St. Mary’s University

Department of Computer Science

Image-Based Rose Leaf Diseases Detection Using Deep Learning

By: Tigist Ashine

Advisor: Dr. Million M.

A thesis submitted to St. Mary’s University, Department of Computer Science, in partial fulfillment of the requirements for the degree of Master of Science in Computer Science.

January 19, 2024

Addis Ababa, Ethiopia

St. Mary's University

Department of Computer Science

Image-Based Rose Leaf Diseases Detection Using Deep Learning

A thesis submitted to St. Mary's University, Department of Computer Science, in partial fulfillment of the requirements for the degree of Master of Science in Computer Science.

By: Tigist Ashine

Approval by the Board of Examiners

1.	Advisor	Dr. Million Meshesha	Million	
		Name	Signature	Date
2.	Internal Examiner			
		Name	Signature	Date
3.	External Examiner	Minale Ashagrie (Ph.D)		March 2, 2024

Declaration

I, the undersigned, declare that the thesis comprises my work. In compliance with internationally accepted practices, I have acknowledged and refereed all materials used in this work. I understand that non-adherence to academic honesty, integrity, and misrepresentation/fabrication of any idea/data/fact/source will constitute sufficient grounds for disciplinary action by the University and can also evoke penal action from the sources that have not been properly cited or acknowledged.

Tigist Ashine ______________________

Name Signature

Place: Addis Ababa, Ethiopia

Date of submission: January 19, 2024

This thesis has been submitted for examination with my approval as a university advisor.

million

Dr. Million Meshesha ______________________

Advisor Signature

Abstract

Using image processing techniques, several forms of study can be conducted in the domain of Image of rose leaf disease classification. However, Image of rose leaf disease detection is still a problem for people who do not know about rose leaf diseases. Image-based rose leaf disease detection using deep learning involves training a deep learning model to analyze images of rose leaf and identify signs of diseases such as bacteria, viruses, and fungi. This process can help in the early detection and management of plant diseases, ultimately contributing to improved agricultural productivity and the economy. Deep learning algorithms are trained using a large dataset of images showing healthy and diseased rose leaf. The model learns to recognize patterns and features associated with different diseases, enabling it to accurately classify new images. To classify whether the image is Fresh, Black spot, or Downy mildew we used three classifiers, such as Support Vector Machine (SVM), K-Nearest Neighbor Classifier (KNN), and convolutional neural network (CNN). The datasets are gathered from the Ethio Agri CEEFT PLC Holeta Flower Farm, which is in the Oromia region, Ethiopia. The data is preprocessed through data collection, cleaning, augmentation, image preprocessing, dataset splitting, and data normalization. Feature extraction is performed using the automatic feature extraction capabilities of convolutional layers in CNNs. The total data set used for the experiment is 4342 Rose Leaf Images. The data is split into train and test data sets such that, 20% of the data set is used for testing the model's performance, and 80% for training machine learning as well as deep learning and creating disease detection models from rose leaf images.

Experimental results shows that the model created by SVM, KNN, and CNN registers an accuracy of 80.32%,71.23%, and 98% respectively. The model created by CNN therefore outperforms the other classification algorithms. To effectively train the deep learning model, this approach requires a vast and diverse dataset, which is one of its main weaknesses and limitations. Furthermore, it might be difficult to record all possible combinations of environmental variables and disease symptoms due to the reliance on image-based data and further research needs to be done to combine image based with text based so as to come up with a generic model for rose leaf disease detection.

Keywords: *Rose leaf disease detection; Deep learning, KNN, SVM, CNN.*

Acknowledgment

First and foremost, I would like to sincerely thank the Almighty God and His Blessed Virgin Mary Mother for giving me strength throughout my academic life and bringing me from the beginning to this MSc level. His kindness made me successful in all my academic fields.

I would like to express my heartfelt gratitude to my thesis advisor Dr. Million Meshesha for his patience with continuous guidance and support throughout this thesis work. All the comments and suggestions he gave me throughout my thesis work were very constructive and helped me a lot. Thanks once again. I would also extend my gratitude to Awash Wine S.C. for helping me with this study by providing financial support to go ahead and carry out my study and accomplish my goal.

Finally, I would like to extend my deepest gratitude to my dear families for their continuous support and the roles they played on my behalf. My sincere gratitude especially to Mr. Yonas Techale for His countless support in achieving my thesis work.

My sincere appreciation also goes to my best friends who helped me a lot in many ways for the successful completion of my study. May God bless you all!

Table of Contents

List of Figures

List of Tables

List of Acronyms

ANN	Artificial Neural Networks
API	Application Programming Interface
CNN	Convolutional Neural Networks
CNNs	Convolutional Neural Network
DIP	Digital Image Processing
DL	Deep Learning
FC	Fully Connected
FN	Fales Negative
FP	Fales Positive
GANs	Generative Adversarial Networks
JPEG	Joint Photographic Experts Group
KNN	K-Nearest Neighbors
ML	Machine learning
MP	Multilayer Perceptron
MSC	Multiple scattering correction
RGB	Red, Green & Blue
RNN	Recurrent Neural Networks
SNV	Standard Normal Variable
SVM	Support Vector Machine
TN	True Negative
TP	True positive
VGG16	Visual Geometry Group

Chapter One

Introduction

1.1. Background

Image processing was first proposed in the 1960s as a method for processing, interpreting, and comprehending image data [1].Image processing is being applied in the area of biotechnology, medical diagnosis, biometrics, pattern recognition, remote sensing, film and video production, and security monitoring.

It is the process of enhancing the quality of an image, reducing noise, and extracting feature information from an image. As a result, image processing can be used to improve image quality for humans as well as extract visual features for future study and understanding [2].

Agriculture plays a significant role in the global economy as it provides food and raw materials for various industries. It also contributes to employment, trade, and economic growth in many countries. However, the agriculture industry faces challenges such as climate change, market fluctuations, and resource depletion [3]. It is important to promote sustainable agricultural practices to ensure food security and economic stability. Agriculture production does occupy a significant amount of land, about 40% of the world's land area [4]. The nature of the land used for agriculture varies greatly depending on the location, which affects the choice of inputs, outputs, and technology used. Due to differences in climate and natural resources, farms and farming practices can vary greatly in terms of size, products produced, technology used, inputs employed, farm incomes, and other economic outcomes. This diversity makes agriculture a complex and dynamic industry that requires careful management to ensure sustainability and economic viability [3]. The agriculture sector is a significant contributor to economic growth. Plant diseases and pests can have a significant impact on agriculture, reducing the quality of food production and negatively affecting the economy. Prophylactic treatments are not always effective in preventing outbreaks, so early monitoring and proper diagnosis of crop diseases are crucial for preventing losses in production quality. A proper crop protection system can help to identify and manage plant diseases and pests, which can help to minimize their impact on agricultural production and support economic growth [5]. Ethiopia's economy is mostly reliant on agriculture, which provides 40% of the nation's GDP, 80% of its exports, and over 75% of its labor force. [6]. But, The most challenging issue the nation is currently dealing with is the spread of various illnesses on various crops, which exacerbates socioeconomic issues including food insecurity, market inflation, and a shortage of hard currency due to the need for imports from other nations to make ends meet.[7]. The review of the Ethiopian Ten-Year Development Plan (2021-30) indicated that the country's economy (GDP) has grown rapidly at an average of 9.2% per year between 2010 and 2020 fiscal year [8]. In absolute terms the growth is equivalent to Birr 828 billion to Birr 1.99 trillion. For the period, the average contributions of agriculture, industry and services sectors respectively were 24%, 37.9% and 40.8%. Though declining overtime, the poverty headcount ratio at the national poverty lines (% of population) remains as high as 23.5% in 2015. The rural poverty remains as worse as 25.6% in 2020. This indicated the continued challenge of reducing poverty in the country. Some of the major challenges contributing to high level of poverty include: rising rate of unemployment, persistence of high inflation rate, low agricultural productivity trap, and persistently increased trade deficit [8]. The study conducted by Belay [9] noted that, Ethiopia's floriculture sector has showed significant boom between 2002 and 2008, where the number of flower farms grew about 16-fold and the value of flower exports about 20-fold. The sector made the country the second largest exporter in Africa and the fifth largest supplier of flowers to the global market. The sector also provided employment opportunities for hundreds of thousands of individuals. Further, rose flowers contributed for about 80% of the floriculture sector.

According to Ashine [10] Ethiopia has enormous potential for the production and selling of horticulture, primarily because of its location for export to different markets, plentiful water sources, rich terrain, labor force, and supportive government policies. The sector did not contribute as much as it should have, despite its potential. This is a result of the restricted funding for extension services and research and development. Vegetables and fruits are a significant additional dietary and nutrient source. Low productivity and a high rate of post-harvest losses from improper handling continue to be significant issues in meeting the growing demands of both domestic and international markets. Pests and diseases are two of the main causes of the loss. Improved distribution network and increased technological advancement are required to meet the rising demand for fruits and vegetables in both domestic and international markets. Roses are important to the economy, but they are also quite susceptible to a wide range of diseases caused by bacteria, viruses, and fungi.

Ardasheva, Cheremnykh [11] articulated this vulnerability by saying "if you grow rose, you're almost guaranteed to encounter disease." Accordingly, he identified its nine common diseases, which are: Black Spot, Botrytis Blight, Cankers, Crown Gall, Downy Mildew, Rust, Powdery Mildew, Rose Mosaic Virus, and Rose Rosett.

Generally, Rose is one of the most popular ornamental plants around the world. However, diseases can cause significant damage to rose plants, leading to reduced yield and quality. Early diagnosis and treatment of these diseases are essential to prevent their spread and minimize the damage.

Image-based disease detection in plants using deep learning models is a method of using machine learning algorithms to analyze images of plants and identify whether they are healthy or diseased. By training a deep learning model on a large datasets of images of both healthy and diseased plants, the model can learn to distinguish between the two types of images [12] This technology has the potential to greatly improve the efficiency and accuracy of plant disease diagnosis and treatment. In this thesis, an attempt is made to use deep learning models for image-based disease detection in plants.

1.1.1. Digital Image Processing (DIP)

The study and processing of digital images is the focus of the computer science discipline known as digital image processing, or DIP. DIP techniques are employed to retrieve details from images, including object edges, surface textures, and object colors. DIP is employed in many different fields, such as video surveillance, machine vision, and medical imaging.[12]. Digital Image Processing (DIP) involves three main stages: preparing, analyzing, and comprehending images [12].

- **Image preprocessing** includes operations such as image enhancement, noise reduction, and image restoration. These operations are performed to improve the quality of the image and make it more suitable for analysis.
- **Image analysis** is the process of taking information out of the picture, including feature extraction, segmentation, and object detection. These procedures are employed to recognize and take meaningful data from the image.
- **Image understanding** involves interpreting the extracted information to make decisions or take actions. This can include tasks such as object recognition, scene understanding, and image classification.

1.1.2. Deep Learning (DL)

Artificial neural networks are used in deep learning (DL), a branch of machine learning, to extract knowledge from data. Large datasets of labeled data are used to train DL models, which enable them to recognize patterns and relationships in the data that would be challenging or impossible to find using more conventional statistical techniques. Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and Convolutional Neural Networks (CNN) are the three primary forms of neural networks that fall under the category of Deep Learning (DL) [12].

- **Artificial Neural Networks** (ANN): ANN is a type of DL model that is inspired by the structure and functioning of the human brain. It consists of interconnected layers of artificial neurons called perceptron's. ANN is commonly used for tasks such as image classification, regression, and pattern recognition.
- **Recurrent Neural Networks** (RNN): The RNN's return links are used to process data in a sequential fashion. Because of its memory component, which enables it to hold onto data from earlier inputs, it is appropriate for applications like time series analysis, speech recognition, and natural language processing.
- **Convolutional Neural Networks** (CNN): CNN was created especially to interpret data that looks like a grid, like photographs. It automatically learns hierarchical representations of visual characteristics using convolutional layers. CNNs are widely used in tasks like image classification, object detection, and image segmentation.

DIP and DL can both be used for plant disease detection. DIP techniques can be used to extract features from images of plant leaves that are indicative of disease. These features can then be used to train a model that can classify images of plant leaves as healthy or diseased. DL models can be trained on larger datasets of images than DIP models, and they can learn more complex patterns and relationships in the data. As a result, DL models can achieve higher accuracy than DIP models in classifying images of plant leaves as healthy or diseased.

1.2. Statement of the Problem

Roses are particularly sensitive to several illnesses that can have a substantial impact on their output and quality, despite the economic significance of rose blooms. For agricultural (and rose flower) farms to effectively control illness, early and precise diagnosis and treatment of the diagnosed diseases are essential. Current methods for detecting diseases in rose leaves are Visual diagnosis or Manual inspections, which is time-consuming and subjective, as well as requiring specialized knowledge and skills which is unaffordable for large-scale farming operations.

To tackle this methodological gap in disease diagnosis advanced technologies and practices must be explored and used to improve rose yield and quality. Fortunately, AI-powered tools that use machine learning algorithms can now be used to efficiently and accurately identify plant diseases based on images of plant parts. Deep learning algorithms allow automatic detection of diseases in rose leaves.

Researchers have conducted and are still conducting studies to investigate reasonable ways to improve the studies automatic means for rose disease diagnosis: These studies are conducted by local and foreign scholars.LC Ngugi, M Abdelwahab, M Abo-Zahhad [13] conducted a study on the detection of black spot, a serious plant disease in China that affects rose production. The researchers utilized hyperspectral technology to analyze both external features and internal structure information of infected roses. Two diseased roses with black spots were used to build a convolutional neural network (CNN) model utilizing their spectral and visual properties. The spectrum data underwent preprocessing using multiple scattering correction (MSC) and Standard Normal Variable (SNV) methods, whilst the hyperspectral images underwent cropping, median filtering, and binarization. The evaluation criteria for three CNN models were classification accuracy and loss function (Alexnet, VGG16, and NDDR). The outcomes demonstrated that the CNN model with feature fusion had more accuracy. The highest accuracies of detection of black spot in different roses were achieved using the NDDR-CNN model. The study's findings suggest that CNN-based spectral analysis can effectively detect black spots in roses and serve as a reference for detecting other plant diseases. The research holds significant potential for further development in the field of plant disease detection. The study conducted by S. Alqethami, B. Almtanni, W. Alzhrani, and M. Alghamdi [14] utilized three prediction models (CNN, SVM, and KNN) and various image processing methods to accurately identify and categorize healthy and diseased apple plant leaves. The models were evaluated using the Kaggle New Plant Diseases database. The primary objective of the study was to assist farmers in detecting and preventing disease spread, providing appropriate solutions based on the classification results. This research emphasizes the potential of image processing and machine learning techniques in addressing crop diseases, contributing to improved agricultural practices and food security. According to J. A. Ruth, R. Uma, A. Meenakshi, and P. Ramkumar, [15] the paper proposes a framework using an optimal deep neural network (ODNN) to automatically detect plant leaf diseases. The framework utilizes Convolutional Neural Network for feature extraction and ODNN for disease detection. Weight optimization is performed using the Improved Butterfly Optimization Algorithm, which incorporates the Genetic Algorithm to enhance convergence rate. Sensitivity, Accuracy, and

Specificity metrics are used for evaluating the proposed method, which achieved an overall accuracy of 99 percent. The experimental results suggest that this approach outperforms existing techniques. The focus of this research is therefore to develop an automated system that can accurately and efficiently detect and diagnose rose leaf diseases based on images using deep learning.

Based on the above research problem, we have formulated the next research questions:

- Which image processing methods are appropriate for creating images of roses?
- What are the suitable deep learning techniques for constructing an optimal model for detecting and diagnosing rose leaf diseases?
- To what extent does the classification model perform in detecting rose leaf disease?

1.3. Objective of the study

1.3.1. General Objective

The general objective of the research is to develop an image-based Rose leaf diseases identification system by using deep learning.

1.3.2. Specific Objectives

A set of specific objectives that all together lead to the realization of the overall objectives are identified as follows:

- ✓ To review related literature to identify appropriate methods and techniques.
- ✓ To gather and prepare healthy and unhealthy Rose leaf images from different sources.
- ✓ To select suitable digital image processing and deep learning algorithms
- ✓ To develop a model for a rose leaf diseases detection system by using a best accurate algorism
- ✓ To improve the object detection and classification methods for Rose leaf images.
- ✓ To evaluate and test the suggested model’s effectiveness.

1.4. Scope of the Study

Review of literature in the field showed that types of rose diseases are wide ranging, so the focus of this research is limited to the most common diseases. Symptoms of rose flower diseases are also manifested in different parts of the plant as primary and secondary symptoms. The potential of image-based rose leaf diseases identification using deep learning-based algorithms is enormous. This research limits itself to the most common rose leaf diseases, such as black spot, powdery mildew, and rust. These diseases have primary symptoms such as discoloration, spots, and deformities on the leaves, while secondary symptoms may include stunted growth and reduced flower production. By using deep learning-based algorithms for image-based identification, this research has the potential to provide a cost-effective and efficient solution for early detection and management of rose leaf diseases. The wide range of types of rose diseases makes it important to narrow the focus of this research to the most common ones.

The images for this research are captured of rose leaves affected by common diseases such as black spot, powdery mildew, and rust. These diseases have primary symptoms such as discoloration, spots, and deformities on the leaves. The images are used to train deep learning-based algorithms for cost-effective and efficient identification of rose leaf diseases to aid in early detection and management.

Further, deep learning algorithms have shown to be quite successful at identifying plant diseases through image analysis. Deep learning algorithms can develop an understanding of the patterns and features that separate healthy leaves from diseased ones by using vast data of images. The focus of this method is therefore to aid early illness diagnosis, which can stop the spread of the disease and reduce crop losses.

1.5. Significance of the Study

Deep learning makes it possible to identify plant diseases accurately and quickly, making it a very advantageous method for identifying diseases in rose leaves. Deep learning algorithms are useful for detecting plant diseases because they can evaluate massive amounts of data and spot patterns that may not be visible to the human eye. Early disease diagnosis enables farmers to take prompt action to stop the disease's progress and reduce crop losses, thereby enhancing crop yields and produce quality without requiring of Experts. Generally automated detection of diseases in rose leaves enables faster, more accurate results to increase the quality and quantity of the rose flower. As well, a cost-effective and effective tool for farmers, image-based disease detection can also be performed remotely. Deep learning-based image-based methods for detecting rose leaf diseases have a wide range of applications that could fundamentally alter how we identify and treat plant diseases.

Early and accurate diagnosis and treatment of the identified diseases are crucial for effective disease management of agricultural (and rose flower) farms. The common method of identifying plant diseases is Visual Examination, which is both subjective and time-consuming.

In recent years, deep learning algorithms and techniques have shown great potential in detecting and diagnosing various diseases. In the field of agriculture, it has been applied to detect and diagnose diseases in crops.

In this study, a deep learning model is built to use images to identify and diagnose diseases in rose leaves. Convolutional Neural Network (CNN) architecture, the foundation of the suggested system, was trained using a set of annotated photos of both healthy and damaged rose leaves. The input image is first processed to improve its characteristics and remove noise, after which it is fed into CNN to extract pertinent information and categorize the image as healthy or unhealthy. The proposed system has the potential to be an effective tool for farmers and agricultural experts to identify and treat diseases timely and accurately, thus improving the yield and quality of roses. This automated system will help farmers and experts in rose flower farming to identify and treat rose diseases in a timely and effective manner, thus ultimately improving yield and quality of roses.

1.6. Methodology of the Study

This research involves collecting a data set of rose leaf images, per-processing the images to remove noise and enhance features, utilizing the data set to train a convolutional neural network (CNN) model, then assessing the model's performance using a range of measures, including accuracy, precision, and recall.

1.6.1. Research design

This study follows experimental research. Experimental research is a type of research design that involves the manipulation of variables to observe the effects on the outcome variable. This type of research is usually conducted in a controlled environment to minimize the impact of extraneous factors on the results. Experimental research is important because it allows researchers to establish causal relationships between variables and to replicate studies. This makes experimental research a powerful tool for advancing scientific knowledge in various fields.

In this experimental study for deep learning-based rose leaf disease diagnosis from images, data preparation, implementation and evaluation are the basic steps followed, as discussed below.

1.6.2. Data collection and preparation

One of basic steps in this study is to collect datasets images of rose leaves that have been classified as either healthy or sick. This is followed by making sure the images have a consistent background and lighting, as well as excellent quality and resolution. The data set is also split to create training, validation, and testing sets from the data sets. To enhance the dataset's size, we apply the task of augmentation by flipping, rotating, and scaling images. Also, to make sure that the image pixel values are in the same range, images are normalized to the same image size.

1.6.3. Implementation tools

Deep learning can be used to achieve image-based rose leaf disease detection using several programming and packaging techniques.

- ✓ Python: Due to its simplicity, usability, and accessibility of several libraries like TensorFlow, Keras, and PyTorch, Python is a popular programming language for deep learning.
- ✓ Keras: TensorFlow is built upon the Keras high-level neural networks API. It offers a user-friendly interface for developing and improving deep learning models.
- ✓ Tensor flow: A flexible framework for developing and improving deep learning models is provided by its free TensorFlow deep learning package.
- ✓ Open CV: It offers a variety of tools and features that may be used for interpreting and composing images and videos, image filtering, feature detection, object recognition, and other image and video processing activities.
- ✓ Scikit learn is a well-known Python deep learning library that provides a variety of tools and functions for deep learning applications like classification, regression, clustering, and more.
- ✓ NumPy: is a Python library for scientific computing that offers support for big multi-dimensional arrays and matrices as well as several mathematical operations that may be performed on them.

1.6.4. Evaluation method

To ensure the reliability and effectiveness of image-based disease detection models in practical use, it is important to evaluate their ability to work with new datasets and identify rare or uncommon diseases. These evaluation methods are crucial in determining the overall performance of the models and their potential for real-world applications.

In this study accuracy, precision, recall, and F1-score are metrics used to evaluate the performance of classification models. Here is the difference between accuracy, precision, recall, and F1-score:

- ✓ **Accuracy** evaluates how accurate the model is overall at predicting both positive and negative classifications. The ratio of accurately predicted samples to the total number of samples is used to compute it.

- ✓ **Precision** determines the proportion of correct positive predictions among all the positive predictions made by the model. It is calculated as the ratio of genuine positives to the sum of true positives and false positives.
- ✓ **Recall** calculates the proportion of correct positive predictions made by the model out of all of its positive predictions. To calculate it, divide the total number of true positives by the number of false positives.
- ✓ **F1-score** is the precision and recall harmonic mean. When working with unbalanced datasets, it is frequently utilized as it offers a measure of both metrics that is balanced. To summarize, the F1-score offers a compromise between precision and recall, accuracy evaluates total correctness, precision represents positive predictive value, and recall measures sensitivity.

1.7. Thesis Organization

The remainder of the thesis is organized as follows: Chapter 1 introduces the research, presents the problem statement, outlines the thesis objectives, describes the research methods, and reviews related works. Chapter 2 discusses Rose Leaf Diseases, Digital Image Processing, Deep Learning, provides an overview of the literature review, and related works. In Chapter 3, the application of Deep Learning for detecting Rose Leaf Diseases is explored, covering fundamental concepts in Neural Networks and Deep Learning, the proposed architecture and system model, processing steps, the use of Convolutional Neural Networks, and evaluation metrics. Chapter 4 summarizes the main results of the thesis and discusses their implications. Finally, Chapter 5 offers conclusions based on the results obtained and recommendations for future work.

CHAPTER TWO

LITERATURE REVIEW

2.1. Overview

There are different rose diseases causing leaf, stem, flower, and root. Leaf diseases are the most common type of plant disease. They are brought on by numerous diseases, such as viruses, bacteria, and fungus. Leaf diseases may appear as spots, blights, mildew, rust, or viruses, among other manifestations. Plants that have leaf diseases may be weaker and more vulnerable to other pests and diseases. They may also decrease their capacity to produce fruit and flowers.

Stem Diseases that are afflicted by stem diseases may become weakened and eventually perish. They are frequently brought on by bacteria or fungi. Various symptoms of stem diseases include cankers, galls, and rot. Plants can become less strong and more susceptible to various diseases and pests because of stem diseases. Plants might die or fall over because of them.

Flower diseases that suffer from plants may see the flowers fade, change, or die off. They usually get passed on by bacteria or fungi. The symptoms of flower diseases can take many different forms, including blight, mildew, rust, and viruses. A plant's ability to produce the flowers can be reduced by flower diseases, which additionally make the flowers unattractive.

Root diseases can affect the roots of plants and might develop bacterial or fungal infections in their roots. Root infections can show up in many ways, such as wilting, yellowing, and death. Plants that are damaged by root infections are more vulnerable to other pests and diseases. They may also result in plant death.

Leaf diseases are the most common type of rose disease. This is because leaves are the most exposed part of the plant and are therefore more susceptible to infection. Leaf diseases can be caused by a variety of pathogens, including fungi, bacteria, and viruses.

2.2. Type of Rose leaf diseases

Rose leaf diseases are a major problem for rose growers. They can cause significant damage to the leaves, flowers, and stems of roses, and can even kill the plant. Some of the most common rose leaf diseases includes [16]:

- ✓ Black spot: The disease is characterized by dark, rounded patches on the leaves and is caused by a fungus.
- ✓ Downy Mildy: The disease is characterized by a white, powdery coating on the leaves and is caused by a fungus.

- ✓ Rust: This disease is caused by a fungus, and it is characterized by orange or red spots on the leaves.
- ✓ Leaf spot: This disease is caused by a variety of pathogens, and it is characterized by small, round spots on the leaves.
- ✓ Rose mosaic: A viral disease that causes the leaves to appear streaked or mottled.
- ✓ Rose rosette: A viral disease which results in excessive growth and leaf deformation.

2.3. Digital image processing

A digital image is a 2D array of brightness numbers, typically stored in binary format. Deep learning models can classify, detect, segment, and synthesize images, enabling them to classify, detect objects, segment, and synthesize images. These models can also be used to create new faces or objects [17].

The study and processing of digital photographs is the focus of the computer science discipline known as "digital image processing." [18].

It is used in a wide variety of applications, including rose leaf disease detection. Digital image processing techniques can be used to extract features from images that can be used to identify different types of rose leaf diseases. For example, the color and texture of the leaves can be used to identify different diseases.

2.3.1. Steps of Digital image processing

The basic procedures for processing digital images are [19]:

- ✓ Image acquisition: This is how an image from the real world is captured. A digital camera, scanner, or other image equipment can be used for this.
- ✓ Image preprocessing: This is the procedure that gets the picture ready for additional processing. This could include operations like image scaling, contrast improvement, and noise reduction (or image restoration). Enhancing an image is the process of making an image better. This may involve tasks such as sharpening, blurring, and color correction. Image restoration is the process of repairing an image that has been damaged or corrupted. This may involve tasks such as denoising, deblurring, and inpainting.
- ✓ Image segmentation: This is the procedure for segmenting an image into various areas according to its characteristics. This could entail edge detection, grouping, and thresholding.

- ✓ Feature extraction: This is the process of extracting features from an image that can be used for further processing. This may involve tasks such as shape descriptors, texture features, and color features.
- ✓ Object detection: This is the process of identifying objects in an image. This may involve tasks such as template matching, feature matching, and machine learning.
- ✓ Image recognition: This is the process of classifying an image into a particular category. This may involve tasks such as supporting vector machines, neural networks, and deep learning.
- ✓ Image compression: This is the method of lowering an image file's size without noticeably compromising its quality. This may involve tasks such as JPEG compression, PNG compression, and lossless compression.
- ✓ Image analysis: This is the process of extracting information from an image. This may involve tasks such as counting objects, measuring distances, and identifying patterns.

Digital image processing and deep learning are two fields that interact significantly and are closely related.

2.4. Deep learning

Artificial neural networks are used in deep learning, a kind of machine learning, to extract knowledge from data [20],It is a powerful tool for image classification, and it has been used to develop successful rose leaf disease detection models. Deep learning models are trained on large datasets of images of healthy and diseased roses. The models learn to identify the features that distinguish between healthy and diseased leaves.

Deep learning has changed digital image processing through improving rose leaf diseases image analysis to be more accurate and efficient.

A subset of machine learning methods called deep learning (DL) algorithms is based on the architecture and operation of the human brain. Below are a few of the most widely used DL algorithms [21]:

- ✓ Convolutional Neural Networks (CNNs): CNNs are used to recognize images and videos. Convolutional layers are utilized for feature extraction from images, and filters are applied to detect patterns.

- ✓ Recurrent Neural Networks (RNNs): Sequential data analysis tasks like speech recognition and natural language processing are performed by RNNs. To enable information to endure and be utilized across the network, they employ loops.
- ✓ Generative Adversarial Networks (GANs): Generating new data like a given dataset is the application of GANs. They are made up of two neural networks: a tool for discrimination that assesses the veracity of the information generated and a generator that produces new data.
- ✓ Autoencoders: For unsupervised learning tasks like dimensionality reduction and data compression, autoencoders are employed. They are made up of a decoder that uses the compressed representation to recreate the original data and an encoder that compresses the incoming data into a lower-dimensional form.

2.5. Related works

Researchers have conducted and are still conducting studies to investigate reasonable ways to improve the performance of Image Based Rose Leaf Diseases Detection Using Deep Learning. So far, much research has been conducted on different aspects of obtaining new ways to improve the performance of these models, and it is likely that the performance of these models will continue to improve in the future.

Researchers have investigated several ways to improve the performance of image-based rose leaf diseases detection using deep learning. These methods include:

- ✓ **Data augmentation:** By generating fresh images from existing images, the size of the data collection is manipulated in this way. By strengthening the model's resistance to changes in the images, this can aid in enhancing its performance.
- ✓ **Feature extraction:** This involves extracting features from the images that are relevant to the task of disease detection. This can help to improve the performance of the model by making it more efficient at identifying the features that are associated with different diseases.
- ✓ **Model architecture:** This involves choosing the architecture of the deep learning model that is used for disease detection. Different architectures have different strengths and weaknesses, so it is important to choose an architecture that is well-suited for the task at hand.
- ✓ **Training parameters:** This involves setting the parameters of the deep learning model, such as the learning rate and the number of epochs. These parameters can have a significant impact on the performance of the model, so it is important to tune them carefully.

In addition to these methods, researchers are also exploring other ways to improve the performance of image-based rose leaf diseases detection using deep learning. For example, some researchers are investigating the use of transfer learning, which involves training a model on a large data set of images of other objects and then fine-tuning the model on a smaller data set of images of rose leaves. Other researchers are investigating the use of reinforcement learning, which involves training a model to learn how to detect diseases by trial and error. Hereunder we present a literature survey of some of the selected works.

Verma, [22] proposed a deep learning-based image classification system for rose leaf disease detection. The system uses a convolutional neural network (CNN) to classify images of rose leaves into healthy and diseased classes. The system was trained on a data set of 2,000 images of rose leaves, and it achieved an accuracy of 97.3% on a test set of 500 images.

Batchuluun, [23] proposed a novel approach for classifying crop and plant diseases using thermal camera images, and a number of experiments were carried out. The studies made use of two different kinds of datasets: the open data set for paddy crops and the self-collected thermal plant data sets. While the Paddy Crops Open Data Set included photographs of both healthy and ill leaves, the Self-Collected Thermal Plant Data Set featured a variety of roses and rose leaf images. CNN-16 outperformed the current approaches with a 98.55 % accuracy rate utilizing the self-collected thermal plant data set, as reported in the research.

Parded,[24] explored a deep learning framework for rose leaf disease detection. The framework uses CNN to classify images of rose leaves into healthy and diseased classes. The framework was trained on a data set of 10,000 images of rose leaves, and it achieved an accuracy of 98.5% on a test set of 1,000 images.

Li [25], tried to detect rose leaf disease using a transfer learning strategy. Using a CNN that was pre-trained on a sizable dataset of pictures of different objects, the method operates. Then, using a dataset of photos of rose leaves, CNN was refined. 97.8% accuracy was attained by the method on a test set consisting of 500 images.

Poornam and Devaraj [26] provided a review of the literature on image-based rose leaf diseases detection using image processing and machine learning techniques. The paper discusses the different techniques that have been used for this task, and it provides an overview of the results that have been achieved.

Zogan, H [27] created a hybrid deep learning model to identify rose leaf disease. A CNN and a recurrent neural network (RNN) are combined in this model. RNN is utilized to categorize the photos into classifications of healthy and unhealthy, while CNN is used to extract features from the images. A test batch of 500 photos yielded a 99.5% accuracy rate for the model.

2.5.1. Research gap

There have been several research papers published on the use of digital image processing and deep learning for rose leaf disease detection. These papers have shown that these techniques can be used to develop effective models for detecting different types of rose leaf diseases. The authors, Sawarkar and Kawathekar [28], studied rose plant disease detection using digital image processing and discussed various image processing techniques for detecting diseases in rose plants. The importance of early detection of plant diseases and the disadvantages of traditional methods were highlighted. The paper also covered the advantages and limitations of different image processing techniques, such as segmentation, feature extraction, and classification. The authors provided a comprehensive comparison of various studies on rose plant disease detection and explored how digital image processing can be used to save rose plants from various diseases. Overall, the paper focused on the use of image processing techniques for detecting and preventing losses in the yield and quantity of agricultural products caused by plant diseases.

Ekka and Behera [29] studied disease detection in plant leaf using digital image processing technique. The aim of their research is to ensure good and disease-free production of crops by using software to automatically detect the affected area in a leaf and provide a better solution. The paper outlines various image processing techniques, such as image acquisition, per-processing, segmentation, and feature extraction, to identify the affected area of a leaf. This will help farmers to take necessary actions to prevent the spread of diseases in crops and ensure better quality and quantity of yield. Generally, this study examines the use of k-means clustering for accurately identifying leaf diseases through image processing. The process involves five stages: image acquisition, per-processing, segmentation, feature extraction, and characterization. By determining the extent of disease present in the leaf, farmers can apply the necessary number of pesticides to control the pests and increase crop yield. This methodology can be expanded by using different algorithms for segmentation and classification. By identifying the affected area of the leaf in percentage, farmers can easily and inexpensively correct the issue.

Halder, [30] conducted a literature review on plant disease detection through image processing. The review highlights the importance of using big data, decision making, ICT, and IoT to solve real-life problems related to agriculture. The authors emphasize the need for early detection of plant diseases to prevent harm to agricultural plant production and ensure sufficient crop growth to feed the increasing world population. The paper surveys various disease classification techniques that can be used for plant leaf disease detection and emphasizes the benefits of using automatic techniques for detecting plant diseases.

Table 1 below shows the summary of the above existing research works of section 2.4 for rose leaf detection with noticeably higher accuracy, although their efforts have certain limits as compared to the current situation. The top four methods used for comparison are listed here. [23-25, 27].

Table 1. 1: Summary of existing work and their limitations

Authors	**Problem**	**Accuracy**	**Limitations**
Pardede, [24]	Used a CNN to classify images of rose leaves into healthy and diseased classes.	98.5%	The study did not compare the performance of the CNN-based framework with other existing methods for rose leaf disease detection
Li, Y, [25]	Applied a CNN that was pre-trained on a large data set of images of other objects.	97.8%	The study covered limited types of plants or crops the approach may not be applicable to detecting other types of plant diseases as the pre-trained CNN was only trained on images of other objects and not specifically on plant-related images.
Batchuluun, [23]	The tests made use of the open data set for paddy crops and the self-collected thermal plant data sets.	98.55 %	The paper does not discuss the potential impact of environmental factors, such as temperature and humidity, on the accuracy of the classification method
Zogan,[27]	Used a combination of a CNN with a recurrent neural network (RNN)	99.5%	The paper does not discuss the potential challenges or limitations of using a hybrid model combining CNN and RNN

2.5.2. The research gaps this study attempted to fill.

The research gap addressed by the study on this Image-Based Rose Leaf Diseases Detection Using Deep Learning was the lack of a comprehensive and accurate method for detecting rose leaf diseases through image-based analysis using deep learning techniques. This study aimed to fill the gap by developing a deep learning model that could accurately identify and classify different types of rose leaf diseases based on images. In the previous study the researchers are also aimed to address the situational differences caused by weather conditions and humidity, which can impact the appearance of different disease types on rose leaves. By focusing on these factors, the study aimed to contribute to the development of a more robust and effective method for detecting and diagnosing rose leaf diseases using image-based analysis [31].

As noted by researchers, environmental factors influencing the visual appearance of rose leaf diseases are the following:

- ✓ Weather conditions such as temperature, rainfall, and sunlight can influence the growth and spread of various diseases on rose leaves.
- ✓ High humidity levels can create a favorable environment for the development of certain types of fungal diseases, such as powdery mildew and black spot, on rose leaves.
- ✓ Extreme weather events, such as heavy rain or drought, can stress rose plants and make them more susceptible to certain diseases.
- ✓ Variations in temperature and humidity can affect the visual characteristics of disease symptoms on rose leaves, making it challenging to accurately identify and classify the diseases based on visual inspection alone.
- ✓ The impact of weather and humidity on disease appearance underscores the need for a robust and reliable method, such as image-based analysis using deep learning, to accurately detect and diagnose rose leaf diseases under varying environmental conditions.

CHAPTER THREE

Modeling for Image Based Rose Leaf Diseases Detection Using Deep Learning

3.1. Overview

Modeling for Image Based Rose Leaf Diseases Detection Using Deep Learning involves using deep learning techniques to develop a model that can accurately detect diseases in rose leaves from images. This typically involves training a deep learning model, such as a convolutional neural network (CNN), on a large dataset of images of healthy and diseased rose leaves. The model learns to identify patterns and features in the images that are indicative of different diseases.

The process usually involves several steps, including data collection, preprocessing, model training, validation, and testing. Once the model is trained and validated, it can be used to automatically analyze new images of rose leaves and classify them as healthy or diseased, and if diseased, identify the specific disease present.

Deep learning models have shown great promise in image-based disease detection across various crops, including roses.

They have the potential to assist farmers and researchers in early and accurate disease diagnosis, which can help in timely intervention and management of plant diseases.

3.2. Proposed Model

Developing a neural network model that uses deep learning algorithms to detect diseases in rose leaves. The system will analyze images captured using a camera and identify fresh, black spot and Downy mildew. The model will be trained using a large dataset of labeled images, and once an image is analyzed, the system will provide a prediction. This system will be useful for farmers, gardeners, and anyone who wants to monitor the health of their roses and prevent disease spread.

The proposed system for image-based rose leaf diseases detection using deep learning consists of the following components:

1) Rose leaves image capturing - To capture images of rose leave Smartphone Samsung Galaxy A70 Android Os, 6.7 Display size inch and 32MP back and front camera is used.
2) Preprocessing module - This module is responsible for preparing the images for analysis. To prepare data for training and improve the model's performance, image

processing is essential to image-based deep learning. The following are some important tasks:

- ✓ Data augmentation: Using methods like rotations, scaling, cropping, and color jittering, one can artificially increase the size and diversity of the dataset. This prevents the model from overfitting on certain patterns and improves generalization.
- ✓ Normalization: To improve convergence and stability during training, scale pixel values to a common range (such as 0-1 or -1 to 1).
- ✓ Removing undesired distortions and aberrations that can confuse the model and affect its accuracy is known as noise removal.

3) Feature extraction module - This module will extract relevant features from the preprocessed images.
4) Deep learning model - This model will be trained using a large dataset of labeled images and will be able to identify different types of diseases that affect rose leaves.

To deploy the proposed model for image-based rose leaf diseases detection using deep learning, the following steps may be involved:

- ✓ Collect and label a large dataset of images of rose leaves with and without diseases.
- ✓ Train a deep learning model using the labeled dataset. This may involve selecting an appropriate architecture for the model, choosing hyperparameters, and optimizing the model's performance.
- ✓ Test the trained model on a separate dataset of labeled images to evaluate its accuracy and performance.
- ✓ Integrate the trained model that can accept images from a user and provide a prediction of healthy and disease.
- ✓ Monitor the performance of the system and update the model as needed to improve accuracy and performance.

To improve recognition and classification accuracy and reduce time complexity, a comparative performance analysis of several transfer learning models using deep CNNs is attempted in this work. In Figure 3.1, the workflow architecture is shown. CNN models that had already been trained were used in the tests, which made use of the rose leaf dataset.

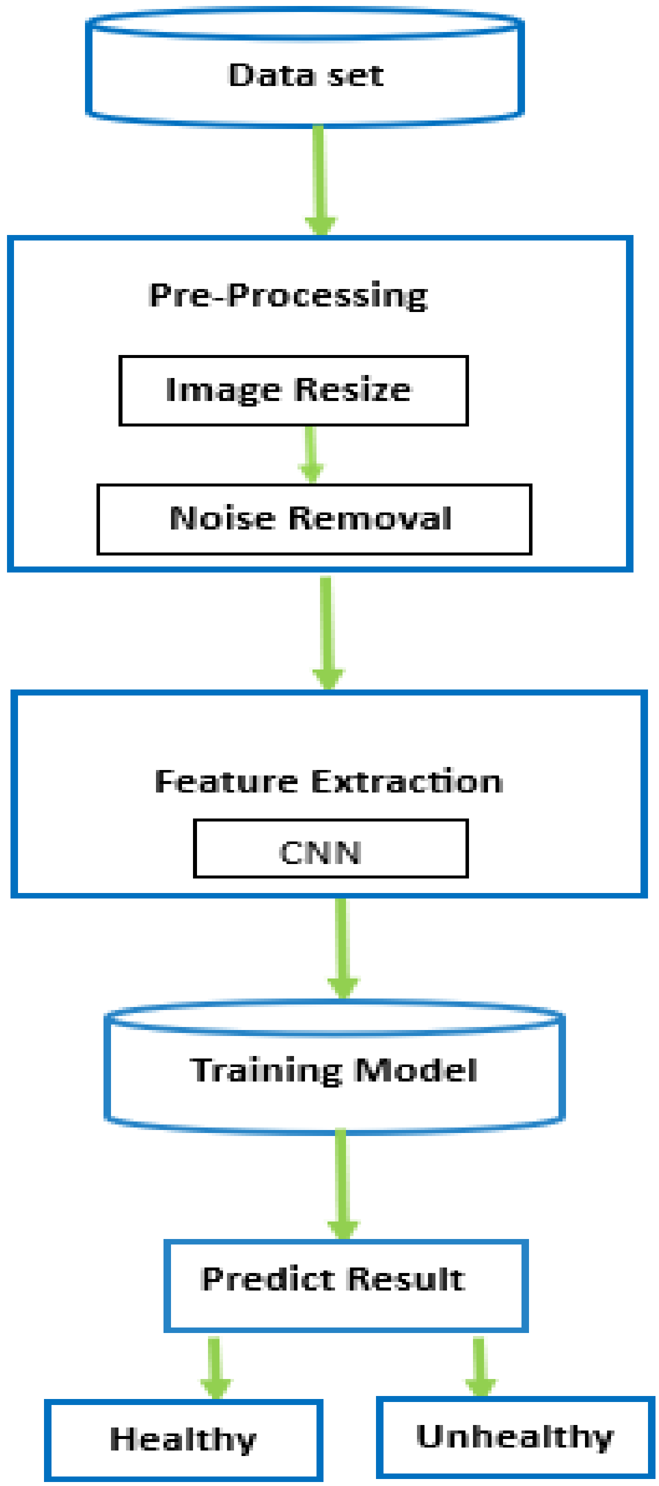

Figure 3. 1:The model Architecture Diagram

The proposed Model in this study is presented in figure 3.1. It has three components, such as preprocessing, feature extraction, and classification.

3.3. Image Processing

The workflow diagram for rose leaf disease detection encompasses a sequence of actions that commence with obtaining images and conclude with evaluating the results. The primary objective is to precisely identify the health status of a given rose leaf by analyzing its visual characteristics. A typical workflow diagram for rose leaf disease detection might include the following steps:

3.3.1. Image Acquisition

The procedure begins with image acquisition, which entails taking a picture of an entire rose leaf from a predetermined image collection. The technique is implemented using both public and private picture collections. Taking pictures of rose leaves that can show disease symptoms is the first stage in this process. Using a camera or any other imaging equipment, photos can be taken to accomplish this. While a self-developed soybean image database is made by overlaying damaged soybean leaves on a white background to eliminate background complexity, the Rose Leaf image database of Soybean, Potato, and Tomato is gathered from the Plant Village picture database. Next, a high-resolution smartphone camera is used to take a picture of the rose leaf. [32].

The process of converting electronic data from a sensor into a numerical representation using a device like a camera or scanner is known as image acquisition. Choosing the right image acquisition method is the first step in creating any machine vision system. Information is acquired through images, and technology attempts to replicate the function of human vision by electronically interpreting and perceiving an image. [33]. The Machine 11 vision system can do computations that enable the automatic extraction and analysis of pertinent object information from captured images. Having a source-input image that functions inside a measured and controlled boundary is the aim of image acquisition. Since the image collected in image acquisition is the raw image or unprocessed, it requires further image processing techniques.

In this study, we used a smartphone camera (To capture images of rose leave Smartphone Samsung Galaxy A70 Android Os, 6.7 Display size inch and 32MP back and front camera is used). The data was collected from The Ethio Agri CEEFT PLC Holeta Flower Farm in the Oromia region, Ethiopia, And the total numbers of images collected from those places are 115 fresh,95 Blackspot, and 105 Downy Mildy. The total number of images taken for this study was 315 and augmented to 945 and the rest images were generated from the public dataset, from each class fresh, Blak Spot and Downy Mildy, which are shown in Table 3.1 below.

Table 3. 1: Number of images taken from each category.

Category	Total number of images
Fresh	1430
Blak spot	1434
Downy Mildy	1478

(A) (B) (C)

Figure 3. 2:Samples of acquired rose leaf images from dataset (A) Healthy (B) Downy mildew (C) Black spot.

The above Figure 3.2 displays the acquired images of healthy leaf samples of rose from the defined image datasets.

3.3.2. Image Preprocessing

Images need to be preprocessed before they can be used for inference and model training. Image processing is a method of enhancing the quality of photos by manipulating them in different ways. [34].

The practice of adjusting digital images to improve their quality, lower noise, or address lighting problems is known as image preprocessing. Furthermore, photo analysis explains how to extract information from pictures by sorting useful areas from uninteresting ones. [35]. Image processing techniques can be used to improve the quality of images. A method can be used to group together various image processing operations. These consist of the following: preprocessing, image enhancement, image restoration, image data compression, image analysis, and image representation. There are a lot of extraneous details in many photographs. The process of enhancing image and pixel quality is also done after preprocessing. Image preprocessing raises the quality of the image by eliminating noise and minimizing distortions in the input photos. Photographs consequently require preprocessing techniques because of the obtrusive components they contain. [36]. The pre-processing or purification of the picture data is the fundamental phase, and before creating the model, most of the machine learning (ML) engineers spend more time on this stage. The management of missing values, outlier detection, and the removal of unwanted or noisy data will be the main inputs for the model we want to build. Image preprocessing is a technique to convert raw image data into clean image data because most raw picture data has noise and has some missing or partial values, inconsistent values, and erroneous values.

Once the images are acquired, they undergo preprocessing to eliminate noise and enhance the relevant features necessary for disease detection. This step ensures that the subsequent analysis is based on clean and accurate data. An improved image is produced as an output of the preprocessing stage from a raw image. Various preprocessing procedures were employed in this work to remove noise, equalize, and resize images. The subsequent subsections provide a description of the preprocessing algorithms that were employed in the execution of this investigation.

- Image resizing

An image's size can be changed without losing any content by resizing it. Resizing an image modifies its dimensions, which frequently affects the file size and image quality. The most common reason for reducing the size of large photographs is to make them smaller so that they can be shared online or via email. Image processing offers a variety of methods for resizing images. such as Box sampling, Mipmap and Fourier-transform techniques, Sinc and Lanczos resampling, bilinear and bicubic algorithms, and nearest-neighbor interpolation. We apply Bicubic Interpolation since it computes the final interpolated value by averaging 16 pixels at various distances between a known and unknown pixel, bicubic interpolation is used because it is the best. The calculation gives pixels that are closer together a greater value. The image produced by this method is sharper than the one created by the earlier methods. It might be the ideal balance between performance quality and improved results [37].

- Histogram equalization

  A technique for adjusting contrast in image processing that makes use of the image's histogram is called histogram equalization. By modifying the input histogram, this technique generates an output with a uniform histogram one in which the different pixel intensities are dispersed equally across the whole dynamic range. The intensities on the histogram can be more evenly spread by making this adjustment. This makes it possible for regions with less local contrast to become more contrasty. An essential pre-processing stage in image-based deep learning for the diagnosis of rose leaf disease is histogram equalization. This can improve the model's capacity to detect sick areas if done correctly.

1) Highlighting Subtle Variations

Rose leaf diseases frequently show one another as small variations in pattern, color, and texture. These small variations can be enhanced by histogram equalization, which makes them easier for the deep learning model to detect and analyze.

2) Enhanced Feature Extraction and Contrast:

Contrast is improved, making darker areas—possible lesions—and lighter, healthy areas more noticeable, by extending the intensity distribution across the range that is detectable. This helps the model extract relevant characteristics required to identify diseases.

3) A decrease in overfitting

Images have different brightness levels by nature. Unbalanced pictures may cause the model to pick up characteristics unique to those changes rather than true disease indications. A more uniform distribution is produced by equalization, which decreases overfitting and enhances adaptability to new tests.

Generally, Rose leaves with diseases like powdery mildew or black spot frequently have darker or lighter colored areas on them. By sharpening the focus of these color changes, equalization can help the model identify these sick spots.

- The following graphs present the results of the fresh, black spot and downy mildew detection after histogram equalization, showcasing both the original and equalized images.

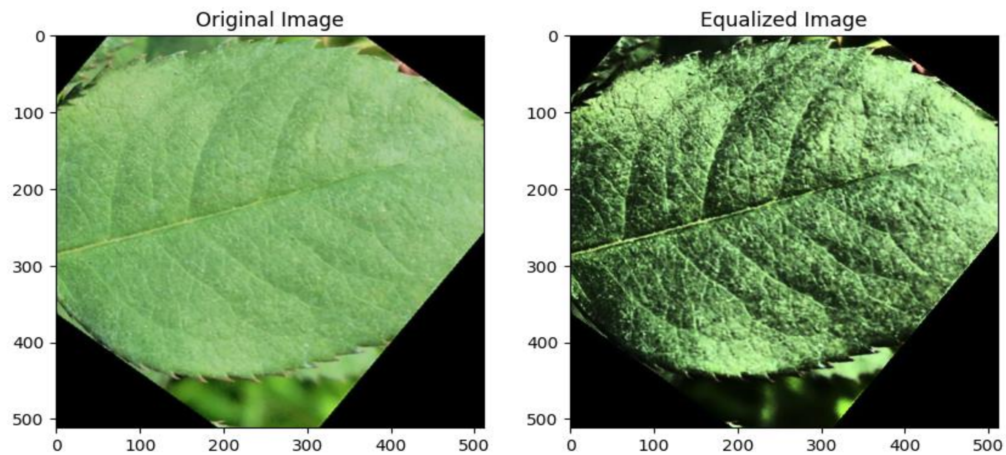

Figure 3. 3:Histogram equalization image for a healthy rose leaf

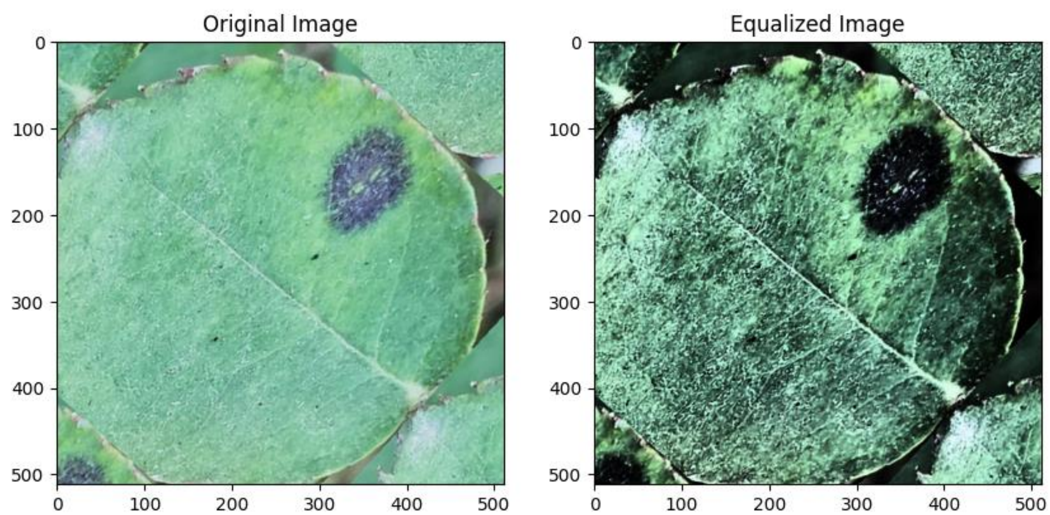

Figure 3. 4:Histogram equalization image for a Black spot rose leaf.

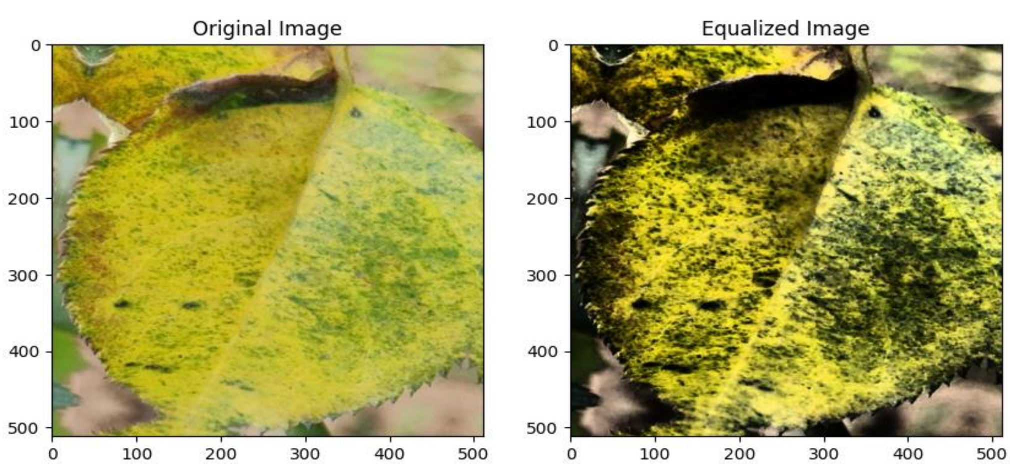

Figure 3. 5:Histogram equalization image for a downy mildew rose leaf. Noise removal

There is a danger of image deterioration when an image is acquired or transferred for image processing applications. There may be noise in the deterioration. Image processing includes the task of noise removal. Different sorts of noise can make a picture difficult to read and provide a barrier in many image-processing applications [38]. The kind of noise affecting the image determines the degree of the noise reduction issue. We take into account the median filtering strategies in this study since they have benefits such maintaining contrast across stages, maintaining boundaries, and being less susceptible to extreme values or outliers than the mean. [38]. The median filter is included in a non-linear filter because it does not use a convolution process. The intensity value of each pixel is sorted to do the calculation. The original pixels will then be changed out by fresh ones that have undergone the counting procedure. The median filter selects pixels whose brightness is equal to or greater than the median value of all the pixels by evaluating each pixel's brightness level. The neighborhood window's pixel values are ranked by intensity during the median filtering procedure, and the middle value (the median) is used as the output value for the pixel being evaluate [39].

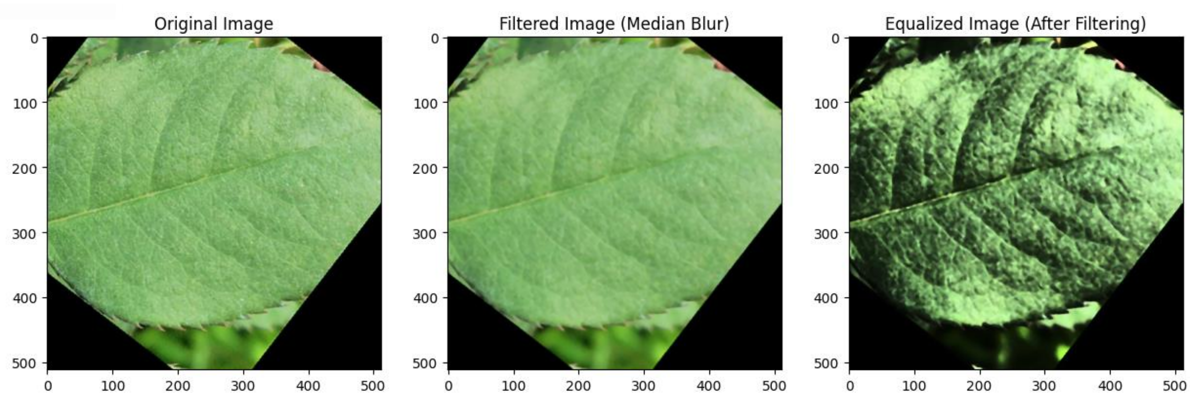

Figure 3. 6: Median filtering noise removal for Healthy rose leaf image

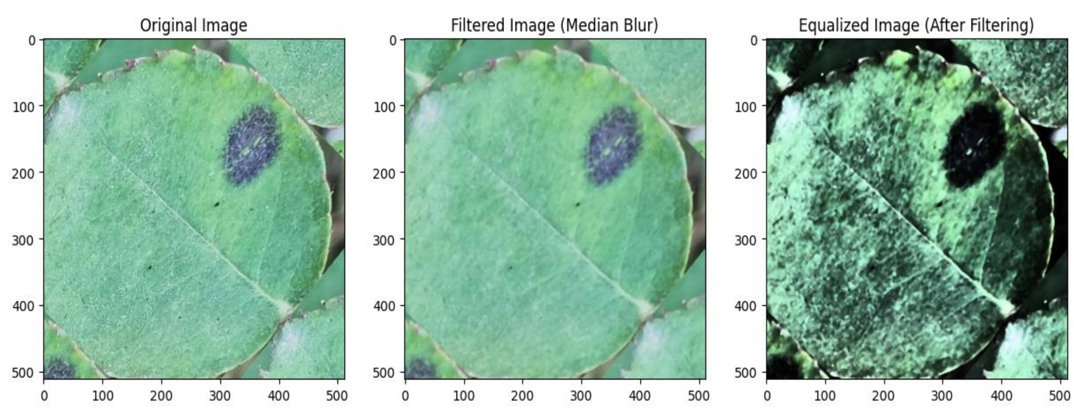

Figure 3. 7:Median filtering noise removal for black spot rose leaf image.

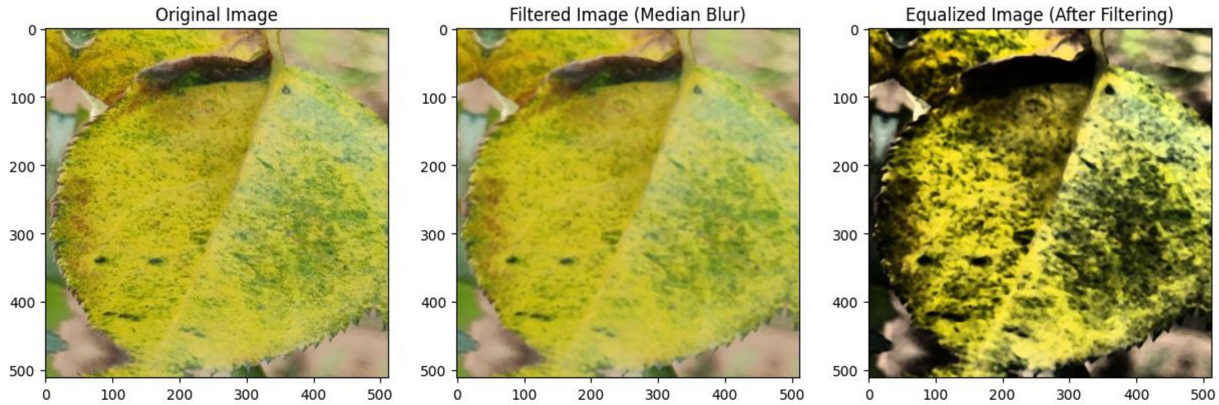

Figure 3. 8: Median filtering noise removal for Downy mildew rose leaf image.

3.4. Image Feature Extraction

Following preprocessing, relevant features are extracted from the preprocessed images. These features encompass various aspects such as color, texture, shape, and size, which are crucial for identifying disease patterns.

In the image processing system, feature extraction comes after preprocessing. It is a crucial stage in the creation of any pattern categorization and attempts to extract the important details that define each class. During this procedure, pertinent characteristics are taken out of objects or alphabets to create feature vectors. Classifiers then employ these feature vectors to identify the input unit with the desired output unit. By examining these characteristics, the classifier finds it easier to differentiate between various classes because it can do so with relative ease. The technique of extracting the most significant information from the raw data is known as feature extraction. Finding the collection of factors that accurately determine a character's shape is known as feature extraction. Every character is represented by a feature vector during the feature extraction stage, which serves as its identity. Finding a set of features that maximizes identification rate with the fewest amount of components and creates a similar feature set for several instances of the same symbol is the main objective of feature extraction. [40].

It's a step in the dimensionality reduction process, which breaks down a large raw data collection into smaller, easier-to-manage groupings. It will therefore be simpler to process when you want to. This huge data sets' abundance of variables is by far their most significant feature. Processing these variables takes a lot of computer power. Hence, feature extraction efficiently reduces the amount of data by choosing and combining variables into features to help extract the optimal feature from such large data sets. These characteristics are simple to handle while accurately and creatively describing the actual data set [41].

3.4.1. CNN for Feature Extraction

CNN for feature extraction: - One use of deep learning for signal processing problems is the convolutional neural network. Feature extraction and picture classification are its two main uses in image processing. Multiple convolutional layers are used in feature extraction, which is then followed by pooling and an activation function. Each of the several comparable steps in the CNN feature extraction process is composed of the pooling function and the convolution layer, two cascading layers. Using CNN for feature extraction is a very efficient way to automatically extract the best features from a large number of training datasets.[42]. CNNs have contributed significantly to advancements in leaf segmentation applications by revolutionizing the field with their unique capabilities. Here are some ways in which CNNs have made a significant impact [42].

Enhanced Feature Extraction: CNNs excel at learning hierarchical representations of images, automatically extracting meaningful features from raw pixel data. This ability allows them to capture both low-level details and high-level semantic information, enabling accurate identification of distinguishing characteristics of leaves, such as textures, shapes, and patterns. This enhanced feature extraction is crucial for precise leaf segmentation.

- ❖ Contextual Understanding: CNNs could capture spatial relationships and contextual information within an image. By analyzing local pixel neighborhoods through convolutional operations, CNNs can understand the context of each pixel in relation to its surroundings. This contextual understanding is invaluable for accurately segmenting complex leaf structures, including overlapping leaves and intricate vein patterns.
- ❖ Adaptability and Generalization: CNNs are trainable models that can adapt to different leaf variations and handle diverse datasets. By training on large, labeled datasets, CNNs learn to generalize and recognize common leaf characteristics across different species, shapes, and sizes. This adaptability allows CNNs to perform well on unseen leaf images, making them suitable for real-world applications where variability in leaf appearance is expected.
- ❖ Automation and Efficiency: CNN-based segmentation approaches automate the leaf segmentation process, eliminating the need for manual intervention or handcrafted features. This automation significantly reduces the time and effort required for analyzing large datasets of leaf images. The efficiency of CNNs enables rapid and accurate leaf

segmentation, facilitating advancements in various applications such as disease detection, trait extraction, and phenotyping.

- Continuous Research and Advancements: Researchers are continuously exploring new architectures, techniques, and optimization methods for CNN-based leaf segmentation. These advancements aim to further enhance accuracy, speed, and robustness. For example, the integration of additional modules like skip connections, attention mechanisms, or advanced loss functions has shown promising results in improving segmentation performance.

3.5. Classification

The technique of identifying and describing features from a leaf image is known as feature extraction and description [43] The intensity, texture, shape, and spectral information of the leaf image can all be used to extract features. Then, models based on mathematics or statistics are used to describe the features. In many plant image analysis applications, such as leaf disease detection, leaf trait extraction, and phenotyping, feature extraction and description is a crucial step.

3.5.1. Support vector machine (SVM)

The following uses of SVMs for image-based disease identification are shown using images and formulas [44].

1. Extraction of Features:
   - ✓ To extract pertinent elements like texture, color variations, and forms, images are pre-processed and evaluated. Techniques like edge detection, feature identifiers, and histogram equalization may be used in this.
2. Representation of Feature Vectors:
   - ✓ A numerical vector that represents the image is created using the features that were extracted. The SVM algorithm uses this vector as its input.
3. Labeling Data:
   - ✓ By labeling images as either healthy or unhealthy, a supervised learning environment is established for the SVM.
4. Hyperplane Separation:
   - ✓ The SVM algorithm looks for an edge in the high-dimensional feature space that best divides the labeled data points into distinct classes (healthy vs. unhealthy).

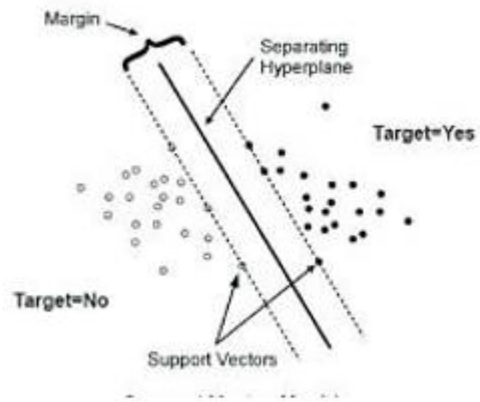

Figure 3. 9:SVM Hyperplane separating diseased [45]

5. Maximizing Margin:

- ✓ The best hyperplane maximizes the margin, or the separation between the best hyperplane and the support vectors—the nearest data points for each class. This provides a strong boundary for decisions.

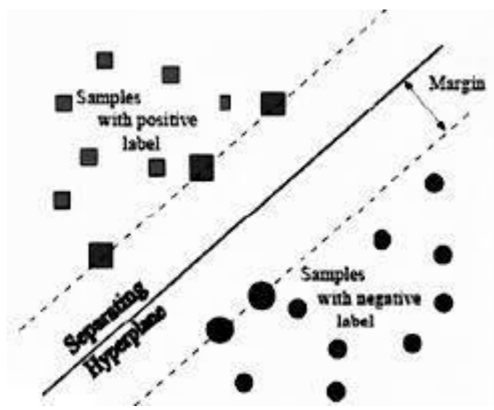

Figure 3. 10:SVM Hyperplane with Maximal Margin [45]

Hyperplane Equation:

$w^T x + b = 0$ (3.3)

Where:

w: weight vector, x: feature vector, and b: bias term

- ➢ Support Vector Machin (SVM) Algorithm used for Classification.
  - ✓ Import SVM module: Load the SVM classifier from a library.
  - ✓ Initialize model: Create an SVM object with a chosen kernel (default RBF) and other hyperparameters.

- ✓ Train model: Fit the SVM to the training data.
- ✓ Test model: Predict labels for the test data.
- ✓ Evaluate accuracy: Calculate the accuracy score using a metrics library.
- ✓ Print Accuracy: Display the model's test accuracy.

3.5.2. K-Nearest Neighbor

KNN (K-Nearest Neighbors) provides insightful information and possible uses in image-based plant disease identification, currently, it is not the driving force. KNN is valuable since it is easy to understand and can help with deep learning growth. It also provides unique insights into disease detection. This is a brief overview [46].

- ➢ The Operation of KNN:
  1) Feature extraction involves looking for pertinent features in photos, such as textures, colors, and forms.
  2) Determine the distances between each training image and the query image using the distance measurement method.
  3) Find the K Closest Neighbors: Determine which K data points resemble the query image the most.
  4) Majority Vote: Label the query image with the condition that is most prevalent among those K neighbors.
- ➢ KNN's merits
  1) Easy to understand and comprehend: The influence of neighbors on classification is easily understood.
  2) useful for little datasets Effective and precise when working with smaller datasets.
  3) Finding anomalous patterns in disease that deviate from the norm is known as anomaly identification.
- ➢ KNN's limitations:
  1) Scalability: Computationally complex for huge datasets.
  2) The dimensionality curse: In high-dimensional feature areas, accuracy can decrease.
  3) K selection: Determining the ideal K value necessitates careful adjustment.
- ➢ K-Nearest Neighbors (KNN) Algorithm is used for Classification.
  - ✓ Prepare data: Clean and split into training and test sets.

- ✓ Initialize model: Create a KNN classifier with a chosen number of neighbors (k).
- ✓ Train model: Fit the KNN model to the training data.
- ✓ Evaluate training accuracy: Predict labels for training data and calculate accuracy.

Evaluate test accuracy: Predict labels for test data and calculate accuracy.

3.5.3. CNNs

Another popular type of Neural Networks is the CNNs that are concrete case of Deep Learning Neural Networks. They are widely applicable in image processing and computer vision related applications. A simple CNN has a sequence of layers, and every layer is responsible for some specific tasks. There is an input layer, an output layer, and several hidden layers to form CNN. Some layers are fixed for convolutional operations; hence, it is named Convolutional Neural Network. Generally, there are three main types of layers. Convolutional layers, Pooling Layers, and Fully Connected Layers are present in a stacking to build a full CNN architecture. These networks have the potential to reduce the number of training parameters and increase the computational efficiency for 2D and 3D images. They have trained using the back propagation and gradient decent as in the standard Neural Networks. Similarly, they also preserve the spatial relationships in an image that are very important factor in medical imaging.

The benefits of a convolutional neural network include the following [47].

- ✓ Computationally, CNN is effective.
- ✓ It makes use of special algorithms for pooling and convolution in addition to parameter sharing. CNN models are becoming more and more popular globally because they can now function on any platform.
- ✓ Without the assistance of a human, it finds the necessary features.
- ✓ It can be used in a range of businesses to carry out important activities including, among others, object identification, document analysis, facial recognition, and climate comprehension.
- ✓ Extract useful features from a trained CNN by giving it your data at each level and adjusting the CNN significantly for a specific task.

Because of the above reason CNN is the best candidate for the development Image based rose leaf diseases detection system. The building blocks of CNN are convolutional layer, pooling layer and fully connected layer, as shown in figure 3.4 below.

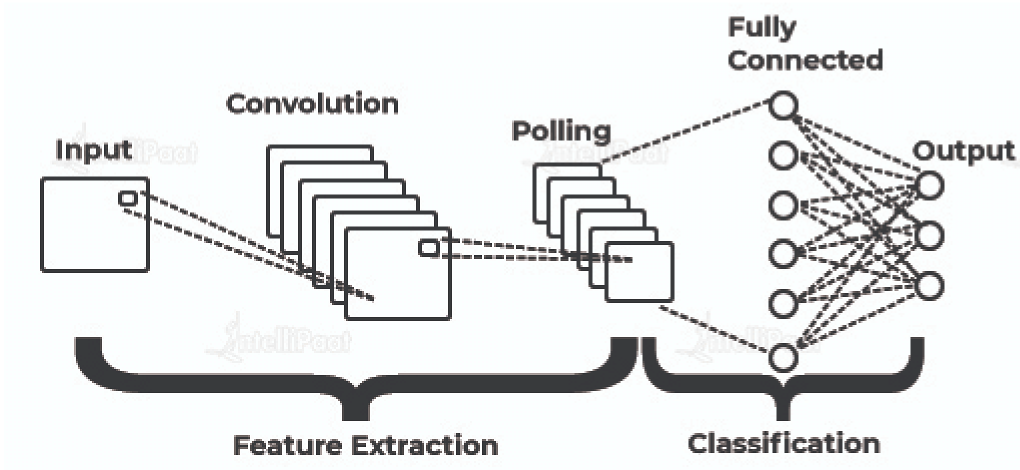

Figure 3. 11:CNN architecture [47]

3.6. Layers of a CNN

3.6.1. Convolutional Layer

The aim of this layer is to learn or detect features in the input image such as lines, edges, colors, etc. It can also learn spatial hierarchies of patterns in an image by preserving spatial relationships. This can be achieved through a set of different convolutional layers.

Generally, Convolutional layer produces an activation map by scanning the pictures several pixels at a time using a filter. Fig 3.12 shows the internal working of the convolution layer.

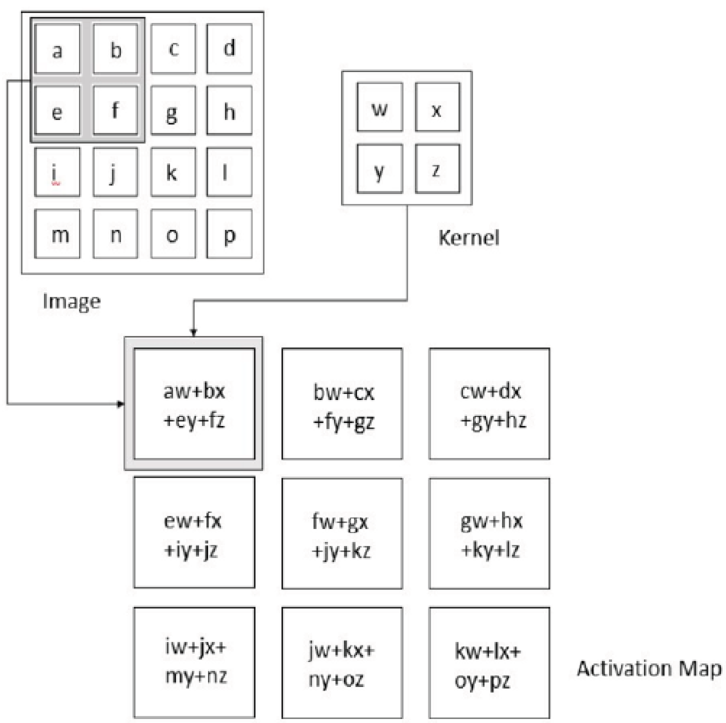

Figure 3. 12:Convolution Layer [47]

In the above figure the first convolutional layer learns the basic elements like edges, lines and the second convolutional layer is responsible for learning the patterns composed of basic elements learned in the previous layer. In this way, very complex patterns are learned through different convolutional layers. In convolution operation, a small window, i.e., kernel / filter moves over an input image and corresponding elements of two windows are multiplied and summed-up to compute a scalar value. This process continues until the whole image is covered. A convolution operation starts with 180-degree rotation of kernel and then element wise multiplication to get the result. The key hyper parameters of this layer are the size and number of the kernels / filters, the stride and padding.

3.6.2. Pooling Layer

The Pooling Layer is used to reduce the size of the feature maps created through convolutional layers. In pooling operation, a single value is computed for a small window in the feature map. Based on the pooling operation, it is named as Average Pooling or Max Pooling. Importantly, there are no learnable parameters associated with a Pooling Layer [48].

In general, the pooling layer minimizes the volume of data produced by the convolutional layer to improve storage efficiency. Figure 3.13 shows the internal working of the pooling layer.

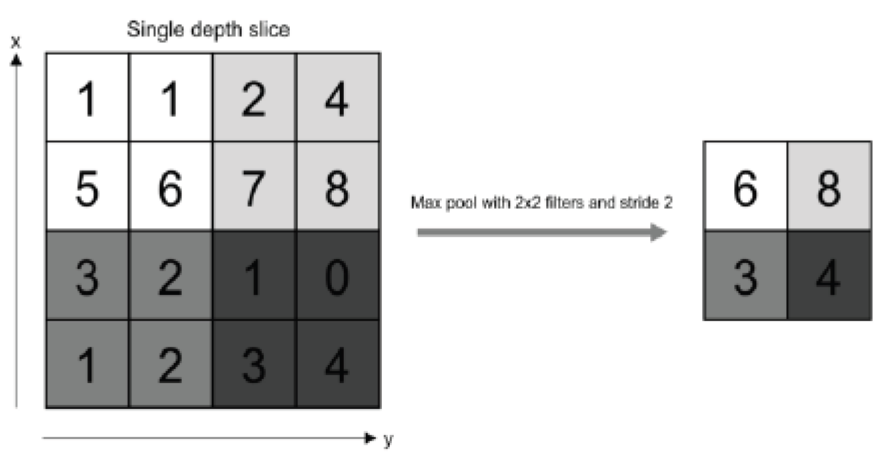

Figure 3. 13:Pooling Layer [49]

3.6.3. Fully Connected Layer

The final layers of a CNN are frequently fully connected. These layers add up the weighted sum of the previous layer's characteristics, indicating the exact mix of "ingredients “needed to get a given intended output result. In the case of a completely connected layer, each element of each output feature is calculated using all of the elements of the previous layer's features. The number of fully connected layers considered in designing the convolutional layer is obtained by using empirical studies.

Through a series of convolution and pooling operations, the size of resulting feature maps gradually decreases. Finally, these feature maps are rearranged into a vector that is fed to Fully Connected Layers. Figure 3.14 shows the internal working of fully connected layer.

- ✓ The output of the previous layers is "flattened" into a single vector, which serves as an input for the following stage in a fully connected input layer.
- ✓ To predict the correct label, the first fully connected layer applies weights to the inputs from the feature analysis.
- ✓ Ultimately, a fully connected output layer provides the probability for every label.

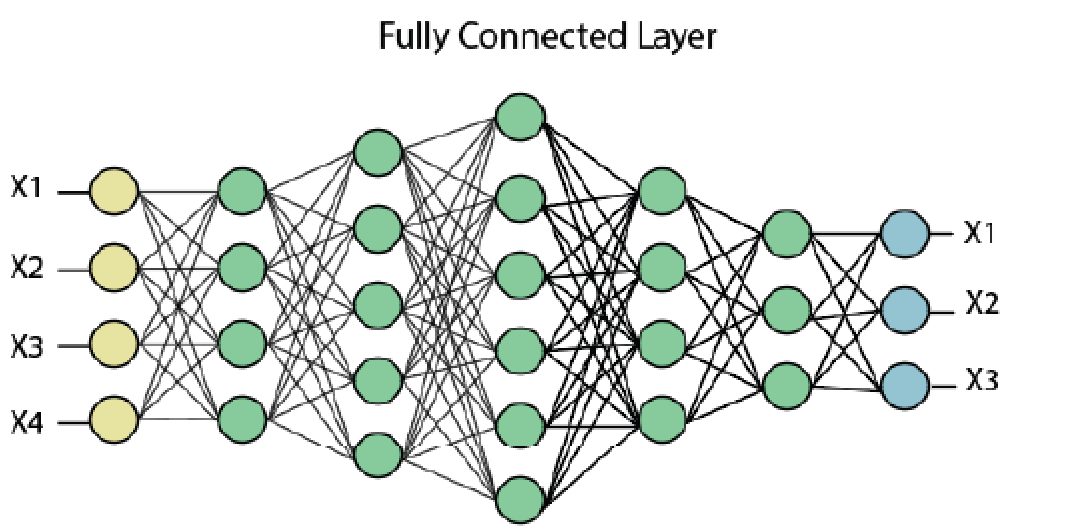

Figure 3. 14:Fully connected layer [47]

fully connected (FC) layers function in deep learning-based image-based rose leaf disease detection:

The deep learning model for leaf disease detection uses FC layers. They take on the role of ultimate decision-makers, converting extracted information into well-informed forecasts and, in the process, pushing the limits of automated and accurate disease detection in plant health monitoring and agriculture.

FC players. These layers modify features extracted from the image, such as textures, forms, and color changes. Convolutional layers extract complex elements, pooling layers reduce complexity while preserving important details, and flattening summarizes the features into a one-dimensional vector. The FC layer then functions as a neural network, with weighted connections allowing each neuron to receive information from every other neuron. Depending on the classification problem, activation functions like sigmoid and SoftMax are utilized for decision-making. Backpropagation and a loss function are used during the model's training process to modify the weights and biases. The model becomes more adept at distinguishing between healthy and damaged leaves as well as recognizing disease kinds through recurrent training.

A model is produced by Fully connected layers performs as a strong decision-maker in image-based deep learning for disease detection, taking the extracted features from CNNs and turning them into predictions:

- ✓ Feature transformation: An image that has been condensed, like a summary report from the CNNs, is sent to the FC layers.
- ✓ Decision-Making Web: Using weighted connections, neurons in the FC layer examine this report, concentrating on various elements (such as leaf textures or color differences).
- ✓ Activation and Prediction: Neurons use functions like sigmoid or SoftMax to determine the presence of a disease (e.g., via likelihood score).

  Model Building: Training fine-tunes the model's knowledge by adjusting weights and biases in the FC layer based on how well predictions match actual labels.

  Iterative Refinement: As a result of several training examples, the FC layer gains proficiency in identifying disease patterns.

3.6. Basic Concepts in Neural Networks and Deep Learning

The fundamental components of the Deep Learning framework are neural networks, which are merged to create complex neural architectures. These architectures are extensively utilized in various areas of computer vision and image processing, such as character recognition, face recognition, and automated language processing. Various neural network architectures, including Multilayer Perceptron (MP), CNN, and Recurrent Neural Network, exist. These models are designed to acquire abstract knowledge by transitioning from low-level features to high-level representations. Importantly, this learning process requires minimal human intervention and eliminates the need for manual feature engineering [50].

3.6.1. Deep Neural Networks

Deep Neural Networks are inspired by the complex neural nets of the human brain.[51]This section covers the fundamental ideas behind Deep Neural Networks, including neurons, weights, and activation functions. There are various types of deep neural architectures, such as MP, Long Short-Term Memory, and CNN. Among these, MP and CNN are discussed in detail since they are the simplest model to comprehend and form the basis of the main contributions in this thesis, particularly CNNs.

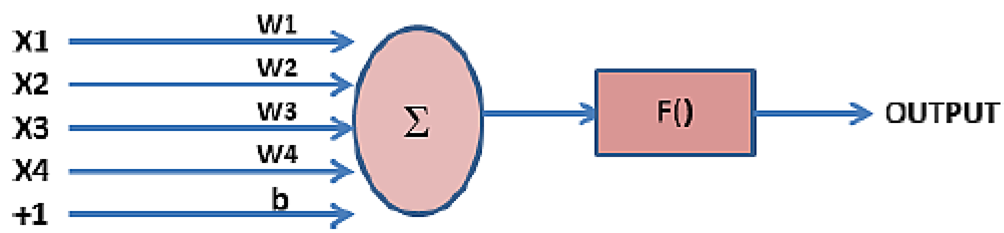

Figure 3. 15: A single neuron with five inputs, x1, x2, x3, x4, and b [51]

3.6.2. Neurons, Weights, and Activation Functions

A neuron is a crucial component of neural networks, which are created by linking multiple neurons together. Figure 3.15 displays the fundamental structure of a single neuron that has five inputs. When a neuron receives several inputs along with their respective weights, it computes the final output, as demonstrated in Figure 3.1.

$$Output = f(\sum_{1}^{i=4} (w_i x_i + b)) \qquad (3.1)$$

The non-linear activation function is represented by f () and the bias is denoted by the parameter b. The bias parameter allows an additional value to be added to the data. After the weighted inputs and bias are calculated, a non-linear activation function (f ()) is applied to further transform the data and compute the final output of a neuron. Several well-known activation functions include Sigmoid, Hyperbolic tangent, SoftMax, Rectified Linear Unit, Exponential Linear Units, and Gaussian Error Linear Unit.

a) SoftMax: For classification tasks, a neural network's output layer frequently uses the function SoftMax. It takes an arbitrary real-valued score vector and squashes it into a vector of values that add up to one, ranging from 0 to 1. For multiclass classification issues, this is helpful.

b) Rectified Linear Unit (ReLU):

- ✓ ReLU is an activation function that introduces non-linearity to the network.

- ✓ It is defined as f(x) = max (0, x), which means that it returns 0 for negative input values and the input value for positive input values.
  - ✓ ReLU is widely used in neural networks and helps address the vanishing gradient problem.

c) Exponential Linear Units (ELU):

- ✓ ELU is an activation function that has a similar form to ReLU but has some advantages, particularly for negative input values. ELU has a non-zero gradient for negative input values, which can help with the vanishing gradient problem and improve learning in deep networks.

d) Gaussian Error Linear Unit (GELU):

An activation function known as GELU, which is based on the Gaussian cumulative distribution function, has demonstrated strong performance in some neural network architectural types, especially transformer models.

- ➢ The activation functions such as SoftMax and Rectified Linear Unit are considered in this thesis and their descriptions are presented as follows:

3.6.2.1. SoftMax

SoftMax is utilized in multiclass systems to determine the probability that a given data point belongs to a specific class. This type of activation function produces probability distribution values for multiclass identification problems. The output layer consists of one neuron for each class, and the sum of all probabilities for each neuron must be equal to one. The mathematical formula for SoftMax is as follows [52]:

$$f_{Softmax}(a_i) = \frac{exp(a_i)}{\sum_j exp(a_j)} \qquad (3.2)$$

Where:

a shows the values from different neurons of the output layer.

exp (a_i) denotes the exponential function, which raises the mathematical constant e to the power of the argument.

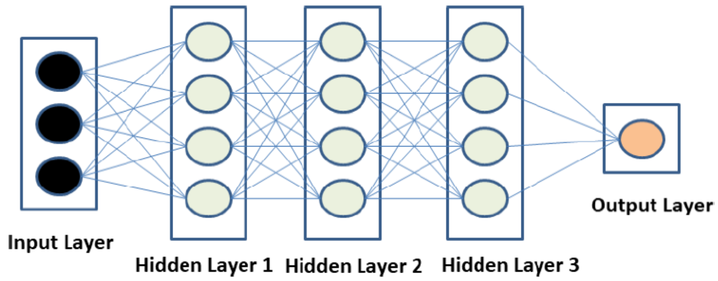

Figure 3. 16:A basic structure of MP with one input layer, three hidden layers, and one output layer [53]

3.6.2.2. Multilayer Perceptron

Multilayer perceptron (MP) consists of an input layer, some hidden layers, and an output layer. During physical construction in MP, the output of neurons of a layer becomes the input of a neuron of the next layer and these neurons are connected in such a way that neurons of one layer are the inputs of neurons of another layer. This network is also an example of Feed forward Neural Networks. It is observed that each layer is connected to the next layer and these connections do not form any closed cycle. However, in some cases, the output of neurons can be the input of neurons of the same layer.

3.7. Evaluation metrices

Evaluation metrics are used to assess the model's or algorithm's efficiency. We used four well-known measures to evaluate The Rose leaf diseases identification system detection model performance, namely, accuracy, precision, recall, and f1-score. Among the whole dataset, we used 20% of the dataset to test the accuracy of the model. The accuracy of the model has been computed using measurement metrics such as precision, recall, f1-score, and accuracy, which are described as follows [54].

- recall

The recall is the ratio of

$$Recall = \frac{TP}{TP+FN} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.3)$$

Where TP is the number of true positives, whereas the number of false negatives is FN. The recall refers to the classifier's capacity to locate all positive samples.

- Precision

Precision is the ratio of

$$Precision = \frac{TP}{TP+FP} \text{ .... (3.4)}$$

Where TP is the number of true positives and FP is the number of false positives. The capacity of the classifier to not categorize as positive a sample that is negative is referred to as precision.

- f1-score

The f1-score is an improvement criterion that's often used to fine-tune the threshold in binary decision-making, and it's measured in terms of precision and recall. Which is interpreted as follows:

$$\mathrm{f1-Score} = \frac{2*(\mathrm{Recall+Precision})}{\mathrm{Recall+Precision}} \text{ .... (3.5)}$$

A weighted average is used to measure the overall accuracy, recall, and F1-score of a classifier or model. The weighted average is determined by adding all of the scores from each class, multiplying them by the number of measurements, and dividing the result by the total number of measurements.

- Accuracy

The ratio of accurately anticipated observations to total observations is known as accuracy. It is the most frequent indicator of classifier efficiency. It is a rate defined for accurate data forecasts and can be calculated as follows:

$$\mathrm{Accuracy} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \text{ .... (3.6)}$$

The ratio of TN/(TN+FP) is defined as TN, which is a true negative. The precision, recall, f1- score, and support for each class are displayed in the classification report () function.

CHAPTER FOUR

Result And Discussion

4.1. Overview

In this chapter, the experimental evaluation of the proposed model for Image-Based Rose Leaf Diseases Detection Using Deep Learning classification has been discussed. In this thesis work, we have used a total sample of 4342 Image of Rose Leaf for fresh, Black spot and Downy Mildy classes. We have created a feature vector in the form of CSV by concatenating the texture feature for the classifiers. We have used SVM, KNN and CNN for classification and we have compared each classifier algorithm.

4.2. Data set

The dataset consists of images of rose leaves collected from The Ethio Agri CEEFT PLC Holeta Flower Farm in the Oromia region, Ethiopia. It includes a total of 4342 images, with 1430 images for "Fresh," 1434 images for "Black Spot," and 1478 images for "Downy Mildy." The dataset was augmented to 945 images, and additional images were sourced from a public dataset. The dataset is used for Image-Based Rose Leaf Diseases Detection using Deep Learning and includes classes for fresh, black spot, and downy mildy.

From the dataset consists of the following:

- Fresh: 1430 images
- Black Spot: 1434 images
- Downy Mildy: 1478 images

After splitting the data into training and testing sets with an 80/20 ratio, we have:

- Training Data: 80% (3473 images)
- Testing Data: 20% (869 images)

4.2.1. Features in CSV format

After features are extracted, we can create a CSV file with some image features.

	0	0.1	1.567608833	0.2	15.27379704	4.823367596	0.3	5.524809837	5.699455261	13.34026718	...
0	0.0	5.284828	0.000000	0.0	16.482744	5.367146	0.0	3.120102	9.356409	13.108521	...
1	0.0	0.000000	1.953162	0.0	18.007290	9.620459	0.0	1.167924	8.486197	14.742878	...
2	0.0	0.000000	4.533111	0.0	15.870173	5.385690	0.0	0.000000	8.255919	12.489344	...
3	0.0	1.757668	2.026471	0.0	16.903875	6.750482	0.0	6.470347	6.239218	16.666819	...
4	0.0	3.731833	5.907321	0.0	17.056330	2.378255	0.0	0.000000	9.201713	16.624939	...

Figure 4. 1:CSV feature vectors extracted from the image (5 rows × 4312 columns of the data frame)

- ✓ This format provides a comprehensive overview of the final images used for classification per class and the proportion used for training and testing.

4.3. SVM classifier for CNN feature extraction

As we see in the data set above, an 80/20 split of the dataset was made into training and testing sets. Accordingly, 20% of the data was used for testing and the remaining 80% was used for training. CNN was used to extract features from the pre-processed image, and the SVM classifier was given the features that were extracted. and we have got an accuracy of 80.32% accuracy without parameter tuning.

	precision	recall	f1-score	support
Fresh	0.77	0.73	0.75	287
Black spot	0.74	0.74	0.74	296
Downy Mildiw	0.91	0.95	0.93	286
accuracy			0.81	869
macro avg	0.81	0.81	0.81	869
weighted avg	0.80	0.81	0.81	869

Figure 4. 2: Evaluation metrics for the SVM model with CNN features

Precision, recall, and f1 score are basic evaluation metrics that are used to get a fine-grained idea of how well the classification algorithm is doing with our dataset. Precision shows how the classifier classifies the audio utterance to its desired class. as we have seen from the above picture the precision for class Fresh is 77% which indicates that the false positive is 33%, which means those images Fresh were not fresh, but the classifier classifies them as fresh. There is a 37% false negative which implies that 37% of the total fresh are predicted as not fresh being them fresh image. For black spot class, the precision is 74% percent which indicates that 26% of the classified black spot was not black spot but they are classified as black spot image. The classification shows that 26% of actual black spot are predicted with the classifier as not black spot. For the case of Downy Mildy image, the SVM classifier attains 91% precision which indicates that 9% of the classified image were not actually under Downy Mildy image but SVM are predicted as Downy Mildy as well 5% of the classified image which were under Downy Mildy category has been predicted as not Downy Mildy.

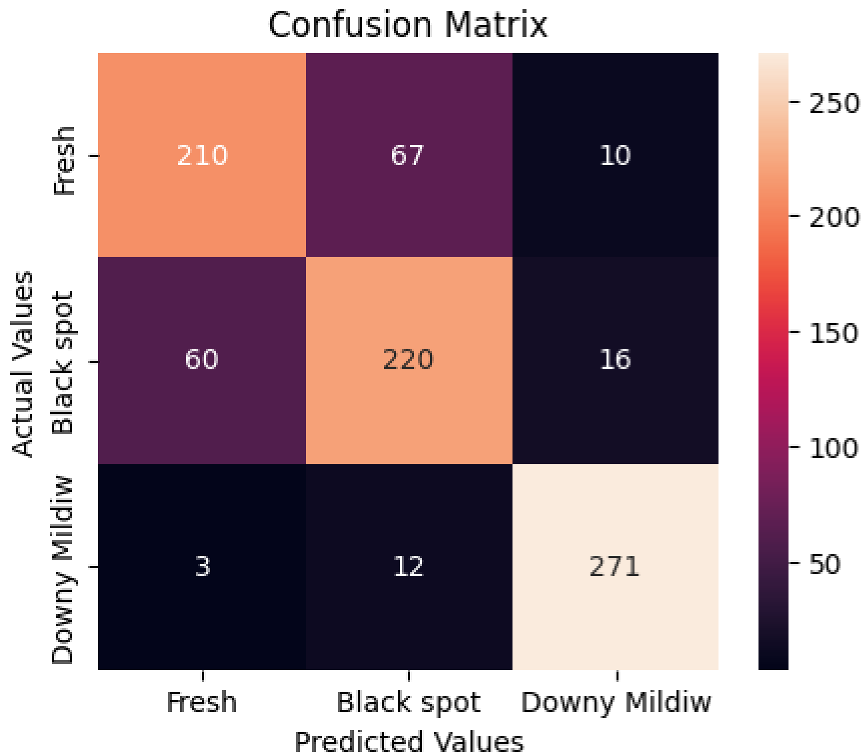

Figure 4. 3:Confusion matrix for SVM classifier

As clearly shown from the confusion matrix 210 (73.171%) of 287 are correctly classified as Fresh, 23.344% are classified wrongly as Blak spot and 3.484% are classified wrongly as Downy mildew.

In case of Black spot 220(74.32%) of 296 are classified correctly, 60 (20.27%) of actual Blak spot image are classified wrongly as Fresh,16 (5.405%) are classified wrongly as downy mildew. In the case of downy mildew 271(94.75%) of 286 are classified correctly as downy mildew 3(1.049%) of the Downy mildew are classified wrongly as fresh,12(4.195%) are classified wrongly as Black spot.

4.3.1. Why Black Spot and Fresh Rose Leaves Confuse SVM Models

The reason black spot and fresh leaves are frequently confusing for the SVM model could be due to the similarity in visual features between the two. Black spot and fresh leaves may have similar color tones, textures, and patterns, which can make it challenging for the SVM model to accurately distinguish between them. Additionally, the presence of shadows, lighting variations, and other environmental factors may further contribute to the confusion.

SVM is a linear classifier that works by finding the optimal hyperplane that separates different classes. If the visual features of black spot and fresh leaves are not linearly separable in the feature space, it can lead to misclassification and confusion between the two classes.

Furthermore, the SVM model's performance may be affected by the quality and diversity of the training data. If the training dataset does not sufficiently capture the variations and nuances in the visual characteristics of black spot and fresh leaves, the model may struggle to differentiate between them accurately.

In contrast, the CNN classifier, which achieved a higher accuracy, is capable of learning hierarchical features and capturing complex patterns in the images, making it better suited to handle the visual complexities and variations between black spot and fresh leaves.

4.4. KNN classifier for CNN feature extraction

From the pre-processed image we have used CNN for feature extraction, and we have given the features extracted from the image to the KNN Classifier and we have got an accuracy of 71.23% accuracy without parameter tuning.

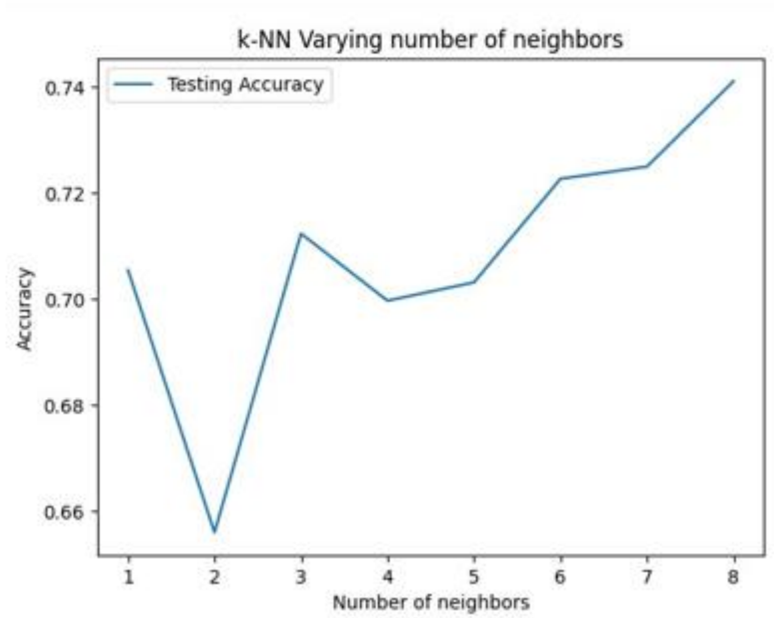

Figure 4. 4: KNN   K value versus Number of neighbors.

Figure 4.4 above shows the relationship between the K value (number of neighbors) and the performance measure (such accuracy or error rate) for the KNN algorithm (K-Nearest Neighbors). The impact of the K value selection on the KNN algorithm's performance is depicted in this graphic. It assists in figuring out the ideal value of K for the given dataset in order to get the best classification or prediction accuracy.

As we have seen from the above figure the KNN algorithm has been evaluated with the Neighboring range of 1 to 9, its best performance is where k=8. The reason why the performance is lesser is that KNN performance degrades when the dataset number increases because it will face difficulty in calculating Euclidean distance.

	precision	recall	f1-score	support
Fresh	0.68	0.61	0.64	287
Black spot	0.61	0.66	0.63	296
Downy mildew	0.85	0.87	0.86	286
accuracy			0.71	869
macro avg	0.71	0.71	0.71	869
weighted avg	0.71	0.71	0.71	869

Figure 4. 5:KNN classifier evaluation metrics

As we have seen from the above picture the precision for class Fresh is 68% which indicates that the false positive for the Fresh image is 32 %, which means those images Fresh were not fresh, but the classifier classifies them as Fresh. There is a 39 % false negative which implies that 39 % of the total Fresh image are classified as not Fresh being them Fresh.

For black spot class, the precision is 61% percent which indicates that 39 % of the classified black spot was not black spot but they are classified as black spot image. The classification shows that 34% of actual Black spot are predicted with the classifier as not black spot. For the case of Downy Mildy image, the KNN classifier attains 85% precision which indicates that 15 % of the classified image were not actually under Downy Mildy image but KNN are predicted as Downy Mildy as well 13 % of the classified image which were actually under Downy Mildy category has been predicted as not Downy Mildy.

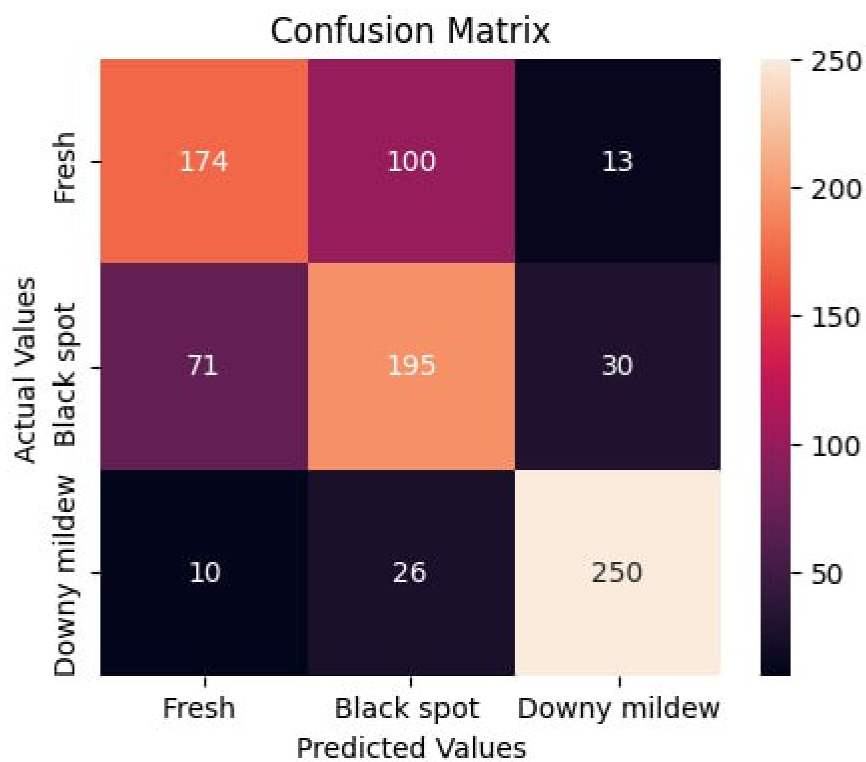

Figure 4. 6: KNN confusion matrix result

As clearly shown from the confusion matrix 174 (60.627%) of 287 are correctly classified as Fresh,34.843% are classified wrongly as Blak spot and 4.529% are classified wrongly as Downy mildew.

In case of Black spot 195(65.878%) of 296 are classified correctly,71 (23.986%) of actual Blak spot image are classified wrongly as Fresh,30 (10.135%) are classified wrongly as Downy mildew. In the case of Downy mildew 250(87.412%) of 286 are classified correctly as Downy mildew 10(3.496%) of the Downy mildew are classified wrongly as fresh,26(9.09%) are classified wrongly as Black spot.

4.4.1. Why Black Spot and Fresh Rose Leaves Confuse KNN Models

The reason the KNN model is frequently confused by black spot and fresh leaves could be due to the nature of the KNN algorithm and the visual similarity between the two classes. KNN makes predictions based on the majority class of its k-nearest neighbors in the feature space. If the feature space for black spot and fresh leaves overlaps significantly, it can lead to confusion and misclassification.

The visual features of black spot and fresh leaves, such as color, texture, and shape, may not have distinct boundaries in the feature space, making it difficult for the KNN model to accurately separate them. Additionally, if the training dataset does not adequately represent the variations and complexities in the visual characteristics of black spot and fresh leaves, the KNN model may struggle to differentiate between them effectively.

Furthermore, KNN is sensitive to noisy and irrelevant features, and if the feature space contains irrelevant or noisy visual features that are not informative for distinguishing between black spot and fresh leaves, it can lead to misclassification and confusion.

In contrast to KNN, the CNN classifier, with its ability to learn hierarchical features and capture complex patterns in images, is better equipped to handle the visual complexities and variations between black spot and fresh leaves, leading to higher accuracy in classification.

4.5. CNN classifier for CNN feature extraction

For image classification tasks, the CNN (Convolutional Neural Network) technique is commonly used. It is made up of different layers, including fully connected, pooling, and convolutional layers.

4.5.1. CNN Algorithm for Image classification and Model Performance Evaluation

The CNN algorithm used for image classification:

- ✓ Input Layer: The input image's raw pixel values are sent to the input layer.
- ✓ Convolutional Layer: This layer uses a series of filters to extract features from the input image. A convolution operation is carried out by each filter to generate feature maps.
- ✓ Activation Function: To add non-linearity to the network, a non-linear activation function such as ReLU (Rectified Linear Unit) is usually applied to the feature maps.

- ✓ Pooling Layer: The most significant information is retained in the feature maps while the spatial dimensions are decreased by the pooling layer. Max pooling and average pooling are two common pooling methods.
- ✓ Fully Connected Layer: After several convolutional and pooling layers, the feature maps are flattened and passed to one or more fully connected layers. These layers perform classification based on the extracted features.
- ✓ Output Layer: The output layer produces the final classification predictions, usually using a SoftMax activation function for multi-class classification.
- ✓ Training: The CNN is trained using a labeled dataset through a process called backpropagation, where the model learns to minimize the difference between its predictions and the true labels.
- ✓ Optimization: An optimization technique, like the Adam optimizer, is used to optimize the model's parameters by modifying the network's weights to minimize the loss function.
- ✓ A dataset that has been divided into training and testing sets is used to train the CNN algorithm, and measures like training accuracy and validation accuracy are used to assess the model's performance.

```
Epoch 33/40
55/55 [==============================] - 1s 10ms/step - loss: 0.0567 - accuracy: 0.9798 - val_loss: 0.8843 - val_accuracy: 0.7917
Epoch 34/40
55/55 [==============================] - 1s 10ms/step - loss: 0.0537 - accuracy: 0.9813 - val_loss: 0.9259 - val_accuracy: 0.7745
Epoch 35/40
55/55 [==============================] - 1s 10ms/step - loss: 0.0444 - accuracy: 0.9842 - val_loss: 0.9742 - val_accuracy: 0.7791
Epoch 36/40
55/55 [==============================] - 1s 10ms/step - loss: 0.0539 - accuracy: 0.9816 - val_loss: 0.9510 - val_accuracy: 0.7802
Epoch 37/40
55/55 [==============================] - 1s 10ms/step - loss: 0.0644 - accuracy: 0.9767 - val_loss: 1.0266 - val_accuracy: 0.7802
Epoch 38/40
55/55 [==============================] - 1s 10ms/step - loss: 0.0716 - accuracy: 0.9770 - val_loss: 0.9559 - val_accuracy: 0.7906
Epoch 39/40
55/55 [==============================] - 1s 10ms/step - loss: 0.0441 - accuracy: 0.9842 - val_loss: 0.9776 - val_accuracy: 0.7860
Epoch 40/40
55/55 [==============================] - 1s 10ms/step - loss: 0.0471 - accuracy: 0.9862 - val_loss: 0.9320 - val_accuracy: 0.7906
```

Figure 4. 7: CNN layer classification algorithm (The iteration of the above experiment)

We have trained the model with a dataset proportion of 80% percent training and 20% testing using a learning rate of 0.001 with Adam optimizer finally the CNN classifier achieved 98% training accuracy and 93% validation accuracy. Thus, CNN achieves high performance when compared with SVM and KNN.

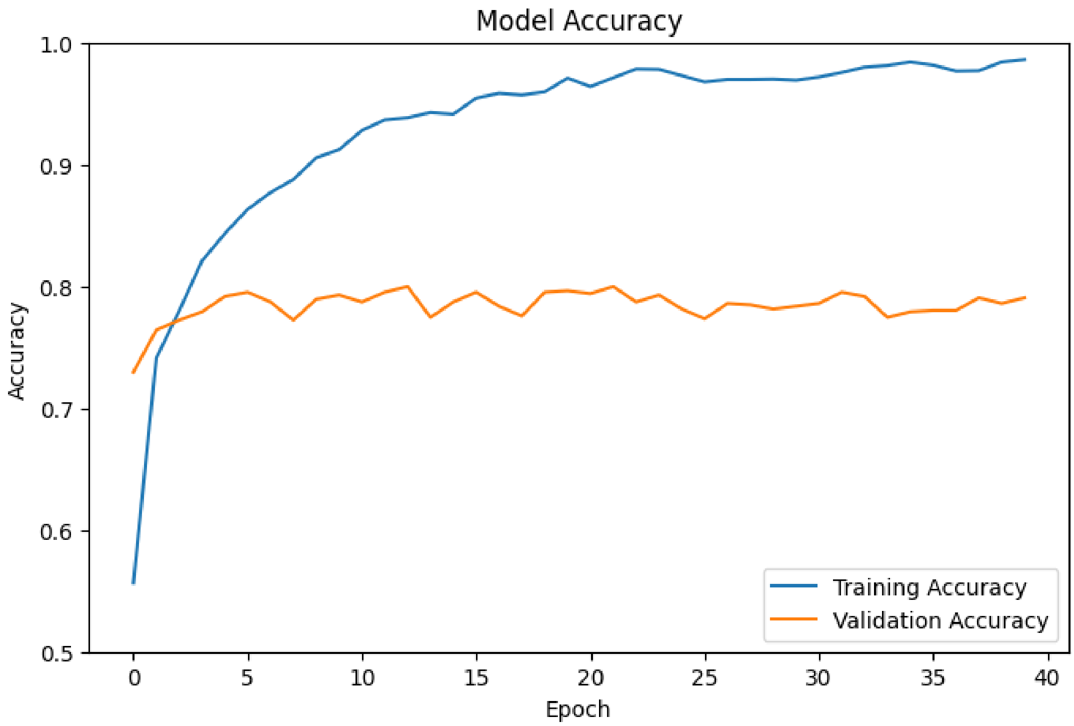

Figure 4. 8: Model accuracy CNN Classifier for CNN feature extraction (Training, testing accuracy, loss, and validation accuracy)

The graph of the model accuracy for the CNN classifier with CNN feature extraction shows that the validation accuracy starts off higher than the training accuracy, crossing initially at epoch 0 with an accuracy of 78. As the training progresses through the epochs, the training accuracy continues to increase, reaching 98% accuracy, while the validation accuracy also increases but at a slower rate, reaching 93% accuracy.

As clearly shown from the above graph, the initial crossing of the training and validation accuracy at epoch 0 with an accuracy of 78 indicates that the model is initially overfitting to the training data. The training accuracy of the model gets better as the number of epochs goes up, but the validation accuracy reaches a point of no return, suggesting that the model is overfitting to the training set.

Furthermore, a low validation accuracy shows that the model is not generalizing to new data as well as it could be, while a high training accuracy shows that the model has learned the training data well. Factors like model complexity, dataset size, and data quality could be to blame for this.

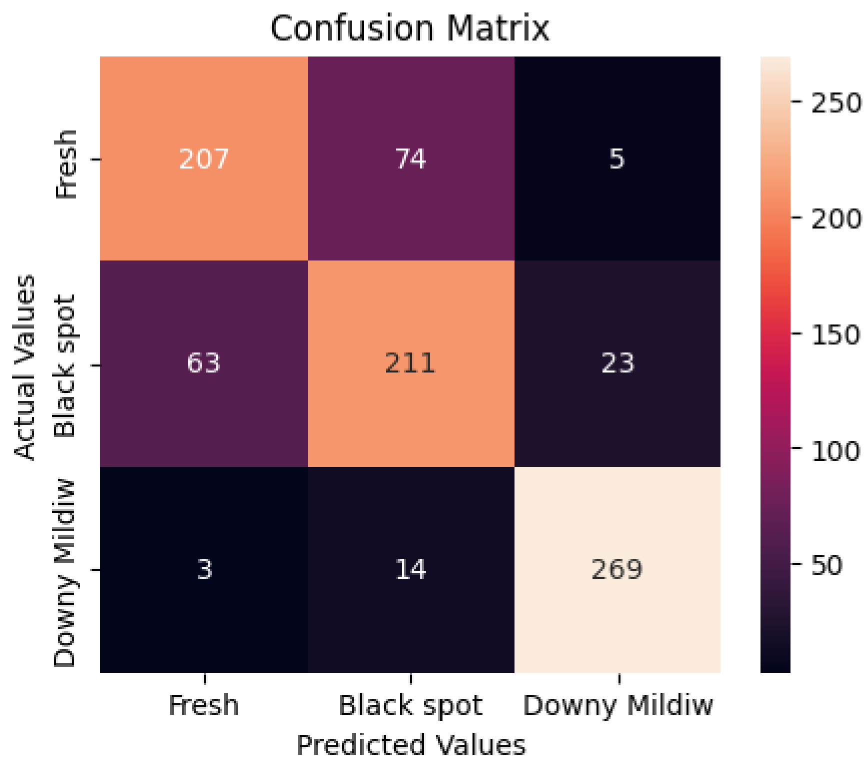

Figure 4.8: CNN confusion matrix result

As clearly shown from the confusion matrix 207 (72.37%) of 286 are correctly classified as Fresh,25.874% are classified wrongly as Blak spot and 1.748% are classified wrongly as Downy mildew.

In case of Black spot 211(71.04%) of 297 are classified correctly,63 (21.212%) of actual Blak spot image are classified wrongly as Fresh,23 (7.744%) are classified wrongly as Downy mildew. In the case of Downy mildew 269(94.055%) of 286 are classified correctly as Downy mildew 3(1.048%) of the Downy mildew are classified wrongly as fresh,14(4.895%) is classified wrongly as Black spot.

4.6. Model Comparison

This section illustrates the model we have used throughout our research as well as determines which model achieves a better result. As we have stated in the previous sections, we have done 3 experiments to get which model achieves better performance.

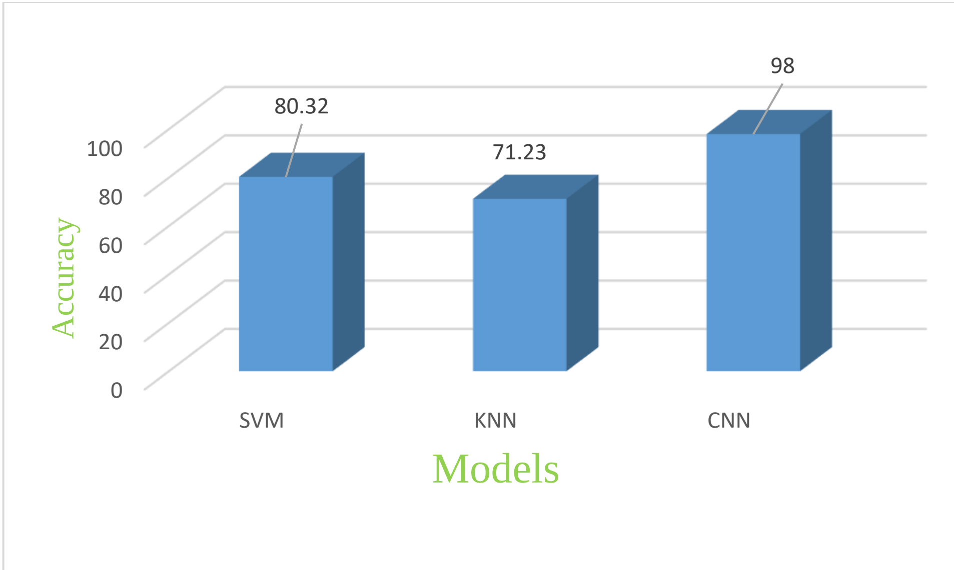

Figure 4. 9:Performance analysis of the model

As we have seen from the chart CNN model is greater in accuracy than SVM model and KNN. Also, SVM model is greater in accuracy than KNN model. Because KNN uses Euclidean distance when perform classification. As the data set increases the model faced problem in calculating the Euclidean distance. So, performance of KNN is highly decreasing as compared to SVM models because the dimensionality of the feature increases that challenge the KNN to calculate the Euclidean distance.

4.7. Discussion of result

The research questions are addressed to a significant extent, and the findings of the study provide valuable insights into the domain of rose leaf disease classification. The study successfully demonstrates the application of image processing techniques for the detection and classification of rose leaf diseases, leveraging deep learning algorithms such as Support Vector Machine (SVM), K-Nearest Neighbor Classifier (KNN), and convolutional neural network (CNN).

The experimental results show that the CNN model outperforms the other classification algorithms, achieving an impressive accuracy of 98%. This indicates the effectiveness of using deep learning for the accurate detection of rose leaf diseases.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

Image-based rose leaf diseases detection using deep learning involves training a deep learning model to analyze images of rose leaf's and identify signs of diseases such as Black spot, Downy mildew. This process can help in early detection and management of plant diseases, ultimately contributing to improved agricultural productivity and the economy. Deep learning algorithms are trained using a large dataset of images showing healthy and diseased rose leafs. for these we have used a total of 4342 images, with 1430 images for "Fresh," 1434 images for "Black Spot," and 1478 images for "Downy Mildy." The dataset was augmented to 945 images. The model learns to recognize patterns and features associated with different diseases, enabling it to accurately classify images.

Here we have used CNN features extraction and we have used different classifier algorithms SVM, KNN, and CNN. Using the KNN classifier we have achieved an accuracy of 71.23%, and SVM has achieved 80.32% accuracy and CNN achieved 98% accuracy.so we conclude that the CNN classifier is better suited to the model we have developed.

In feature engineering for image-based disease detection, the steps involve collecting a large dataset of rose leaf images, preprocessing the data, extracting features using CNN, and training and evaluating the model with different classifiers. The process includes data cleaning, resizing, normalization, and dataset augmentation. Feature extraction is performed using CNN to capture patterns and features from the images, and the model's performance is evaluated using SVM, KNN, and CNN classifiers. The CNN classifier is found to be the most suitable for the developed model, achieving 98% accuracy.

The convolutional layer in CNN creates feature maps by applying filters to the input image and looking for patterns and features. To decrease computational complexity and increase the robustness of the learnt features, the pooling layer shrinks the spatial dimensions of the feature maps while maintaining pertinent information. To extract hierarchical patterns and characteristics from the input images, both layers are essential, enabling CNN to learn discriminative representations for disease classification.

Generally, Image-based rose leaf diseases detection using deep learning is a promising approach to early detection and management of plant diseases. The study successfully developed a deep learning model that utilizes Convolutional Neural Networks (CNN) to achieve high accuracy in classifying images of healthy and diseased rose leaves. The CNN classifier outperforms other traditional machine learning algorithms, such as Support Vector Machines (SVM) and K-Nearest Neighbors (KNN). The study's strengths lie in its innovative application of deep learning to plant disease detection, its large dataset of rose leaf images, and its comprehensive evaluation of different classifier algorithms. However, the study acknowledges limitations in dataset availability and the need for further experimentation with noisy images before the model can be widely applied in real-world settings.

5.2. Recommendation

Future work should focus on addressing the limitations identified in this study. This includes expanding the dataset to include a wider range of rose plant species and images with varying degrees of noise. Additionally, exploring image segmentation techniques and developing a user-friendly interface could further enhance the model's practicality.

There is still a gap for expanding the proposed work. As a result, when doing this research, the following are some noteworthy future work recommendations:

- ❖ One of the challenges in this study is getting the necessary data for experimentation. We recommend the preparation of standardized data set for further research in agriculture problem domain.
- ❖ Even if it registers high accuracy in this study, the CNN model needs an extensive experiment to see its effectiveness in a variety and noisy images of rose plant images.
- ❖ Use segmentation techniques which are appropriate for the color image of the Rose leaf to investigate its effect on the performance of the model.
- ❖ There are different plants that are easily affected by different diseases. We suggest more research needs to be conducted to construct a model that can detect plant diseases.
- ❖ Design a user interface to enable users to interact with the system and input images for analysis, view the predict results either Healthy or Diseased.
- ❖ We suggest considering the Environmental Factors Influencing the Visual Appearance of Rose Leaf Diseases.

Reference

[1] C. Luo, Y. Hao, and Z. Tong, "Research on Digital Image Processing Technology and Its Application," in 2018 8th International Conference on Management, Education and Information (MEICI 2018), 2018, pp. 587-592.

[2] S. Gupta, S. Mahajan, and A. K. Pandit, "A Review On Image Processing Techniques," in 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), 2020, pp. 20-24.

[3] J. M. Alston and P. G. Pardey, "Agriculture in the global economy," Journal of Economic Perspectives, vol. 28, pp. 121-146, 2014.

[4] D. Whittlesey, "Major agricultural regions of the earth," Annals of the Association of American Geographers, vol. 26, pp. 199-240, 1936.

[5] J. Eunice, D. E. Popescu, M. K. Chowdary, and J. Hemanth, "Deep learning-based leaf disease detection in crops using images for agricultural applications," Agronomy, vol. 12, p. 2395, 2022.

[6] M. Sollai, "How to Feed an Empire? Agrarian Science, Indigenous Farming, and Wheat Autarky in Italian-Occupied Ethiopia, 1937–1941," Agricultural History, vol. 96, pp. 379-416, 2022.

[7] T. G. Abebe, M. R. Tamtam, A. A. Abebe, K. A. Abtemariam, T. G. Shigut, Y. A. Dejen, et al., "Growing use and impacts of chemical fertilizers and assessing alternative organic fertilizer sources in Ethiopia," Applied and Environmental Soil Science, vol. 2022, pp. 1-14, 2022.

[8] G. Shibre, W. Mekonnen, and D. Haile Mariam, "Decomposition analysis of women's empowerment-based inequalities in the use of maternal health care services in Ethiopia: Evidence from Demographic and Health Surveys," Plos one, vol. 18, p. e0285024, 2023.

[9] B. T. Mengistie, A. P. Mol, and P. Oosterveer, "Governance of agro-pesticide through private environmental and social standards in the global cut flower chain from Ethiopia," Ambio, vol. 46, pp. 797-811, 2017.

[10] S. K. Ashinie and T. T. Tefera, "Horticultural crops research and development in Ethiopia: Review on current status," Journal of Biology, Agriculture and Healthcare, vol. 9, pp. 1-14, 2019.

[11] O. Ardasheva, E. Cheremnykh, T. Lekontseva, and D. Zorin, "Comprehensive assessment of varieties of garden roses in the conditions of the Udmurt Republic," in IOP Conference Series: Earth and Environmental Science, 2023, p. 012026.

[12] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, et al., "Deep learning vs. traditional computer vision," in Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1 1, 2020, pp. 128-144.

[13] L. C. Ngugi, M. Abdelwahab, and M. Abo-Zahhad, "A new approach to learning and recognizing leaf diseases from individual lesions using convolutional neural networks," Information Processing in Agriculture, vol. 10, pp. 11-27, 2023.

[14] S. Alqethami, B. Almtanni, W. Alzhrani, and M. Alghamdi, "Disease detection in apple leaves using image processing techniques," Engineering, Technology & Applied Science Research, vol. 12, pp. 8335-8341, 2022.

[15] J. A. Ruth, R. Uma, A. Meenakshi, and P. Ramkumar, "Meta-heuristic based deep learning model for leaf diseases detection," Neural Processing Letters, vol. 54, pp. 5693-5709, 2022.

[16] S. Nuanmeesri, "A hybrid deep learning and optimized machine learning approach for rose leaf disease classification," Engineering, Technology & Applied Science Research, vol. 11, pp. 7678-7683, 2021.

[17] S. Asgari Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: a review," Artificial Intelligence Review, vol. 54, pp. 137-178, 2021.

[18] K. R. Castleman, Digital image processing: Prentice Hall Press, 1996.

[19] I. Pitas, Digital image processing algorithms and applications: John Wiley & Sons, 2000.

[20] J. D. Kelleher, Deep learning: MIT press, 2019.

[21] P. P. Shinde and S. Shah, "A review of machine learning and deep learning applications," in 2018 Fourth international conference on computing communication control and automation (ICCUBEA), 2018, pp. 1-6.

[22] S. Verma, A. Chug, A. P. Singh, S. Sharma, and P. Rajvanshi, "Deep learning-based mobile application for plant disease diagnosis: A proof of concept with a case study on tomato plant," in Applications of image processing and soft computing systems in agriculture, ed: IGI global, 2019, pp. 242-271.

[23] G. Batchuluun, S. H. Nam, and K. R. Park, "Deep learning-based plant classification and crop disease classification by thermal camera," Journal of King Saud University-Computer and Information Sciences, vol. 34, pp. 10474-10486, 2022.

[24] H. F. Pardede, E. Suryawati, D. Krisnandi, R. S. Yuwana, and V. Zilvan, "Machine learning based plant diseases detection: A review," in 2020 International conference on radar, antenna, microwave, electronics, and telecommunications (ICRAMET), 2020, pp. 212-217.

[25] Y. Li, J. Xue, K. Wang, M. Zhang, and Z. Li, "Surface defect detection of fresh-cut cauliflowers based on convolutional neural network with transfer learning," Foods, vol. 11, p. 2915, 2022.

[26] S. Poornam and A. F. S. Devaraj, "Image based Plant leaf disease detection using Deep learning," Inernational journal of computer communication and informatics, vol. 3, pp. 53-65, 2021.

[27] H. Zogan, I. Razzak, X. Wang, S. Jameel, and G. Xu, "Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media," World Wide Web, vol. 25, pp. 281-304, 2022.

[28] V. Sawarkar and S. Kawathekar, "A review: Rose plant disease detection using image processing," IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN, pp. 2278-0661, 2018.

[29] B. K. Ekka and B. S. Behera, "Disease Detection in Plant Leaf Using Image Processing Technique," Int. J. Progress. Res. Sci. Eng, vol. 1, pp. 151-155, 2020.

[30] M. Halder, A. Sarkar, and H. Bahar, "Plant disease detection by image pro-cessing: a literature review," image, vol. 1, 2019.

[31] C. Salgado-Salazar, N. Shiskoff, M. Daughtrey, C. L. Palmer, and J. A. Crouch, "Downy mildew: a serious disease threat to rose health worldwide," Plant disease, vol. 102, pp. 1873-1882, 2018.

[32] R. Thanki and P. Joshi, "Image Processing and Its Applications," in Advanced Technologies for Industrial Applications, ed: Springer, 2023, pp. 33-48.

[33] M. Goyal, R. Goyal, P. Venkatappa Reddy, and B. Lall, "Activation functions," Deep learning: Algorithms and applications, pp. 1-30, 2020.

[34] J. Šťastný and M. Minařík, "A Brief Introduction to Image Pre-Processing for Object Recognition," 2007.

[35] D. I. Patrício and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review," Computers and electronics in agriculture, vol. 153, pp. 69-81, 2018.

[36] S. S. Sann, S. S. Win, and Z. M. Thant, "An analysis of various image pre-processing techniques in butterfly image," International Journal for Advance Research and Development, vol. 6, pp. 1-4, 2021.

[37] T. Azizov, A. Azizova, and S. Al Ghadban, "Construction and calculation of reinforced concrete overlap with a high spatial work effect," International Journal of Engineering & Technology, vol. 7, pp. 567-574, 2018.

[38] M. K. JASIM, R. Najm, E. Kanan, H. Alfaar, and M. Otair, "Image Noise Removal Techniques: A Comparative Analysis," International Journal of Science and Applied Information Technology, pp. 24-29, 2019.

[39] A. Shah, J. I. Bangash, A. W. Khan, I. Ahmed, A. Khan, A. Khan, et al., "Comparative analysis of median filter and its variants for removal of impulse noise from gray scale images," Journal of King Saud University-Computer and Information Sciences, vol. 34, pp. 505-519, 2022.

[40] D. Kaur and Y. Kaur, "Various image segmentation techniques: a review," International Journal of Computer Science and Mobile Computing, vol. 3, pp. 809-814, 2014.

[41] A. O. Salau and S. Jain, "Feature extraction: a survey of the types, techniques, applications," in 2019 international conference on signal processing and communication (ICSC), 2019, pp. 158-164.

[42] M. Jogin, M. Madhulika, G. Divya, R. Meghana, and S. Apoorva, "Feature extraction using convolution neural networks (CNN) and deep learning," in 2018 3rd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT), 2018, pp. 2319-2323.

[43] V. K. Vishnoi, K. Kumar, and B. Kumar, "A comprehensive study of feature extraction techniques for plant leaf disease detection," Multimedia Tools and Applications, pp. 1-53, 2022.

[44] H. Xue, Q. Yang, and S. Chen, "SVM: Support vector machines," The top ten algorithms in data mining, vol. 6, pp. 37-60, 2009.

[45] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, pp. 273-297, 1995.

[46] S. S. Harakannanavar, J. M. Rudagi, V. I. Puranikmath, A. Siddiqua, and R. Pramodhini, "Plant leaf disease detection using computer vision and machine learning algorithms," Global Transitions Proceedings, vol. 3, pp. 305-310, 2022.

[47] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in 2017 international conference on engineering and technology (ICET), 2017, pp. 1-6.

[48] Y. Liu, H. Pu, and D.-W. Sun, "Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices," Trends in Food Science & Technology, vol. 113, pp. 193-204, 2021.

[49] H. Gholamalinezhad and H. Khosravi, "Pooling methods in deep neural networks, a review," arXiv preprint arXiv:2009.07485, 2020.

[50] Y. Bengio, "Learning deep architectures for AI," Foundations and trends® in Machine Learning, vol. 2, pp. 1-127, 2009.

[51] K. Shrivastava, S. Kumar, and D. K. Jain, "An effective approach for emotion detection in multimedia text data using sequence based convolutional neural network," Multimedia tools and applications, vol. 78, pp. 29607-29639, 2019.

[52] X. Qi, T. Wang, and J. Liu, "Comparison of support vector machine and softmax classifiers in computer vision," in 2017 Second International Conference on Mechanical, Control and Computer Engineering (ICMCCE), 2017, pp. 151-155.

[53] S. Cheon, H. Lee, C. O. Kim, and S. H. Lee, "Convolutional neural network for wafer surface defect classification and the detection of unknown defect class," IEEE Transactions on Semiconductor Manufacturing, vol. 32, pp. 163-170, 2019.

[54] T. Kynkäänniemi, T. Karras, S. Laine, J. Lehtinen, and T. Aila, "Improved precision and recall metric for assessing generative models," Advances in Neural Information Processing Systems, vol. 32, 2019.

APPENDIX

1.Code for SVM classifier

```
# SVM classifier

X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.2,random_state=42, stratify=y)
from sklearn.preprocessing import StandardScaler
# While the standard scaler has some options,
# those are rarely used and we usually instantiate it with the default parameters:
scaler = StandardScaler()
scaler.fit(X_train)
X_train_scaled = scaler.transform(X_train)
scalerTest= StandardScaler()
scalerTest.fit(X_train)
X_test_scale=scalerTest.transform(X_test)
#perform  SVM
# importing SVM module
from sklearn.svm import SVC
# kernel to be set radial bf
classifier1 = SVC(kernel='rbf',decision_function_shape='ovo')
# traininf the model
classifier1.fit(X_train_scaled, y_train)
# testing the model
y_pred = classifier1.predict(X_test_scale)
# importing accuracy score
from sklearn.metrics import accuracy_score
# printing the accuracy of the model
print(accuracy_score(y_test, y_pred))
```

2.Code for KNN classifier

```
#import KNeighborsClassifier
from sklearn.neighbors import KNeighborsClassifier

#Setup arrays to store training and test accuracies
neighbors = np.arange(1,9)
train_accuracy =np.empty(len(neighbors))
test_accuracy = np.empty(len(neighbors))

for i,k in enumerate(neighbors):
    #Setup a knn classifier with k neighbors
    knn = KNeighborsClassifier(n_neighbors=k)

    #Fit the model
    knn.fit(X_train_scaled, y_train)

    #Compute accuracy on the training set
    train_accuracy[i] = knn.score(X_train_scaled, y_train)

    #Compute accuracy on the test set
    test_accuracy[i] = knn.score(X_test_scale, y_test)
```