



**A**

**Project Report**

on

**Plant Leaf Disease Prediction Using Machine Learning**

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY

**DEGREE**

SESSION 2020-24

in

**Computer Science and Information Technology**

By

Deepanshu Goel (2000290110058)

Garvit Chauhan (2000290110065)

Aditya Singh(2000290110016)

Anshul (2000290110032)

**Under the supervision of**

Dr. Meeta Chaudhary

**KIET Group of Institutions, Ghaziabad**

Affiliated to

# Dr. A.P.J. Abdul Kalam Technical University, Lucknow

(Formerly UPTU)

# May, 2024

**DECLARATION**

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

Date: 18/12/2023 Team members:

**Deepanshu Goel (2000290110058)**

**Garvit Chauhan (2000290110065)**

**Aditya Singh(2000290110016)**

**Anshul (2000290110032)**

**CERTIFICATE**

This is to certify that Project Report entitled “Plant Leaf Disease prediction using Machine Learning” which is submitted by **Deepanshu Goel, Garvit Singh, Aditya Singh, Anshul** in partial fulfillment of the requirement for the award of degree

B. Tech. in Department of Computer Science and Information Technology of Dr. A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

.

**Date: 18/12/2023 Supervisor**

Dr. Meeta Chaudhary

**ACKNOWLEDGEMENT**

It gives us a great sense of pleasure to present the report of the B.Tech Project undertaken during B.Tech. Final Year. We owe special debt of gratitude to **Dr. Meeta Chaudhary** Assistant Professor, Department of Computer Science and Information Technology ,KIET, Ghaziabad, for her constant support and guidance throughout the course of our work. her sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his/her cognizant efforts that our endeavors have seen light of the day.

We also take the opportunity to acknowledge the contribution of **Dr. Abhinav Juneja,** Head of the Department of Computer Science and Information Technology, KIET, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

Date: 10/12/2023 Team members:

**Deepanshu Goel (2000290110058)**

**Garvit Chauhan (2000290110065)**

**Anshul (2000290110032)**

**Aditya Singh (2000290110016)**

**ABSTRACT**

Agriculture is the primary source of livelihood for about 70% of the rural population in India. The crop variety cultivated in India is very diverse. There are more than 500 crop varieties grown in India. Despite the technological advances, the agricultural practices are still manual and involve less automation than western countries. Most of the diseases affecting a plant will reflect the damage in the leaves. The diseases affecting the plant can thus be identified from the leaf images. This paper presents an automatic plant leaf damage detection and disease identification system. The first stage of the proposed method identifies the type of the disease based on the plant leaf image using RCNN. The RCNN model is trained on images categorized according to their nature, i.e., healthy and the type of the disease. This model is then used for testing new leaf images. The proposed model produced a classification accuracy of 96%, with fewer images used during the training stage. The second stage identifies the damage in the leaf using deep learning-based semantic segmentation. Each RGB pixel value combination in the image is extracted, and supervised training is performed on the pixel values using the 1D Convolutional Neural Network (CNN). The trained model can detect the damage present in the leaves at a pixel level. Evaluation of the proposed semantic segmentation resulted in an accuracy of 97%. The third stage suggests a remedy for the disease based on the disease type and the damage state. The proposed method detects various defects in different plants in the experimental analysis, namely apple, grape, potato, and strawberry. The proposed model is compared with the existing techniques and obtained better performance in comparison with those methods.

## TABLE OF CONTENTS

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | **Page No.** |
| Certificate | |  | ii |
| Declaration | |  | iii |
| Abstract | |  | iv |
| Acknowledgement | | | v |
| Table of contents | | | vi |
| List of Figures | |  | viii |
| List of Tables | |  | ix |
| List of Abbreviations | | | x |
| 1 | INTRODUCTION | | 1 |
|  | 1.1 Relevance of the Project | | 5 |
|  | 1.2 Problem Statement | | 5 |
|  | 1.2.1 | Goal | 5 |
|  | 1.2.2 Comparison with Existing System | | 6 |
|  | 1.2.3 | Solution/Implementation | 6 |
|  | 1.2.4 | Impact | 6 |
| 2 | LITERATURE SURVEY | | 8 |
|  | 2.1 Financial Data Analysis | | 8 |
|  | 2.2 Time series Analysis | | 8 |
| 3 | OBJECTIVE AND METHODOLOGY | | 11 |
|  | 3.1 Understanding Cryptocurrency | | 11 |
|  | 3.2 Twitte | R | 12 |
|  | 3.3 Sentiment Analysis | | 13 |
|  | 3.4 Related work | | 14 |

|  |  |
| --- | --- |
| 4.REQUIREMENTS SPECIFICATION | 17 |
| 1 Functional Requirements | 17 |
| 2 Non-Functional Requirements | 17 |
| 3 Hardware Requirements | 18 |
| 4 Software Requirements | 18 |
| 4.SYSTEM ANALYSIS AND DESIGN | 19 |
| 4.1 System Architecture | 19 |
| 4.2 Flow Chart | 21 |
| 4.3 State Diagram | 21 |
| 4.3 Use-Case Diagram | 23 |
| 5. ALGORITHM | 25 |
| 5.1 Facebook Prophet | 25 |
| 5.2 CNN-LSTM | 26 |
| 6. IMPLEMENTATION | 29 |
| 6.1 Data Representation | 29 |
| 6.2 Data Pre-Processing | 30 |
| 6.3 Data Split | 30 |
| 6.4 Collection of Tweet | 30 |
| 6.5 Model Code | 33 |
| 7. TESTING | 40 |
| 8. RESULTS AND DISCUSIONS | 42 |

|  |  |
| --- | --- |
| 9. CONCLUSION & FUTURE WORK | 47 |

|  |  |
| --- | --- |
| 10. REFERENCES | 49 |

# CHAPTER 1

**Introduction**

In countries like India it is of utmost importance to bring technological advancement in the fields related to crop productivity. Research initiatives and tentative study process in the important domain of qualitative farming is focused towards improving the yield and food crop standard at low cost, with greater monetary outcome. Agricultural building model stands as a result of a compound interlinking of soil with seeds , and chemicals used to enhance growth. Vegetable and fruits exists as one of the present significant agricultural achieved output. In directive for getting surplus and effective worthy products, a product value examination and improvement has always been importantly imperative. Diseases are disablement to the conventional state of the plant that translates or hinders its important roles such as transpiration, photosynthesis, fertilization, pollination, germination etc. They distorting diseases are spawned by pathogens like, fungi, bacteria and viruses, because of unfavorable environmental situations. Accordingly, the preliminary stage for diagnosing of plant disease is a significant task. Farmers need periodic monitoring by professionals which might be prohibitively costly and time absorbing. Thence, looking for quick, less costly and precise ways to smartly detect the diseases from the indicators that look to be on the plant leaf is of great pragmatic importance. In our study we are proposing a system which can be used to identify the particular type of disease a tomato leaf might have. It is of major concern to identify the type of disease an important crop like tomato can have, by implementing upbringing technologies like image recognition, which represent the application functioning visually and it is also an important reason for making digital technologies popular. 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 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.Top of Form

#### Relevance of the Project

Plant leaf diseases are incredibly relevant in agriculture, horticulture, and even ecosystems. They can significantly impact plant health, crop yields, and overall ecosystem balance. Here are a few reasons why they are relevant:

1. Crop Yield Reduction: Diseases affecting leaves can lead to reduced photosynthesis, stunted growth, and, ultimately, lower crop yields. This can have significant economic implications for farmers and food production.

2. Ecosystem Imbalance: Diseases in plant leaves can disrupt the balance within ecosystems by affecting plant populations. This can alter food sources for various animals and insects, potentially affecting entire food chains.

3. Environmental Impact: Some leaf diseases can lead to defoliation, causing environmental issues such as soil erosion, loss of habitat, and changes in microclimates due to reduced shade or altered water cycle dynamics.

4. Food Security Concerns: Leaf diseases can threaten food security by impacting the quantity and quality of crops available for consumption. In regions where agriculture is a primary food source, these diseases can have severe consequences for communities.

5. Global Trade and Economy: Plant diseases affecting leaves can limit trade opportunities due to quarantine restrictions on affected produce. This can impact global agricultural markets and trade relations between countries.

6. Need for Disease Management: Understanding and managing plant leaf diseases is crucial for sustainable agriculture. This involves various strategies like crop rotation, genetic resistance, biological control, and the use of fungicides or pesticides.

7.Research and Innovation:The study of leaf diseases leads to advancements in plant pathology, genetics, and biotechnology. Developing disease-resistant crops or finding environmentally friendly ways to manage these diseases requires ongoing research and innovation.

Overall, the relevance of plant leaf diseases extends far beyond agriculture, touching upon environmental, economic, and societal aspects. Managing and understanding these diseases is crucial for sustainable food production and maintaining ecosystem health.

* 1. **Problem Statement**

##### Goal

The primary goal of plant leaf disease prediction is to anticipate, identify, and manage potential disease outbreaks in plants before they cause significant damage. Predictive models aim to forecast the occurrence, severity, and spread of diseases affecting plant leaves. Here are the key objectives:

Early Detection and Prevention:

Timely Intervention: Predictive models help in early detection, allowing farmers or gardeners to take preventive measures before diseases become widespread.

Reduced Damage: Early identification means early action, potentially minimizing the impact of diseases on crop yields and plant health.

Optimal Resource Management:

Efficient Resource Allocation: Anticipating disease outbreaks helps optimize the use of resources such as pesticides, fungicides, water, and labor, reducing unnecessary applications and costs.

Sustainable Practices: Predictive models encourage sustainable agricultural practices by promoting targeted and limited use of chemicals when necessary, minimizing environmental impact.

Improved Crop Management:

Precision Agriculture: Predictive models assist in precision farming, allowing for targeted treatment application only where and when needed.

Crop Rotation Planning: Anticipating disease outbreaks aids in planning crop rotations to break disease cycles and maintain soil health.

Enhanced Decision-Making:

Informed Decisions: Farmers and growers equipped with predictive models can make informed decisions on planting, cultivation practices, and disease management strategies.

Risk Mitigation: Understanding disease risks allows for proactive measures, reducing the likelihood of crop failure due to diseases.

Research and Development:

Advancing Knowledge: Disease prediction models contribute to ongoing research in plant pathology and genetics, aiding in the development of disease-resistant varieties and innovative management techniques.

Innovation in Disease Control: Predictive models drive innovation in disease control methods, leading to more effective and environmentally friendly solutions.

Ultimately, the goal of plant leaf disease prediction is to empower farmers, researchers, and agricultural communities with the knowledge and tools needed to minimize the impact of diseases, enhance crop productivity, and promote sustainable agriculture.

##### Comparison with Existing System

Existing Systems make use of Machine learning on a given data set to predict the disease.

##### Solution/Implementation

The proposed Solution is an algorithm which analyse the given data set and

predict disease .

##### Impact

This project would help the farmers so that they can use pesticide to save the crop.

* 1. **Summary**

The proposed methodology in the following tomato plant leaf disease detection system focus on generating an advance and efficient system which makes the process of creating high yield of tomato much more easier for the farmers. The project aims to detect the most common diseases occurring on a tomato leaf, namely early blight, bacterial spot and curl using image processing technique under upbringing technology i.e., machine learning. In easier terms, the farmer will be able to accurately detect the type of disease a particular plant is having using the image of the plant. The proposed system is based on four important modules namely: • Pre-processing. • Segmentation • Feature extraction. • Classification using KNN. In this study, we describe the comparison of our system with preexisting systems with proper methodology and implementation. The proposed systems functionality is better than existing disease detection system as it is able to generate a more accurate and precise result with easier and faster implementation. It aims to make the life of farmers easier. The system can be a boon to the agricultural sector as it advances the crop production and management process, as agriculture is of the major reason.

**CHAPTER 2**

# LITERATURE SURVEY

# 1.A Systematic Literature Review on Plant Disease Detection: Motivations, Classification Techniques, Datasets, Challenges, and Future Trends

In this systematic review, we surveyed studies that present plant disease and pest detection containing IP, ML, DL, and others. It is clearly demonstrated that plant pests and diseases harm the global agricultural industry. Despite the rapid expansion of AI-based solutions, there are still many barriers to overcome before high-performance real-time PDD solutions can be produced, according to a comprehensive assessment of PDD research employing imaging applications. This system review presented a comprehensive overview of current plant pest and disease detection studies. ML, IP, and DL-based plant disease models and monitoring technologies have shown promising results. The study considered 176 articles published between 2012 and 2022. These studies were selected after applying rigorous inclusion criteria from five academic databases, including ACM, Springer, Scopus, IEEE Xplorer, and Google Scholar. Our analysis presented significant relevant findings from the considered studies providing adequate responses to the research review questions. Most studies centered extensively around CNN-based disease detection systems for numerous crops, notably citrus, have been studied. Lightweight and TL algorithms, CNNs, GANs, attention mechanisms, and autoencoders have been investigated for high-functioning model construction, and a more comprehensive range of modifications can still be done in this paradigm to reduce the computation complexity. However, the present training paradigm for DL models require Systematic Literature Review on Plant Disease Detection data sets, making finding remedies for many plant diseases difficult. There are a limited number of publicly accessible datasets on this topic.

PCA reduced features, and SVM classified cucumber disease using SLIC, EM algorithm, and PHOG. Their model used in-field photographs to have the highest cucumber disease identification rate of 65.41% [87]. The image background was removed from the maximum correlation coefficient and global thresholding based on OTSU [162]. Their model predicts 95% okra YVMV disease and 82.67% bitter gourd disease. Stretching contrast, a top-hat filter, besides the Gaussian function, enhances images In addition, the bulk of DL models is created using data collected under laboratory circumstances, which may hinder their performance in real-time utilization. This study further resolves that industry and academia have many computation complexities and an excellent opportunity to avail models practically visible for realfield implementation. Finally, forthcoming reviews should scrutinize the works and considerable limitations of innovative agricultural applications, especially on plant pests and detection, which is another crucial research area. This survey covered a more comprehensive range of plant pests and diseases methods, challenges, and dataset checks and demonstrated that future research sparks new ideas and the concepts of relevant theories, methods, and practices in industries and academia.

**A Survey on Plant Leaf Disease Detection**

Sneha Patel1 | Dr. U.K. Jaliya1 | Pranay Patel1

Deep learning constitutes a recent, modern technique for image processing with accurate results. Many techniques of deep learning and image processing are used for leaf disease detection and classification. Deep learning techniques such as CNN, Fast RCNN, Faster RCNN, and Mask RCNN, and image processing techniques such as image preprocessing, segmentation, feature extraction etc. are used for disease detection. As per the survey, deep learning technique provides high accuracy than image processing technique. Plant leaf disease detection has wide range of applications available in various fields such as Biological Research and in Agriculture Institute. Agricultural productivity is something on which economy highly depends. This paper provides an overview of various techniques that are used for Plant Leaf Disease Detection. It also covers survey on different diseases classification techniques that can be used for plant leaf disease detection. Some authors are describing to find leaf diseases using various methods and to recommend the various implementations. Many researchers have used different techniques of image processing to detect the leaf disease and that follows steps: Image Acquisition, Image Pre-processing, Image Segmentation, Feature extraction, and classification. Abirami Devaraj has worked as a disease of Alternaria, Alternate, Anthracnose, Bacterial Blight and Cercospora Leaf Spot this disease is detected using image processing techniques that involve loading an image, image preprocessing, image segmentation, feature extraction, and classification. Velamakanni Sahithya is a ladies' finger plant leaves that are chosen and examined to find an early stage of various diseases such as a yellow mosaic vein, leaf spot, powdery mildew. Leaf images are captured, processed, segmented, features extracted, and classified and show the healthy or unhealthy. kMeans clustering is used for segmentation and for classification, SVM and ANN are used. This work uses PCA to reduce the Feature set. Priyanka Soni. This paper is defined specifically for leaf disease identification. The work is here divided into two major stages. In the first stage, the ring project-based segmentation model is defined to explore the features of leaf images. Once the features are identified, the next work is to apply the PNN classifier to identify the existence of a disease. The work is about to identify the health and infected disease based on featured region identification. The work is applied to randomly collect leaf images from the web for different plants. The simulation results show a clear and accurate identification of diseased leaf. Sujatha image processing used for the identification of leaf diseases. This is used k-means clustering and SVM. There are five steps for the leaf disease identification which are said to be image acquisition, image pre-processing, segmentation, feature extraction, classification. This approach by using different algorithms for segmentation, classification. By using this concept the disease identification is done for all kinds of leaves and also the user can know the affected area of leaf in percentage by identifying the disease properly the user can rectify the problem very easily and with less cost.

In this paper, survey on various techniques for Leaf Disease Detection is done. In the leaves, disease is the main reason for less production of vegetables and fruits. To overcome that issue using Deep Learning and Image Processing techniques. Different author used that techniques and different datasets for accurate result. After reviewing techniques we can conclude that there are number of ways by which we can detect disease of plants. Each has some advantages and limitations. According to survey Deep Learning Techniques is more accurate than Image Processing Techniques.

**3. Plant Leaf Disease Prediction**

Vaishnavi Monigari1 , G. Khyathi Sri2 , T. Prathima3 1, 2, 3Department of Information Technology, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India

The paper presents a deep learning model called the plant disease detector, which is able to detect different diseases of plants based on images of their leaves. This model is developed by applying advance neural network techniques. Initially dataset is augmented to increase sample size, and subsequently Convolution Neural Network (CNN) with multiple convolution and pooling layers is applied. A model is trained and then tested properly to validate its results. Proposed model has achieved 98.3% testing accuracy.85% of the data collected is used for training and 15% of the data is used for testing from PlantVillage dataset. These images show healthy and diseased plants. This research focuses on deep learning models to detect disease in plant leaves. But in future, these models can be integrated with drones or other systems to locate diseases living in plants in order to treat them accordingly.

A public dataset of 54,306 images of healthy and diseased plant leaves has been used to train a deep convolutional neural network to identify 14 crops and 26 diseases. An accuracy of 99.35% was achieved for this model on a held-out test set, showing the success of this approach. The general approach of training deep learning models on increasingly large and publicly accessible image datasets presents a path toward the mass deployment of smartphone-aided crop disease detection[2] Image processing and machine learning can be used to improve plant diseases detection techniques, thereby reducing the time, effort, and knowledge necessary for the detection of infected plants. It involves image acquisition, filtering, segmentation, feature extraction, and classification. This paper proposes a way to best detect disease by detecting its appearance from plant images and, if present, evaluating its type among Alternaria Alternata, Anthracnose, Bacterial Blight and Cercospora Leaf Spot. As the minimum accuracy is 95.774 percent and the maximum accuracy is 99.874 percent, this process gives almost accurate results. The process detects the diseases by the area of disease, although it has a low affected region[3] A neural network was trained on simple leaf images of healthy and diseased plants in this study using deep learning to detect and diagnose plant diseases. The models were trained on an open database of 87,848 images from 25 different plants in 58 distinct plantdisease combinations. The best performing model architecture had a success rate of 99.53% in indicating the corresponding plantdisease combinations (or healthy plants). Since the model has a high success rate, it is an excellent early warning tool that could be further developed to support the implementation of an integrated disease identification system in real-time[4]. A mathematical model is proposed that detects and recognizes plant diseases through deep learning, improving its accuracy, generality, and training efficiency. After recognizing leaves placed in complex surroundings, the region proposal network (RPN) is applied to extract symptom features from the pictures following Chan-Vese algorithm. The segmented images are then input into the transfer learning model with the training dataset of diseased leaves provided. Using three types of diseases (black rot, bacterial plaque, and rust), the model shows higher accuracy than the traditional method, thus reducing the influence of disease on production and making it more beneficial to sustainable agriculture. This paper presents a deep learning algorithm that is of great significance to intelligent agriculture, environmental protection, and agricultural production[5] The current shortcomings of current plant disease detection models are discussed. The new dataset contains 79,265 leaf images with the aim of being the largest dataset to contain leaf images. The images were taken in various weather conditions, under various lighting conditions and during daylight hours with an unreliable background resembling realistic scenarios. Traditional augmentation methods and state-of-the-art style generative adversarial networks were used to augment the number of images in the dataset. Tests were conducted to verify the effectiveness of training in a controlled setting and usage in the real world to accurately identify diseases of plants on natural and detection of multiple diseases in a single leaf. The trained model achieved an accuracy rate of 93.67%. Finally, a new two-stage architecture of a neural network was proposed for plant disease classification in a real environment[6] In this paper, a system was proposed for classifying three diseases affecting grapes– Anthracnose, Powdery Mildew and Downy Mildew and identifying the severity of these diseases using image processing and machine learning algorithms. U 900 images of disease infected grapes leaves were acquired by the farmers and field workers from the fields. Images of single leaf or bunch of leaves were captured with background from different distances and at different angles using mobile phone cameras with varying resolution starting from less than 1 megapixel to 13 megapixel. This proposed disease detection algorithm consists of 4 main stages: (a)Pre-processing of the input images, (b) Leaf extraction from the background, (c) Disease patch identification and (d)Background removal. Performance of four machine learning algorithms namely, PNN, BPNN, SVM and Random Forest are compared, for separating the background from disease patches and classifying between the different diseases. The performance of different texture features like local texture filters, Local Binary Patterns, GLCM features, and some statistical features in RGB plane for classification are also observed. It is observed that the proposed system achieves best classification accuracy of 86% using Random Forest and GLCM features. [7] In this paper, a real-time decision support system integrated with a camera sensor module is designed and developed for the identification of plant disease. The performance of three machine learning algorithms, Extreme Learning Machine (ELM) and Support Vector Machine (SVM) with linear and polynomial kernels was analyzed. A real-time decision support system using extreme learning machine was designed and developed using Raspberry PI hardware. Results demonstrated that the performance parameters, namely accuracy and sensitivity of the extreme learning machine, is 95% and is higher when compared to the other adopted classifiers. It is also observed that the developed real-time hardware with Extreme Learning Machine classifier is highly capable of detecting three different plant diseases and can be extended to detect many more plant diseases by training it with wide range of train datasets.

**4.An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques.**

A Proliferation of literature is available in plant leaf disease detection. We will highlight some of the key contributions. A methodology for detecting plant diseases early and accurately using diverse image processing techniques has been proposed by Anand H.Kulkarni , where Gabor filter has been used for feature extraction and ANN based classifier has been used for classification with recognition rate up to 91%. F. Argenti proposed a fast algorithm for calculating parameters of co-occurrence matrix by supervised learning and maximum likelihood method for fast classification. Homogenize techniques like sobel and canny filter has been used to identify the edges by P.Revathi . These extracted edge features have been used in classification to identify the disease spots. The proposed homogeneous pixel counting technique for cotton diseases detection (HPCCDD) algorithm has been used for categorizing the diseases. They claim the accuracy of 98.1% over existing algorithm. Tushar H Jaware proposed a novel and improved k-means clustering technique to solve low-level image segmentation. Spatial gray-level dependence matrices (SGDM) method has been used for extracting statistical texture features by Sanjay B. Dhaygude. RGB images have been converted into Hue Saturation Value (HSV) color space representation and showed the H, S and V components. Mokhled S. Al-Tarawneh presented an empirical investigation of olive leaf spot disease using auto-cropping segmentation and fuzzy c-means classification. Rgb to Lab colorspace and median filter used for image enhancement. At end present comparative assessment of fuzzy c-means and k-mean clustering. The fuzzy feature selection approach namely fuzzy curves (FC) and fuzzy surfaces (FS) have been proposed to select features of cotton leaf disease by Yan-Cheng Zhang. This has been resulted in reduced dimensional feature space. Back-propagation (BP) networks have been used to classify the grape and wheat diseases by Haiguang Wang. Also by using principal component analysis (PCA) dimensions of the feature data has been reduced. Texture features based on the local power spectrum of Gabor filters has been proposed by Simona E. Grigorescu where complex moments, Gabor energy and grating cell operator features have been discussed. They concluded that grating cell operator responded only to texture features. Detection of unhealthy region and classification using texture features has been proposed by S. Arivazhagan. Their algorithm has been tested on ten species of plants namely banana, beans, jackfruit, lemon, mango, potato, tomato and sapota. 94.74% accuracy has been achieved by Support vector machine (SVM) classifier. Dheeb Al Bashish, developed neural network classifier based on statistical classification and could successfully detect and classify the diseases with a precision of around 93%. A Research of maize disease image recognition of corn based on BP networks effectively identified by Song Kai where YCbCr color space technology is used to segment disease spot, Co-occurrence matrix (CCM) spatial gray level layer is used to extract disease spot texture feature, and BP neural network has been used to classify the maize disease. The applications of K-means clustering, BP neural networks had been formulated for clustering and classification of diseases that affect on plant leaves by H. Al-Hiary. They provide adequate support for accurate detection of leaf diseases. The proposed algorithm has been tested on five diseases viz. Early and late scorch, cottony and ashen mold , tiny whiteness. Menukaewjinda tried another ANN, i.e. back propagation neural network (BPNN) for efficient grape leaf color extraction with complex background. They also explore modified self organizing feature map (MSOFM) and genetic algorithm (GA) and found that these techniques provide automatic adjustment in parameters for grape leaf disease color extraction. Support vector machine (SVM) has been also found to be very promising to achieve efficient classification of leaf diseases. 21 color, 4 shape and 25 texture features has been extracted by Haiguang Wang and principal component analysis (PCA) has been performed for reducing dimensions in feature data processing, then back-propagation (BP) networks, radial basis function (RBF) neural networks, generalized regression networks (GRNNs) and probabilistic neural networks (PNNs) has been used as the classifiers to identify diseases.

## CHAPTER 4

**REQUIREMENTS SPECIFICATION**

## Functional Requirements

The system should be able to provide these functionalities efficiently.

* Resource Visualization: The visualizations should be self-explanatory which can be easily understood by the user. There will be line plots and graphs which can be used as an effective measure while devising any new program.
* ML algorithm should be able to predict the output efficiently and accurately.
* On exceeding the critical conditions, alert should be sent to the aqua farmers.
* Predict the disease in monthly manner. i.e., predict for next data based on past data.

## Non-Functional Requirements

Non-functional requirements are requirements that specifies criteria that can be used to judge the operation of a system rather than the behaviour.

* **Usability:** System has been made user friendly by including a readme file in the program so that any user facing difficulty can refer it and easily solve there problem.
* **Scalability**: If more parameters required, it can be added easily. Number of visualizations can be increased. Currently the system predicts for hourly manner this interval can be changed accordingly.
* **Reliability:** System should give reliable predicted results.

**Performance**: Our LSTM model will have improved performance because of the use of datasets with lowest time intervals and has high precession. For checking the accuracy, we have shown the performance metrics using RMSE.

**Documentation**: Coding standards are maintained throughout the project.

**Maintainability**: This project has easy maintainability of the web application, can be modifiable and integrated with advanced computational and operational technologies.

## Hardware Requirements

* System: Core i5 Processor
* Hard Disk :1 TB.
* GPU: for matrix calculation
* RAM: 8GB

## Software Requirement

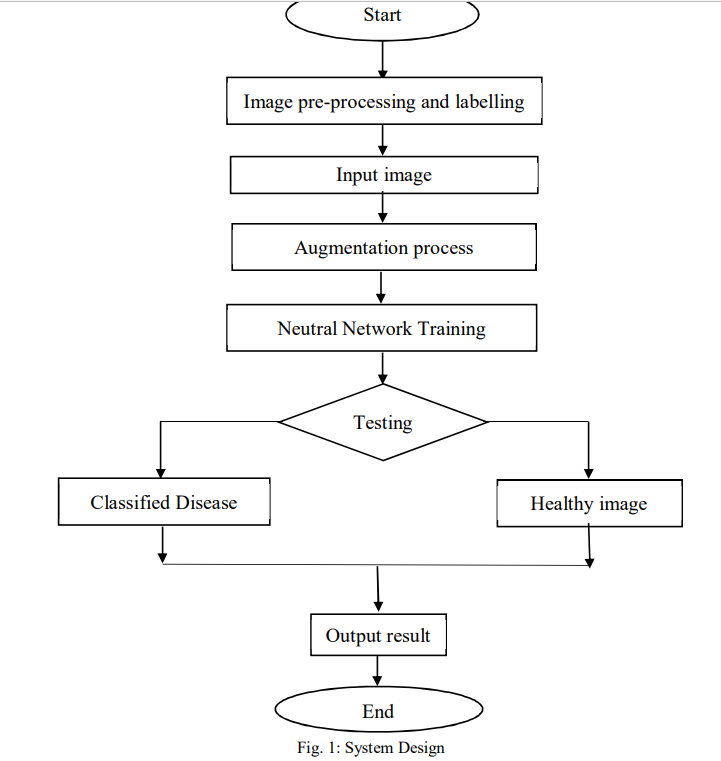
* + - IDE: Anaconda(with Tensor flow)
    - Programming language: Python
    - Library: Numpy, pandas, keras, Tensorflow

## Summary

This chapter gives an insight into the functional and non-functional requirements that the system provides. It also describes the hardware and software requirements that are required for building the system.

**PROPOSED METHODOLOGY**

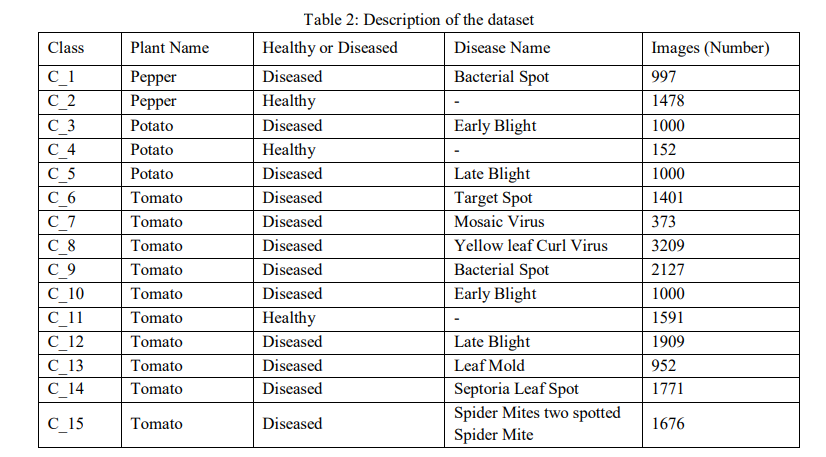
RCNN algorithms analyze an image and extract its features. Convolutional neural networks are deep learning algorithms that can process large datasets containing millions of parameters, modeled on 2D images, and connect the resulting representations to the corresponding outputs. A RCNN is a supervised multilayer network that can dynamically learn new features from datasets. In nearly all significant classification challenges, RCNNs have achieved state-of-the-art results recently. In the same architecture, they are also able to systematically isolate features and categorize them.

****

Collect images of plants with and without disease. A Python script calculated the training time by automatically resizing the images, which was calculated using the OpenCV framework. By augmenting the dataset and adding distortion to the images, overfitting can be reduced during the training period. The Deep Neural Network is trained on datasets of healthy and diseased crop leaves. It serves its purpose by classifying images of leaves into diseased or healthy categories based on their pattern of defect. As the leaves have texture and visual similarities, they are attributes for identifying disease types. Hence, computational vision applied to deep learning provides an efficient way to solve the problem.

**B. Dataset Description**

This dataset consists of 20,639 images of diseased and healthy plant leaves, which were classified into 15 classes to train a deep convolutional neural network which can identify the diseases.

****

C. Data Preprocessing

The dataset included images that were resized to minimize training time, which was calculated automatically by a Python script that uses the OpenCV framework. The input data is pre-processed by scaling the data points from [0, 255] (the minimum and maximum RGB values of the image) to [0, 1]. The dataset is divided into two parts, one for training and one for testing. 80% of the dataset is for training, and 20% for testing. A training dataset consists of 16,511 images and testing is made of 4,128 images. The training dataset is used to train the model while the testing dataset is kept unseen so that accuracy of the model can be tested.

D. Data Augmentation

Data augmentation is a technique for increasing the number of images in a database. Various operations such as shifting, rotating, zooming, and flipping are applied to image datasets to diversify our dataset. By augmenting the dataset and adding distortion to the images, overfitting can be reduced during the training period. The Keras Image Data Generator class implements in-place data augmentation or on-the-fly data augmentation. Through this type of augmentation of data, we can make sure that our network, when trained, sees new variations every time epoch. It allows us to come up with high results utilizing a smaller dataset.

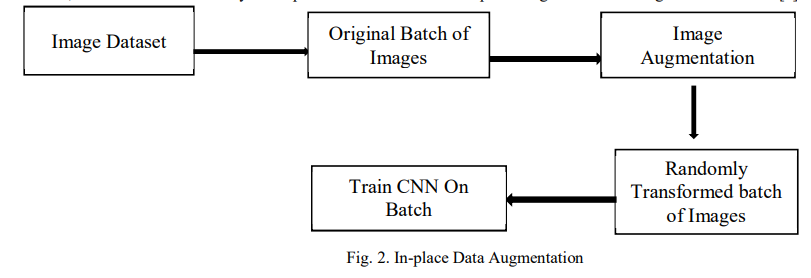
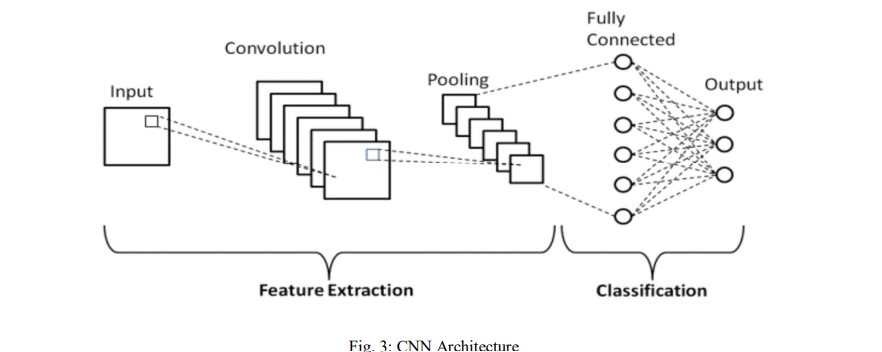
****

Figure 2 demonstrates the process of applying in-place data augmentation

1) Step 1: The ImageDataGenerator is presented with an input batch of images.

2) Step 2: Next, the ImageDataGenerator transforms each image into a random series of rotations, flipping , cropping etc.

3) Step 3: The randomly transformed batch is trained by using CNN. E. Architecture of Convolutional Neural Network CNN architecture is divided into two main parts : 1) A convolution tool that separates and categorizes the various features of images for analysis in what is called Feature Extraction. 2) Convolution is applied to the output of the fully connected layer and predicts the class of image based on the features extracted before.

****

## F. Convolution Layers

## The three layers that make up the CNN is the convolutional layer, pooling layer, and fully-connected layer (FC layer). When these layers are stacked, a CNN is formed. There are two additional layers to these three which are the dropout layer and the activation function. Convolutional Layer is the first layer that focuses on extracting features from the input images. In this layer, the convolution mathematical process is performed between each input image and a set of convolution filters of a particular size. By sliding the filter over the input image, a dot product is computed between the filter and the parts of the input image corresponding to the size of the filter. Feature maps represent the output. Later, the Feature map can be used to as input to other layers The Conv2D function takes the following arguments:

## 1) Filters - The number of different filter methods (feature detectors) that will be applied to the original image for creating the feature map. There are different types of filters, such as the Edge Detection Filter and Blur Filter.

## 2) Kernel Size - It gives the dimension of the (n x n) matrix of a convolution filter. 3) Activation - The activation function for the neurons. We use a Rectifier Linear Unit (Relu) function as an activation function for every layer besides the output layer. We have also added non-linearity to our network using ReLU. This is essential in identifying any linear relationships within the feature map.

## 4) Input Layer - It takes the shape of the Input Images and the number of channels (3 for color)

The proposed system architecture (Figure 4.1) shows the complete working of the system starting from training the model using the collected dataset to showing the predicted result and appropriate message on the web application.

G. Pooling Layer

In our convolutional neural network, the next layer is called the pooling layer. One of the main objectives of the pooling layer is to minimize the spatial dimension of the data propagating through the network. Pooling can be achieved in two different ways in convolutional neural networks. Max- pooling and average pooling. In Max Pooling which is the most common in two, for each section of image we scan the highest value. Average Pooling calculates the average of an image's elements within a predefined sized region. Pooling Layer serves as the bridge between Convolutional Layer and the Fully Connected Layer.

H. Fully Connected Layer

Fully Connected Layers (FC) consist of weights and biases as well as neurons, and they are used to connect neurons between different layers. In this layer, we flatten the output of the last convolutional layer and connect every node of the current layer with every other node of the next layer. This layer basically takes its input from the preceding layer, whether it is a convolutional layer, ReLU, or pooling layer. At this stage, the classification process begins.

I. **Dropout** When all the features are connected to the FC layer, it can lead to overfitting of the training dataset. A model is said to be overfitted if it can perform proficiently on training datasets but then shows negative performance when applied to new datasets. To solve this problem, a dropout layer is used wherein a few neurons are removed during the training process, thus reducing the size of the neural network model. On passing a dropout of 0.2, 20%of the nodes are removed randomly from the neural network.

J. **Activation** Activation functions plays a major role in the process of neural network. It determines what information from the model should be fired in the forward direction, and which information should not at the end of the network. Hence, it adds nonlinearity to the network. It has been observed that there are quite a few widely used activation functions. The most frequently utilised activation functions are Sigmoid, tanH, Softmax, and ReLU. Each activation function has its own specific application. For a multi class classification we generally use ReLU and Softmax functions.

1) **ReLU**: The rectified linear unit (ReLU) function is the most widely used activation function in today's networks. There is an advantage of using the ReLU function compared to the other activation functions in that it does not activate all the neurons at once. If the input is negative, then it is converted to 0, and the neuron is not activated. If the input is positive, it returns the positive value of x and the neurons get activated. Consequently, only a few neurons are activated at a time, making the network sparse and very efficient. The ReLU function also served as a significant advancement in the field of deep learning by overcoming the vanishing gradient problem. ReLU = max(0,x)

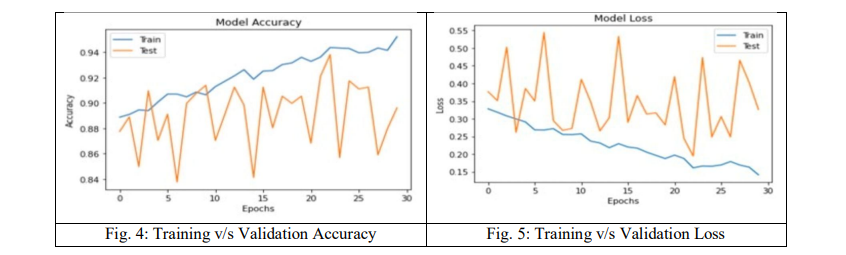
2) **Softmax**: The softmax function is ideally used in the output layer of the classifier where we are actually trying to get the probabilities to define the class of each input. As a result, it is easier for us to categorize data points and determine to which category they belong A convolutional neural network will be used to classify images without relying on pre-trained models. There are a number of popular pre-trained models available that can tell the difference between hundreds of classes without training each of them. These models have relatively complex architectures that help them handle hundreds of thousands of classes. The architecture can be difficult for a beginner to visualize. Keras make building of custom CNN’s easier. We developed this project using Custom CNN.

**K. Model** We now make use of Sequential model. Sequential model API is a way to build deep learning models in which sequential classes and model layers are created and added. The input to a convolutional neural network, is an (n x m x 3) for colored images, where the number 3 represents the red, green, and blue components of each pixel in the image. For this model, we first create a 2D convolutional layer with 32 filters of 3 x 3 kernels and a Rectified Linear Unit (ReLU) activation. In the following layers, we perform batch normalization which is used to scale data by a certain factor and pooling we use maximum pooling with a pooling size of two. Next, two blocks of 2D Convolutional layer are created with 64 filters and ReLU activation followed by a pooling layer. Finaly we add a layer of Convolutional layer with 32 filters followed by a layer of ReLU activation and pooling. Then we flatten the output from these layers so the data can proceed to fully connected layers. Flatten is used to convert data into a 1-Dimensional form. We add another 512 dense layers with a dropout of 0.2. Finally, we use the softmax activation function to convert the outputs into probability values.

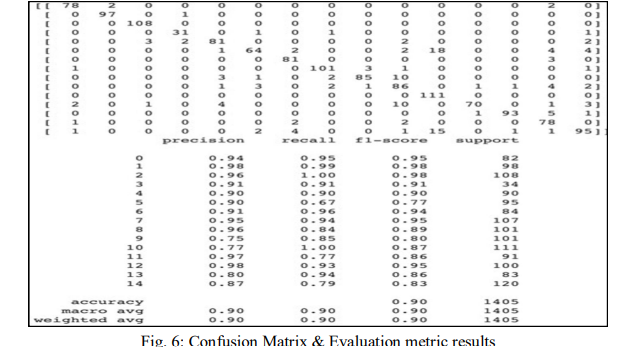
L. **Training** the Model The next step is to compile the model and then train it on a training dataset. The following parameters need to be specified for compiling the model: Optimizer - An optimizer is an algorithm or methodology used to reduce losses by modifying the weights and learning rate of a neural network. Optimizers train models faster and work more efficiently. As we have a multi class classification problem ,we use the Adam optimization technique because it always leads to a smoother way than other optimization techniques. Adam is an optimization algorithm that uses adaptive moment estimation to generate more efficient neural network weights[12]. loss → The cost function that calculates the difference between predicted and actual values. In our case we will be using “ sparse categorical\_cross entropy”. Sparse categorical cross entropy can be used for integer targets instead of categorical vectors[13] In order to fit the model, we have to specify the following parameters: batch\_size → Number of images to be used for training our CNN model before back-propagating the weights. epochs → An epoch is a measure of how many times the whole training set of images is used once.We train our model over 20 epochs and 30 epochs. A higher number of training epochs increases its accuracy along with lowering the loss.

**IV. RESULTS**

We plot a graph to illustrate the maximum accuracy the model achieved during training and validation while minimizing the loss.

****

**Confusion Metrix**



All the pairs both having disease and not having disease were plotted on a confusion matrix. A confusion matrix measures the degree of accuracy of a classification model with respect to each classification category. A trained model's evaluation and output is determined by True positives(TP), True negatives(TN), False positives(FP), False negatives(FN). For evaluation we also used F1, which combines both precision and recall in one term. The higher the F1-Score, the better the model. For all three metrics, models with 0 perform the worst while models with 1 perform the best[14]. Figure 5.3 displays the precision, recall, F1 and support for each class. The overall accuracy reported is 90%.

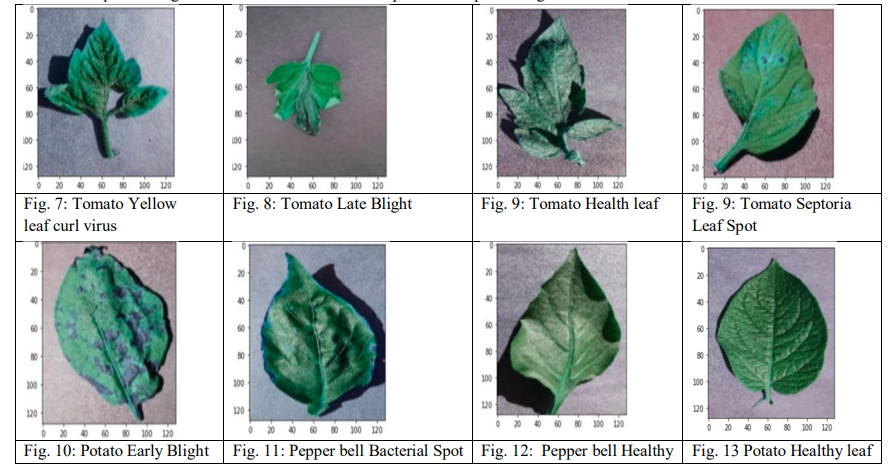
**1) Precision:** Precision describes all the positive classes correctly predicted by the model; how many of those are actually positive. The precision is calculated by taking the number of correctly classified positive examples divided by the number of predicted positive examples. The equation can be written



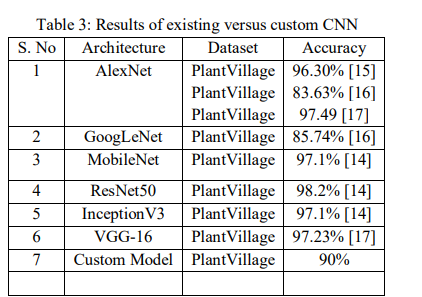
2) **Recall:** It defines how much the model predicted correctly among all positive classes. Recall is the ratio between the number of correctly classified positive examples and the total number of positive examples. The equation can be written as:



**C. Outputs Screenshots** A random sample of images is taken from the dataset and predicts the plant image's disease and class.



**D. Comparison of Results**

****

**CONCLUSION**

Even though there are various methods for detecting and classifying plant diseases using automatic or computer vision, research into this field has been lacking. In addition, there are few commercial options, with the exception of those focusing on the identification of plant species via photographs. Over the last few years, there has been tremendous progress in the performance of convolutional neural networks. The new generation of convolutional neural networks (CNNs) has shown promising results in the field of image recognition. A novel approach to automatically classifying and detecting plant diseases from leaf images was examined through this project utilizing deep learning techniques. With an accuracy of 90%, the developed model could distinguish healthy leaves from eight diseases that could be observed visually. On the basis of this high level of performance, it becomes apparent that convolutional neural networks are highly suitable for automatic diagnosis and detection of plants.

**FUTURE SCOPE**

The main goal for the future project is to develop a complete system comprising a trained model on the server, as well as an application for mobile phones that display recognized diseases in fruits, vegetables, and other plants based on photographs taken from the phone camera. This application will aid farmers by facilitating the recognition and treatment of plant diseases in a timely manner and help them make informed decisions when utilizing chemical pesticides[6]. Also, future work will involve spreading the use of the model across a wider land area by training it to detect plant diseases on aerial photos from orchards and vineyards captured with drones, in addition to convolution neural networks for object detection. Drones and other autonomous vehicles, such as smartphones, to be used for real-time monitoring and dynamic disease detection in large-scale open-field cultivations. A future possibility for agronomists working at remote locations could be the development of an automated pesticide prescription system that would require the approval of an automated disease diagnosis system to allow the farmers to purchase appropriate pesticides. Thus, the uncontrolled acquisition of pesticides could be severely restricted, resulting in their excessive use and misuse, with their potentially catastrophic effects on the environment.

# REFERENCES

[1] Sankaran, S.; Mishra, A.; Ehsani, R. “A review of advanced techniques for detecting plant diseases”. Comput.Electron. Agric. 2010, 72, 1–13. [CrossRef]

[2] Chaerani, R.; Voorrips, R.E. “Tomato early blight (Alternariasolani): The pathogens, genetics, and breeding for resistance”, J. Gen. Plant Pathol. 2006, 72, 335–347. [CrossRef]

[3] Alvarez, A.M.” Integrated approaches for detection of plant pathogenic bacteria and diagnosis of bacterial diseases”, Annu. Rev. Phytopathol. 2004, 42, 339–366. [CrossRef] [PubMed]

[4] Gutierrez-Aguirre, I.; Mehle, N.; Delic, D.; Gruden, K.; Mumford, R.; Ravnikar, M. “Real-time quantitative PCR based sensitive detection and genotype discrimination of Pepino mosaic virus”,. J. Virol. Methods 2009, 162, 46–55. [CrossRef] [PubMed]

[5] Martinelli, F.; Scalenghe, R.; Davino, S.; Panno, S.; Scuderi, G.; Ruisi, P.; Villa, P.; Stropiana, D.; Boschetti, M.; Goudart, L.; et al. “Advanced methods of plant disease detection. A review” ,Agron.Sust. Dev. 2015, 35, 1–25. [CrossRef]

[6] Bock, C.H.; Poole, G.H.; Parker, P.E.; Gottwald, T.R. “Plant Disease Sensitivity Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging”. Crit. Rev. Plant Sci. 2007, 26, 59–107.

[7] Down To Earth's print edition dated April 1-15, 2019

[8] Lefeuvre, Pierre, Darren P. Martin, Gordon Harkins, Philippe Lemey, Alistair JA Gray, Sandra Meredith, Francisco Lakay et al. "The spread of tomato yellow leaf curl virus from the Middle East to the world." PLoS pathogens 6, no. 10 (2010).

[9] India Agriculture GDP report from Ministry of Statistics and Programme Implementation (MOSPI)

[10] 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.

[11] J. D. Pujari, R. Yakkundimath and A. S. Byadgi, “Identification and classification of fungal\

[12] 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.

[13] R. Hari Shankar, A. K. Veeraraghavan, Uvais, K. Sivaraman and S. S. Ramachandran, "Application of UAV for Pest, Weeds and Disease Detection using Open Computer Vision," 2018

[14] International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India,2018, pp. 287-292.

[15] Kumar, Meeta, et al. “Survey on techniques for plant leaf classification.” International Journal of Modern Engineering Research 1.2 (2011): 538-544.

[16] Fuentes, Alvaro & Lee, Yujeong& Hong, Youngki& Yoon, Sook& Park, Dong. (2016). Characteristics of Tomato Plant Diseases - A study for tomato plant disease identification. Babatunde, Oluleye H., et al. "A genetic algorithm-based feature selection." (2014).

[17] F. Y. A. Rahman, S. R. M. S. Baki, A. I. M. Yassin, N. M. Tahir and W. I. W. Ishak, "Monitoring of Watermelon Ripeness Based on Fuzzy Logic," 2009 WRI World Congress on Computer Science and Information Engineering, Los Angeles, CA, 2009, pp. 67-70.

[18] P. Pawar, V. Turkar and P. Patil, "Cucumber disease detection using artificial neural network," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, 2016, pp. 1-5.

[19] Zhang, Harry. "The optimality of naive Bayes." AA 1.2 (2004).

[20] 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 ICCPET 2020 Journal of Physics: Conference Series 1712 (2020) 012012 IOP Publishing doi:10.1088/1742-6596/1712/1/012012 13 and Information Communication Technology (ICEEICT), Dhaka, 2016, pp. 1-6.

[21] 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.

[22] Meenakshi K, Safa.M, Krthick T, Sivaranjani N.” A novel study of machine learning algorithms for classifying health care data”, Research Journal of Pharmacy and Technolgy, Research Journal of Phamacy and Technology, 2017

[23] Meenakshi K, Maragatham, G, Agarwal, N., Ghosh, I, “ A Data mining Technique for analysing and predicting the success of movie”, Journal of Physics: Conference Series, Vol 1000, Issue 1, 2018

[24] Saranya, G., & Pravin, A. (2020). A comprehensive study on disease risk predictions in machine learning. International Journal of Electrical and Computer Engineering (IJECE), 10(4), 4217.

[25] Saranya, G., Geetha, G., & Safa, M. (2017). E-antenatal assistance care using decision tree analytics and cluster analytics based supervised machine learning. 2017 International Conference on IoT and Application (ICIOT)

[26] G Geetha, M Safa, C Fancy, D Saranya “A hybrid approach using collaborative filtering and content based filtering for recommender system”, Journal of Physics: Conference Series Vol 1000, Issue 1, 2018.

[27] M Srivastava, S Pallavi, S Chandra, G Geetha,”Comparison of optimizers implemented in Generative Adversarial Network (GAN)” International Journal of Pure and Applied Mathematics Volume 119 Issue 12,2018