## A Project/Dissertation Final Report

## on

**Implementation of Classification and Segmentation for Plant Diseases Using a Deep Learning Model**

Submitted in partial fulfillment of the requirement for the award of the degree of

BACHOLOR OF COMPUTER SCIENCE AND ENGINEERING



Under the supervision of Mr. Anandhan K AP/SCSE

Galgotias University

Submitted By

Maya Shankar Jha 19SCSE1010573 Muskan Rahuja 19SCSE1010074

**SCHOOL OF COMPUTING SCIENCE AND ENGINEERING DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING / DEPARTMENT OF COMPUTERAPPLICATION**

**GALGOTIAS UNIVERSITY, GREATER NOIDA**

**INDIA**

**MAY, 2022**

**A picture containing clipart

Description automatically generatedSCHOOL OF COMPUTING SCIENCE AND ENGINEERING**

**GALGOTIAS UNIVERSITY, GREATER NOIDA**

**CANDIDATE’S DECLARATION**

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **“**Implementation of Classification and Segmentation for Plant Diseases Using a Deep Learning Model**”** in partial fulfillment of the requirements for the award of the Project review submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of MAY,2022, under the supervision of Mr. Anandan K ,Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

Maya Shankar Jha 19SCSE1010573  
Muskan Rahuja

19SCSE1010074

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Mr. Anandan K AP/SCSE

Galgotias University

**CERTIFICATE**

The Final Thesis/Project/ Dissertation Viva-Voce examination of Maya Shankar Jha(19SCSE1010573) & Muskan Rahuja(19SCSE1010074) has been held on 10-05-2022 and his/her work is recommended for the award of

### Signature of Examiner(s) Signature of Supervisor(s)

**Signature of Project Coordinator Signature of Dean**

Date: May, 2022

Place: Greater Noida

ACKNOWLEDGMENT

We would like to thank our guide and reviewers for their valuable suggestions on our work.

PG. 2

# ABSTRACT

Misdiagnosis of many diseases affecting agricultural crops can lead to the misuse of chemicals that lead to the emergence of resistant pathogens, increased input costs, and further outbreaks with significant economic losses and environmental impacts. Current diagnostic tests based on human testing are time-consuming and expensive, and although computer-based models promise increased efficiency, significant differences in symptoms due to age of infected tissue, genetic variation, and light conditions in trees reduce the accuracy of detection. We are going to build a model based on deep learning which can accurately identify the diseases type in apple leaves that either it is affected by scab or rust or by multiple disease and help farmers to take related correct step. The models will be trained by using different data sets of images collected from many agricultural locations by the Kaggle team to better understand the disease.Model uses deep learning convolutional neural network training to classify the images.

PG. 3

### Contents

### Title

**Chapter No. Page No.**

### Candidates Declaration I

Acknowledgement II

[Abstract III](#_TOC_250005)

[Contents IV](#_TOC_250004)

[List of Table V](#_TOC_250003)

[List of Figures VI](#_TOC_250002)

[Acronyms VII](#_TOC_250001)

[Chapter 1 Introduction 8-11](#_TOC_250000)

* 1. Introduction
  2. Formulation of Problem
     1. Tool and Technology Used

### Chapter 2 Literature Survey/Project Design 12-21

**Chapter 3 Functionality/Working of Project 22-25**

### Chapter 4 Modules

* 1. Modules description

### Images

* 1. Flowchart

### 23-34

### Chapter 5 Conclusion and Future Scope 36

* 1. Conclusion
  2. Future Scope

### Reference 37-50

**Publication/Copyright/Product**

PG. 4

### List of Table

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Caption** | **Page No.** |
| **1** | **Comparison Table** | **13** |
|  |  |  |
|  |  |  |
|  |  |  |

PG. 5

### List of Figures

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Title** | **Page No.** |
| 1 | Datasets | 26-27 |
| 2 | Flowchart | 28 |
| 3 | CNN MODEL ARCHITECTURE | 24 |
| 4 | EPOCH & PREDICTION | 32-37 |

PG. 6

### Acronyms

|  |  |
| --- | --- |
| CNN | Convolutional Neural Network |
| MT | Million Tones |
| VI | Venturia Inaequalis |
| KNN | k-nearest neighbors |
|  |  |
|  |  |
|  |  |

PG. 7

# CHAPTER- 1 INTRODUCTION

Agriculture, a major occupation all over the world plays a lead role in Indian economy. The area under apple cultivation in India increased by 24% from 1.95 lakh ha. in 1991-92 to 2.42 lakh ha. in 2001-02 although production increased by less than 1% (i.e. from 11 to 12 lakh tones). It is mostly grown in the states of Jammu & Kashmir, Himachal Pradesh, Uttaranchal, Arunachal Pradesh and Nagaland.

Agriculture, with its allied sectors, is the largest source of livelihoods in India. 70 percent of its rural households still depend primarily on agriculture for their livelihood, with 82 percent of farmers being small and marginal. In 2017-18, total food grain production was estimated at 275 million tonnes (MT).

Apple farming is an important activity and profession of farmer communities in the Himalayan states of India. At present, the traditional apple farming is under stress due to changes in climate. Apple (Malus pumila) is commercially the most important temperate fruit and is fourth among the most widely produced fruits in the world after banana, orange and grape. China is the largest apple producing country in the world. The area under apple cultivation in India increased by 24% from 1.95 lakh ha. in 1991-92 to 2.42 lakh ha. in 2001-02 although production increased by less than 1% (i.e. from 11 to 12 lakh tones). It is mostly grown in the states of Jammu & Kashmir, Himachal Pradesh, Uttaranchal, Arunachal Pradesh and Nagaland.

An agri-export zone has been established in Himachal Pradesh covering the districts of Shimla, Siramour, Kullu, Mandi, Chamba and Kinnaur considering the vast potential for increasing exports. Initial targets are for exporting apples to neighbouring countries as well as to west Asia and to the south-east Asian countries. With increase in production of apples in the State, the State government is taking effective steps for its marketing besides providing packing material,

PG. 8

transportation, ensuring adequate procurement of apples under Market Intervention Scheme, so that the growers get remunerative prices in the markets. Apple can be grown at altitudes 1,500- 2,700 m. above m.s.l. in the Himalayan range which experience 1,000-1,500 hours of chilling (the no. of hours during which temperature remains at or below 70 C during the winter season). The temperature during the growing season is around 21-240 C. For optimum growth and fruiting, apple trees need 100-125 cm. of annual rainfall, evenly distributed during the growing season. Excessive rains and fog near the fruit maturity period result in poor fruit quality with improper colour development and fungal spots on its surface. Areas exposed to high velocity of winds are not desirable for apple cultivation.

Loamy soils, rich in organic matter with pH 5.5 to 6.5 and having proper drainage and aeration are suitable for cultivation.

Arunachal Pradesh is one of the ideal locations for apple cultivation. If high yielding varieties of apple are introduced in the state it can make a big headway by exporting the produce to Bangladesh which presently depends on Bhutan besides meeting the internal demand of the northeastern markets.

Efforts are being made by the NHB and the State Govts. for imparting modern technologies to the growers in apple producing states.

To date, Apple's best year of manufacturing in India was in 2017, when it shipped 3.2million units. In 2020, it had shipped 3.1million units despite the pandemic.

## Formulation of Problem

Diseases have adverse effects on the plant and agricultural fields. The main causes of these diseases are viruses, genetic disorders, and infections agents such as bacteria, fungi, and viruses. Therefore, to identify and evaluate these plant diseases of such importance impel us to builds an automated intelligence that can be built yield , enhances farmer profits, and more contribution to

PG. 9

the national economy. Earlier many researchers in the field of computer science image processing and photographic proposal is proposed for use traditional imaging techniques such as LBP, K- means the combination of detecting these leaf diseases. In-depth learning models are better at mapping and therefore are better feature generators. Therefore, we created an in-depth learning model to find a leaf diseases that use several classifications in this paper.

List of most common apple leaf diseases :-

* + 1. Apple scab (Venturia inaequalis)
    2. Marssonina leaf blotch (pre mature leaf fall)
    3. Black rot canker
    4. Collar rot
    5. Powdery mildew
    6. Sooty blotch and fly speck
    7. Apple mosaic and other virus diseases
    8. Alternaria leaf spot/blight
    9. Core rot 10.Brown rot 11.White rot / root rot 12.Seedling blight 13.IPM for Apple

Many people and technological groups are involved in the field of agriculture to increase the yield and throughput. There has been various techniques used in the past to solve problems related to disease spread in a tomato plant. With the advancement in technology tomato plant disease detection have become more easy and precise. In our system a different approach, i.e. KNN algorithm is used for the same. Various kind of methods have been used recently to

PG. 10

determine the type of plant disease. Some of these involves analysis and study of chemical analysis method to determine plant diseases, and ways which are indirect by implementing physical techniques, like spectroscopy of the leaf and imaging, to get information related to properties of tomato plant. Following this, the merits of the project contrasted with the existing technologies are related to the underlying points : • The system avoids the process involved in gathering inputs for studying them in the laboratory, because of pre-existing images taken in place of the plant diseases. It examines the chances where a particular plant is concurrently simulated with higher than one pest or disease in the unchanged recorded input. The outlook deploy inputting of various images apprehended by various cameras with diverse resolutions, like mobile phone and the other available cameras devices. The project is systematically pact with different conditions related to illuminations, the size of actors in an image, and surrounding distinction, etc., holding across the neighboring part of that particular plant. It imparts a feasible functioning approach that is able to maneuver in the domain by not using costly and complex and compound technologies.

PG. 11

### CHAPTER- 2 LITERATURE SURVEY

Studies says that most of the fungus, viruses and bacterial disease in plants leaf and roots can easily be studied and identified using the image processing techniques like RGB to grey scale conversion.

With a purpose of providing more accurate and faster solution for detecting disease in plants leaf, this system uses k means clustering approach for detecting a more accurate data for plant disease with a precision between 83% and 94%.[1]

Model uses logistic regression over the pre-trained models like VGG19 for extracting the relevant features from the dataset for fine-tuning(transfer learning) and outperformed the other existing models with a big classification accuracy around 97.8% over the test dataset.[2]

By monitoring and diagnosing police diseases, a number of advances have been made including RGB imaging, X-ray, ultrasound, and many hyperspectral technologies. The method proposed by Macedo-Cruz et al aims to measure the damage caused by ice on oat plants. The authors used three blocking techniques: the Otsu method, the Isodata algorithm, and the ambiguous blockchain. A program proposed by Yao et al. aims to identify and classify three types of diseases affecting rice crops. This image is divided in the manner of Otsu, after which the diseased areas are separated. Color, shape and texture features are extracted, the latter appearing in the HSV color space. The components are sent to the Vector Support Machine, which performs the final separation.

The method proposed by Padikar and Sil detects and separates the two diseases affecting the rice plants, transforming the image into a HSI color spectrum, an entropy-based boundary used for differentiation. A detector is used on the edge in a split image, and spots are detected using the size of the green material.

PG. 12

The images are separated by a K-means algorithm, and then 50 colors, textures and texture elements are extracted. The method proposed by Wiwart et al aims to identify and discriminate between the four types of mineral deficiencies - nitrogen, phosphorus, potassium and magnesium. Prior to color analysis, images are converted to HSI and L \* a \* b \* color spaces. Those differences are calculated by Euclidean ranges, which are calculated in both color spaces. In their two papers, Kurniawati et al. proposed a way to identify and label three different types of diseases that affect powdered crops. Recently, Sharada Prasanna Mohanty et al. use a deep convolutional neural network to identify 14 plant species and 26 diseases. Our work will use key features of RGB images and machine learning to detect disease in plants.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| NAME | YEAR | ACCURACY | DATASET | ALGORITHM |
| Fast and Accurate Detection and Classification of Plant Diseases[1] | 2011 | 83-94 | Own generated | K means clustering |
| Detection of potato diseases using image segmentation and  multiclass support vector machine [3] | 2017 | 95 | Plant village, kaggle | Multi class vector support machine |
| Plant Diseases Classification using Machine Learning [5] | 2021 | 84.94 | plant village dataset, kaggle | extreme learning machine(elm) |
| Plant Leaf Disease Classification and Detection System Using Machine  Learning [6] | 2020 | 70% | Self collected | KNN |
| Identification of Potato Late Blight Disease from Crop Images Captured under Uncontrolled Environment[7] | 2014 | 93% | Plant village, kaggle | BPNN |

Table 1 - Comparision Table

PG. 13

"Apple fruit quality test: A study" by Komal Sindhi, Jaymit pandya when they focused on fruit quality testing. First the author outlines various factors such as color, texture, and basis in which the classification of diseases is performed. Diagnosis is made using this database, which is categorized according to the factors.

Detection And Classification Using Random Forest Classifier

In this paper we have considered three common diseases of apple fruit namely apple peel, apple rot and apple color. The color and texture features are combined together and the random forest section is used for disease classification. Therefore, if the fruit is infected with any one disease then the infected part is separated using the k-means combination method.

Automatic Fruit Quality Inspection System

The main purpose of this program is to replace the manual testing program. This helps speed up the process to improve accuracy and efficiency and reduce time. This program collects animated images and the image is processed to determine the characteristics of the fruit such as texture, color and size. Then the Disable Fruit is based on the acquisition of the blob, the color acquisition is based on achievement and the size acquisition is based on the binary image of the tomato. And sorting and grading are done based on color and size.

Mr. Abhijeet, professor. A.P. patil (2017). In this study paper, the diagnosis of apple disease is done by kmeans clustering techniques and the Learning vector Quantization neural network. First, we fix the database and split the image and remove the feature. The final step is a neural network test that results in the apple disease detection rate of this algorithm being more than 95% accuracy.

Miss.Kambale, Anuradha Manik, Dr.Mrs.Chougule (2015). In this paper the author conducts image processing to classify and classify images of apple fruit. First, we created two databases -

PG. 14

one with standard images and the other with featured images. Second, we do pre-image processing and image classification. Then the feature removal is done and after that the splitting and editing is done. The results of the experiments suggest that the proposed solution could significantly support the automatic division into categories and disease of apple fruit diseases.

### Machine learning in detecting and classifying diseases of a plant leaf

Identification of diseases is one of major area in agriculture which needs to be taken care of, though many practices have been done and implemented to cope up with this issue, rapid and quick identification of the diseases still remains in state of inchoate. The use machine learning in facilitating the identification and detection helps to counter this problem to a much greater extent.

### Reviewing classification and detection on plants using ML.

The paper gives an elaborate view about the techniques which can be implemented for detecting and classifying the various plant leaf diseases caused by bacteria, viruses and fungi. Based on their morphology i.e., their particular form, shape, or structure the diseases detected through classification are categorized. The techniques used in classification aids in automatic detection of the diseases of plant leaf [7].

### Machine learning in detection of stem diseases of jute plant.

In this paper detection of diseases of the stem plant is done, using the HSV algorithm, GLCM algorithm and SVM to perform and initiate the segmentation process, followed with feature extraction and classification respectively. It discusses the removal of noise, conversion from RGB to HSV and vice versa [8].

### Detection of abnormalities of the leaves of plants and training using papaya leaves.

PG. 15

This paper talks about the detection and recognition of abnormalities of plants for training and study papaya leaves were taken. Random forest classifier was used for classification and it got trained using images of leaves with an almost seventy percent accuracy [9].

### Apple leaf disease detection.

The common Apple leaf diseases like rust, grey spot, brown spot were discussed and found out with the help of deep learning algorithms and improved CNNs. The dataset for diseased leaves were generated, processed and collected. New deep CNN model designed to identify small diseased spots [10].

### Improvisation in our work:

We understood the functioning of the classification algorithms, feature extraction algorithms, segmentations algorithms etc. We studied how detection of the disease is done automatically and the same is implemented effectively in the real time project. We chose tomato plant leaves for study, training, testing and detection of diseases.

### Study of plant leaf diseases and use of digital image processing

The digital image processing provides a vast area for identification of diseases through the various algorithms it supports.

### Image processing techniques in identifying fungal crop diseases.

The most common bacterial, fungal and viral diseases are studied which affects the plant leaves and roots on wide scale and reduces the productivity of the plants can be easily studied and identified through RGB to grey scale conversions [11].

PG. 16

UAV for pests and weed using open computer vision In this paper drone cameras systems are used for detection of pests or diseased weed or crops. When the diseased areas are identified, those particular areas only are treated with fertilizers or pesticides and not the whole region so that the control can be done swiftly and in the minimum time possible [12].

### Severity measurement using image processing.

Simple threshold methods and triangle threshold methods are used so that the lesion area and are of the leaf can be segmented. Categorization is done calculating quotient of lesion and leaf regions. Plants such as sugarcane have various kinds of such diseases which can affect the amount and the quality of crop and its yield. To avoid this it‟s important to know the severity of the diseases so the appropriate amount of fertilizers can be used in time.

### Improvisation in our work:

Understanding the pain and efforts of the farmers and how much time is devoted to cultivate a crop for one season, we employed the method which monitors the plants and its leaves from very beginning i.e., by using digital image processing. Early detection can save the farmers from huge losses hence our work focuses on early and correct detection of diseases using digital image processing so that neither the crop nor its yield is affected. Also the algorithms which are supported by image processing are quite useful in segmentation and in feature extractions which forms an essential part of the project.

### Selection of algorithms

In machine learning various algorithms are available for feature extraction, clustering, segmentation. Selecting the most suitable as per needs of the task can be tough at times. To reduce the complexity and improve the time of response selection of the most suitable algorithm is

PG. 17

required. We did a comparative study of the algorithms used in various previous projects and reasoned out the best one suitable for the project.

### Pre-processing and filters.

Before we start working on the image we need to smooth out the image and resize it for further working. Various kinds of noises are also available in the images like Gaussian noise, shot noise, noise related to salt and pepper representation, Quantization Noise (Uniform Noise), film noise etc. [22] There are different methods for removing different kinds of noise in our case we have salt and pepper noise hence to remove such kind of noise median filter is considered to be one the best algorithms, so we went ahead with the same.

### Segmentation

The functioning and well known and significant are of image processing has to be segmentation as it derives useful and more appropriate to say meaningful data set from, from meaningless data by segmenting the parts of the image into several parts. It can be classified on the basis of region, edge threshold, feature and model. Out the above ours project deals with feature extraction so segmentation on such basis is important to simplify the further work and, we found that K-Mean algorithm which works well for feature extraction segmentation also is the one of the mostly widely used and easy to implement algorithm. Hence we decided to go with it.

### Feature extraction

The basic geometric features which are extracted in feature extraction are diameter, physiological width, leaf area, leaf perimeter, morphological features, rectangularity etc. The GLCM or the Grey level Co-occurrence Matrix a traditional algorithm is very often used to extract the spatial dependency of the texture. Similarly another algorithm used for feature extraction is HOG i.e.

PG. 18

Histograms of Oriented Gradients in the word we proposed we have used this system because this promises performance. Also, another advantage of HOG is that the very general structure of the object is captured by this algorithm. Hence we chose to work with GLCM and HOG.

### Classification

The algorithms used for classification are quite commonly used for almost all similar projects like SVM i.e., Support Vector Machine, CNN, KNN. The classification is a necessary step as it compares the values received after the feature extraction step is compared with a pre-calculated set of data here we went ahead with KNN algorithm though both are supervised algorithms the KNN is simple easy and flexible to implement. It is also quite robust to data containing noise.

### Improvisation in our work:

Different plants require different algorithm depending upon various factors some even upon the morphology [7]. We compared and contrasted each of the above algorithms before finally selecting it with test and trial methods.

ELM is proposed by Vijayalakshmi & Murugan [13] to detect plant diseases. Their model follows the basic steps of machine learning model such as feature extraction, training the classifier and classification. In their work, the artificial bee colony clustering (ABC) [14] is used isolate infected areas using the collective behaviour of insects such as bees. Second, the feature extraction is completed using transform encoded local patterns (TELP) to compute the texture analysis [15], gradient features and color histogram techniques from the data. For the training and testing, the author tested both classifiers namely Support Vector Machine (SVM) [16] and ELM, which resulted in ELM leading by 2% at the rate of accuracy at 97%. This work proves that ELM can surpass the conventional model, SVM in terms of classification accuracy. Another approach on ELM is suggested by Saragih [17] with Simulated Annealing (SA) to classify the Jatropha Curcas

PG. 19

Disease. In this work, the dataset of jatropha diseases is used in the process of model training, testing, and the optimising the output of the weights of SA [18]. The important role of SA is to tune the best weight value of neuron on ELM classifier. The model is further enhanced using the decision tree classifier to determine the jatropha seed disease. The aim of the decision tree classifier is to create a model capable of making a judgement according to the specialised knowledge. The result of this method gives out a much more desirable rate of accuracy at 94.74% compared to pure ELM at 66.67%. Manso [19] proposed a smartphone application to classify the coffee leaf diseases with ELM. The system recognises diseases such as leaf miners and leaf rust and calculates the rate of its severity. The image processing step involves the identification of a leaf from the image by separating the background and foreground of the image to obtain the Hue Saturation Value (HSV) and YCbCr colour space. This method is done via several segmentation methods that include histogram distribution, edge detection, neighbourhood detection, and clustering. Otsu algorithm [20] and the iterative threshold algorithm are used to perform a comparison to k-means in the process to segment the foliar damages. The results show the practicality of the model to identify feeding damage by obtaining the of the severity rate of 99.095%. Thus, the ability of the ELM model is proved for able to produce a significant result with the given image features of the coffee leaves. Maniyath [21] recommends a comparison of various machine learning models such as Support Vector Machine (SVM), K-Nearest Neighbours (KNN), RF and naïve bayes algorithm. It uses the public datasets of diseased leaves for training and testing purposes. Second, the histogram of an orientated gradient (HOG) is used to extract features such as hu moments, Haralick textures and colour histogram. The datasets of diseased and healthy leaves are then collectively trained under the aforementioned models. The comparison resulted with random forests having the highest accuracy at 70.14%. The model can be improved with implementing the ELM model as the random tree does not scale well with the amount features compared to ELM. Panchal, et al. [22] uses Random Forest (RF) as a classifier

1. to detect plant diseases, which share similarities with the previous titles mentioned. Images

PG. 20

of leaves with plant diseases such as bacterial spot, early blight and late blight are acquired using a camera or gadgets with similar capabilities to create a dataset. The image processing step comprises of converting the images to grey scale as well as HSV and smoothing of the images using a smoothing filter. Then, the research has utilised the k-means segmentation method to isolate the infected portion of the leaf and Grey Level Co-occurrence Matrix (GLCM) [24] to extract features from leaf. The proposed methodology can recognise and differentiate the plant diseases with the rate of 98% accuracy when the RF classifier is used to determine the plant diseases. Besides, HSV histogram features can be considered the alternative of the GLCM. It is capable of describing an image in a general presentation which is required for a classification model. Bhatia, et al. [25] suggests, “Application of ELM in Plant Disease Prediction for Highly Imbalanced Dataset”, that implements ELM classifier for plant disease prediction according to an imbalanced dataset consisting of real time images being the Tomato Powdery Mildew Disease (TPMD) dataset. To balance this dataset, several re-sampling techniques such as Random over Sampling (ROS), Synthetic Minority Over-Sampling Technique (SMOTE), Random under Sampling (RUS) and Importance Sampling (IMPS) ahead of performing the classification using ELM. The sample models are further evaluated using the Classification Accuracy (CA) and area under the curve (AUC). This technique resulted to ELM performing better for TPMD dataset after the resampling techniques were applied. The optimum outputs are found from the IMPS technique that shows 88.57% in CA and 89.19% in AUC. The result can be improved; hence this work will use tomato dataset. According to the literature review, it is observed that image features play a significant role in improving classification result. In addition, ELM, despite using single RBF can give good classification results with suitable features.

PG. 21

# CHAPTER- 3 FUNCTIONALITY/WORKING OF PROJECT

### MACHINE LEARNING

Plant diseases affect the growth of their respective species, therefore their early identification is very important. Many Machine Learning (ML) models have been employed for the detection and classification of plant diseases but, after the advancements in a subset of ML, that is, Deep Learning (DL), this area of research appears to have great potential in terms of increased accuracy. Many developed/modified DL architectures are implemented along with several visualization techniques to detect and classify the symptoms of plant diseases. Moreover, several performance metrics are used for the evaluation of these architectures/techniques.

### DEEP LEARNING

The Deep Learning (DL) approach is a subcategory of Machine Learning (ML), introduced in 1943 when threshold logic was introduced to build a computer model closely resembling the biological pathways of humans. This field of research is still evolving; its evolution can be divided into two time periods-from 1943–2006 and from 2012–until now. During the first phase, several developments like backpropagation, chain rule, Neocognitron, hand written text recognition (LeNET architecture), and resolving the training problem were observed However, in the second phase, state-of-the-art algorithms/architectures were developed for many applications including self-driving cars, healthcare sector, text recognition, earthquake predictions, marketing, finance, and image recognition. Among those architectures, AlexNet is considered to be a breakthrough in the field of DL as it won the ImageNet challenge for object recognition known as ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in the year 2012. Soon after, several architectures were introduced to overcome the loopholes observed previously. For the evaluation of these algorithms/architectures, various performance metrics were used. Among these metrics, top-1%/top-5% error, precision and recall, F1 score, training/validation accuracy and loss,

PG. 22

classification accuracy (CA) are the most popular. For the implementation of DL models, several steps are required, from the collection of datasets to visualization mappings.

In this program, we aim to identify and identify diseases that affect plant leaves. Initially, a leaf image is taken with a camera and stored in a specific location. This is a picture pre-processing is done, starting with the conversion of the RGB image into a Grayscale image. The the image is taken in RGB format, and the captured image is converted from the RGB color model to Gray. This image is stored in the destination of the same directory. Next, comes the file reproduction of image, which is performed to feed input data to the neural network at a specific pixel width. Pixels the width described here is 256 \* 256. Next, we calculate the mean values, variations and standard deviations. The following is the process of removing the feature. This can be described as a form of reduction in size they represent well the enhanced parts of the image in the form of a vector feature. Photo at the beginning darkened to some degree, so that it slips easily. Now, this smooth image is stored in the file direction. This is taken from the training data index. This image is used for neural training network. The image could be resized and read for better precision or accuracy. In addition, this image is included in the file training guide. Our proposed system having four modules: Acquisition of Image, Pre-processing on images, Extracting features from images after pre-processing step and finally Classification.

### CNN Model

The CNN model is used for disease detection and classification, the model includes an Input layer, Convolutional layers, Pooling layers, fully connected layers, and output layer. The model is implemented using python programming language and Tensorflow library, which “is an interface for expressing machine learning algorithms, and an implementation for executing such algorithms”. The model architecture consists of four convolutional layers. The convolutional layer is a core layer on the CNN, and it is the reason behind the CNN name. The convolutional

PG. 23

layer used to extract the feature map form the input image, then creating a new matrix with a smaller size using filters. In the proposed model, the first layer uses a 7×7, which is the largest filter size on the model to represent the general features from the input image. The second convolutional layer uses a 5×5 filter, then in the last two layers,3×3 filter used. The ReLu activation function was used after the matrix convoluted through all convolutional layers. The pooling layer condenses the output matrix from the previous convolutional layer. Max pooling function used at each pooling layer. The first pooling layer filter size is 3×3, and the next pooling layers use 2×2 filters. The dropout function utilized in the model to avoid model overfitting. Lastly, there is one fully connected layer with 128 neurons, activated using ReLu activation function. Lastly, the output layer is activated using the Softmax function, which contains three neurons. For design and implementation of the model architecture, TensorFlow and Keras libraries were used, which based on python programing language. The model architecture was implemented on the cloud-based kernel to achieve high computing power and graphical computing power.

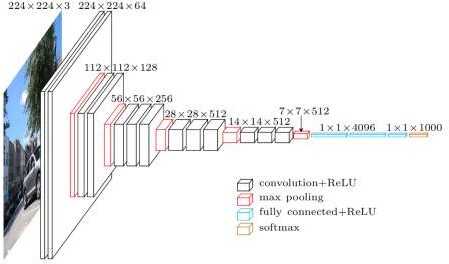


Fig1-CNN MODEL ARCHITECTURE

PG. 24

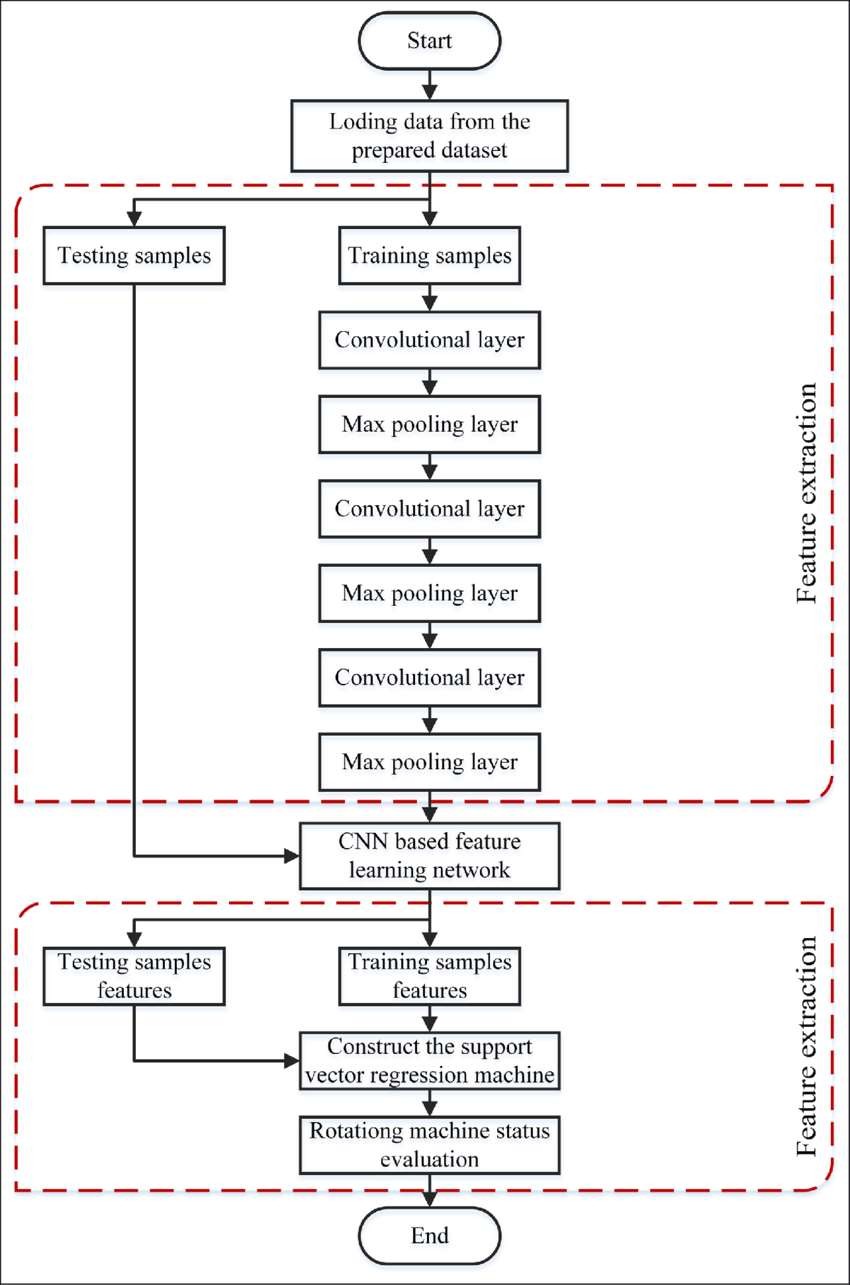


Fig 2. Flow Diagram Of CNN

PG. 25

# CHAPTER 4 MODULES



* Collecting and importing dataset
* Preparing and training data
* Data processing
* Model preparation
* Making the prediction on a single image

COLLECTING AND IMPORTING DATA

Dataset contains 3,642 images and 3 csv files containing the train and test images with affected disease with the value of 0 and 1 in train.csv file.

This dataset have been collected from Kaggle platform. Here with importing the PIL, OpenCV, TQDM and NATSOR, we have declared the train variable with train.csv file, test variable with test.csv and DIR with image local directory path.

## IMAGES

PG. 26



## Fig.3-Sample image of leaves in different condition

PREPARING AND TRAINING DATA

In this module we are training the dataset or in other word we can say that we are preparing the data in required manner to train our model. Here, firstly the collection of train images is labelled with the value with 1, 2, 3, and 4 for the healthy, multiple disease, scab and respectively using the train.csv file. With the help of natsort library, we have sorted the train images as per their name and number. After that the train images is divided into four different folders with their characteristics or affected disease.

Diagram

Description automatically generated

Fig 4 Architecture diagram for the proposed system

PG. 27

FLOWCHART

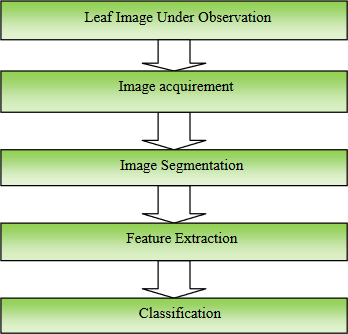


Fig. 5 -A proposed flowchart for image processing

DATA PREPROCESSING

This step involves processing the test image in order to bring it to the size, colour and quality of the images comprising our dataset. This involves various stages through which the image goes. These stages are:

* 1. Image Resizing: The dimensions of the image are brought equal to the dimensions of the training images by using the „imresize ()‟ method in MATLAB. Image resizing is a crucial step as the pixel values may change if the dimensions of all the training as well as the testing images are not the same.

PG. 28



Fig-6-Training of Images

* 1. Smoothening: The smoothening of the image renders the pixel values to gradually even out to all the points of the image so as to allow a smooth image. Along with this the image also gets converted from colored to grayscale image using the function „ RGB2GRAY ()‟. 3. Noise Filtering: The noise is the unwanted extras that are present in the images that make feature determination and extraction difficult. Thus the process of noise filtering involves removal or averaging of the pixel values that add noise to the image. The process used in our system to ensure noise removal is „Median Filter

PG. 29



Fig-7-Training of Images

In this part, all the images extracted after the preparation of training data with categorised folders, the images is converted from BGR (Blue, Green, Red) to RGB (Red, Green, Blue) with the help of OpenCV liberary. One more important thing is size declaration of image.

All images which are used for training the model is resized.

PG. 30

MODEL PREPARATION

Neural networks is used to create deep learning models. It contains an input layer, output layer and an hidden layer. The input layer takes the input and these inputs are processed under hidden layer and after that output layer provide the prediction based on the data provided to model. This model is ongoing.

MAKING THE PREDICTION ON A SINGLE IMAGE

In this part a particular image is selected and its path is declared as value of an variable and the model predict that either that leaf is affected by any disease of not and if that is affected then which type of disease it is. This module is also ongoing.

To make sure that the proposed CNN model achieves the desired performance accuracy, the model was validated using large number of leaf images, which are different images than the images used for the training process includes healthy leaf images, infected by scab, and infected by rust. For model testing, the process started with testing data entered to the model then followed by step to be classified. The testing data pass into the model every epoch to calculate the validation accuracy. The validation and training accuracies through 30 epochs were represented. After finishing from all epochs, the total validation accuracy will be calculated. The total accuracy for the proposed model is 93.8%.

PG. 31

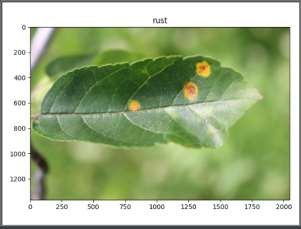
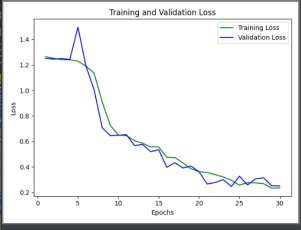
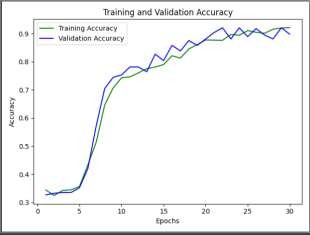


Fig .8-training and validation accuracy epoch graph

Fig 9- training and validation loss epochs

Fig 10 -leaf affected by rust

PG. 32

Dataset

Visualization

Technique

Plant Disease Detection

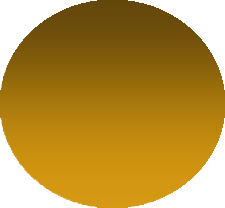
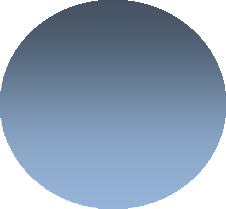
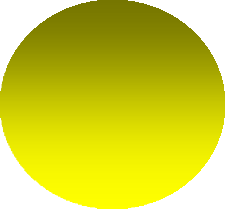
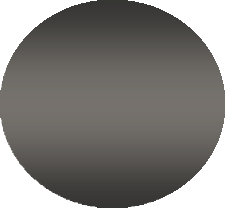
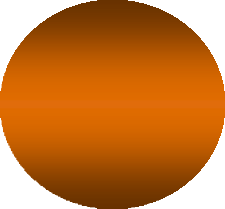
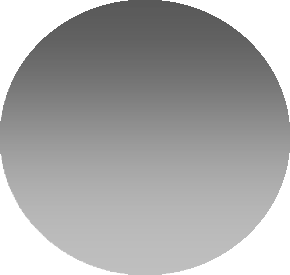
Splitting

the data

Performance Metrices

Train Model

Architecture



PG. 33

# CHAPTER 5 CONCLUSION AND FUTURE SCOPE

## CONSLUSION

An intensively designed CNN-based model attention is to detect the familiar apple leaf disease to maintain and enhance the effective management of Orchards. A large number of apple leaf images have been collected by several cameras and cell phone devices. A total of 3,642 images have been acquired from the public database from the Kaggle platform. It contains healthy apple leaves, affected by rust and scab, or are affected by many diseases. In this paper, an apple pest control program has remained proposed which stands based on the technology of Convolutional Neural Network (CNN) having a well-trained architectural design. The primary purpose of this proposed model is to diagnose potted plant diseases early in order to increase plant production and reduce losses due to diseases such as apple peel and rust on the leaves of the apple plant. The model has been trained and validated using a dataset containing 3642 colored apple leaf pictures. The dataset is divided into training and testing. Models are tested in 30 epochs and each epoch contains 91 total steps. The accuracy rate of this proposed model is around 93.8%.

PG. 34

## REFERENCES

1. H. A. Hiary, S. B. Ahmad, M. Reyalat, M. Braik, and Z. Alrahamneh, “Fast and Accurate Detection and Classification of Plant Diseases,” International Journal of Computer Applications, vol. 17, no. 1, pp. 31–38, 2011.
2. Potato Leaf Diseases Detection Using Deep Learning May 2020 DOI:[10.1109/ICICCS48265.2020.9121067](http://dx.doi.org/10.1109/ICICCS48265.2020.9121067) Conference: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)
3. M. Islam, Anh Dinh, K. Wahid and P. Bhowmik, "Detection of potato diseases using image segmentation and multiclass support vector machine," 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, 2017, pp. 1- 4.

[4 ] Samer I. Mohamed, “Potato Leaf Disease Diagnosis and Detection System Based on Convolution Neural Network”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-9 Issue-4, November 2020

[5 ] Tan Soo Xian & Ruzelita Ngadiran, Plant Diseases Classification using Machine Learning [doi:10.1088/1742-6596/1962/1/012024, 1962 (2021) 012024]

[6 ] G. Geetha, S.Samundeswari, G.Saranya ,K.Meenakshi and M. Nithya Plant leaf disease classification and detection system using machine learning, Journal of Physics: Conference Series 1712 (2020) 012012 doi:10.1088/1742-6596/1712/1/012012

PG. 35

1. L. S. Puspha Annabel, T. Annapoorani and P. Deepalakshmi, “Machine Learning for Plant Leaf Disease Detection and Classification – A Review,” 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2019, pp. 0538-0542.
2. Zhang, Harry. "The optimality of naive Bayes." AA 1.2 (2004): 3.
3. Z. N. Reza, F. Nuzhat, N. A. Mahsa and M. H. Ali, "Detecting jute plant disease using image processing and machine learning," 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Dhaka, 2016, pp. 1-6.
4. S. Ramesh et al., "Plant Disease Detection Using Machine Learning," 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, 2018, pp. 41-45.
5. J. D. Pujari, R. Yakkundimath and A. S. Byadgi, “Identification and classification of fungal
6. disease affected on agriculture/horticulture crops using image processing techniques,” 2014 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, 2014, pp. 1-4.
7. S.Vijayalakshmi and D.Murugan, “An Effective Approach for Diagnosis of Plant Disease using ELM,” vol. 4, no. 12, pp. 327–335, 2017.
8. A. Kumar, D. Kumar, and S. K. Jarial, “A review on artificial bee colony algorithms and their applications to data clustering,” Cybern. Inf. Technol., vol. 17, no. 3, pp. 3–28, 2017, doi: 10.1515/cait-2017-0027.

PG. 36

1. Z. Pan, Z. Li, H. Fan, and X. Wu, Feature based local binary pattern for rotation invariant texture classification, vol. 88. Elsevier Ltd, 2017.
2. B. Kalantar, B. Pradhan, S. Amir Naghibi, A. Motevalli, and S. Mansor, “Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN),” Geomatics, Nat. Hazards Risk, vol. 9, no. 1, pp. 49–69, 2018, doi: 10.1080/19475705.2017.1407368.
3. T. H. Saragih, W. F. Mahmudy, A. L. Abadi, and Y. P. Anggodo, “Application of extreme learning machine and modified simulated annealing for jatropha curcas disease identification,” Int. J. Adv. Soft Comput. its Appl., vol. 10, no. 2, pp. 108–119, 2018.
4. L. M. R. Rere, M. I. Fanany, and A. M. Arymurthy, “Simulated Annealing Algorithm for Deep Learning,” Procedia Comput. Sci., vol. 72, pp. 137–144, 2015, doi: 10.1016/j.procs.2015.12.114.
5. G. L. Manso, H. Knidel, R. A. Krohling, and J. A. Ventura, “A smartphone application to detection and classification of coffee leaf miner and coffee leaf rust,” arXiv, 2019.
6. C. Huang, X. Li, and Y. Wen, “AN OTSU image segmentation based on fruitfly optimization algorithm,” Alexandria Eng. J., vol. 60, no. 1, pp. 183–188, 2020, doi: 10.1016/j.aej.2020.06.054.
7. S. R. Maniyath et al., “Plant disease detection using machine learning,” Proc. - 2018 Int. Conf. Des. Innov. 3Cs Comput. Commun. Control. ICDI3C 2018, no. April, pp. 41–45, 2018, doi: 10.1109/ICDI3C.2018.00017.

PG. 37

1. P. Panchal, V. C. Raman, and S. Mantri, “Plant Diseases Detection and Classification using Machine Learning Models,” CSITSS 2019 - 2019 4th Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. Proc., vol. 4, pp. 1–6, 2019, doi: 10.1109/CSITSS47250.2019.9031029.
2. M. Belgiu and L. Drăgu, “Random forest in remote sensing: A review of applications and future directions,” ISPRS J. Photogramm. Remote Sens., vol. 114, pp. 24–31, 2016, doi: 10.1016/j.isprsjprs.2016.01.011.
3. O. R. Indriani, E. J. Kusuma, C. A. Sari, E. H. Rachmawanto, and D. R. I. M. Setiadi, “Tomatoes classification using K-NN based on GLCM and HSV color space,” Proc. - 2017 Int. Conf. Innov. Creat. Inf. Technol. Comput. Intell. IoT, ICITech 2017, vol. 2018-Janua, pp. 1–6, 2018, doi: 10.1109/INNOCIT.2017.8319133.
4. A. Bhatia, A. Chug, and A. Prakash Singh, “Application of extreme learning machine in plant disease prediction for highly imbalanced dataset,” J. Stat. Manag. Syst., vol. 23, no. 6, pp. 1059–1068, 2020, doi: 10.1080/09720510.2020.1799504.

PG. 38