Plant Leaf Disease Classification using Deep Learning

A PROJECT REPORT

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
SOFTWARE ENGINEERING

Submitted by

AMIT RAMANI (2K22/SWE/14)

Under the supervision of

Mr. Sanjay Patidar

SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi 110042

MAY, 2024

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

We wish to express our sincerest gratitude to Mr. Sanjay Patidar for his continuous guidance and mentorship that he provided us during the project. He showed us the path to achieve our targets by explaining all the tasks to be done and explained to us the importance of this project as well as its industrial relevance. He was always ready to help us and clear our doubts regarding any hurdles in this project. Without his constant support and motivation, this project would not have been successful.

Place: Delhi | Amit Ramani

Date: | (2K22/SWE/14)

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Amit Ramani, Roll No. 2K22/SWE/14 student of M.Tech (Software Engineering),hereby declare that the project Dissertation titled "Plant Leaf Disease Classification using Deep Learning" in partial fulfilment of the requirement for the award of degree of Master of Technology, submitted in the Department of Software Engineering, Delhi Technological University, Delhi is an authentic record of my own work carried out during the period from 2022 to 2024 under the supervision of Mr. Sanjay Patidar.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Candidate's Signature

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled “Plant Leaf Disease Classification using Deep Learning” which is submitted by Amit Ramani, Roll No. 2K22/SWE/14, Department of (Software Engineering), Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of degree of Master of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Mr. Sanjay Patidar
Assistant Professor
Department of Software Engineering
DTU, Delhi

Date:

Abstract

The prevalence of plant diseases is in fact one of the main factors that lower the quality and quantity of agricultural products. The diseases keep emerging in the leaves of the plants with the development in plant structure and change in cultivation methods. Usually, the diseases first attack the leaves and then spread to the whole plant; hence the variety and yield of the crops that can be grown get highly influenced. Plant diseases are, in fact, one of the leading prevailing points of attacks on the global food supply and funds. This work has developed a system using EfficientNetV2 for plant leaf disease classification. The model has been trained on the PlantVillage dataset, which now contains 61,486 manually labeled images showing 14 different classes of healthy or unhealthy crop leaves and categorized over 39 distinct classes. Extensive testing and comparison showed that the model properly identifies plant leaf diseases. This all is going to draw the conclusions able to revolutionize the strategy for disease detection and control in plants. The experimental results revealed that the EfficientNetV2 model was able to give an accuracy of 99.40% in training and 99.24% in testing, suggesting its high effectiveness for early diagnosis of leaf diseases. In addition to that, with the implementation of deep learning and lately designed EfficientNetV2, it offers an effective way to timely disease detection for the improvement of agricultural practice, which aims at global food security.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

1.1 Overview

Plant diseases are one of the most threatening factors for agriculture all over the world, as they cause drastic distortion of the crop growth process with considerable yield losses. Most traditional management practices highly depend on human expertise; therefore, they are slow, laborious, and error-prone. As a consequence, control at its initial stage, when it is possible, might not be achieved by these methods. This inefficiency underlines the very necessity of next-generation automated solutions that could enhance plant disease diagnosis and management. Higher-level learning technologies, principally deep learning algorithms, have provided a promising opportunity for the solution of these tasks. In recent times, researchers have increasingly been exploring the potential for increasing the identification accuracies of various leaf diseases with these algorithms to be critical in the administration of proper treatment and preventive measures. Early and accurate identification of diseases is critical because any delay in the process itself alludes to heavy losses in terms of reductions in harvest, which affects food security as well as agricultural profitability.

Diseases attract attention in countries like India where agriculture is the key and a commanding field toward national development and food production. The agriculture sector forms a backbone to the means of livelihood for millions of people and significantly contributes to the national GDP. However, the sector also faces a myriad of threats and challenges, with plant diseases being the most conspicuous, as they could destroy crops, consequently hurting food security. As the world's population is projected to hit about ten billion by the year 2050 [1], coupled with the fact that environmental conditions are not stable, this may require new perspectives in how agriculture is conducted. In the achievement of this objective, deep learning models and associated technologies play an important role. The algorithms in question also performed well at photo classification jobs, which involved using images of leaves from plants to determine a pattern or characteristic that could be correlated with different diseases manifested on the crops. It is from these pictures that the algorithms can therefore deduce accurately, giving a farmer [2] an opportunity to, in detail, implement a targeted treatment strategy.

As a result, the application of precision agricultural technology in disease control is becoming increasingly important. It helps farmers make early decisions, identify outbreaks, and prevent further spread as a result of diseases. Such proactive measures shall be key in getting the crops taken care of and also useful in protecting the economic interests of the farmers. This is about monitoring crop health and environmental conditions using sensors, drones, and Internet of Things devices round the clock. This data, in turn, is fed to advanced data processing algorithms that convert it into real-time insights on crop conditions. This knowledge helps farmers apply units of water, fertilizers, and pesticides effectively to minimize resource waste while also minimizing effects on the environment [3]. Used in the context of plant disease management, it can identify the hot spots of a disease and predict an outbreak to take action that reduces impact.

The augmentation of agricultural technologies now becomes a must because of the increased global population and ecological pressures. Deep learning solutions, by their very nature, encourage interdisciplinary collaboration among scientists in agronomy, computer science, and environmental science. This will be a key realization for the implementation of more comprehensive and sustainable solutions to complex modern agriculture problems. Reduction of chemical inputs through these technologies boosts agricultural productivity with higher sustainability and lesser environmental pollution. This fact makes it possible to detect and control diseases at the very early stages, so deep learning technologies would be a great addition to any agricultural system that could operate under pressure.

It is not only the benefits that are for an individual farm but an influence beyond that. Consequently, there will be enhanced global food security through better detection and management of plant diseases ensured by the use of advanced technology. These technologies have the capacity to increase crop yield and reduce losses, hence contributing greatly to the reduction of poverty and increasing living conditions. It has been seen, too, as managing environmental sustainability by promoting better resource use efficiency in the reduction of an ecological foot mark from agriculture. Conclusions: Integrating deep learning models into the existing precision agricultural technologies would pave transformational ways forward in plant disease management. Technologies provide the powerful tools of detection and controlling plant diseases more effectively to the farmers to further the productivity of agriculture in a sustainable manner. There will have to be an important place for agriculture to adopt these advanced technologies with an increased population and environmental challenges across the globe to ensure food security under sustainable agriculture. Interdisciplinary collaboration and innovation could bring robust solutions for serving the demands of modern agriculture, ensuring a bright future for all.

1.2 Background

Detection of plant leaf diseases in agriculture has been very challenging, with potential to wreak havoc by causing crop failures over wide areas. Physical methods do not, in most cases, particularly at early stages, offer room for intervention. All these conventional techniques are quite labor-intensive and subject to human errors, with consequent delays in response and control measures that help diseases to spread. It is in this context that the introduction of advanced technologies has fundamentally changed the domain with the ability to quickly and accurately analyze substantial volumes of visual data in agriculture by using sophisticated algorithms. Automated disease recognition systems can now recognize the type of disease that is affecting the plant and predict potential future occurrences, hence enabling preemptive actions.

It has further enriched precision agriculture with the unique opportunities of real-time monitoring and intervention against plant diseases. As of today, the farming practices have since employed sensors, drones, and very many other IoT devices in their modern-day farming so as to get full-fledged information on crop diseases, environmental factors, and crop conditions. Additionally, the data can be analyzed at all times, allowing for a live view that enables real-time insights. For instance, such analysis can expose subtle changes in plant health which are not visually detectable or otherwise hence give insight into disease detection early enough. With these insights put in place, farmers can go much further by putting harmonized treatment on their crops, applying specific fungicides, or even changing irrigation patterns to lower the losses of the crops by reducing certain risks that are spread in a disease. It will not just gain in management of the disease through a professional way of data-driven approach but will dispel the gain by creating a sustainable practice in agriculture through the reduction of chemical abuse.

There are, however, still considerable challenges to be tackled despite this progress. The major challenge here is the unavailability of high-quality, annotated training data sets required during the development of robust machine learning models. Most existing data are either too small or do not manifest enough diversity for the data to be used in training across algorithms, which would generalize well for the many crops and conditions of the disease. In addition, practical and scalable solutions in plant disease detection remain underdeveloped. This is therefore a challenge in the quest to identify diseases in plants from images: lighting, occlusions, and multiple diseases, occurring on one leaf. Therefore, it is very promising for effective deployment in agriculture but poses a challenge in this domain of computer vision and machine learning technologies.

This is because in the successful detection and management of plant diseases in the future, transdisciplinary approaches will need to be in place that cannot occur without involvement by all actors, including researchers, agronomists, technologists, and policymakers. This will require a global approach by developing comprehensive datasets and improving the algorithms to make this tool implementable by farmers around the world. Innovations through collaboration can be put forth in this respect in order to develop an improved practice for the use of a plant disease detection technique. The utilization of state-of-the-art technology ensures the increase in productivity to be sustainable, hence guaranteed food security in the future. It is expected that further research and development in the area will result in a resilient agricultural system able to meet challenges that lie ahead for the growing world population under unpredictable and rapidly changing climatic conditions.

More, it is in this field that advances can be significantly more than the summed crops harvested within a perimeter of one farm. Global food security will benefit from better detection and management of plant disease to enhance stable crop yields around the world. It can also reduce the poor economic status of crop failure, mainly affecting smallholder farmers in developing regions. These technologies may discourage environmental health and biodiversity by reducing dependency on chemical inputs. Clearly, the potential benefits revealed only underline the importance of sustained investment and innovation in agricultural technology. By looking forward, the massive integration of immense technologies in plant disease detection is going to provide a sustainable and resilient agricultural system that can feed the increasing world population and change in environmental conditions.

Deep learning has permeated application domains in many other fields or sectors related to pattern recognition, computer vision, among others. Convolutional neural networks (CNN) have shown great results in the task of classifying images, which is the reason why they are part of applications for classifying plant diseases. A CNN consists of many layers of artificial neurons that will learn diverse representations of the input image at different levels of the process. A bestiary of local patterns and features (such as edges, textures, shapes) is slowly built one layer at a time until they jointly become part of a global classification at a higher level. By the nature of incorporating ReLU (rectified linear units) , one type of activation function allowing non-linearity, it is possible that a neural network can learn complex interactions in the data [4].

The convolutional neural network architectures, taking all possible stages in the search for an efficient neural network that would be possible, crumbling the least in computational resources to bring about maximal accuracy within a certain scale of variance. For instance, while the continuations through remote coupling escalation parameterizations serve to enhance model accuracy, they have wisely chosen to scale up network depth, width, and resolution adequately with respect to one another. Therefore, the accuracy optimization maintains a strategy of convergence toward the limit that fits some hardware platform specifically. Furthermore, the deep learning design has also introduced new technology to include traditional methods of embedding with attention mechanisms and self-attention mechanisms to represent features and data long-range dependencies better.

Hyperparameters form a major part of deep learning models, especially during the training phase when researchers "search" the hyperparameter space of their objective function in an effort to come up with the correct configuration that will maximize their model's performance. Oftentimes, users resort to techniques like grid searches, random searches, and Bayesian optimization techniques for the best setup when modeling any new problem. More advanced are the designs used for NAS and AutoML, automating the searching process for optimal sets of model hyperparameters, ensuring that the model set up is better and more

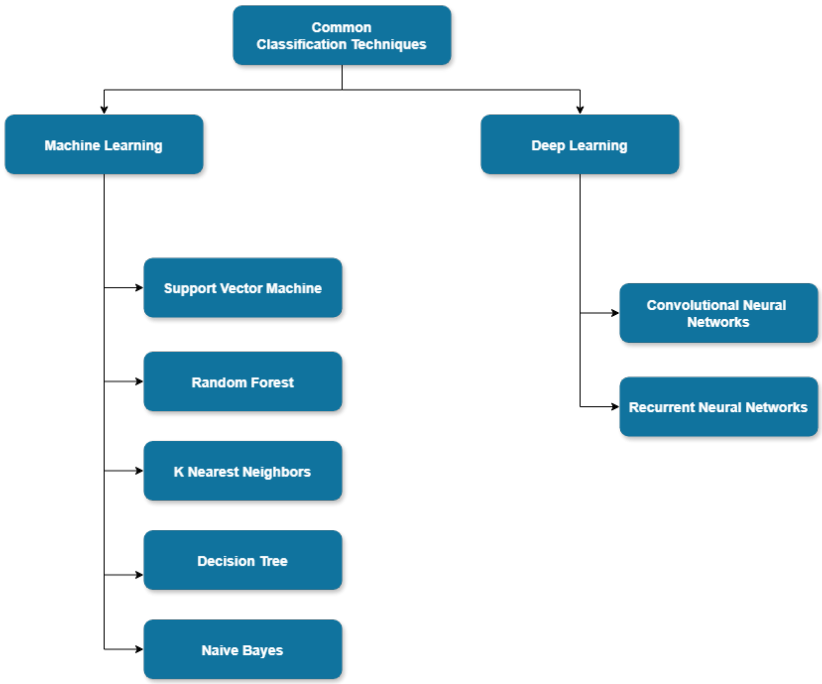

Figure 1.1: Types of common classification techniques

effective than the researcher's model.

Data preprocessing and normalization must be carried out in order for the input data to be considered ready for use by deep learning models. For example, classification of an image necessitates data augmentation and normalization to feed the data set with both enriching and scaling of the data set in such a way that trainability and generalization potentials would be high. Some common operations during this process are resizing images to fixed size, pixel value normalization to fall within a fixed range, and data augmentation in order to introduce randomness and reduce overfitting. Normalization makes the inputs to the model consistent, which in turn keeps the variation between the inputs low under different lighting conditions and color distributions. Key data augmentation techniques are random rotation, flipping, cropping, and zooming, which in turn learn robust features from their variations and hence generalize well on new samples.

Deep learning models cannot be evaluated without using performance metrics such as F1-score, accuracy, precision and recall. With this knowledge of technicalities, researchers can solve problems related to disease identification in agriculture using modern methodologies that will improve crop health and yield.

Plant leaf disease classification involves a plethora of hundreds of machine learning and deep learning technologies as shown in shown in 1.1. Basically, the machine learning techniques include Support Vector Machines, Random Forests, K-Nearest Neighbors, Decision Trees, and Naive Bayes because of their different capabilities for handling complex datasets and giving strong classifications. The deep learning models proposed herein, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), link into obtaining superior performance in feature extraction and image analysis compared to the other conventional models. Therefore, this kind of holistic approach may eventually lead the way to high accuracy and efficiency in disease detection for agricultural applications.

These techniques are discussed below:

1. **Support Vector Machines (SVM):** Support Vector Machines (SVM) is an effective algorithm for supervised learning and used mainly in classification tasks. It may locate the best hyperplane in feature space such that it cleanly separates data points into different classes, wherein the best hyperplane bounds the margin of closest data points from each class, which are denoted as support vectors. The SVM fits well to very high-dimensional feature spaces, since the amount of dimensions is allowed to go up much more than one would have a chance to handle with such simple methods as simple decision boundaries. This is just the kind of situation that comes about in complex and intricate feature spaces, such as the space used in plant leaf disease classification. Thus, the kernel functions such as the radial-based function (RBF), polynomial, and sigmoid kernels increase the feature dimension, transforming the non-linear features in input space into linear separations, thereby making them classify well [5]. It is very important for proper classification of healthy and diseased plant leaves as SVM has strong capabilities to handle the complex nature of patterns with subtle variations in the leaf image. One of the main advantages is robustness when a good separation margin exists between classes, even with data samples that are relatively small. This kind of robust performance stems from the SVM's ability to zero in on the most important information only and disregard that which brings about a minimum amount of information, making it less prone to overfitting in practice and better at generalization over unseen data. A kernel together with its appropriate parameters is in practice the most important factor affecting the performance of many SVMs, for example, the regularization parameter (C) and the kernel-specific—such as the gamma parameter of RBF. The choice of appropriate kernel along with these parameters often involves very close experimentation with the models evaluated through cross-validation, which in turn can be quite time-consuming and computationally costly. Furthermore, SVMs are intensive in computation during the training stage, since they address complex quadratic programming problems. Resources and training times might largely increase when considering huge datasets. Despite these drawbacks, SVMs are still popular and effective for plant leaf disease classification, forming the basis for most current applications not only in agriculture but also in other fields.

2. **Random Forests:** Random Forests are an important and versatile ensemble learning technique: the technique creates many decision trees and outputs the mode of the class (classification) of the individual trees. This reduces overfitting, the most popular problem of a single decision tree, hence boosting accuracy and model stability. The algorithm works by creating

many uncorrelated decision trees trained on different subsets of the data and also on a random subset of features. Aggregation methods usually provide more accurate and general models, and the classification decision for a vector of features will be realized as the average vote of the trees in the forest. One more extra advantage of Random Forests over the other methods would be that it copes well with the huge number of input features required for classification of plant leaf diseases. On essential occasions in dealing with high-dimensional data—i.e., images—the selection feature can be done automatically to identify relevant features for making predictions. This feature selection adds up the robustness properties of the models against noisy and irrelevant data, normally existing in agricultural data sets. Thus, another important advantage of Random Forests is their ability to capture complex feature interactions. Elaborate patterns and combinations in the case of plant leaf diseases can be missed by a single decision tree. The present model can averagely better model these relationships through all possible multiple fits of decision trees with different parts of data, thereby leading to better classification accuracy. Random Forests do have some disadvantages, however. The more trees are added to the forest, the more complex the model will become overall, meaning less interpretability could be given to the model. In an agricultural context, understanding the process involved for decision-making could be important for diagnosing and managing diseases [6]. Even though performance is strong in Random Forests, with inherent characteristics for the ensemble, little parameter tuning is required. Nevertheless, the production problems go beyond the size and complexity of the model in the context of the deployment required by resource-constrained environments.

3. **K-Nearest Neighbors (KNN):** The K-Nearest Neighbors, or KNN, is a very simple and intuitive algorithm. It is non-parametric and used for classification and regression tasks. Majority voting of the chosen k-nearest neighbors in the feature space assigns the class. For such simplicity, KNN becomes the favorite in most applications, one of which is the classification of plant leaf diseases. K Nearest Neighbours is capable of making direct classification of new samples similar to the other known ones in the dataset based on the similarity between feature vectors. That is, for a plant leaf disease classification scenario, K-NN shines in handling the variations and complex patterns on display over the feature space. These feature vectors, based on leaf images, can capture slight differences in color, texture, and outlines manifested by different diseases in plants. Because KNN is based on the training of a local neighborhood for classification, it will adapt to intricate patterns found in data without an explicit training phase. In this respect, the algorithm becomes very useful in those cases when functional relationships between features are either nonlinear or not easily modelled by parametric approaches. The main advantage of KNN is in its simplicity and ease of implementation. In other words, unlike other machine learning algorithms, like support vector machine or random forests, KNN does not

introduce a separate training process. This algorithm just stores the whole dataset and uses it for the time of prediction to find closest neighbors of some sample. This very characteristic eliminates parameter estimation and model fitting that are quite computation-intensive and hard in some other models. Moreover, KNN is quite flexible in that it can be easily adapted to different types of data or distance metrics. However, KNN also contains a few major drawbacks that can greatly influence its performance and scalability. The choice of k, meaning the number of neighbors to be considered, is really a major decision in that respect, one which has a great effect on the accuracy of the model. Moreover, for very small values of k, the algorithm is highly sensitive to noise and outliers, while for large values of k, the dilution of local structure may misclassify a point. Usually, deeper experimentation and validation are required for the choice of k that might be considered highly appropriate. Moreover, KNN is quite sensitive to the choice of a distance metric. Euclidean, Manhattan, and Minkowski distances are some common metrics applied. Depending on the nature of the data, each can be strong in some applications and weak in others. Another major limitation of KNN is the computational cost and inefficiency during the prediction phase, especially for large datasets. As a con, KNN requires computing the distance between the testing sample and all of the training samples; this results in an increase of computational cost fully proportional to the dataset itself. This can result in slow prediction times and high memory usage; therefore, KNN is not the best approach for real-time applications or for big data scenarios [7]. Efficient implementations and optimizations, such as KD-trees or ball trees, to some extent mitigate these issues, but in turn tend to have a lot of algorithmic complexity. Besides, KNN is very sensitive to irrelevant or redundant features that may degrade its performance. The curse of dimensionality in high-dimensional spaces may wear down the distance metric and, as a result, classify poorly. This means that effective feature scaling and selection techniques are very important with KNN. Normalizing feature vectors will make all features share equally in contributing toward the distance calculation, while feature selection helps to reduce data dimensionality and eliminate noise.

4. **Decision Trees:** Decision Trees are one of the oldest classification algorithms, the first to come out and still widely used. They classify by recursively partitioning a dataset into partitions that depend on input features. This process develops a tree-like structure, where each internal node represents a decision about a feature, each branch indicates the result of the decision, and each leaf node means a class label. Decision trees are best suited for plant leaf disease classification because of its simplicity and interpretability, with added major benefits. These models are easily visualized and can be easily understood and interpreted; thus, leading to the clear observation of how rules are taken at each step. This is especially desirable in understanding the underlying patterns of plant disease data. In view of classification of diseases in a plant leaf, the decision tree is flexible

in representing features because it could handle both numerical and categorical data. They have little problems handling large datasets, and the process during the training phase is quite efficient—done with a sort in data space—toward finding the best split using criteria like Gini value or information gain. This makes decision trees very appropriate for quick prototyping and exploratory data analysis. In addition, they are non-parametric in the sense that they do not specifically assume any given form for input data, so they may adapt to different types of distributions and relationships among features [8]. On the other hand, though, the overfitting problem, especially when decision trees get deeper and more convoluted, is what plagues them. Overfitting is a situation in which one captures noise and fluctuations from the training set rather than underlying patterns, so it generalizes poorly to held-out data. However, it can be adjusted with several mitigation techniques. Pruning is the way through which a tree's complexity is reduced by sections of the tree that have little power in predicting target variables. Depth of tree, number of samples to split a node, or the number of samples in a leaf node can also be generalized by limiting them.

5. **Naive Bayes:** Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying the Bayes theorem with strong independence assumptions between the features. This is highly unrealistic for practical scenarios, but it hugely helps to simplify computations, making Naive Bayes one of the most efficient and scalable algorithms. Naive Bayes can be very useful in the classification of plant leaf disease because it is robust in handling large datasets and very fast for computations. Therefore, it looks like a quite attractive solution to real-time determination problems where the processes of quick decision-making should be taken, for instance, to detect diseases at the earliest possible stage. Bayes' theorem is all about updating some estimate of the probability of a hypothesis. Naive Bayes applies this theorem to be able to compute the posterior probability of a class given a set of features. Again, in practice, the Naive Bayes classifiers do surprisingly well, often matching or exceeding the performance of more complex algorithms despite the independence assumption. This is partly due to their robustness with regard to useless features, which do not degrade performance as that of other algorithms. Able to handle both continuous and discrete data, the Naïve Bayesian classifier is thus versatile in nature with respect to the input feature types. One of the major strengths of Naive Bayes is its simplicity. It is easy to implement the algorithm, which generally requires very little size of training data to estimate the necessary parameters. For Gaussian naive Bayes, these are only the means and variances of the variables. This simplicity translates into computational efficiency, making the models of Naive Bayes scalable with large numbers of datasets. In particular, the classification of diseases in plant leaves is done using a large amount of data and high-dimensional feature space. Both the training and prediction processes with the algorithm are fast; hence, it would be perfect for real-time application and scenarios with

limited computational resources. The main strength of Naive Bayes is thus the assumption of independence. That might not be so realistic for some practical problems, especially plant leaf disease classification where features will have high correlation, for example, different symptoms of a disease on a leaf. The assumption of independence between the features can—and in most cases does—result in less-than-optimal performance because the model will not be able to capture properly the true underlying dependencies in the data. In most cases, however, Naive Bayes is effective in cases where some approximations to independence are correct or when the inter-dependencies only slightly affect the classification task.

6. **Convolutional Neural Networks (CNN):** Convolutional Neural Networks are a technology breakthrough in the machine learning field and are best suited for tasks related to classifying images or visual-based data. By definition, CNNs are part of the deep neural networks class, which marked a revolution with respect to the way features of images would be done, becoming very favored also in plant leaf disease classification. The advancement of CNN used convolutional layers that, sacrosanct to the use of different filters to input images, automatically learn and extract spatial hierarchies of features. These features go from low-level details, such as edges and textures, to high-level representations, like shapes and very particular patterns that are important for plant diseases. CNNs work very well at the classification of different plant leaf diseases since they are capable of grasping and interpreting the very detailed features of the images. Other specific symptoms in recorded incidences of diseases on leaves include spots, discoloration, and changes in texture. These patterns are learned directly from the training data. This capability allows CNNs to easily handle very large-scale datasets of images, making them very convenient for use in agriculture, where very large image datasets can easily be collected using modern technologies like drones and sensors. One of the great advantages of CNNs is their automatic feature extraction. Classical machine learning will require one to have prior knowledge of the domain so as to manually derive features, while in CNN, the most relevant features are derived directly from the data. This makes modeling quite easier because generalization is good enough to make a hypothesis that generalizes about novel unseen data. In other words, CNNs are designed to suss out spatial and temporal dependencies in images, which makes them well-fitted to capturing the complex visual patterns associated with plant disease. In fact, despite its belonging to plant pathology, CNNs outperform other machine learning and deep learning models in the image classification task. Their deep architecture, characterized by many layers of convolutional and pooling operations, allows CNNs to learn appropriate hierarchical feature representations for accurate and robust classification. The stacking ability of multiple layers enables the CNN to build more and more complex representations of input data. This is quite important for distinguishing, just as an example, between a healthy or infected plant [9]. But there are also a few challenges to the application of CNNs. First is

the dependence on massive amounts of labeled data. The training of CNN for high accuracy purposes would necessarily involve a considerably large dataset of annotated images, a process by itself that is quite laborious and time-consuming to collect. This requires extensive and wide-ranging dataset collection and labeling for success, in which this can be a logistical challenge itself for identification of different plant leaf diseases.

7. **Recurrent Neural Networks (RNN):** The basic RNN and their advanced variations like LSTM networks and GRUs are designed for the processing of sequential and temporal data, holding good in itself for many patterns-over-time tasks. Traditionally, RNNs have been found useful in the domain of natural language processing, speech recognition, and forecasting time series data. When utilized in the detection of disease in plants, they demonstrate enormous potentials, especially when integrated with Convolutional Neural Networks (CNNs) into a hybrid model for improving the diagnosis and monitoring of diseases. Due to their recurrent connections, an RNN, as a consequence, may store a memory of previous inputs in the sequence; hence, it can capture dependencies across time. This capability would be invaluable when the information in the dynamics of symptom development associated with disease progression are critical. An application would be a classification of diseases in plants leaves for an early detection process in order to take effective action. Therefore, an RNN applied into processed sequential patches from an image passed through a CNN will monitor the development and spread of disease symptoms, which is more indicative of comprehensive analysis than static image classification. Here we combine CNNs and RNNs to use the strengths of both architectures. CNNs work great with spatial-feature extraction; they really help capture local patterns such as spots, lesions, and discoloration on plant leaves. These features can then be put into an RNN in order to capture the progression of disease over time. This hybrid approach is particularly useful in agricultural scenarios where monitoring is unceasing, and the early detection of plant diseases leads to timely and effective treatment. By its nature, one of the interesting advantages of using RNNs for a sequence-generative model is the modeling of dependencies in a way that is potentially unlimited. A member of the RNN family in general, LSTM networks are designed to counteract the vanishing gradient problem and, therefore, retain information down long sequences. This will be quite helpful in watching out for diseases and pathogens since the symptoms of these diseases show up after some time. The hybrid CNN-RNN model is able to prescribe subtle changes by the use of LSTMs and predict when it is likely that a disease will spread, thereby putting in place timely interventions for better crop health management. It should be noted that the blending of RNNs with plant disease detection allows the system to have more interpretability. It helps a researcher in understanding the different temporal patterns associated with some diseases by checking the sequence of feature activations over time. To a great extent, this not only fast-tracks activity in the determi-

nation of diseases but also unravels the mechanism of disease development and increases its determination precision [10]. This sequential capability can also be considered, for example, in real-time crop monitoring, where it gives a continuous update on the status of a plant and forewarns on proactive measures to be taken to alleviate the impact of disease. There is one main challenge of using RNNs in plant disease detection: the models are computationally intensive and demand substantial resources mostly during training, for long sequences and large datasets. The training process is rather time-consuming, and it is easy for the models to overfit, especially with high-capacity models not well regularized. Further, hyperparameters like the number of layers, number of hidden units in each layer, and learning rates require tuning for setup; hence, the setup with RNN architectures is complex.

1.3 Problem Statement

Plant diseases do pose a critical threat to crop health and overall productivity around the world. Traditional approaches as far as plant diagnosis and control is concerned often rely on physical appearance and human interaction. This always leads to errors, inefficiencies, as well as delayed responses towards the detection of diseases and contaminations. The reliance on human familiarity increases the prospects of encountering variations in the process of diagnosing the diseases and actual remedies. The impacts of plant diseases on the agricultural systems have intensified with population growth and climate variability on a global scale.

Traditional approaches are inadequate for scaling up and do not present significant possibilities in combating unforeseen risks of the diseases or monitoring wide farm fields. At the same time, casual methodologies of symptoms' visual interpretation by humans lead to inconsistency in the diseases' typology. It makes any valuable control actions impossible . Hence, the current situation needs state-of-art technologies and procedures to promote automation and refinement in the plant disease control process.

Computer vision, machine learning, and deep learning can be utilized in the development of smart systems that could accurately recognize, classify and forecast real time plant diseases. These technologies have the ability to revolutionize agriculture by enabling prompt detection and proactive treatment of disease outbreaks as well as safeguarding agricultural yields through minimizing any financial loss while ensuring food security for a rapidly growing population.

However, having marked datasets for machine learning model training is important in order for such technologies to become useful. Elaborating advanced diagnostic tools usable in agriculture necessitates collaboration among scientists, farmers as well as stakeholders.

1.4 Contributions

The contributions made by this research are substantial in terms of agricultural technology and plant disease classification which is a great advancement.

Use of deep learning based classification model: The approach toward deep learning thus arrives at a novel classification model of research focus. In this case, it has proven more effective and accurate enough to identify some kind of advancement in the field of plant disease identification through image classifications. As a result, the technique was proposed in precisely differentiating between healthy and infected leaves. More importantly, it applies the above technique to plant disease recognition based on deep learning approaches, which advances the application of automation in disease diagnosis in agriculture and increases the level of accuracy while decreasing time complexity in the specified area.

Hyperparameter optimization: This research goes a step further than just the model building and builds on the model to optimize the architecture for better performance. The model is considered at its peak performance when fine-tuned for some of the parameters, like the learning rate, choice of optimizer, batch size, and many more, using the hyperparameter optimization approach. The paramount thing for further enrichment of performance and applicability of the developed model in the agricultural setup is the optimization of parameters at the best possible level of precision. It generalizes really well, and with reasonable accuracy of classification, due to the optimal way in which the hyperparameters affect the architecture of the model in this research project.

These steps of deep learning and agriculture directed toward a better understanding someday will help in laying the building blocks of precision farming, sustainable food production, and global food security.

Chapter 2

LITERATURE REVIEW

The recent past has seen much concern in the identification and control of plant diseases through integration of learning techniques with practices. This is a comprehensive account of research review that seeks to put together studies that give insights on the ways to classify plant diseases and on the development of radical educational structures.

In their recent paper, Hassan et al. [11] point out the current drawbacks of deep learning models in terms of their large parameters, which need considerable processing power to be effective and may not be so effective in practical usage. To fill up this gap, the authors have used a novel model that actually trades parameter size for performance. The deep features in the images of corn plant leaves are extracted using two pre-trained models: EfficientNetB0 and DenseNet121. The selection of these CNNs lies within such criteria as high accuracy and efficiency in tasks of image classifications. This makes the features from those two networks concatenate into one, thereby making this a very complex single feature set with the technique of fusing both to improve learning capability and performance. They have also performed some data augmentation techniques to bring variations in the training images. The approach is to have a very high number of varying images that should be available for training. The model also gets adapted to coping with even complex cases and this strategy demonstrated its effectiveness since the model equally distinguished between healthy and unhealthy corn plant leaves. To begin with, Hassan et al. tested their model against some of the best pre-trained CNN models, ResNet152 and InceptionV3, for the purpose of model validation. The present model attains a classification accuracy of 98%.

The real importance of Sunil et al. [12] lies in the importance of worrying about diseases in plants like Colletotrichum Blight and Phyllosticta Leaf Spot, which are major threats to cardamom cultivation, and thereby agricultural productivity in this country. This study demonstrates that deep-learning techniques through a combination of the EfficientNetV2 model with the U2-Net model can effectively help detect and classify plant diseases for timely interventions to prevent them. The high detection accuracy of 98.26% clearly depicts the effectiveness of the approach in diagnosing diseases reliably and efficiently; it attests to a promising solution to tackle the challenges that farmers face. The findings from this study underline the radical impact of deep learning towards the revolution of plant disease diagnosis practices. Artificial intelligence and machine learning technologies will bring more accuracy, efficiency, and scalability to disease management solutions for plant health.

Afterward, Yang Zhang et al. [13] implemented a modified model of Faster RCNN for further improvement in tomato disease leaf recognition and localization accuracy. The major modifications of the approach include the extraction of image features using a depth residual network instead of VGG16 so that deeper features of the disease can be extracted. In more detail, the authors applied the k-means clustering algorithm to bounding-box clusters to make the anchoring procedure, based on the clustering results, much closer to real bounding boxes. Finally, experiments using the different feature extraction networks indicated that when receiving enhancements from the latter method, the recognition accuracy of the original Faster RCNN model was raised by more than 2.71%. Besides this, the proposed improved approach also conferred faster speed of detection, further validating it for effectiveness in faster detection not only of healthy tomato leaves but also of other diseases like powdery mildew, blight, leaf mold fungus, and ToMV. This work therefore extends works related to deep learning object detection in the field of agricultural disease management and offers some optimistic prospects of enhancements of those practices monitoring crop health and its management.

In their work, Nazki et al. [14] presented an unsupervised novel approach in image translation to improve plant disease recognition. This research is aimed to improve the accuracy of crop disease leaf recognition and localization of the diseased leaves. It builds over enhanced Faster RCNN intended for the identification of healthy tomato leaves and four most prevalent diseases: powdery mildew, blight, leaf mold fungus, and Tomato Mosaic Virus. Key Enablers: Replace VGG16 with a depth residual network to extract features from images to enhance the extraction of deeper features of diseases by the network. The authors also adapted the k-means clustering algorithm for the clustering of bounding boxes, and further fine-tuned the process of anchoring with the result of the methods for clustering to make anchor frames fit better with real bounding boxes. The experimental results show that the recognition accuracy of the improved method can increase by 2.71%, compared to the original model, Faster RCNN, with a faster detection speed. This work demonstrates how unsupervised image translation with adversarial networks opens the new way toward plant disease recognition evolution and presents promising ways for intensification in the practice of agricultural disease management.

Zhang et al. [15] addressed the difficult problem of cucumber leaf disease classification using high-level image processing methods. They acknowledged the fact that a diseased leaf is extremely complex to analyze and tried to overcome these difficulties in a large way by using deep learning models coupled with AlexNet. In the model they found some drawbacks of having excessive parameterization and limited feature scales. In response to above shortcomings, they offered a new model known as Global Pooling Dilated Convolutional Neural Network (GPD-

CNN). The proposed model unifies dilated convolution with global pooling, making it possesses some salient features in comparison with classical state-of-the-art convolutional neural network (CNN) and AlexNet. Firstly, the fully connected layers of GPDCNN are replaced by a global pooling layer, therefore increasing the convolutional receptive field without increasing computational complexity. The model restores spatial resolution by the dilated convolutional layers without adding the training parameters. In the last part, GPDCNN integrates the merits of dilated convolution and global pooling. Experiments conducted on datasets of six common cucumber leaf diseases show the effectiveness of the proposed model in the accurate recognition of cucumber diseases. This work reiterates the potential of GPDCNN in becoming a powerful tool for disease identification in plants, encompassing great strides in leveraging deep learning techniques for agricultural applications.

Uday et al. [16] continued their research into the dire consequences of the development of fungal diseases on mango trees, paying very close attention to Anthracnose disease, which affects both the fruits and leaves of mangoes. Theirs was a quest to find an effective way of diagnosing Anthracnose disease and cost-effective, early remedy to the challenge in agriculture. The authors proposed a Multilayer Convolutional Neural Network for classification of Mango leaves infected by Anthracnose fungal disease. The validation was done on a real-time dataset consisting of 1070 numbers of images of leaves of mango trees taken at Shri Mata Vaishno Devi University, Katra, J&K, India, containing both healthy and infected images of tree leaves. These results demonstrated that the MCNN model had higher classification accuracy than the others based on this state-of-the-art approach, which justifies MCNN effectiveness in accurate Anthracnose-infected mango leaf identification. The study further revealed that it is possible to effectively use the MCNN as a robust tool for early detection of the fungal diseases of mango trees to improve the agricultural practice for increased crop yield.

[17]'s research relates to the deadly effect of some diseases, insects and nematodes, and other pests attacking the sunflower crop with a high degree of loss in production. Some of the infections and infestations may be detectable due to the symptoms using the naked eye, but such methods are not practical for an extensive monitoring system of large farms. Segmentation and classification system for images of the sunflower leaf were proposed by the authors. The paper provides a rough survey of various disease classification techniques adopted in detecting sunflower leaf diseases. It was observed, particularly in the disease classification process, that segmentation in images of sunflower leaves—indeed, a crucial step—can be effected using the Particle Swarm Optimization algorithm. Performed experiments have shown promising results: the developed algorithm gives an average classification accuracy of 98.0% on leaf images, which is better than the accuracies reported for state-of-the-art methods, at 97.6% and 92.7%. The latter is a research study on the effectiveness of Image Segmentation based on Particle Swarm Optimization in proper identification and categorization of diseases that appear on the sunflower leaf, hence development of tools for effective

disease management in the crop.

Mishra et al. [18] studied important effects on the Indian economy and food availability caused by corn disease, setting up a context in which an automated diagnosis is in severe need to avoid severe crop loss. In this paper, we present a real-time automatic corn leaf disease recognition technique using deep convolutional neural networks. Fine-tuned hyperparameters, pooling combinations, and optimized numbers of parameters enhanced the performance of the deep neural network toward real-time inference. The pre-trained deep CNN model was successfully deployed on a Raspberry Pi 3 using the Intel Movidius Neural Compute Stick added with dedicated CNN hardware blocks. The overall average accuracy for the recognition of corn leaf disease with the deep-learning model is at 88.46%, which reconfirms its feasibility in all the previous works for the recognition and detection of diseases in corn plants. However, this type of model may be further implemented on standalone smart devices with additional use in farming environments such as a Raspberry Pi, smartphones, or drones.

According to Sharma et al. [19], their diagnosis is in time, and therefore crop damage is minimal on case when the diseased leafs are detected. Most importantly, according to them, a general weak point of most automated deep learning models is that performance generally drops drastically after having been tested on independent data. The CNN models here can be trained using segmented image data to fill this gap. Results indicate that the S-CNN model, trained by segmented images, performs much better than the F-CNN model trained by full images. This is done by the far superior generalization on independent data never seen before by the models at up to 98.6% accuracy for 10 disease classes between them. It also shows that the S-CNN model has superior self-classification confidence over the F-CNN model, for example, on tomato plant and target spot disease type. On the whole, this research has advanced ways of application, such as automation, that bring them closer to non-experts and are underway for timely disease identification in plants.

In Agarwal et al. [20], it was noted the way diseases have a massive effect on the tomato, allowing for less quality and quantity of yield. They proposed to tackle this issue using a deep learning approach for disease detection and classification, mainly using Convolutional Neural Networks. The proposed model consists of three convolutional layers followed by three max-pooling layers and. Their proposed model experimentally proved good classification accuracy in comparison to the pre-trained models like VGG16, InceptionV3, and MobileNet. The classification accuracy of their model during classification was between 76% to 100% across different classes, with an average accuracy of 91.2% for the nine disease classes and one healthy class. The research singles out that CNNs can be helpful in the accurate detection and classification of tomato leaf diseases and that the potential contributes to the advance of crop management practices, ensuring higher yields.

In this paper, Khamparia et al. [21] address an important question in crop-related diseases that cause a productivity reduction in agriculture and make a strong point for diagnosis and subsequent management of crop diseases. Although it has largely focused on many such techniques, including support vector machines and various image processing techniques, there still are scopes to innovate, particularly in approaches dealing with vision. This paper, through its realization, proposes adopting a hybrid approach to crop disease detection from leaf images that couples the technique of autoencoders with convolutional neural networks. A novel technique is proposed in this paper for combined use of Convolutional Encoder Networks and evaluated on a dataset consisting of 900 images obtained on three different crops showing five types of crop diseases. It can be observed through the experimentation process in the proposed architecture that significant accuracies are yielded, with some variations in the different epochs undertaken and in the sizes of convolution filters used. For instance, a model attains an accuracy of 97.50% by taking the size of the convolution filters as 2x2 for 100 epochs; with the filter size increased to 3x3, the improvement is seen to 100% in performance. This study shows that the potential for the detection and classification of crop diseases by far with deep learning techniques can be promising and further make a strong point toward the improvement of its sustainability and productivity.

In a more general sense, most works aimed to cover disease detection and classification in plants using a wide range of deep learning models and techniques. The proposed methodologies presented promising results from 86% to 98%, varying with the dataset used and disease classes as summarized in table []. Among the topics discussed are better disease recognition, real-time inference, and image segmentation, as well as the use of synthetic images for data augmentation, where the models were enhanced.

Table 2.1: Summary of Papers on Plant Disease Detection

Sr. No.	**Paper**	**Dataset Used**	**Technique Used**	**Accuracy**
1.	"End-to-End Deep Learning Model for Corn Leaf Disease Classification" by Hassan et al. (2022) [?]	PlantVillage (Corn)	EfficientNetB0 and DenseNet121	98.56%
2.	"Cardamom Plant Disease Detection Approach Using EfficientNetV2" by Sunil et at. (2021) [12]	Collected cardamom plant leaves	EfficientNetV2	98.26%
3.	"Deep Learning-Based Object Detection Improvement for Tomato Disease" by Yang Zhang et al. (2020) [13]	AIChallenger training dataset	Faster RCNN	98.54%

Continued on next page

Table 2.1 – Continued from previous page

Sr. No.	Paper	Dataset Used	Technique Used	Accuracy
4.	"Unsupervised image translation using adversarial networks for improved plant disease recognition" by Nazki et al. (2020) [14]	tomato plant disease recognition dataset	AR-GAN	86.10%
5.	"Cucumber leaf disease identification with global pooling dilated convolutional neural network" by Zhang et al. (2019) [15]	600 leaves of cucumber having common and infected leaves	GPDCNN	94.65%
6.	"Multilayer Convolution neural network for the Classification of mango leaves infected by Anthracnose Disease" by Uday Pratap Singh et al. (2019) [16]	Images acquired at SMVDU	MCNN	97.13%
7.	"Sunflower leaf diseases detection using Image Segmentation based on Particle swarm optimization" by Vijai Singh (2019) [17]	Self-acquired sunflower leaves	Particle Swarm Optimization Algorithm	98.00%
8.	"Deep Convolutional neural network based detection system for real time corn plant disease recognition" by Mishra et al. (2019) [18]	PlantVillage Dataset (Corn)	DCNN	88.46%
9.	"Performance analysis of deep learning CNN models for disease detection in plants using image segmentation" by Sharma et al. (2019) [19]	Tomato leaves images for disease detection	CNN	98.60%
10.	"ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network" by Agarwal et al. (2019) [20]	Tomato leaves images taken from PlantVillage Dataset	CNN	91.20%

Continued on next page

Table 2.1 – Continued from previous page

Sr. No.	Paper	Dataset Used	Technique Used	Accuracy
11.	"Seasonal Crops Disease Prediction and classification Using Deep Convolutional Encoder Network" by Khamparia et al. (2019) [21]	PlantVillage Dataset (Seasonal Crops)	DCEN	97.50%

In this area, the works related to plant disease detection and classification are coming up with several critical constraints which should be looked into and resolved in future research. These limitations play a crux role in fostering development in this field and designing better, stronger, and more efficient deep learning models catering to the subtle issues of plant disease detection.

The most visible limitation in nearly all current studies is the use of small datasets, which considerably reduces the generalization of the results to large and diversified populations. Small datasets are inherently unrepresentative and insufficiently diversified to even remotely capture the whole spectrum of variations and complexities found in real life. Therefore, models trained on such datasets will have limited generalization capacity when confronted with novel or unseen data. Small datasets might not capture the variability properly, especially in disease expression when signs may be subtle or incidences are rare. The performance of disease detection algorithms used on a small dataset can compromise effective performance when applied in practice.

Another important issue is that due to the subtlety of many symptoms and the tendency for diseases to hit in small parts of a plant, small data set sizes also will complicate the task of accurately detecting the affected region. In the agricultural setup, however, a lot of diseases affect only relatively small parts of a plant, and in such situations, timely action could require the determination of early signs of infection. However, subtle signs are often overlooked by algorithms trained on a very lean dataset; as a result, they are diagnosed late or wrongly. Furthermore, limited diversity in training data may hurt the model's generalization capability, and hence its performance on new or unseen patterns of diseases could be completely different from what is suggested by the validation set performance.

However, the other strong limitation is that this research focuses on the detection of a disease rather than identification of the disease at an early stage. Diseased detection is for sure so important in assessing the condition of plant health. On the other hand, early intervention is very important to this effect so that crop losses are lessened or should not occur. This will help them take timely measures for the control of the disease at an opportune moment with minimum chances of being spread and causing more damage to the crop.

Although most studies already conducted prioritize the development of models for detection of the disease in its advanced stage of infection, they miss the important feature of early identification. In this regard, therefore, there arises a need for urgency in terms of research priorities for the early detection of diseases and delving into innovative ways in which subtle symptoms of a disease can be captured at its inception. Also, current investigations in general do not make consideration of overall impacts of lighting and environmental conditions on image quality and disease perception. For example, day/night settings or wearing highly textured clothing or direct sunlight exposure might have a powerful influence on plant tissue appearance and symptoms' visibility. Most experiments lack statistical consideration of these factors; this often makes it a biased or unreliable prediction of models for other environmental conditions.

Moreover, disease development and progress depend on the environmental variables: humidity, temperature, and soil composition, which make disease diagnosis an ever more complicated task. With this in mind, future works will truly consider a comprehensive way for accounting for significance on the interplay of the different factors. There is general agreement that these are formidable limitations, so several attempts have been made in gaining the attention of the research community in trying to develop far more comprehensive and robust methodologies for the diagnosis of plant diseases. To this end, setting up larger and more diverse saving datasets with a massive range involving most of the species of plants and diverse disease types and leaf environmental conditions could be a big future line of research. The second step in this direction would be by using some advanced data collection techniques, such as remote sensing and crowdsourcing sources, to harness full-range variations available in real agricultural setups. Datasets from all these can be used to train the classifiers and to evaluate deep learning models that can then make one able to develop more accurate and easily generalizable algorithms for disease detection. Early disease detection can be done through advanced image processing techniques combined with domain knowledge, apart from the increase in dataset size. For instance, integration of the multimodal imaging methods like hyperspectral and thermal imaging can offer their complementary information in plant health and disease status. At the same time, the use of domain-specific features such as leaf texture and chlorophyll fluorescence additionally improves the discriminative power of deep learning models for disease detection. It would also improve the fusion of these approaches in research with impacts on developing more robust and effective algorithms for early identification of conditions to open up avenues for proactive management strategies in agriculture. Moreover, it is recommended that future studies focus on the lighting condition-invariant model and the model with different environmental conditions. This can be achieved by adopting strong image preprocessing techniques that must normalize the image intensity and help alleviate the effect of uneven lighting. In addition, one must investigate the adversarial training method to boost model robustness for environmental variations against domain shift. By introducing adversarial examples that emulate the changes in environmental conditions, adversarial training might enable the model to learn more robust and generalizable representations of disease patterns than are estimable from the original data. Interdisciplinary collaboration regarding the relevance of the research endeavors, agronomists, and agricultural practitioners should be taken to furthering the research works by finding result-oriented and practically feasible answers. Involvement of stakeholders in the entire research process may further help to shed light on practical challenges and constraints, thereby enabling the researchers to design measures to be in tune with farmers' needs. Besides, the collaboration of agriculture extension services and other collaborating stakeholders within the industry will enhance translating research information into implementable recommendations and technologies of value to farmers while being useful for realizing high productivity in agriculture. To summarize, although research in the area of plant disease detection has advanced significantly, there are some critical hitches and limitations that remain to be addressed for the development of robust and practically effective detection models. Addressing these issues will necessarily require a multidimensional approach, including data set expansion, innovative algorithmic techniques, and interdisciplinary collaboration. They provide important opportunities for proactive disease management in agriculture and important contributions to global efforts to guarantee food security and sustainability.

From such studies, it is clear that deep learning has a great potential to transform this domain of plant disease identification. With improved deep learning techniques, more in convolutional neural networks and autoencoders, researchers are quite arguably better placed to spearhead a new era in precision agriculture through the effective and efficient early detection and classification of diseases of plants. Deep learning provides the power to surmount the long-standing problems of traditional methods in disease detection, which are usually laborious, subjective, and many times visual inspection by human experts. With this power from deep learning, the algorithms can extract successfully complex patterns and features from digital images of plant leaves for rapid and reliable disease diagnosis. That is quite helpful because the challenges brought by plant diseases are always continuously diverse and in a way evolving. The deep learning algorithms, being retrainable, are thus up-to-date and effective in dynamic agricultural and environmental settings compared with conventional methods, which could lag behind the onset of new diseases or variations in the disease pattern. Another domain in which this deep learning could be utilized with regard to these new emerging technologies in general, ranging from remote sensing and drones to the Internet of Things, is for better tracking and monitoring of the state of crops at larger scales. These new technologies enable the rapid collection of high-resolution imagery and environmental data, which can then be integrated into deep learning models being used for fast real-time disease detection and predictive analytics. It has revolutionized disease detection, optimized agricultural practices, and maximized crop production while minimizing resource input and environmental impact. Deep learning algorithms extract insight from the analysis of enormous amounts of data related to plant health, soil, weather, crop performance, among others, to provide actionable recommendations that assure productivity and sustainability at the farm level. Such deep learning-based analyses can thus help to inform development and deployment of interventions such as the precise application of fertilizers, pesticides, and irrigation to realize optimized resource allocation, decreased input costs, and minimized environmental pollution. This means that this aspect of data-based farming will tilt toward the empowerment of a farmer in the optimal use of resources, reduction in risk, and building resilience to adverse environments, which will ultimately result in the best farming systems that are improved and made sustainable. The detection of plant diseases by using deep learning has a huge application in other important areas associated with global food security and human welfare. Its technology allows early detection of plant diseases and their control, assuring crop yields, reducing post-harvest losses, and ensuring the provision of good quality food for increasing populations. Further, the deep learning will target productivity optimization of agriculture to conserve nature by minimizing waste of resources. Deep learning for agriculture holds promises of revolutionizing the identification and management of plant-related diseases. By using AI techniques and a neural network, it can bring new opportunities to establish agricultural sustainability, increasing the ability of resilience in crop production and, therefore, food production security for the future generation. With current innovation and improvement happening in this very fast-growing field of techniques and methodologies toward deep learning, we are surely poised to take flight into a new era of agricultural sustainability and prosperity.

Healthy plants are central to ensuring global food security and, therefore, underpin the lives of billions across the world. It is estimated that the global population will reach ten billion by 2050, with food consumption required to increase by 70 to 100. Crop production has not only to remain constant but must increase and, therefore, the demand for agricultural growth is high in any possible way. However, these traditional ways of detection and management of plant diseases are often labor-intensive and inefficient, lagging fast expansion of diseases; the conventional approaches would give rise to serious economic losses, with harmful environmental impacts by overreliance on chemical treatments. Therefore, there is an urgent call for poor, rapidly changing technologies to bring about a revolution in detecting and managing plant diseases. Among these novel technologies are deep learning, computer vision, and machine-learning-based systems. These innovative solutions can help to automate the process of the disease detection that will be much more precise and in time in relation to multiple plant diseases. Therefore, the technology integration should allow developing DSSs that can not only differentiate between several types of diseases but also be able to learn on large datasets and predict outbreaks. This enables on-the-spot intervention, thereby reducing the proliferation of the disease and crop losses. Not only do these technologies enhance agricultural productivity and crop yields, but they also have other implications that are very crucial. It reduces dependency on chemical fertilizers and pesticides, thereby lessening environmental pollution. Early detection and control of plant diseases using these advanced technologies guarantees sustainability in agricultural practices while safeguarding ecosystems and biodiversity. Furthermore, these technologies safeguard food security even in vulnerable populations. It strengthens the capacity of an agricultural system to resist and be able to cope with changes like climate change and other environmental stresses. Advanced tools in disease detection and management also increase rural people's access to modern agricultural technologies. Such an access tends to level the playing ground and, with time, reduces economic inequalities that come as a result of technology adoption. These technologies support relatively small-scale farmers with low crop loss and better yields, thus supporting livelihoods and partially reducing poverty. In essence, it integrates deep machine learning, computer vision, and additional practices to further express transformational shifts toward greater efficiencies, sustainability, and equitableness of food systems that will enable guaranteed adequate and stable food production for future generations.

Chapter 3

METHODOLOGY

The proposed stages of the systematic planned approach are shown in 3.1. To identify and classify the category of plant leaf diseases. The major steps are summarized briefly as follows:

3.1 Proposed System

Although a number of methods are adopted to detect the disease of the leaves of the plant species, in this research, a number of machine and deep learning methods are employed to test if they would give much more accurate results. Indeed, very high resulting procedures for machine learning such as Random Forest, Decision Trees, and Support Vector Machines show correct results in classification performance of leaves being healthy or infected. But no doubt, they do not represent complex patterns of a disease in leaf images. On the other hand, CNN-based deep learning techniques have proven far better in dealing with image recognition issues, among which is diseased plant leaf identification. The identifications of many researchers have been alerted to the classification of various plant diseases using huge tagged database images of leaves. By integrating the strengths of the two paradigms, a comprehensive plant leaf disease location method can be developed. It is a multi-disciplinary method that not only improves accuracy but also provides a simple way to track and prevent disease epidemics in live and in real time in agro-systems. This approach supports food security in the world's communities, improves plant health and increases the yield in agriculture.

For the development of the model for detection of plant leaves disease follows the steps given below of the proposed system.

Step 1: Used PlantVillage dataset and distributed into the train, valid and test set..

Step 2: Performed image scaling, image normalization and corruption detection to ensure consistency and integrity within the data.

Step 3: Integrated conventional image processing techniques and data augmentation to accurately classify and carefully extract relevant features.

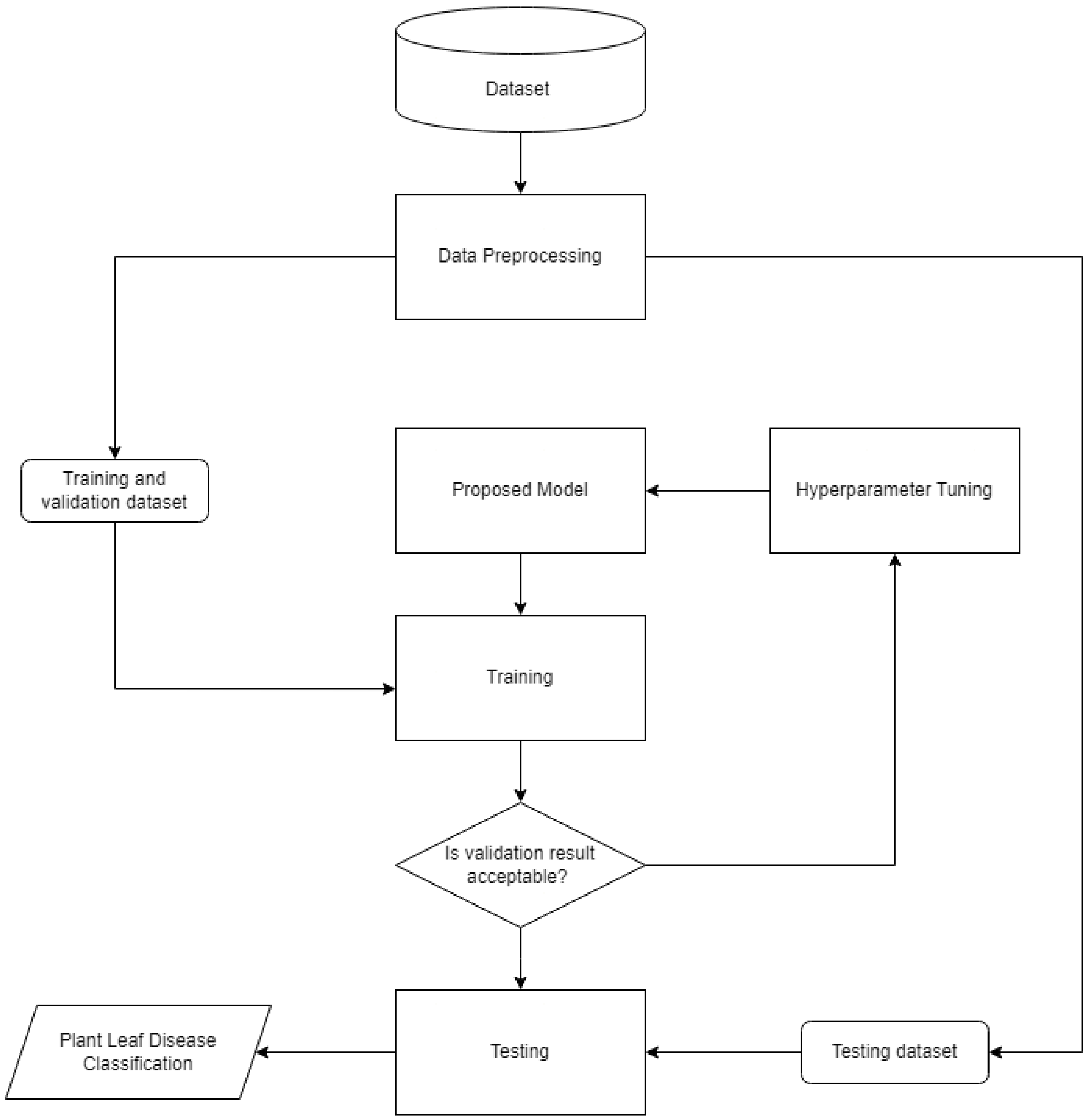

Figure 3.1: Proposed system

Step 4: Combined EfficientNetV2 architecture with extra layers to further increase its flexibility and effectiveness.

Step 5: Used transfer learning and hyperparameter optimization to enable model improvement.

Step 6: Comprehensive assessment to make the reader understand how well it performs in real world scenarios.

3.2 Data Preprocessing

Before training the model, the dataset was preprocessed in order to achieve uniformity and consistency. This was basically done through eliminating corrupted files, scaling images, and normalization of lighting. It was later split into a training dataset, testing dataset, and validation dataset. The entire reasoning was to cause less inconsistency and bias in data types which were going to facilitate better training of performance in the model. Doing this, we guaranteed that the dataset was standardized as a clean input dataset with consistent values so that the model could learn well generally on unseen data [22] [23] [24]. Data preprocessing encompasses a lot of steps that one has to go through in preparing the dataset for model training.

The steps we followed in our research is as follows:

Step 1: Detection and cleaning of corrupt files is a major step involving the search for image files in the dataset to determine whether they are in the right format or if any corruption is found. Corrupted files are then removed from the dataset.

Step 2: Rescale All Images to 224x224 Shape For consistency, images in the dataset across different fashion items will all be rescaled to one common size of 224x224. This scales up to a common size for machine learning algorithms that take into account image proportions (e.g., Convolutional Neural Networks).

Step 3: Normalization to Mitigate Lighting Variations Varying intensity normalizes effect by first mapping images to uniform intensity range usually [0-1]. Normalization enhances the behavior of CNNs with respect to [0-255] which may feel more natural.

Step 4: Dataset split further into the training dataset, testing dataset, and validation dataset. Many standard splits give one 60% for a training set, 20% for a testing set, and 20% for a validation dataset to enable the separation of the datasets regarding model training, evaluation, and performance assessment.

Down the line, these preparation steps are taken to standardize and normalize a dataset in a way that it can enhance the performance and generalization properties of a model.

3.3 Data Augmentation

Data augmentation is a form of feature representation that increases variation and prevents overfitting by extending the type of training data. We used data-augmentation transformation, which includes things like random rotation, random horizontal and vertical flip and random color jitter as shown in 3.2. These augmentations bring in more diverse instances, making the model able to generalize for wider scenarios.

3.4 Model Description

EfficientNetV2 [25] architecture is a hybrid model which has been known for its low computational cost in comparison to other state-of-the-art architectures for

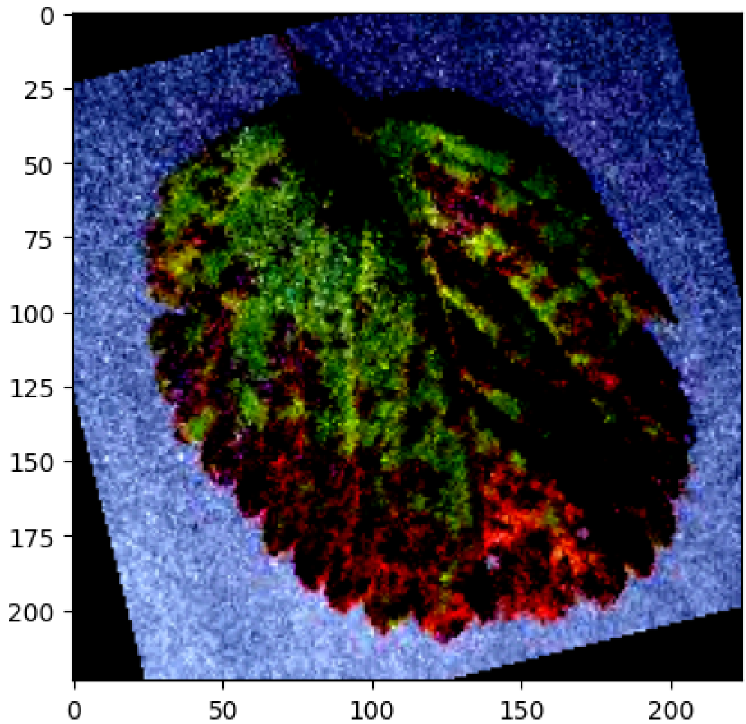

Figure 3.2: Sample of image after pre-processing and data augmentation

images. Other layers include Batch Normalization, Fully Connected, Dropout, and Output, which add to the ordinal performance of this model as shown in 3.3. Further to that, it concatenates an Output layer that has a softmax activation function added to it. It is of output size 38 corresponding to the count of classes in the classification problem. After applying the Softmax activation function, it converts the raw output values of its preceding layer into a probability distribution across classes in such a way that each class probability means that any given input belongs to this derived probability class. This pretrained model can easily distinguish between actual healthy leaves and leaves with diseases such as spot disease, late blight disease, and others through low-level texture patterns and high-level semantic features captured from input images. Therefore, the architecture would ensure positive trade-offs between computational efficiency and classification accuracy, of paramount importance in establishing a system of classification that can be trusted for disease diagnosis in leaves.

3.5 Feature Selection and Hyperparameter Tuning

The solutions in conventional image processing were combined with deep learning methods to extract features useful in capturing low-level texture patterns and high-level semantic information from input images for the effective classification of diseases. This helps the model to recognize complex small details and the pattern

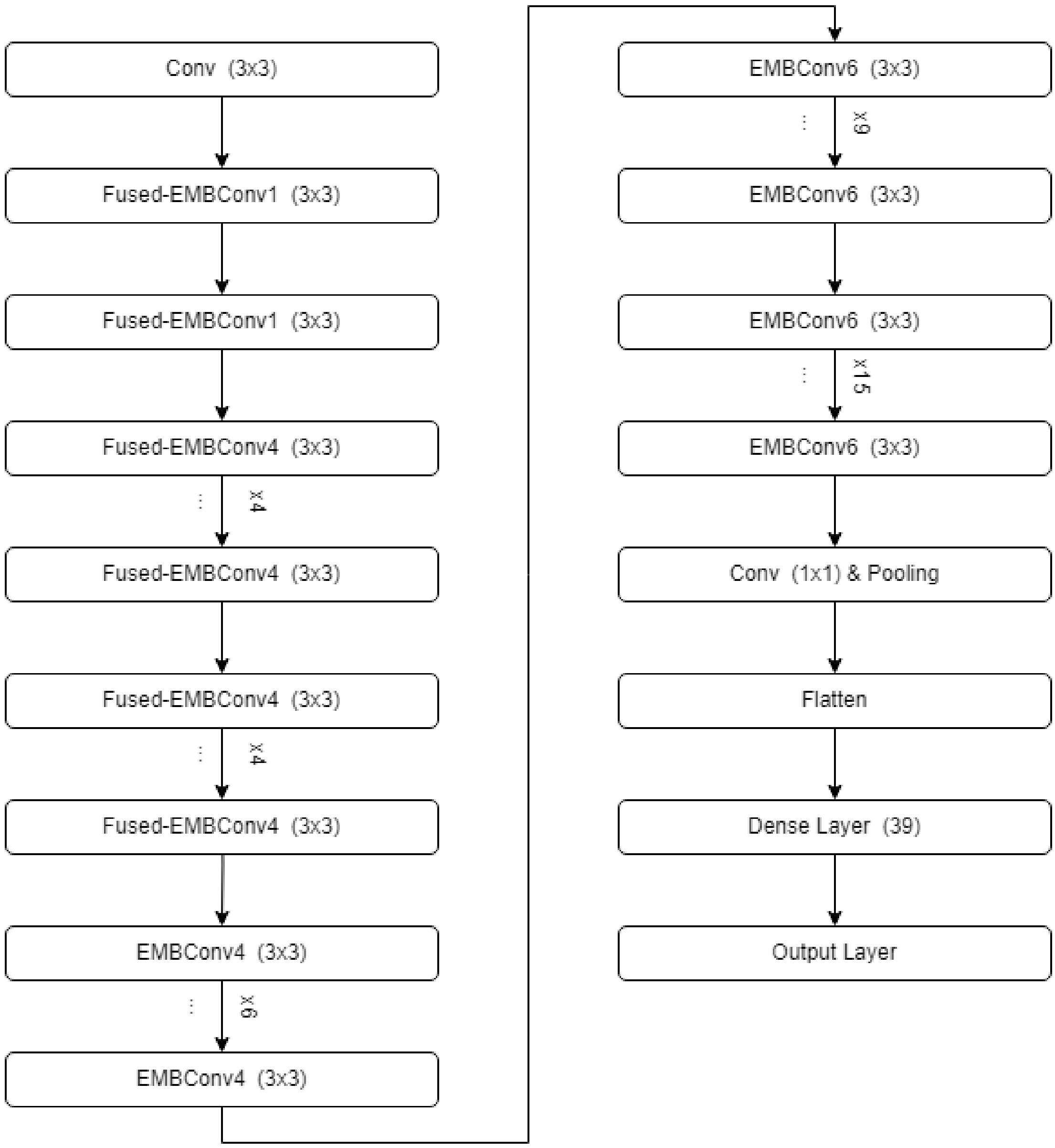

Figure 3.3: EfficientNetV2 architecture

nature of the disease, thus improving the efficiency of classification. All-inclusive feature selection was conducted, where the model learns important distinguishing features in leaves of healthy and diseased plants, thus contributing to the overall effectiveness and reliability of this classification system. The current study fine-tuned hyperparameters through iterative experimentation and validation toward improved performance and generalization in a classification model. The learning rate, optimizer, batch size, and dropout rate were tuned to get high accuracy and strong robustness against changes among agricultural conditions. In addition, it is also better adapted for different data distribution and complexity, which thus improves how well it can determine disease in actual cases.

The following are the hyperparameters that we assign:

1. **Learning Rate:** The learning rate is an important hyperparameter because it determines the size of steps that the optimizer takes to update the weights of the model during training. In the demo below, one would set a learning rate of 0.0005, and at each step of training, a slight adjustment in the weights would be made.

2. **Loss Function:** The loss function will be taken to account for the difference between true labels and the predictions. Our loss function is a categorical cross-entropy loss, which would be proper for multi-class classification with mutually exclusive classes.

3. **Optimizer:** It nods at the optimization algorithm for updating the model weights based on the computed gradients. We chose Adam as our optimizer since it's pretty good with sparse gradients, and it's mostly in use with deep learning models.

Chapter 4

EXPERIMENTATION DETAILS

4.1 Environmental Setup

All preprocessing tasks were done using an Intel(R) Xeon(R) CPU@2.20 GHz and are quite strong and high-performing in information handling. This robust setup of the CPU enabled effective data manipulation that would allow for the processing of large and complicated datasets with absolute ease and speed. This involves data cleaning, normalization, augmentation, and transformation. All of this is done with a highly efficient Intel Xeon processor. This stage is thus critical in putting raw data in preparation state for further tasks of deep learning, such that it feeds through the networks with optimum quality and relevance. Parallel to this, the GPU setup with the Nvidia Tesla P100 was very core in accelerating the process of training deep neural networks. With excellent parallel processing ability, indispensable time taken by very sophisticated models to train for practical use with neural networks has been considerably reduced. The Nvidia Tesla P100 architecture can be described as a heavily compute-intensive task and tailored for deep learning purposes, which makes it the best option for us. Its architecture permits numerous components to be carried out at the same time, causing massive parallelization in heavy computation carried out during training deep learning models. Combination of CPU and GPU resources on Kaggle delivers a very effective and stable work environment in modeling. All preprocessing was done using an Intel Xeon CPU to ensure that the data was in the best condition possible before it was fed into the neural networks. The Nvidia Tesla P100 GPU accelerated the process of training to allow for experiments with a high number of hyperparameter variations. Such a combination allowed rapid iteration within various models and their respective configurations. The preparation stage also included the resizing of the images to a unified size, cropping, and data normalization. Augmentations for rotate, flip, and color adjust were incorporated to augment data and make the training set more diverse, hence the model would generalize better with new unseen data. This made possible several capacities, such as running many threads at once, which considerably sped up the whole of the preprocessing process. Then, after preprocessing, it was fed to the deep neural network for training. The huge memory bandwidth and capacity of Nvidia Tesla P100 make it quite easy to handle large batches of data, thereby leading to time-efficient training, even with extensive datasets. Such an advanced architecture made it capable of running neural network operations very effectively in dealing with the complex computation types like forward and backward propagation. The classification results obtained by the set of blazing, high-performance computational resources accelerate the very process of model development and overall productivity of a team when combined. Both CPU and GPU resources are considered critical for the development of high-accuracy and robust deep learning models in smooth iterations, coupled with large experimentations. In so doing, hyperparameters are further optimized, network architectures are optimized, and there is extensive validation of the models in a manner that consequently makes sure there are better performance metrics.

4.2 Dataset Description

This research was carried out on the large dataset of plant leaf images in PlantVillage [26], which approximately has 61,486 plant leaves images of 39 different classes as shown in 4.1. The model can learn from different agricultural conditions to enhance its capacity in distinguishing leaf blight from healthy leaves. The representative dataset used for this study revealed that the dataset to be used might have the potential to address actual issues coming up in agriculture. The investigation was based on the PlantVillage dataset, which is an all-inclusive repository with various images for training and evaluating classification models as shown in 4.2. A comprehensive coverage is required in the training and evaluation processes, whereby the classification models must be well reliable and effective under different numerous agricultural conditions.

4.3 Evaluation Metrics

The performance is measured with the study by means of quality metrics: accuracy, precision, recall, or F1-score, among others. This has served to know how well the model could separate healthy leaves from diseased ones. Additionally, it was shown what the model did well and at which weak points, making step-by-step improvements possible. It is through such careful assessment that a plant disease identification system based on sound classification would be reliable and effective. The following three metrics were used in such experiments to assess the model which uses True Positive (TF), True Negative (TN), False Positive (FP), and False Negative (FN).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

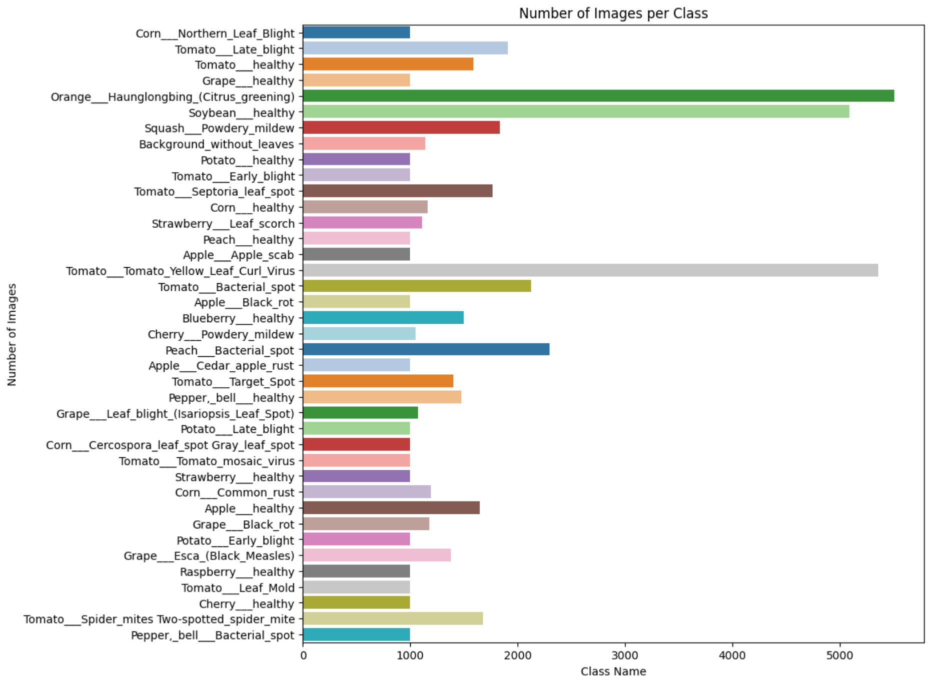

Figure 4.1: Visualization of distribution of 39 classes of plant leaf disease

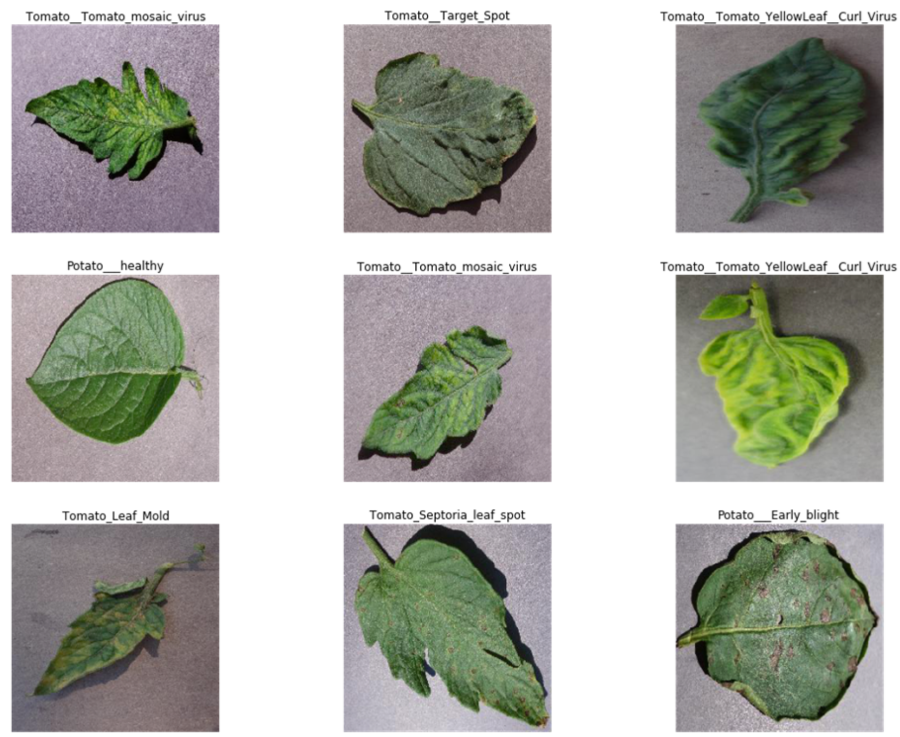

Figure 4.2: PlantVillage dataset samples

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{4.4}$$

A confusion matrix summarizes the prediction of a classification problem. It holds count values containing numbers of correct and false predictions for each class and its diagonal is where the value is the actual accuracy. All other deviations from this nominal indicate incorrect predictions for the diagonal value. The classification report presents the representation of the main classification metrics on a per-class basis. This allows for more intuitive inference with respect to the behavior of the classifier over global accuracy, where it would mask functional weaknesses that apply to just one class in a multi class problem.

Chapter 5

RESULTS AND DISCUSSION

The model was trained for about 5.7 hours in the experimenting phase by running 70 epochs every time as shown in 5.1. Now, each point of time is closely monitored, and this model was able to learn to be at a wonderful training accuracy of 99.40%. Equally important was the accuracy of the testing 99.24%, way higher than the performance benchmarks reported by the last state-of-the-art methods. Such a result attesting to the strategy's robustness and effectiveness in making exact classifications of plant diseases is regarded as the major stride in agriculture technologies. We monitored the change of accuracy and loss of this model as training proceeded through more and more epochs. These learning dynamics of our model become more intuitive if one visualizes this information through time. The constant increase in accuracy with a small decrease in every epoch loss function can be observed. This is further evidence that our model was really able to capture the base of hidden patterns and features within the training data, providing great classification efficacy. We extended the performance verification of our model to go far beyond mere accuracy metrics. We further analyzed the confusion matrix deeply to bring out more insights into the classification capability across the different classes of diseases. The latter allows performance to be gauged in classifying correctly instances from each class, with a revelation of places where misclassification might occur. From a close look at the confusion matrix, we were able to identify some specific classes in which the model was doing exceptionally well and others where we could work on improving it. We have also done qualitative analysis upon the model prediction. On checking correct and incorrect predictions, it was established that patterns and trends were repeated in different samples. Qualitative investigation, therefore, allowed understanding the model decision-making process and pinpointing areas on which we suggest laying strategies about how to improve. Understanding the driving forces between those correct and incorrect predictions has enabled the model to focus on classification performance and accuracy.

Table 5.1: Training and Validation Loss per Epoch

Epoch	**Training Loss**	**Validation Loss**	**Elapsed Time**
1	0.3391	0.0967	4m 58s
2	0.0899	0.0511	9m 49s
3	0.0698	0.0606	14m 42s
4	0.0610	0.0426	19m 33s
5	0.0585	0.0447	24m 24s
6	0.0576	0.0395	29m 16s
7	0.0511	0.0524	34m 8s
8	0.0486	0.0371	39m 1s
9	0.0439	0.0400	43m 54s
10	0.0485	0.0347	48m 46s
11	0.0419	0.1048	53m 38s
12	0.0465	0.0383	58m 31s
13	0.0400	0.0488	63m 24s
14	0.0411	0.0469	68m 17s
15	0.0404	0.0542	73m 9s
16	0.0391	0.0555	78m 2s
17	0.0362	0.0544	82m 56s
18	0.0406	0.0330	87m 51s
19	0.0377	0.0546	92m 44s
20	0.0359	0.0719	97m 37s
21	0.0368	0.0355	102m 33s
22	0.0340	0.0394	107m 28s
23	0.0353	0.0473	112m 22s
24	0.0348	0.0291	117m 17s
25	0.0378	0.0623	122m 12s
26	0.0357	0.0327	127m 7s
27	0.0333	0.0396	132m 2s
28	0.0334	0.0526	136m 58s
29	0.0343	0.0407	141m 55s
30	0.0320	0.0381	146m 50s
31	0.0312	0.0317	151m 46s
32	0.0333	0.0320	156m 44s
33	0.0318	0.0371	161m 41s
34	0.0276	0.0263	166m 37s
35	0.0324	0.0343	171m 34s
36	0.0328	0.0250	176m 30s
37	0.0322	0.0349	181m 28s
38	0.0256	0.0307	186m 22s
39	0.0313	0.0346	191m 17s
40	0.0285	0.0497	196m 12s

Continued on next page

Table 5.1 – Continued from previous page

Epoch	Training Loss	Validation Loss	Elapsed Time
41	0.0289	0.0406	201m 8s
42	0.0323	0.0412	206m 3s
43	0.0270	0.0439	210m 58s
44	0.0305	0.0441	215m 55s
45	0.0296	0.0354	220m 50s
46	0.0283	0.0602	225m 45s
47	0.0306	0.0326	230m 41s
48	0.0310	0.0271	235m 37s
49	0.0301	0.0358	240m 32s
50	0.0265	0.0215	245m 29s
51	0.0276	0.0342	250m 24s
52	0.0285	0.0340	255m 20s
53	0.0243	0.0365	260m 16s
54	0.0307	0.0344	265m 10s
55	0.0282	0.0278	270m 5s
56	0.0283	0.0337	275m 0s
57	0.0266	0.0329	279m 55s
58	0.0264	0.0389	284m 51s
59	0.0252	0.0357	289m 46s
60	0.0255	0.0480	294m 41s
61	0.0276	0.0265	299m 36s
62	0.0259	0.0326	304m 32s
63	0.0270	0.0285	309m 27s
64	0.0262	0.0268	314m 22s
65	0.0272	0.0287	319m 18s
66	0.0290	0.0289	324m 12s
67	0.0249	0.0318	329m 8s
68	0.0257	0.0303	334m 3s
69	0.0240	0.0319	338m 57s
70	0.0281	0.0360	343m 52s

During the training process, detailed logs of the loss and accuracy values at each epoch were made in order to properly make notes on model development and the evolution of performance. With this, we could fine-monitor the model in respect to both the optimization process and learning dynamics. The logged metrics are further visualized to describe a clear view of the learning curve of the model. In Figure 5.1, the development of the accuracy of the proposed model is given, while the correspondent trends in the loss function are shown in Figure 5.2. Such graphical views explain a lot about how the convergence of the model occurs over time and in what way it molds around the training data. For the training accuracy, it starts increasing slowly as a function of epochs, representing the model's initial learning on the data. With progress in the epoch, steeper growth in the trend shows that the model would probably capture deeper insights from the training dataset. The curve of accuracy on testing data generally follows the same arc. Small higgling around may also be seen in both aspects because of the variations in validation data, where the amplitude of such fluctuations dies down with an increase in the number of iterations. It indicates generalization over unseen data by the model and not just overfitting on the training set. At the same time, the corresponding trend in the loss values for both the training and testing phases. What is evident here is the loss function, which is the method of error measurement between the predicted and actual values, monotonically decreases in a logarithmic manner with an increase in the number of epochs. A model loses, implying that as the number of epochs increases, it has a constantly improving capability to minimize prediction errors. Further convergence of loss values not only demonstrates proficiency in data learning but also efficiency in optimizing the parameters. These detailed logs and visualizations enabled the obtaining of fast feedback, which is quintessential for hyperparameter tuning and adjusting the model architecture. Fast feedback on the convergence rate, performance, and potential overfitting greatly influences changes to the learning rate or even reforms in the network structure to squeeze out greater performance from the model. In sum, trends in accuracy and loss values can give an indication of some problem during training: either overfitting or underfitting. If the training accuracy keeps on increasing, and at the same time, the testing accuracy starts to decrease or stays constant, this might indicate overfitting. This would need some kind of regularization technique or further data augmentation. Furthermore, the epoch-log level allows probing into specific training dynamics. For example, individual spikes or plummets in accuracy/loss curves will allow us to find out what actually causes these anomalies, which could be due to certain batches of data or some prevalent features of the dataset.

This fine-grained monitoring not only ensures transparency in the training process with the model but also reacts better to any types of anomalies during the training process. In further addition, performance key metrics of Accuracy and Loss, Precision, Recall, and F1-Score for different classes are also monitored and analyzed. Such multifaceted evaluation throws more coherent light on model performance. The precision is the measure of the proportion of true positive predictions out of all positive predictions that were made. Remember, on the other hand, measures how good the model is at finding all relevant occurrences in the dataset. In turn, F1 is the harmonic mean of precision and recall—a balanced metric with respect to false positives and false negatives. Observation of these metrics across a variety of classes would enable the identification of specific strong and weak points of the model. It is known that most of the plant diseases where symptoms are clear or characteristic show a higher precision and recall, but diseases with weak or overlapping symptoms generally don't perform well. A close look at class-specific performance is very important for further tuning the model in order to be all-class-suitable. The other is that the confusion matrix provided quite a fine-grained description of how the predictions of the model were, classifying into each class. Insight into the occurrence of most common misclassifications

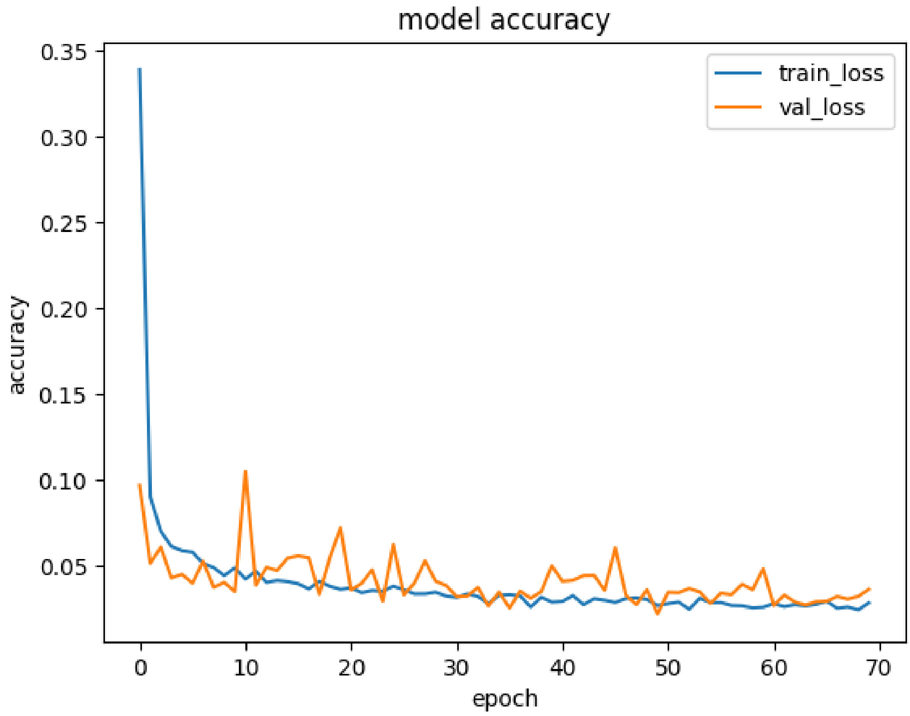

Figure 5.1: Train accuracy vs Test accuracy

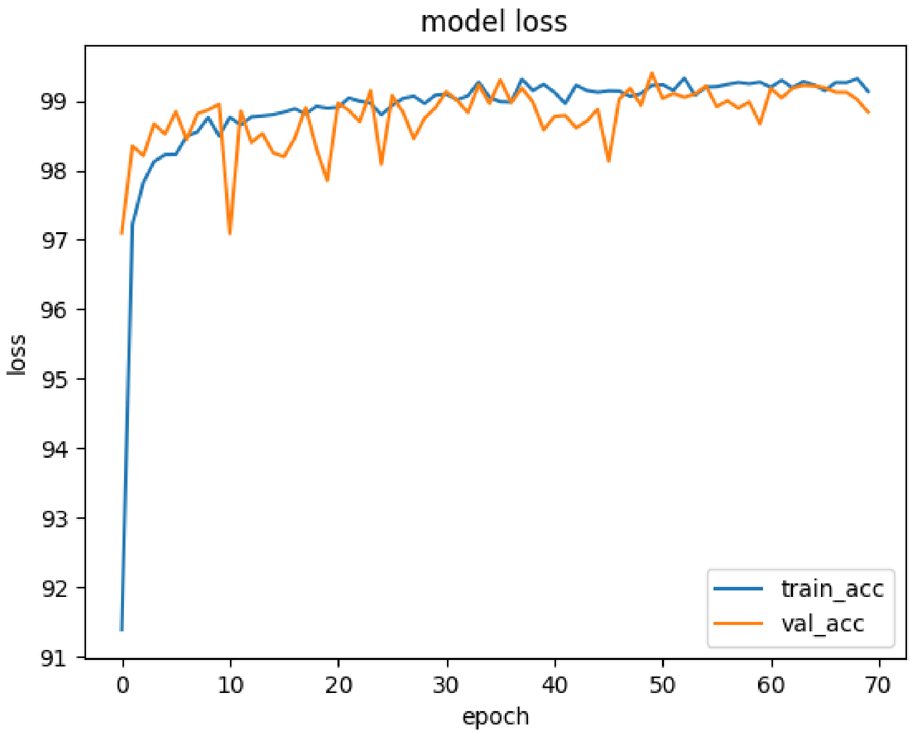

Figure 5.2: Train loss vs Test loss

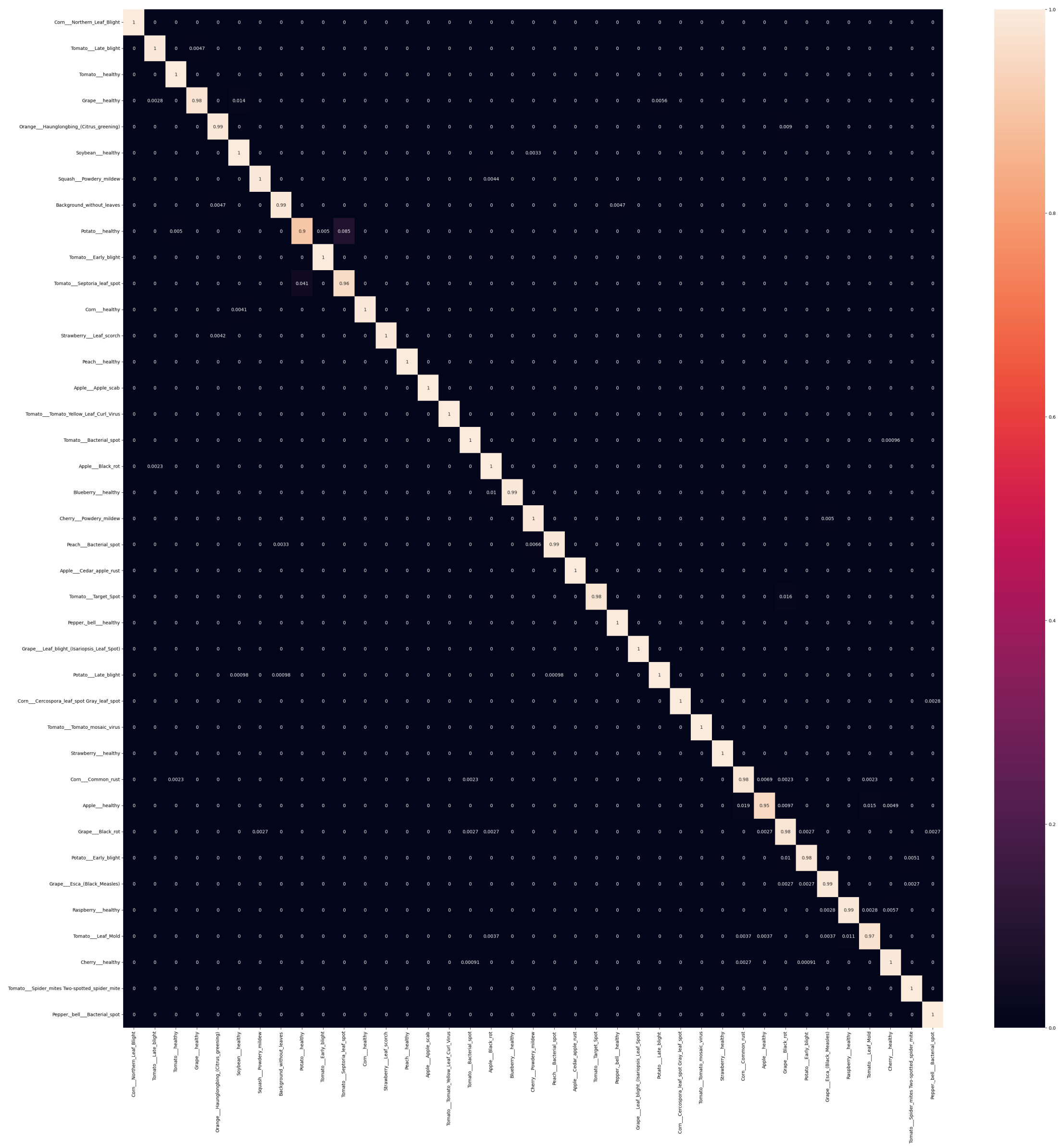

Figure 5.3: Confusion Matrix

	precision	recall	f1-score	support
Corn___Northern_Leaf_Blight	1.00	1.00	1.00	213
Tomato___Late_blight	0.99	1.00	0.99	211
Tomato___healthy	0.99	1.00	1.00	202
Grape___healthy	1.00	0.98	0.99	354
Orange___Haunglongbing_(Citrus_greening)	0.99	0.99	0.99	222
Soybean___healthy	0.98	1.00	0.99	302
Squash___Powdery_mildew	1.00	1.00	1.00	227
Background_without_leaves	0.99	0.99	0.99	215
Potato___healthy	0.96	0.90	0.93	199
Tomato___Early_blight	1.00	1.00	1.00	222
Tomato___Septoria_leaf_spot	0.92	0.96	0.94	196
Corn___healthy	1.00	1.00	1.00	241
Strawberry___Leaf_scorch	1.00	1.00	1.00	238
Peach___healthy	1.00	1.00	1.00	270
Apple___Apple_scab	1.00	1.00	1.00	214
Tomato___Tomato_Yellow_Leaf_Curl_Virus	1.00	1.00	1.00	196
Tomato___Bacterial_spot	1.00	1.00	1.00	1038
Apple___Black_rot	0.99	1.00	0.99	430
Blueberry___healthy	1.00	0.99	0.99	198
Cherry___Powdery_mildew	0.99	1.00	0.99	202
Peach___Bacterial_spot	1.00	0.99	0.99	301
Apple___Cedar_apple_rust	1.00	1.00	1.00	201
Tomato___Target_Spot	1.00	0.98	0.99	186
Pepper,_bell___healthy	1.00	1.00	1.00	206
Grape___Leaf_blight_(Isariopsis_Leaf_Spot)	1.00	1.00	1.00	213
Potato___Late_blight	1.00	1.00	1.00	1020
Corn___Cercospora_leaf_spot Gray_leaf_spot	1.00	1.00	1.00	360
Tomato___Tomato_mosaic_virus	1.00	1.00	1.00	212
Strawberry___healthy	1.00	1.00	1.00	180
Corn___Common_rust	0.98	0.98	0.98	437
Apple___healthy	0.98	0.95	0.96	206
Grape___Black_rot	0.97	0.98	0.98	377
Potato___Early_blight	0.98	0.98	0.98	197
Grape___Esca_(Black_Measles)	0.99	0.99	0.99	366
Raspberry___healthy	0.99	0.99	0.99	352
Tomato___Leaf_Mold	0.98	0.97	0.98	271
Cherry___healthy	1.00	1.00	1.00	1093
Tomato___Spider_mites Two-spotted_spider_mite	0.99	1.00	0.99	199
Pepper,_bell___Bacterial_spot	0.99	1.00	1.00	331
accuracy			0.99	12298
macro avg	0.99	0.99	0.99	12298
weighted avg	0.99	0.99	0.99	12298

Figure 5.4: Classification Report

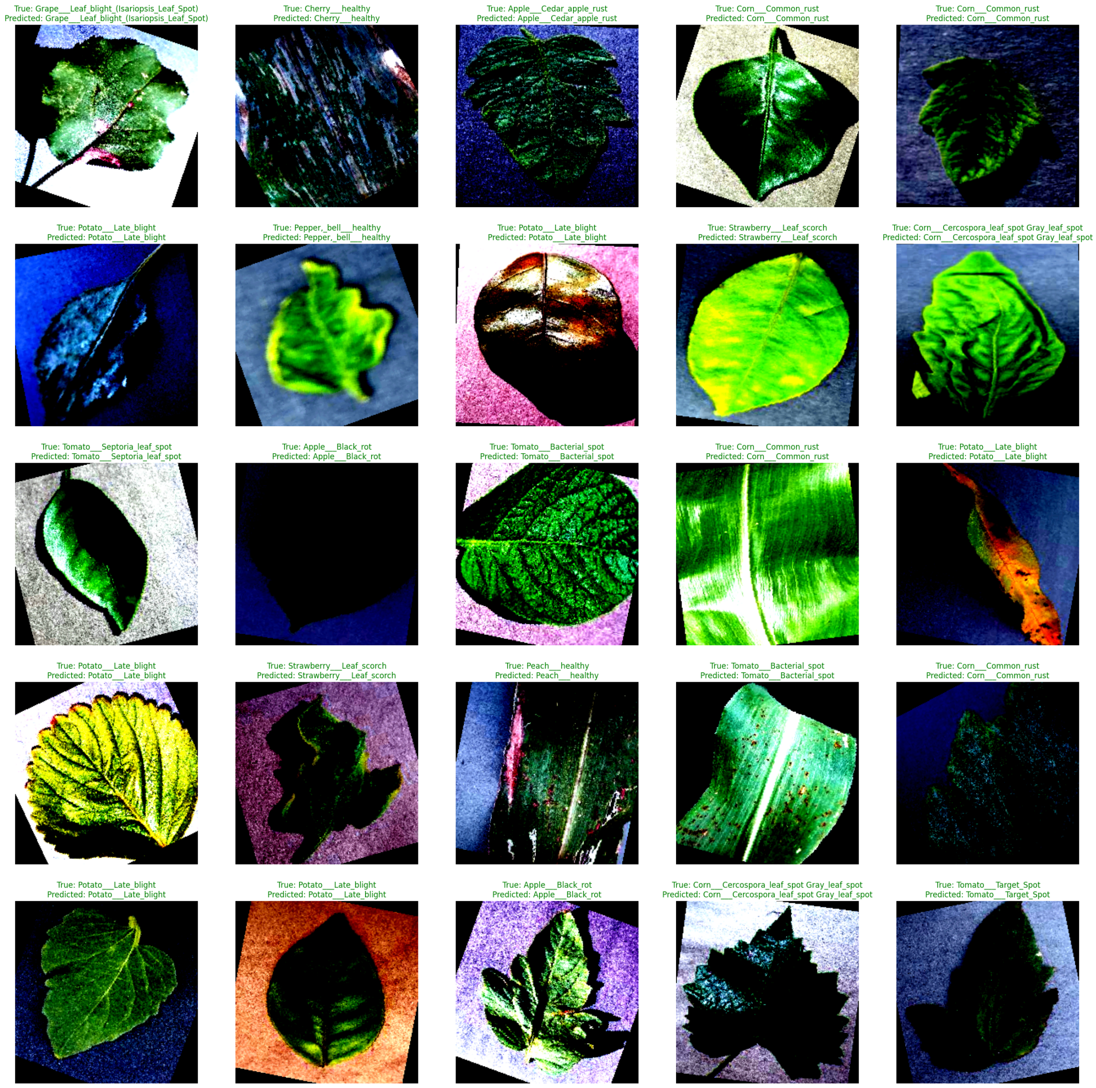

Figure 5.5: Samples of predicted images

enabled guidance on improvements, like if two diseases are very commonly confused, maybe visual similarity was introduced at an additional feature or data preprocessing step. Qualitative assessment was made by examining some of the predictions. Both the right and wrong predictions are checked to understand how the model works and to support quantitative analysis. The analysis helps unveil dependencies or features the model is strongly dependent on during classification. Knowing these would enable one to make informed adjustments to the model, like improving on the specific feature extraction methods or adding domain-specific knowledge to improve the accuracy. This overall view was so insightful that the classification report, with precision, recall, and F1-score for each class, gave such an insights view into the general performance of the model. To be precise, this report has helped point out those areas where strength lies and which need modifications within the classifier model before reiterating the whole processes toward optimization. For example, classes having low F1 scores were further examined to identify underlying challenges and devise strategies for enhancement. In general, this extensive logging and detailed analysis of metrics at each epoch played a pivotal role in fine-tuning the models to achieve high performance. This helps to visualize the trends of change in accuracy and loss, along with detailed metrics on class-specific performances, giving one a good view of model capabilities and room for further improvement. All these holistic measures ensured that the model had a high rate of accuracy and at the same time generalized well with new, unseen data, making for a robust tool for plant disease classification.

The confusion matrix gives a pictorial view of checking the number of times the prediction was right from the real labels, and it helps to check on the strengths and weaknesses of the classifier by how good the model has discriminated against different classes. Hence, the errors that the system commits, such as false positives or false negatives, determine what to work on. In classification models, the confusion matrix is considered a powerful evaluative tool. It gives a breakdown of what a model predicted as well as the real features against each class. Through such comparisons, one gets to know the behavior of the model and highlights which areas are an easier task for the model and which ones it struggles with. In this work, a confusion matrix of the model is exposed, revealing some critical insights to the performance of the model over different classes of plant diseases. Each cell of the matrix is made up of cases that the given model predicted to be part of a certain class; however, in fact, they were assigned from another class or the same one. The diagonal elements of the matrix give the number representing the right decisions for each class, and the off-diagonal elements give the misclassifications. We can then find which classes are actually confusing the model by observing the confusion matrix. More number of misclassifications between any two classes may indicate some kind of similarity between these two classes, making it difficult for the model to distinguish. This perhaps will lead toward further refinements of the model, say features or selection and tuning of the hyperparameters of the classification algorithm. The confusion matrix is also instrumental in the quantification of model performance in terms of precision, recall, and F1-score for each class. In this regard, the precision of a measure states how well the model's positive predictions are in alignment with true positivity, characterizing the accuracy of positive prediction from the model. Recall describes a measure of true positive prediction from all genuinely positive instances, which reflects that the potential of the model can capture all relevant instances from any particular class. The F1-score harmonizes precision and recall, hence provides equilibrium in the evaluation of both false positives and false negatives.

Thus, the confusion matrix 5.3 in this study resulted in high precision and high recall for most classes, therefore meaning the model recalled and was able to identify almost all types of plant diseases. However, the drop in precision and recall values from some classes might be indicative that the model has to face more difficulty in relation to those particular classes. Such information becomes relevant in guiding future modifications of the model, say data collection specifically for the most challenging classes or more advanced feature extraction methods. This is also a very important aspect of the confusion matrix, revealing how to consider the class distribution in training data. If the under-representation of low classes is found in the training data, the model may learn poorly with respect to these classes, which will eventually drive down the precision and recall for these classes. Possible ways to deal with class imbalance would be through the use of the confusion matrix or methods such as data augmentation for generating more instances for the underrepresented classes or an implementation of a class-weighted loss function, which gives more importance to the underrepresented classes. Therefore, in addition to showing the performance of the model, a confusion matrix will point out specifically what types of errors a model can do. An example of a false positive is when some class is predicted by the model while in the actual label it's an entirely different label. A false negative is also a case where a particular class is in question, but the model fails to predict it, which is the actual label. One can deduce from the confusion matrix which classes have high false positives or high false negatives and then remedy these errors. For example, a high rate of false positives for one class in the model may translate to a lack of distinctiveness of features being used for recognition with respect to that class; hence, the model may misclassify instances of other classes as that class. This is a problem in need of repair through feature extractions refinement—for example, by using sophisticated techniques in capturing more distinctive features for that class. In the event that false negatives for a class are many, most likely the model is insensitive to representing that class, which in turn will lead to the missed instances of that class. We can do that by adjusting the classification threshold of the model or techniques like over-sampling and data augmentation to add more examples of that class in the training set. Further, the confusion matrix details summarized views regarding how the model performed across all classes. This gives a more holistic view of how the model is doing when computing measures that average the measures across all classes. These metrics provide a worldwide perspective on the strengths and weaknesses of a model that leads to further improvements in a holistic way to enhance performance.

More detailed sample prediction analysis was performed to understand the working and performance of the model for both the correct classification and the misclassification of the samples. Figure 5.5 walks the reader through a complete look at the model's capability to distinguish between a diseased and a healthy plant leaf accurately. This step becomes quite essential because practical examples of the success and failure of the model have a significant meaning in understanding the overall model performance. The correct predictions are proof of the capability of the model to identify different plant diseases under various situations. For example, being provided with clear and distinct images of diseased leaves, the model will easily make the correct identification of the disease with high confidence. These examples should be seen as proof of the excellent performance and high reliability of the model in cases where disease symptoms are well-visible and can be distinguished from healthy leaf patterns. However, the misclassifications are as crucial as the ones that show the model's limits and the cases in which it can fail. Closer scrutiny of these incorrect predictions allows the unearthing of various patterns and possible pitfalls. For instance, the disease-related symptoms may be subtle and hence hardly detected by the model or partially masked by confounding factors like overlapping leaves, shadows, or different lighting conditions and misclassified in the images. Herein, however, lie the challenges: the model must detect and classify diseases, but most times, it will be from less salient visual evidence, usually resulting in misclassification. This is supported also by the observation that some diseases are more easily confused than others; this is the case if the symptoms of different diseases are, in nature, visually similar, say, through showing similar patterns of leaf discoloration or similar lesions. For example, diseases that yellow leaves or make spots on them at times can be easily confused with each other, and some erudition is critical as it points to the need for increased levels of the model's performance to capture and distinguish finer detail and subtle difference between disease symptoms. To do this, we can use the following approaches:. One of the possible ways is by increasing the diversity of the training dataset—including more challenging images with a variety of light, angle, or occlusions of the objects. This will make the model more robust and adaptive toward real-world conditions, as such variations are common. Further, including advanced image pre-processing techniques that enhance the visibility of the disease symptoms can also help improve the model's performance. For instance, such techniques as image segmentation, contrast adjustment, and noise reduction bring to light the critical features of any diseased leaf, which makes it much more accessible for the model to catch those key features and classify them accordingly. Another viable improvement is creating a more complete heterogeneous deep model for the processing of complex and delicate features. For example, multiscale feature extraction helps capture both the details and the fine points of the images. This will enable the model to better distinguish between diseases that tend to show very similar visual symptoms. By further using the attention mechanisms within the model, the latter will be able to focus on areas of the image where the disease can manifest itself subtly and otherwise might not be attended to. The feedback from classification analysis is saying that the labeling of training data should be more detailed and refined. In some cases, the current labels might not be sufficiently granular, leading to confusion between diseases with overlapping symptoms. By providing more detailed and specific labels, the model can learn to distinguish between these subtle differences more effectively.

This fact is further expounded through the classification report given in Figure 5.4, generalizing a detailed breakdown of the model's performance across several classes. This report churns out important metrics, including precision, recall, and f1-score for every individual class, thereby making quite meaningful and nuanced the view on the capacity of the model overall. An exploration of these metrics will put us in a good position to know specifically in which classes the model is performing well and in some other classes where more optimization or improvement is needed. Precision is a measure that states what proportion of the positive predictions given by a model are true positive out of all predictions for positives. If precision is high, the model will make fewer false-positive mistakes. Given the fact that high precision is achieved for certain classes, from there it means that detection of those diseases is done with high accuracy. For example, the classification of diseases that give out specific visual symptoms can be done with a great deal of precision since the model is able to clearly differentiate from healthy leaves or other classes of diseases. Recall, on the other hand, measures the number of true positive predictions in relation to all actual positives, emphasizing how well the model can capture all relevant instances of a class. High recall means that the model indicates fewer real instances of the disease being skipped, which is very important in cases where the model needs to catch all occurrences of the disease to be stopped. The classification report shows that only for some classes is the recall really high; this means the model will be very effective in identifying almost all the instances of the said diseases. However, it is more appropriate for other classes to have a less recall, which implies that the model misses instances, either because of subtle disease symptoms or less distinctive visual features. An F1-score is a measure that is balanced between precision and recall. Particularly useful when the balance between precision and recall is needed, since it's a single metric taking into account false positives and false negatives. From the classification report, one can see that the F1 score is pretty high for some of the classes, which means performance is quite balanced. This might suggest possible trade-offs between precision and recall for classes that have lower F1 scores, which may then be resolved during fine-tuning of the model or data augmentation. Some of the trends and patterns can also be further indicated by a deeper look into the classification report. For example, classes with low precision may be facing some mix-up from other classes that share similar symptoms. This happens quite often when diseases are rich in visual characteristics so it will confuse them with one another. Again, by checking the confusion matrix and inspecting the components of the classification report we are able to identify which classes potentially cause confusion with what classes most frequently and what kinds of errors are made. Along with this, the classification report shows how the performance metrics of the model show disparity against classes due to different factors, which include quality and quantity of training data, discriminative disease symptoms, and efficiency in feature extraction. Classes with many clear examples of training will probably have high precision and recall, while those with fewer or unclear such instances may record lower performance metrics. The significance that comes with a diverse and representative set of training data cannot be overstated. Another quite useful work is the classification report in the task of iterative improvement of any model.

Strategies to increase the accuracy of a model targeting classes with underperforming metrics include augmentation of training data with more exemplars from classes represented in lesser amount, fine-tuning feature extraction to capture subtle disease symptoms better, or adjustment of the model architecture to handle complex cases better. A proper look at precision, recall, and F1-score of each class might give an insight into a potential bias. If the model is failing repeatedly on a few classes, that may actually show bias on the training data for those classes or show that data for those classes is too less to support proper classification, and the other class data needs to be properly balanced out in the training dataset, or else it will be prejudiced. Essentially, dealing with these biases is key to developing a robust and fair model capable of being good across all types of plant disease performances.

1. **Impact of Dataset Size and Quality:** The great performance of the model in the classification of the plant diseases was proven by size and quality of the data set. High-quality, well-annotated images provide the model with details on which to learn and differentiate features of different plant diseases accurately. The clarity and accuracy of the annotations are very important for each image if correctness in identifying and classifying diseases is to be undertaken. However, in the process of experimentation, it became clear that the incorporation of an expansive dataset with diversified samples would lead to not only well-designed models but also enhanced robustness and generalization capabilities for the varied environmental conditions and disease manifestations. A diverse dataset that captures the variability in symptoms of diseases at various stages of disease progression and variations in lighting conditions can be key to making a model generalize well in real-world applications [27] [28]. For instance, pictures on a variety of light intensities, various shots, and background conditions will make the model learn to differentiate diseases well even with such variations. High resolution is yet another critical point towards quality since it helps the model to capture the fine details in the picture that can prove to be of much importance to distinguish between similar diseases.

2. **Data Augmentation on dataset:** Since the dataset was not enough for effective training, data augmentation techniques were used to artificially increase the dataset size and variety. These include random rotations, flips, shifts, and adjustments of colors. When one inflates their data in training, it therefore means that the model is exposed to a larger number of variations, hence preventing overfitting and improving the generalization of new unseen data. In other words, data augmentation involves creating synthetic variations of the underlying images—a way to give the model a rounded learning

experience. Data augmentation was such an important game changer for the model to be robust. Random rotations ensure the model has learned the leaves of the disease from random angles, while flips and shifts vary the position. This color adjustment changes the model to be invariant to lighting conditions and color variations due to either different camera settings or other environmental nuisances. In this holistic approach, generic learning in the domain made the model, which increases the performance gradually in the test set over real-world applications. Model optimization and hyperparameter tuning Consequently, this is the hyperparameter optimization process integrated properly to obtain the best performance from the model. The hyperparameters, including learning rate, batch size, number of epochs, and architecture of the neural network layers, went through fine-tuning in an appropriate and systematic way [29]. Naturally, the values and various combinations seek optimal settings across combined, extensive fine-tuning that give rise to the best performance metrics in this study. Advanced optimization algorithms, with enhancements for Adam and RMSprop, were used to further improve the training process by giving faster convergence and a steadier dynamic during training. The final very important hyperparameter was the learning rate, which was tuned so that the model would learn effectively but not overshoot the optimal solutions. It became necessary to adjust the batch size in order to strike a balance between training time and the stability of the training process. Furthermore, the number of layers and their configurations were optimized to the development of a model that has high enough capacity to learn complex features in the data but without making it unwieldy for training. We did hyperparameter tuning with great rigor, and the performance of the model took a dramatic increase. A fine balance is achieved while setting these parameters to ensure high accuracy yet to maintain stability and efficiency during training.

3. **The use of transfer learning:** Transfer learning was a major key to the strategy in making the training process fast while attaining high accuracy. First, by using the weights obtained from pre-trained models, the number of epochs trained was considerably decreased. The transfer learning is to have a pre-trained model on a huge dataset—for instance, ImageNet—then fine-tuned in the framework. With this approach, the model can utilize all pre-learned features over a wide scope of recognition tasks, which enhances the effect of learning complex features: effective and efficient. Strong base in the model of classification of plant disease under investigation is the use of pre-trained models, for instance, VGG16, ResNet, and InceptionV3. These models were pretrained with very diverse data sets of images and hence had a very rich set of features that one could leverage well for plant disease recognition [30]. Fine-tuning from this data set on plant diseases is consolidated, at a very high pace, after just a few epochs. This becomes very beneficial since training deep neural networks really demands a lot from computation resources. This not only sped up the training process but also

helped in attaining higher accuracy levels. The model was able to leverage those intrinsic features and patterns that have already been learned from the large-scale datasets and are relevant to plant disease identification. This seemed very effective and efficient, especially in the scenario where limiting factors are computational resources and time.

Chapter 6

CONCLUSION AND FUTURE SCOPE

In conclusion, the experimental results show that the accuracy achieved by the proposed model for plant leaf disease classification with EfficientNetV2-based architecture is 99.24%, which makes the model suitable for tasks with practical applications where high accuracy is needed.

Such results will be useful for real applications in the field of plant leaf disease detection. Automated systems can further be implemented with deep learning models so that a quick and accurate disease identification graph of plants with symptoms on their leaves, or of the leaves themselves, could be done on an image taken of them. In this way, the potential crop yield and quality would increase, while inspection and manual disease diagnosis in crops would decrease. The development of mobile applications improved access and ease for farmers, especially those far from the market centers. Results of this study will provide very firm practical application-based applications that can be developed in this area of study.

Still, the model's performance can still be further enhanced through more research. A number of future works are as follows:

1. **Expansion of Datasets** What is needed for deep learning models is ever larger and more diverse datasets to ensure generalizing power. This includes taking photographs and making use of them under different light conditions, at different stages in the growth of an object, and from all over Planet Earth.

2. **Early detection of plant diseases** Consequently, with the aid of high-quality imagery, consideration of characteristics like humidity and temperature, to be able to come up with a way diseases can be detected when in their early stages. This will make it easier for the farmer to take precautionary measures with ample time before widespread diseases to save the crops, improve on health, and increase on overall productivity.

3. **Real-Time Disease Monitoring** The use of deep learning models for real-time monitoring of diseases could aid in providing real-time feedback on the

progression of a disease, thus contributing toward timely intervention. In addition, deep learning models can be used with devices employing the Internet of Things, drones, or robotic systems for continuous field monitoring and data collection in agriculture.

Bibliography

[1] F. Zhu, M.-Y. Cao, Q.-P. Zhang, R. Mohan, J. Schar, M. Mitchell, H. Chen, F. Liu, D. Wang, and Z. Q. Fu, "Join the green team: Inducers of plant immunity in the plant disease sustainable control toolbox," *Journal of Advanced Research*, vol. 57, pp. 15–42, 2024.

[2] Y. Kaya and E. Gürsoy, "A novel multi-head cnn design to identify plant diseases using the fusion of rgb images," *Ecological Informatics*, vol. 75, p. 101998, 2023.

[3] A. Ahmad, D. Saraswat, and A. El Gamal, "A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools," *Smart Agricultural Technology*, vol. 3, p. 100083, 2023.

[4] E. Moupojou, A. Tagne, F. Retraint, A. Tadonkemwa, D. Wilfried, H. Tapamo, and M. Nkenlifack, "Fieldplant: A dataset of field plant images for plant disease detection and classification with deep learning," *IEEE Access*, vol. 11, pp. 35 398–35 410, 2023.

[5] G. Geetha, S. Samundeswari, G. Saranya, K. Meenakshi, and M. Nithya, "Plant leaf disease classification and detection system using machine learning," in *Journal of Physics: Conference Series*, vol. 1712, no. 1. IOP Publishing, 2020, p. 012012.

[6] S. S. Chouhan, U. P. Singh, and S. Jain, "Automated plant leaf disease detection and classification using fuzzy based function network," *Wireless Personal Communications*, vol. 121, no. 3, pp. 1757–1779, 2021.

[7] R. Sujatha, J. M. Chatterjee, N. Jhanjhi, and S. N. Brohi, "Performance of deep learning vs machine learning in plant leaf disease detection," *Microprocessors and Microsystems*, vol. 80, p. 103615, 2021.

[8] A. S. Tulshan and N. Raul, "Plant leaf disease detection using machine learning," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2019, pp. 1–6.

[9] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and electronics in agriculture*, vol. 145, pp. 311–318, 2018.

[10] K. Harshavardhan, P. A. Krishna, and A. Geetha, "Detection of various plant leaf diseases using deep learning techniques," in *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*. IEEE, 2023, pp. 1–6.

[11] H. Amin, A. Darwish, A. E. Hassanien, and M. Soliman, "End-to-end deep learning model for corn leaf disease classification," *IEEE Access*, vol. 10, pp. 31 103–31 115, 2022.

[12] C. Sunil, C. Jaidhar, and N. Patil, "Cardamom plant disease detection approach using efficientnetv2," *Ieee Access*, vol. 10, pp. 789–804, 2021.

[13] Y. Zhang, C. Song, and D. Zhang, "Deep learning-based object detection improvement for tomato disease," *IEEE access*, vol. 8, pp. 56 607–56 614, 2020.

[14] H. Nazki, S. Yoon, A. Fuentes, and D. S. Park, "Unsupervised image translation using adversarial networks for improved plant disease recognition," *Computers and Electronics in Agriculture*, vol. 168, p. 105117, 2020.

[15] Z. S. Zhang ShanWen, Z. S. Zhang SuBing, Z. C. Zhang ChuanLei, W. X. Wang XianFeng, and S. Y. Shi Yun, "Cucumber leaf disease identification with global pooling dilated convolutional neural network." 2019.

[16] U. P. Singh, S. S. Chouhan, S. Jain, and S. Jain, "Multilayer convolution neural network for the classification of mango leaves infected by anthracnose disease," *IEEE access*, vol. 7, pp. 43 721–43 729, 2019.

[17] V. Singh, "Sunflower leaf diseases detection using image segmentation based on particle swarm optimization," *Artificial Intelligence in Agriculture*, vol. 3, pp. 62–68, 2019.

[18] S. Mishra, R. Sachan, and D. Rajpal, "Deep convolutional neural network based detection system for real-time corn plant disease recognition," *Procedia Computer Science*, vol. 167, pp. 2003–2010, 2020.

[19] P. Sharma, Y. P. S. Berwal, and W. Ghai, "Performance analysis of deep learning cnn models for disease detection in plants using image segmentation," *Information Processing in Agriculture*, vol. 7, no. 4, pp. 566–574, 2020.

[20] M. Agarwal, A. Singh, S. Arjaria, A. Sinha, and S. Gupta, "Toled: Tomato leaf disease detection using convolution neural network," *Procedia Computer Science*, vol. 167, pp. 293–301, 2020.

[21] A. Khamparia, G. Saini, D. Gupta, A. Khanna, S. Tiwari, and V. H. C. de Albuquerque, "Seasonal crops disease prediction and classification using deep convolutional encoder network," *Circuits, Systems, and Signal Processing*, vol. 39, pp. 818–836, 2020.

[22] K. R. Gavhale, U. Gawande *et al.*, "An overview of the research on plant leaves disease detection using image processing techniques," *Iosr journal of computer engineering (iosr-jce)*, vol. 16, no. 1, pp. 10–16, 2014.

[23] C. G. Dhaware and K. Wanjale, "A modern approach for plant leaf disease classification which depends on leaf image processing," in *2017 International Conference on Computer Communication and Informatics (ICCCI)*. IEEE, 2017, pp. 1–4.

[24] S. R. Kamlapurkar, "Detection of plant leaf disease using image processing approach," *International Journal of Scientific and Research Publications*, vol. 6, no. 2, pp. 73–76, 2016.

[25] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International conference on machine learning*. PMLR, 2021, pp. 10 096–10 106.

[26] G. Geetharamani and A. Pandian, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," *Computers & Electrical Engineering*, vol. 76, pp. 323–338, 2019.

[27] L. S. P. Annabel, T. Annapoorani, and P. Deepalakshmi, "Machine learning for plant leaf disease detection and classification–a review," in *2019 international conference on communication and signal processing (ICCSP)*. IEEE, 2019, pp. 0538–0542.

[28] A. Bhargava, A. Shukla, O. Goswami, M. H. Alsharif, P. Uthansakul, and M. Uthansakul, "Plant leaf disease detection, classification and diagnosis using computer vision and artificial intelligence: A review," *IEEE Access*, 2024.

[29] M. V. Applalanaidu and G. Kumaravelan, "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)*. IEEE, 2021, pp. 716–724.

[30] P. Sharma, P. Hans, and S. C. Gupta, "Classification of plant leaf diseases using machine learning and image preprocessing techniques," in *2020 10th international conference on cloud computing, data science & engineering (Confluence)*. IEEE, 2020, pp. 480–484.

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis PLANT LEAF DISEASE CLASSIFICATION USING DEEP LEARNING

Total Pages 55 Name of the Scholar AMIT RAMANI

Supervisor (s)

(1) SANJAY PATIDAR

(2)

(3)

Department SOFTWARE ENGINEERING

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: TURNITIN Similarity Index: 11%, Total Word Count: 19365

Date: 24/05/2024

Candidate's Signature

Signature of Supervisor(s)

11% Overall Similarity

Top sources found in the following databases:

- 6% Internet database
- 6% Publications database
- Crossref database
- Crossref Posted Content database
- 9% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	**researchgate.net** Internet	<1%
2	**Mayur Desai, Namrata Ansari. "Enhancing Plant Leaf Disease Classific...** Crossref	<1%
3	**Thomas More Hogeschool on 2023-05-31** Submitted works	<1%
4	**Liverpool John Moores University on 2024-03-11** Submitted works	<1%
5	**link.springer.com** Internet	<1%
6	**Sakshi Raina, Abhishek Gupta. "A Study on Various Techniques for Pla...** Crossref	<1%
7	**Buckinghamshire Chilterns University College on 2024-05-21** Submitted works	<1%
8	**Liverpool John Moores University on 2023-05-31** Submitted works	<1%

9	scholar.colorado.edu Internet	<1%
10	Technological Institute of the Philippines on 2024-01-27 Submitted works	<1%
11	"Data Science and Applications", Springer Science and Business Media... Crossref	<1%
12	Aston University on 2024-01-23 Submitted works	<1%
13	slideshare.net Internet	<1%
14	hindawi.com Internet	<1%
15	mdpi.com Internet	<1%
16	techscience.com Internet	<1%
17	University of Essex on 2024-03-22 Submitted works	<1%
18	jetir.org Internet	<1%
19	University of East London on 2024-05-09 Submitted works	<1%
20	University of Liverpool on 2024-03-21 Submitted works	<1%

21	"Statistical Atlases and Computational Models of the Heart. Regular an... Crossref	<1%
22	Meroua Belmir, Wafa Difallah, Abdelkader Ghazli. "Plant Leaf Disease ... Crossref	<1%
23	Shanwen Zhang, Subing Zhang, Chuanlei Zhang, Xianfeng Wang, Yun S... Crossref	<1%
24	University of Cambridge on 2018-05-17 Submitted works	<1%
25	annalsofrscb.ro Internet	<1%
26	University of Wales, Bangor on 2024-05-30 Submitted works	<1%
27	"Internet of Things and Machine Learning in Agriculture", Walter de Gru... Crossref	<1%
28	Arba Minch University on 2024-03-14 Submitted works	<1%
29	Bournemouth University on 2023-09-01 Submitted works	<1%
30	Intercollege on 2023-02-26 Submitted works	<1%
31	University of Westminster on 2023-05-02 Submitted works	<1%
32	ijraset.com Internet	<1%

33	researchsquare.com	Internet	<1%
34	Heriot-Watt University on 2023-04-19	Submitted works	<1%
35	Liverpool John Moores University on 2023-05-24	Submitted works	<1%
36	Queen Mary and Westfield College on 2019-07-31	Submitted works	<1%
37	dspace.daffodilvarsity.edu.bd:8080	Internet	<1%
38	scribd.com	Internet	<1%
39	5dok.org	Internet	<1%
40	Laureate Higher Education Group on 2019-06-09	Submitted works	<1%
41	University of Hull on 2023-05-03	Submitted works	<1%
42	cnaa.md	Internet	<1%
43	Middlesex University on 2023-10-08	Submitted works	<1%
44	Pranavi Reddy, Andrew J. "Diagnosis of Autism in Children Using Deep ...	Crossref	<1%

45	Sheffield Hallam University on 2024-01-17 Submitted works	<1%
46	University of North Texas on 2023-07-21 Submitted works	<1%
47	University of Westminster on 2023-10-18 Submitted works	<1%
48	Vishal Menon, V Ashwin, Raj K Deepa. "Plant Disease Detection using ... Crossref	<1%
49	static.frontiersin.org Internet	<1%
50	arxiv-vanity.com Internet	<1%
51	Junde Chen, Jinxiu Chen, Defu Zhang, Yuandong Sun, Y.A. Nanehkaran... Crossref	<1%
52	Southern University And A & M College on 2022-11-04 Submitted works	<1%
53	bookdown.org Internet	<1%
54	webapps.unsworks.library.unsw.edu.au Internet	<1%
55	Athlone Institute of Technology on 2024-05-11 Submitted works	<1%
56	Hassan Amin, Ashraf Darwish, Aboul Ella Hassanien, Mona Soliman. "E... Crossref	<1%

57	Igdır Universitesi on 2024-05-22 Submitted works	<1%
58	Liverpool John Moores University on 2023-06-06 Submitted works	<1%
59	Liverpool John Moores University on 2023-10-18 Submitted works	<1%
60	London Metropolitan University on 2023-12-07 Submitted works	<1%
61	University of Hertfordshire on 2022-08-15 Submitted works	<1%
62	University of Lincoln on 2022-09-22 Submitted works	<1%
63	ictactjournals.in Internet	<1%
64	mafiadoc.com Internet	<1%
65	newsghana.com.gh Internet	<1%
66	Aashish Jha, Madhavi Purohit, Vivek Maurya, Amiya Kumar Tripathy. "... Crossref	<1%
67	Arab Academy for Science, Technology & Maritime Transport CAIRO o... Submitted works	<1%
68	Asia Pacific University College of Technology and Innovation (UCTI) on... Submitted works	<1%

69 University of Glamorgan on 2023-09-14 <1%
Submitted works

70 Yang Zhang, Chenglong Song, Dongwen Zhang.. "Deep Learning-based... <1%
Crossref

71 tb.chinasmp.com <1%
Internet

72 frontiersin.org <1%
Internet

73 "Recent Innovations in Computing", Springer Science and Business Me... <1%
Crossref

74 Arunangshu Pal, Vinay Kumar, Khondekar Lutful Hassan, Binod Kumar ... <1%
Crossref

75 Babes-Bolyai University on 2023-06-22 <1%
Submitted works

76 Dr. S. P. Mukherjee International Institute of Information Technology (I... <1%
Submitted works

77 Federal University of Technology-Nigeria on 2024-01-22 <1%
Submitted works

78 Indira Gandhi Delhi Technical University for Women on 2023-09-13 <1%
Submitted works

79 Kabenla, Philippe. "Automation of Loan Eligibility Process with Deep M... <1%
Publication

80 Karabük Üniversitesi on 2023-01-11 <1%
Submitted works

81	Liverpool John Moores University on 2022-06-08 Submitted works	<1%
82	National College of Ireland on 2019-12-12 Submitted works	<1%
83	South Dakota Board of Regents on 2018-12-05 Submitted works	<1%
84	Technical University of Košice on 2024-05-28 Submitted works	<1%
85	Tilburg University on 2024-05-20 Submitted works	<1%
86	University of East London on 2024-05-08 Submitted works	<1%
87	University of Johannsburg on 2023-10-31 Submitted works	<1%
88	University of Northumbria at Newcastle on 2023-05-07 Submitted works	<1%
89	University of Strathclyde on 2024-03-25 Submitted works	<1%
90	University of Sunderland on 2022-12-12 Submitted works	<1%
91	University of Warwick on 2017-09-14 Submitted works	<1%
92	Wright State University on 2023-12-15 Submitted works	<1%

93	Yang Wu, Xian Feng, Guojun Chen. "Plant Leaf Diseases Fine-grained C... Crossref	<1%
94	Yeditepe University on 2023-06-07 Submitted works	<1%
95	dokumen.pub Internet	<1%
96	download.hrz.tu-darmstadt.de Internet	<1%
97	dspace.vsb.cz Internet	<1%
98	ia.spcras.ru Internet	<1%
99	iris.uniroma1.it Internet	<1%
100	subscription.packtpub.com Internet	<1%
101	thefreelibrary.com Internet	<1%
102	"Applications of Computer Vision and Drone Technology in Agriculture ... Crossref	<1%
103	"Data Engineering and Intelligent Computing", Springer Science and Bu... Crossref	<1%
104	"Proceedings of International Conference on Communication and Com... Crossref	<1%

105	"Proceedings of International Conference on Data Science and Applica... Crossref	<1%
106	Anna University on 2024-04-25 Submitted works	<1%
107	Army Institute of Technology on 2024-04-05 Submitted works	<1%
108	Aulia Ikvanda Yoren, Suyanto Suyanto. "Tomato Plant Disease Identific... Crossref	<1%
109	Babes-Bolyai University on 2023-06-20 Submitted works	<1%
110	Bathula Nagachandrika, R. Prasath, I.R. Praveen Joe. "An automatic cla... Crossref	<1%
111	City University on 2023-05-19 Submitted works	<1%
112	D. Srinivasa Rao, Ramesh Babu Ch, V. Sravan Kiran, N. Rajasekhar et al... Crossref	<1%
113	Dini, Shadi. "Speech Act Classification in Computational Linguistics Usi... Publication	<1%
114	Folasade Olubusola Isinkaye, Michael Olusoji Olusanya, Pramod Kuma... Crossref	<1%
115	Galileo Global Education on 2023-05-01 Submitted works	<1%
116	Heriot-Watt University on 2021-04-21 Submitted works	<1%

117	Hoang-Tu Vo, Nhon Nguyen Thien, Kheo Chau Mui. "Tomato Disease R... Crossref	<1%
118	Igdır Universitesi on 2024-05-24 Submitted works	<1%
119	Indian Institute of Technology Guwahati on 2021-06-22 Submitted works	<1%
120	Jahan, Nusrat. "Wheat Diseases Analysis Using Machine Learning and ... Publication	<1%
121	Liverpool John Moores University on 2021-01-22 Submitted works	<1%
122	Liverpool John Moores University on 2022-06-05 Submitted works	<1%
123	Liverpool John Moores University on 2023-03-22 Submitted works	<1%
124	Liverpool John Moores University on 2023-06-05 Submitted works	<1%
125	Manchester Metropolitan University on 2023-05-26 Submitted works	<1%
126	Pankaj Kumar, Sunidhi Ashtekar, S. S Jayakrishna, K P Bharath, P. T Va... Crossref	<1%
127	Reem Ibrahim Hasan, Suhaila Mohd Yusuf, Laith Alzubaidi. "Review of t... Crossref	<1%
128	Universiti Teknologi MARA on 2018-06-08 Submitted works	<1%

129	Universiti Teknologi Malaysia on 2022-09-12 Submitted works	<1%
130	University of Bristol on 2023-12-28 Submitted works	<1%
131	University of Glamorgan on 2023-10-24 Submitted works	<1%
132	University of Greenwich on 2023-12-22 Submitted works	<1%
133	University of Hertfordshire on 2017-04-25 Submitted works	<1%
134	University of Leicester on 2023-09-11 Submitted works	<1%
135	University of Lincoln on 2023-06-29 Submitted works	<1%
136	University of Moratuwa on 2021-12-07 Submitted works	<1%
137	University of Northumbria at Newcastle on 2024-05-23 Submitted works	<1%
138	University of Nottingham on 2023-09-07 Submitted works	<1%
139	University of Sunderland on 2023-07-17 Submitted works	<1%
140	University of Surrey on 2023-09-12 Submitted works	<1%

141	University of Teesside on 2023-05-09 Submitted works	<1%
142	William Macdonald, Yuksel Asli Sari, Majid Pahlevani. "Grow-light smar... Crossref	<1%
143	dataaspirant.com Internet	<1%
144	dergipark.org.tr Internet	<1%
145	files.osf.io Internet	<1%
146	ijece.iaescore.com Internet	<1%
147	publications.eai.eu Internet	<1%
148	sjuoz.uoz.edu.krd Internet	<1%
149	diva-portal.org Internet	<1%
150	ijisae.org Internet	<1%
151	ijrte.org Internet	<1%

ANNEXURE-VII

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India

CERTIFICATE OF FINAL THESIS SUBMISSION

(To be submitted in duplicate)

1. Name: AMIT RAMANI
2. Roll No: 2K22/SWE/14
3. Thesis title: PLANT LEAF DISEASE CLASSIFICATION USING DEEP LEARNING
4. Degree for which the thesis is submitted: MTECH
5. Faculty (of the University to which the thesis is submitted) SANJAY PATIDAR
6. Thesis Preparation Guide was referred to for preparing the thesis. YES ☑ NO ☐
7. Specifications regarding thesis format have been closely followed. YES ☑ NO ☐
8. The contents of the thesis have been organized based on the guidelines. YES ☑ NO ☐
9. The thesis has been prepared without resorting to plagiarism. YES ☑ NO ☐
10. All sources used have been cited appropriately. YES ☑ NO ☐
11. The thesis has not been submitted elsewhere for a degree. YES ☑ NO ☐
12. All the correction has been incorporated. YES ☑ NO ☐
13. Submitted 2 hard bound copies plus one CD. YES ☑ NO ☐

(Signature(s) of the Supervisor(s))
Name(s): Mr. Sanjay Patidar

(Signature of Candidate)
Name: AMIT RAMANI
Roll No: 2K22/SWE/14