using a support vector regression (SVR) algorithm disease of the leaves. Agricultural plant disease has been hypothesized by (Sanjay B. Dhaygude, 2013); a prototype software approach for detecting rice disease based on diseased photos of diverse rice plants is described in Identification Using Pattern Recognition Techniques. Photos of diseased rice plants are taken using a digital camera, and sick sections of the plants are identified using image growth and image segmentation algorithms. After that, a neural network is used to classify the affected leaf tissue. Cotton Leaf Spot Diseases Detection Using Feature Selection with Skew Divergence Approach has been suggested by (Dheeb AI Bashish, Malik Braik, 2010). In this study, the upgraded PSO feature selection method uses the skew divergence method and extracts Edge, CYMK color, GA, Color, and Texture variations features. The collected features were analyzed using SVM, BPN, and Fuzzy with Edge Selection & Classification. (Yan-Cheng Zhang, Han-Ping Mao, Bo hu, 2007), A panel of experts analyzed the images on the remote server. Computer vision methods are used to identify and classify damaged areas in a picture. The disease-affected lesions are segmented using a straightforward color difference technique. With the use of a mobile phone notification, the expert may assess the study's findings and offer comments to the farmers. An image recognition system capable of identifying crop diseases is the purpose of this study. The scanned color picture of a diseased leaf is the first step in image processing. These photos are segmented using a mathematical morphology technique. These attributes were then utilized in conjunction with a classification approach called the membership function to separate the three categories of illnesses. (Haiguang Wang, Guanlin Li, Zhanhong Ma, 2012a) cotton leaf disease was identified and classified using image processing and machine learning approaches. The poll results on background removal and segmentation approaches were also brought up throughout the discussion. A translation from RGB to HSV color space for background removal was discovered during this research. In addition, we discovered that thresholding produces superior results compared to other methods of removing background noise. We used thresholding to the masked image after performing color segmentation by masking green pixels in the background removed image to produce a binary picture. This segmentation process is important in extracting disease-specific characteristics. For illness categorization, we observed that SVM performed well in accuracy. A total of three of our recommended processes have been implemented: Image Acquisition, Image Preprocessing, and Image Segmentation. Our proposed work has five key parts. (H. AI-Hiary, S. Bani-Ahmad, M. Reyalat, 2011) After taking a picture, the collected pictures are initially enhanced using Image Edge detection and segmentation algorithms. Image segmentation is conducted out using the R, G, and B

Color Feature images (disease spots). An image feature extraction process is then used to identify disease areas and implement a pest management strategy based on these attributes. Cotton leaf spot, cotton leaf color segmentation, and image segmentation based on edge detection are included in this study's analysis and categorization of illness. (M. Meunkaewjinda, P. Kumsawat, K. Attakitmongcol, 2008) We propose and test a new software method for detecting and classifying plant illnesses using Image Processing. Few agricultural professionals in rural India can examine crop photographs and provide advice. Farmers often get expert solutions to their inquiries too late because of the delay in expert responses. Image processing algorithms based on color, texture, and shape will be developed in this article to identify crop illnesses or other issues that may impact crops from photos and provide farmers with the most up-to-date information through SMS. As a result, more efficient use of chemicals will lead to increased production and a better-quality output due to these technologies. (Lili li, Shujuan zhang, 2021), Using deep learning and advanced imaging techniques, the disadvantages in plant disease recognition get removed, and the efficiency of the proposed algorithm is also increased by using a particular methodology in processing the application. Certain visualization techniques are used to organize deep learning architectures. A roll of metrics is used for training and validation. Mean Average Precision (MAP) is an indicator for evaluating the framework. (Haiguang Wang, Guanlin Li, Zhanhong Ma, 2012b) In this proposed technique, Principal component analysis (PCA) and Backpropagation networks (BP) are used to predict two kinds of wheat diseases. Generalized Regression Networks GRNNs and Probabilistic Neural Networks PNNs are used for classification to identify the grape diseases. The fitting and prediction accuracy of the class labels are identified via these processing algorithms. (Song kai, Liu zhikun, Su hang, 2011) Based on image processing and analysis for maize crop disease prediction, image recognition of corn leaf. Texture features of corn disease and segmentation of disease spots are used to recognize the disease. The backpropagation neural network (BP) mechanism is also implemented to coordinate disease detection input/output relations. A spatial gray matrix is used for fast identification of the maize disease. (Mokhled S, 2013) image analysis and classification techniques for spotted disease detection of Olive leaves are used for the research. Various images of leaves were collected from the field and taken as samples. C- means clustering is used to detect the defect. The severity of the disease is determined via the classification of leaf areas. As a result, 86% accuracy is obtained through the processed algorithm. (Argenti et al., 1990) An algorithm for fast analysis is proposed for calculating the parameters of matrices. A supervised learning mechanism is used for reading the input data. Statistical features get extracted through classification and segmentation.

3 Objectives

The primary contributions of this research are:

1. We understand the features of infected and uninfected leaves of different fruits and vegetable crops.
2. We explore the existing machine learning and deep learning models used in plant leaf disease detection.
3. We are developing deep transfer learning models to detect plant leaf diseases and customizing the model to work with different class labels.
4. We are testing the model with randomly selected plant leaf images to predict whether the given leaf is infected or uninfected, finding the disease name.

4 Methodology

The basic steps involved in plant leaf disease detection are given below.

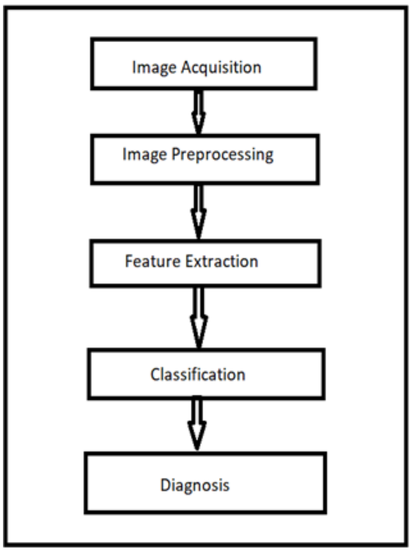

Fig. 2 Basic Steps for Disease Detection Algorithm

4.1 Acquisition of images

High-quality RGB color photographs are taken using a digital camera with the necessary resolution. The project explicitly dictates how an image database is built. The classifier, which determines the algorithm's resilience, is made more efficient by the picture database.

4.2 Processing and segmentation of images

During the pre-processing stage, background and noise are removed from the picture data, and unwanted distortions are suppressed. It improves picture processing and analysis capabilities.
HIS and CIELAB color spaces are used to transform the RGB color picture. When HSI and CIELAB were developed as color-independent space models, they were also based on human perception.
Segmentation is the first step in locating the area that is contaminated.
K-means clustering and edge detection algorithms are often used for segmentation.
A variety of characteristics are used to define the contaminated area after segmentation. For describing a location, color, texture, and form are often employed.
Color characteristics are essential for determining the visual surroundings, identifying objects, and conveying information.
An object's texture is one of its most critical identifying characteristics. Image retrieval benefits from its use as a robust regional descriptor.
The texture is described, by contrast, as homogeneity, dissimilarity, energy, and entropy. An image's shape is one of the most fundamental properties that may be used to describe its content.

4.3 Classification

At this point, no other diseases may be detected. Classifiers such as the Support Vector Machine and the Artificial Neural Network are often employed.
According to specified traits and allocating each illness to one of the preset groups, it is identifying a rule.

The transfer learning model VGG-19 was customized and used in plant leaf disease prediction. The following diagram depicts the basic architecture of the VGG-19 model. The overall architecture of our research work is represented in the following flowchart.

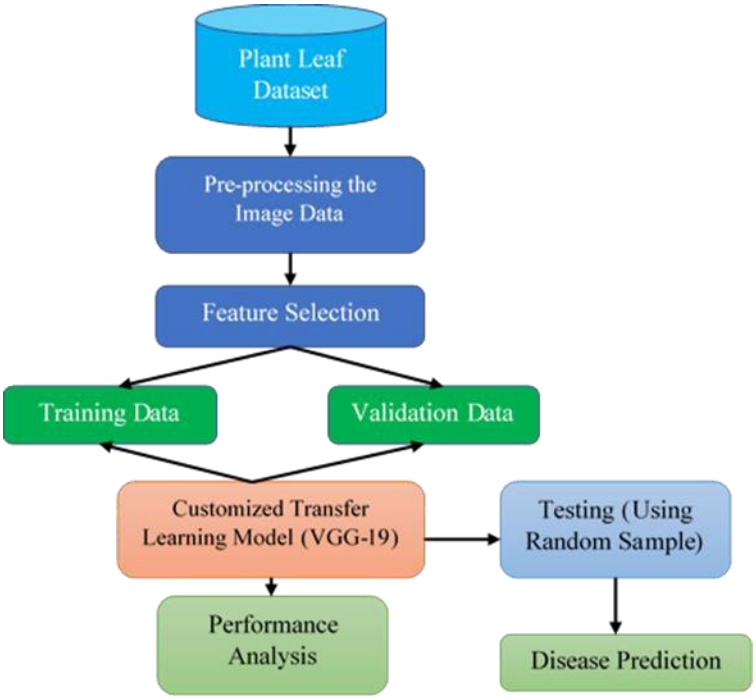

Fig.3 Overall Architecture of Proposed Model

4.4 VGG-19 Model

VGG stands for Visual Geometry Group. As the name VGG19 indicates, it consists of 19 layers. Out of those 19 layers, 16 are convolution layers, 3 fully connected layers, 5 Max pool layers, and 1 Softmax layer. It is mostly used for transfer learning mechanisms. It is also named a good classification architecture for processing the input data.

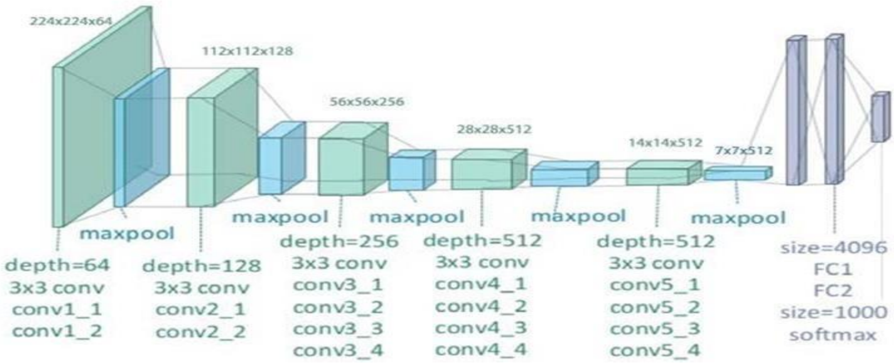

Fig. 4 VGG19 Architecture Diagram

Source: https://www.researchgate.net/figure/llustration-of-the-network-architecture- **of-VGG-19-model-conv-means-convolution-FC-means_fig2_325137356**

5 Experimentation Results

This section presents the dataset description and experimental results obtained from our implementation.

5.1 Dataset Description

We analyzed various images of the leaves of the plant. We rescaled all the images of leaves by using image processing and segmentation. The dataset consists of 38 class labels categorized into trained and validation data; as per the features and texture of the image input sent through the class labels, the plant's disease gets detected. We run all our experiments across a range of train-test set splits, namely 80–20 (80% of the whole dataset used for training and 20% for testing). The following diagram shows the different class labels considered for experimentation.

Fig. 5 Samples of Different Class Labels in the Dataset

5.2 Experimental setup

The system is configured with Intel(R) Core (TM) i7-8550U CPU @ 1.80GHz-1.99GHz and 16GB Random Access Memory (RAM). It has a 64bit operating system with X64 based processor. Here we used the Google Collaboratory platform to execute the code.

5.3 Results and Discussions

The results obtained from the implementation of the VGG-19 model are represented in this section. We obtained 81.1% accuracy in plant leaf disease classification and prediction. The following snapshot gives the code implemented to print the model's accuracy.

```
acc = model.evaluate_generator(val)[1]

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

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: UserWarning: `Model.evaluate_generator` is deprecated and will be removed in a future
  """Entry point for launching an IPython kernel.
The accuracy of your model is = 81.09492659568787 %
```

Fig. 6 Accuracy of the Model

The following figures show the comparison of training accuracy and validation accuracy.

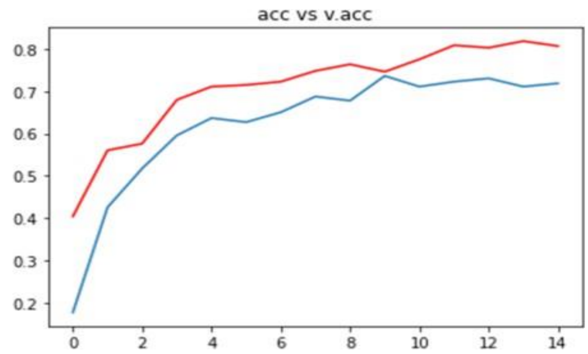

Fig. 6 Comparison of Training Accuracy and Validation Accuracy

The following figures show the comparison of training loss and validation loss.

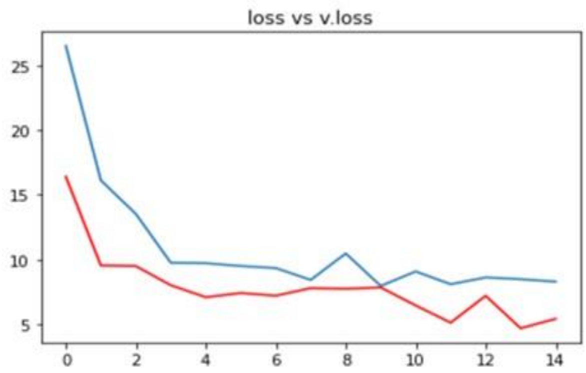

Fig. 7 Comparision of Training Loss and Validation loss

We tested the customized VGG-19 model by giving a random plant leaf image to predict whether it is infected or not, if it is infected, and displaying the name of the disease.

```
path ="/content/drive/MyDrive/archive/test/test1/TomatoYellowCurlVirus1.JPG"
prediction(path)

the image belongs to Tomato___Tomato_Yellow_Leaf_Curl_Virus
```

Fig. 8 Testing with a Random Image (Tomato_Yellow_Leaf_Curl

The customized VGG-19 model efficiently predicts different types of plant leaf diseases. The model can predict 38 types of diseases from the plant leaves of different types of fruits and vegetables. The model accuracy is 81.1%. In the future, we will improve the accuracy of the model.

6 Conclusion

The central theme of this research is to predict the disease of plant leaves caused by different types of pathogens. In this approach, we used a transfer learning mechanism that initiates the usage of knowledge obtained from the previously implemented model. We took the dataset as input which consists of various images of leaves of certain fruits and vegetable plants. We take 38 class labels indicating various diseases of the train and validation datasets. Under pre-processing, the images of leaves are normalized and rescaled, and their size is customized. The transfer learning model VGG19 is implemented in 38 classes. The model's accuracy on the given dataset is 81.1%. In the future, we will improve the accuracy of the model.

References

1. **Anand.H.Kulkarni, A. P. R. K. (2012). Applying Image Processing technique to detect plant diseases. International Journal of Modern Engineering Research(IJMER), 28(1–2), 54–69. https://doi.org/10.1177/0958305X16685471**
2. **S.Arivazhagan, R. Newlin Shebiah, S. Ananthi, & Vishnu Varthini, S. (2013). Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Agricultural Engineering International: CIGR Journal, 15(1), 211–217. https://doi.org/10.1049/ip-f-2.1990.0064**
3. **Al Bashish, D., Braik, M., & Bani-Ahmad, S. (2010). A framework for detection and classification of plant leaf and stem diseases. Proceedings of the 2010 International Conference on Signal and Image Processing, ICSIP 2010, 113–118. https://doi.org/10.1109/ICSIP.2010.5697452**
4. **Nandini Kakran, Pratik Singh, K. P. J. (2019). Crop disease detection using image segmentation. International Journal of Computer Applications, 2(4), 190–194.**
5. **Sanjay B. Dhaygude, N. P. K. (2013). Agricultural plant Leaf Disease Detection Using Image Processing. 2(1), 599–602.**
6. **Revathi, P., & Hemalatha, M. (2012). Classification of cotton leaf spot diseases using image processing edge detection techniques. International Conference on Emerging Trends in Science, Engineering and Technology, 169–173. https://doi.org/10.1109/incoset.2012.6513900**
7. **Yan-Cheng Zhang, Han-Ping Mao, Bo hu, M.-X. L. (2007). Features selection of cotton disease leaves image based on fuzzy feature selection techniques. Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR '07, 1, 124–129. https://doi.org/10.1109/ICWAPR.2007.4420649**
8. **Haiguang Wang, Guanlin Li, Zhanhong Ma, X. L. (2012). Image recognition of plant diseases based on backpropagation networks. 2012 5th International Congress on Image and Signal Processing, CISP 2012, Cisp, 894–900. https://doi.org/10.1109/CISP.2012.6469998**
9. **H. AI-Hiary, S. Bani-Ahmad, M. Reyalat, M. B. and Z. Alr. (2011). Fast and Accurate Detection and Classification of Plant Diseases. International Journal of Computer Applications, 17(1), 31–38. https://doi.org/10.5120/2183-2754**
10. **M. Meunkaewjinda, P. Kumsawat, K. Attakitmongcol, A. S. (2008). Grape leaf disease detection from color imagery using hybrid intelligent**

system. Proceedings of ECTI-CON, 1, 513–516. https://doi.org/10.1109/ECTICON.2008.4600483

11. **Lili li, Shujuan zhang, B. wang. (2021). Plant Disease Detection and Classification by Deep Learning - A Review. IEEE Access, 9(Ccv), 56683–56698. https://doi.org/10.1109/ACCESS.2021.3069646**
12. **Haiguang Wang, Guanlin Li, Zhanhong Ma, X. L. (2012b). Image recognition of plant diseases based on principal component analysis and neural networks. International Conference on Natural Computation, Icnc, 246–251. https://doi.org/10.1109/ICNC.2012.6234701**
13. **Song kai, Liu zhikun, Su hang, G. chunhong. (2011). A research of maize disease image recognition of corn based on BP networks. Third International Conference on Measuring Technology and Mechatronics Automation, ICMTMA, 1(2009921090), 246–249. https://doi.org/10.1109/ICMTMA.2011.66**
14. **Mokhled S, A.-T. (2013). An Emperical Investigation of Olive Leave Spot Disease Using Auto-Cropping Segmentation and Fuzzy C-Means Classification. World Applied Science Journal, 23(9), 1207–1211. https://doi.org/10.5829/idosi.wasj.2013.23.09.1000**
15. **Argenti, F., Alparone, L., & Benelli, G. (1990). Fast algorithms for texture analysis using co-occurrence matrices. IEEE Proceedings, 137(6), 443–448. https://doi.org/10.1049/ip-f-2.1990.006**