PLANT DISEASE DETECTION USING DEEPLEARNING TECHNIQUES

Dissertation submitted in fulfillment of the requirements for the Degree of

MASTER OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

By

ADITI SINGH

Registration number: 11604614

Supervisor

HARJEET KAUR

School of Computer Science and Engineering (14 Bold)

Lovely Professional University

Phagwara, Punjab (India)

Month: April Year: 2021

i

TOPIC APPROVAL PERFORMA

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education, Transforming India

School of Computer Science and Engineering (SCSE)

Program : P192-ND::Integrated B.Tech. - M.Tech. (Computer Science & Engineering)

COURSE CODE : CSE549   **REGULAR/BACKLOG :** Regular   **GROUP NUMBER :** CSERGD0468

Supervisor Name : Harjeet Kaur   **UID :** 12427   **Designation :** Assistant Professor

Qualification : _____   **Research Experience :** _____

SR.NO.	NAME OF STUDENT	Prov. Regd. No.	BATCH	SECTION	CONTACT NUMBER
1	Aditi Singh	11604614	2016	K1637	9779421432

SPECIALIZATION AREA : Programming-I   **Supervisor Signature:** _____

PROPOSED TOPIC : Plant Disease Detection using Techniques of Deep Learning

Qualitative Assessment of Proposed Topic by PAC		
Sr.No.	Parameter	Rating (out of 10)
1	Project Novelty: Potential of the project to create new knowledge	8.23
2	Project Feasibility: Project can be timely carried out in-house with low-cost and available resources in the University by the students.	8.00
3	Project Academic Inputs: Project topic is relevant and makes extensive use of academic inputs in UG program and serves as a culminating effort for core study area of the degree program.	7.14
4	Project Supervision: Project supervisor's is technically competent to guide students, resolve any issues, and impart necessary skills.	8.77
5	Social Applicability: Project work intends to solve a practical problem.	8.23
6	Future Scope: Project has potential to become basis of future research work, publication or patent.	8.00

PAC Committee Members		
PAC Member (HOD/Chairperson) Name: Sawal Tandon	UID: 14770	Recommended (Y/N): Yes
PAC Member (Allied) Name: Ravi Kant Sahu	UID: 16920	Recommended (Y/N): Yes
PAC Member 3 Name: Virrat Devaser	UID: 14591	Recommended (Y/N): Yes

Final Topic Approved by PAC: Plant Disease Detection using Techniques of Deep Learning

Overall Remarks: Approved

PAC CHAIRPERSON Name: 11024::Amandeep Nagpal   **Approval Date:** 29 Apr 2020

ICCRDA-2020

1st INTERNATIONAL CONFERENCE ON COMPUTATIONAL RESEARCH AND DATA ANALYTICS (ICCRDA-2020)

Certificate

This is to certify that **Prof./Dr./Mr./Ms. Aditi Singh** is a presenter/co-author of the paper title Potato Plant Leaves Disease Detection and Classification using Machine Learning Methodologies presented in the 1st International Conference on Computational Research and Data Analytics (ICCRDA-2020), jointly organized by Suleyman Demirel University and Isparta Applied Sciences University, Turkey and College of Engineering Roorkee, India on 24th October, 2020.

Prof (Dr.) Tuncay Yigit
General Chair
Suleyman Demirel University,
Turkey

Prof (Dr.) B.M. Singh
General Chair
Director, College of Engineering Roorkee,
India

iii

ICICC-2021

INTERNATIONAL CONFERENCE ON INNOVATIVE COMPUTING AND COMMUNICATION (ICICC-2021)

Certificate

This is to certify that **Prof. / Dr./ Mr. / Ms. Aditi Singh** is a presenter/co-author of the paper titled **Comparative study on Identification and classification of Plant Diseases with the support of transfer learning** presented in the **4ᵗʰ International Conference on Innovative Computing and Communication (ICICC-2021)**, organized by Shaheed Sukhdev College of Business Studies, University of Delhi, New Delhi, India in association with National Institute of Technology Patna, India and University of Valladolid, Spain on **20ᵗʰ-21ˢᵗ February, 2021**.

Poonam Verma
Principal, SSCBS,
University of Delhi, New Delhi

Prabhat Kumar
General Chair
National Institute of Technology, Patna

Siddhartha Bhattacharyya
General Chair
CHRIST University, Bangalore

ABSTRACT

Agriculture is one of the essential sectors for the survival of humankind. At the same time, digitalization touching across all the fields that became easier to handle various difficult tasks. Adapting technology as well as digitalization is very crucial for the field of agriculture to benefit the farmer as well as the consumer. Due to adopting technology and regular monitoring, one can able to identify the diseases at the very initial stages and those can be eradicated to obtain a better yield of the crop.

Crop growth and yield are essential aspects that influence the field of agriculture as well as farmer economically, socially, and in every possible way. So, it is necessary to have close monitoring at various stages of crop growth to identify the diseases at right time. But, humans naked may not be sufficient and sometimes it would be misleading scenarios arise. In this aspect, automatic recognition and classification of various diseases of a specific crop are necessary for accurate identification. This thought gave inspiration for the present proposed framework. The proposed framework mainly concentrated on the transfer learning phenomenon based on three different pre-trained models such as VGG-16, ResNet-50, and ResNet-50 v2, and then compared the three models based on transfer learning models based on various standard evaluation metrics. VGG-16 based transfer learning model achieved an accuracy of 98.74%, ResNet-50 based transfer learning model achieved an accuracy of 98.84%, and ResNet-50 v2 based transfer learning model achieved an accuracy of 98.21%. The dataset is considered for the implementation is the "PlantVillage" dataset which includes the various diseased and healthy leaves of Pepper, Potato, and Tomato and it is an openly available dataset through Kaggle.

DECLARATION STATEMENT

I hereby declare that the research work reported in the dissertation/dissertation proposal entitled "PLANT DISEASE DETETCTION" in partial fulfilment of the requirement for the award of Degree for Master of Technology in Computer Science and Engineering at Lovely Professional University, Phagwara, Punjab is an authentic work carried out under supervision of my research supervisor Mr./Mrs. Research Guide's Name. I have not submitted this work elsewhere for any degree or diploma.

I understand that the work presented herewith is in direct compliance with Lovely Professional University's Policy on plagiarism, intellectual property rights, and highest standards of moral and ethical conduct. Therefore, to the best of my knowledge, the content of this dissertation represents authentic and honest research effort conducted, in its entirety, by me. I am fully responsible for the contents of my dissertation work.

Signature of Candidate

Name of the Candidate: Aditi Singh

R.No: 11604614

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the M.Tech Dissertation/dissertation proposal entitled **"PLANT DISEASE DETECTION SYSTEM"**, submitted by **Aditi Singh** at **Lovely Professional University, Phagwara, India** is a bonafide record of his / her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree.

Signature of Supervisor

(Harjeet Kaur)
Date:

Counter Signed by:

1) **Concerned HOD:**
   HoD's Signature: _____

   HoD Name: _____

   Date: _____

2) **Neutral Examiners:**

   External Examiner

   Signature: _____

   Name: _____

   Affiliation: _____

   Date: _____

   Internal Examiner

   Signature: _____

   Name: _____

   Date: _____

ACKNOWLEDGEMENT

I have taken efforts in this project. However, it would not have been possible without the kind support and help of my mentor Mrs. Hrajeet Kaur. I would like to extend my sincere thanks to all of them.

I am highly indebted to her for her guidance and constant supervision as well as for providing necessary information regarding the project & also for her support in completing the project.

TABLE OF CONTENTS

CONTENTS   **PAGE NO.**

TABLE OF CONTENTS

CONTENTS	PAGE NO.

LIST OF TABLES

LIST OF FIGURES (16 BOLD)

CHAPTER 1

INTRODUCTION

Crop growth and yield are essential aspects that influence the field of agriculture as well as farmer economically, socially, and in every possible way. So, it is necessary to have close monitoring at various stages of crop growth to identify the diseases at right time. But, humans naked may not be sufficient and sometimes it would be misleading scenarios arise. In this aspect, automatic recognition and classification of various diseases of a specific crop are necessary for accurate identification. In this chapter, I will provide a tour d'horizon for the proposed methodology of my research work. This chapter consists of the background, problem statement, objectives, and scope of the proposed methodology.

One of the severe causes of increased microbial infection is illiteracy among farmers in India. Once a crop is infected by some disease it is difficult for farmers to find out the real cause of diseases. pathogens and pests are affecting the crops badly. The crop produces for 5 main food crops by 10 -40%, This data is in reference to a study report which is published by UC Agriculture and Natural Resources.

In the Indian context where agriculture contributes to 16% of GDP and engages almost 60% of the population requires great measures to be taken to avoid plant diseases. According to the Ministry of Food Processing Industries in the year 2016 agricultural loss was 13 billion dollars. One of the helpful measures in plant disease detection methods can be done with the help of image processing and neural network. Neural network and deep learning in some recent research have proved its worth in doing such classification tasks efficiently.

Agriculture is an essential sector in countries like India as those countries' economy directly or indirectly dependent on agriculture. It indicates the necessity of taking care of plants from seedling until the expected crop obtains. Through this process, the crop needs to cross a lot of phases to obtain the expected crop such as weather conditions, the survival of the crop from various diseases, and the survival of the crop from various animals. Of these major phases, the crops can be protected from the various animals by providing proper protection for the field and this issue can be solvable. The next major issue is weather conditions which will not be in the control of humans, humans can only pray for better weather conditions to obtain a better crop. Finally, The major issue is very crucial to protect the crop from various diseases as these diseases can impact the complete growth and yield of the crop. If one can able to identify these diseases in time, then the crop can be protected using appropriate fertilizers. If this process of identification and classification of diseases able to digitalize which would be helpful for the agriculturists. It will decrease the time for the identification of diseases and precision in classifying the diseases.

1.1 BACKGROUND:

Agriculture is the oldest profession. Humans started cultivation even science there was no civilization. With the development of science, plants were identified as living-thing. It could also respire, reproduce, and even get prone to various diseases. These are different types of diseases by various microorganisms may it be bacteria, viruses, or fungi. Plant diseases can damage crops to a great extent. It can even be fatal to human beings. One such situation emerged in 1840 when a large amount of potato crop was destroyed due to a disease called the Late blight of potato. This is also known as the Irish famine and this was a darker phase of European history, where people were dying of hunger. We all know that plants are very important in our lives so we need to protect them not only from deforestation but also from various plant diseases. In India still, a large population is engaged in agriculture. It ranks 2nd in the world in terms of agriculture production This is again a major challenge. Crops are damaged in India because of various factors one of the major causes is natural calamities and other is the microbial diseases. According to United Nations data 96-billion-dollar loss was in agriculture in a decade (starting from 2005). As humans do not have any control over natural disasters but we can control microbial infections in plants.

Once a crop is infected by some disease it is difficult for farmers to find out the real cause of the disease. pathogens and pests are affecting the crops badly. The crop produces for 5 main food crops by 10 -40%, This data is about a study report which is published by UC Agriculture and Natural Resources.

In the Indian context where agriculture contributes to 16% of GDP and engages almost 60% of the population requires great measures to be taken to avoid plant diseases. According to the Ministry of Food Processing Industries in the year 2016 agricultural loss was 13 billion dollars. One of the helpful measures in plant disease detection methods can be done with the help of image processing and neural network. Neural networks and deep learning in some recent research have proved their worth in doing such classification tasks efficiently.

1.2 PROBLEM STATEMENT

The proposed research work is carried on three crops: Potato, tomato, and pepper. In a survey, 61.33% of farmers cultivating potatoes reported bright as one of the major reasons for their crop failure. According to The Hindu, in the year 2020, Maharashtra faced a huge crisis for tomato crops approx. 60% of crops failed due to a virus. India has a 40% share of total pepper production globally. Pepper has also many herbal benefits in throat infections. India is the 2nd largest producer of tomatoes and potatoes followed by China which is the largest producer of these crops. Improving fertilization and automating the system for disease detection can improve the crop production of our country.

1.3 OBJECTIVES

There are three objectives of the proposed methodology:
  i. To develop a prototype for a plant disease detection system.
  ii. To apply image processing techniques to identify the disease pattern.
  iii. Use machine learning algorithms to predict disease.
  iv. Use transfer learning techniques to predict disease.

1.4 SCOPE

The proposed methodology is used for the precise detection of disease in crops. Which can provide controlled fertilization to farmers. Accurate identification of disease also helps farmers to identify the infection and do relatively controlled fertilization to avoid any future crop failures.

CHAPTER-2: REVIEWS OF LITERATURE

Sardogan, M., et al. in 2018 [1] presented a model with a combination of convolutional neural networks (CNN) along with learning vector quantization(LVQ) for the identification and categorization of diseases of tomato plant leaves. The presented framework was implemented on the data size of 500 images with the four categories of diseases considered for tomato plant leaves. The convolutional neural network is utilized for the extraction of vital attributes from the images as well as for the classification. Wallelign, S., et al. in 2018 [2] discussed the viability of convolutional neural network architecture for the classification of various plant diseases with the aid of leaf images. The mentioned framework is implemented by utilizing the LeNet, one of the popular CNN architecture, for disease classification in the aspect of soybean plants. The soybean plant leaf images of 12,763 samples are obtained from the standard database called PlantVillage. The mentioned framework able to achieve an accuracy of 99.32% indicating the viability of CNN with plant disease classification utilizing the leaf images. Sladojevic, S., et al. in 2016 [3] concerned the generation of the new-age model for the identification of various diseases of 13 plant diseases out of the healthier plant leaf images. The deep learning architecture called Caffe was utilized for training the data. The results were obtained from the mentioned framework with a precision of 91percent to 98percent. Fuentes, A., et al. in 2017 [4] proposed a framework and can be applied in two stages. At first, the meta architectures of Faster R-CNN, R-FCN, and SSD will be combined to form a single meta-architecture. Lastly, certain methodologies such as VGG-16, VGG-19, and ResNet-50 will be attached to extract the features from more depth and these models' efficiency was estimated. When compared to many other models, the proposed framework efficiency is better. Arivazhagan, S. and Ligi, S. V. in 2018 [5] proposed a framework based on automated deep learning for the recognition and classification of various diseases in mango plants. The dataset utilized for this framework consists of 1200 images which include both diseased and healthy leaves of mango. The accuracy obtained from the proposed framework is 96.67%. Oppenheim,

D. and Shani G. in 2017 [6] proposed a framework based on convolutional neural network architecture for the recognition and classification of various diseases in potato plants. The dataset

utilized for this framework consists of 2465 potato images. Barbedo, J. G. A. in 2018 [7] investigated and identified the pros and cons through various factors that affect the model and efficiency of deep learning neural networks which are used for the recognition as well as the classification of various plant diseases. The investigation carried out on the literature as well as the experiments carried out with the image database consists of 50000 images of various plant diseases. Brahimi, M., et al. in 2017 [8] proposed a framework based on a convolutional neural network for the detection and classification of various diseases in the tomato crop. The dataset utilized for this framework consists of 14,828 tomato leaf images with almost nine diseases from the plant village image database. The proposed framework able to achieve an accuracy of 99.18%. Shrivastava, V. K., et al. in 2019 [9]focused on the detection and classification of various diseases in the rice plants using a framework with the aid of CNN architecture along with SVM. The framework was implemented on the dataset consists of 619 rice plant leaf images with all four categories of diseases. The accuracies are evaluated for various proportions of training and testing datasets and the maximum accuracy achieved is 91.37%. Ozguven, M. M. and Adem, K. in 2019 [10] updated an existing faster region-based CNN architecture by varying the parameters for the identification of disease-affected regions in the case of sugar beet. The dataset consists of 155 sugar beet images and an accuracy rate of 95.48% achieved using the proposed framework.

Uguz, S. and Uysal, N. in 2020 [11] considered a comparison of a transfer learning scenario with CNN architectures such as VGG-16 and VGG-19 along with proposed CNN architectures in case of Olive plant diseases. The framework implemented on the dataset consists of 3400 Olive plant leaf images. In this framework, a data augmentation methodology was implemented for improving the size of the dataset. Before data augmentation, the accuracy attained about 88% and after data augmentation, the accuracy attained about 95%. Agarwal, M. et al. 2020 [12] proposed a customized model based on CNN identification of disease of tomato leaves. Also, compared the proposed model with machine learning models and VGG-16. The proposed model attained an accuracy of 98.4%, the KNN model attained an accuracy of 94.9%, and the VGG-16 model attained an accuracy of 93.5%. The tomato leaf images dataset utilized for this framework is extracted from the Plant village dataset. Wang, J. et al. in 2018 [13] considered a transfer learning scenario based on CNN architecture for detection and classification of diseases with the aid of leaf images of 2 crops such as cucumber and rice. The proposed framework was implemented on 2430 images of cucumber as well as rice with eight diseases extracted from the plant village dataset.

The proposed framework attained an accuracy of 90.84%. Toda, Y., and Okura, F. in 2019 [14] reviewed the scenario of deep learning methodology impact the diagnosis of plant diseases utilizing the leaf images. CNN architecture works as a black box model for the diagnosis of diseases of the plant. It is also discussed the various aspects of hyperparameters that affect classification accuracy.

So far the various models and research have been identified in terms of identification as well as the classification of various categorical diseases in the specific plant using deep learning scenarios. Deep learning can also be utilized for the identification and classification of macro-nutrients in a specific plant. For instance, Tran, T. T. et al. in 2019 [15] proposed a system based on a deep learning scenario providing a monitoring system that monitors across various stages from the seedling stage to the yielding stage to achieve an enhanced rate of yield. The proposed framework was implemented utilizing a dataset consists of 571 images include tomato leaf images and tomato fruit images of various stages of growth of the crop. The inception-ResNet v2 and autoencoder attained the accuracies of 87.27% and 79.09% respectively. This literature work shows that the effect of transfer learning on the detection and classification of plant diseases through leaf images.

According to Ehsan Kiani et.al, 2017 [16] image segmentation done with the help of colors i.e. color image segmentation techniques helps to better understand and solve the problem. One can first find out the three-color image components of an image which are Red, Green, and Blue components. The red and Green components help to identify the yellow components of the image which is usually marked as an infected part. Fuzzy logic is a good technique to solve a disease classification problem.

Vijai Singh et.al [2017] An advancement of genetic algorithm is proposed by the author named minimum distance algorithm to find the infected plant part of the plant that is to perform image segmentation. After the image segmentation step the author has checked the accuracy of the algorithm with other classification algorithms like k mean clustering and SVM [17].

Konstantinos 2018[18] In this paper author has used a convolution neural network technique to identify various plant diseases. A detailed study has been done by the author. Images of various plant leaves are taken which includes both the infected leaves images and healthy leaves images and then the author has classified it in various classes and all CNN architectures gave more than

7

97% accuracy. The CNN architectures include AlexNet, AlexNetOWTBn, GoogLeNet, Overfeat, VGG.

Kamlesh Golhani et.al, 2018 In this paper author has done a detailed review of various deep learning algorithms along with their advantages and disadvantages also their optimization techniques. A comparison has also been made for these techniques about the related work.[19]

Channamallikarjuna et.al, [2018]: In this paper, the author has explained in detail the algorithm proposed. The 1st step was image acquisition followed by image enhancement and then image segmentation. Color image segmentation was done with the help of the HSV technique. The tool made for testing plant disease was integrated with sensors that could also find out the real-time atmospheric and climatic conditions which could affect the plant disease.[20]

X.E. Pantazi et.al, 2019 In this paper author provides a detailed study of each step of the algorithm. Grabcut algorithm for image segmentation. LBP algorithm for thresholding. One Class Classification method and its hybridization with the SVM algorithm for conflict resolution.[21]

Sindhuja Sankaran et.al, 2010, This paper shows one of the early works in plant disease detection technology. This paper does a comparative analysis of disease prediction in plants with the help of molecular techniques, imaging and spectroscopic techniques, and also volatile organic compound profile study.[22]

Anjna et.al [2020]: In this paper, the author has explained the algorithm he used for the detection of plant disease. The image segmentation is performed with the help of k mean classification and has given a very satisfactory result or the process. Feature extraction for classification was done with the help of a grayscale co-occurrence matrix. Then the author has compared various classification techniques and has performed their comparison. The tree can also categorize data efficiently into classes. Then SVM using its kernel function is also a good approach and K nearest neighbor is also a good algorithm. Out of these algorithms, KNN and SVM had proved their worth. [23]

Monzurul Islam et al. 2017[24] proposed an automated system for potato disease detection. The author in this work used plant Village dataset for his work. SVM algorithm was used to predict diseases. An accuracy of 95% was achieved in classing the three potato disease which are late bright, early bright and healthy leaves. The use of SVM algorithm followed by image segmentation helped the author to achieve such a high accuracy. The approach would provide the agrigulturists with reliable, and precise toolkit which could save time in identifying the disease and doing classification.

Harshal Waghmare and Radha Kokare 2016[25] the author in this paper proposed a framework where one could utilize leaf textures for identification of diseases in plants. A coloured image is provided as input to system which is segmented to find the infected region in the image and special section of leaf is obtained. On the basis of features a model based on texture is derieved. The texture of leaf is very much unique of every new category of the leaf disease. This is the data which SVM classifier process to identify the disease. The author in this work used multiclass SVM classifier to categorize and identify the disease in grapes plant leaf images. The image pattern is then classified as an SVM designation for multiclass in groups who are healthy or ill classes respectively. The proposed study mainly concentrates commonly and worst affecting disease which are downy mildew disease & black red. The recommendation system proposed in this work quickly provides tenants with expert guidance with 96.6 percent of accuracy.

Shima Ramesh and Mr. Ramachandra Hebbar et al. 2018[26] proposes techniques for leaf-based disease detection and approaches which have efficient findings. Random forest is the algorithm which included in this proposed work for the detection of healthy or infected leaves images for creating dataset. The proposed work contains phases: Identifying dataset, feature extraction from leaf-images, Identifying from dataset, extraction of function, classifier, & classification. The datasets created infected and stable or healthy leaves & are trained collectively under Random forest algorithm & grouping of infected and healthy videos. Extracting useful properties from images the histogram-oriented gradient(HOG) has been utilized. In common, use of MLto train huge available datasets provides us with a simple and effective technique to detect the occurring various diseases in plants. The model in proposed work is trained with Random Forest Classifier algorithm 160 papaya leaves images. The model can be used to categorize an approximate 70%

percent accuracy. Accuracy of model can be enhanced with large number of images and certain other local features & global

Mrs. Shruthi U et al. 2019 [27] the proposed work on machine learning algorithms to classify diseases & it applies on mainly data & it prioritizes automatically the accuracy of all those tasks. This proposed work demonstrates the different phases where general identification of common plant diseases and use of machine learning algorithms on comparative research work which is for plant disease identification and classification such as: acquisition of image dataset, processing and cleaning the obtained data which is present in the image dataset, the segmentation of image data in the image dataset, extraction of essential features, and classification of input image in image dataset is based on features extracted and patterns which are identified. Proposed work demonstrate the Convolutional Neural Network(CNN) has high accuracy when more number of diseases are there in.

Rajleen Kaur and Dr. Sandeep Singh Kang 2015[28] proposed a framework for automatic plant disease detection and classification and the infected part of the input plant leaf images & also crop from agricultural production. It is achieved with the computer advancement and technology which allows agriculturists to enhance production. SVM algorithm is recent classifier in the neural network approach & problem which involves detection and accuracy in classification. SVM algorithm which is introduced in proposed work includes two types of datasets: 1)training data. Original input image is taken and is then used processed. It usually uses input image pixels in black and have a background, and the hue section is separated with saturation Third is disease diagnosis where unhealthy components are identified & stable components is segmented. This approach also provides % of region in leaf image where diseases is infected and will give classified disease name.

Pooja V et al. 2017[29] proposed a framework where machine learning techniques are utilized for classification & disease detection & uses tools for image processing. 1) It will capture the damaged or infected region in the leaf image and 2) it will perform image processing. Segments in which images are derived & it identifies region where extraction of leaf features is performed. At final step, SVM algorithm classified results is sent, and new results are derieved. Leaf disease classification performed by a support vector machine(SVM), methodology & provides the best possible results than techniques used previously.

Jitesh P. shah et al. 2016 [30] proposed a technique for examination of infected leaves image processing techniques & ML algorithm for identification of affected plant leaves using images.The author not only inspected, but also had briefly discussed key concepts of digital image processing & machine learning for identification and detection of infected plant diseases.This proposed work was implemented to the depth where the author considered 19 experiments on various categories of diseases. This work is also utilized in the survey which was conducted for future research and upgrade of system. This work includes classification of various diseases related to paddy plants. Mckay, et.al 2008[31] the author has divided the whole algorithm into three parts. First, the acquired image from any source needs to be denoised. Then the wavelet analysis of the image is performed and the last step is neighborhood pixel analysis. In this step, the analysis of pixels is done that is pixel comparison with neighbor pixel is done so as to check the accuracy of image enhancement. This is done to check whether in the image smoothening process some important data is lost or not.

Zahid Ullah et.al, 2020 [32]: In this paper author has explained a detailed methodology for image enhancement. There are various scenarios where some amount of useful information about the image is damaged, distorted, or lost. So, the first author has advised us to used filtering. A median filter is passed through the image to remove some amount of noise. A good contrast of an image is required for good analysis. Good contrast makes all image object visible. In this paper, the author has suggested a very effective methodology by limiting contrast in histogram equitization. This method is better than the histogram equalization method because it removes the over-amplification issue. And then it is good to do the wavelet transform process. 2D discrete wavelet transform is an efficient tool to do image enhancement.

In the paper Moumita Chanda et.al 2019[33]: the author has efficiently shown the use of the backpropagation technique to find the effective weight of the network. [19]The local gradient of any neuron is the product of the corresponding error signal of that neuron and derivative of the associated activation function.

An effective method for multi-dimension data. The input space is transformed into a higher dimensional feature space with the help of the kernel function. Some types of SVM include: Radial basis, two-layer perceptron and all have different kernel functions. Sachin D. Khirade et.al 2015[34]: The author has shown a detailed methodology of disease identification and detection with help of SVM.

S. Hern´andez et.al, 2020 [35] In this paper the author had raised a major concern about uncertainties in prediction. It is very difficult and also uncertain to make a prediction on an unknown sample, for which the model has not been trained. This can be considered with the help of adding uncertainties in prediction. For this problem, the author has proposed to use Bayesian deep learning techniques. And here the miss classified output can be considered as uncertainties. Deep convolutional neural network architecture is used to train the model to identify the infected part. Three optimization techniques were used in this paper namely: Stochastic Gradient Descent, Stochastic Gradient MCMC, and MC dropout. Out of these three MC dropout algorithm and SGD gave overconfident predictions and SGLD as less confident predictions and these predictions are based on the probabilistic entropy. It has the best accuracy out of three. Image entropy was the main parameter used to train the deep neural network.

Parul Sharma et.al, [2019]: In this paper, the author has addressed a major problem with the success of a model in the real world. Most of the algorithms proposed work efficiently on training data but fail to achieve the desired output in the real-world scenario. The author made a change in training a convolutional neural network. Instead of training the convolutional neural network with a full image the author trained the model with the help of a segmented image and got very efficient output. [12].

CHAPTER3: PRESENT WORK

The research work is done in two sections. In the first section, disease prediction is done using machine learning algorithms. In the second section, a complete discussion of the proposed framework is done using transfer learning.

3.1. ML IMPLEMENTATION:

In the automation of multiple processes, machine learning plays a critical role. The proposed architecture was designed with that goal in mind, and it is based on machine learning methodologies. Especially in the case of detecting and categorizing images into various disease categories. This section has been structured such that the topic begins with the device specifications and data acquisition for the data used in the currently suggested approach. The second point of discussion would be image segmentation. Feature extraction would be the focus of the debate. The fourth point of discussion would be the classification method
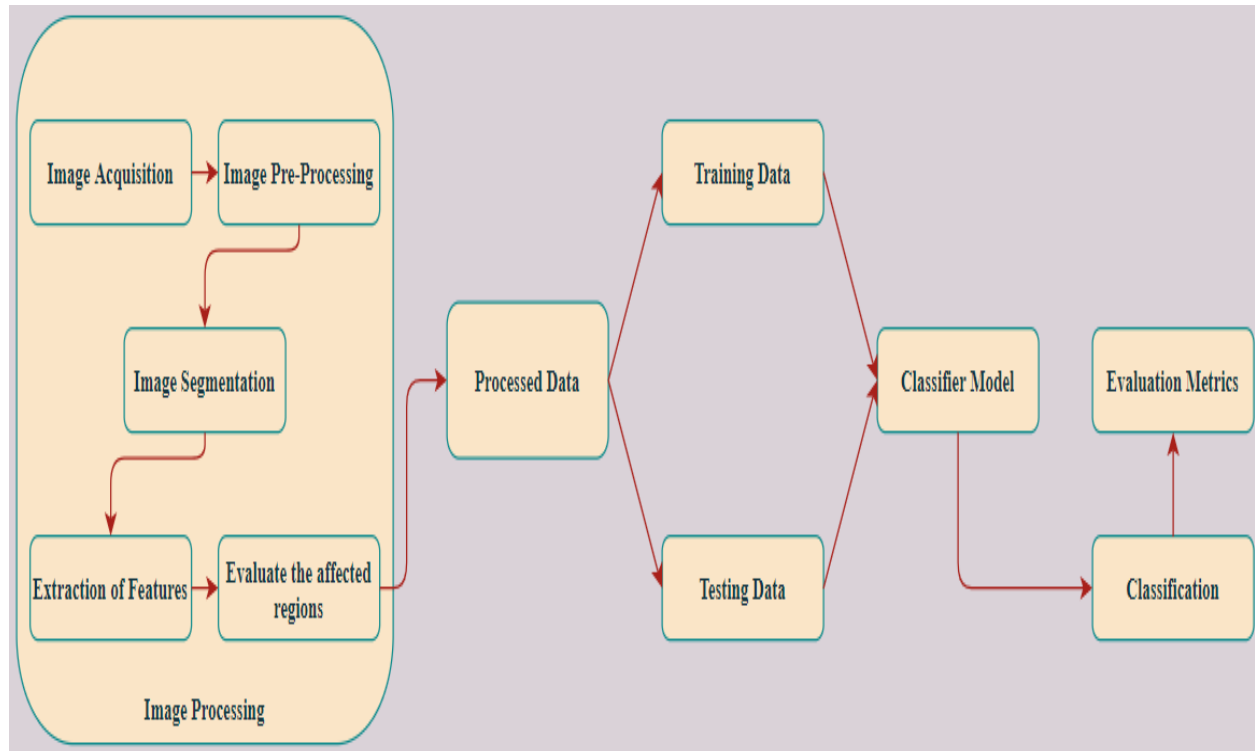
Figure 1. Flow Chart of the Machine Learning Methodology
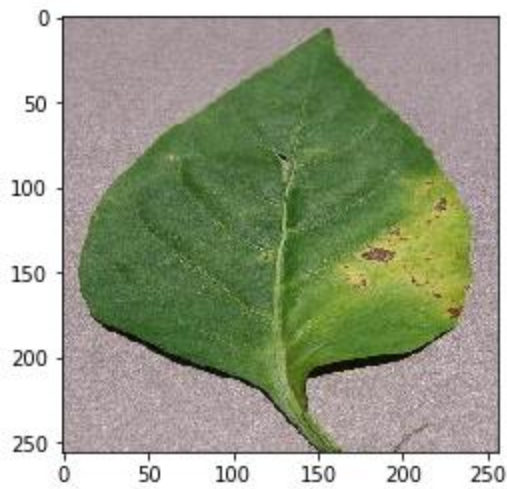
1.1. Image Processing Steps:

1. Import image:

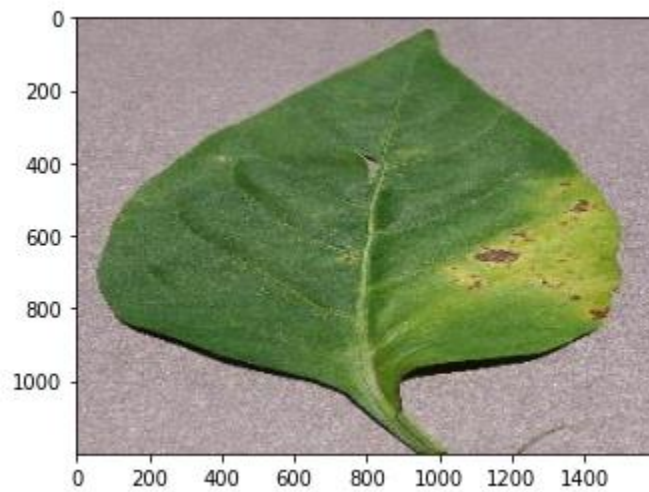**Figure2.1** Input Image

2. Resize image:

Figure2.2: Resized input image

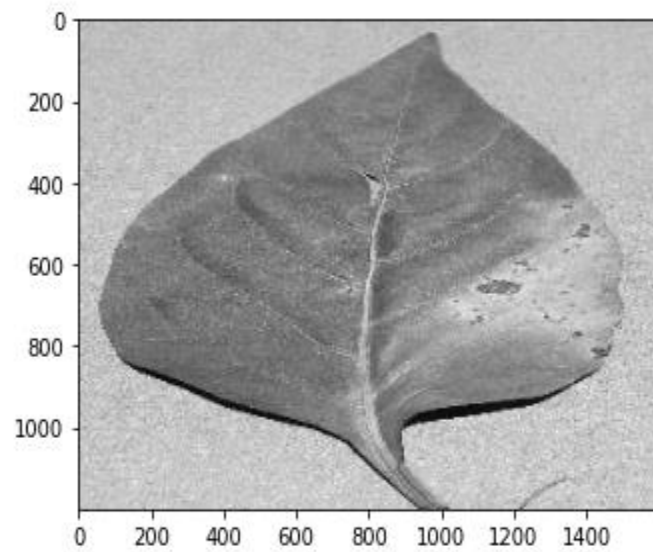### 3. Blurring the image

Figure2.3: Blurred Image

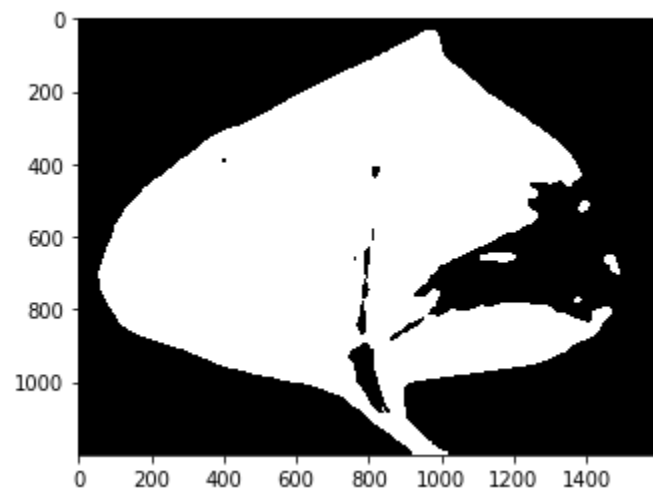### 4. Otsu segmentation to identify the infected region

Figure2.4: Segmented Image

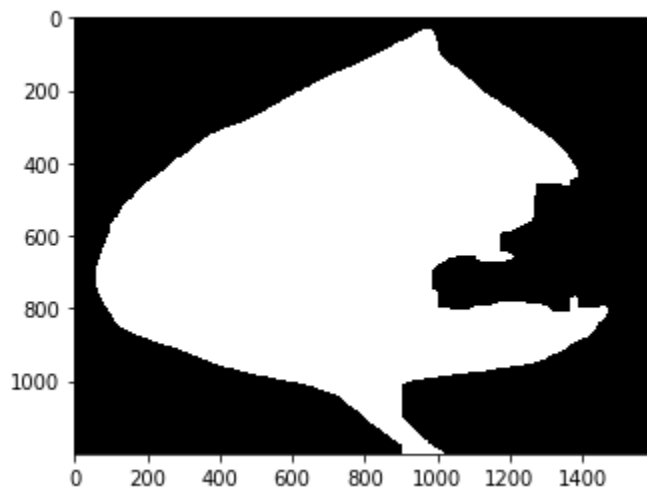**5. Create a mask image for background subtraction using leaf contour**

Figure2.5: Image mask for background removal.

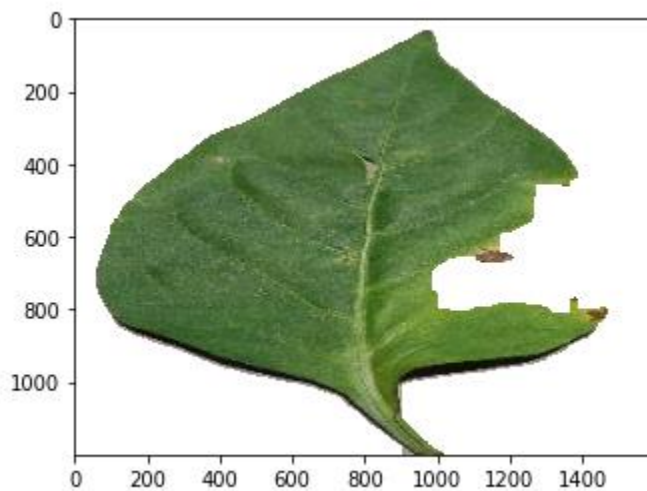**6. Performing masking operation on original image:**

Figure2.6: Original image after applying image mask.

17

6.1. Feature Extraction

When image segmentation provides infected region to do analysis. There are lot of data that can be extracted from the image. Dimensionality reduction is a crucial step to be followed. So as to avoid model confusion and conflict Also it is very important to consider all necessary features and avoid any miss. Feature Extraction is one of the most crucial steps in machine learning. Extraction of essential features is very important. Features need to be selected to avoid overfitting and under-fitting. In this project, various image features are extracted such as randomness, mean, entropy, the standard deviation of the colored image.

6.2. Classification Process

Depending on the features extracted, the extracted feature dataset classification need to be performed for various categories: Early Blight, Late Blight & healthy leaves. To perform classification support vector machine algorithm is used. Which involves a supervised learning approach. It is also considered as a part of an optimization technique and which is defined as in the below-mentioned equation. The training set here is represented as (X_k, Y_k), $K = 1, 2, \ldots \ldots, i$ $X_k \in R^n$ and $Y_k \in \{1, -1\}^i$.

$$\min_{w,f,\xi} \frac{1}{2} w^T w + C \sum_{k=1}^{i} \xi_k \text{ subjected to } Y_k = (w^T \phi(X_k) + e) \geq 1 - \xi_k, \xi_k \geq 0$$

(1)

The equation represented ablove is used in binary classification context. Whereas, in this proposed methodology, the classification performed is not a binary classification, so the above mentioned equation can be generalized to perfrom multiclass classification. The generalization will then be performed considering n-dimentional space where n is no of categories.

Proposed Methodology

A series of steps need to be carefully followed for the process need to be followed in a diseplined manner:

Step-1: Image Acquisition for dataset creation: This step involves exploring various data sources from where data can be extracted for training the model and further how the test image input is to be provided.

Step-2: Image Pre-processing and background removal: This is most important phas, as it involves the quality assurance of the data. In the image pre-processing phase image is processed to desired color format, resized to desired size and images are denoised.

Step-3: Image Segmentation to obtain infected region: Region of interest that is the infected part of the leaf is identified. This is again one of the most crucial step, as entire analysis is dependent on the infected refion identified by the process of segmentation.

Step-4: Extraction of Features from images: On the basis of obtained region of interest which is the infected part of the leaf various image features like standard deviation, mean of red, blue and green channels, the entropy of image is extracted.

Step-5: Evaluate and identification of the affected region: By comparing the extracted region of interests & features which are extracted from the image, an efficient model is derieved.

Step-6: Processed Dataset creation: The data which are processed in previous stages are processed and extracted and converted to a csv file format and stored. This stored data is further utilized for analysis purpose.

Step-7: Training Data Extraction: Randomly the data in csv file is split. The 70% of the split data is used for training the proposed model.

Step-8: Testing Data Extraction: Randomly the data in csv file is split. The 30% of the split data is used for training the proposed model.

Step-9: Classification: Test data has labels such as: Late Blight, Early Blight, and Healthy, based on which classification is performed.

Step-10: Evaluation of proposed model: Depending on the obtained results from the classifier model, the evaluation metrics such as precision, recall, F1-score, and accuracy will be obtained.

Figure 3. The three Sample leaves of potato are (1): leaf affected by Light Blight (2): leaf affected by Early Blight (3): leaf unaffected (Healthy)

3.2 TRANSFER LEARNING IMPLEMENTATION:

The proposed framework was built on the concept of transfer learning. For transfer learning, the pre-trained model needs to be considered. In this framework, the pre-trained models considered are various CNN architectures such as VGG-16, ResNet-50 which can be considered as version1, and ResNet-50 v2.

VGG-16 architecture: Before constructing VGG-16, two more standard models exist before VGG-16 such as LeNet-5, and AlexNet. VGG-16 is in no way different than those methodologies except for the depth of the model. When compared to LeNet-5 and AlexNet, the number of convolutional layers are very high. In the case of LeNet-5, the number of convolutional layers is 2, the number of layers in the case of AlexNet is 5, whereas the number of convolutional layers in the case of VGG-16 is 13. VGG-16 proved to be more efficient and consistent in performance when compared with the other two models. It also consists of 5 maximum pooling layers, 1 flatten layer and 3 fully connected layers. This model alone consists of 138 million parameters which will be covering up a good amount of storage. By the time VGG-16 evolved, a major part of the

20

researchers believed that improving the depth of the model will improve the performance of the model, but on contrary, that proved to be wrong. As an extension of VGG-16, VGG-19 was developed which is a deeper model when compared to VGG-16. VGG-19 consists of 16 convolutional layers, 5 maximum pooling layers, and 3 fully connected layers. According to the assumption of researchers, VGG-19 has to be performed better than the VGG-16 but in reality, it is not showing any considerable difference in efficiency and consistent performance when compared to VGG-16. At the same time, as the network size was increased the number of parameters also increased so the usage of storage also improved. So, VGG-16 can be considered as the most standard CNN architecture till VGG-19 was evolved. The VGG-16 architecture can be mentioned as shown in figure-1.
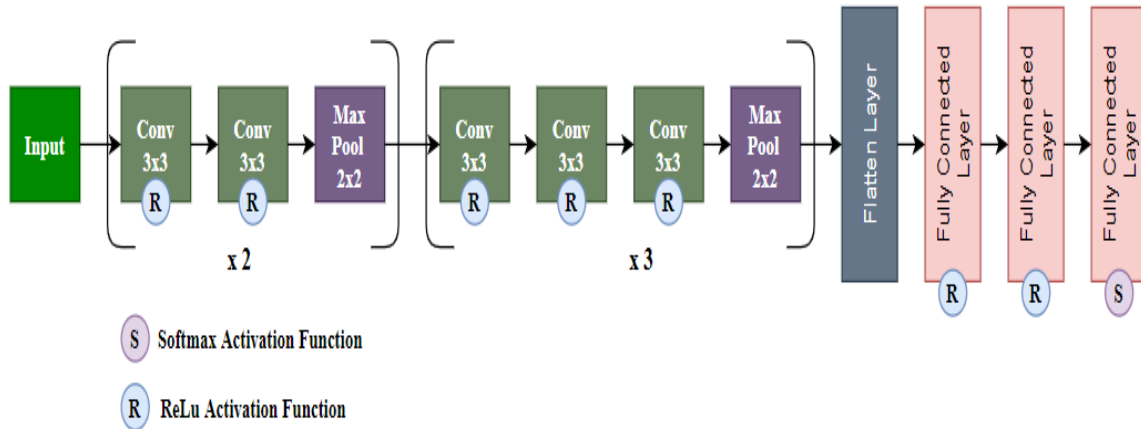

Figure. 4. VGG-16 basic architecture

ResNet-50: It is the first model that utilized the batch normalization concept very effectively. Before ResNet-50 and post VGG-19 there is a model developed known as Inception or GoogleNet. This model is also concentrated on the depth of the model but it was implemented in blocks in the form of parallel towers. It might be efficient and consistent when compared to VGG-16 but it has its negative implications in terms of stabilized learning as well as the storage aspects. From there, the newer concept was evolved and it is popularly known as "Batch Normalization". Due to this concept, the models able to attain a stabilized learning process during training as well as the reduction in epochs i.e, a smaller number of epochs are sufficient to train the model which will be time-effective, cost-effective, as well as storage effective. This concept was successfully inserted into ResNet-50. So, post-VGG-16, ResNet-50 is considered to be one of the greatest landmarks in the case of CNN architecture development. This model consists of 26 million parameters and the

basic building block of ResNet-50 consists of 49 convolutional layers as well as one fully connected layer. The overall architecture of ResNet-50 can be represented as mentioned in figure-2.
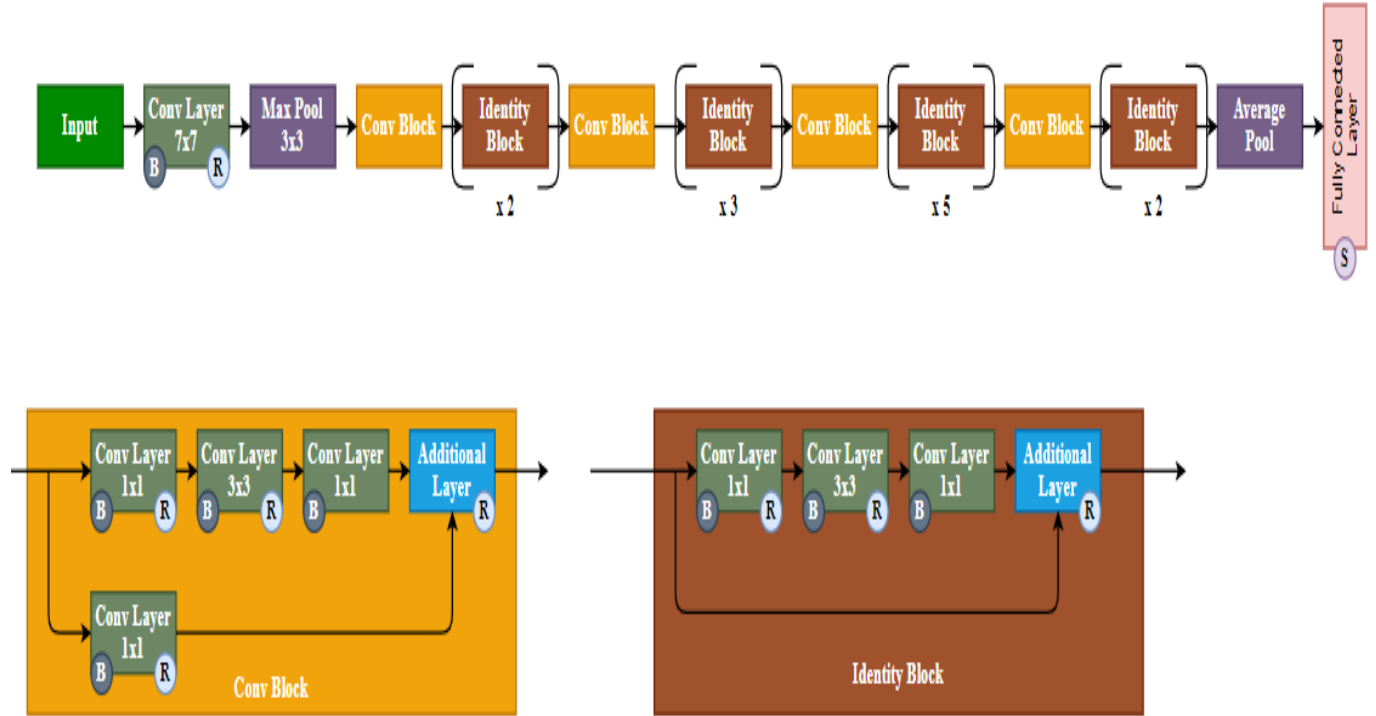

Figure 5: ResNet-50 basic architecture

ResNet-50 v2: The model ResNet-50 v2 is almost the same as the Resnet-50 in terms of the total number of convolutional layers, fully connected layer as well as the number of parameters. The major difference between the ResNet-50 v1 and ResNet-50 v2 is the way of applying batch normalization and activation function. ResNet-50 v1 is considered to be a post-activation model due to the concept of batch normalization and activation functions applied to post the weight layers whereas ResNet-50 v2 is considered to be a pre-activation model due to the concept of batch normalization and activation function applied before the weight layers. The basic structural difference of ResNet-50 v1 and Resnet-50 v2 can be represented as mentioned in figure-3. These three models are considered as the pre-trained models for the concept of transfer learning.
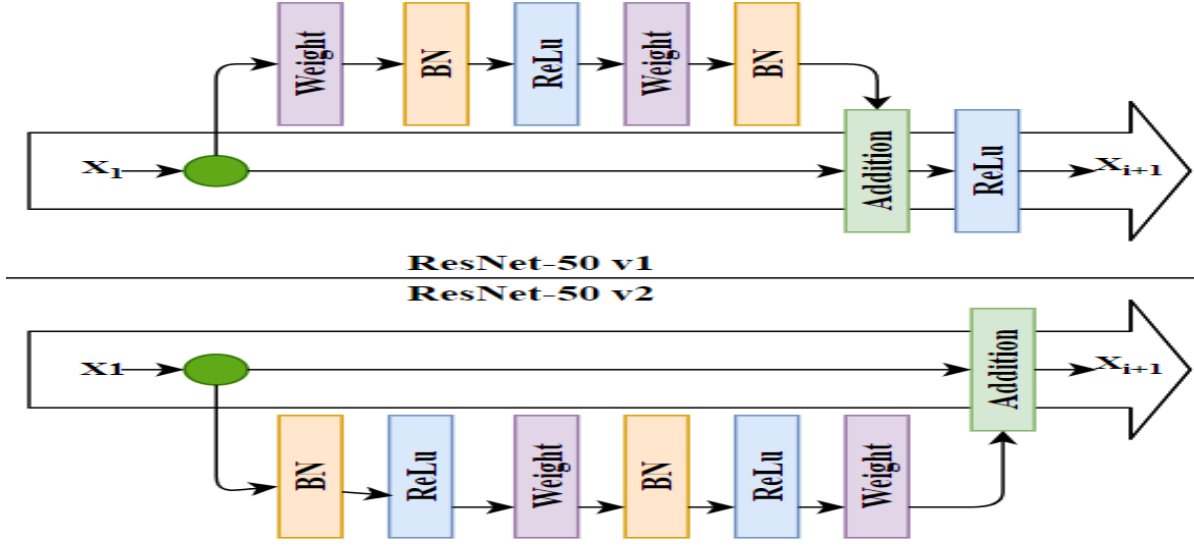
Figure 6: Structural variation between ResNet-50 v1 and ResNet-50 v2

Transfer Learning: To understand the concept of transfer learning in laymen's language is the utilization of the existing knowledge from the resource learner in the objective job. In the process of transfer learning, the researchers need to take care of three aspects as mentioned below.

- Try to identify the answers for the kind of knowledge that needs to be transferred from resource learner to the objective for improvisation of performance of the objective job.
- Need to identify when the identified knowledge to be transferred to enhance the performance of the objective job rather than degrading it.
- Need to identify the ways in which knowledge can be transferred across various jobs.

The proposed framework deals with all these three steps in a specific way. The standard CNN architecture can extract the features from the objective i.e, images, and keep on training the model. Once the training of the model is done the knowledge will be transferred, it can be utilized as a pre-trained model for various other similar problems. Through this pre-trained model, the knowledge will be transferred to accomplish all the jobs that exist in the problem. The proposed framework can be implemented as mentioned in figure-4.
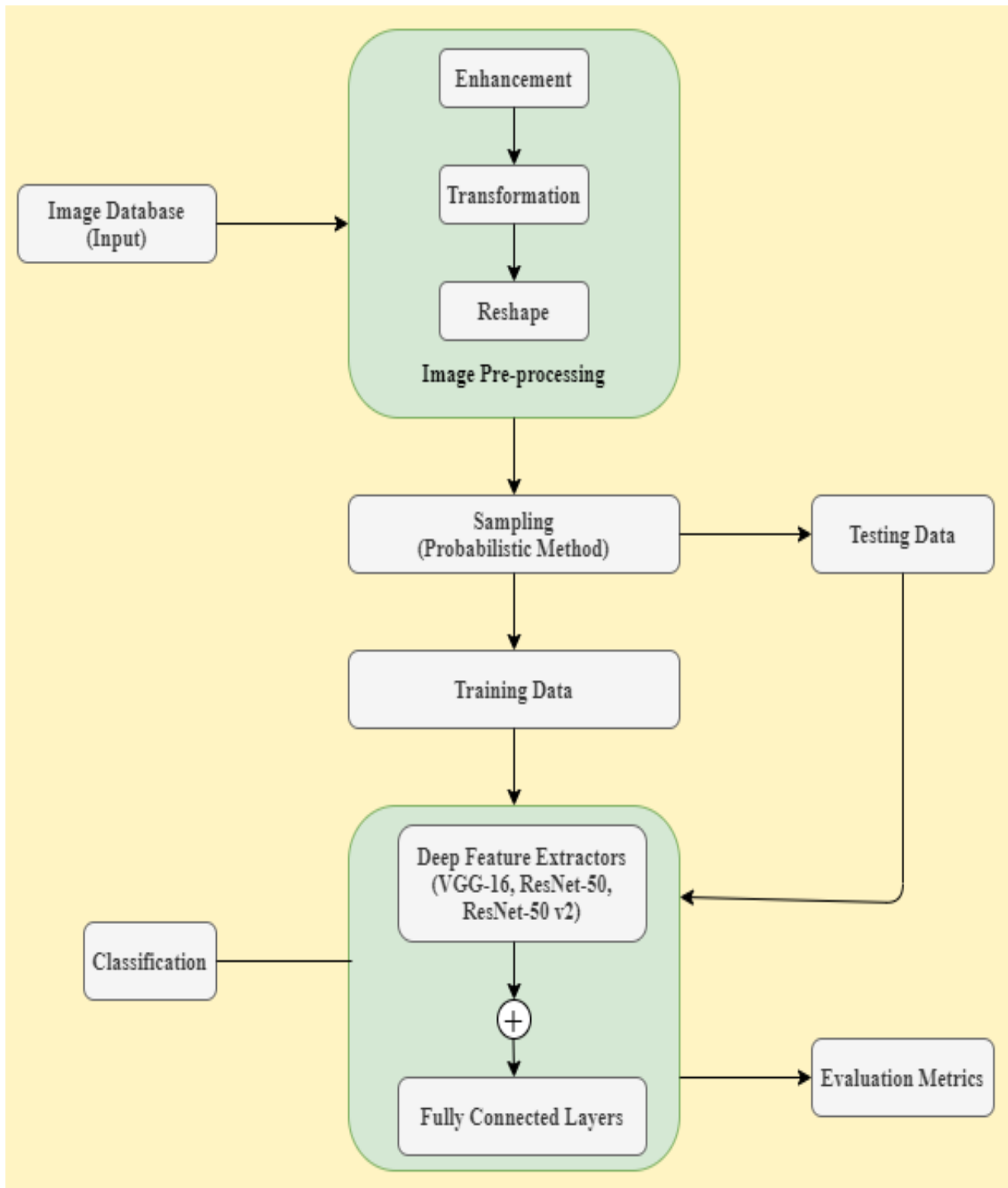
Figure 7: The flowchart for the proposed model

CHAPTER 4: RESULTS AND DISCUSSION

4.1. Disease detection using ML:

At first machine learning algorithm was implemented to detect plant disease. This was done in two phases. I.) Implementing ML algorithm on potato dataset. II) Implementing ML algorithm on the entire dataset.

    I.   Implementing ML algorithm on potato dataset.

The dataset which is considered in the proposed work is an openly accessed dataset & it was randomly divided into the training dataset consists of 1820 images and the testing dataset consists of 780 images. The otsu algorithm was utilized for the binary image segmentation and infected region identification this was done with the help of preparing an image mask. The Gray Level Co-occurrence Matrix is the main tool that implements the concepts learned from extracted features. utilized for feature extraction, & multi-class support vector machine(SVM) methodology was utilized for the classification of potato leaves. The model derived is evaluated using certain evaluation metrics: precision, recall, F1-score, and accuracy.

Table 1. Evaluation Metrics of Potato Disease analysis using SVM

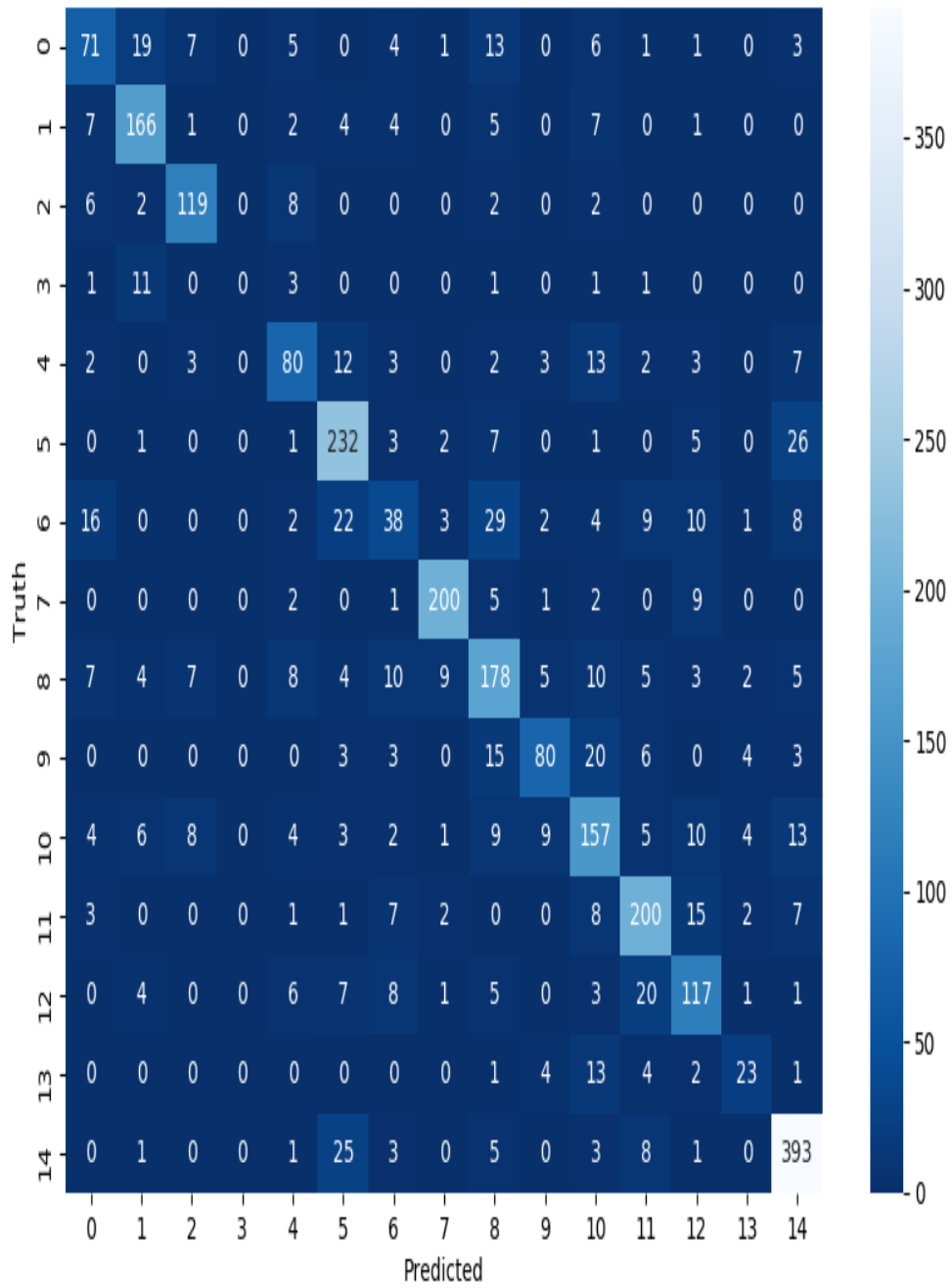| Category Name	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Late Blight	91.07	95.41	93.29	94.71
Early Blight	98.36	94.71	96.43	96.84
Healthy	98.93	98.62	98.76	96.43
Overall	96.12	96.25	96.16	95.99

Figure 8 : Confusion matrix obtained for SVM based transfer learning approach

PLANT DISEASE DETECTION USING TRANSFER LEARNING:

The proposed framework is implemented on the three various models of transfer learning and pre-training models considered are VGG-16, ResNet-50, and ResNet-50 v2. The confusion matrices are obtained for each of the models as mentioned in figures- 5, 6, and 7 respectively. Evaluation metrics obtained from each of the models are represented in the Table-1 and this evaluation metrics comparison is visualized by using graphical representation as mentioned in figure-8. The accuracy obtained from VGG-16 is 98.74%, ResNet-50 is 98.84%, and ResNet-50 v2 is 98.21%.
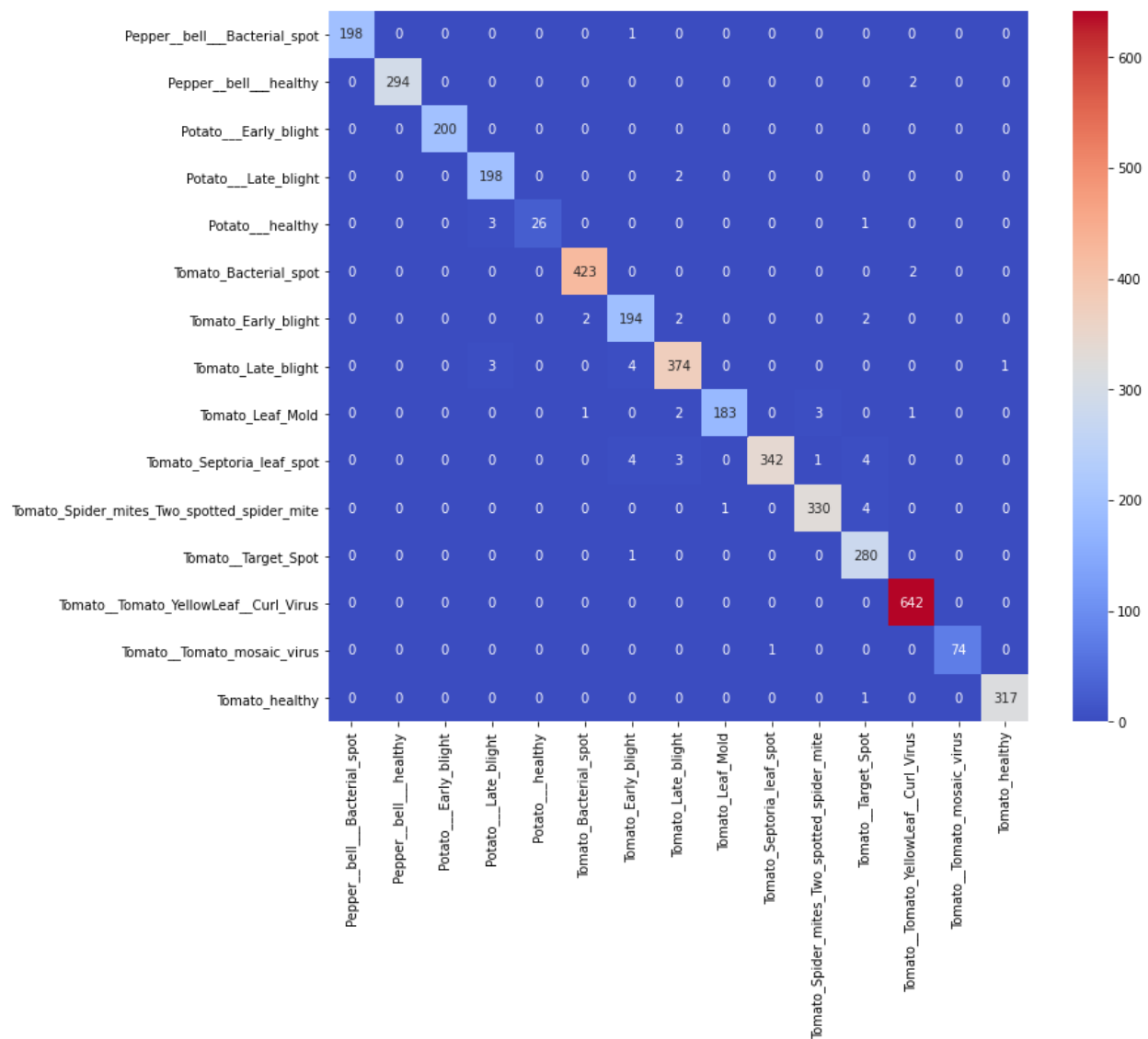

Figure 9 : Confusion matrix obtained for VGG-16 based transfer learning approach
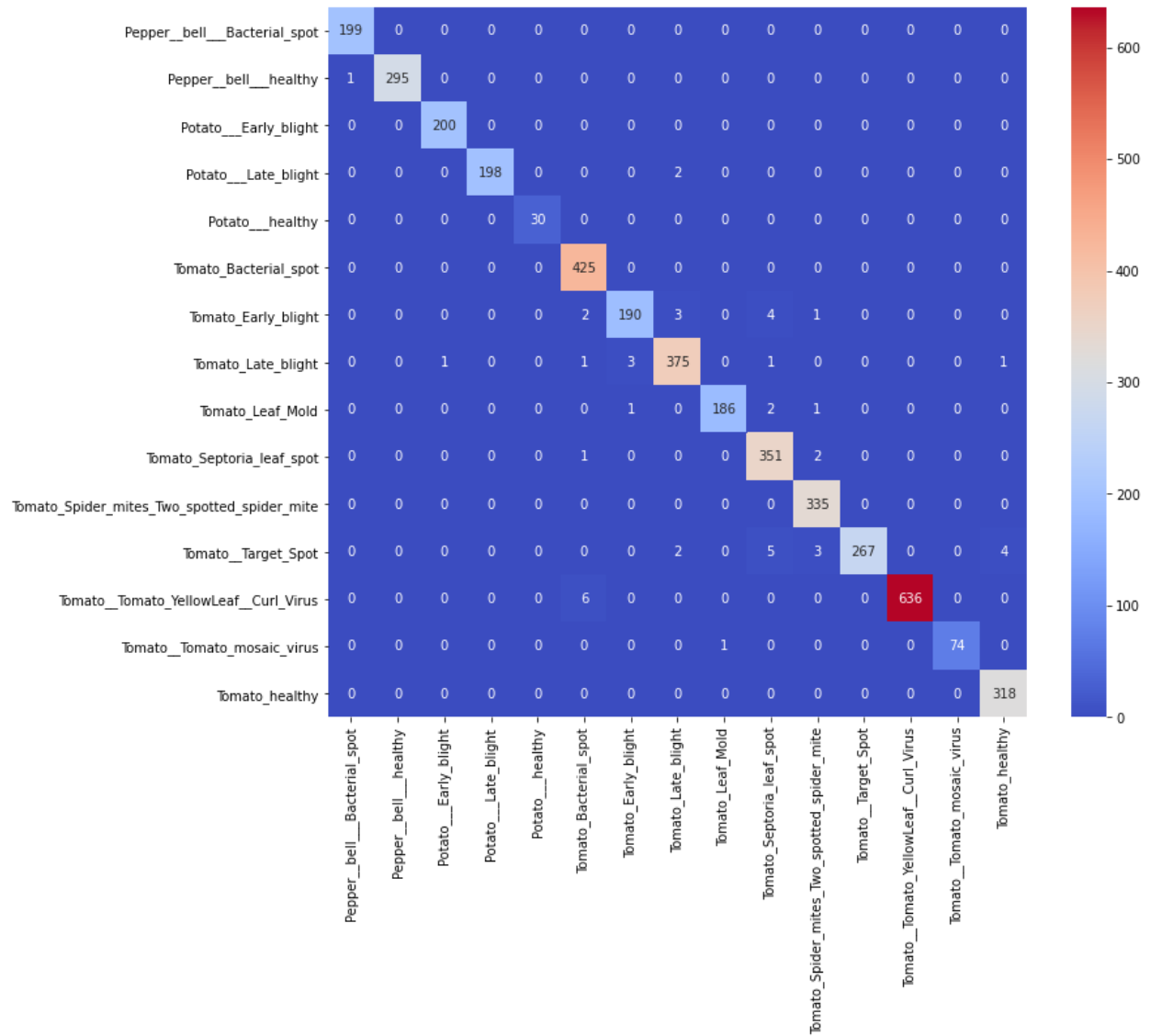
Figure 10 : Confusion matrix obtained for ResNet-50 based transfer learning approach
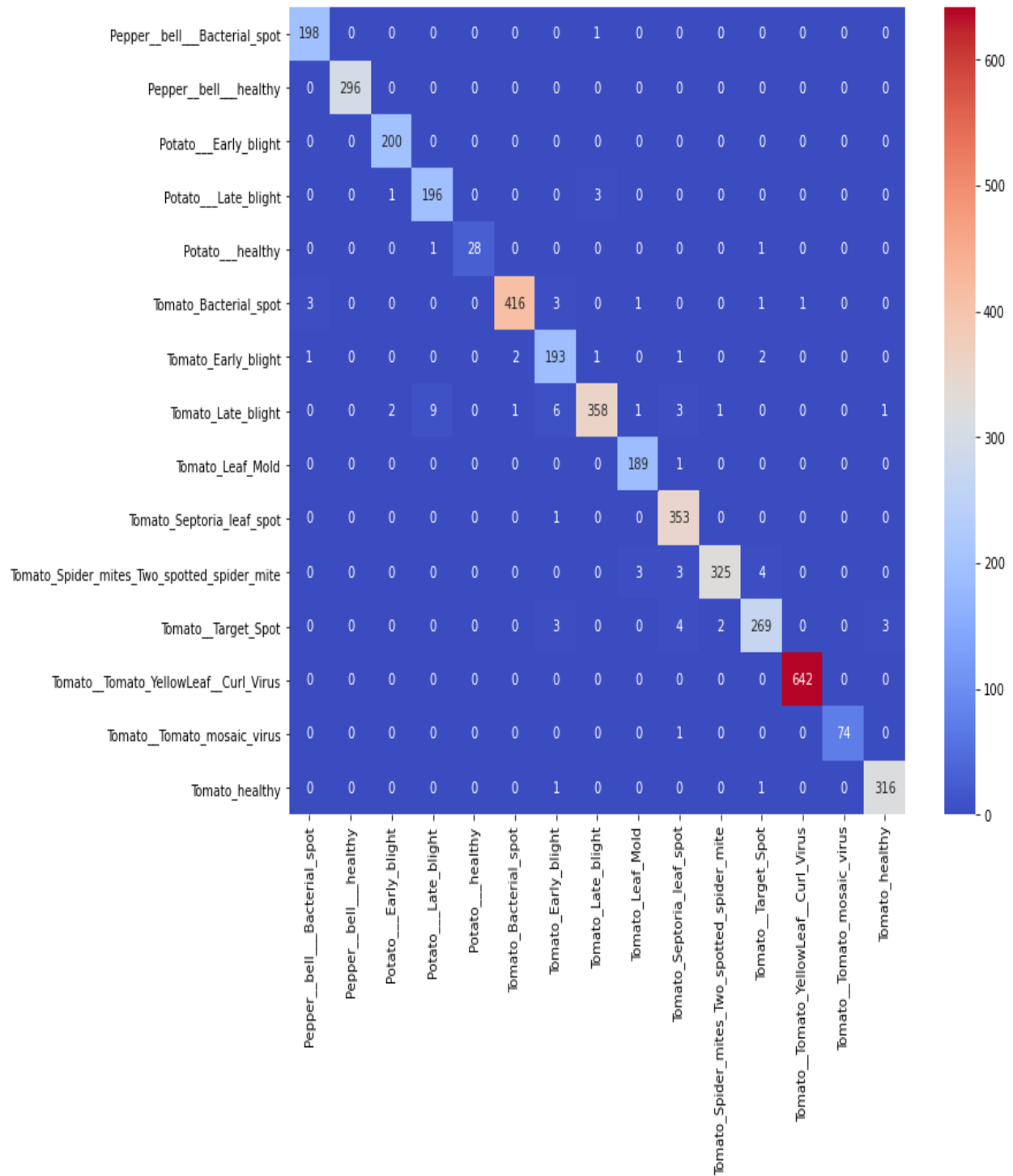
Figure 11: Confusion matrix obtained for ResNet-50 v2 based transfer learning approach

Table 2: Details of evaluation metrics obtained for the transfer learning models implemented

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
VGG-16	98.74	98.76	98.74	98.74
ResNet-50	98.84	98.85	98.84	98.83
ResNet-50 v2	98.21	98.23	98.21	98.20

Figure 12: Comparison of evaluation metrics across the models implemented

(1)

(2)

Figure 13: (1) Input Interface (2) Interface output

CONCLUSION

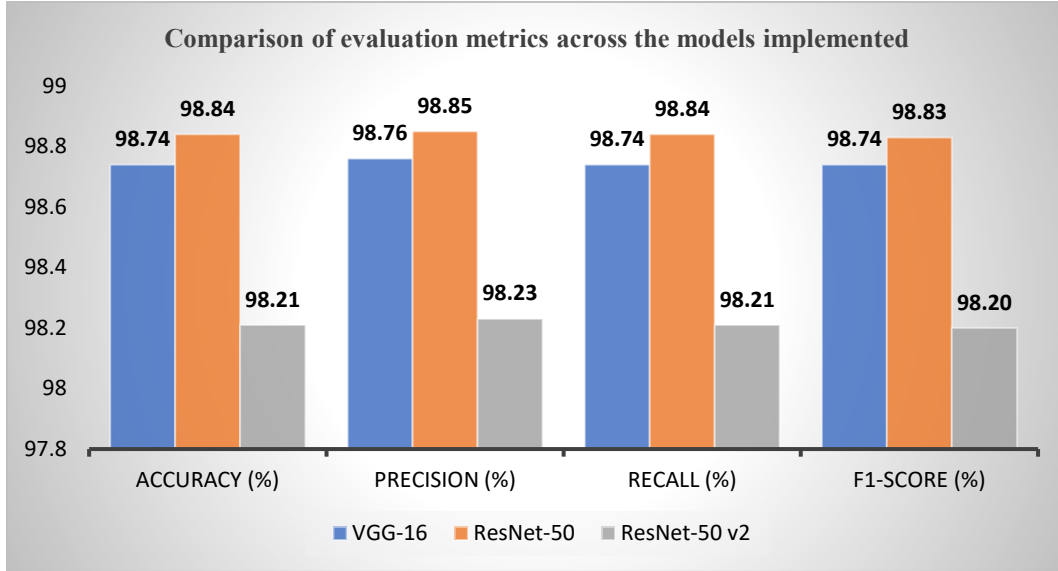The machine learning algorithm used in proposed work is SVM. SVM gave good result in when the detection categories were less. As the no of disease categories increased it failed to achieve the accuracy. Transfer learning is the current effective research for obtaining the better performance of the models with a minimal and faster training phase. It proved very true with the proposed framework. The proposed framework able to attain better accuracy with all the three models such as VGG-16, ResNet-50, and ResNet-50 v2, yet ResNet-50 based transfer learning model a bit more efficient when compared to the other models. The proposed framework efficient with the multi-class classification of various diseases along with healthy leaves that include crops of pepper, potato, and tomato.

LIMITATIONS:

1. The three most crucial points in the selection of any transfer deep learning model are:
   - The proposed framework is utilized for the classification of diseases across the various species of crops.
   - The proposed framework utilized the concept of deep learning.
   - The proposed framework also adopted a trending research concept of transfer learning and able to achieve a better efficient model.

   If any of these three are neglected will result in a negative transfer or overfitting problem.
2. Machine learning models are not very efficient in predicting diseases from leaf images when the no of categories is increased.

FUTURE SCOPE:

1. The disease detection system can be integrated in cloud system for efficient result processing.
2. Integration of automated disease detection system with sensos to measure soil p

REFERENCES:

[1] Sardogan, M., Tuncer, A., and Ozen, Y.: Plant Leaf Disease Detection and Classification Based on CNN with the LVQ Algorithm. In: 3rd Int. Conf. Comput. Sci. Eng. (2018) 382–385

[2] Wallelign, S., Polceanu, M., and Buche, C.: Soybean plant disease identification using a convolutional neural network. In: Proc. 31st Int. Florida Artif. Intell. Res. Soc. Conf. FLAIRS 2018 (2018), 146–151

[3] Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., and Stefanovic, D.: Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. Comput. Intell. Neurosci. 2016(2016)

[4] Fuentes, A., Yoon, S., Kim, S. C., and Park, D. S.: A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors (Switzerland) 17 (2017).

[5] Arivazhagan, S. and Ligi, S. V.: Mango Leaf Diseases Identification Using Convolutional Neural Network. Int. J. Pure Appl. Math., 120(2018) 11067–11079

[6] Oppenheim, D. and Shani, G.: Potato Disease Classification Using Convolution Neural Networks. Adv. Anim. Biosci., 8 (2017), 244–249

[7] Barbedo, J. G. A.: Factors influencing the use of deep learning for plant disease recognition. Biosyst. Eng., 172 (2018) 84–91

[8] Brahimi, M., Boukhalfa, K., and Moussaoui, A.: Deep Learning for Tomato Diseases: Classification and Symptoms Visualization. Appl. Artif. Intell., 31 (2017) 299–315

[9] Shrivastava, V. K., Pradhan, M. K., Minz, S., and Thakur, M. P.: Rice plant disease classification using transfer learning of deep convolution neural network. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch., 42 (2019) 631–635

[10] Ozguven, M. M. and Adem, K.: Automatic detection and classification of leaf spot disease in sugar beet using deep learning algorithms. Phys. A Stat. Mech. its Appl., 535 (2019) 122537

[11] Uğuz, S. and Uysal, N.: Classification of olive leaf diseases using deep convolutional neural networks. Neural Comput. Appl., 5 (2020)

[12] Agarwal, M., Gupta, S. K., and Biswas, K. K.: Development of Efficient CNN model for Tomato crop disease identification. Sustain. Comput. Informatics Syst., 28 (2020) 100407

[13] Wang, J., Chen, L., Zhang, J., Yuan, Y., Li, M., and Zeng, W. H.: CNN transfer learning for automatic image-based classification of crop disease. Springer Singapore 875 (2018)

[14] Toda, Y., and Okura, F.: How Convolutional Neural Networks Diagnose Plant Disease. Plant Phenomics, 2019 (2019) 1–14

[15] Tran, T. T., Choi, J. W., Le, T. T. H. and Kim, J. W.: A comparative study of deep CNN in forecasting and classifying the macro-nutrient deficiencies on development of tomato plant. Appl. Sci., 9 (2019)

[16] Kiani, E., & Mamedov, T. (2017). Identification of plant disease infection using soft-computing: Application to modern botany. *Procedia Computer Science*, *120*, 893–900. https://doi.org/10.1016/j.procs.2017.11.323

[17] Singh, V., & Misra, A. K. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture*, *4*(1), 41–49. https://doi.org/10.1016/j.inpa.2016.10.005

[18] Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, *145*(January), 311–318. https://doi.org/10.1016/j.compag.2018.01.009

[19] Golhani, K., Balasundram, S. K., Vadamalai, G., & Pradhan, B. (2018). A review of neural networks in plant disease detection using hyperspectral data. *Information Processing in Agriculture*, *5*(3), 354–371. https://doi.org/10.1016/j.inpa.2018.05.002

[20] Mattihalli, C., Gedefaye, E., Endalamaw, F., & Necho, A. (2018). Plant leaf diseases detection and auto-medicine. *Internet of Things*, *1–2*, 67–73. https://doi.org/10.1016/j.iot.2018.08.007

[21] Pantazi, X. E., Moshou, D., & Tamouridou, A. A. (2019). Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers. *Computers and Electronics in Agriculture*, *156*(July 2018), 96–104. https://doi.org/10.1016/j.compag.2018.11.005

[22] Sankaran, S., Mishra, A., Ehsani, R., & Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture*, *72*(1), 1–13. https://doi.org/10.1016/j.compag.2010.02.007

[23] Sood, M., & Kumar, P. (2020). ScienceDirect Stood ScienceDirect Hybrid System for Detection and Classification of Plant Disease Using for Qualitative Texture Features Analysis Hybrid System Detection and Classification of Plant Disease Using Qualitative Texture Features Analysis. *Procedia Computer Science*, *167*(2019), 1056–1065. https://doi.org/10.1016/j.procs.2020.03.404

[24] M. Islam, A. Dinh, K. Wahid, and P. Bhowmik, Detection of potato diseases using image segmentation and multiclass support vector machine 2017, *Can. Conf. Electr. Comput. Eng.*, 8–11.

[25] H. Waghmare, R. Kokare, and Y. Dandawate, Detection and classification of diseases of Grape plant using opposite color Local Binary Pattern feature and machine learning for automated Decision Support System 2016, *3rd Int. Conf. Signal Process. Integ. Networks,* 513–8.

[26] S. R. Maniyath *et al.*, Plant disease detection using machine learning 2018, *Proc. 2018 Int. Conf. Des. Innov. 3Cs Comput. Commun. Control.*, 41–5.

[27] U. Shruthi, V. Nagaveni, and B. K. Raghavendra, A Review on Machine Learning Classification Techniques for Plant Disease Detection 2019, *2019 5th Int. Conf. Adv. Comput. Commun. Syst.*, 281–4.

[28] R. Kaur and S. S. Kang, An enhancement in classifier support vector machine to improve plant disease detection, *Proc. 2015 IEEE 3rd Int. Conf. MOOCs, Innov. Technol. Educ.*, 135–40.

[29] V. Pooja, R. Das, and V. Kanchana, Identification of plant leaf diseases using image processing techniques 2018, *Proc.- 2017 IEEE Technol. Innov. ICT Agric. Rural Dev.*, **1**, 130–3.

[30] J. P. Shah, H. B. Prajapati, and V. K. Dabhi, A survey on detection and classification of rice plant diseases 2016, *2016 IEEE Int. Conf. Curr. Trends Adv. Comput.*

[31] Mckay, C., Swaminathan, A., Gou, H., & Wu, M. (2008). 04517945. 1657–1660.

[32] Ullah, Z., Farooq, M. U., Lee, S. H., & An, D. (2020). A hybrid image enhancement-based brain MRI images classification technique. *Medical Hypotheses*, *143*(April), 109922. https://doi.org/10.1016/j.mehy.2020.109922

[33] Ramya, V., & Lydia, M. A. (2016). Leaf Disease Detection and Classification using Neural Networks. *International Journal of Advanced Research in Computer and Communication Engineering*, *5*(11), 207–210. https://doi.org/10.17148/IJARCCE.2016.51144

[34] Chanda, M., & Biswas, M. (2019). Plant disease identification and classification using back-propagation neural network with particle swarm optimization. *Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019*, *2019-April*(Icoei), 1029–1036. https://doi.org/10.1109/icoei.2019.8862552

[35] Hernández, S., & López, J. L. (2020). Uncertainty quantification for plant disease detection using Bayesian deep learning. *Applied Soft Computing Journal*, *96*, 106597. https://doi.org/10.1016/j.asoc.2020.106597
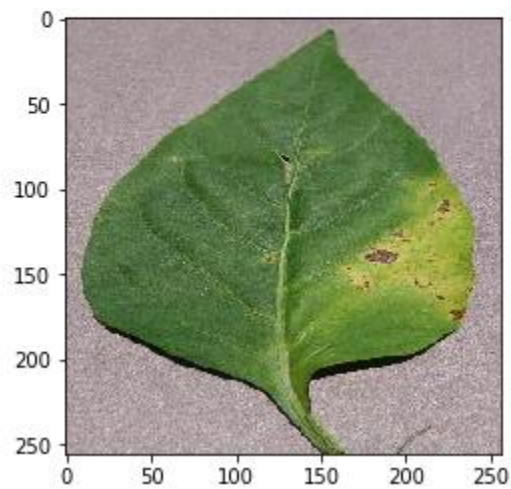
APPENDIX

SOME CODE SIPPNETS:

1. DISEASE PREDICTION USING MACHINE LEARNING:

   Step1: Background removal from image:


```
In [20]: main_img = cv2.imread(test_img_path)
         img = cv2.cvtColor(main_img, cv2.COLOR_BGR2RGB)
         plt.imshow(img,cmap="Greys_r")

Out[20]: <matplotlib.image.AxesImage at 0x1f7bd5bcf48>
```
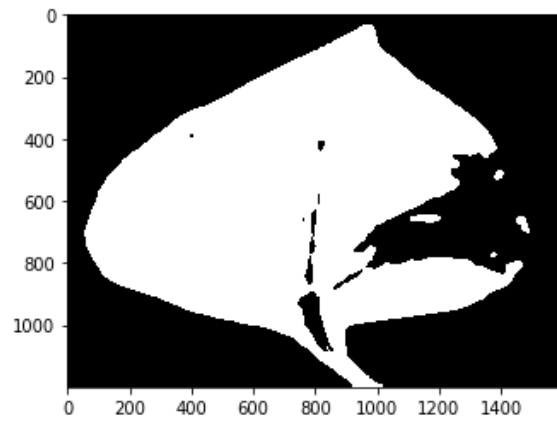
```
ret_otsu,im_bw_otsu = cv2.threshold(blur,0,255,cv2.THRESH_BINARY_INV+cv2.THRESH_OTSU)
plt.imshow(im_bw_otsu,cmap='Greys_r')
```

<matplotlib.image.AxesImage at 0x1f7bb230888>
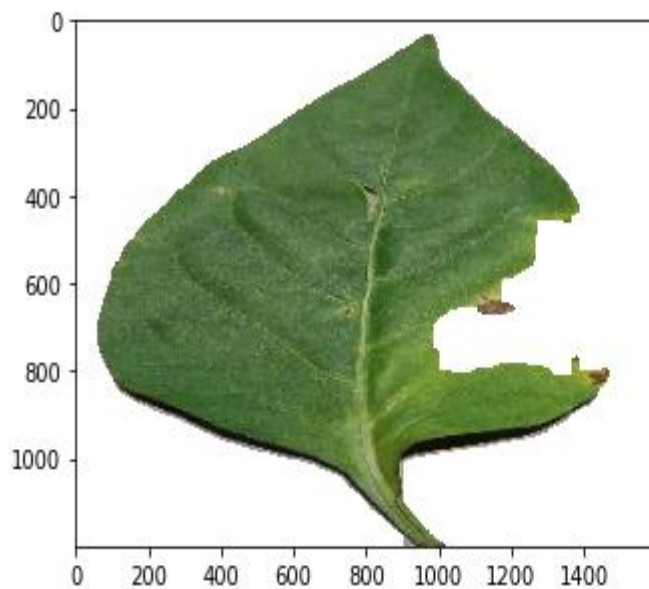
```
: white_pix = [255,255,255]
  black_pix = [0,0,0]

  final_img = maskedImg
  h,w,channels = final_img.shape
  for x in range(0,w):
      for y in range(0,h):
          channels_xy = final_img[y,x]
          if all(channels_xy == black_pix):
              final_img[y,x] = white_pix
```

```
: plt.imshow(final_img)
```

```
: <matplotlib.image.AxesImage at 0x1f7bca31f48>
```

Step2: Dataset Creation

```python
def load_image_files(container_path, dimension=(64, 64, 3)):
    image_dir = Path(container_path)
    folders = [directory for directory in image_dir.iterdir() if directory.is_dir()]
    categories = [fo.name for fo in folders]

    names = ['mean_r','mean_g','mean_b','stddev_r','stddev_g','stddev_b', \
             'contrast','correlation','inverse_difference_moments','entropy','label'
            ]
    df = pd.DataFrame([], columns=names)
    for i, direc in enumerate(folders):
        for file in direc.iterdir():

            label=i
            print(file)
            print(type(file))
            imgpath=file.__str__()
            print(type(imgpath))
            main_img = cv2.imread(imgpath)
            #print(main_img)

            #Preprocessing
            img = cv2.cvtColor(main_img, cv2.COLOR_BGR2RGB)
            gs = cv2.cvtColor(img,cv2.COLOR_RGB2GRAY)
            blur = cv2.GaussianBlur(gs, (25,25),0)
            ret_otsu,im_bw_otsu = cv2.threshold(blur,0,255,cv2.THRESH_BINARY_INV+cv2.THRESH_OTSU)
            kernel = np.ones((50,50),np.uint8)
            closing = cv2.morphologyEx(im_bw_otsu, cv2.MORPH_CLOSE, kernel)
```

```python
            #Color features
            red_channel = img[:,:,0]
            green_channel = img[:,:,1]
            blue_channel = img[:,:,2]
            blue_channel[blue_channel == 255] = 0
            green_channel[green_channel == 255] = 0
            red_channel[red_channel == 255] = 0

            red_mean = np.mean(red_channel)
            green_mean = np.mean(green_channel)
            blue_mean = np.mean(blue_channel)

            red_std = np.std(red_channel)
            green_std = np.std(green_channel)
            blue_std = np.std(blue_channel)

            #Texture features
            textures = mt.features.haralick(gs)
            ht_mean = textures.mean(axis=0)
            contrast = ht_mean[1]
            correlation = ht_mean[2]
            inverse_diff_moments = ht_mean[4]
            entropy = ht_mean[8]

            vector = [red_mean,green_mean,blue_mean,red_std,green_std,blue_std,\
              contrast,correlation,inverse_diff_moments,entropy,label
              ]

            df_temp = pd.DataFrame([vector],columns=names)
            df = df.append(df_temp)
            #print(img)
    return df
```

Step 4: Applying SVM to the dataset

```
import numpy as np
import pandas as pd
import os
import string
from sklearn.model_selection import train_test_split
```

```
dataset = pd.read_csv("plant_village_svm_image.csv")
```

```
data=pd.read_csv("plant_village_svm_image.csv")
```

```
dataset.head(5)
```

	Unnamed: 0	mean_r	mean_g	mean_b	stddev_r	stddev_g	stddev_b	contrast	correlation	inverse_difference_moments	entropy	label
0	0	119.115326	121.576431	98.027527	50.152874	50.609478	57.638393	793.659110	0.843263	0.093317	13.587035	0
1	0	143.174530	146.939346	123.799118	54.253350	51.836093	64.900211	603.173061	0.893223	0.139993	13.163057	0
2	0	136.387131	144.797287	128.792007	50.822961	45.731046	56.488960	1016.151532	0.778726	0.065172	13.895159	0
3	0	108.445770	128.540268	128.710266	43.976010	39.104082	48.979839	635.675258	0.813031	0.103604	13.344537	0
4	0	136.695541	144.784637	130.345825	42.548343	38.495790	53.776973	843.987074	0.743673	0.100887	13.413923	0

Step 5: Accuracy analysis.

TRANSFER LEARNING IMPLEMENTATION:

Step1: Import dataset

Step2: Split data into train and test directories.

Step3: Load pre-trained neural network (VGG/ ResNet50/ResNet50 v2)

Laoding the Neural Network architecture (pre-trained model)

```
[ ]  img_height, img_width = 224,224
     conv_base = vgg16.VGG16(weights='imagenet', include_top=False, pooling='max', input_shape = (img_width, img_height, 3))

     Downloading data from https://storage.googleapis.com/tensorflow/keras-applications/vgg16/vgg16_weights_tf_dim_ordering_tf_kernels_notop.h5
     58892288/58889256 [==============================] - 1s 0us/step
```

```python
import shutil
import os
import numpy as np
import argparse

def get_files_from_folder(path):

    files = os.listdir(path)
    return np.asarray(files)

def main(path_to_data, path_to_test_data, train_ratio):

    _, dirs, _ = next(os.walk(path_to_data))

    data_counter_per_class = np.zeros((len(dirs)))
    for i in range(len(dirs)):
        path = os.path.join(path_to_data, dirs[i])
        files = get_files_from_folder(path)
        data_counter_per_class[i] = len(files)
    test_counter = np.round(data_counter_per_class * (1 - train_ratio))

    for i in range(len(dirs)):
        path_to_original = os.path.join(path_to_data, dirs[i])
        path_to_save = os.path.join(path_to_test_data, dirs[i])

        if not os.path.exists(path_to_save):
            os.makedirs(path_to_save)
        files = get_files_from_folder(path_to_original)


            for j in range(int(test_counter[i])):
                dst = os.path.join(path_to_save, files[j])
                src = os.path.join(path_to_original, files[j])
                shutil.move(src, dst)

    if __name__ == "__main__":
        main('./PlantVillage', './testData/', 0.8)
```
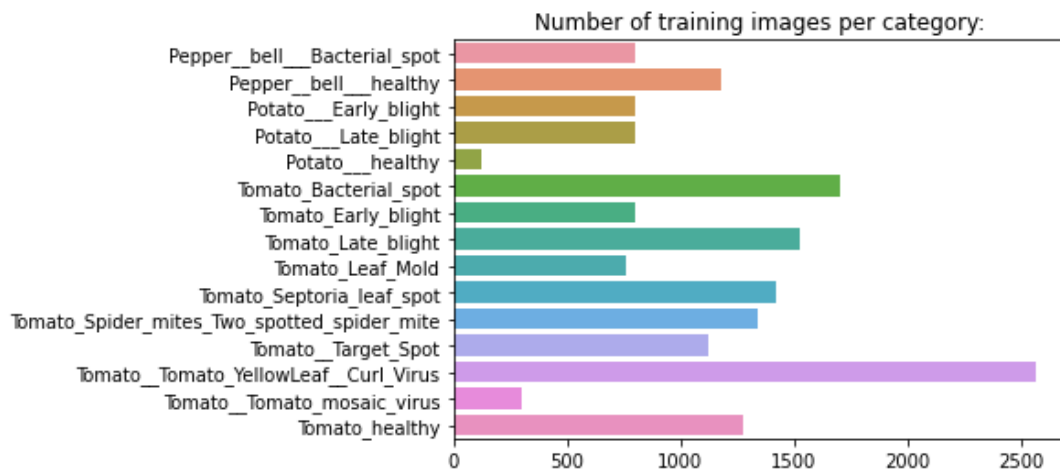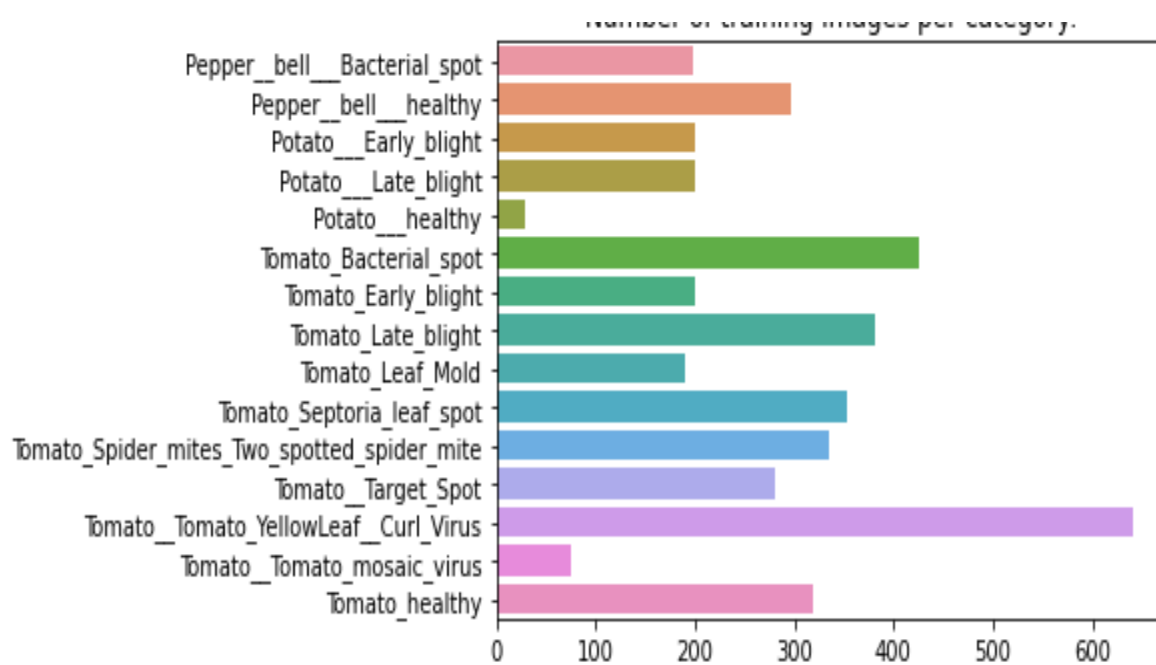
Analysis of training and testing images per categories:

Number of training images per category:

Analysis of test images per category:

Step4: Checking all the layers in loaded Neural network

Cheking all the layers in loaded Neural network

```
for layer in conv_base.layers:
    print(layer, layer.trainable)
```

```
<tensorflow.python.keras.engine.input_layer.InputLayer object at 0x7f8f71871ef0> True
<tensorflow.python.keras.layers.convolutional.Conv2D object at 0x7f8f4067e128> True
<tensorflow.python.keras.layers.convolutional.Conv2D object at 0x7f8f4067e438> True
<tensorflow.python.keras.layers.pooling.MaxPooling2D object at 0x7f8f4067e860> True
<tensorflow.python.keras.layers.convolutional.Conv2D object at 0x7f8f303d64e0> True
<tensorflow.python.keras.layers.convolutional.Conv2D object at 0x7f8f303dd400> True
<tensorflow.python.keras.layers.pooling.MaxPooling2D object at 0x7f8f303dd7b8> True
<tensorflow.python.keras.layers.convolutional.Conv2D object at 0x7f8f303e6080> True
<tensorflow.python.keras.layers.convolutional.Conv2D object at 0x7f8f303e6e10> True
<tensorflow.python.keras.layers.convolutional.Conv2D object at 0x7f8f303ee198> True
<tensorflow.python.keras.layers.pooling.MaxPooling2D object at 0x7f8f303eeb00> True
<tensorflow.python.keras.layers.convolutional.Conv2D object at 0x7f8f303f8630> True
<tensorflow.python.keras.layers.convolutional.Conv2D object at 0x7f8f303ff4a8> True
<tensorflow.python.keras.layers.convolutional.Conv2D object at 0x7f8f303ff8d0> True
<tensorflow.python.keras.layers.pooling.MaxPooling2D object at 0x7f8f304094a8> True
<tensorflow.python.keras.layers.convolutional.Conv2D object at 0x7f8f30409c88> True
<tensorflow.python.keras.layers.convolutional.Conv2D object at 0x7f8f30392b00> True
<tensorflow.python.keras.layers.convolutional.Conv2D object at 0x7f8f30392ba8> True
<tensorflow.python.keras.layers.pooling.MaxPooling2D object at 0x7f8f3039b7f0> True
<tensorflow.python.keras.layers.pooling.GlobalMaxPooling2D object at 0x7f8f303a3320> True
```

Step 5: Connecting loaded network to output layer.

Connecting the Loaded Neural Network to the Output Layer

```
[ ]  model = models.Sequential()
     model.add(conv_base)
     model.add(layers.Dense(nb_categories, activation='softmax'))
     model.summary()
```

```
Model: "sequential"

_____
Layer (type)                 Output Shape              Param #
=================================================================
vgg16 (Functional)           (None, 512)               14714688
_____
dense (Dense)                (None, 15)                7695
=================================================================
Total params: 14,722,383
Trainable params: 14,722,383
Non-trainable params: 0
_____
```

Step6: Set Hyperparameters and creatning Keras Image Generators for training and testing.

```
#Number of images to load at each iteration
batch_size = 32# only rescaling
train_datagen =  ImageDataGenerator(rescale=1./255)
test_datagen =  ImageDataGenerator(rescale=1./255)
# these are generators for train/test data that will read pictures #found in the defined subfolders of 'data/'print('Total number of images for "training":')
train_generator = train_datagen.flow_from_directory(
  train_data_dir,
  target_size = (img_height, img_width),
  batch_size = batch_size,
  color_mode='rgb',
  class_mode = "categorical"
)
print('Total number of images for "validation":')
val_generator = test_datagen.flow_from_directory(
  val_data_dir,
  target_size = (img_height, img_width),
  batch_size = batch_size,
  color_mode='rgb',
  class_mode = "categorical",
  shuffle=False
)
```

Step7: Compile and fit the model.ss