Classification of Potato and Corn Leaf Diseases Using Deep Learning

By

Sheikh Alima Mahbub
20101517
Mayisha Siddique
20101395
Tasmia Hasan
21101325
Nazia Tasmeem
19101151
Rubaba Aziz Proma
20101606

A thesis submitted to the Department of Computer Science and Engineering in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
June 2024

© 2024. Brac University All
rights reserved.

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing a degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Sheikh Alima Mahbub
20101517

Mayisha Siddique
20101395

Tasmia Hasan
21101325

Nazia Tasmeem
19101151

Rubaba Aziz Proma
20101606

Approval

The thesis titled ”Classification of Potato and Corn Leaf Diseases Using Deep Learning” submitted by:

1. Sheikh Alima Mahbub (20101517)
2. Mayisha Siddique (20101395)
3. Tasmia Hasan (21101325)
4. Nazia Tasmeem (19101151)
5. Rubaba Aziz Proma (20101606)

Of Spring, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on May 2024.

Examinig Committee:

Supervisor:

(Member)

Dr. Md. Ashraful Alam, PhD
Assistant Professor
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:

(Member)

Md. Golam Robiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Departmental Head:

(Chair)

Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

One of the major hindrances to sustainable agriculture and an imminent threat to food security is plant disease. Constantly monitoring a plant's health and spotting the problems in it is quite painstaking because it demands a lot of work, human resources for visualization, and knowledge of plant diseases. However, deep learning can be extremely useful in the early diagnosis of plant disease, which will minimize productivity loss and help to achieve the objective of sustainable agriculture. In this study, we will use image processing of the leaves to detect plant illness using a vision-based automatic method that uses deep learning models for disease classification such as ResNet50,Densenet121, VGG-16, Inception V3 and Vision Transformers. These techniques are plant image based algorithms.

Keywords: Image processing, Deep-learning, Disease Classification, ResNet50, Densenet121, VGG-Inception, Vision Transformers

Acknowledgment

Firstly, all praise to the Great Allah for whom our thesis has been completed without any major interruption. By the grace of Allah, we were able to put in our best efforts and complete it on time successfully. Secondly, we thank our supervisor Dr. Md.Ashraful Alam for giving us this opportunity to research on this topic and assisting us through the process. We would like to express our special gratitude and appreciation to him for guiding us throughout the whole phase and giving us such attention and time.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AI Artificial Intelligence

ANN Artificial Neural Network

CNN Convolutional Neural Network

DL Deep Learning

GDP Gross Domestic Product

KNN k-nearest neighbors

ML Machine Learning

NB Naive Bayes

NN Nearest Neighbour

SVM Support Vector Machine

ViT Vision Transformer

Chapter 1

1.1 Introduction

Agriculture plays a crucial role in fostering socio-economic development in Bangladesh. Within the rural regions, over 87% of the population relies on agricultural activities as their primary source of income, often supplemented by non-agricultural wages. The main food staples of Bangladesh are rich in carbohydrates, and farmers cultivate a variety of carbohydrate-rich crops like corn and potatoes to meet the nation's dietary needs. Carbohydrates are essential for the human body, providing a multitude of health benefits. They serve as a primary source of energy, assisting in the regulation of insulin and blood sugar levels, participating in the metabolism of triglycerides and cholesterol, and aiding in the fermentation process. When carbohydrates are consumed these are broken down by the digestive system into glucose which the body uses as fuel. Any extra glucose that is stored in the muscles and liver for later use when more energy is needed. The term "carbohydrates" covers a wide range of foods including sugars, fruits, vegetables, fibers and legumes. For instance, potatoes are a carbohydrate-rich vegetable with a medium-sized potato (about 5.3 ounces with the skin) containing roughly 26 grams of carbohydrates. Similarly, corn is another important crop in Bangladesh that offers numerous health benefits making it a valuable part of the diet. Therefore, growing carbohydrate-rich foods like corn and potatoes are essential not only for the nutrition of the Bangladeshi population but also for the country's health and economic stability. These agricultural practices help rural people thrive and contribute significantly to the nation's socio-economic development.

1.2 The historical background of potato diseases

The history of early and late blight diseases in potato crops started back to the late 19th century. Early blight which is caused by the fungal pathogen Alternaria solani primarily affects the foliage, stems and tubers of potato plants. Late blight is a severe plant disease caused by Phytophthora infestans that became notorious during the Irish potato famine in the mid-19th century and impacts multiple parts of the potato plant. These diseases pose significant challenges to potato producers worldwide leading to production losses ranging from 30% to 100%, with considerable economic consequences.

1.3 The historical background of corn leaf diseases

Northern corn leaf blight (NCLB) produced problems for New England maize producers in the early 20th century. These difficulties progressively subsided in the 1930s before reappearing in the 21st century. It has been discovered that the presence of "Exsero hilumturcicum," also referred to as NCLB, significantly reduces agricultural productivity. A plant disease Gray leaf brought on by the fungus "Cercospora zeae maydis," is blamed for the decrease in maize yields observed worldwide. The damaging effects of common rust are a chronic and well-researched agricultural problem that can affect maize.

1.4 The statistical data on losses

The economic ramifications resulting from corn illnesses exhibit variability contingent upon the prevalence of the diseases, the susceptibility of maize, prevailing climatic conditions and the implementation of control methods. The absence of effective management of the Northern Corn Leaf Blight (NCLB) policy has the potential to result in a decrease in crop yields ranging from 5% to 30% or even more. The occurrence of Common Rust and Corn Cercospora leaf diseases has been seen to lead to a decrease in crop output ranging from 5% to 10%. In the year 2022, there was a decline in yield of 2.8% in the United States, while Ontario experienced a yield decrease of 2.7%. These figures were lower than the average yields observed in the preceding five-year period. For the case of early blight disease it can cause yield losses ranging from 10% to 30% or even higher in severe cases if not correctly controlled. With a global loss of $6.7 billion, Late Blight Disease (Phytophthora infestans) is the most significant potato disease commercially. Across worldwide the yearly crop losses range from 15% to 30%.

1.5 Disease management

The precise identification and categorization of plant diseases are of utmost importance in order to effectively minimize agricultural yield reductions. Preventive and mitigative measures encompass a range of tactics such as the utilization of resistant cultivars, implementation of crop rotation practices, effective residue management, judicious application of fungicides, the adoption of balanced fertilization techniques, rotation of resistance genes, removal of crop residues and careful consideration of prevailing weather conditions.

The contributions of the work are as follows: Construction of a unique dataset of images from seven different classes of potato and corn healthy and diseased leaf.
Proposal of an efficient custom CNN model that performs well to classify leaf diseases

Chapter 2

Problem Statement

The current restriction of classic Convolutional Neural Network (CNN) models is a challenge in the detection of plant leaf diseases. While these models have some benefits, they frequently fall short of offering a practical and extremely efficient way to precisely distinguish between blight infections in potato plants and leaf diseases of maize crop. Our research intends to take on this problem through providing an improved technology model that gets over the limitations of conventional approaches. Our goal is to create an innovative method that can rapidly and precisely identify hazardous diseases of leaves. Contemporary techniques such as Vision Transformers (ViT), conventional CNN models and individualized CNN architecture that have shown outstanding accuracy in other areas will be included in this methodology. In order to improve the precision and durability of the model, our approach will look into other approaches which involve data augmentation and transfer learning. Our goal is to leverage these cutting-edge methods by creating a trustworthy and precise disease identification system.

Chapter 3

Research Objectives

- **Investigate Deep Learning Models:** Evaluate the performance of a custom-designed 14-layer CNN model in detecting diseases in potato and corn crops from leaf images, compared to several other deep learning models such as ResNet50, VGG16, DenseNet121, Inception V3, and Vision Transformer (ViT).

- **Assess Accuracy and Efficiency:**Analyze the accuracy of these models while considering their computational efficiency and overall robustness. The analysis will determine which model provides the best accessibility and serviceability to the practical implementation in the agricultural world.

- **Develop a Diverse and Comprehensive Dataset:** Generate a dataset which holds different categories of images of potato and corn leaves with the label of disease name. This dataset has been used to train and assess the validity of the models that has been used.

- **Optimize Model Performance:** Adjusted the hyperparameters and the architecture of the models to reduce the false positives and false negatives to improve the accuracy of models in general.

- **Test and Implement on Real-World Data:** To recognize the efficiency of the models in the real agricultural environment we applied it to authentic data. Besides, appraise their performance and usefulness for improving the farmers and others in the industrial part of agriculture.

- **Provide a Deployment Framework:** deploy the most useful and effective model which gives more accuracy through the phase of testing to expand a system which will be user-friendly and helpful to diagnose the diseases earlier and create sustainability to the work of the farmers.

By advancing these objectives our aim is to enhance the work stability of the farmers via introducing an application of deep learning in the sector of agriculture. This will provide a reliable apparatus to identify the diseases of potato and corn leaves earlier.

Chapter 4

Literature Review

The potato is a highly significant agricultural crop cultivated by farmers in many regions worldwide encompassing diverse weather conditions. Given its status as a primary food source for a significant portion of the global population the yield and production of this particular resource are subjects of considerable importance and attention. Late Blight and Early Blight diseases are widely recognized as two of the most widespread ailments that afflict potato plants. Among these two Late Blight disease is particularly notorious for its heightened lethality, resulting in significant global losses. The potato plant affliction known as Early Blight is mostly caused by the fungal pathogen Alternaria solani, resulting in detrimental effects on the foliage, stems, and tubers of potato plants. The more severe Late Blight disease also impacts analogous regions of the plant resulting in a blackened appearance and the presence of warty lesions. The etiological agent responsible for the occurrence of disease is the pathogen known as Oomycete Phytophthora infestans. The imperative to regulate and identify innovative approaches for mitigating potato plant diseases arises from the potential for a severe and abrupt reduction in annual potato production which can lead to hunger as evidenced by the historical Irish Potato hunger in the 1840s. Farmers prioritize early sickness identification as it enables them to implement necessary steps to prevent the spread of diseases among their crops. However, farmers encounter difficulties in visually discerning these two diseases. The good news is that recent advancements in the fields of computer vision and deep learning have provided novel opportunities for the timely detection and diagnosis of the aforementioned diseases that impact potato plants. Similarly, there are several corn leaf diseases that are also important to discover early as have a significant portion of carbohydrate nutrient. The corn can readily mitigate hunger as they are easy to serve quickly.

This study illustrates the methodology employed for infecting potato leaves through the utilization of image analysis. The methodology encompasses the sequential stages of image segmentation, feature extraction, and disease categorization. The segmentation technique demonstrates a mean intersection over union (IoU) of 93.70% across the five classes. The four classifiers used for sickness recognition particularly k-NN, SVM, ANN, and RF, exhibit high overall accuracies. Among these classifiers SVM achieves the highest accuracy rate of 97.4%. The Support Vector Machine (SVM) algorithm obtained a high accuracy of 91% in classifying the severe disease. The application of the SVM classifier has yielded significant betterment in the classification performance resulting in an overall accuracy of 92.2%. [1] The improvement in this study was achieved through the utilization of a finely chosen set of features. Thus, the recommended strategy provides farmers with a precise and practical method to monitor and help eradicate crop diseases in potato cultivation. This study presents an object recognition model using deep learning techniques to efficiently detect potato blight disease which has a significant risk to global food security. The model deployed in this study utilized a deep convolutional neural network architecture. The network was trained on a total of 2,152 pictures from open source dataset including both healthy and diseased leaves. Using a data splitting of 60% for training, 20% for testing, and 20% for validation the optimal detection resulted in an accuracy of 99.31% and a confidence level of 99%. The proposed model exhibits a notable level of accuracy. The paper acknowledges the potential issue of overfitting when analyzing the outcomes of different train-test splits. The usability of the model renders it appropriate for future deployment in Android applications with the aim of aiding farmers in efficiently recognizing and managing diverse crop plant diseases.[2] The researchers explore a range of deep learning (DL) models, including AlexNet, GoogLeNet, and VGGNet, alongside machine learning (ML) techniques such as SVM, NN, KNN, and NB, in order to analyze photos of plant leaves. The research highlights the efficacy of these techniques in accurately detecting and categorizing plant diseases including a particular emphasis on the potential of mobile-based apps to enhance agricultural productivity. In general, the study report offers a comprehensive examination of the cutting-edge methodologies being utilized in this domain and presents suggestions for future investigation.[3] The present research uses deep convolutional neural networks to provide a possible classification architecture for potato leaf blight. The fully linked layers are utilized to integrate diversity into the design and convolutional layers are used to extract features. By using data augmentation techniques to increase the amount of the image collection, testing accuracy is significantly improved. In comparison to previous studies in this field, the suggested methodology acheived an overall mean evaluation accuracy of 98%. This study shows how deep learning and other artificial intelligence technologies can improve plant protection and disease prevention.[4] Early identification is crucial in mitigating the impact of plant diseases on the growth of plants. Despite the existence of multiple machine learning (ML) models utilized for the purpose of diagnosing and categorizing diseases, advancements in deep learning (DL) have demonstrated potential in terms of enhancing accuracy. This review places emphasis on the evaluation of DL models and visualization techniques for the purpose of visualizing plant diseases while also considering the utilization of performance metrics. This statement highlights the need of precise disease detection in plants even in the absence of visible symptoms and outlines areas where further research is needed. The study examines the utilization of established deep learning architectures such as AlexNet and GoogLeNet, as well as novel or modified models in order to demonstrate their potential in the detection and classification of plant diseases.comparable studies involve focusing on the implementation of visualization methods such as heat maps to investigate and visibility.[5] The research project examines the practical use of machine vision and image processing methods to agricultural products, with an emphasis on potatoes, in order to detect defects a convolutional neural network (CNN) was implemented by a team of researchers to classify five different potato diseases. They analyzed their findings compared to those from other methods. Through expanding the precision of their deep learning methodology the group accomplished remarkable success rates of 100% and 99% for specific disease classifications. This study demonstrates the potential applications of AI and image processing in the agricultural sector, in the detection and categorization of disease and pests.[6] To detect early indications of potato blight in actual farming conditions this research used deep learning and automated vision methods. Using convolutional neural networks (CNNs) researchers divided the nineteen potato plants some healthy and some diseased—into groups of two, four, or six. Testing the CNNs accurately identified early stages of blight illness EfficientNet beats other models. Based on the results of this research, farmers who successfully recognize plants with early blight infections may benefit financially from using lesser agrochemicals.[7] Convolutional neural networks is a kind of deep learning technology which are used by the suggested model to interpret images of potato leaves (CNN). CNN can determine whether the leaves exhibit symptoms of early or late blight two prevalent diseases that affect potatoes. On a dataset developed explicitly for the intent of determining potato leaf diseases the model achieved an excellent accuracy rate of 99.75% outperforming previous techniques in terms of accuracy and computing performance. The model further includes YOLOv5 techniques for image segmentation to further maximize accuracy by detecting the potato leaves in pictures from the rest of the plant.[8] This research study examines the importance of crop monitoring and smart farming in the field of agriculture specifically highlighting the role played by disease monitoring algorithms based on artificial intelligence. The objective of this study is to develop a disease detection model using deep learning techniques for bell peppers, potatoes, and tomatoes which are three significant agricultural products. The present model employs a multi-stage approach to gain detailed insights on both the individual crop and the type of illness which stands in contrast to previous research endeavors that aggregated diseases affecting specific crops under a single category. The accuracy of classification models is evaluated using pre-trained Convolutional Neural Network (CNN) models with EfficientNet consistently demonstrating the highest level of accuracy. Additional experimentation is conducted with non-model crops, resulting in a notable fall in accuracy. However, the inclusion of non-model crops in the training datasets leads to an improvement in accuracy. The study also examines Based on research examining the classification model's ability to detect the same disease in many crops it has been found that early blight may be predicted with greater accuracy compared to late blight. Based on the analysis conducted it is evident that the characteristics of crop diseases hold significant importance in the classification of diseases. The essay's conclusion emphasizes the possible applications of the disease detection model in phenotypic research, seed purity testing, and crop development. Furthermore, this emphasizes the potential application of CNN analysis in the examination of plant phenotypes and the implementation of intelligent farming techniques, thereby facilitating automated data collection and event identification within the agricultural industry. To enhance the classification model and establish a robust framework for its practical implementation, it is suggested that the study incorporates a greater number of high-quality photographs depicting diverse crops.[9] The paper examines corn's role in the food chain, its many applications, including its usage in biofuels, as well as its significance for small-scale farming in underdeveloped areas. However, it draws attention to how susceptible maize harvests are to illnesses, which are made worse by extreme weather. It is well known that the development of technology, especially artificial intelligence, has the ability to solve these problems. The study focuses on classifying three common maize illnesses (Cercospora leaf spot, common rust, and northern leaf blight) together with healthy plants using corn leaf photos as input using deep transfer learning with convolutional neural networks. Due to the efficiency of deep learning models, preprocessing and feature extraction are not required. The research's thorough analysis, which included numerous data splitting situations and repeated testing, produced an amazing result.These findings raise the possibility of creating software that can help plant pathologists and farmers recognize corn infections and treat them effectively.[10] Convolutional neural networks, particularly VGG-16, were used in this paper to forecast the severity of the prevalent rust disease in maize. In order to calculate the proportion of leaf area impacted by the disease, they developed a novel method incorporating picture segmentation. The researchers developed fuzzy decision methods for categorizing disease photos into Early stage, Middle stage, Late stage, and Healthy stage using these percentages. They built a VGG-16 neural network to automate this process, with outstanding validation (95.63%) and testing (89%) accuracy. In conclusion, this work proves the efficiency of AI, particularly CNNs, in plant pathology and provides insightful information for managing disease in maize crops.[11] Deep convolutional neural networks (DCNNs) are used in this study to identify Northern corn leaf blight (NCLB), a damaging maize foliar disease, Through the use of data augmentation techniques, the researchers were able to increase a database that originally contained 985 photos of both healthy and sick maize leaves to 30,655 images. To locate the NCLB in the maize(corn) plants, a variety of Deep CNN models, including denseNet, ResNet50, VGG16, and InceptionV3, were used. The best performance was achieved by the DenseNet121 architecture with the Softmax loss function, which had a remarkable accuracy rate of 95.19% for NCLB diagnosis. The Pytorch and Keras deep learning frameworks were implemented by the researchers using Python.[12]

Chapter 5

5.1 Methodology

The current study adopts an approach that involves the evaluation and comparison of different deep learning models, namely Vision Transformers (ViTs), ResNet50, DenseNet121, VGG-16, Inception-V3 and a 14-layer customized CNN model. In order to assess their effectiveness in accurately identifying diseases in potato and corn (maize) plants. The objective of this study is to assess the efficacy of various models in differentiating between healthy and diseased potato and corn plant leaf images across seven distinct categories, namely: Early Blight potato disease, Late Blight Potato Disease, Corn Gray leaf spot, Corn Common rust, Corn Northern Leaf Blight, Healthy Potato, and Healthy corn. Vision Transformers (ViTs) are a specific type of deep learning models that utilize the transformer architecture, originally designed for natural language processing, to address computer vision tasks. Visual Transformers (ViTs) differ from conventional Convolutional Neural Networks (CNNs) in their approach to picture interpretation. While CNNs have historically been the prevailing architecture in computer vision tasks, ViTs interpret images as sequences of patches. This allows them to encompass both the local and global context inside one image. Given that ViTs are a relatively infrequently employed model for the detection of blight disease in potatoes, this study aims to assess its accuracy in comparison to the more widely utilized models. The models ResNet50, DenseNet121, VGG-16 and Inception V3 are commonly used in the field of computer vision for various tasks such as image classification and object detection. Along with these traditional CNN models we are going to implement a lightweight custom CNN model consisting 14 layers. The method of disease detection involves three essential components: data acquisition, categorization and prediction.

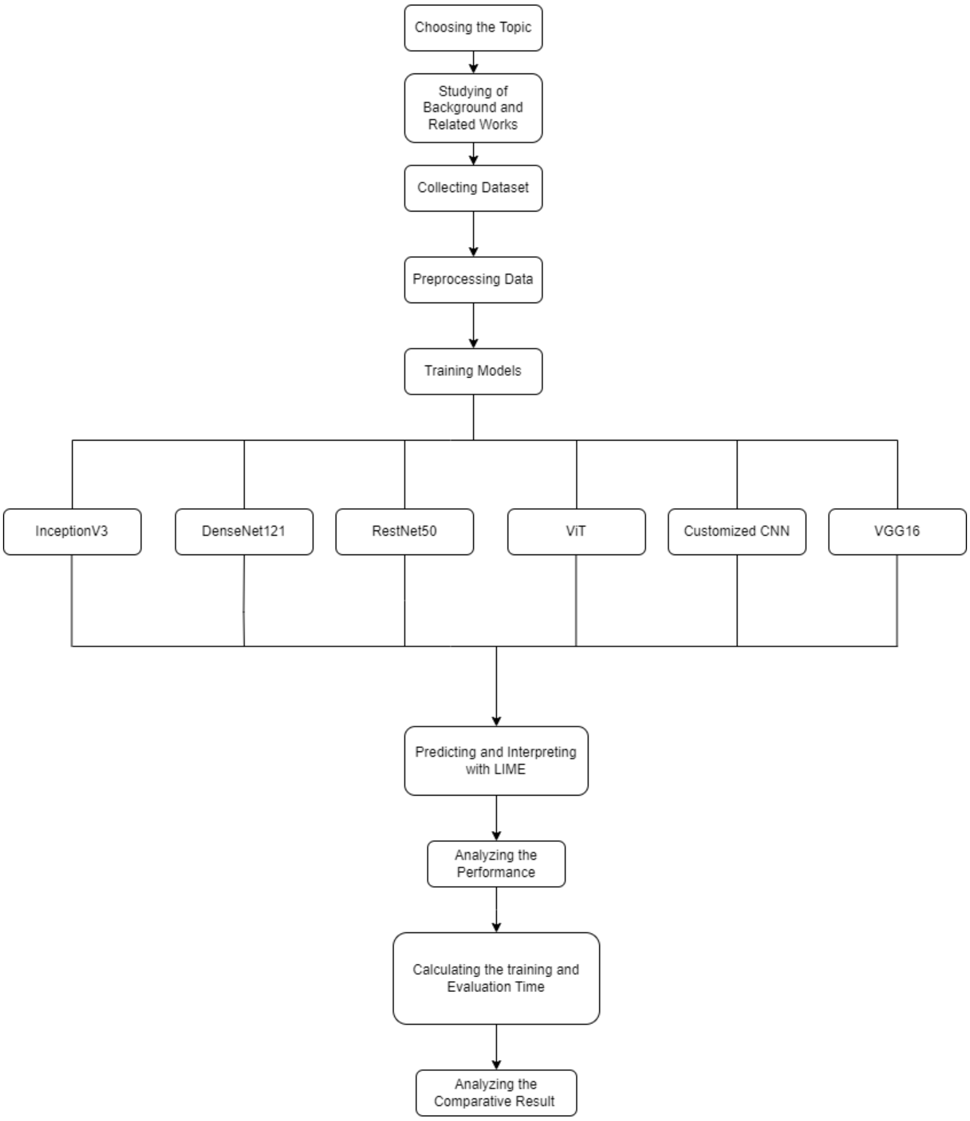

Figure 1: Proposed Workplan

5.2 Data Acquisition

The dataset was obtained from the "Plant Village" website [22]. The dataset comprises images that have been categorized into seven distinct classes, namely Corn Gray leaf spot, Corn Common rust, Corn Northern Leaf Blight, Healthy Potato, Healthy corn, Potato Early blight and Potato Late blight. A total of 6004 images were distributed throughout these seven categories. Of these categories 3 classes belong to corn leaf diseases, 1 class to corn healthy leaf, 2 classes to potato leaf diseases and 1 class to potato healthy leaf.

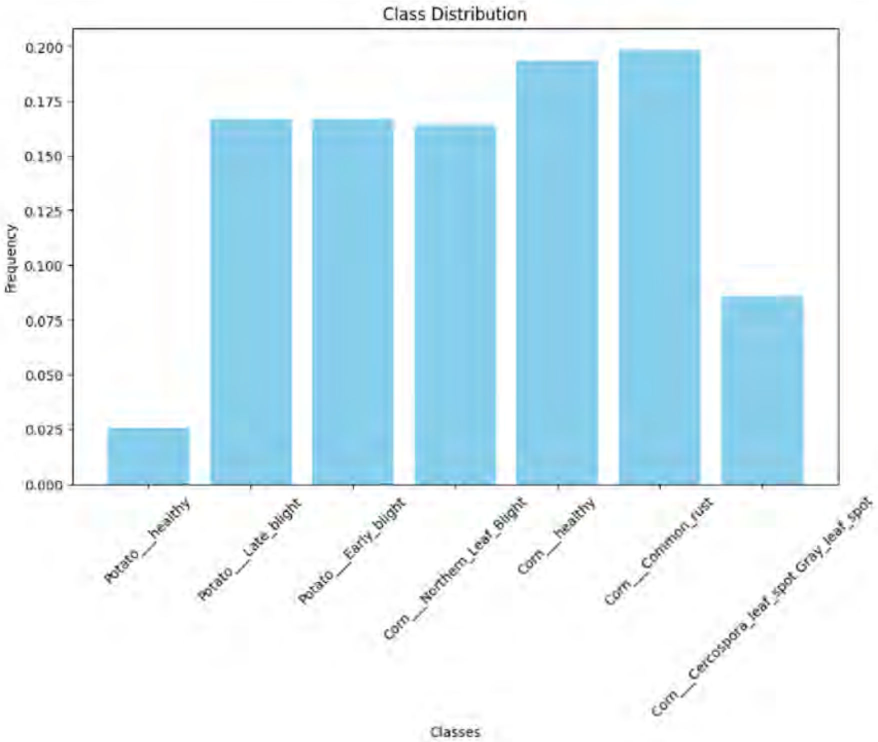

Figure 2: Unbalanced Dataset

5.3 Data Augmentation for Balancing the Dataset

Data augmentation is a methodology used in machine learning where the training dataset is artificially enlarged by implementing diverse modifications on the pre-existing images. This facilitates the enhancement of the model's capacity to generalize across diverse changes of the data. In this context, the ImageDataGenerator class provided by the Keras API in TensorFlow is employed for the purpose of data augmentation. The CNN classifiers, specifically ResNet50, DenseNet121, VGG-16, Inception V3 and the custom model were subjected to various image augmentation techniques including:

- **Rotation:** Images were rotated within a range of 20 degrees.
- **Width Shift:** Images were shifted horizontally within a range of 20%.
- **Height Shift:** Images were shifted vertically within a range of 20% of the height.
- **Shear:** Images were sheared within a range of 20 degrees.
- **Zoom:** Images were zoomed in and out within a range of 20%.
- **Horizontal Flip:** Images were randomly flipped horizontally.
- **Fill Mode:** New pixels generated by the above transformations were filled using the nearest pixel values.

The dataset for the 7 classes has been balanced through data augmentation. The target was to balance the dataset by having roughly the same number of images in each class. In particular, the target number of images per class was set to almost (10000)/7 which is approximately 1428. Here, the initial unbalanced dataset has been balanced later on through data augmentation in the next image fig. The bar chart

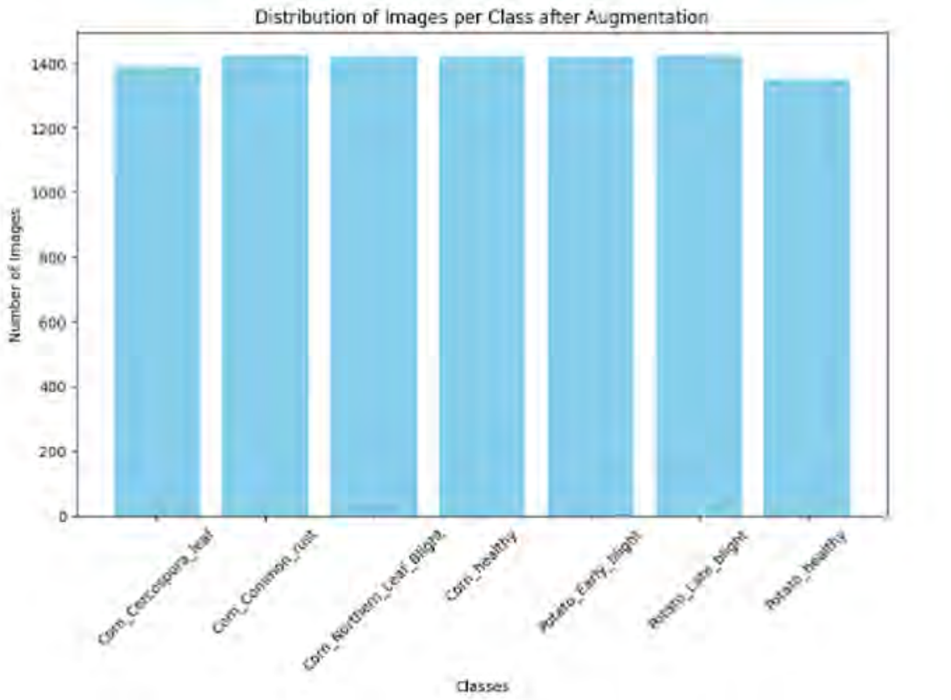

Figure 3: Balanced Dataset

shows the distribution of images per class after augmentation. Each class has approximately 1428 images, indicating a successful balancing of the dataset. The x-axis and y-axis represent the different classes the number of images in each class respectively.

5.4 Data Pre-Processing

The pre-processing is an crucial step in preparing data for training convolutional neural networks (CNNs). It ensures that the input data is in a suitable format for the model and often includes several steps such as resizing, normalization and augmentation. Here, we outline the pre-processing steps employed for training various CNN models (InceptionV3, ResNet50, DenseNet, VGG16 and custom model) along with the ViT model in our study.

Loading and Preprocessing Images which is done through images reading from the file system using TensorFlow's **tf.io.read_file** function which reads the entire contents of the file as a string. Then comes decoding the images where the string content is then decoded into a numeric tensor representing the image with three color channels (RGB). After that the decoded image is resized to a standard size 224 x 224 pixels ensuring a consistent input shape across the dataset. This size is chosen because it is a common input dimension for many pre-trained CNN architectures. Then comes the type casting data type to match the expected input type for TensorFlow models. After that model specific preprocessing is done such as
For InceptionV3: to scale pixel values to the range [-1, 1].
For ResNet50: to normalizes pixel values by subtracting the mean and scaling according to ImageNet standards. DenseNet to pixel values similarly to ResNet50. VGG16 to normalizes pixel values to the range [0, 255]. After that the label for each image is extracted from the file path. The file path is split and the parent directory name, which represents the class label is retrieved. Data augmentation is applied to increase the diversity of the training data and improve the model's generalization ability. The entire dataset is divided into training, validation, and testing subsets. TensorFlow's tf.data.Dataset API is used to efficiently handle large datasets enabling operations such as shuffling, batching and prefetching.

The dataset is split into three parts: training (80%), validation (10%), and testing (10%).
For ViT Model preprocessing: In Data Augmentation for training Only these techniques were used such as

- **Random Horizontal Flip:** Flips the image with a 0.5 probability.
- **Random Rotation:** Rotates the image within a range of -20 to 20 degrees.
- **Color Jitter:** Adjusts brightness, contrast, saturation, and hue.

After that normalization was done that normalizes pixel values to have a mean of [0.485, 0.456, 0.406] and a standard deviation of [0.229, 0.224, 0.225]. Finally converts images into PyTorch tensors.

Chapter 6

6.1 Model Architecture

The model architecture of the traditional CNN model, Vision Transformer model and the Custom CNN models have been described here. In the pretrained CNN models some extra layers of global average pooling, dropout and dense layer was added in order to enhance the model performance on the existing dataset.

Model Architecture of ResNet50

The model utilizes the ResNet50 architecture for feature extraction, followed by a global average pooling layer to reduce dimensionality. A dropout layer is added for regularization, and the final dense layer outputs the class probabilities. This combination leverages the powerful feature extraction capabilities of ResNet50 while reducing the risk of overfitting and providing a final classification output.

- **ResNet50 (Functional):** The base model used is ResNet50, a powerful convolutional neural network pre-trained on the ImageNet dataset.
- **Output Shape:** The output shape from ResNet50 is (None, 7, 7, 2048). This tensor represents the feature maps extracted by the ResNet50 model.
- **GlobalAveragePooling2D:** A global average pooling layer is applied to the output of ResNet50. This layer reduces the spatial dimensions (7x7) by computing the average of each feature map.
- **Output Shape:** The output shape after this layer is (1, 2048), which significantly reduces the dimensionality while preserving the most important features.
- **Dropout:** A dropout layer is added to prevent overfitting. It randomly sets a fraction of the input units to zero at each update during training time.
- **Output Shape:** The output shape remains (1, 2048), as dropout does not alter the shape but only deactivates some neurons temporarily.
- **Dense:** The final dense layer is a fully connected layer with 7 units, corresponding to the 7 classes in the classification task.
- **Activation Function:** Typically, this layer would use a softmax activation function to output probabilities for each class.
- **Output Shape:** The output shape is (1, 7), representing the probability distribution over the 7 classes.

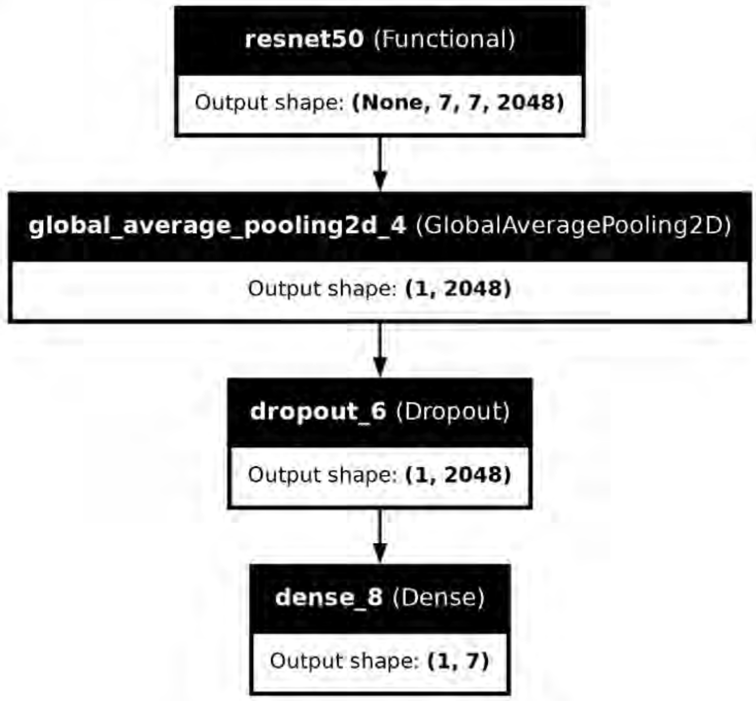

Figure 4: Resnet50 Model

Model Architecture of DenseNet121

The model uses a global average pooling layer to minimize dimensionality after using the DenseNet121 architecture for feature extraction. The class probabilities are output by the final dense layer after a dropout layer is added for regularization.

This combination lowers the chance of overfitting, makes use of DenseNet121's potent feature extraction capabilities, and produces a final classification output. DenseNet121, a convolutional neural network that has been pre-trained on the ImageNet dataset, serves as the foundational model.

- **Output Shape:** The output shape from DenseNet121 is (None, 7, 7, 1024). This tensor represents the feature maps extracted by the DenseNet121 model.
- **GlobalAveragePooling:** A global average pooling layer is applied to the output of DenseNet121. This layer reduces the spatial dimensions (7x7) by computing the average of each feature map.
- **Output Shape:** The output shape after this layer is (1, 1024), which significantly reduces the dimensionality while preserving the most important features.
- **Dropout:** A dropout layer is added to prevent overfitting. It randomly sets a fraction of the input units to zero at each update during training time.
- **Output Shape:** The output shape remains (1, 1024), as dropout does not alter the shape but only deactivates some neurons temporarily.
- **Dense:** The final dense layer is a fully connected layer with 7 units, corresponding to the 7 classes in the classification task.
- **Activation Function:** Typically, this layer would use a softmax activation function to output probabilities for each class.
- **Output Shape:** The output shape is (1, 7), representing the probability distribution over the 7 classes.

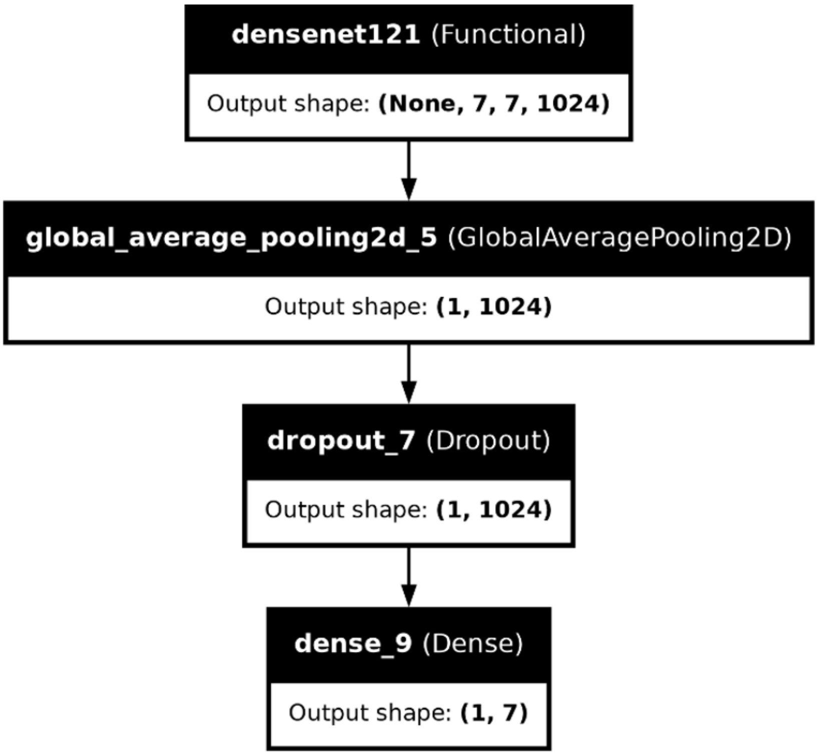

Figure 5: Densenet121 Model

Model Architecture of VGG-16

- **VGG16 (Functional):** This is the base model, VGG-16, pre-trained on ImageNet.
- **Output Shape:** (None, 7, 7, 512), meaning for each input image, it outputs a feature map of shape 7x7 with 512 channels.
- **Global Average Pooling 2D (global_average_pooling2d_6):** This layer performs global average pooling on the feature map, reducing the spatial dimensions while keeping the depth.
- **Output Shape:** (1, 512), compressing the 7x7 spatial dimensions into a single vector of 512 features.
- **Dropout (dropout_8):** This dropout layer helps in regularizing the model by randomly setting a fraction of input units to 0 at each update during training time, which helps prevent overfitting.
- **Output Shape:** (1, 512), remains the same as the input to this layer.
- **Dense (dense_10):** This fully connected (dense) layer is the output layer with 7 units, corresponding to the 7 classes for classification.
- **Output Shape:** (1, 7), where each unit represents the probability of the input image belonging to one of the 7 classes.

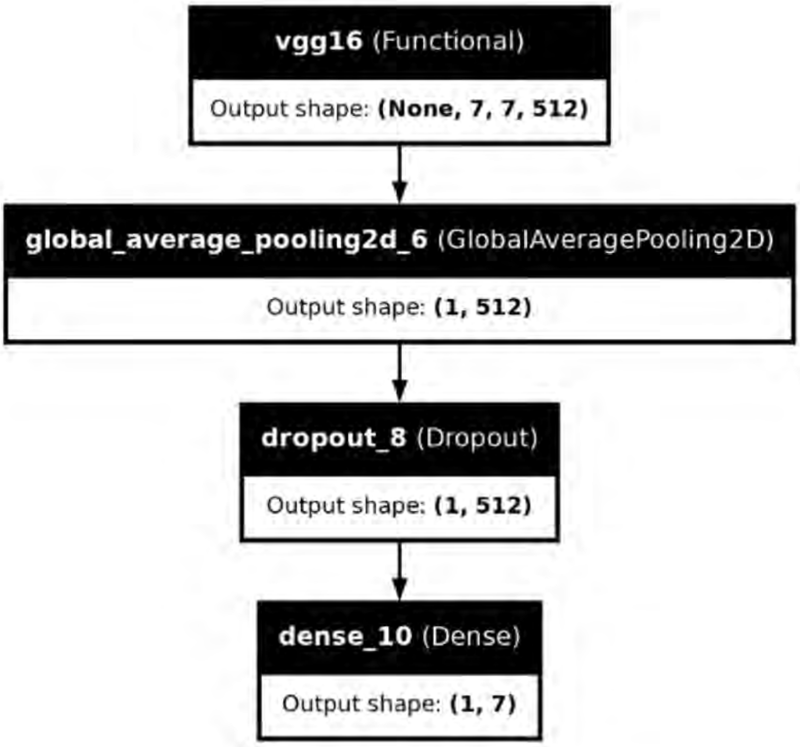

Figure 6: VGG-16 Model

Model Architecture of Inception V3

- **Inception_v3 (Functional):**
  Output shape: (None, 5, 5, 2048): This is the output from the InceptionV3 model which processes input images and outputs feature maps of shape (5, 5, 2048) for each input image in the batch (None indicates the batch size).

- **Global_average_pooling2d_7 (GlobalAveragePooling2D):**
  Output shape: (1, 2048) This layer performs global average pooling on the feature maps, reducing each feature map to a single value. The output is a 1D tensor with 2048 features.

- **Dropout_9 (Dropout):**
  Output shape: (1, 2048) This dropout layer is used to prevent overfitting by randomly setting a fraction of input units to 0 at each update during training time.

- **Dense_11 (Dense):**
  Output shape: (1, 7) This fully connected (dense) layer is the output layer of the model. It maps the 2048-dimensional input to 7 output classes corresponding to the 7 classes in the classification task.

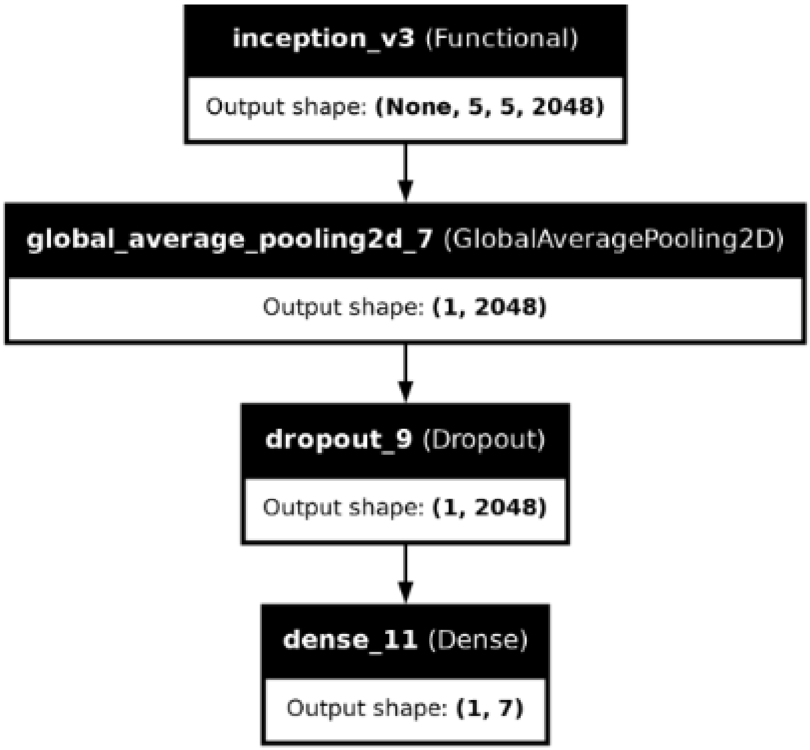

Figure 7: Inception V3 Model

Model Architecture of Custom Model

- Conv2D (32 filters, 3x3 kernel): Convolution operation and Activation (ReLU)
- MaxPooling2D (2x2 pool size): Max pooling operation
- Conv2D (64 filters, 3x3 kernel): Convolution operation and Activation (ReLU)
- MaxPooling2D (2x2 pool size): Max pooling operation
- Conv2D (128 filters, 3x3 kernel): Convolution operation and Activation (ReLU)
- MaxPooling2D (2x2 pool size): Max pooling operation
- Conv2D (256 filters, 3x3 kernel): Convolution operation and Activation (ReLU)
- MaxPooling2D (2x2 pool size): Max pooling operation
- Flatten: Flatten operation
- Dense (512 units): Fully connected layer and Activation (ReLU)
- Dropout: Dropout operation
- Dense (7 units): Fully connected layer (output layer) and Activation (Softmax)

Total Layers: 4 (Conv2D + Activation) + 4 (MaxPooling2D) + 1 (Flatten) + (2 Dense + 2 Activation) + 1 (Dropout)

This sums up to 14 layers when considering all operations individually.

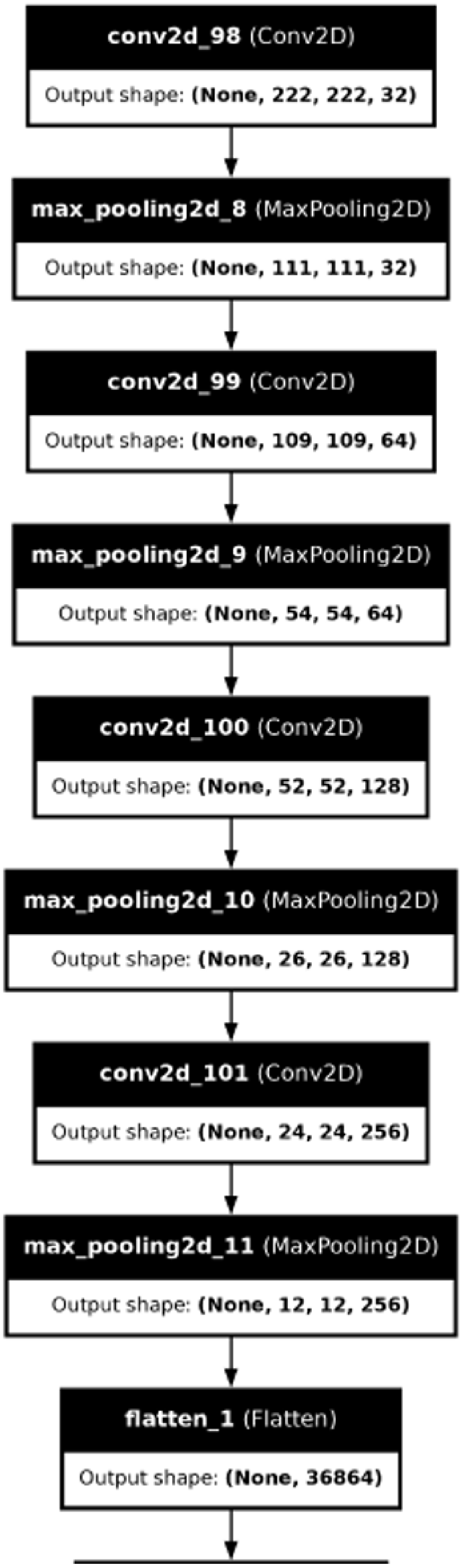

Figure 8: Custom Model

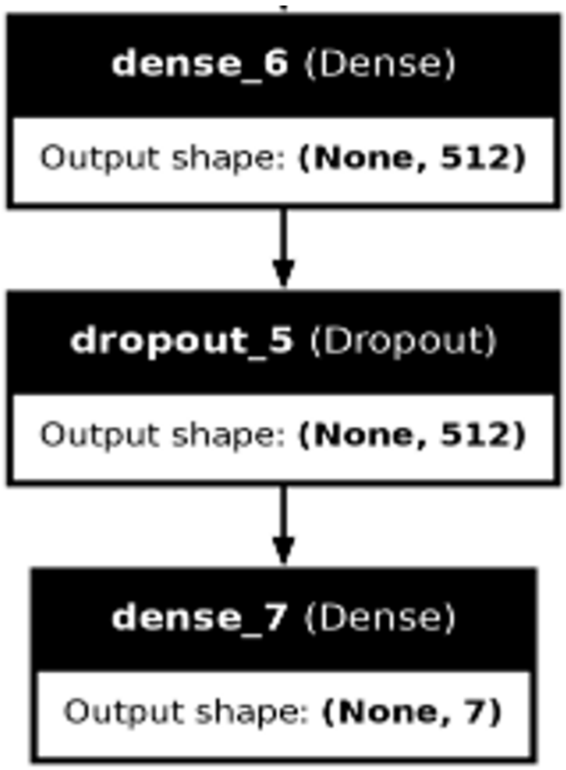

Figure 9: Custom Model

Model Architecture of ViT Model

- **ViT (Functional):** Output shape: (None, 224, 224, 3) this indicates the input to the Vision Transformer which consists of images with dimensions 224x224 and 3 color channels (RGB).
- **Patch_embedding:** Output shape: (None, 196, 768) This layer divides the input image into patches and embeds each patch into a 768-dimensional vector. The image is divided into 196 patches.
- **Positional_encoding:** Output shape: (None, 197, 768) This layer adds positional information to the patch embeddings. An additional token is included for the classification task resulting in 197 tokens.
- **Transformer_blocks:** Output shape: (None, 197, 768) This module consists of several transformer layers that process the embedded patches and positional encodings to capture relationships between patches.
- **Cls_token:** Output shape: (None, 1, 768) The classification token (`cls_token`) is extracted from the output.
- **mlp_head:** Output shape: (None, 7) This is the final classification layer that maps the 768-dimensional `cls_token` output to 7 classes corresponding to the classification task.

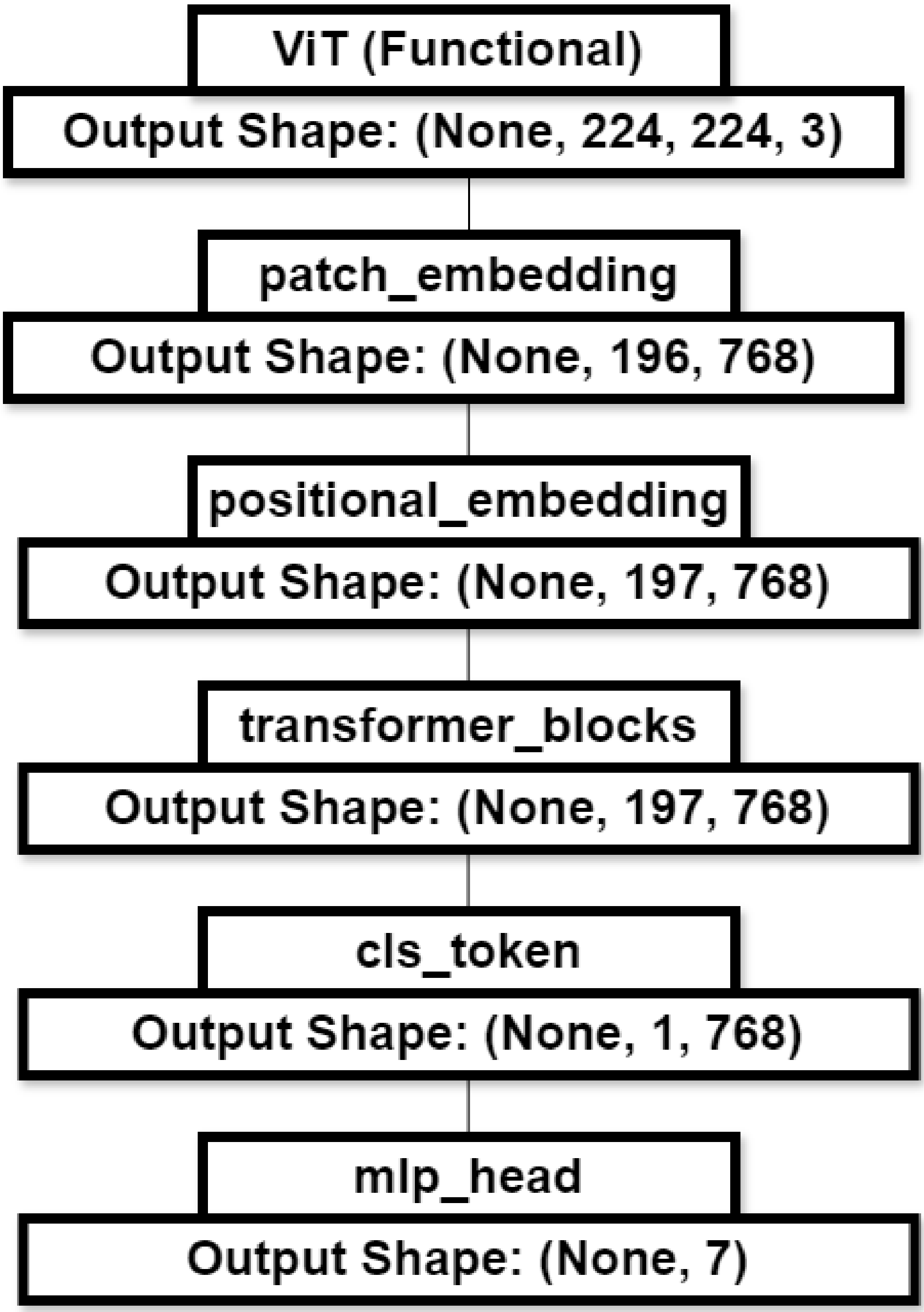

Figure 10: ViT Model

Chapter 7

7.1 Training Process and Performance Evaluation

ResNet50 Model

The ResNet50 model demonstrated excellent performance, achieving high accuracy (98%) on the test set with consistent precision, recall, and F1-scores across all classes. The training process was efficient, with a total time of 422.31 seconds for 20 epochs. The steady improvement in validation accuracy and reduction in loss further confirms the model's robustness and generalization capability.

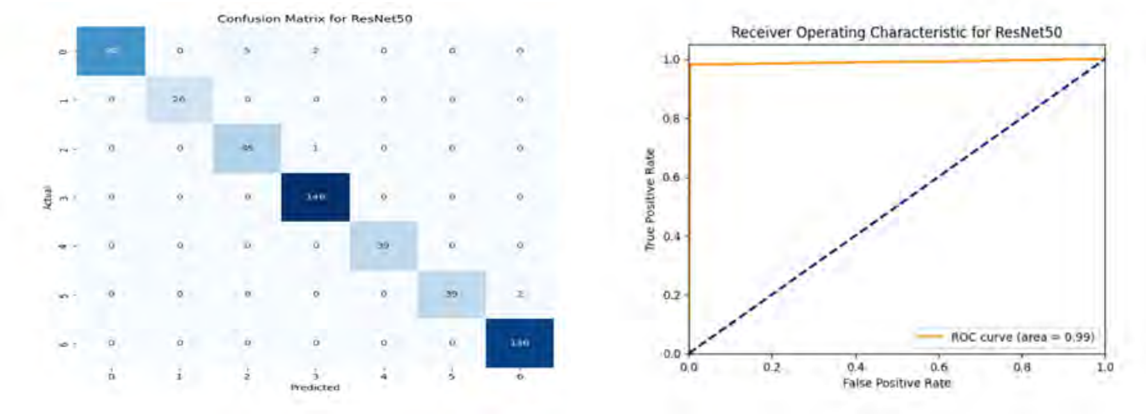

Figure 11: Confusion Matrix (Left) and ROC (Right)

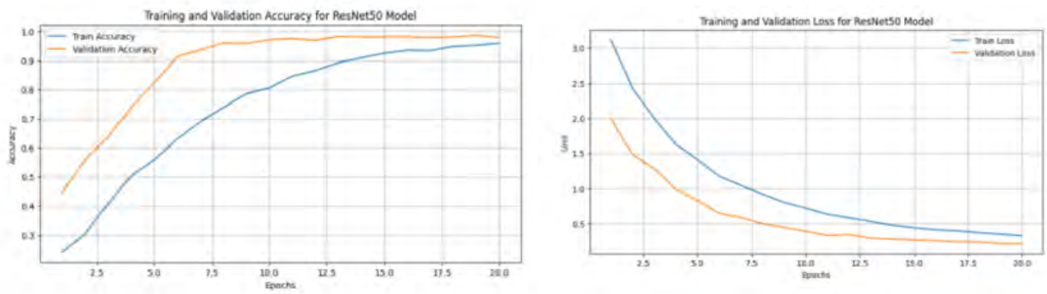

Figure 12: ResNet50 Model Training Accuracy (Left) and Loss (Right)

DenseNet121 Model

Over the course of 20 epochs the DenseNet121 model was trained; the training process took about 415.23 seconds per epoch. From 0.0994 to 0.6087, training accuracy increased steadily, and validation accuracy rose from 0.0866 to 0.7458, suggesting that generalization and learning were successful. The AUC of 0.85 on the ROC curve indicates good performance in differentiating across classes. The classification report shows the model's performance across many parameters with an overall accuracy of 0.75, weighted precision of 0.72, recall of 0.75 and F1-score of 0.69.

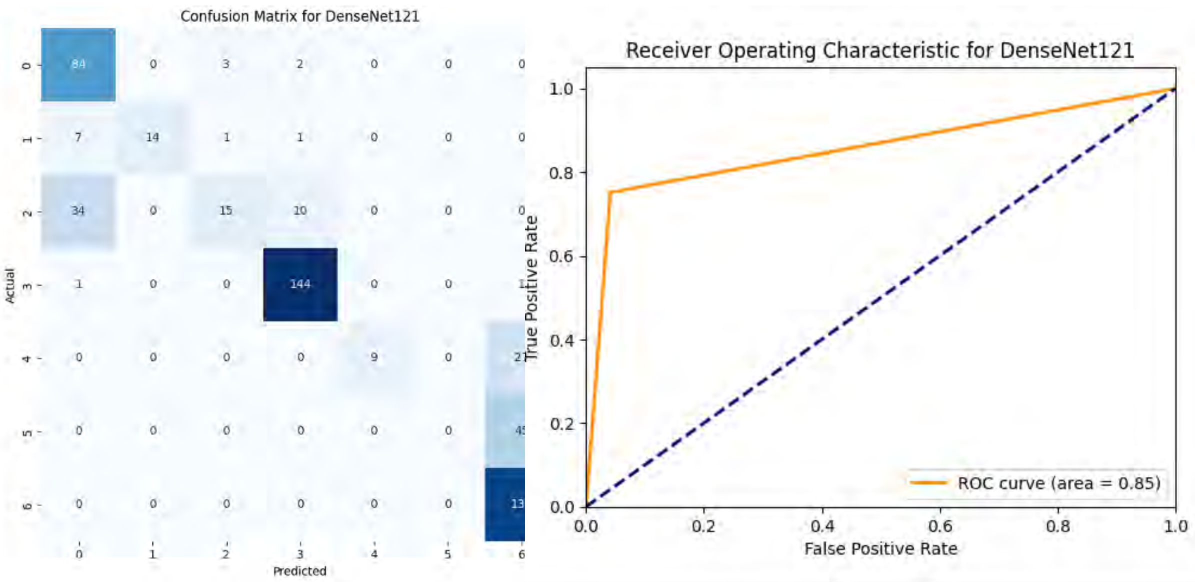

Figure 13: Confusion Matrix (Left) and ROC (Right)

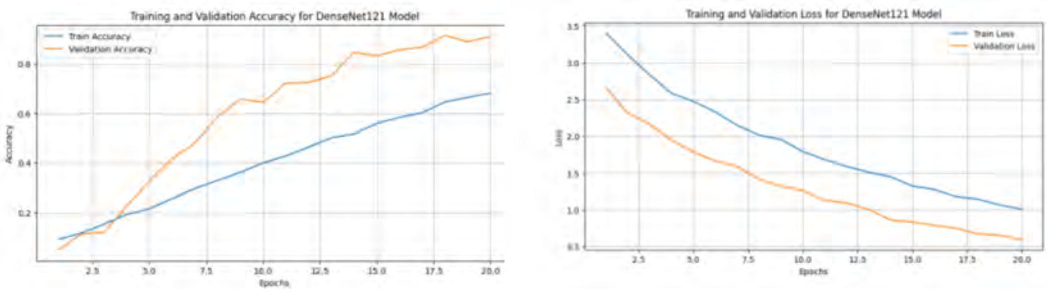

Figure 14: DenseNet121 Model Training Accuracy (Left) and Loss (Right)

VGG-16 Model

The VGG-16 model was trained over 20 epochs taking around 547.17 seconds approximately for the entire training process. Training accuracy improved from 0.0112 to 0.3627 and validation accuracy increased from 0.0433 to 0.6352 indicating that the model is effective learning and has done generalization. The ROC curve with an AUC of 0.79 suggests decent performance in distinguishing between the healthy and disesed classes. The classification report shows an overall accuracy of 0.64, with a weighted precision of 0.66, recall of 0.64 and an F1-score of 0.63 representing the model's performance across various metrics.

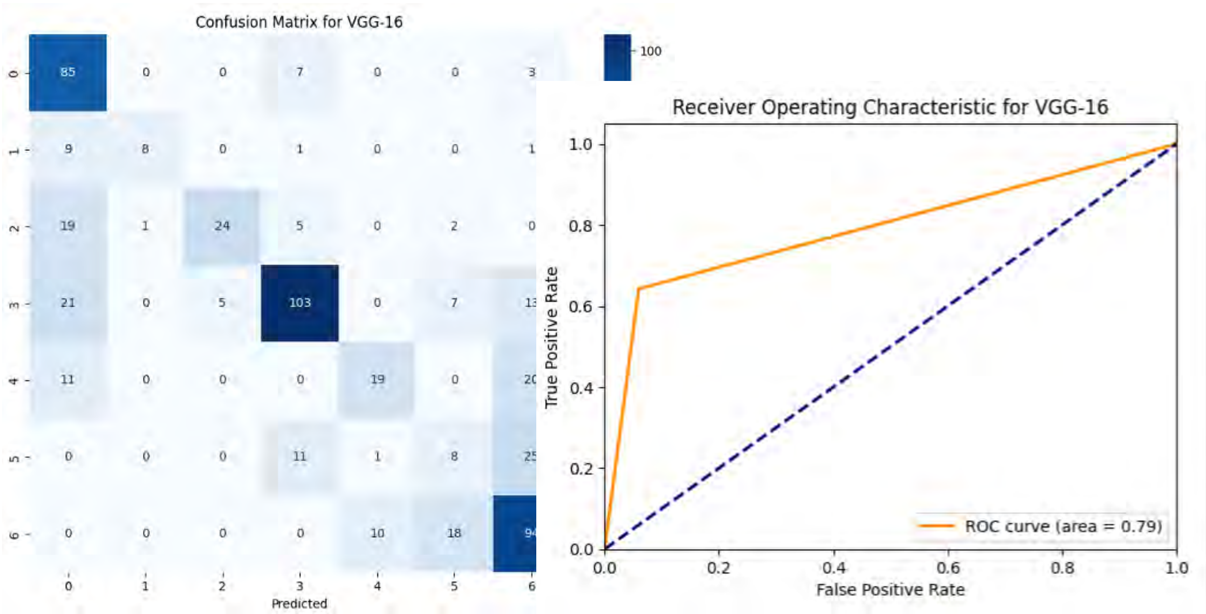

Figure 15: Confusion Matrix (Left) and ROC (Right)

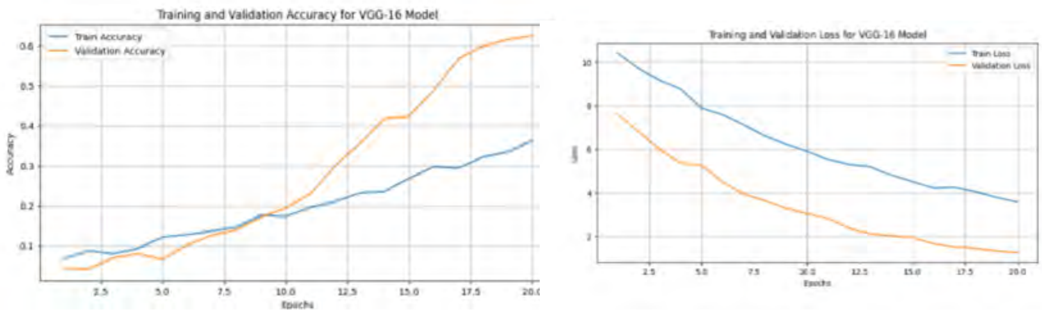

Figure 16: VGG-16 Model Training Accuracy (Left) and Loss (Right)

ViT Model

The ViT model was trained over 10 epochs with an almost consistently decreasing training and validation losses which indicates an effective learning and generalization. The model achieved a fine classification metrics with an overall accuracy of 0.99. The total training time was approximately 1952.57 seconds for 25 epochs showcased the model's efficiency.

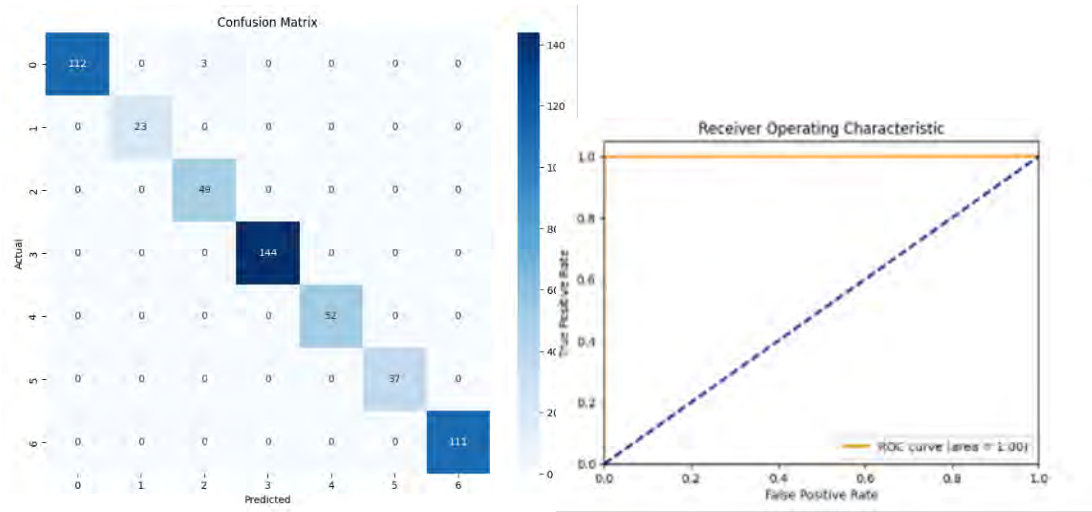

Figure 17: Confusion Matrix (Left) and ROC (Right)

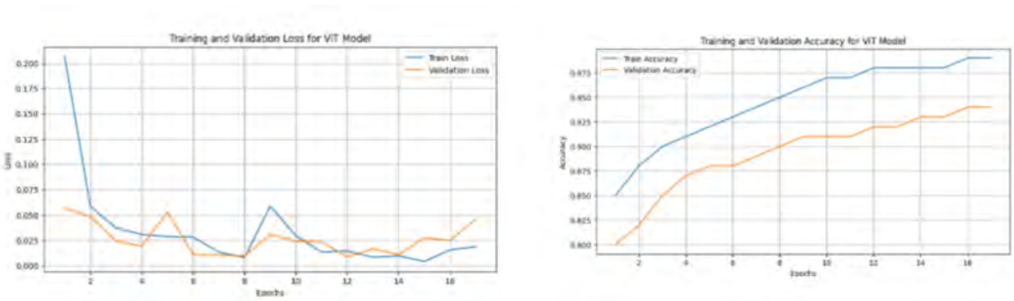

Figure 18: ViT Model Training Accuracy (Left) and Loss (Right)

Inception V3 Model

The Inception V3 model was trained for 20 epochs with a total training time of 503.45 seconds this shows effective learning and improved generalization. The model achieved high classification metrics with an overall accuracy of 0.96 and the ROC curve indicates an AUC of 0.98.

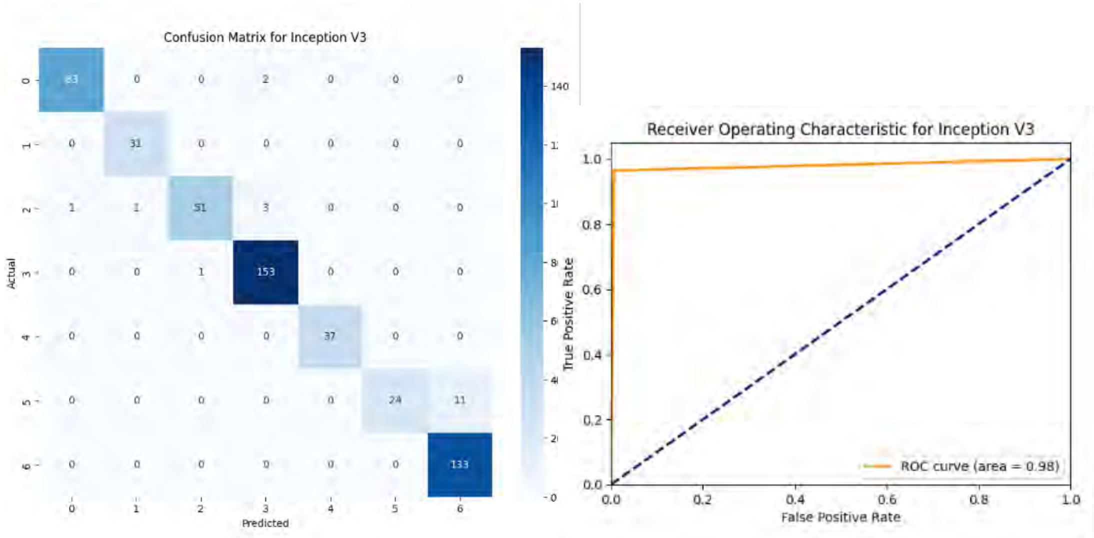

Figure 19: Confusion Matrix (Left) and ROC (Right)

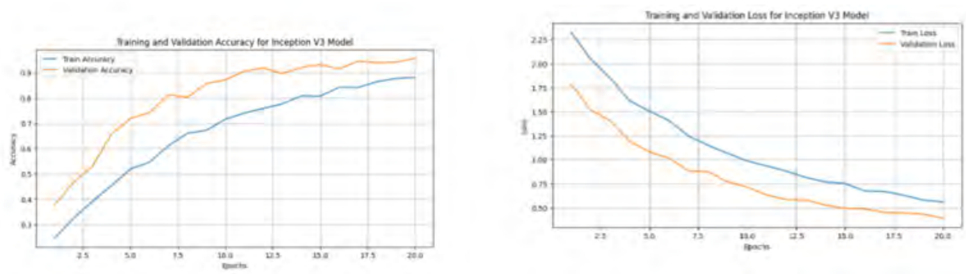

Figure 20: Inception V3 Model Training Accuracy (Left) and Loss (Right)

Custom CNN Model

Time taken for 20 epochs is around 229.12 seconds. The custom 14-layer CNN model was trained for 20 epochs, showing significant improvement in both training and validation accuracy and a corresponding decrease in loss. Training improves from 0.2119 in the first epoch to 0.8893 in the 20th epoch. Validation accuracy increases from 0.4855 to 0.9574. The ROC curve with an AUC of 0.99 and the high classification report metrics indicate that the model performs very well in classifying the images.

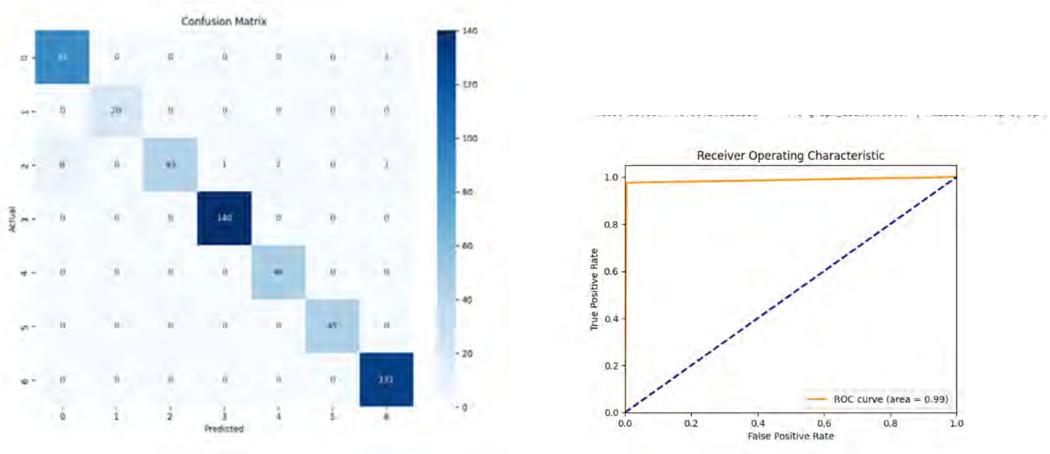

Figure 21: Confusion Matrix (Left) and ROC (Right)

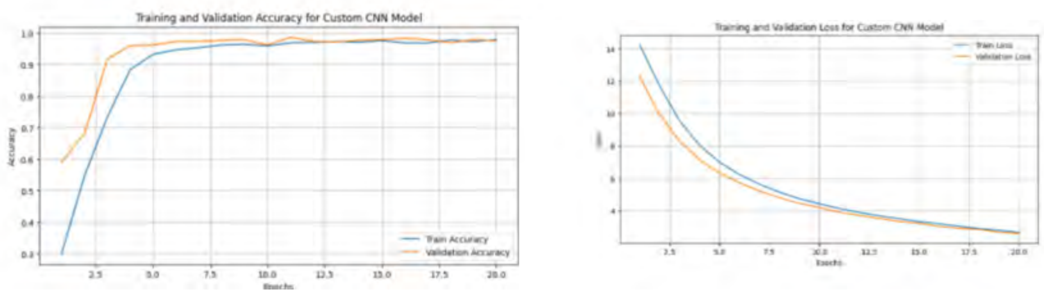

Figure 22: Custom Model Training Accuracy (Left) and Loss (Right)

The ROC curve for the given models shows the distinguishing ability of the models between healthy and diseased class.

Chapter 8

8.1 Result Analysis

The performance metrics for all the models are given in Table 1 for comparative analysis

Table 1: Result Analysis with Comparison Table

Model	F1-Score	Accuracy	Training Time (s)	Evaluation Time (s)	AUC
ResNet50	0.98	422.31	15.76	0.99	0.98
DenseNet121	0.75	415.23	32.34	0.85	0.69
VGG-16	0.64	547.17	10.45	0.79	0.63
Inception V3	0.96	503.45	15.93	0.98	0.95
Custom CNN	0.98	229.12	4.89	0.98	0.97
ViT	0.99	1952.57	7.18	0.99	0.99

Based on the existing paper review on different plant leaf disease detection using images the highest accuracy obtained for all the traditional CNN models and ViT models along with the obtained accuracy for these models used in our study has been shown here.
The test accuracy achieved for ResNet50 is 98% whereas the highest accuracy found for potato and corn leaf disease detection is around 87.51%-90.28% [6,20]. Again, for DenseNet121 achieved accuracy is 75% and the highest accuracy exists is around 97%[6,11,16]. For VGG-16 we obtained 64% accuracy but found the highest accuracy to be about 97% [13]. Simialrly, for Inception V3 model we got test accuracy 96% and existing accuracy is also around 96% [20]. Finally, for ViT obtained accuracy is 99% but highest accuracy found in paper review is around 88.86%[21].Thus, it is observed that the accuracy obtained for Resnet50 and ViT model in our study is highest whereas Inception V3 model has also shown a competitive accuracy. However, the accuracy obtained for DenseNet121 and VGG-16 is lower than the accuracy from the existing literature. This variation is found due to different approaches for model training, fine-tuning hyperparameter and different dataset sizes.

Based on the F1-classification report and training times provided, the ViT (Vision Transformer) and Custom CNN models demonstrate the best performance, each achieving a 0.99 and 0.98 accuracy. However, the ViT model has a significantly higher training time cost of 1952.57 seconds compared to the Custom CNN's 229.12 seconds, making the Custom CNN more efficient in terms of training time. Similarly, we find evaluation time of custom model faster than all other models.The ResNet50 model also performs well with a 0.98 accuracy and macro average, and a moderate training time of 422.31 seconds. DenseNet121 and VGG-16 show relatively lower performance with 0.75 and 0.64 accuracy, respectively, and have varying training times of 415.23 and 547.17 seconds. With 0.96 accuracy and 503.45 seconds of training time, Inception V3 strikes a good balance between performance and training duration. When taking into account training and evaluation efficiency as well as performance, the Custom CNN turns out to be the most efficient model.

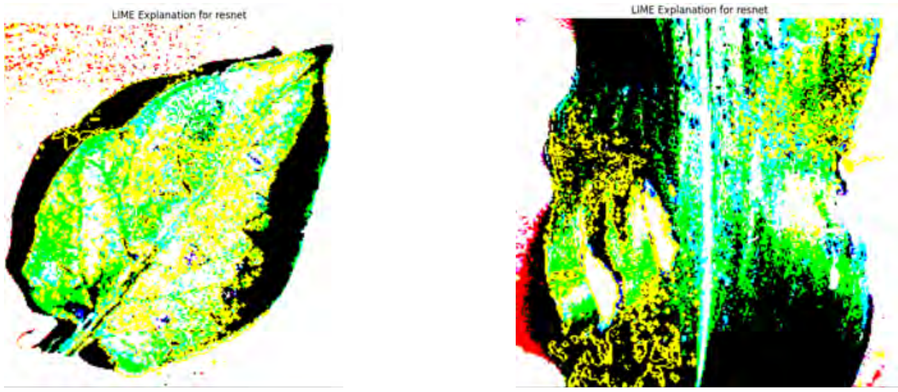

Figure 23: LIME for ResNet50

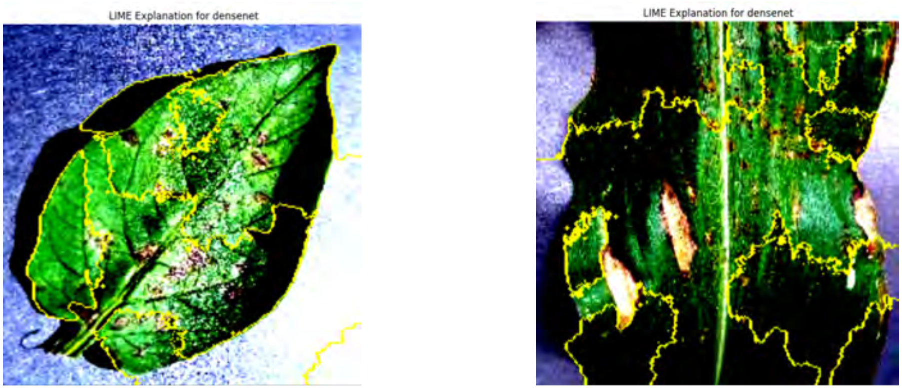

Figure 24: LIME for DenseNet121

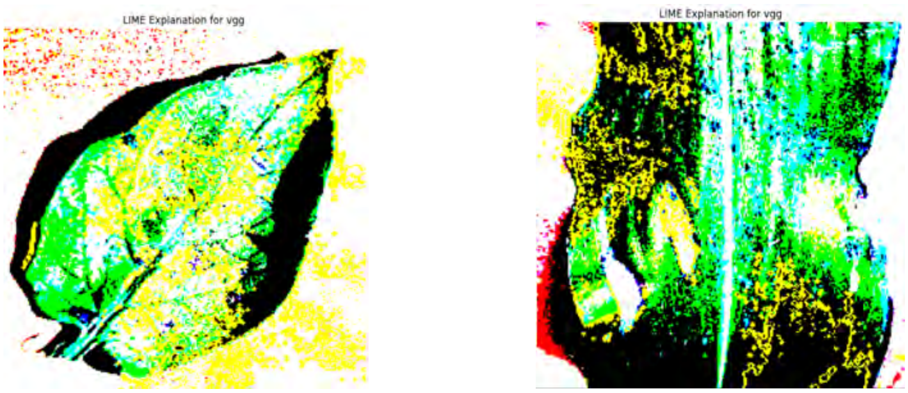

Figure 25: LIME for VGG-16

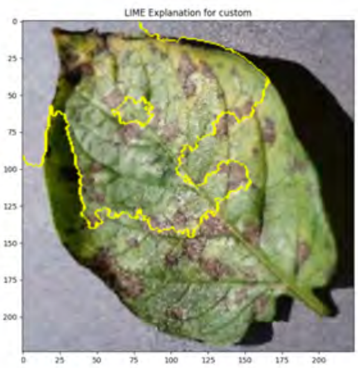

Figure 26: LIME for Custom Model

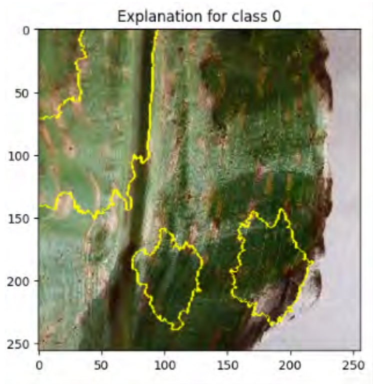

Figure 27: LIME for ViT

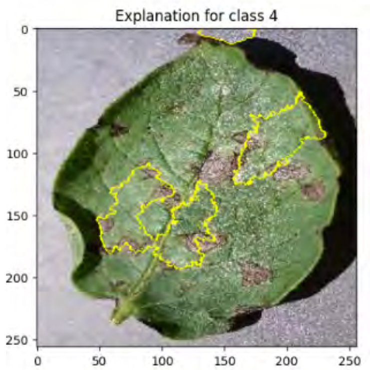

Figure 28: LIME for ViT Model

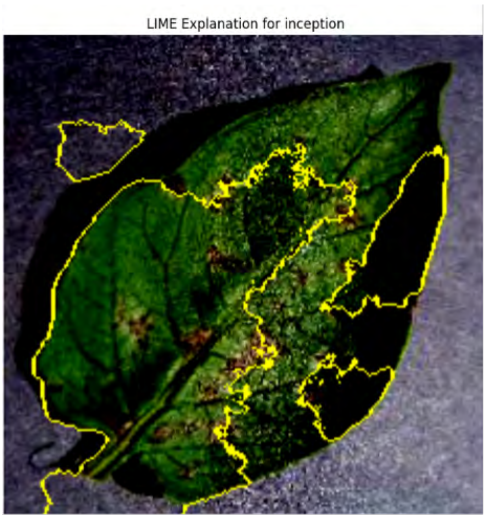

Figure 29: LIME for Inception V3 Model

Figure 30: LIME for Inception V3 Model

The highlighted areas of the LIME interpretation shows the region with significant features such as spots and discoloration, which the model used to classify the image for corn and potato diseases. It shows that the model focused on meaningful parts of the image to make its prediction.

Chapter 9

Conclusion

After comparing a number of models, it is shown that Vision Transformer (ViT) and Custom CNN perform the best. They both obtain a macro average of 0.99 and 0.98 and overall accuracy of 0.99 and 0.98, demonstrating their stability and dependability in all classes.

With a macro average of 0.98 and an accuracy of 0.98, Resnet50 likewise shows good performance. With an accuracy of 0.96 and a macro average of 0.94, Inception V3 is a dependable substitute that comes in close second. By comparison DenseNet 121 and VGG-16 perform far worse, with macro averages of 0.58 and 0.65 and accuracy of 0.75 and 0.64 respectively and very poor performance in several classes. Consequently, Resnet50 and Inception V3 are strong alternatives to ViT and Custom CNN for this classification job, while DenseNet 121 and VGG-16 are less appropriate because of their patchy performance.

Future Works

The plan for our future works in this study includes:

- **Data Collection and Annotation:** Collect a comprehensive dataset of potato and corn leaves affected by various diseases.Ensure accurate annotation of the dataset to include different disease types and stages.
- **Model Training and Fine-Tuning:** Train the high-performing models (ViT and Custom CNN) on the new agricultural dataset. Fine-tune the models to optimize their performance specifically for potato and corn disease detection.
- **Advanced Techniques:** Implement transfer learning to leverage pre-trained models for improved accuracy and reduced training time. Apply data augmentation techniques to enhance the dataset diversity and improve model robustness.
- **Performance Evaluation:**

  Evaluate the models' effectiveness in detecting potato and corn diseases through various metrics (accuracy, precision, recall, F1 score). Compare the performance against existing methods to validate improvements.
- **Real-Time Application Development:** Develop a mobile application integrating the trained models for real-time disease detection. Provide a user-friendly interface for farmers to upload leaf images and receive instant disease diagnosis.
- **Field Testing and Validation:** Conduct field tests to validate the application's performance in real-world conditions. Gather feedback from farmers and agricultural experts to refine and improve the application.
- **Extension to Other Crops:** Explore the possibility of extending the disease detection framework to other crops. Collect and annotate datasets for additional crops and diseases to generalize the application's utility.

References

[1] https://www.canr.msu.edu/biotechpp/about-late-blight-disease#:~:text=Late%20Blight%20Disease%20(Phytophthora%20infestans,range%20from%2015%2D30%25

[2] Hou, C., Zhuang, J., Tang, Y., He, Y., Miao, A., Huang, H., Luo, S. (2021). Recognition of early blight and late blight diseases on potato leaves based on graph cut segmentation. Journal of Agriculture and Food Research, 5, 100154. https://doi.org/10.1016/j.jafr.2021.100154

[3] Lakhani, J., Harwani, D. (2022). DEEP CONVOLUTION NEURAL NETWORK CLASSIFICATION OF BLIGHT DISEASE OF POTATO. Journal Name, 34(2), 177-185.

[4] Applalanaidu, M. V., Kumaravelan, G. (2021). A Review of Machine Learning Approaches in Plant Leaf Disease Detection and Classification. In 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) (pp. 716-724). Tirunelveli, India. https://doi.org/10.1109/ICICV50876.2021.9388488

[5] Khalifa, N. E., Taha, M., Abou El-Magd, L., Hassanien, A. E. (2021). Artificial Intelligence in Potato Leaf Disease Classification: A Deep Learning Approach. https://doi.org/10.1007/978-3-030-59338-4

[6] Saleem, M. H., Potgieter, J., Arif, K. M. (2019). Plant Disease Detection and Classification by Deep Learning. Plants, 8(11). https://doi.org/10.3390/plants8110468

[7] Arshaghi, A., Ashourian, M., Ghabeli, L. (2023). Potato diseases detection and classification using deep learning methods. Multimedia Tools and Applications, 82, 5725-5742. https://doi.org/10.1007/s11042-022-13390-1

[8] Afzaal, H., Farooque, A. A., Schumann, A. W., Hussain, N., Esau, T., Abbas, F., Acharya, B. (2021). Detection of a Potato Disease (Early Blight) Using Artificial Intelligence. Remote Sensing, 13(3), 411. https://doi.org/10.3390/rs13030411

[9] Mahum, R., Munir, H., Mughal, Z.-U.-N., Awais, M., Khan, F. S., Saqlain, M., Mahamad, S., Tlili

[10] Chakraborty, K. K., Mukherjee, R., Chakroborty, C., Bora, K. (2022). Automated recognition of optical image-based potato leaf blight diseases using deep learning. Physiological and Molecular Plant Pathology, 117, 101781. https://doi.org/10.1016/j.pmpp.2021.101781 38

[11] Fraiwan, M., Faouri, E., Khasawneh, N. (2022). Classification of Corn Diseases from Leaf Images Using Deep Transfer Learning. Plants (Basel), 11(20), 2668. https://doi.org/10.3390/plants11202668

[12] Setiawan, W., et al. (2022). Title of the Paper. Journal of Physics: Conference Series, 2406, 012019.

[13] Pan, S., Qiao, J., Wang, R., Yu, H., Wang, C., Taylor, K., Pan, H. (2022). Intelligent diagnosis of northern corn leaf blight with deep learning model. Journal of Integrative Agriculture, 21(4), 1094-1105.

[14] Rashid, J., Khan, I., Ali, G., Almotiri, S. H., Al Ghamdi, M. A., Masood, K. (2021). Multi-Level Deep Learning Model for Potato Leaf Disease Recognition. Electronics, 10(17), 2064. https://doi.org/10.3390/electronics10172064

[15] Khan, A., Rauf, Z., Sohail, A., Rehman, A., Asif, H., Asif, A., Farooq, U. (2023). A survey of the Vision Transformers and its CNN-Transformer based Variants. ArXiv. https://arxiv.org/abs/2305.09880

[16] Arnaud, S., Rehema, N., Aoki, S., Kananu, M. (2022). Comparison of Deep Learning Architectures for Late Blight and Early Blight Disease Detection on Potatoes. Open Journal of Applied Sciences, 12, 723-743. https://doi.org/10.4236/ojapps.2022.125049

[17] World Bank. (2016, May 17). Bangladesh's Agricul39 ture: A Poverty Reducer in Need of Modernization. Retrieved from https://www.worldbank.org/en/news/feature/2016/05/17/banglagri

[18] Agrico. (n.d.). Agrico in Bangladesh. Retrieved from https://www.agricopotatoes.com/agrico-in-bangladesh

[19] Xia, Y., Tang, M., Tang, W. (2023). Fine-grained Potato Disease Identification Based on Contrastive Convolutional Neural Networks. Applied Artificial Intelligence, 37(1). https://doi.org/10.1080/08839514.2023.2166233

[20] Brown, D. L., De Silva, M. (2023, September). Plant Disease Detection on Multispectral Images using Vision Transformers. Paper presented at the Irish Machine Vision and Image Processing Conference (IMVIP), Galway, Ireland. DOI: 10.5281/zenodo.8232619. https://www.researchgate.net/publication/377585574_Plant_Disease_Detection_on_Multispectral_Images_using_Vision_Transformers

[21] Alidev, A. (2020). PlantVillage Dataset. Kaggle. https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset