A Report

On

IMAGE BASED PLANT LEAF DISEASE DETECTION USING DEEP LEARNING

Submitted in the partial fulfilment of requirements for The award of degree of

BACHELOR OF TECHNOLOGY

In

INFORMATION TECHNOLOGY

Submitted By

A. Razak (191FA07051)

N. Jyothi Swaroop (191FA07135)

Under the guidance of

B NAGA SUDHEER

Assistant Professor

Department of IT & CA

VIGNAN'S

Foundation for Science, Technology & Research

(Deemed to be UNIVERSITY)

-Estd. u/s 3 of UGC Act 1956

NAAC Accredited A+

DEPARTMENT OF IT & CA

2022-2023

VIGNAN'S
Foundation for Science, Technology & Research
(Deemed to be UNIVERSITY)
-Estd. u/s 3 of UGC Act 1956
NAAC Accredited A+

CERTIFICATE

This is to certify that the Socio Centric Project Report entitled **"IMAGE-BASED PLANT DISEASES DETECTION USING DEEP LEARNING"** is being submitted by **Abdul Razak (191FA07051), N. Jyothi Swaroop (191FA07135)** in partial fulfilment for the award of B. Tech Degree in Information Technology at Vignan's Foundation for Science, Technology and Research, deemed to be University. It is a record of bonafide work carried out by them in Department of Information Technology, Vignan's Foundation for Science Technology and Research under the supervision of **B NAGA SUDHEER.**

Signature of Project Guide
(B Naga Sudheer)

Signature of HOD
(Dr. N. Veeranjanyulu)

Signature of External

TABLE OF CONTENTS

Topics	Page No
Declaration	i
Acknowledgements	ii
Abstract	iii
1. Introduction	
1.1 Plant Leaf Disease	2
1.2 Need of Plant Leaf Disease Detection	3
1.3 Technologies Behind Disease Detection	4
1.4 Machine Learning	5
1.5 Machine Learning Techiques	7
1.6 Why Deep Learning	8
1.7 Convolutional Neural Network	11
1.8 Image Classification Techniques	12
1.9 Motivation, Objectives and Scope of the work	15
1.10 Organization of Report	17
2. Literature Survey	
2.1 Tomato Leaf Disease Detection Using Deep Learning	19
2.2 Reviewing Important Aspects of Plant Leaf Disease Detection and Classification	19
2.3 Plant Pathology Disease Detection in Apple Leaves Using CNN	20
2.4 Plant Leaf Disease Detection and Classification based On CNN	20
2.5 Plant Leaf Detection and Disease Recognition Using Deep Learning	21

2.6 Classification of Plant Leaf and Disease Using ML 21
2.7 Detection of plants Disease from Unhealthy Plant Leaves Using ML 22
2.8 Detection of Unhealthy Region of Plant Leaves Using Image Processing 23
2.9 Application Research of Plant Leaf Pests and Disease Base On Unsupervised Learning 23
2.10 Rose Plant Leaves Disease Detection Using CNN 24
2.11 Plant Leaf Disease Detection Using Deep Learning 25
2.12 Plant Disease Detection and Classification Using Deep Learning ConvNets 25
2.13 Machine Learning for Plant Leaf Disease Detection and Classification 26
2.14 Plant Leaf Disease Classification and Detection System Using Machine Learning 26
2.15 Identification of Plant Leaf Disease Using Machine Learning Algorithms 27

3. Plant Leaf Disease Detection
3.1 About Model and Learning 29
3.2 Overview of Existing Work 32
3.3 Selection of Classifiers 37
3.4 Proposed System 41
3.5 Experiment Analysis 42
3.6 Baseline Results 46

4. Results
4.1 Results of a Web Application 50

5. Conclusion
5.1 Summary 53
5.2 Conclusion 54
5.3 Future Work and Extension 55
5.4 References 56

TABLE FOR FIGURES

Topics	Page No.
Figure 1.1 Disease of a Plant	2
Figure 1.2 Agriculture Losses Comparison	4
Figure 1.3 Difference Between AL, ML and DL	6
Figure 1.4 Types of Machine Learning	7
Figure 1.5 Convolutional Neural Networks Architecture	9
Figure 1.6 Image Classification	12
Figure 3.1 Block Diagram of Image Classifier	30
Figure 3.2 Samples Images from Datasets	31
Figure 3.3 Image after Preprocessing	31
Figure 3.4 Image Converted to HSV Color Space	32
Figure 3.5 Images After K Means Clustering	32
Figure 3.6 Images After Removal of Background	33
Figure 3.7 Operation of Convolution Layer without Padding	35
Figure 3.8 Operation of Convolution Layerwith Padding	36
Figure 3.9 Pooling Operation	36
Figure 3.10 Naive inception module	37
Figure 3.11 Statistics of PlantDoc Dataset	40
Figure 3.12 Data Collection	41
Figure 3.13 Training Vs Validation Accuracy	45
Figure 4.1 Home Page	47
Figure 4.2 Selection of Images from a Datasets	47
Figure 4.3 Result and Analyze of a plant	48

DECLARATION

We hereby declare that the Socio Centric Project report entitled **“IMAGE-BASED PLANT DISEASES DETECTION USING DEEP LEARNING”** submitted to the Department of Information Technology, Vignan’s Foundation for Science, Technology and Research, deemed to be University. This report is the work done by us in the Department of Information Technology.

Place:

Date: **Signature of Students**

ACKNOWLEDGEMENTS

We express our gratitude towards the Management for providing opportunity to work and implement Socio Centric project.We feel it our responsibility to thank **B Naga Sudheer** under whose valuable guidance that the project came out successfully after each stage. It is a great pleasure for us to express our sincere thanks to Prof. Dr. N. Veeranjanyulu, HOD, IT of VFSTR Deemed to be University, for providing me an opportunity to do Socio Centric project.We extend our wholehearted gratitude to all our faculty members of Department of Information Technology who helped us in our academics throughout course.Finally, we wish to express thanks to our family members for the love and affection overseas and forbearance and cheerful depositions, which are vital for sustaining effort, required for completing this work.

With Sincere regards,

B. Gowtham Sai
(191FA07098)

T. Sai Swaroop
(191FA07126)

G. Bhanu Prakash
(191FA07132)

ABSTRACT

Plant diseases are a major threat to farmers, consumers, environment and the global economy. In India alone, 35% of field crops are lost to pathogens and pests causing losses to farmers. Indiscriminate use of pesticides is also a serious health concern as many are toxic and bio magnified. These adverse effects can be avoided by early disease detection, crop surveillance and targeted treatments. Most diseases are diagnosed by agricultural experts by examining external symptoms. However, farmers have limited access to experts. Our project is the first integrated and collaborative platform for automated disease diagnosis, tracking and forecasting. Farmers can instantly and accurately identify diseases and get solutions with a mobile app by photographing affected plant parts. Real-time diagnosis is enabled using the latest Artificial Intelligence (AI) algorithms for Cloud-based image processing. The AI model continuously learns from user uploaded images and expert suggestions to enhance its accuracy. Farmers can also interact with local experts through the platform. For preventive measures, disease density maps with spread forecasting are rendered from a Cloud based repository of geo-tagged images and micro-climactic factors. A web interface allows experts to perform disease analytics with geographical visualizations. In our experiments, the AI model (CNN) was trained with large disease datasets, created with plant images self-collected from many farms over 7 months. Test images were diagnosed using the automated CNN model and the results were validated by plant pathologists. Over 95% disease identification accuracy was achieved. Our solution is a novel, scalable and accessible tool for disease management of diverse agricultural crop plants and can be deployed as a Cloud based service for farmers and experts for ecologically sustainable crop production. Not only as an agricultural economy but also with a large amount of population to feed, it is necessary that leaf diseases in plants are detected at a very early stage and predictive mechanisms to be adopted to make them safe and avoid losses to the agri-based economy.

Chapter - 1

1. INTRODUCTION

1.1 Plant Leaf Disease

India is a growing economic giant and more than 65% of the population is either directly associated with agriculture or agriculture products. The crops are suffered the major losses due to plant diseases and insect damage. In the approximate figures, the worldwide annual production tonnages lost due to various pests at the start of the 21st century. The Losses due to plant diseases contribute to around 15 to 17% of the total accumulated losses over the annual production range and this is highly alarming. A total of 68% average annual loss of crop production is spoilt as loss that is being caused by several factors such as pests, weeds, and plant leaf diseases. This causes a big blow to the economy. Crop enhancement and protection results, based on paramount global practices and the new technologies available are the answer. There are many emerging trends and promising solutions for sustainable crop protection which include fustigation, agronomy, chemicals, seed treatment, and bio-technology growth and use of sustainable expertise to identify the crop disease at the earliest.

Figure 1.1 Disease of a Plant

In the country, the next generation of agriculture has to include all possible promising solutions using in a specified situation. The sector has enormous unrealized possible for growth and incredibly low-level application of crop protection chemicals, as compared to the universal norms joined with the class for educated farming, fast increasing the awareness in young, etc. The best solution to the problem is to identify the disease of the plant so that precautionary steps can be taken to safe guard the same. This paper implements the concept of applying convolutional neural network implementation to the detection of leaf disease in the tomato plant and suggests a suitable solution to the farmer to recover the same.

The Losses due to plant diseases contribute to around 15 to 17% of the total accumulated losses over the annual production range and this is highly alarming. A total of 68% average annual loss of crop production is spoilt as lossthat is being caused by several factors such as pests, weeds, and plant leaf diseases. This causes a big blow to the economy. Crop enhancement and protection results, based on paramount global practices and the new technologies available are the answer. There are many emerging trends and promising solutions for sustainable crop protection which include fustigation, agronomy, chemicals, seed treatment, and bio-technology growth and use of sustainable expertise to identify the crop disease at the earliest.

1.2 Need of Plant Leaf Disease Detection

In India nearly 70% of the residents makes livelihood based on the farming practices. Numerous technologies and studies were urbanized to deliver smart agricultural systems and enrich the quantity and quality of productivity. In order to overcome the requirement of growing population, it is required to create a major advancement in agricultural productiveness. Rice is the most common vital food for the people in India. Aiming on paddy crop disease management is important to analyze the occurrence of paddy crop diseases such as Brown spot, Sheath blight, Leaf smut and Bacterial leaf blight.

These diseases terminates 10 to 12% of production in India. So early detection of these diseases is very important to achieve a proficient production. The existing method of disease detection is a manual approach which takes a lot of time. Apprehending the images of infested leaves and process those by using an automated system could be a smart solution for farmers. Many number of Image processing techniques are regularly using in agricultural sector for the exposure, detecting and analyzing disease invasions of plants. At present, deep learning methods are becoming popular in this area.

Deep learning is a subclass of Artificial Intelligence. It is a progressive and innovative technique which makes use of a popular mechanism called neural networks. A neural network is popular among various deep learning modules which are extensively using with the combination of image processing technique. A CNN with minimal procedure will identify as well as catalogue. It is proficient in assessing images and the features needed by using its layered structure. The CNN comprises of 4 layers which are images as input, convolutional and pooling layers, fully connected layer and finally the required output image.

Economic growth of agricultural countries like India is mostly dependent on food crops, but due to several deadly diseases found on leaves of plants causes major loss in the production of several agricultural products. Also, many plants have importance from the medicinal point of view. Ayurvedic medicine industry requires identifying and saving Ayurvedic plants from the diseases. Numerous stateof-the-art approaches for accurate plant leaf disease detection and its classification were invented and studied by researchers in the past. The conventional methodology of the said process was very time consuming and at the same time not always accurate, since farmers or researchers required to appoint an expert in the field to monitor the growth of the plant and suggest remedies. Today is the era of technology and Artificial Intelligence (AI) is the need of the day in every sector. Recently, AI has achieved tremendous appreciation in agricultural field, especially in identifying plants, their diseases and classification. Researchers have employed several approaches in literature to gain good amount of accuracy in the mentioned work. They have researched and touched different aspects of plant diseases under various domains and inspire nowadays researchers to research further considering the challenges they have encountered. This paper reviews notable aspects in the field of plant leaf disease detection and its classification, which can be a guide for future researchers in the said field.

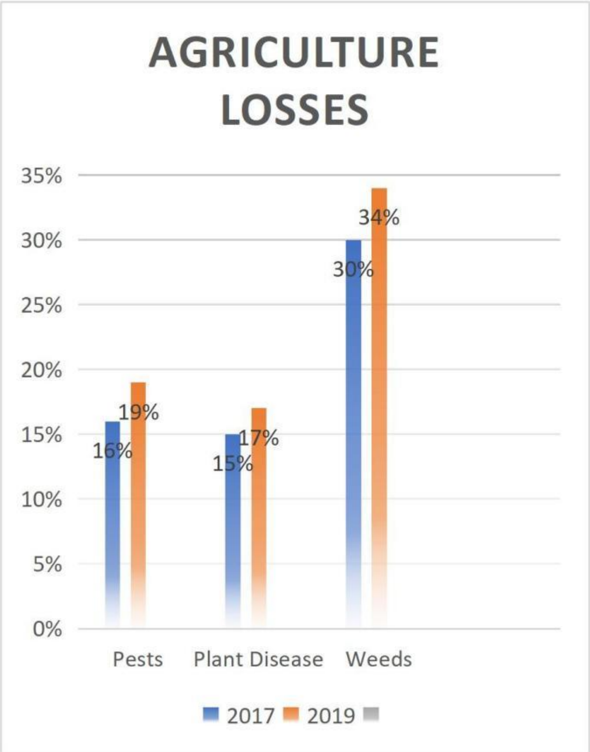


Figure 1.2 Agriculture Losses Comparison

The entire paper is designed as follows. The general concept of plant leaf disease detection and classification. A detailed taxonomy of different diseases that occurs in plants. Several existing methodologies from the literature under different domains. Dedicated to dataset that is predominantly used for experimentation in the said area which concludes the work giving future directions.

1.3 Technologies Behind Disease Detection

New development in computer science have recently been developed field of Machine Learning, which also helps to automate assessment and processing carried out by humans so their manual efforts can be reduce to a minimum. The goal of Machine Learning is to enable computers to learn without explicitly programming, depending on the technological challenge. This is because machine learning depends on computer programs that can be changed as new data are released. The diseases self-recognition in the is based on the symptoms of the disease. This helps farmers, analysts and scholars to collect reports on the incidence of the disease rapidly and reliably. This limits the tracking of human beings in the broader sector. The aim is to derive the characteristic of the area in disease identification from image. The various other diseases include rust, kole roga, yellow leaf disease, leaf rot, leaf curl, angular leaf spot, leaf spot, late blight, bacteria wilt, etc., which all affect the growth of the plant. Plant diseases depend on climatic conditions too. For instance, reddish purple leaf spots on the older leaves of the plant are due to the entomosporium leaf spot fungus. Furthermore, this disease is very dangerous because in cool and wet weather conditions.

We also proposed in this paper a novel paradigm for detecting plant leaf diseases based on a deep Convolutionary neural network (strong CNN). A free data collection of 39 separate plant leaf groups and context photos trains the Deep CNN model. Six types of data increase methods were used: picture fl-flipping, gamma correction, injection noise, PCA color increase, rotation and scaling. We have observed that data increase can increase the model's performance.

Apples are grown all across the world and are one of the most efficiently and widely cultivated fruits contributing majorly to global yield. However, various diseases cause substantial loss to apple production. Therefore, accurate and timely detection of apple leaf diseases is crucial. In this paper, the dataset is provided by Cornell Initiative for Digital Agriculture (CIDA) consisting of about 3600 training images categorised and classified into 4 classes, namely healthy, scab, rust and multiple diseases on the same leaf. While scab is a fungal plant disease characterised by crustaceous lesions on leaves and fruits, rust is another fungal ailment with tiny specks that range in many brown shades.

1.4 Machine Learning

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that makes it more similar to humans: The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect. Images from training dataset with class label.

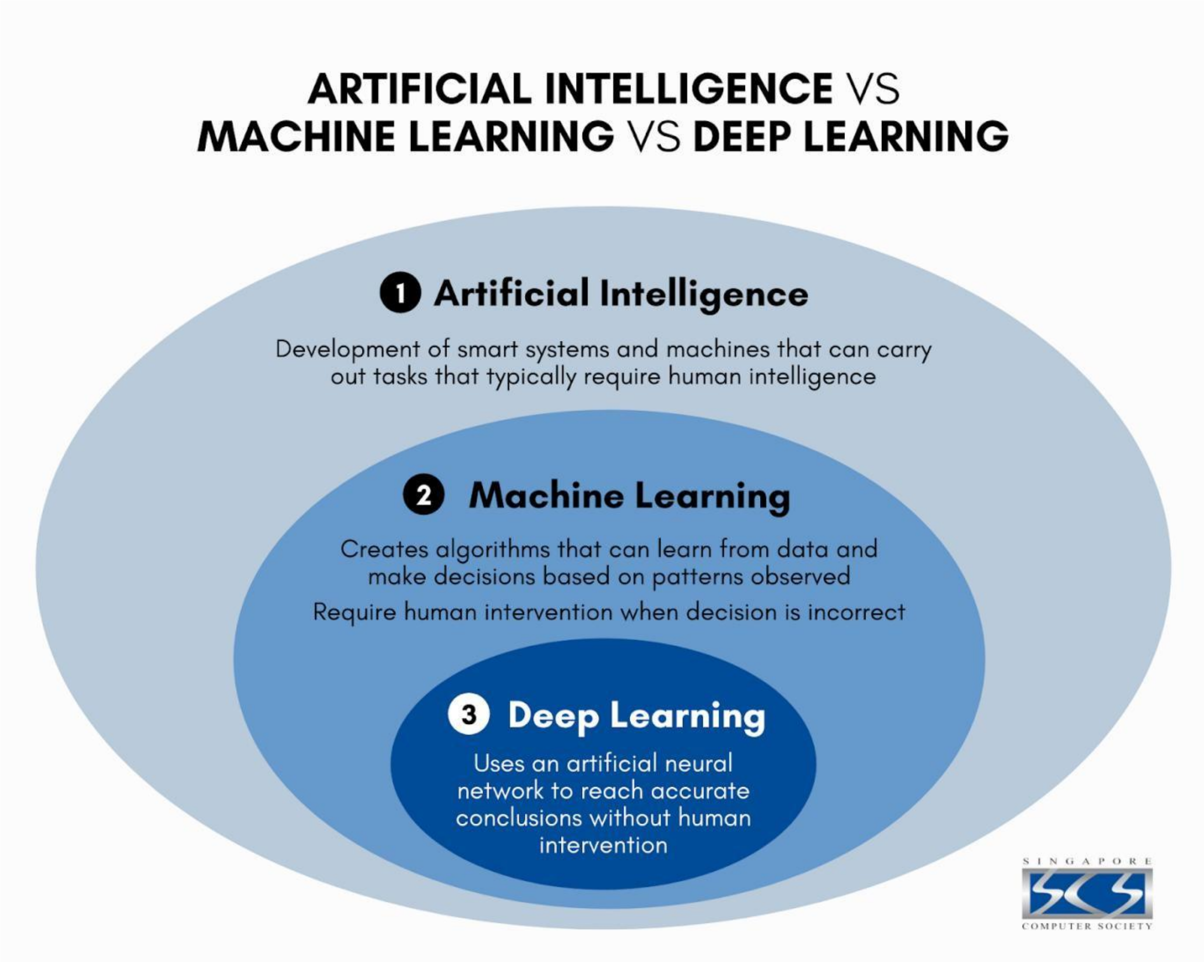


Figure 1.3 Difference Between AI, ML and DL

The conventional means involve human scouting – farmers and plant pathologists and use of pesticides for treatment thereafter. It is a challenging and time-consuming process, and most of the time leads to incorrect diagnosis with unbefitting application of the pesticides.

Computer Vision uses images and its pattern mappings for results while Machine Learning involves wellestablished approaches for predicting output namely, supervised, unsupervised and reinforcement learning, which include data acquisition, pre-processing, training the model and using either for predictions or classifications. Machine learning algorithms have gained popularity in computer vision to improve the accuracy and rapidity of the automated plant pathology diagnosis outcomes.

Machine Learning along with Computer Vision advances the computer capabilities for object detection, data sensing and understanding, classification and feature extraction. Computer Vision uses images and its pattern mappings for results while Machine Learning involves wellestablished approaches for predicting output namely, supervised, unsupervised and reinforcement learning, which include data acquisition, pre-processing, training the model and using either for predictions or classifications. The inputs are either a direct input of pixels or 3D points, or vectorsrepresenting features in terms of color and texture distributions, shape measures and/or edge measures.

We also proposed in this paper a novel paradigm for detecting plant leaf diseases based on a deep Convolutionary neural network (strong CNN). A free data collection of 39 separate plant leaf groups and context photos trains the Deep CNN model. Six types of data increase methods were used: picture fl-flipping, gamma correction, injection noise, PCA color increase, rotation and scaling. We have observed that data increase can increase the model's performance.

Apples are grown all across the world and are one of the most efficiently and widely cultivated fruits contributing majorly to global yield. However, various diseases cause substantial loss to apple production. Therefore, accurate and timely detection of apple leaf diseases is crucial. In this paper, the dataset is provided by Cornell Initiative for Digital Agriculture (CIDA) consisting of about 3600 training images categorised and classified into 4 classes, namely healthy, scab, rust and multiple diseases on the same leaf. While scab is a fungal plant disease characterised by crustaceous lesions on leaves and fruits, rust is another fungal ailment with tiny specks that range in many brown shades.

The goal of Machine Learning is to enable computers to learn without explicitly programming, depending on the technological challenge. This is because machine learning depends on computer programs that can be changed as new data are released. The diseases self-recognition in the is based on the symptoms of the disease. This helps farmers, analysts and scholars to collect reports on the incidence of the disease rapidly and reliably. This limits the tracking of human beings in the broader sector. The aim is to derive the characteristic ofthe area in disease identification from image. The various other diseases include rust, kole roga, yellow leaf disease, leaf rot, leaf curl, angular leaf spot, leaf spot, late blight, bacteria wilt, etc., which all affect the growth of the plant. Plant diseases depend on climatic conditions too. For instance, reddish purple leaf spots on the older leaves of the plant are due to the entomosporium leaf spot fungus. Furthermore, this disease is very dangerous because in cool and wet weather conditions.

1.5 Machine Learning Techniques

The challenge to supervised learning is labeled training data requiring time and expertise and that to unsupervised learning is evolving patterns due to clustering. These approaches use Support Vector Machines, K-Means Clustering, KNN, Naïve Bayes, Probabilistic Models and Neural Networks. These ML approaches heavily depend on data augmentation and pre-processing techniques for feature extraction. However, because features selection is human based, these approaches still have quite low recognition rate and can ignore better features.

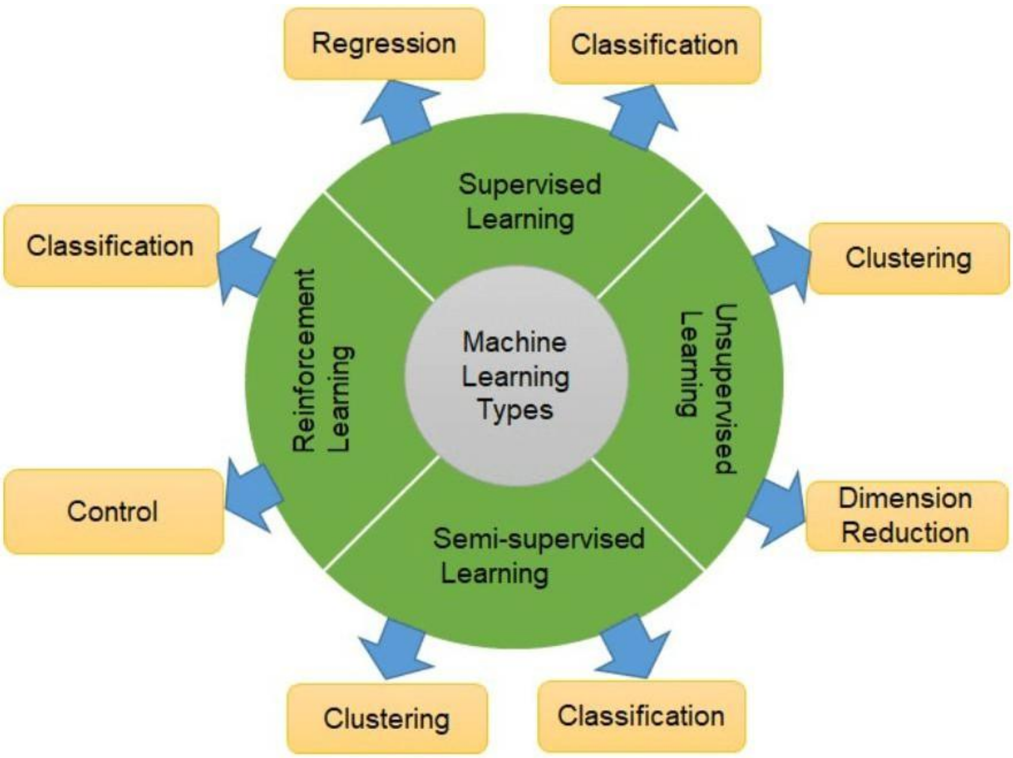


Figure 1.4 Types of Machine Learning

Machine Learning along with Computer Vision advances the computer capabilities for object detection, data sensing and understanding, classification and feature extraction. Computer Vision uses images and its pattern mappings for results while Machine Learning involves wellestablished approaches for predicting output namely, supervised, unsupervised and reinforcement learning, which include data acquisition, pre-processing, training the model and using either for predictions or classifications. The inputs are either a direct input of pixels or 3D points, or vectorsrepresenting features in terms of color and texture distributions, shape measures and/or edge measures.

1.6 Why Deep Learning

Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain—albeit far from matching its ability—allowing it to “learn” from large amounts of data. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy. Deep learning drives many artificial intelligence (AI) applications and services that improve automation, performing analytical and physical tasks without human intervention. Deep learning technology lies behind everyday products and services (such as digital assistants, voice-enabled TV remotes, and credit card fraud detection) as well as emerging technologies (such as self-driving cars).

Early plant disease detection plays a significant role in efficient crop yield. Plant diseases like black measles, black rot, bacterial spot, etc. affect the growth, crop quality of plants and economic impacts in the agriculture industry. To avoid the impact of these diseases, expensive approaches and the use of pesticides are some solutions the farmers usually implement. The use of chemical means damages the plant and the surrounding environment. In addition, this kind of approach intensifies the cost of production and major monetary loss to farmers. Early discovery of diseases as they occur is the most important period for efficient disease management. Manual disease detection through human experts to identify and recognize plant diseases is a usual practice in agriculture. With the improvements in technology, automatic detection of plant diseases from raw images is possible through computer vision and artificial intelligence researches. In this study, the researchers were able to investigate plant diseases and pest’s infestation that affects the leaves of the plants.

The agriculturist in provincial regions may think that it's hard to differentiate the malady which may be available in their harvests. It's not moderate for them to go to agribusiness office and discover what the infection may be. Our principle objective is to distinguish the illness introduce in a plant by watching its morphology by picture handling and machine learning.

Pests and Diseases results in the destruction of crops or part of the plant resulting in decreased food production leading to food insecurity. Also, knowledge about the pest management or control and diseases are less in various less developed countries. Toxic pathogens, poor disease control, drastic climate changes are one of the key factors which arises in dwindled food production.

1.7 Convolutional Neural Network

Over the last few years, Deep Learning, a special class of Machine Learning algorithms having multiple layers for transforming data, has been predominately used in agriculture. Convolutional Neural Networks is an unsupervised deep learning model with capability to extract features automatically from training dataset, thereby avoiding complex pre-processing and ensuring high recognition accuracy. CNN models require very few neurons but large data for their training. Data augmentation and determining the best structures of the network model are difficult tasks. A basic CNN model is a neural network using just convolution and pooling layer. Deep CNN is neural network with a lot of layers, called multi-layer perceptron. DNN models surpass CNN in simplicity, time and accuracy, whereas CNN requires a deeper architecture yet doesn't perform as well. However, CNN provides improved visual connection in evaluation process by taking images as input.

In this paper, more accurate DNN models namely EfficientNet and DenseNet are deployed, which will overcome the drawbacks of popular pre-trained CNN models - excess pooling layers, more parameters, more computation time, loss of feature maps and pattern information. EfficientNet reduces parameters and ensures accuracy through model scaling and DenseNet deploys feature reuse. The paper also evaluates the possible model of relatively new Capsule Networks for classification using simple reduced capsules on images. The proposed CapsNET model enhances the learning capacity of the DL models for apple plants, with few layers from CNNs.

The rest of the paper is organized as follows: We refer the previous works, present systems and their limitations. We briefly review the proposed system, novelty and its objectives. We describe the evaluation methodology and experimental procedures used. We present quantitative results and discuss our experiments. Plant diseases affect the growth and crop yield of the plants and make social, ecological and economical impacts on agriculture. Recent studies on leaf diseases show how they harm the plants. Plant leaf diseases also cause significant economic losses to farmers. Early detection of the diseases deserve special attention. Plant diseases are studied in the literature, mostly focusing on the biological aspects. They make predictions according to the visible surface of plants and leaves. Detection of diseases as soon as they appear is a vital step for

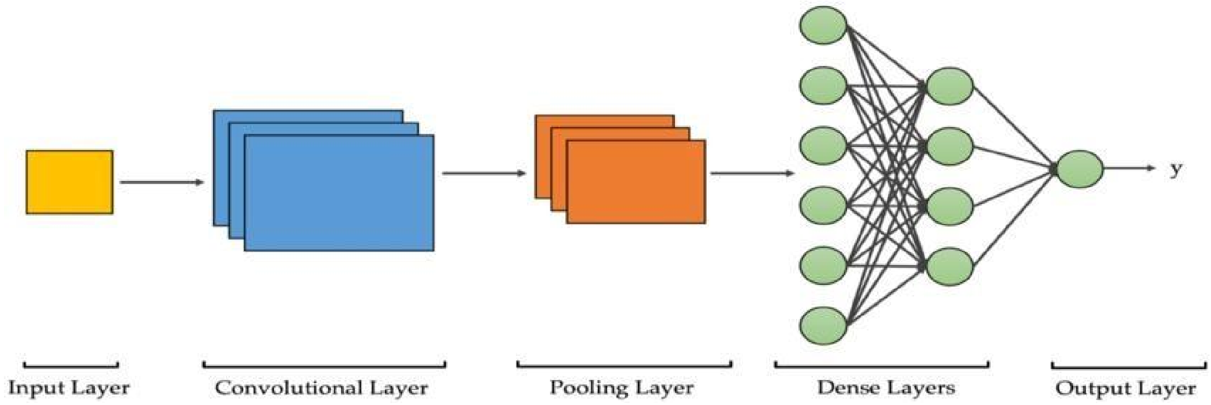


Figure 1.5 Convolutional Neural Networks Architectur

effective disease management. The detection is traditionally carried out by human experts. Human experts identify diseases visually but they faces some difficulties that may harm their efforts. In this context, detecting and classifying diseases in an exact and timely manner is of the great importance.

Advances in artifificial intelligence researches now make it possible to make automatic plant disease detection from raw images. Deep learning can be thought as a learning method on neural networks. One of the advantages of deep learning is that it can extract features from images automatically. The neural network learns how to extract features while training. CNN is a multi-layer feed-forward neural network and is the popular deep learning model.

In recent years, CNN models have been widely used in image classification problems. Introduce a hybrid model to extract contextual information of leaf features using CNN and Deconvolutional Networks (DN). Konstantinos at al. Performed several pre-trained CNN models on a large open leaves dataset. Their studies show that CNN is highly suitable for automatic plant disease identification. Durmus at al. Also used AlexNet and Squeeze pre-trained CNN models on tomato leaves from an open dataset to detect diseases. Atabay at al. Fifine-tuned a pre-trained model and designed a new CNN model to perform tomato leaf disease identifification. Their study indicates that custom CNN model gives better results than the pre-trained model. Setting a suitable CNN model is a challenging issue to produce higher accuracy values. Zhang at al proposed a three-channel CNN model based on RGB colors to detect vegetable leaf diseases.

Plant leaf images are couremplex with its background and the color information extracted from a single color component is limited. It causes the feature extraction method to give lesser accuracy results. Using different color components is promising instead of single one. In the proposed paper we developed a CNN model based on RGB components of the tomato leaf images on PlantVilliage dataset. We preferred Learning Vector Quantization (LVQ) algorithm as classifier due to its topology and adaptive model.

Early plant disease detection plays a significant role in efficient crop yield. Plant diseases like black measles, black rot, bacterial spot, etc. affect the growth, crop quality of plants and economic impacts in the agriculture industry. To avoid the impact of these diseases, expensive approaches and the use of pesticides are some solutions the farmers usually implement. The use of chemical means damages the plant and the surrounding environment. In addition, this kind of approach intensifies the cost of production and major monetary loss to farmers. Early discovery of diseases as they occur is the most important period for efficient disease management. Manual disease detection through human experts to identify and recognize plant diseases is a usual practice in agriculture. With the improvements in technology, automatic detection of plant diseases from raw images is possible through computer vision and artificial intelligence researches. In this study, the researchers were able to investigate plant diseases and pest's infestation that affects the leaves of the plants.

Over the last few years, Deep Learning, a special class of Machine Learning algorithms having multiple layers for transforming data, has been predominately used in agriculture. Convolutional Neural Networks is an unsupervised deep learning model with capability to extract features automatically from training dataset, thereby avoiding complex pre-processing and ensuring high recognition accuracy. CNN models require very few neurons but large data for their training. Data augmentation and determining the best structures of the network model are difficult tasks. A basic CNN model is a neural network using just convolution and pooling layer. Deep CNN is neural network with a lot of layers, called multi-layer perceptron. DNN models surpass CNN in simplicity, time and accuracy, whereas CNN requires a deeper architecture yet doesn't perform as well. However, CNN provides improved visual connection in evaluation process by taking images as input. Advances in artifificial intelligence researches now make it possible to make automatic plant disease detection from raw images. Deep learning can be thought as a learning method on neural networks. One of the advantages of deep learning is that it can extract features from images automatically. The neural network learns how to extract features while training. CNN is a multi- layer feed-forward neural network and is the popular deep learning model.

Image processing techniques are now commonly employed in agriculture and it is applied for the detection and recognition of weeds, fruit-grading, identifying and calculating disease infestations of plants, and plant genomics. Currently, the introduction of deep learning methods turns out to be popular.

Deep learning is the advanced methods of machine learning that uses neural networks that works like the human brain . Traditional methods involve the use of semantic features as the classification method. LeChun et al describes deep learning as a neural network learning process and one feature of deep learning is that it can automatically obtain features through image patterns .

A convolutional neural network (CNN) is a deep learning model that is widely used in image processing. The work of Lee et al presents a hybrid model to obtain characteristics of leaves using CNN and classify the extracted features of leaves. The study of Ferentinos, K.P. uses simple and infected plant leaf images to detect plant diseases using pre-trained CNN model. Durmus et al work on the detection of diseases of the tomato leaves using AlexNet and SqueezeNet pre-trained CNN architectures. While Atabay et al contributed a new CNN architecture to do disease classification and identification.

The methodology in the study involves three key stages: acquisition of data, pre-processing of data and image classification. The study utilized dataset from Plant village dataset that contains plant varieties of apple, corn, grapes, potato, sugarcane, and tomato. There are 11 types of plant diseases identified in the study including healthy images of identified plants. Image pre-processing involves re-sized images and enhancement before supplying it for the classification model.

The agriculturist in provincial regions may think that it’s hard to differentiate the malady which may be available in their harvests. It's not moderate for them to go to agribusiness office and discover what the infection may be. Our principle objective is to distinguish the illness introduce in a plant by watching its morphology by picture handling and machine learning.

Pests and Diseases results in the destruction of crops or part of the plant resulting in decreased food production leading to food insecurity. Also, knowledge about the pest management or control and diseases are less in various less developed countries. Toxic pathogens, poor disease control, drastic climate changes are one of the key factors which arises in dwindled food production.

Various modern technologies have emerged to minimize postharvest processing, to fortify agricultural sustainability and to maximize the productivity. Various Laboratory based approaches such as polymerase chain reaction, gas chromatography, mass spectrometry, thermography and hyper spectral techniques have been employed for disease identification. However, these techniques are not cost effective and are high time consuming.

A convolutional neural network (CNN) is a deep learning model that is widely used in image processing. The work of Lee et al presents a hybrid model to obtain characteristics of leaves using CNN and classify the extracted features of leaves. The study of Ferentinos, K.P. uses simple and infected plant leaf images to detect plant diseases using pre-trained CNN model. Durmus et al work on the detection of diseases of the tomato leaves using AlexNet and SqueezeNet pre-trained CNN architectures. While Atabay et al contributed a new CNN architecture to do disease classification and identification.

Random forests are as a whole, learning method for classification, regression and other tasks that operate by constructing a forest of the decision trees during the training time. Unlike decision trees, Random forets overcome the disadvantage of over fitting of their training data set and it handles both numeric and categorical data.

In recent times, server based and mobile based approach for disease identification has been employed for disease identification. Several factors of these technologies being high

resolution camera, high performance processing and extensive built in accessories are the added advantages resulting in automatic disease recognition

1.8 Image Classification Techniques

Modern approaches such as machine learning and deep learning algorithm has been employed to increase the recognition rate and the accuracy of the results. Various researches have taken place under the field of machine learning for plant disease detection and diagnosis, such traditional machine learning approach being random forest, artificial neural network, support vector machine(SVM), fuzzy logic, K-means method, Convolutional neural networks etc.…

Random forests are as a whole, learning method for classification, regression and other tasks that operate by constructing a forest of the decision trees during the training time. Unlike decision trees, Random forets overcome the disadvantage of over fitting of their training data set and it handles both numeric and categorical data.

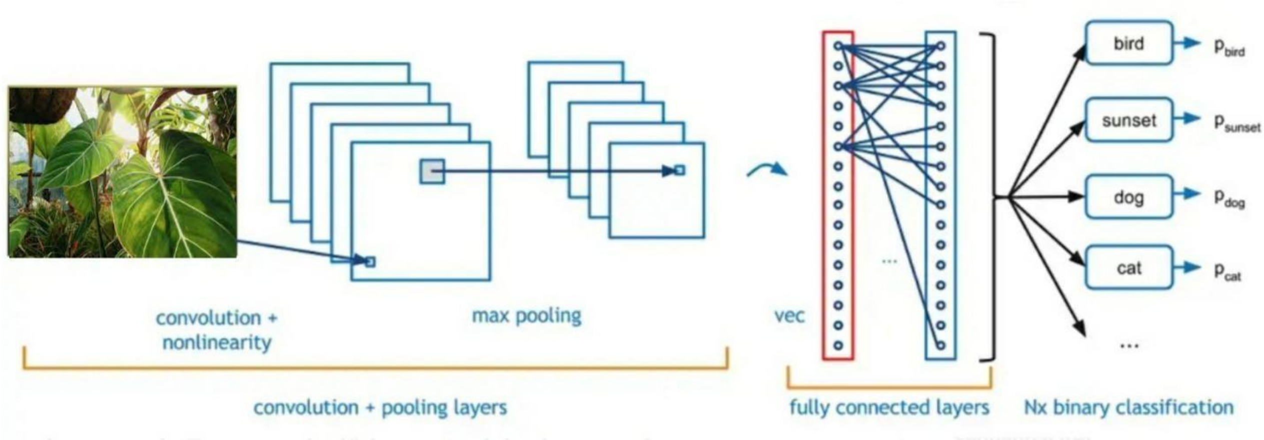


Figure 1.6 Image Classification

The histogram of oriented gradients (HOG) is an element descriptor utilized as a part of PC vision and image processing for the sake of object detection. Here we are making utilization of three component descriptors:

1. Hu moments

2. Haralick texture

3. Color Histogram

Hu moments is basically used to extract the shape of the leaves. Haralick texture is used to get the texture of the leaves and color Histogram is used to represent the distribution of the colors in an image.

The existing method for plant disease detection is simply naked eye observation by experts through which identification and detection of plant diseases is done. For doing so, a large team of experts as well as continuous monitoring of experts is required, which costs very high when farms are large. At the same time, in some countries, farmers don't have proper facilities or even idea that they can contact to experts. Due to which consulting experts even cost high as well as time consuming too. In such condition the suggested technique proves to be beneficial in monitoring large fields of crops. And automatic detection of the diseases by just seeing the symptoms on the plant leaves makes it easier as well as cheaper. This also supports machine vision to provide image based automatic process control, inspection, and robot guidance.

1.9 Motivation, Objectives and Scope of the work

The existing method for plant disease detection is simply naked eye observation by experts through which identification and detection of plant diseases is done. For doing so, a large team of experts as well as continuous monitoring of experts is required, which costs very high when farms are large. At the same time, in some countries, farmers don't have proper facilities or even idea that they can contact to experts. Due to which consulting experts even cost high as well as time consuming too. In such condition the suggested technique proves to be beneficial in monitoring large fields of crops. And automatic detection of the diseases by just seeing the symptoms on the plant leaves makes it easier as well as cheaper. This also supports machine vision to provide image based automatic process control, inspection, and robot guidance.

Image segmentation is the process of separating or grouping an image into different parts. There are currently many different ways of performing image segmentation, ranging from the simple thresholding method to advanced color image segmentation methods. These parts normally correspond to something that humans can easily separate and view as individual objects. Computers have no means of intelligently recognizing objects, and so many different methods have been developed in order to segment images. The segmentation process in based on various features found in the image. This might be color information, boundaries or segment of an image.

Genetic algorithms belong to the evolutionary algorithms which generate solutions for optimization problems. Algorithm begins with a set of solutions called population. Solutions from one population are chosen and then used to form a new population. This is done with the anticipation, that the new population will be enhanced than the old one. Solutions which are selected to form new solutions (offspring) are chosen according to their fitness - the more appropriate they are the more probability they have to reproduce.

Agriculture is major source of livelihood of people in our nation. Surveys reveal that about seventy percent of the Indian population is solely dependent on agriculture for their living, out of which eighty-two percent are poor or marginal farmers. Due to climatic alterations that is occurring across the globe and several other factors that affect agricultural produce, there has been a considerable rise in the number of plant or crop diseases. These diseases are responsible for restricting plant growth and also depreciating the quality as well as the quantity of agricultural produce. Hence, detection of plant diseases or nutrient deficiency in initial stages is of utmost importance. If the disease can be accurately detected, ways can be suggested subsequently to protect the crop from massive damage or crop failure. There are various diseases that may affect crops such as bacterial blight, alternaria alternate, cercospora leaf spot, anthracnose etc. Crop diseases are basically identified by observing different pattern on the parts of the crop like leaf, fruit, and stem. These diseases are mostly found on rice leaves, soyabean leaves, carrot leaves, mango leaves etc. and can be identified from the images of leaves. Often failure of crops also occurs due to nutritional deficiencies. This can also be predicted based on Chlorophyll and Nitrogen content of the leaves. The colour of the leaf is associated with the level of nitrogen and chlorophyll content present in the leaf. Chlorophyll and Nitrogen affects the green colour of the plant and determines their biomass yield. Plants that are sufficiently supplied with Nitrogen are green and very healthy, while those insufficiently supplied with Nitrogen are pale green or yellow in colour and remain small and stunted. Hence, leaf colour has led to exploit this property by using image processing analysis to detect Chlorophyll and Nitrogen content in plants. This can be done using modern technology that will benefit the farmers and save them from suffering major financial losses. Since paddy(rice) is India's most important food and various varieties of paddy is extensively cultivated across country, therefore in this paper we will be mainly concentration upon the detection and classification of diseases in paddy from the images of the unhealthy leaves of the crop by image processing using machine learning technique. Rice suffers from various diseases such as Rice Blast, Brown Spot of rice, Neck Blast, Sheath Blight of Rice etc. caused by their respective pathogens.

Manual disease detection is an extremely cumbersome task requiring a lot of processing time and manpower. Identification of these leaf diseases through naked eye is often prone to high error rates and faulty classification. In this scenario, modern farming techniques play a significant role and thus image processing can be used for the crop disease detection with greater accuracy. In our paper, we aim to capture the image of the diseased part of the plant or the diseased leaf and use this image for the detection. We will be applying image pre-processing to enhance the quality of the captured image and remove noises from the images. This will be followed by segmentation using K-means and subsequent classification done by SVM technique that will finally provide us the result of our search. SVM is adopted in our system because SVM's are very fine when one has no idea on the data and work well with semi-structured data like Images. Once the disease gets detected and properly identified, hence measures can be suggested to save the crop from further damage. It is pertinent to mention that along with the detection of disease, we will be predicting the chlorophyll content of the leaf as well as the nitrogen content to the plant and then subjecting it to SVM to categorize whether it falls under “efficiency” or under “deficiency”. This is help to reflect is the plant is suffering from any kind of nutritional deficiency and the farmers can add the required fertilizers in adequate amount or take other necessary measures. The detailed working of our proposed mechanism is discussed in section III of the paper. Thus, our system aims to provide a high-speed, accurate and inexpensive method in detecting and classifying leaf diseases.

The agriculturist in provincial regions may think that it’s hard to differentiate the malady which may be available in their harvests. It's not moderate for them to go to agribusiness office and discover what the infection may be. Our principle objective is to distinguish the illness introduce in a plant by watching its morphology by picture handling and machine learning.

Pests and Diseases results in the destruction of crops or part of the plant resulting in decreased food production leading to food insecurity. Also, knowledge about the pest management or control and diseases are less in various less developed countries. Toxic pathogens, poor disease control, drastic climate changes are one of the key factors which arises in dwindled food production.

Various modern technologies have emerged to minimize postharvest processing, to fortify agricultural sustainability and to maximize the productivity. Various Laboratory based approaches such as polymerase chain reaction, gas chromatography, mass spectrometry, thermography and hyper spectral techniques have been employed for disease identification. However, these techniques are not cost effective and are high time consuming.

1.10 Pathogens of Leaves

1.10.1. Tomatoes

Fungi is the predominant plant pathogens and it can cause multiple diseases including early blight, septoria leaf spot, target spot, and leaf mold. Fungi can attack plants through different sources such as infected soil and seeds. Fungal infections can spread from one plant to another through animals, humans, machinery, and soil contamination. The fungal attack vectors include plant pruning wounds, insects, leaf stomata, and others. The early blight disease of tomato plants is caused by the fungus, which affects the plant leaves. If it affects the seedlings’ basal stems, adult plant's stem, and fruits, it is called collar rot, stem lesion, and fruit rot, respectively. Numerous methods have been devised for the control of early blight but the most effective methods are cultural control i.e. efficient soil, nutrients, and crop management to reduce infections and also with the use of fungicidal chemicals. Septoria leaf spot of tomato plants is caused by fungus, which releases tomatinase enzyme that speeds up the degradation of tomato steroidal glycoalkaloids α-tomatine.

Fungi is the predominant plant pathogens and it can cause multiple diseases including early blight, septoria leaf spot, target spot, and leaf mold. Fungi can attack plants through different sources such as infected soil and seeds. Fungal infections can spread from one plant to another through animals, humans, machinery, and soil contamination. The fungal attack vectors include plant pruning wounds, insects, leaf stomata, and others. The early blight disease of tomato plants is caused by the fungus, which affects the plant leaves. If it affects the seedlings' basal stems, adult plant's stem, and fruits, it is called collar rot, stem lesion, and fruit rot, respectively. Numerous methods have been devised for the control of early blight but the most effective methods are cultural control i.e. efficient soil, nutrients, and crop management to reduce infections and also with the use of fungicidal chemicals. Septoria leaf spot of tomato plants is caused by fungus, which releases tomatinase enzyme that speeds up the degradation of tomato steroidal glycoalkaloids α-tomatine. The target spot disease of tomato plants is caused by the fungus. Symptoms of target spot disease in tomato plants are necrotic lesions of light brown color in the center. The lesions spread to a larger blighted leaf area and result in early defoliation. The target spot also damages the fruit directly by entering into the fruit pulp. The leaf mold disease of plants is caused by the fungus. It occurs during periods of extended leaf wetness. Bacteria is also a major plant pathogen. Bacteria enter plants through wounds such as insect bites, pruning, cuts, and also through natural openings such as stomata. Plant's surrounding environmental conditions such as temperature, humidity, soil conditions, availability of nutrients, weather conditions, and airflow are important factors in determining the bacterial growth on the plant and the consequent damage. Bacterial spot is a plant disease caused by bacteria. Molds are also a major cause of plant diseases. Late blight disease of tomato and potato plants is caused by mold. The appearance of dark uneven blemishes on leaves tips and plant stems are a few of the symptoms. Tomato yellow leaf curl virus (TYLCV) is a devastating virus causing tomato disease. This virus attacks the plant through another insect. Although tomato plants are unhealthy diseased leaves and iii) ten-class classification of healthy and various diseased leaves. In study II, different types of tomato leaf diseases are classified into the group of diseases while in study III, different classes of unhealthy and healthy leaf images were classified. Similar experiments the primary hosts for the virus, this viral infection has been reported in several other plants including beans and pepper, tobacco, potatoes, and eggplants. In the last few decades, due to the rapid spread of the disease, the research focus has been shifted to damage control of yellow leaf curl disease . Another viral disease that specifically affects tomato plants is caused by Tomato mosaic virus (ToMV). This virus is found worldwide and affects not only tomatoes but other plants as well. Symptoms of ToMV infection include twisting and fern-like appearance of leaves, damaged fruit with yellow patches, and necrotic blemishes.

1.10.2 Apple

The accurate identification of apple leaf diseases is of great significance for controlling the spread of diseases and ensuring the healthy and stable development of the apple industry. In order to improve detection accuracy and efficiency, a deep learning model, which is called the Coordination Attention EfficientNet (CA-ENet), is proposed to identify different apple diseases. First, a coordinate attention block is integrated into the EfficientNet-B4 network, which embedded the spatial location information of the feature by channel attention to ensure that the model can learn both the channel and spatial location information of important features. Then, a depth-wise separable convolution is applied to the convolution module to reduce the number of parameters, and the h-swish activation function is introduced to achieve the fast and easy to quantify the process. Afterward, 5,170 images are collected in the field environment at the apple planting base of the Northwest A&F University, while 3,000 images are acquired from the PlantVillage public data set. Also, image augmentation techniques are used to generate an Apple Leaf Disease Identification Data set (ALDID), which contains 81,700 images. The experimental results show that the accuracy of the CA-ENet is 98.92% on the ALDID, and the average F1-score reaches .988, which is better than those of common models such as the ResNet-152, DenseNet-264, and ResNeXt-101. The generated test dataset is used to test the anti-interference ability of the model. The results show that the proposed method can achieve competitive performance on the apple disease identification task.

1.10.3 Grapes

Black rot, Black measles, Leaf blight and Mites of grape are four common grape leaf diseases that seriously affect grape yield. However, the existing research lacks a real-time detecting method for grape leaf diseases, which cannot guarantee the healthy growth of grape plants. In this article, a real-time detector for grape leaf diseases based on improved deep convolutional neural networks is proposed. This article first expands the grape leaf disease images through digital image processing technology, constructing the grape leaf disease dataset (GLDD). Based on GLDD and the Faster R-CNN detection algorithm, a deep-learning-based Faster DR-IACNN model with higher feature extraction capability is presented for detecting grape leaf diseases by introducing the Inception-v1 module, Inception-ResNet-v2 module and SE-blocks. The experimental results show that the detection model Faster DR-IACNN achieves a precision of 81.1% mAP on GLDD, and the detection speed reaches 15.01 FPS. This research indicates that the real-time detector Faster DR-IACNN based on deep learning provides a feasible solution for the diagnosis of grape leaf diseases and provides guidance for the detection of other plant diseases.

1.11 Organization of Report

Chapter 1 gives the brief introduction of leaf disease detection using convolutional neural

network, its applications, objective of the system and motivation.

Chapter 2 contains literature survey that provide summary of individual paper.

Chapter 3 provide overview of existing work for leaf disease detection using CNN that has been done

using done using feature-based approach.

Chapter 4 presents Implementation and its results, tools and technology used to achieve this and

dataset detail.

Chapter 5 contains conclusion about leaf disease detection using CNN and future work about

Chapter - 2

2. Literature Review

2.1 Tomato Leaf Disease Detection Using Deep Learning Techniques

Surampalli Ashok, Gemini Kishore, Velpula Rajesh, S.Suchitra, S.G.Gino Sophia, B.Pavithra have stated that early Detection of Plant Leaf Detection is a main necessity in a developing agricultural economy like India. Now not simplest as an agricultural economic system however additionally with a massive quantity of populace to feed, it is necessary that leaf illnesses in vegetation are detected at a very early stage and predictive mechanisms to be followed to cause them to secure and avoid losses to the agri-primarily based economy. This paper proposes to identify the Tomato Plant Leaf disease using photograph processing techniques based on photograph segmentation, clustering, and open-supply algorithms, as a result all contributing to a dependable, safe, and correct device of leaf disease with the specialization to Tomato plant life.

2.2 Reviewing Important Aspects of Plant Leaf Disease Detection and Classification

Vishakha A.Metre, Sudhir D.Sawarkar have stated that agriculture, being the dominant enterprise from the point of view of comparatively cheap increase of nations like India, performs a crucial role in satisfying the demand of food. However, extreme climate conditions and numerous climate changes may additionally invite great infectious illnesses in vegetation resulting from fungi, viruses and micro organism. These plant illnesses may be a main risk to meals deliver and subsequently it's miles critical to become aware of and prevent the flora from the illnesses on the early tiers. The conventional approaches have been depending on the professionals within the field and therefore time ingesting. For the reason that era is upgrading each day and has plenty of its blessings in plant leaf ailment detection discipline as nicely, diverse sickness identity processes the usage of distinct domain names were proposed inside the literature to come across and remedy the plant diseases that occur at the plant leaves. Despite the fact that, the various existing tactics have furnished better effects, demanding situations exist to be able to attain optimized effects of plant leaf ailment detection method.

This paper evaluations exclusive methodologies underneath photograph processing, device learning, deep gaining knowledge of and swarm intelligence domain names for plant leaf disorder detection.Information of diverse illnesses that takes place on plant leaves may be very vital so that you can deal with it; subsequently this paper offers a detailed taxonomy about the unique plant illnesses and dataset that is popularly utilized in various existing strategies for schooling and testing motive of plant leaf ailment detection and its classifications.

2.3 Plant Pathology Disease Detection in Apple Leaves Using Deep Convolutional Neural Networks

V V Srinidhi, Apoorva Sahay, K.Deeba have stated that plant pathology is the science of plant sicknesses that tries to improve the probabilities for survival of flowers beneath unfavourable environmental conditions and parasitic microorganisms that purpose sickness. Temperature, pH, humidity and moisture are environmental elements contributing to development of plant illnesses. Misdiagnosis can lead to misuse of chemical substances inflicting financial loss, environmental imbalance and pollution and emergence of resistant pathogen lines. Contemporary sickness analysis is time consuming, pricey and based on human scouting.Automatic disease segmentation and analysis from plant leaf photographs may be moderately beneficial than the present one.Automated plant disorder detection entails photograph acquisition, pre-processing and segmentation, observed by way of augmentation, function extraction and type using models.

This challenge makes use of Deep Convolutional Neural Networks models specifically EfficientNet and DenseNet to come across Apple plant illnesses from snap shots of apple plant leaves and as it should be classify them into 4 training.The categories encompass “wholesome”, “scab”, “rust and “a couple of sicknesses”. On this task, the apple leaf disorder dataset is improved using information augmentation and photo annotation strategies, namely Canny edge Detection, Blurring and Flipping.

Based on augmented dataset, fashions the usage of EfficientNetB7 and DenseNet are proposed providing accuracy of 99.8% and 99.seventy five% respectively and overcoming known shortcomings of convolutional neural networks.

2.4 Plant Leaf Disease Detection and Classification based on CNN with LVQ Algorithm

Melike Sardogan, Adem Tuncer, Yunus Ozen have stated that the early detection of sicknesses is important in agriculture for an efficient crop yield. The bacterial spot, past due blight, septoria leaf spot and yellow curved leaf illnesses affect the crop satisfactory of tomatoes. Automated strategies for classification of plant diseases additionally help taking movement after detecting the signs of leaf diseases. This paper affords a Convolutional Neural network (CNN) version and Learning Vector Quantization(LVQ) set of rules based totally technique for tomato leaf ailment detection and type.

The dataset contains 500 pix of tomato leaves with four signs and symptoms of diseases. We've modeled a CNN for automatic feature extraction and category. Colour records is actively used for plant leaf ailment researches. In our model, the filters are carried out to a few channels based on RGB components.

The LVQ has been fed with the output characteristic vector of convolution component for education the community. The experimental consequences validate that the proposed approach successfully recognizes four one-of-a-kind kinds of tomato leaf illnesses.

2.5 Plant Leaf Detection and Disease Recognition using Deep Learning

Sammy V.Militante1, Bobby D.Gerardoij, Nanette V. Dionisioj have stated that the latest upgrades in laptop vision formulated thru deep learning have paved the approach for a way to detect and diagnose diseases in flora by the usage of a digicam to capture images as foundation for recognizing numerous kinds of plant sicknesses. This examine affords an green solution for detecting a couple of sicknesses in several plant varieties. The machine turned into designed to come across and apprehend several plant varieties especially apple, corn, grapes, potato, sugarcane, and tomato. The device can also come across numerous illnesses of flowers.

Made from 35,000 photos of healthy plant leaves and inflamed with the illnesses, the researchers have been able to educate deep studying fashions to stumble on and apprehend plant illnesses and the absence these of diseases.

The trained version has completed an accuracy fee of 96.5% and the machine turned into able to check in as much as 100% accuracy in detecting and spotting the plant range and the sort of illnesses the plant changed into infected.

2.6 Classification of plant leaf and diseases using machine learning and image preprocessing techniques

Pushkar sharma, Pankaj hans, subhas Chand gupta have stated that agriculture is one of main factor the that decides the growth of the country. Agriculture sector contributes around 16% of GDP of India. In India 15-20% of crops are lost due to diseases, pets,weeds.. Also, we can take reference of the incident of Georgia (USA) in 2007 in which there was loss of around 540 USD due to plant diseases. Hiring experts may cost them heavily and use of pesticides without knowledge will harm the land. Hence in order to solve this problem we have developed the Artificial Intelligence based solution.Accuracy and speed are the two main factors that will decide the success of the automatic plant leaf disease detection and classification model.n, KNN, SVM and CNN are trained and compared on the basis of accuracy and the algorithm that performs best in training as well as testing is taken in account.After using image preprocessing techniques along with k means clustering and comparing various classifiers available.The Logistic regression performs quite well considering number of classes.e. This approach of ours goals towards increasing the productivity of crops in agriculture. In this approach we have follow several steps i.e. image collection, image preprocessing, segmentation and classification.The presented model used the dataset that consists of more than 20,000 images with 19 total classes. The following model can be extended by using even more large dataset with more categories of diseases and the accuracy can also be improved by tuning the hyper parameters.

2.7 Detection of plant diseases and nutritional deficiencies from unhealthy plant leaves using Machine Learning techniques

Debolina Nath, Dr. Pushan Kumar Dutta, Anup Kumar Bhattacharya have stated that the in countries like India, which are majorly dependent on an agricultural economy, detection of crop-disease and classification has immense significance.Manual disease detection requires enormous amount of labour and involves too much processing-time. Therefore, image-processing plays important role for the detection of crop/plant disease from images of diseased leaves.

Due to climatic alterations that is occurring across the globe and several other factors that affect agricultural produce, there has been a considerable rise in the number of plant or crop diseases.Crop diseases are basically identified by observing different pattern on the parts of the crop like leaf, fruit, and stem. Hence, leaf colour has led to exploit this property by using image processing analysis to detect Chlorophyll and Nitrogen content in plants. This can be done using modern technology that will benefit the farmers and save them from suffering major financial losses.By this the farmers can add the required fertilizers in adequate amount or take other necessary measures. The first phase is applying pre-processing to the photographed leaf picture, which includes median filters and contrast stretching to improve contrast and reduce digitization noise, respectively. In the subsequent stage, the leaf image's L*a*b* colour model, which divides pixels into uniform sets, is used to apply the K-means clustering algorithm. The segment with the highest hue value among the other segments is extracted as the illness segment after these segments have been transformed to the HSV colour model. In order to classify the cropped image into different categories of diseases, we will then apply machine learning to the extracted sick segment using a Multi SVM classifier. The Chlorophyll-Nitrogen content of the leaf is predicted using the same approach and the SVM.

2.8 Detection of unhealthy region of plant leaves using Image Processing and Genetic Algorithm

Vijai Singh,Varsha,Prof. A K Misra have stated that the Indian economy heavily depends on agricultural productivity. One of the reasons that disease detection in plants is crucial in the world of agriculture is because diseases in plants are a very common occurrence. Inadequate care in this region has major negative effects on plants, which have an impact on the quality, volume, or productivity of the corresponding products.Due to which consulting experts even cost high as well as time consuming too. In such condition the suggested technique proves to be beneficial in monitoring large fields of crops. And automatic detection of the diseases by just seeing the symptoms on the plant leaves makes it easier as well as cheaper. Plant disease identification by visual way is more laborious task and at the same time less accurate. Whereas if automatic detection technique is used it will take less efforts, less time and more accurately. With very less computational efforts the optimum results were obtained, which also shows the efficiency of proposed algorithm in recognition and classification of the leaf diseases. Another advantage of using this method is that the plant diseases can be identified at early stage or the initial stage. This paper presents an algorithm for image segmentation technique used for automatic detection as well as classification of plant leaf diseases and survey on different diseases classification techniques that can be used for plant leaf disease detection. Image segmentation, which is an important aspect for disease detection in plant leaf disease, is done by using genetic algorithm.

2.9 Application research of plant leaf pests and diseases base on unsupervised learning

Mingjing Pei, Min KongMaosheng Fu, Xiancun Zhou,Jieru Xu have stated that the in agricultural productivity, detecting plant pests and diseases is extremely crucial.Plant disease target detection is the application of computer vision technology to detect plant disease infested areas and their specific positions in complicated natural settings, which is required for accurate categorization and identification of plant diseases, as well as assessment of disease damage severity. Existing methods based on the self-supervised approach to identify normal samples from abnormal samples by synthesizing aberrant samples and training a binary classifier.exist, however, the synthesized samples are still a long way from being accurate the detection of abnormal regions of plant leaves is by unsupervised method, which saves a lot of labor and does not require labeled data. There are more than 100 samples in normal and ten faulty samples in the other, and the relevant samples are depicted. This paper utilizes the idea of image restoration and uses a deep learning correlation model to detect and localize the abnormal regions of plant leaves. The experimental results show that the img_AUCROC and pixel_AUCROC level anomaly detection and localization achieve good results, which bring influence and reference to other peers.the location of plant leaf damage is located, delivering relevant and crucial information to workers and plant experts. In order to detect plant leaf disorders, this study evaluates unique techniques in the areas of image processing, machine learning, deep learning, and swarm intelligence.

2.10 Rose Plant Leaves: Disease Detection and Pesticide Management using CNN

Mohammad Ibrahim Khaleel, Paineni Gowtham Sai, Adabala Uday Raghavendra Kumar, Purushothamman Raja, Vinh Truong Hoang have stated that the agriculture is one of the main livelihoods and major source of income for farmers. Rose cultivation is one of the highest contributing income sources to India. But the main problem in the cultivation of rose plant is the diseases on the plant leaves which will affect its growth. The most commonly occurring diseases on rose plant leaves are Blackspot, Rose mosaic, Rose aphids, and Powdery mildew. Manual methods for disease detection take more time and expensive.deep learning uses matrix multiplication for building neural networks. Instead of using this technique, CNN uses convolution. It is combining two functions mathematically to produce a third function in order. CNN is made up of artificial neurons that are comprised of multiple layers. Artificial neurons are mathematical functions that emulate their biological equivalent. CNN uses tensor as input.The most important advantage of CNN compared to its previous versions of neural network technique is that it automatically detects the significant features without any human intervention.

Training accuracy indicates the ability of the model to classify images during training of the model. The pre-processing such as image resizing and image augmentation of the data set were performed before training with CNN architecture using the optimized parameters. The trained architecture was tested using the test data set and the results were analyzed. Based on the results, an appropriate pesticides was prescribed.

2.11 Rose Plant Leaves Disease Detection and Pesticide Management using CNN

Mohmmad Ibrahim Khaleel, Papineni Gowtham sai, Adabala Uday Raghavendra Kumar, Purushothmman Raja, Vinh Truong Hoang have stated that theagriculture is the major source of income for farmers in India. Flower cultivation is one of the major income sources for farmers. Rose cultivation is one of the major income sources for farmers due its heavy demand for its usage in cosmetics and decoration purposes in temples etc. One of the major drawback in cultivation of rose plant is its Diseases which effects its growth and production of flowers which cause loss in cultivation. The major diseases in rose plant leaves are Rose mosaic, Rose aphids, Powdery mildew. The detection of those diseases take more time where it causes major damage to the plant health. So we need an technology to detect the diseases before damage happens. Here is a need for an intelligent system based on advanced Deep Learning (DL) architectures which are the sub-field of artificial intelligence. Convolution Neural Network (CNN) is one of the deep neural networks used for many image based object detection tasks. The proposed work employs the CNN model for the detection of the above- said four types of diseases on rose plant leaves. The images were acquired from the real field and some are from an existing data set. The pre-processing such as image resizing and image augmentation of the data set were performed before training with CNN architecture using the optimized parameters. The trained architecture was tested using the test data set and the results were analyzed. Based on the results, an appropriate pesticides was prescribed.

2.12 Plant Disease Detection and Classification Using Deep Learning ConvNets

A Lakshmanrao, M Raja Babu, T Srinivasa Ravi Kiran have stated that agriculture is the major occupation in India, Agriculture is hugely important to humans as a food source. As a result,Plant diseases detection has become a major concern. Traditional methods for identifying Plant diseases are available. However agriculture professionals or plant pathologists have traditionally employed empty eye inspections to detect leaf disease this approach of detecting leaf disease traditionally can be subjective, time consuming , as well as expensive, and requires a lot of people and a lot of information about plant diseases. It is also possible to detect plant leaf diseases using an experimentally evaluated software solution.Currently ,machine learning and deep learning are using in recent years the agriculture sector is also not a exception for machine learning, in this paper, we applied"convents" for plant disease detection and classification.We collected a plant village dataset from Kaggle. It contains images of 15 different cases of plant leaves of three different plant potato, pepper, tomato. We divided the dataset into three datasets and applied convents on three datasets we achieved an accuracy of 98.3% for potato plant disease detection ,pepper plant disease detection, tomato plant disease detection . Experimental results have Shown that our model achieved a good accuracy rate for plant leaf disease detection and classification.

2.13 Machine Learning for plant leaf Disease Detection and Classification

L.Sherly Pushpa Annabel, T Annapoorani, P. Deepalakshmi have stated that plants are considered to be important as they are the source of energy supply to mankind. Plant diseases can affect the leaf any time between sowing and harvesting which leads to huge loss on the production of crop and economical value of market. Therefore, leaf disease detection plays a vital role in agricultural field. However, it requires huge manpower, more processing time and extensive knowledge about plant diseases. Hence, machine learning is applied to detect diseases in plant leaves as it analyzes the data from different aspects, and classifies it into one of the predefined set of classes. The morphological features and properties like color, intensity and dimensions of the plant leaves are taken into consideration for classification. This paper presents an overview on various types of plant diseases and different classification techniques in machine learning that are used for identifying diseases in different plant leaves.

2.14 Plant leaf disease classification and detection system using machine learning

Geetha Gunasekara, S.Samundeswari, Saranya Gangadhara Moorthy, Meenakshi have stated that the in a developing country like India agriculture plays a noteworthy role. Agricultural intervention in the livelihood of rural India indulges by about 58%. Among the agricultural products, tomato is one of the most used crops. Thus, preventing significant loss in quantity and yield of tomato is majorly dependent on recognition and classification of diseases a tomato plant might possess. Latest and fostering technologies like Image processing is used to rectify such issues using different types of techniques and algorithms. Initially, the leaves of a tomato plant get affected, when plant develops a particular type of disease. In this project, four consecutive stages are used to discover the type of disease. The four stages include pre-processing, leaf segmentation, feature extraction and classification. To remove the noise we are doing the pre-processing and to part the affected or damages area of the leaf, image segmentation is used. The k-nearest neighbors (KNN) algorithm, which is a guided, supervised and advance machine learning algorithm, is implemented to find solutions for both the problems related to classification and regression. During the terminal stage, user is recommended with the treatment. Mostly live plants are adversely affected by the diseases. This paper imparts representation of leaf disease detection employing image processing that can identify drawbacks in tomato plant from images, based on color, bound and texture to give the brisk and reliable results to the farmer.

2.15 Identification of Plant Leaf Diseases using Machine Learning Algorithms

M Geetha Yadav, Rajasekhar Nennuri, Devarakondaa Rajeshwari, Voggu Rishitha Tandur Punnut have stated that plant diseases are therefore becoming a major problem for humans. Plant diseases can happen at any time, It may be happen between harvesting and sowing. It is a great loss of economic value in the market. Therefore, the detection of leaf diseases plays an important role in agriculture. Therefore, disease detection has been carried out using traditional methods. However, the traditional method of detecting leaf diseases is empty-eye observation by agricultural experts or plant pathologists. Detecting diseases of the leaves of plants using this method, which can be subjective, time consuming, and profitable, requires a large amount of manpower and an extensive knowledge of plant diseases. With the help of an experimental evaluation software solution, plant leaf diseases can be automatically recognized and classified. Machine learning is being used for a new development. Machine learning is used to detect diseases in plants. Machine learning is one of the subsections of artificial intelligence to work automatically or to give instructions for a specific task. The main goal of machine learning is to grasp the training data and incorporate it into models that will be useful to humans. As a result, we can use machine learning to detect plant diseases. It helped make good decisions and predict the large amount of data. The color of the leaves, the extent of damage to the leaves, the leaf area of the unhealthy plant is used in the classification. In this regard, various machine learning algorithms are reviewed to identify various diseases of the leaves of plants and identify the best precision.

Chapter - 3

3. Plant Leaf Disease Detection

3.1 About Model and Learning

3.1.1 OpenCV

OpenCV is a Python open-source library, which is used for computer vision in Artificial intelligence, Machine Learning, face recognition, etc. In OpenCV, the CV is an abbreviation form of a computer vision, which is defined as a field of study that helps computers to understand the content of the digital images such as photographs and videos.

The purpose of computer vision is to understand the content of the images. It extracts the description from the pictures, which may be an object, a text description, and three-dimension model, and so on. For example, cars can be facilitated with computer vision, which will be able to identify and different objects around the road, such as traffic lights, pedestrians, traffic signs, and so on, and acts accordingly.

Computer vision allows the computer to perform the same kind of tasks as humans with the same efficiency. There are a two main task which are defined below:

Object Classification - In the object classification, we train a model on a dataset of particular objects, and the model classifies new objects as belonging to one or more of your training categories.

Object Identification - In the object identification, our model will identify a particular instance of an object - for example, parsing two faces in an image and tagging one as Virat Kohli and other one as Rohit Sharma.

Why open cv is used for computer Vision?

OpenCV is available for free of cost. Since the OpenCV library is written in C/C++, so it is quite fast. Now it can be used with Python. It requires less RAM to usage, it maybe of 60-70 MB.
Computer Vision is portable as OpenCV and can run on any device that can run on C.

3.1.2 What is Computer Vision?

The term Computer Vision (CV) is used and heard very often in artificial intelligence (AI) and deep learning (DL) applications. The term essentially means giving a computer the ability to see the world as we humans do.

Computer Vision is a field of study which enables computers to replicate the human visual system. As already mentioned above, It's a subset of artificial intelligence which collects information from digital images or videos and processes them to define the attributes. The entire process involves image acquiring, screening, analysing, identifying and extracting information. This extensive processing helps computers to understand any visual content and act on it accordingly.

Computer vision projects translate digital visual content into explicit descriptions to gather multi-dimensional data. This data is then turned into a computer-readable language to aid the decision-making process. The main objective of this branch of artificial intelligence is to teach machines to collect information from pixels.

How does a computer read an image?

How does a human mind apprehend an image? When you see the image below, what do you actually see and how do you say what is in the Image?

You most probably look for different shapes and colours in the Image and that might help you decide that this is an image of a dog. But does a computer also see it in the same way? The answer is no.

A digital image is an image composed of picture elements, also known as pixels, each with finite, discrete quantities of numeric representation for its intensity or grey level. So the computer sees an image as numerical values of these pixels and in order to recognise a certain image, it has to recognise the patterns and regularities in this numerical data.

Here is a hypothetical example of how pixels form an image. The darker pixels are represented by a number closer to the zero and lighter pixels are represented by numbers approaching one. All other colours are represented by the numbers between 0 and 1.

But usually, you will find that for any colour image, there are 3 primary channels – Red, green and blue and the value of each channel varies from 0-255. In more simpler terms we can say that a digital image is actually formed by the combination of three basic colour channels Red, green, and blue whereas for a grayscale image we have only one channel whose values also vary from 0-255.

3.1.3 Uses of OpenCV

OpenCV uses image processing technology in order to process images in nanoseconds. The five main ways in which OpenCV is used are :

Object Detection

Using Object Detection, we can detect any objects such as detecting faces, detecting human bodies, and also detecting animals. It is done using a pattern of geometry technique.

Image Reconstruction

Using Image Reconstruction, we can get the reconstruction of the image properties. This can be done by using the samples provided by passing them into multiple frames.

Object Recognition

Object Recognition is similar to object detection. In object recognition, we need not detect the objects, but we need to match the results to those stored in the database previously.

Object Tracking

Object Tracking is based on video analysis. It means that we can get the recordings or recorded videos from the webcam. And then, by performing image processing techniques on the photo itself, we can learn in detail about the photo and objects in the image.

Video Processing

Video Processing can be comparable to object tracking. It is because of this process that we also need to analyze by seeing the video frames in order to detect the objects in the image. This is followed by tracking them in the video frames. The speed of the object's movement can also be tracked using this.

3.2 Overview of Existing Work

The model that is proposed by us to detect and classify the infected plant leaves consists of 4 phases.

Those phases are: -

- Dataset Collection
- Image Preprocessing
- Segmentation
- Selection of Classifier

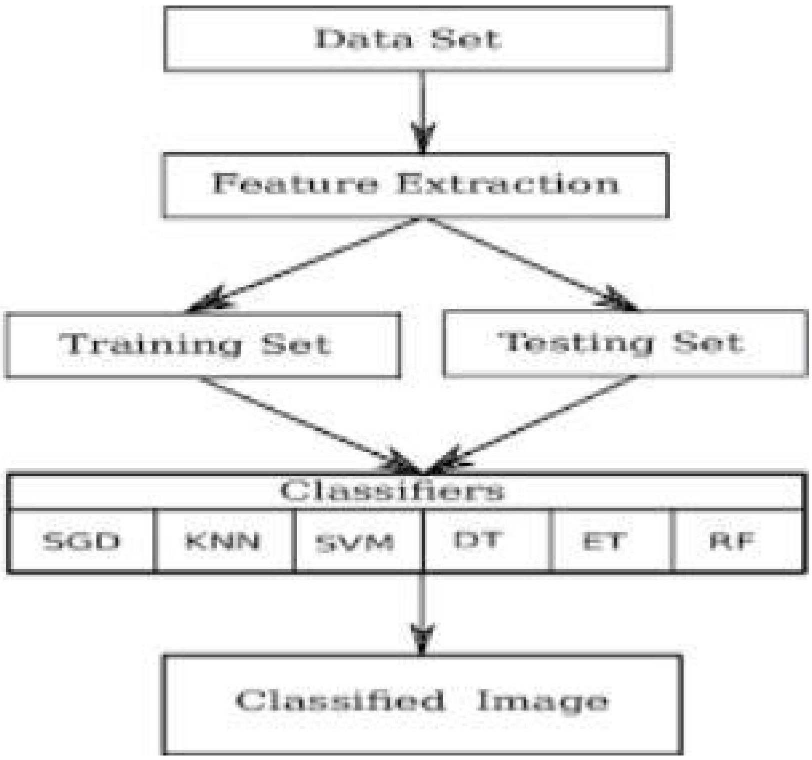


Figure 3.1 Block Diagram of Image Classification

A. Dataset Collection

Firstly, the images of leaves were collected from online sources such as GitHub, Kaggle and also some of the image's dataset consists of 20,000 images divided into 19 different classes.The dataset consists of both healthy and infected leaves which covers diseases like black rot, rust, bacterial spot, early blight, late blight, leaf scorch, target spot, mosaic virus of different crops like apple, potato, tomato, grape, strawberry, corn.

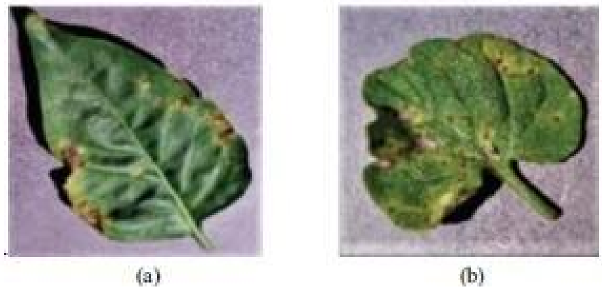


Figure 3.2 Sample Images from Datasets

Dataset contains a lot of separate pieces of data but can be used to train an algorithm with the goal of finding predictable patterns inside the whole dataset.

Data is an essential component of any AI model and, basically, the sole reason for the spike in popularity of machine learning that we witness today. Due to the availability of data, scalable ML algorithms became viable as actual products that can bring value to a business, rather than being a by-product of its main processes. Your business has always been based on data. Factors such as what the customer bought, the popularity of the products, seasonality of the customer flow have always been important in business making. However, with the advent of machine learning, now it's important to collect this data into datasets. Sufficient volumes of data allow you to analyze the trends and hidden patterns and make decisions based on the dataset you've built. However, while it may look rather simple, working with data is more complicated since it requires, first of all, proper treatment of the data you have, from the purposes of using a dataset to the preparation of the raw data for it to be actually usable.

Splitting Your Data: Training, Testing, and Validation Datasets in Machine Learning

Usually, a dataset is used not only for training purposes. A single training dataset that has already been processed is usually split into several parts, which is needed to check how well the training of the model went. For this purpose, a testing dataset is usually separated from the data. Next, a validation dataset, while not strictly crucial, is quite helpful to avoid training your algorithm on the same type of data and making biased predictions.

Features of the Data: Raw data is a good place to start but you obviously cannot just shove it into a machine learning algorithm and hope it offers you valuable insights into your customers' behaviors. There are quite a few steps you need to take before your dataset becomes usable.

1. **Collect:** The first thing to do when you're looking for a dataset is deciding on the sources you'll be using to collect the data. Usually, there are three types of sources you can choose from: the freely available open-source datasets, the Internet, and the generators of artificial data. Each of these sources has its pros and cons and should be used for specific cases.

2. **Preprocess:** There's a principle in data science that every experienced professional adheres to. Start by answering this question: has the dataset you're using been used before? If not, assume this dataset is flawed. If yes, there's still a high probability you'll need to re-appropriate the set to fit your specific goals.

3. **Annotat:** After you've ensured your data is clean and relevant, you also need to make sure it's understandable for a computer to process. Machines do not understand the data the same way as humans do (they aren't able to assign the same meaning to the images or words as we). This step is where a lot of businesses often decide to outsource since keeping a trained annotation professional is not always viable.

B. Image Preprocessing

In this step images are resized to smaller pixel size in order to speed up the computations. The acquired images contain some noise. This noise is removed using some filtering techniques like Gaussian Blur. After that images are present in RGB format which is not appropriate for further work as RGB format is unable to separate image intensity. Hence it is converted to another color space that is HSV which separate color from intensity. Also, RGB color space is noisier than HSV**.**

Figure 3.3 Image after Preprocessing

RGB to HSV conversion: -

First R, G, B values are divided by max value that is 255 So, R’ = R/255
G’ = G/255 B’ = B’/255
Then Cmax = max (R’, G’, B’)

Cmin = min (R’, G’, B’) Delta = Cmax – Cmin

Hue: -

$$H = \begin{cases} 60^\circ * \left(\frac{G'-B'}{\Delta} mod 6\right), Cmax = R' \\ 60^\circ * \left(\frac{B'-R'}{\Delta} + 2\right), Cmax = G' \\ 60^\circ * \left(\frac{R'-G'}{\Delta} + 4\right), Cmax = B' \end{cases}$$

Saturation: -

$$S = \begin{cases} 0, & Cmax = 0 \\ \frac{\Delta}{Cmax}, & Cmax \neq 0 \end{cases}$$

Value: -

$$V = C_{max}$$

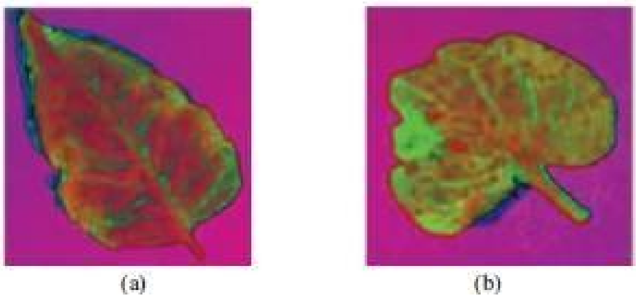


Figure 3.4 Image Converted to HSV Color Space

C. Segmentation

In this step, segmentation of images is done in order to separate the leaves from the background. Segmentation is performed using K-means clustering with 2 cluster centers, one for background and one for foreground. K-means clustering is unsupervised learning technique that is used to segregate the datapoints in the predefined number (k) of clusters or groups on the basis of their similarities.

K –Means algorithm works as follows: -

Set of inputs: - number of clusters(k), set of datapoints

1. Put k centroids in random location in space.
2. Repeat the following steps until none of cluster location changes: -

a. For every datapoint x(i) -

i. Find nearest centroid c(i) by argmax

ii. Assign x(i) to the cluster with nearest centroid

b. For every cluster, new centroid is assigned by taking mean of all datapoints assigned to that cluster.

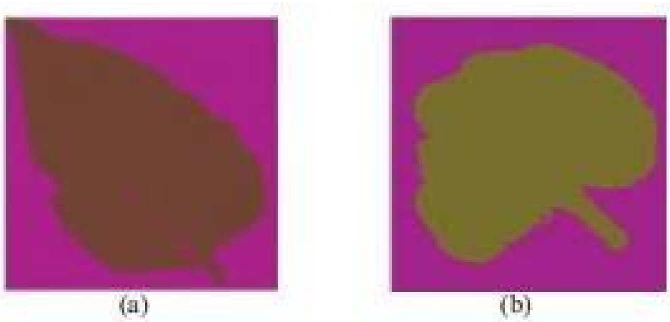


Figure 3.5 Images After K Means Clustering

After finding the two clusters, one with background and other one with leaf part, the clustered image is used to change the pixel value of the background of the leaf to black. By doing so the useless information from the image is eliminated which in turn increases accuracy.

Figure 3.6 Images After Removal of Background

D. Feature Extraction

After clusters are formed texture features are extracted using GLCM [13]. (Gray-Level Co-occurrence Matrix).

E. Classification

In classification is used for testing the leaf disease. The Random forest classifier is used for classification.

3.3 Selection of Classifier

This is the classification problem as we have to classify the type of disease on the leaf of the plant. So, we have plenty of machine learning as well as deep learning algorithms that we can apply on this dataset. We have decided to start with low complex algorithms and increasing the complexity level in order to increase accuracy of the model. We have selected four classifiers namely – logistic regression, KNN, SVM and CNN.

A. Logistic Regression

It is the simplest classification algorithm available but yet powerful enough to make some good results. The logistic regression makes the use of logistic function that is sigmoid function to squeeze the output in range of 0 and 1. After training on training set, the model gives the accuracy of 66.4% on testing set which is not that bad considering complexity of algorithm and number of classes in dataset.

B. KNN (K Nearest Neighbors)

It is the algorithm can be used in both classification as well as regression problems. It is very simple and easy algorithm to implement. Here we plot all the datapoints in space and then find the k nearest neighbors of the datapoint that we want to classify by finding the distance between all other datapoints and the input datapoint. Then k datapoints are chosen which are nearest to that datapoint and their classes are taken then predicted class of input is the class with maximum occurrence. On our dataset the knn model was able to give accuracy of 54.5%.

C. SVM (Support Vector Machine)

SVM is another machine learning algorithm that we have used to classify the diseases. In this algorithm all points are mapped in space so that points of different class can be divided by gap. Gap should be as wide as possible so that boundary can separate them. This boundary is called decision boundary and the extreme data points of classes called support vectors. Kernel tricks are used for nonlinear dataset. The kernels that are available are linear, nonlinear, polynomial and RBF.

The svm in our case worked poorly as it gives accuracy of only around 53.4% after using linear kernel.

D. CNN (Convolutional Neural Network)

This is the far most complex deep learning model that we have used to classify the diseases. As it is very complex hence it requires good computational power as well. It is the most common neural network that is applied to image classification problems. CNN is a neural network which comprises of four layers namely-Convolutional layer, Pooling layer, Activation function layer and Fully connected layer.

a) Convolutional Layer

It is most important layer in CNN model and also responsible for the naming of this network. In this layer, some mathematical operations are performed to get the features of the image. It consists of filters which have width and height less than input image and depth same as input image. If image of size 64*64*3 is feeding in the CNN and we have total of 10 filters then output of this layer will have dimension of 64*64*10. There are total of 5 convolutional layers where number of filters are 32,64,64,128,128 respectively. The kernel size is 3*3 for all layers.

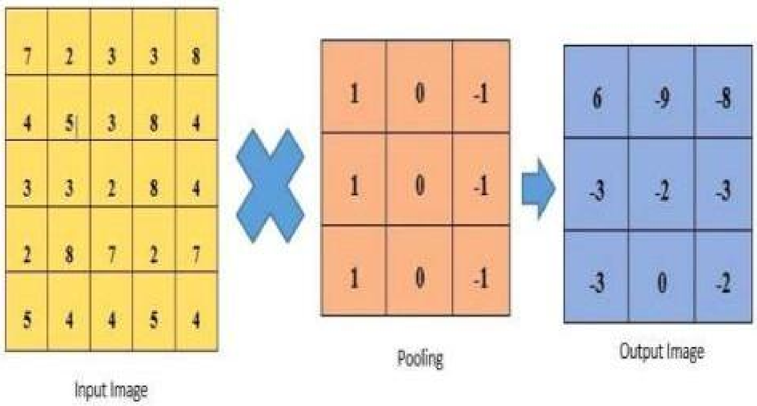


Figure 3.7 Operation of Convolution Layer without Padding

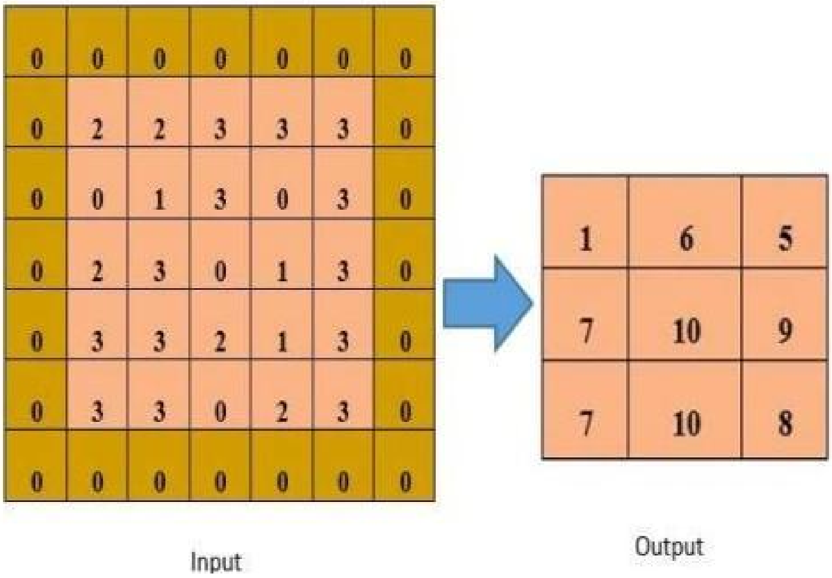


Figure 3.8 Operation of Convolution Layerwith Padding

b) Pooling Layer

This is the layer which is majorly responsible for size reduction of output of previous layer. Filters of different sizes can be used in this layer but generally 2*2 size is preferred. There are two major kind of pooling layers that are used namely- max pooling and average pooling. As name suggest max pooling take the maximum value out of filter and average pooling takes the average.

It is the algorithm can be used in both classification as well as regression problems. It is very simple and easy algorithm to implement. Here we plot all the datapoints in space and then find the k nearest neighbors of the datapoint that we want to classify by finding the distance between all other datapoints and the input datapoint. Then k datapoints are chosen which are nearest to that datapoint and their classes are taken then predicted class of input is the class with maximum occurrence. On our dataset the knn model was able to give accuracy of 54.5%.

There are two major kind of pooling layers that are used namely- max pooling and average pooling. As name suggest max pooling take the maximum value out of filter and average pooling takesthe average.

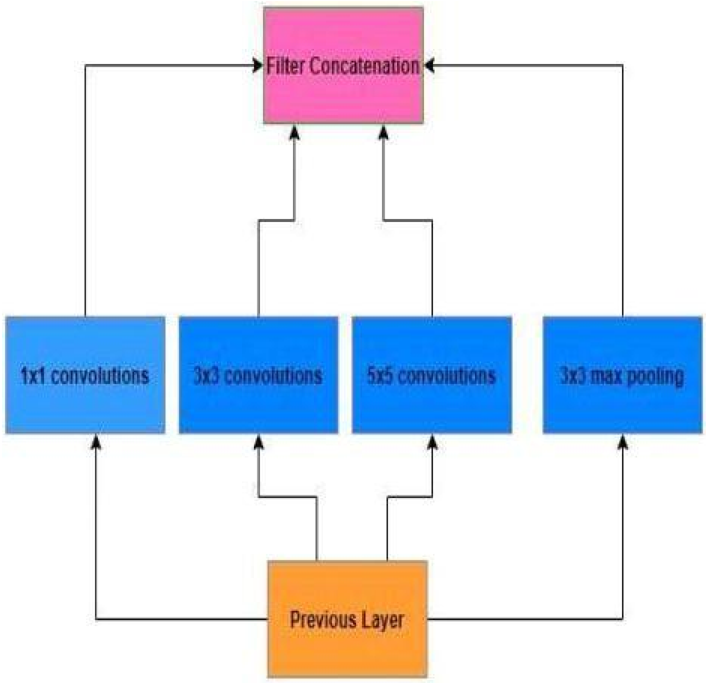


Figure 3.9 Pooling Operation

c) Activation Layer

In any neural network activation layer plays an important role as it is responsible for nonlinear learning of the network. There are different types of activation functions such as sigmoid, tanh, ReLU, LeakyReLU.In our model we have used ReLU for all the layers except output layer for which we have used softmax.

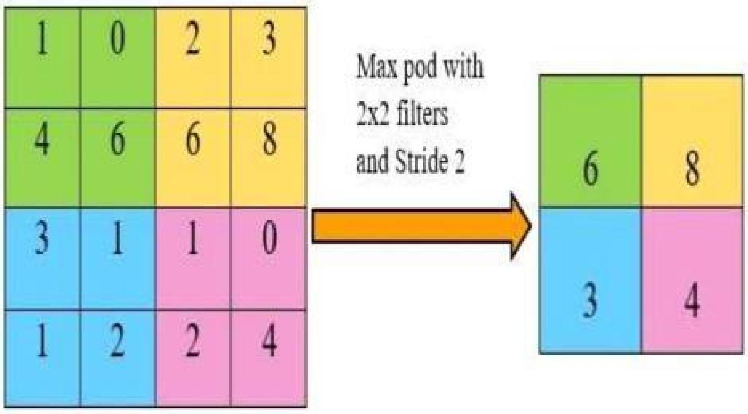


Figure 3.10 Naive inception module

d) Fully Connected Layer

After performing all the computations in previous layers, the output is feed into normal neural network for classification purpose. Model have 2 dense layers with 1024 and 19 unitsrespectively.

3.4 PROPOSED SYSTEMS

Plant Disease Detection Model An automated system designed to help identify plant diseases by the plant's appearance and visual symptoms could be of great help to amateurs in the gardening process and also trained professionals as a verification system in disease diagnostics. Advances in computer vision present an opportunity to expand and enhance the practice of precise plant protection and extend the market of computer vision applications in the field of precision agriculture. Exploiting common digital image processing techniques such as color analysis and threshold will be used with the aim of detection and classification of plant diseases. The acquired image encloses condensed information that is extremely difficult for the computer to process, it requires a pre-processing step to extract a certain feature (e.g., color and shape) that is manually predefined by experts. In such situations, deep learning is typically used because it allows the computer to autonomously learn the most suitable feature without human intervention

Among various network architectures used in deep learning, convolutional neural networks (CNN) are widely used in image recognition.

Hardware Requirements:

- ✓ RAM: 4 GB
- ✓ Storge: 500GB
- ✓ CPU: 2 GHz or faster
- ✓ Architecture: 32-bit or64-bit

Software Requirements:

- ✧ Python 3.5 in Google Colab is used for data pre-processing, model training and predection.
- ✧ .Operating System: windows 7 and above or Linux based OS or MACOS

Functional Requirements:

Functional requirements describe the system functionality, while the non-functional requirements describe system properties and constraints. Functional requirements capture the intended behavior of the system. This behavior may be expressed as services, tasks, or the functions the system is required to perform. This lays out important concepts and discusses capturing functional requirements in such a way they can drive architectural decisions and be used to validate the architecture. Features may be additional functionality, or differ from basic functionality along some quality attribute. In the proposed system, concert assesses the compliance of a workflow by analyzing the five established elements required to check for the rule adherence in workflows: activities, data, location, resources, and time limits.

Non-functional requirements:

1. Security System needs to control the user access and session. It needs to store the data in a secure location and stored in a secure format

It requires a secure communication channel for the data.

2. Concurrency and Capacity System should be able to handle multiple computations executing simultaneously, and potentially interacting with each other.

3. Performance is generally perceived as a time expectation. This is one of the most important considerations especially when the project is in the architecture phase.

4. Reliability It is necessary to ensure and notify about the system transactions and processing as simple as keep a system log will increase the time and effort to get it done from the very beginning. Data should be transferred in a reliable way and using trustful protocols.

5. Maintainability Well-done system is meant to be up and running for long time. Therefore, it will regularly need preventive and corrective maintenance. Maintenance might signify scalability to grow and improve the system features and functionalities.

6. Usability End user satisfaction and acceptance is one of the key pillars that support a project success. Considering the user experience requirements from the project conception is a win bet, and it will especially save a lot of time at the project release, as the user will not ask for changes or even worst misunderstandings.

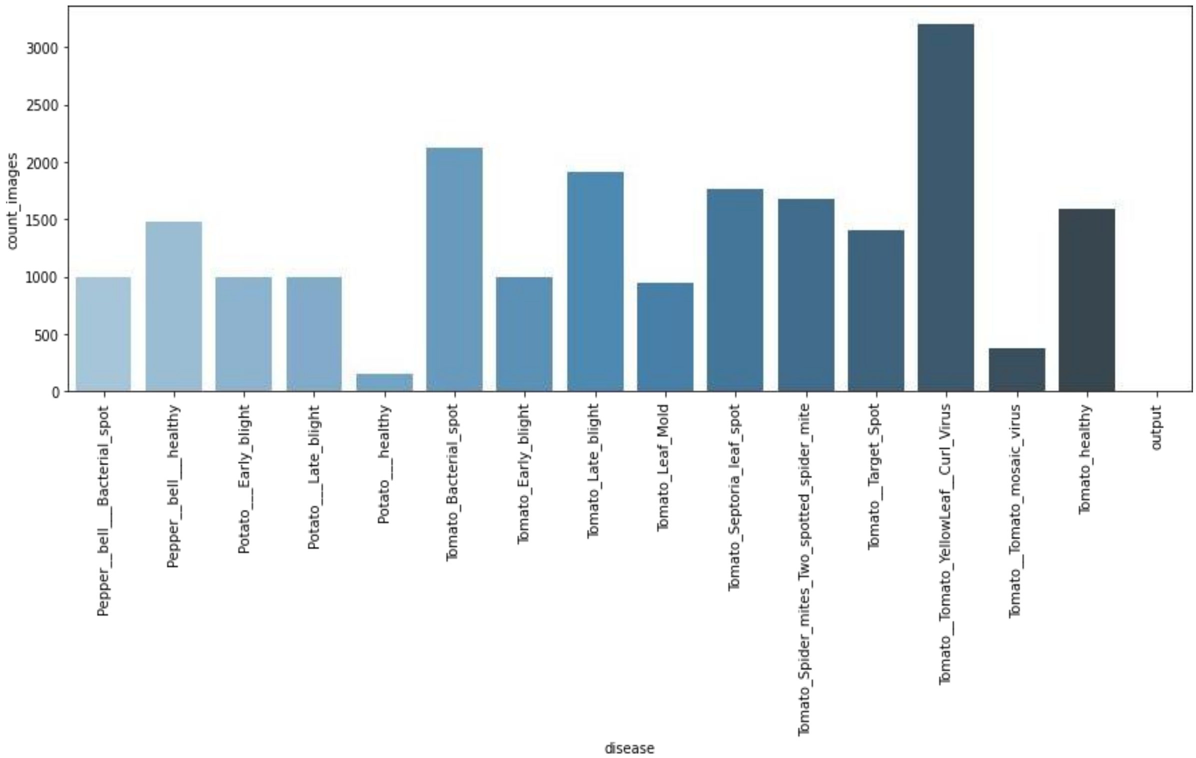
7. Documentation All projects require a minimum of documentation at different levels. In many cases the users might even need training on it, so keeping good documentation practices and standards will do this task spread along the project development; but as well this must be establishing since the project planning to include this task in the list.

3.5 Experiment Analysis

Datasets for plant disease detection The Plant Village dataset(PVD) is the only public dataset for plant disease detection to the best of our knowledge. The data set curators created an automated system using Google Net and Alex Net for disease detection, achieving an accuracy of 99.35%. However, the images in Plant Village dataset are taken in laboratory setups and not in the real conditions of cultivation fields, due to which their efficacy in real world is likely to be poor. In contrast, we curate real-life images of healthy and diseased plants to create a publicly available dataset.

3.5.1 THE PLANTDOC DATASET

The Plant Village dataset contains images taken under controlled settings. This dataset limits the effectiveness of detecting diseases because, in reality, plant images may contain multiple leaves



with different types of background conditions with varying lighting conditions.

Figure 3.11 Statistics of PlantDoc Dataset

3.5.2 Data Collection

To account for the intricacies of the real world, we require models trained on real-life images. This fact motivated us to create a dataset by downloading images from Google Images and Eloisa for accurate plant disease detection in the farm setting. We downloaded images from the internet

since collecting large-scale plant disease data through fieldwork requires enormous effort. We collected about 20,900 images by using scientific and common names of 38 classes mentioned in the dataset by Mohanty. Four users filtered the images by selecting images based on their metadata on the website and guidelines mentioned on APSNet . APS compiled a list of peer-reviewed literature corresponding to each plant disease. We referred APS' prior literature and accordingly classified images. Some of the most important factors for classification were the color, area and density of the diseased part and shape of the species. We removed inappropriate (such as nonleaf plant, lab controlled and out-of-scope images) and duplicate images across classes downloaded due to web search. Every image was checked by two individuals according to the guidelines to reduce labeling errors. Finally, to have sufficient training samples, we removed the classes with less than 50 images. Figure 2 shows the statistics of the final dataset having a total of 27 classes spanning over 13 species with 2,598 images. To build an application for the object detection task, we need exact bounding regions containing the leaf in the entire image. Hence, we used the Labeling tool to make the bounding boxes around the leaves (Figure 3) in all the images. In real scenarios, the image may have multiple leaves or a combination of diseased and healthy leaves. We labeled all the leaves in the image explicitly with their particular classes. While labeling the boxes, we made sure that the entire leaf should be present inside the box and the area of the bounding box should not be smaller than 1/8th (approximately) of the image size. After labeling, the information about all the coordinates of boxes in an image and their respective class label were stored separately in an XML file corresponding to each image.

Figure 3.12 Data Collection

3.5.3 BENCHMARKING PLANTDOC DATASET

We now discuss two benchmark set of experiments on our dataset: i) plant image classification; and ii) detecting leaf within an image. PreTrained Weights Training Set Test Set Accuracy F1-Score (Set %) (Set %) ImageNet PlantDoc (80) PlantDoc (20) 13.74 0.12 ImageNet PVD PlantDoc (100) 15.08 0.15 ImageNet+PVD PlantDoc (80) PlantDoc (20) 29.73 0.28 Table 1: Transfer Learning doubled the accuracy after finetuning on Uncropped PlantDoc datase.

PreTrained Weights	**Training Set**	**Test Set Accuracy (Set %)**	**Accuracy**	**F1-Score**
ImageNet	PlantDoc (80)	PlantDoc (20)	13.74	0.12
ImageNet	PVD	PlantDoc (100)	15.08	0.15
ImageNet+PVD	PlantDoc (80)	PlantDoc (20)	29.73	0.28

(Table Transfer Learning doubled the accuracy after finetuning on Uncropped PlantDoc dataset)

3.5.6 Experimental Setup

For training the networks, we used stochastic gradient descent with momentum 0.9, categorical crossentropy loss, and a learning rate of 0.001. All weights were initialized with the orthogonal initializer. We applied common data augmentation techniques such as rotation, scaling, flipping etc. on the input images. All images were resized to 100 × 100, before feeding into the networks. For pre-trained models, we used the weights provided in Keras trained on ImageNet.

Plant image classification using raw images (uncropped). Our first experiments aim to understand classification accuracy on the uncropped Plant Doc dataset. We evaluated the performance of VGG16 using different training sets on Plant Doc as shown in Table. Plant image classification using cropped images. Further, we evaluate the performance of several popular CNN architectures on the Cropped-Plant Doc dataset that have recently achieved state of-the-art accuracies on image classification tasks on the popular datasets, such as ImageNet, CIFAR-10, etc. gives the complete list of the architectures that we used for benchmarking our Cropped-Plant

Doc dataset. This experiment was conducted to verify the performance of Plant Village in real-setting.

The aim of our next experiments is to evaluate the performance of Faster R-CNN with InceptionResnetV2 model and MobileNet model on our PlantDoc Dataset as shown in Table 3. We use mean average precision (mAP: higher is better) to evaluate the models and compare it with scores on COCO dataset since no evaluation exists in the domain of plant disease.Object Detection models require training for a much longer duration. For training Faster R-CNN with Inception Resnet v2 network, we used Momentum optimizer keeping a degrading learning rate with an initial value of 0.0006. For training the MobileNet network, with RMSprop as optimizer – we took an initial learning rate of 0.0005 with decay steps as 25000 and decay factor as 0.95. While training, data augmentation like random horizontal flip and random SSD crop was applied on input images. We split our dataset into 2,360-238 based on training-testing. We took the pre-trained weights and fine-tuned on training set of PlantDoc. As aforementioned, weprovide train-test splits of the dataset for consistent evaluation and fair comparison over the dataset in future.

3.6 Baseline Results:

The following results present the Accuracy, Precision, Recall and F1-Score of both the CNN and KNN model. In order to understand the concept of the abovementioned evaluation metrics, it is important to understand what True positives (TP), True negatives (TN), False positives (FP) and False negatives (FN) are. For example, for an image of a tomato leaf that contains the disease ‘Early blight’, the confusion matrix looks like as shown in Fig 4. The confusion matrix shown below is for one of the 10 categories/diseases.

If the model correctly predicts the image of the plant as containing the disease, then the outcome is known as a True positive (TP) outcome. If the model correctly predicts the image of the plant as not containing the disease, then the outcome is known as True negative (TN) outcome.

If the model incorrectly predicts the image of the plant as containing the disease, then the outcome is known as a False positive (FP) outcome. If the model incorrectly predicts the image of the plant as not containing the disease, then the outcome is known as a False negative (FN) outcome.

3.6.1 Accuracy

This metric is calculated by dividing the total number of correct predictions which is the True positive and True negative outcomes, with the overall total number of samples (Liu et al., 2014). As we can observe from Table 6 and Table 7, the CNN model scored a high accuracy of 98.5% in comparison to the KNN model which scored 83.6%. In this report since the dataset chosen at hand contains labels that contain the same number of samples, accuracy proves as a useful measure in comparing both models. However, in scenarios where the dataset is imbalanced accuracy cannot be proven useful due to the paradoxical finding known as the Accuracy paradox.

3.6.2 Precision This metric is calculated by dividing all the correctly predicted predictions (True positives) with the total positive predictions (True positives and False positives) (M and M.N, 2015). This metric highlights the exactness of a classifier and answers the question, out of all the samples that the classifier predicted as being ‘Early blight’, how many of them were actually ‘Early blight’. As we can observe from Table 6 and Table 7, the CNN model scores a precision score of 93% and the KNN model scores 90%. This proves that among the two classifiers, the CNN model will significantly return more relevant results than irrelevant ones in comparison with the KNN model. 5.1.3 Recall This metric is calculated by dividing all the correct predictions (True positives) with the total number of True positives plus the False negatives (M and M.N, 2015). This metric is used to highlight the sensitivity of the model, meaning among the samples labelled as ‘Early blight’ in the dataset, how many did the classifier recognize as ‘Early blight’ in its predictions. As we can observe from Table 6 and Table 7, the CNN model scores a recall of 93% and the KNN model scores 84%, proving that the CNN model is able to return most of the relevant results in the plant village dataset in comparison to the KNN model.

3.6.3 F1-Score

Since in both precision and recall the CNN model has produced better scores than the KNN model, comparing the F1-Score of the models is hardly necessary as F1-Score is a harmonic mean of both precision and recall and is used to serve as alternative metric that helps in deciding if a model that produces high precision and low recall is more suitable than a model that produces low precision and high recall. However, from Table 6 and Table 7 we can see that the CNN model even in the F1-score metric performs better than the KNN model, producing an F1-Score of 93% and the KNN model producing a score of 86%. The training vs validation accuracy and the training vs validation loss were also plotted for the CNN model in which we can observe from

Fig 5 and Fig 6 that as the training accuracy increases the validation accuracy increases and as the training loss decreases the validation loss also decreases signifying that the model is neither under fitting nor overfitting and that the model is constantly learning. Under fitting in machine learning models occurs when a model performs poorly on the training dataset and performs poorly on the validation/testing dataset as well. Overfitting in machine learning models occurs when the model performs really good on the training dataset but performs poorly on the validation/testing dataset.

In conclusion, to answer our first research question, from the above evaluation metrics we can observe that the results are as hypothesized and CNN does in fact produce better results than the KNN model, which makes the CNN model a better suit for application in plant disease detection. The explanations as to why CNN performed better than the KNN model can be found from the core architecture that the CNN model is equipped with, namely, the convolutional layers, pooling layers and fully connected layers. In a CNN the image goes through many learnable layers in order to understand the different aspects of the image such as edges, shapes, objects, etc. Meanwhile, the KNN model only looks at the neighbors present near the image that is to be classified in order to predict its label. The model complexity that lies behind both these models is one of the reasons why the CNN model outperformed the KNN model. However, when comparing the training time of both the models, the KNN model easily wins over the CNN model. The KNN model took approximately an hour to finish its training and the CNN model took approximately 4 hrs. to finish its training. If the KNN model had produced similar or closer results to the CNN model, in that case, considering the training time the KNN model would have been the best pick.

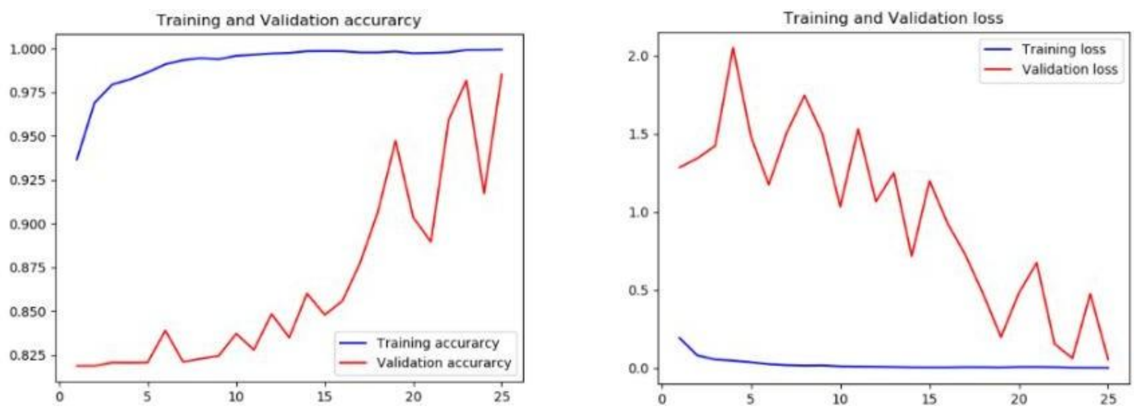


Figure 3.13 Training Vs Validation Accuracy

Chapter - 4

4. RESULTS

4.1 Results of a Web Application

Home Page:

Depicts the first look of our front end. We have a option to choose a image “Choose File”. It has a button called “Predict” which can be used to browse the images on the system’s hard disk.

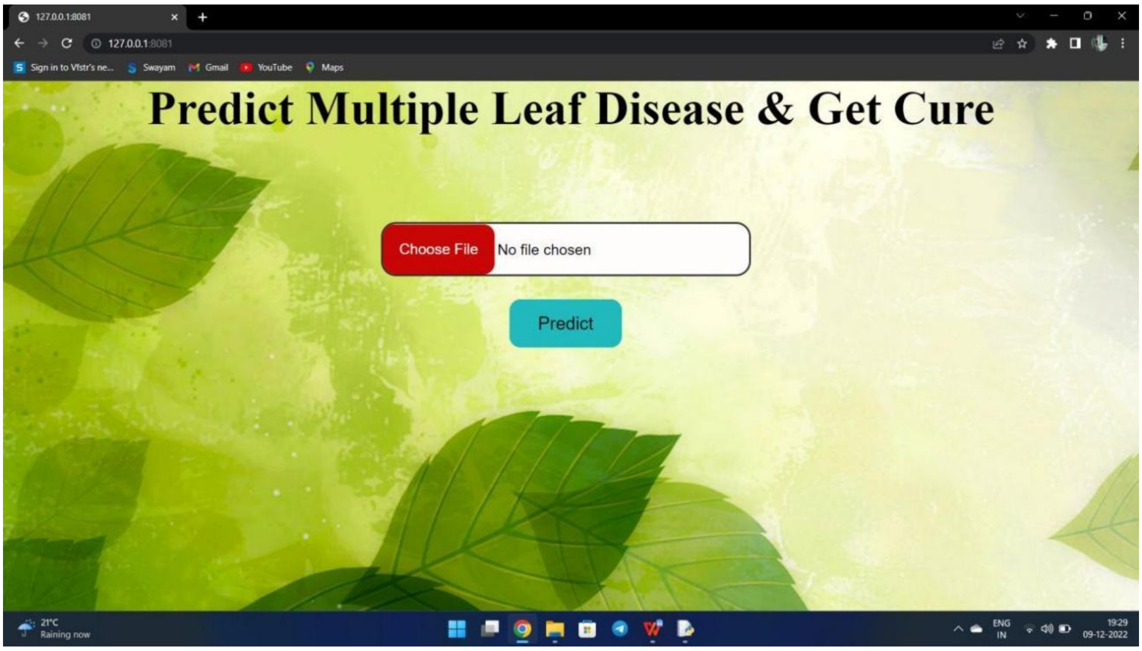


Figure 4.1 Home Page

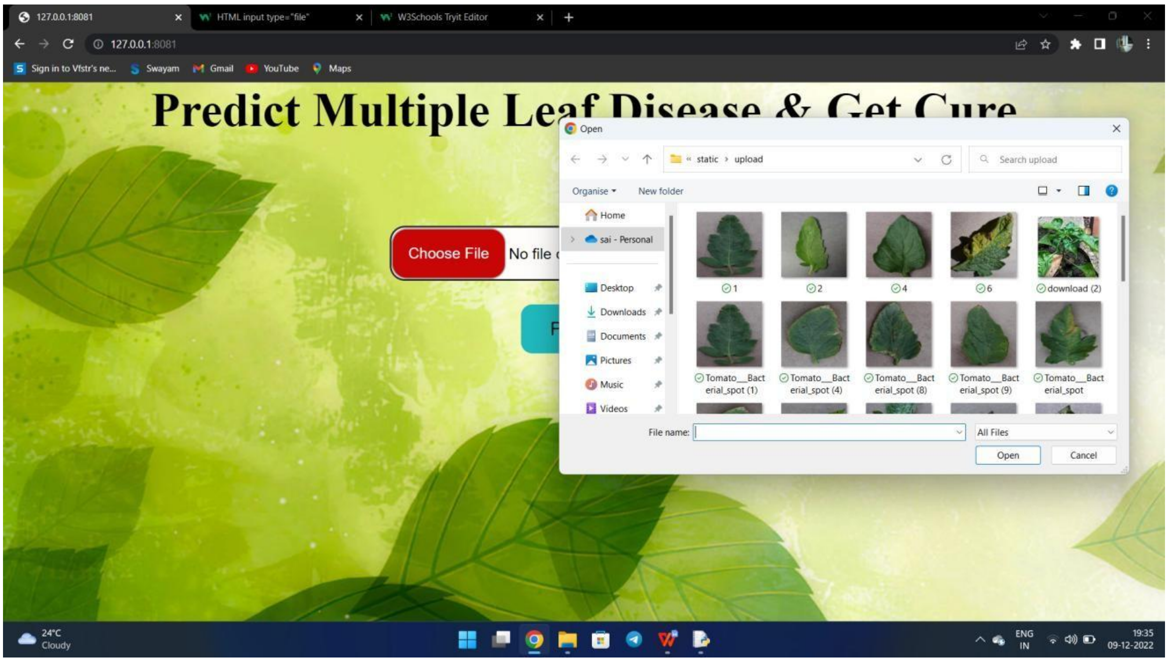


Figure 4.2 Selection of Images from a Datasets

The above fig, shows the popup which appears when user clicks on 'get photo' button. The popup window will be having number of input images to be selected, this action should be confirmed with a double click or an open button.

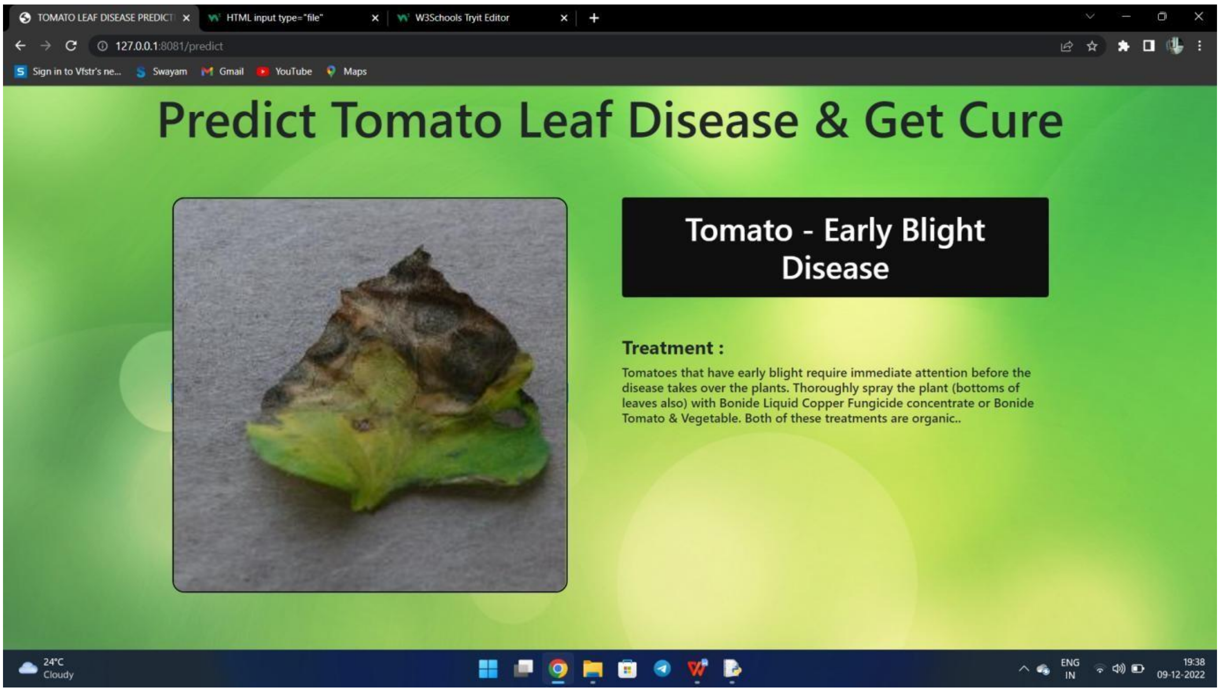


Figure 4.3 Result and Analyze of a plant

The above image, displays the status of leaf i.e. Healthy or Unhealthy, if it is unhealthy it displays the particular disease name. Here is a button called 'Remedies' by clicking this button it displays the Remedies for particular diseases.

Chapter - 5

5. Summary, Conclusion & Future Work and Extension

5.1 Summary

There are number of ways by which we can detect disease of plants and suggest remedies for them. Each has some pros as well as limitations. On one hand visual analysis is least expensive and simple method, it is not as efficient and reliable. Image processing is a technique which is most spoken for very high accuracy and least time consumption are major advantages offered. The applications of K-means clustering and Neural Networks (NNs) have been formulated for clustering and classification of diseases that effect on plant leaves. Recognizing the disease accurately and efficiently is mainly the purpose of the proposed approach. The experimental results indicate that the proposed approach is a valuable approach, which can significantly support an accurate detection of leaf diseases in a little computational effort. Alongside the supply of cultivation tools, the farmers also need access to accurate information that they can use for efficient crop management and there is no better way than providing them a service that they can use through the software.

The aim of this paper was to implement two different machine learning models for the detection of plant diseases and provide explainability for the predictions made by each machine learning model. The models that were chosen for this study were the, Convolution Neural Network and K-nearest Neighbor. In order to evaluate the models, metrics such as Accuracy, Precision, Recall and F1-score were used in this study. The expandability of models was done using the Explainable Artificial Intelligence technique, Local Interpretable Model-agnostic Explanations. Both the models were tested on the same dataset called the plant village dataset. On comparing both the models using the performance metrics, the results show that the CNN model performs better than the KNN model in the application of plant disease detection. The CNN model scored an accuracy of 98.5%, precision of 93%, recall of 93% and f1-score of 93%, while the KNN model managed to score an accuracy of 83.6%, precision of 90%, recall of 84% and f1-score of 86%. Explanations were generated for both CNN and KNN model's predictions using LIME and a user study was conducted to evaluate if the farmers trust and understand the implemented model's predictions and explanations. Results from the user study highlight the areas that farmers look for when diagnosing plant diseases and the results also indicate that the predictions and explanations from AI and XAI models are not adequate enough to trust the models for the purpose of detecting plant diseases. The feedback from the farmers help the study in identifying areas that might help boost trust in farmers and in turn help improve future iterations of the study. In short the key contributions of this thesis were, 1. Implementation of LIME to generate explanations for plant diseases 2. Execution of a user study to evaluate user trust of farmers 3. Identification of areas that could possibly help user trust grow in farmers 4. Identification of LIME's instability for image data.

5.2 CONCULSION

This paper presents an automated, low cost and easy to use end-to-end solution to one of the biggest challenges in the agricultural domain for farmers – precise, instant and early diagnosis of crop diseases and knowledge of disease outbreaks - which would be helpful in quick decision making for measures to be adopted for disease control. This proposal innovates on known prior art with the application of deep Convolutional Neural Networks (CNNs) for disease classification, introduction of social collaborative platform for progressively improved accuracy, usage of geocoded images for disease density maps and expert interface for analytics. High performing deep CNN model "Inception" enables real time classification of diseases in the Cloud platform via a user facing mobile app. Collaborative model enables continuous improvement in disease classification accuracy by automatically growing the Cloud based training dataset with user added images for retraining the CNN model. User added images in the Cloud repository also enable rendering of disease density maps based on collective disease classification data and availability of geolocation information within the images. Overall, the results of our experiments demonstrate that the proposal has significant potential for practical deployment due to multiple dimensions – the Cloud based infrastructure is highly scalable and the underlying algorithm works accurately even with large number of disease categories, performs better with high fidelity real-life training data, improves accuracy with increase in the training dataset, is capable of detecting early symptoms of diseases and is able to successfully differentiate between diseases of the same family.

5.2 FUTURE WORK AND EXTENSIONS

Future work involves expanding the model to include more parameters which can improve the correlation to the disease. We can augment the image database with supporting inputs from the farmer on soil, past fertilizer and pesticide treatment along with publicly available environmental factors such as temperature, humidity and rainfall to improve our model accuracy and enable disease forecasting. We also wish to increase the number of crop diseases covered and reduce the need for expert intervention except for new types of diseases. For automatic acceptance of user uploaded images into the Training Database for better classification accuracy and least possible human intervention, a simple technique of computing the threshold based on a mean of all classification scores can be used. Further application of this work could be to support automated time-based monitoring of the disease density maps that can be used to track the progress of a disease and trigger alarms. Predictive analytics can be used to send alerts to the users on the possibility of disease outbreaks near their location.

Reference

[1] R. Nalawade, A. Nagap, L. Jindam and M. Ugale, "Agriculture Field Monitoring and Plant Leaf Disease Detection," 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA), 2020, pp. 226-231, doi: 10.1109/CSCITA47329.2020.9137805.

[2] K. K. Singh, "An Artificial Intelligence and Cloud Based Collaborative Platform for Plant Disease Identification, Tracking and Forecasting for Farmers," 2018 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), 2018, pp. 49-56, doi: 10.1109/CCEM.2018.00016.

[3] P. Sharma, P. Hans and S. C. Gupta, "Classification Of Plant Leaf Diseases Using Machine Learning And Image Preprocessing Techniques," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2020, pp. 480-484, doi: 10.1109/Confluence47617.2020.9057889.

[4] D. Nath, P. K. Dutta and A. K. Bhattacharya, "Detection of plant diseases and nutritional deficiencies from unhealthy plant leaves using machine learning techniques," 4th Smart Cities Symposium (SCS 2021), 2021, pp. 351-356, doi: 10.1049/icp.2022.0368.

[5] V. Singh, Varsha and A. K. Misra, "Detection of unhealthy region of plant leaves using image processing and genetic algorithm," 2015 International Conference on Advances in Computer Engineering and Applications, 2015, pp. 1028-1032, doi: 10.1109/ICACEA.2015.7164858.

[6] S. Ramesh et al., "Plant Disease Detection Using Machine Learning," 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), 2018, pp. 41-45, doi: 10.1109/ICDI3C.2018.00017.

[7] A. Lakshmanarao, M. R. Babu and T. S. R. Kiran, "Plant Disease Prediction and classification using Deep Learning ConvNets," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), 2021, pp. 1-6, doi: 10.1109/AIMV53313.2021.9670918.

[8] S. V. Militante, B. D. Gerardo and N. V. Dionisio, "Plant Leaf Detection and Disease Recognition using Deep Learning," 2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), 2019, pp. 579-582, doi: 10.1109/ECICE47484.2019.8942686.

[9] M. Sardogan, A. Tuncer and Y. Ozen, "Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm," 2018 3rd International Conference on Computer Science and Engineering (UBMK), 2018, pp. 382-385, doi: 10.1109/UBMK.2018.8566635.

[10] V. V. Srinidhi, A. Sahay and K. Deeba, "Plant Pathology Disease Detection in Apple Leaves Using Deep Convolutional Neural Networks : Apple Leaves Disease Detection using EfficientNet and DenseNet," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1119-1127, doi: 10.1109/ICCMC51019.2021.9418268.

[11] T. R. Ganesh Babu, S. Priya, J. G. Chandru, M. Balamurugan, J. Gopika and R. Praveena, "Prediction and Analysis of Plant-Leaf Disease in Agricultural by using Image Processing and Machine Learning Techniques," 2021 International Conference on Computational Performance Evaluation (ComPE), 2021, pp. 540-544, doi: 10.1109/ComPE53109.2021.9751855.

[12] V. A. Metre and S. D. Sawarkar, "Reviewing Important Aspects of Plant Leaf Disease Detection and Classification," 2022 International Conference for Advancement in Technology (ICONAT), 2022, pp. 1-8, doi: 10.1109/ICONAT53423.2022.9725870.

[13] S. Ashok, G. Kishore, V. Rajesh, S. Suchitra, S. G. G. Sophia and B. Pavithra, "Tomato Leaf Disease Detection Using Deep Learning Techniques," 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 979-983, doi: 10.1109/ICCES48766.2020.9137986.

[14] B. Demir and S. Ertürk, "Improving SVM classification accuracy using a hierarchical approach for hyperspectral images," 2009 16th IEEE International Conference on Image Processing (ICIP), 2009, pp. 2849-2852, doi: 10.1109/ICIP.2009.5414491.

[15] A. Lakshmanarao, M. R. Babu and T. S. R. Kiran, "Plant Disease Prediction and classification using Deep Learning ConvNets," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), 2021, pp. 1-6, doi: 10.1109/AIMV53313.2021.9670918.