**A** **REPORT**

**ON**

# Crop Disease Detection Using Machine Learning and Deep Learning: A Comprehensive Analysis

## **By**

**Name of the students** **Registration No.**

1) Poojitha Ramireddygari AP21110011296

2) Khyathi Devi Kotipalli AP21110011327

3) Hema Poojitha Chandu AP21110011331

4) Vyshnavi Gayam AP21110011347

**Prepared in the partial fulfillment of the**

Undergraduate Research Opportunities Project (UROP)



|  |
| --- |
|  |

**AT**

**SRM University, AP**

**Neerukonda, Mangalagiri, Guntur**

**Andhra Pradesh – 522 240**

**Under the Guidance of Dr. Murali Krishna**

**[November 2023]**

**Certificate**

Date: 25-11-2023

# This is to certify that the work present in this Project entitled “Crop disease detection using machine learning and deep learning: a comprehensive analysis” has been carried out by [‘Poojitha Ramireddygari’, ‘Khyathi Devi Kotipalli’, ‘Hema Poojitha Chandu’, ‘Vyshnavi Gayam’] under my supervision. The work is genuine, original, and suitable for submission to the SRM University – AP for the award of Bachelor of Technology in School of Engineering and Sciences.

**Supervisor**

Dr. Murali Krishna Enduri,

Assistant Professor.

**Acknowledgements**

I would like to express my profound gratitude to Dr Niraj Upadhayaya, of CSE department, and Dr.Murali Krishna Enduri for their contributions to the completion of my project titled “**Crop disease detection using machine learning and deep learning: a comprehensive analysis**”.

I would like to express my special thanks to our mentor Dr.Murali Krishna Enduri for his time and efforts, he provided throughout the year. Your useful advice and suggestions were helpful to me during the project’s completion. In this aspect, I am eternally grateful to you.

We would like to take this opportunity to express my gratitude to all of our group members. The project would not have been successful without their cooperation and inputs.

**Poojitha Ramireddygari**

**Khyathi Devi Kotipalli**

**Hema Poojitha Chandu**

**Vyshnavi Gayam**

# Table of Contents

Abstract 5

Abbreviation 6

List of Figures 7

List of Tables 8

Introduction 9

1.1significance of plant disease detection 9

1.2 usage of report 9

1.3 Overview or Background 9

1.4 Problem Statement 9

1.5 What are we trying to solve 9

1.6 Why are we trying to solve 9

1.7 How we collected the data 10

1.8 How are we going to solve 10

1.9 Research goals and objectives 10

1.10 What knowledge patterns are we observing 10

1.11 What insights do we gain 10

1.12 What do we give back to society 11

Methodology 11

2.1 Literature review 13

2.2 Architecture 14

2.3 Implementation 15

Discussion 17

Results 20

Conclusion 31

Future Works 31

References 32

**ABSTRACT**

Agriculture provides livelihood for nearly two and half billion of the world’s population. It employs around 58 percent Indians making it the highest employment sector in India. Despite having highest employment rate, India’s agricultural sector has low crop yields than global average. This is due to many factors like unlikely rains, excessive use of pesticides and fertilizers, and diseases, etc. Pests and diseases cause over Rs 290 billion per annum losses of crops in India. Crop diseases can have a notable impact on crop productivity leading to loss for farmers. This is a worldwide problem. Early detection of the diseases is crucial to prevent crop damage. Mostly, detection of these diseases is done manually, which is time taking and may not be accurate.

Embracing automatic crop disease detection becomes imperative for identifying diseases in their early stages efficiently. Integration of technologies in agriculture help farmers overcome various challenges. Using machine learning and deep learning to detect crop diseases can assist farmers to keep a close eye on their crops as they grow, ensuring healthier plants and better yields. Our main objective is to employ machine learning models for crop disease detection. We have used popular Plant Village and Plant Pathology datasets, consisting images of different crops and Random Forest, CNN and SVM algorithms are implemented for classification. The results obtained are promising in detecting the crop diseases.

**Abbreviations**

* CNN – Convolutional Neural Network
* ML – Machine Learning
* DL – Deep Learning
* SVM – Support Vector Machine
* ReLu – Rectified Linear Unit
* RBF – Radial Basis Function
* RGB – Red, Green, Blue
* ROC - receiver operating characteristic curve
* TP – True Positive
* TN – True Negative
* FP – False Positive
* FN – False Negative
* IoT – Internet of Things
* AUC – Area Under Curve

**List of Figures**

* Fig.1 Flow chart for SVM+CNN Model
* Fig 2. Class distribution of different datasets
* Fig 3. Sample images of Plant Pathology Dataset
* Fig 4. Sample images of Apple Dataset
* Fig 5. Sample images of Potato Dataset
* Fig 6. Sample images of Corn Dataset
* Fig 7. Sample images of Grape Dataset
* Fig 8. Sample images of Tomato Dataset
* Fig 9. Object Detection and Segmentation of an image
* Fig 10. Comparison of the accuracies of different algorithms for different datasets.
* Fig 11. Confusion matrix of Plant Pathology dataset
* Fig 12. Confusion matrix of Apple dataset
* Fig 13. Confusion matrix of Potato dataset
* Fig 14. Confusion matrix of Corn dataset
* Fig 15. Confusion matrix of Grape dataset
* Fig 16. Confusion matrix of Tomato dataset

**List of Tables**

Table 1. Details of the dataset

Table 2. Performance metrics for all datasets and models

Table 3. Accuracy and loss graphs of different datasets

Table 4. ROC curves of different datasets in CNN+SVM model

1. **Introduction**

**1.1 Significance of Plant Disease Detection**

The impact of agriculture on human life and the economies of nations worldwide is self-evident. With a significant portion of the global population relying on agriculture for their livelihoods, maintaining the health and productivity of crops is paramount. Plant diseases pose a constant threat, and if left unaddressed, they can lead to severe declines in crop yields. This not only disrupts the food supply chain but also results in food insecurity, economic instability, and environmental consequences [[1]](#_Alexandre_Meybeck,_Vincent). So, detecting plant disease is one of the major areas to focus.

**1.2 Usage of Report**

This report delves into the realm of plant disease detection, focusing on innovative technologies such as computer vision, machine learning, deep learning, and extensive agricultural datasets. The objective is to enhance the capability of crop disease detection systems, utilizing cutting-edge tools to mitigate the impact of diseases on crop yields.

**1.3 Overview or Background**

Traditionally, plant diseases have been identified through time-consuming visual examinations conducted by seasoned agricultural specialists. However, the complexities of modern agriculture render this approach impractical. In response, the integration of data analysis and agricultural science has gained prominence in recent years [[2]](#_Albert_Khakimov,_Alisher) [[3]](#_Role_of_Modern).

**1.4 Problem Statement**

The traditional methods of disease detection in plants are proving inadequate in the face of evolving agricultural challenges. This research addresses the need for more efficient and accurate methods by leveraging advanced technologies. By utilizing ML and deep learning techniques, we tried to improve the yields of the crop.

**1.5 What are we trying to solve**

The central focus of this research is to enhance crop disease detection systems through the integration of computer vision, machine learning, and extensive agricultural datasets. This aims to improve the accuracy and speed of disease identification, thus minimizing crop damage.

**1.6 Why are we trying to solve**

In the context of the Fourth Industrial Revolution (4IR), the adoption of intelligent systems, the Internet of Things (IoT), and artificial intelligence (AI) in agriculture is crucial. Early disease detection, transitioning from manual to automated processes, is pivotal for sustainable and productive agriculture [[4]](The#_Jehoon_Sung,_).

**1.7 How we collected the data**

Datasets utilized in this research were sourced from reputable repositories, including Plant Village [[5]](Plant_Village_Dataset#_) and Plant Pathology [[6]](Plant_Pathology_2020-fgvc7#_). These datasets serve as a foundation for the development and training of machine learning models.

**1.8 How are we going to solve**

The methodology involves the application of deep learning algorithms, particularly convolutional neural networks (CNN), to advance the field of plant disease detection. This builds upon the work of previous research, enhancing the capabilities of automated detection systems [[7]](#_Shujuan_Zhang,_Bin).

**1.9 Research goals and objectives**

Our primary goal is to revolutionize agriculture by improving the accuracy and efficiency of crop disease detection. Specific objectives include implementing computer vision techniques, leveraging machine learning algorithms, and contributing to the body of knowledge in this evolving field.

**1.10 What knowledge patterns are we observing**

Observations will center around patterns and anomalies identified in crop images. Through machine learning, the system aims to recognize distinctive features indicative of various diseases, contributing to a growing knowledge base.

**1.11 What insights do we gain**

The insights gained from this research contribute to a deeper understanding of the intersection between advanced technologies and agriculture. By automating disease detection, we aim to provide farmers with timely information to protect their crops.

**1.12 What do we give back to society**

The shift towards automated and data-driven crop disease detection has the potential to revolutionize agriculture, ensuring global food security. This research contributes to the development of sustainable practices and tools that empower farmers to safeguard their crops and livelihoods.

1. **Methodology**

Machine Learning and Deep learning models can be employed in classification of crop images collected from various datasets. The models that we have used in our research are CNN, SVM and Random Forest.

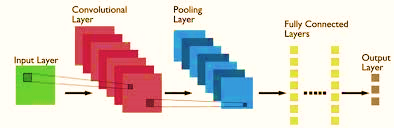
**a) CNN:**

CNN or ConvNet is designed for working with grid-like data, like images. Each part of the CNN works on a certain area of the image. A typical CNN consists of three main layers: a layer for convolution, another for pooling, and a layer for complete connections.

**Convolution Layers:** These layers undertake extensive computational tasks. It achieves this by executing a dot product between a set of learnable parameters (kernel) and a confined segment of the receptive field. This kernel, which is smaller than the image, possesses greater depth.

**Pooling Layers:** The network's complexity is decreased by using these layers. During forward propagation, blocks of size k×k are minimized to a single value. Eventually, the error associated with this singular value is calculated through backward propagation from the previous layer and relayed back to its source. As this value originates from a single location within the k×k block, errors from pooling layers exhibit a notably sparse nature.

**Dense Layers:** It is a fully connected layer, is commonly used in final phases of the neural network. These are deeply interconnected with the preceding layer, with each neuron linked to every neuron in the anterior layer. In this layer, the neurons execute matrix-vector multiplication, taking input from every neuron of the prior layer. This multiplication follows a simple rule: the row vector's columns must match the column vector's length.

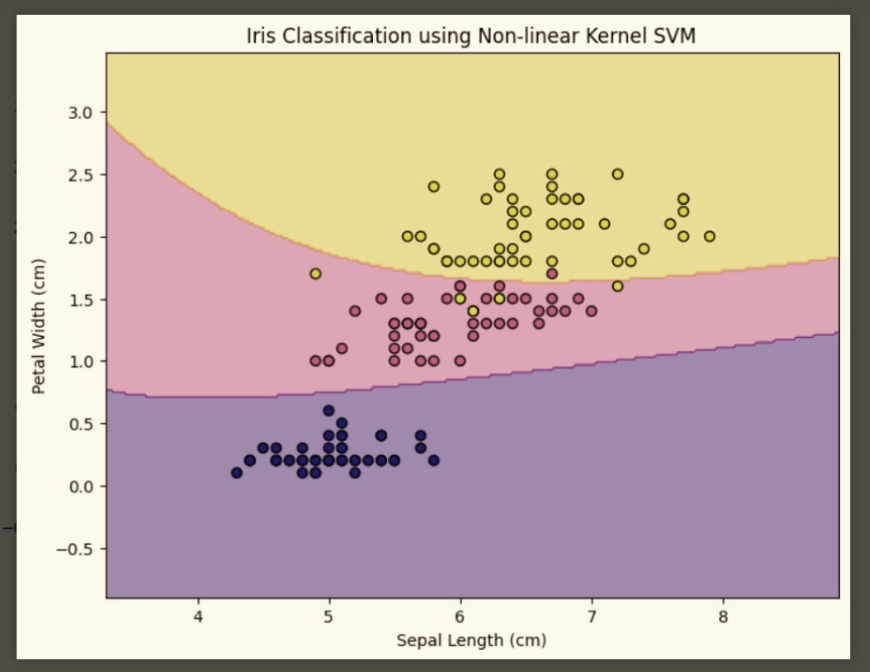


**b) SVM:**

It is one of the popular supervised learning algorithms that can be used for classification and regression. SVM finds an optimal hyperplane that separates the classes from one another. We choose support vectors which are data points nearer to hyperplane from both the classes. We try to widen the distance between hyperplane and support vectors, which is called margin to get better performance of SVM.

SVM’s hyperplane can separate only linearly separable data. For making non linearly separable data to linearly separable data we convert the original data to higher dimensional space, like φ(v)=v 2 and adding this dimension to our feature space, the classes become linearly separable.

The kernel trick avoids the computational burden caused by adding extra features in higher dimensional feature space. It operates through pairwise similarities using a kernel function, allowing efficient dot product computations without transforming data into higher dimensions. This method helps in handling complex datasets and simplifies model training.



**c) Random Forest:**

It belongs to a category of algorithms for supervised learning, builds an ensemble of decision trees referred to as a "forest," commonly trained using the bagging technique. The underlying concept of the bagging method is that the combination of multiple learning models amplifies the overall predictive performance. It introduces added randomness during tree growth by selecting features from a random subset for node splitting. This enhances the overall model performance.

Gini Index: It is a measure of impurity or the degree of non-uniformity in a dataset. It is used as criterion for splitting the data during the construction of decision tree.

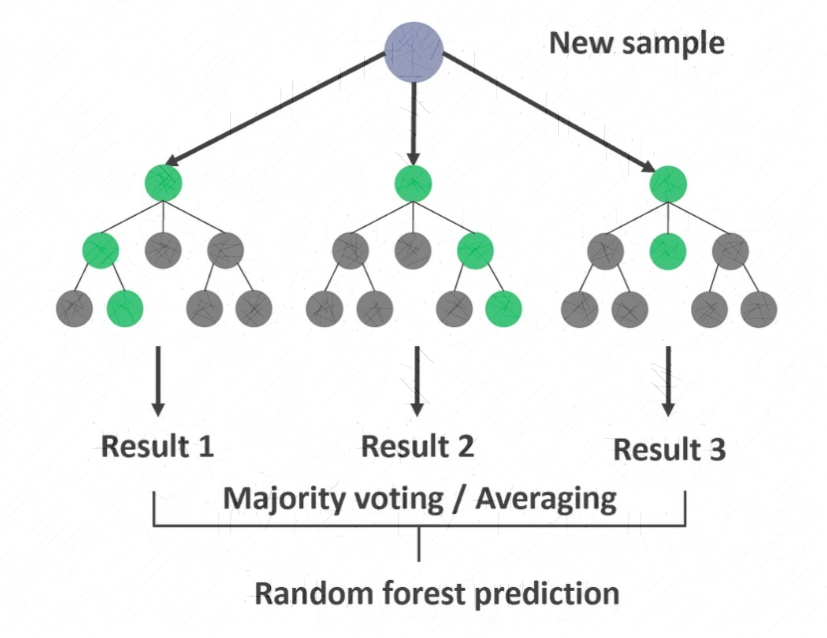
Gini Index=1-∑(ji=1) (prob)2 = 1-[(prob+)2+(prob-)2]

prob+ = probability of a positive class,

prob- = probability of a negative class

A Gini index 0 represents that dataset is perfectly classified, 0.5 implies that the classes are uniformly distributed. Our Goal is to minimize Gini index while constructing decision tree.

Entropy is also used to measure the impurity of split.

Entropy(S)= -prob (+) log prob (+) - prob (-) log prob (-)

* 1. **Literature Review:**

Previous studies on crop disease detection have shown that a variety of image processing methods and classification algorithms work well. One such study conducted by [[8]](#_Niveditha_M,_Ramachandra) Shima Ramesh et Al. in 2018 used papaya leaves dataset with 160 images and implemented various Machine learning models. Random Forest gave highest accuracy of 70.14%. They have done feature extraction with the help of histogram of oriented gradients (HOG), which is a feature descriptor used for object detection. Hu Moments, Haralick texture, Color Histogram are also utilized. Accuracy can be raised by using larger dataset and using features like SIFT, SUTF, etc.

In 2016, Srdjan Sladojevic et Al. utilized a diverse set of images from various sources [[9]](#_Srdjan_Sladojevic,_Marko). Data augmentation was effectively applied to the dataset to enhance its diversity and reduce overfitting. By introducing slight distortions through techniques like affine transformation, perspective transformation, and image rotations, the model's ability to generalize was improved. The deep CNN architecture chosen for this study, based on the Caffe framework, featured a modified CaffeNet architecture with multiple convolutional and pooling layers. The use of ReLU activations and dropout layers was effective in addressing overfitting, and the model demonstrated a remarkable overall accuracy of 96.3% after fine-tuning.

Pranesh Kulkarniet Al. [[10]](#_Pranesh_Kulkarni,_Atharva) performed pre-processing on image dataset from Plant Village which is a crucial step to ensure precise results in classification problems. Using grey level co-occurrence matrix (GLCM) they have extracted features like color, shape, and texture and accomplished task of selecting features. For classification they have implemented Random Forest with multiple decision trees and acquired an average accuracy of 93%.

In 2022, Sunil S. Harakannanavar et al. [[11]](#_Sunil_S._Harakannanavar,jayasri) suggested a model for the detection of leaf diseases based on the integration of Image processing and Machine learning techniques. They considered tomato plant dataset from Plant Village database which consists of six disorders and performed hierarchical clustering and K means clustering to enhance image nature and facilitate segmentation of samples of leaves. Using the Principal Component Analysis (PCA), Discrete Wavelet Transform (DWT), GLCM, significant regions and the features are extracted. Subsequently, Support Vector Machine (SVM), K Nearest Neighbors (KNN), and Convolutional Neural Network (CNN) algorithms are employed to classify the extracted features and record the model's performance. They combined DWT, GLCM, PCA, CNN and got 99.09% accuracy.

Godfrey Winster Sathianesan et AI. [[12]](#_Gnanavel_Sakkarvarthi,_Godfrey) introduces a powerful deep learning technique for identifying tomato crop diseases. CNN with a hidden, a flattening, two max-pooling layers, and two convolutional layers is used in the study. The authors improved accuracy significantly by employing CNN, beating well-known models such as InceptionV3 and ResNet 152. The paper discusses crop diseases in the agriculture industry, with a focus on India in particular, highlighting the importance of finding quick solutions. They have achieved 98% accuracy for the training set and 88.17% accuracy for the testing set. It provides growers and agricultural production with an accurate and efficient tool for early disease identification in tomato crops. This researach demonstrates the possible uses of deep learning methods in agriculture by providing unique insights into the field of plant disease identification.

* 1. **Architecture:**

In our research paper, we took on the important job of detecting and categorizing plant diseases that harm different types of crops. We gathered images of different crops like apple, potato, corn, tomato, grape which were sourced from Plant Village dataset and images from Plant Pathology dataset for detecting diseases. The dataset comprises images, each uniquely associated with specific disease classes across different crops.

1. Details of the datasets

|  |  |  |
| --- | --- | --- |
| Datasets | Crop Diseases | No of Images |
| Plant Pathology | Healthy  Rust  Scab  Multiple Diseases | 1032  1244  1184  182 |
| Apple | Healthy  Scab  Black Rot  Cedar Apple Rust | 1645  630  621  275 |
| Potato | Healthy  Early Blight  Late Blight | 152  1000  1000 |
| Corn | Healthy  Common Rust  Northern Leaf Blight  Cercospora/Gray Leaf Spot | 1162  1192  985  513 |
| Grape | Healthy  Grape Black Rot  Leaf Blight  Black Measles (Esca) | 423  1180  1076  1383 |
| Tomato | Healthy  Bacterial Spot  Early Blight  Late Blight | 1273  1702  800  1527 |

* 1. **Implementation:**

1. **Data Collection and Preprocessing**

The dataset consists of images of plant leaves with labels indicating the presence of various diseases. The images are loaded using OpenCV, resized to a standardized dimension of 256x256 pixels, and converted into arrays. Data augmentation is applied to the training set using techniques such as rotation, shifting, shearing, and flipping to enhance model generalization. Labels are binarized to facilitate multi-class classification.

1. **Convolutional Neural Network (CNN) Architecture**

A CNN model is employed for feature extraction and classification. The model consists of convolutional layers with Leaky ReLU activation, batch normalization, and max-pooling, followed by fully connected layers. The last layer utilizes a softmax activation function for multi-class classification. The model is compiled using binary cross-entropy loss and the Adam optimizer with a learning rate decay.

1. **Model Training and Evaluation**

The CNN model is trained using the augmented dataset, and the training process is monitored for accuracy and loss. Learning curves are plotted to visualize the model's performance on both training and validation sets. After training, the model is evaluated on the test set to assess its generalization ability.

1. **Feature Extraction for SVM**

Features are extracted from the intermediate layer of the trained CNN model. The dense layer named 'my\_dense' is used for feature extraction, and the resulting features are used as input for a Support Vector Machine (SVM) model. The SVM is trained using a radial basis function (RBF) kernel.

1. **SVM Model Training and Evaluation**

The SVM model is trained on the features extracted from the CNN. The accuracy of the SVM model is assessed on both the training and test sets. Additionally, precision, recall, and F1-score are calculated to provide a comprehensive evaluation of the SVM model's performance.

1. **Model Comparison and Analysis**

The performance of both the CNN and SVM models is compared in terms of accuracy and other relevant metrics. The strengths and limitations of each model are discussed, providing insights into their effectiveness for plant disease classification.

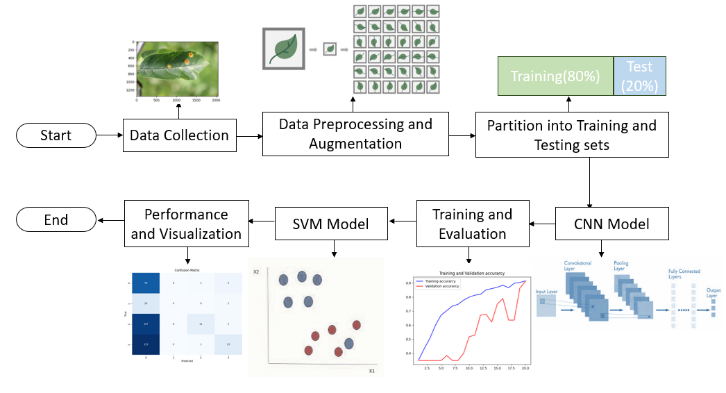


Fig.1 Flow chart for SVM+CNN Model

1. **Discussion:**

**Class Distribution of the Datasets:**

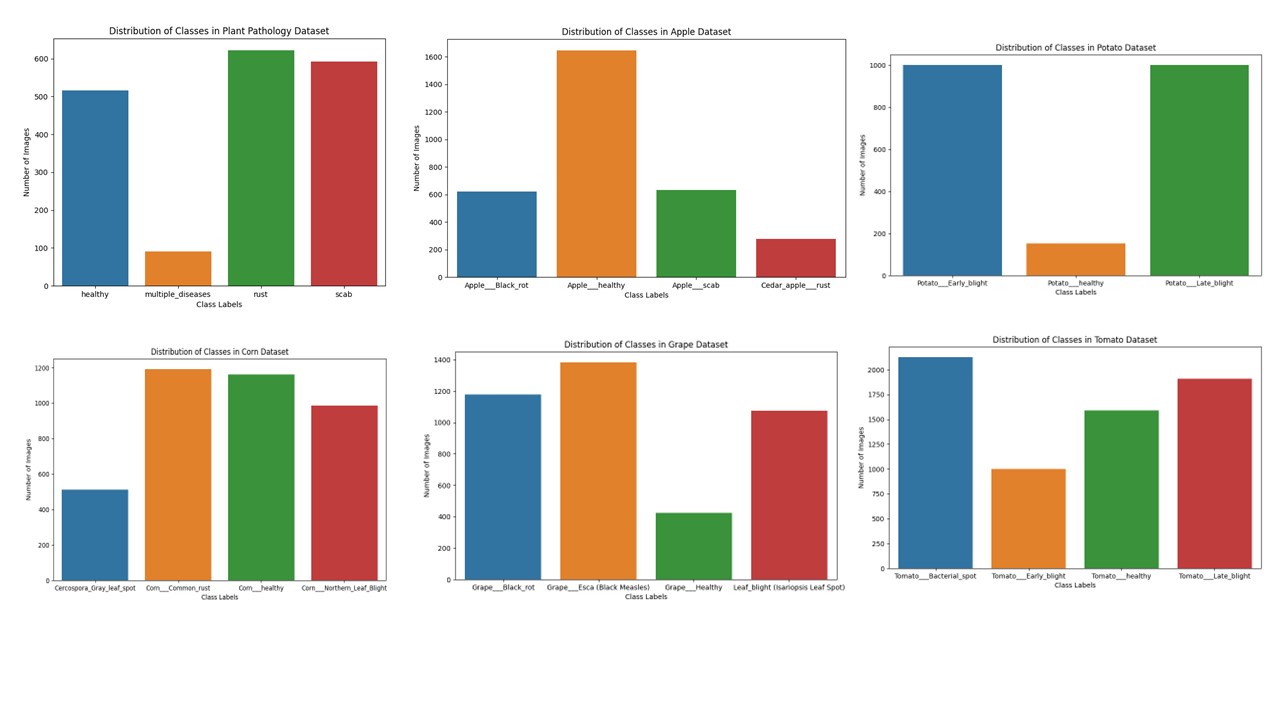
Utilizing the Matplotlib and Seaborn libraries in Python, we crafted a bar graph to shed light on the distribution of classes within our dataset. By importing the os library, we calculated the number of images in each class dynamically. Subsequently, this data was visualized using a bar graph, where the x-axis represents different class labels, and the y-axis signifies the corresponding number of images. Such visualizations are pivotal in understanding the distribution of classes, aiding in subsequent steps of our analysis and machine learning endeavors for crop disease detection.

Fig 2. Class distribution of different datasets

**Visual Exploration of Dataset Classes:**

For a visual introduction to the dataset, we utilized Matplotlib to showcase a subset of images from each class. The resulting figure is a grid of images, where each column corresponds to a different class. The title of each subplot identifies the specific class, providing an immediate visual representation of the diverse types of images in our dataset.

A.

Fig 3. Sample images of Plant Pathology Dataset

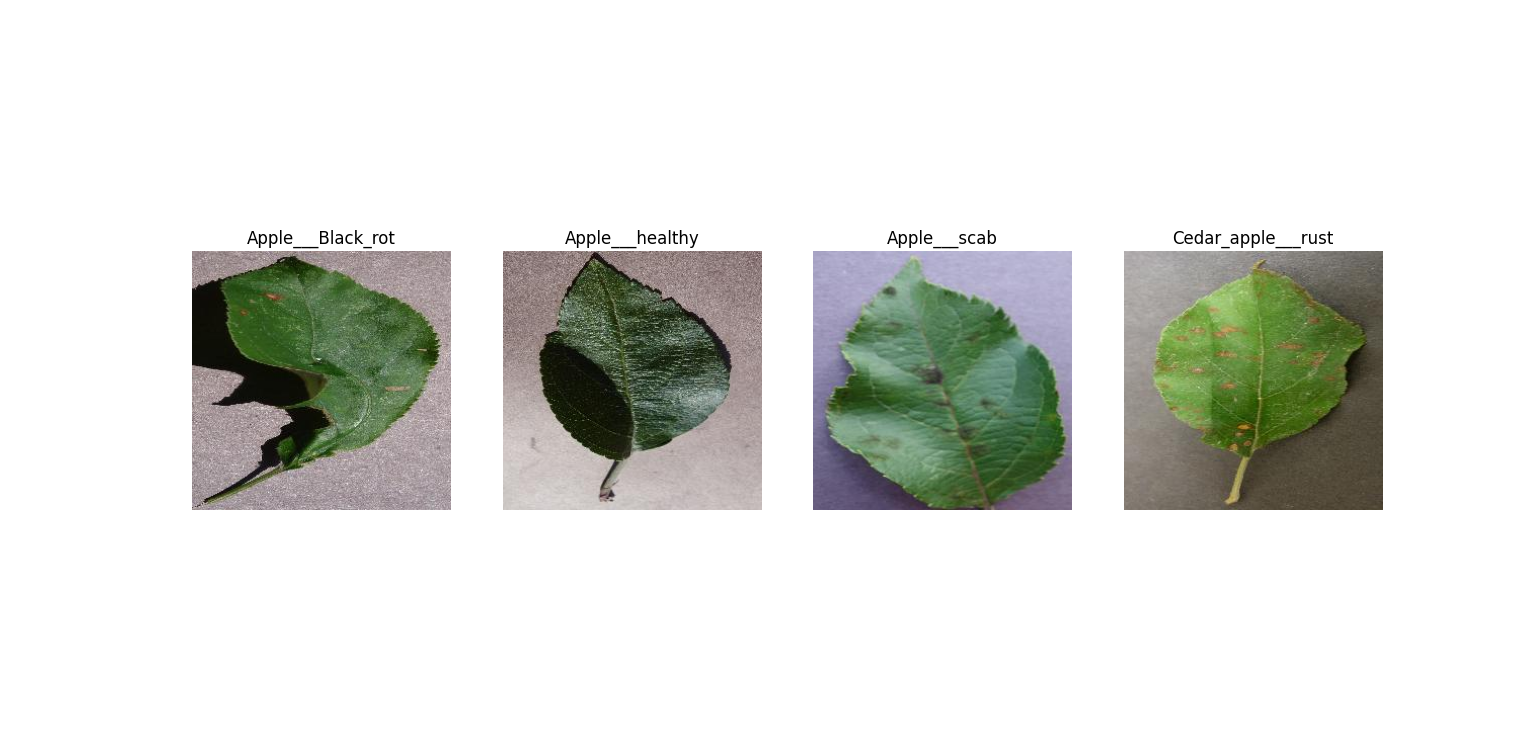
B.

Fig 4. Sample images of Apple Dataset

C.

Fig 5. Sample images of Potato Dataset

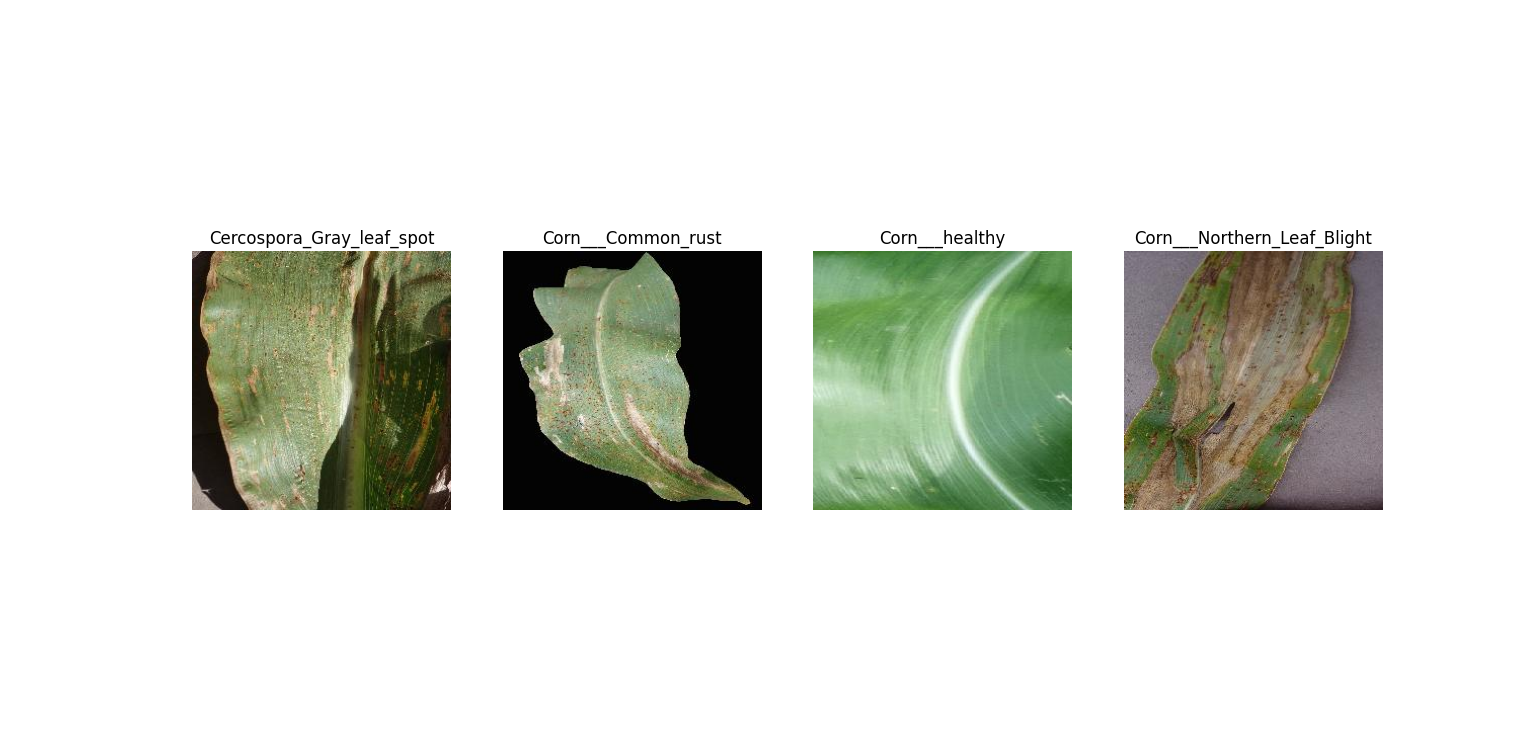
D.

Fig 6. Sample images of Corn Dataset

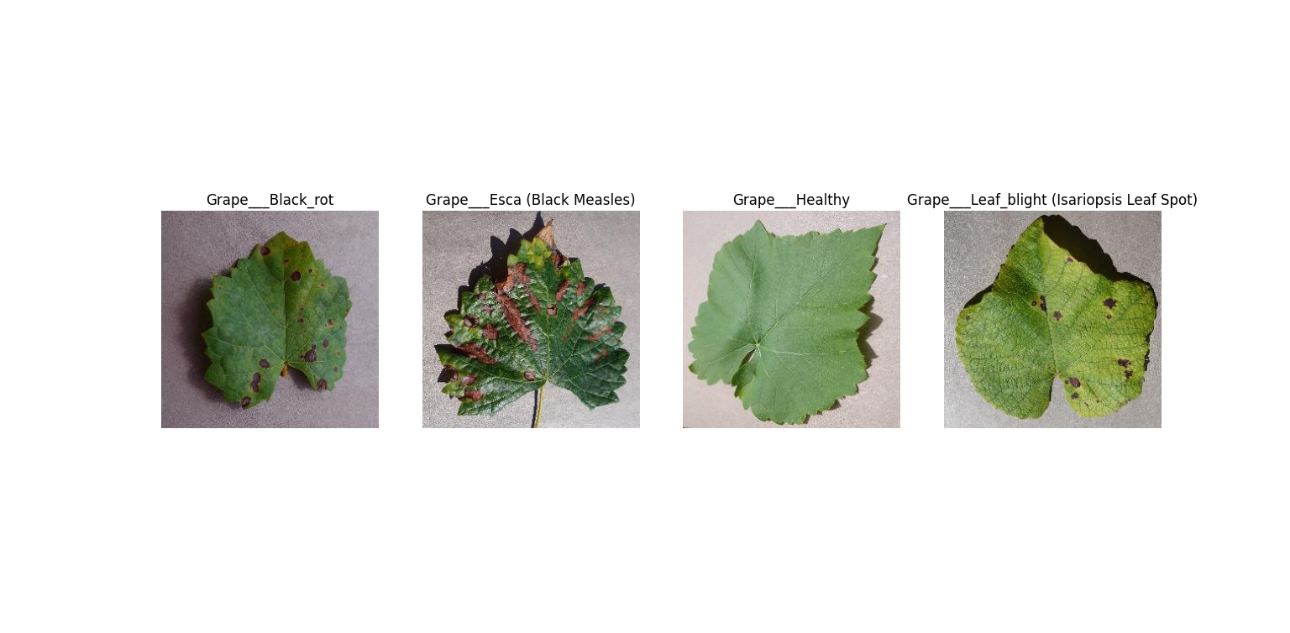
E.

Fig 7. Sample images of Grape Dataset

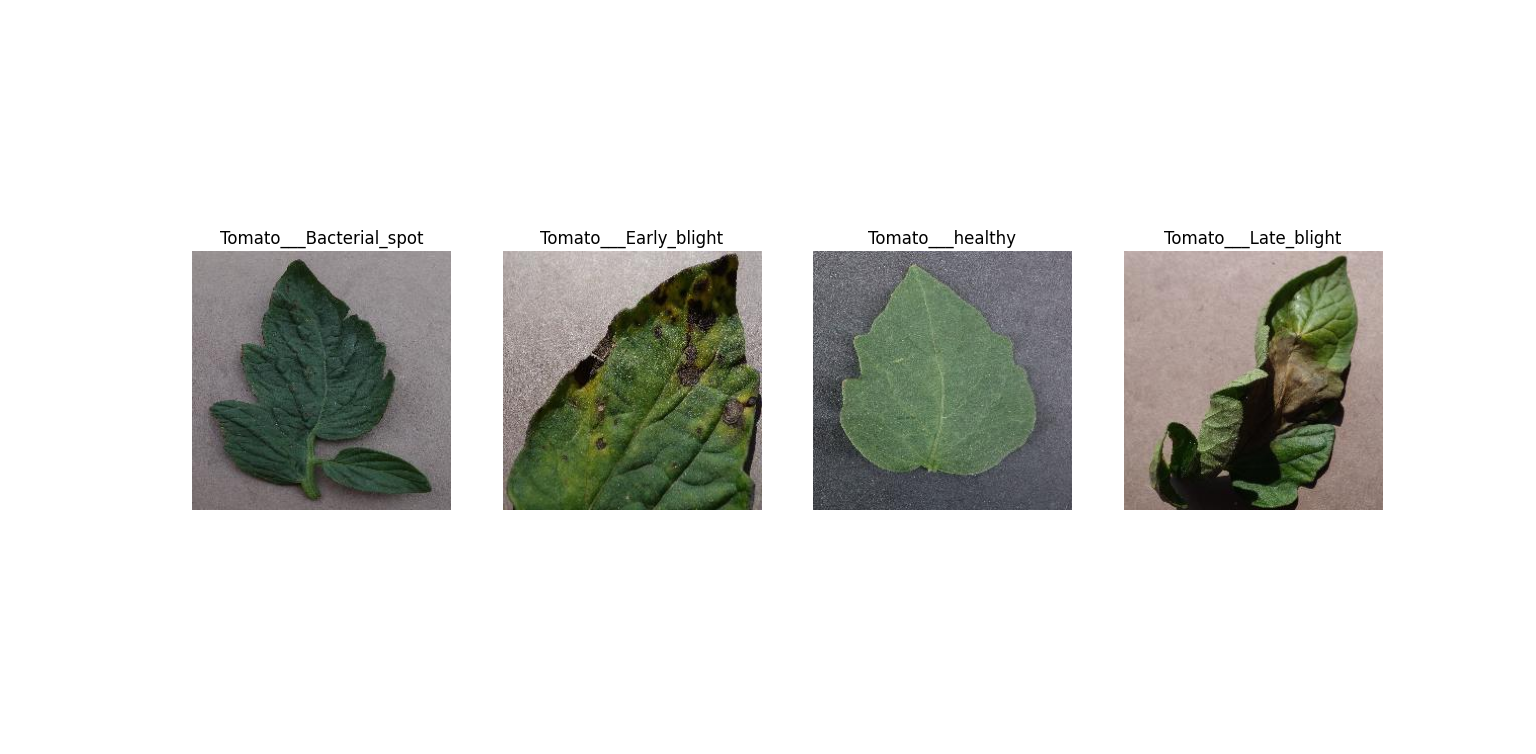
F.

Fig 8. Sample images of Tomato Dataset

**Class Imbalance:**

In our crop disease datasets, a notable observation pertains to class imbalances among different crop diseases. The distribution of images across disease categories within each crop type reveals variations in the number of instances, which can impact the training dynamics of machine learning models. For instance, in the Plant Pathology dataset, the 'Multiple Diseases' class has a substantially lower number of images (182) compared to 'Rust' (1244) and 'Scab' (1184). Similarly, in the Apple dataset, the 'Cedar Apple Rust' class exhibits a lower count (275) compared to 'Healthy' (1645) and 'Scab' (630). Class imbalances can pose challenges during model training, as the algorithm may exhibit a bias towards the majority class. Addressing these imbalances is crucial for ensuring that the model learns to generalize effectively across all disease categories, thereby enhancing its predictive capabilities. This consideration becomes especially pertinent when deciding on appropriate strategies for data augmentation, class weighting, or other techniques to mitigate the impact of class imbalances on model performance.

**Object Detection and Segmentation:**

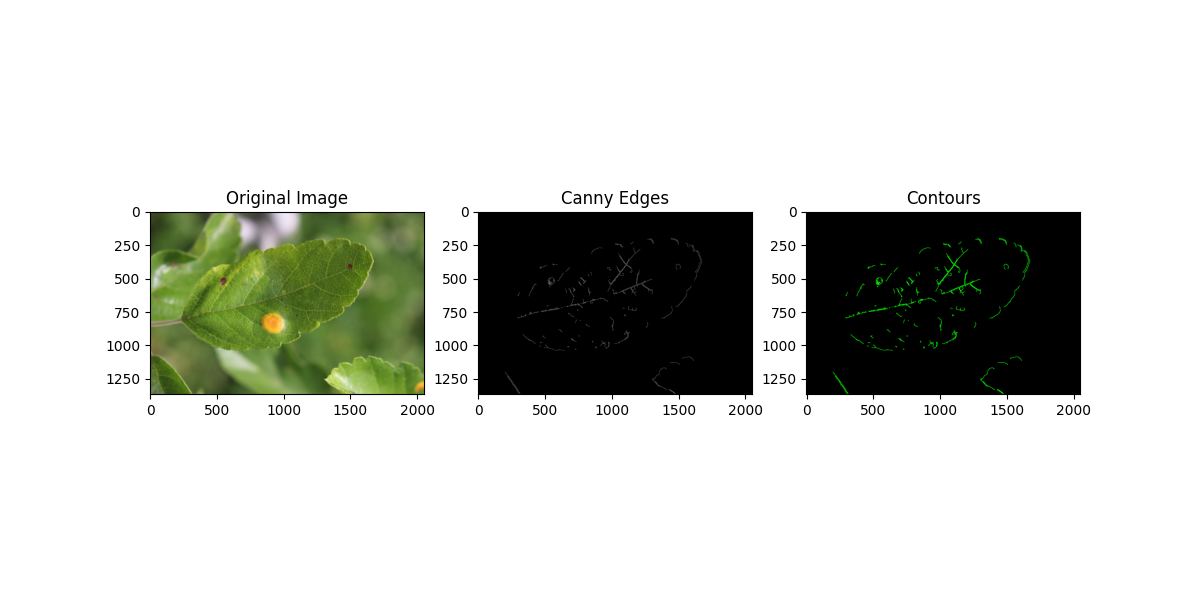
We harnessed the capabilities of computer vision through OpenCV to execute object detection and segmentation on crop images. Beginning with image loading and preprocessing, including grayscale conversion and Gaussian blurring, we then applied the Canny edge detector to identify key edges within the image. Extracting contours from these edges enabled the creation of a visual representation of distinct objects present in the image. The side-by-side comparison of the original RGB image, Canny edges, and identified contours showcased the effectiveness of our object detection and segmentation techniques, providing a valuable foundation for understanding crop structures.

Fig 9. Object Detection and Segmentation of an image

1. **Results:**

We have employed various evaluation metrics to measure the effectiveness of the deep learning and Machine Learning models classify crop diseases. To measure the effectiveness of the algorithms, we considered multiple performance metrics, each providing unique insights into the model’s capabilities to accurately predict crop diseases.

These metrics include:

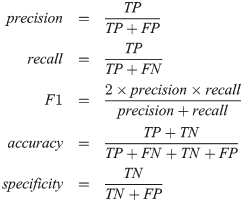
**Confusion Metrix:** A confusion matrix is a comprehensive table that vividly captures the performance of a classification model. It presents a holistic view of predicted and actual classifications, allowing us to assess the model's effectiveness in distinguishing between different classes. The matrix is structured into four quadrants: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

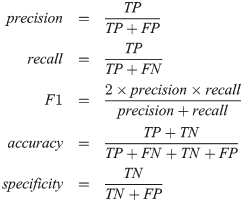
**TP** – Cases in which the model accurately identifies and predicts the presence of the positive class.

**TN** – Classes in which the model correctly identifies and predicts the absence of the negative class

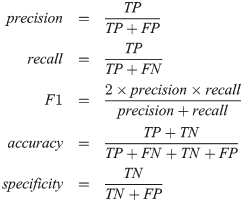
**FP** – Classes in which the model incorrectly predicts the positive class

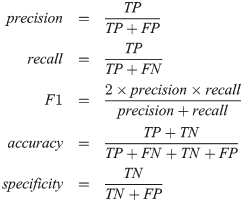
**FN** – Classes in which the model incorrectly predicts the negative class

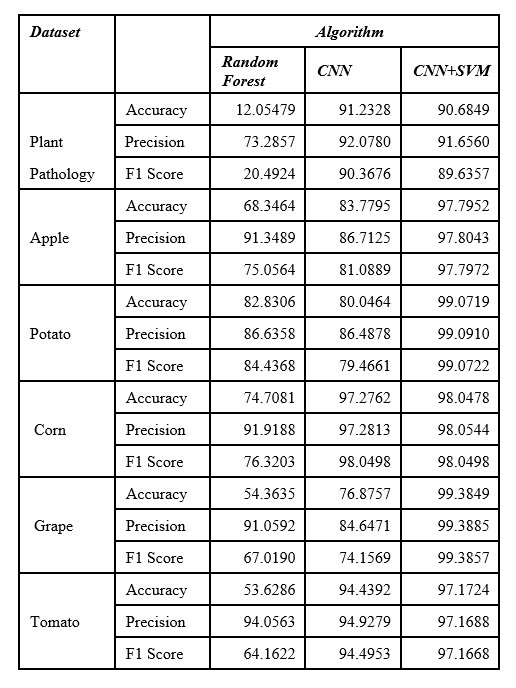
**Test Accuracy:** It quantifies the model's proficiency in correctly classifying test data, expressed as a percentage. It gauges how accurate the model's predictions are overall serving as a fundamental indicator of its general performance.

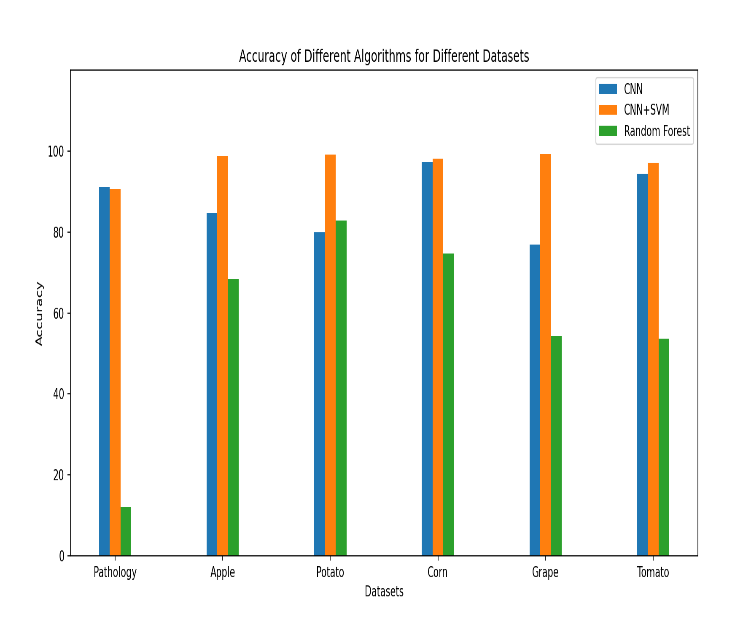
**Precision:** It assesses how well the model identified positive cases among the occurrences that it predicted to be positive. High precision reduces false positives by indicating that the model is likely to be accurate when predicting a disease.

**Recall:** Recall is another crucial metric which determines how well the model can distinguish true positive cases from all real positive cases among all actual positive cases. It helps in assessing the model's ability to avoid missing true positive cases, reducing false negatives.



**F1-Score:** The harmonic mean of recall and precision is known as the F1-Score. By considering both false positives and false negatives, it offers a fair assessment of the model's effectiveness. An F1-Score that is high suggests that a model is generally good at handling imbalanced datasets.

**Table 2. PERFORMANCE METRICS FOR ALL DATASETS AND MODELS**

****

**Fig 10. Comparison of the accuracies of different algorithms for different datasets.**

We have presented the results of our evaluation, representing the performance of each algorithm on the various datasets. We plotted the accuracy values achieved by these algorithms against their corresponding names for the different datasets. Notably, the average accuracy values obtained for CNN+SVM, CNN, Random Forest were 97.18, 87.43, 57.65 respectively. Remarkably, the CNN+SVM algorithm exhibited the highest average accuracy of 97.18 among the three methods, indicating its superior performance. The findings, depicted in Figure [Fig.10], clearly demonstrates the outstanding performance of the CNN+SVM algorithm, suggesting its potential as a highly effective tool for accurately predicting crop diseases.

**Confusion Matrices of CNN+SVM Model for different Datasets:**

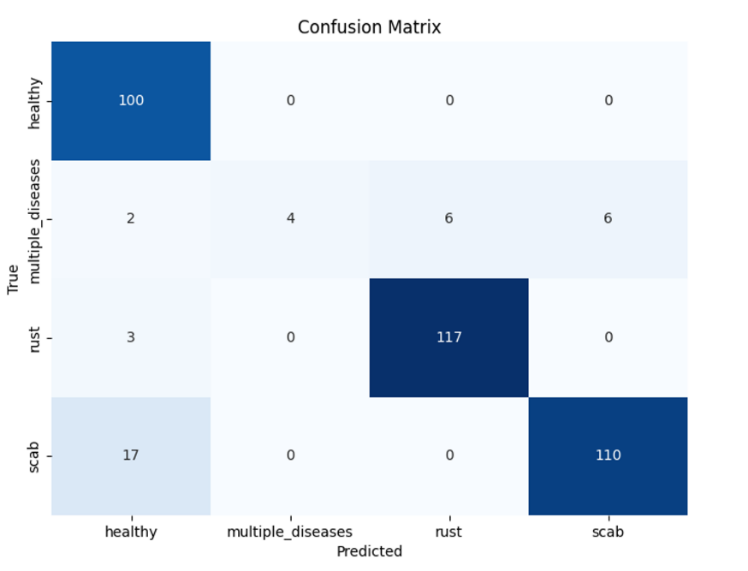
A.

Fig 11. Confusion Matrix of Plant Pathology Dataset

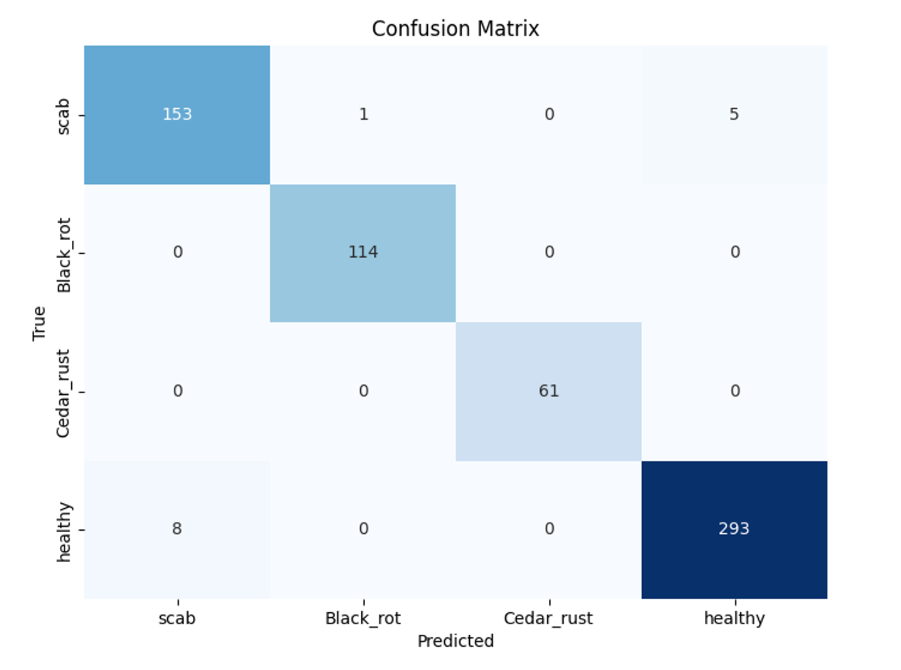
B.

Fig 12. Confusion Matrix of Apple Dataset

C.

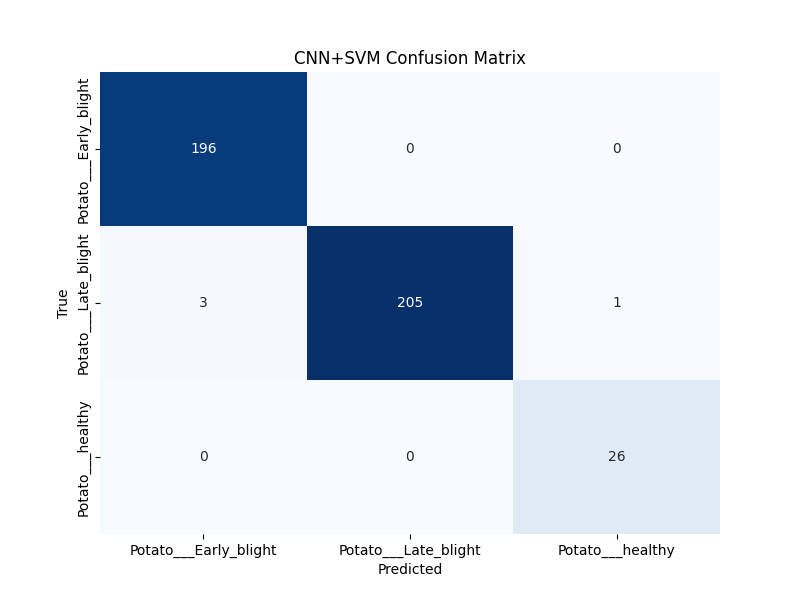


Fig 13. Confusion Matrix of Potato Dataset

D.

Fig 14. Confusion Matrix of Corn Dataset

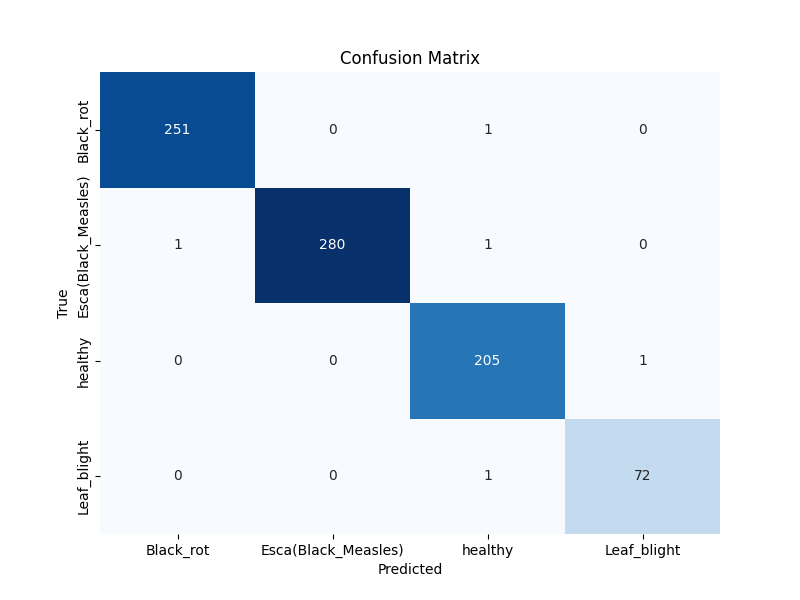
E.

Fig 15. Confusion Matrix of Grape Dataset

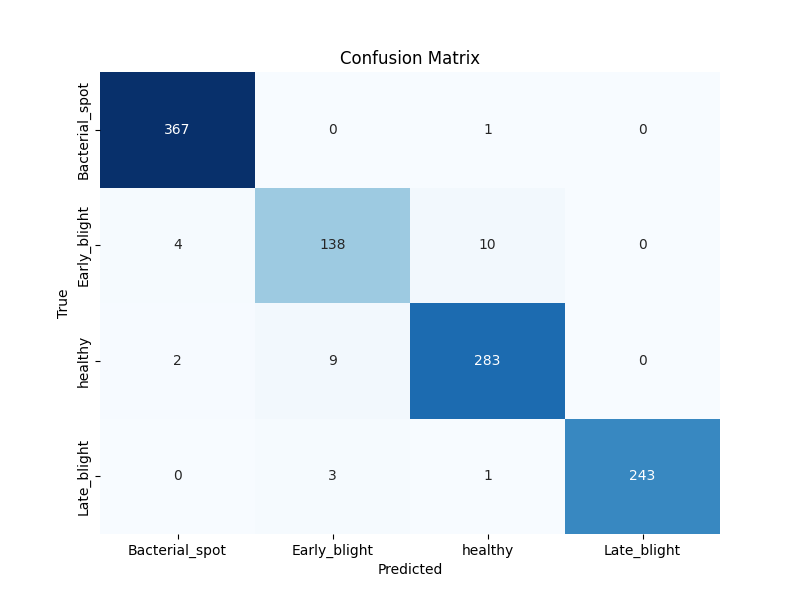
F.

Fig 16. Confusion Matrix of Tomato Dataset

**Accuracy and Loss over Epochs:**

Accuracy and loss over epochs are performance metrics used in the training of machine learning models, particularly neural networks. These metrics are typically visualized in a plot to monitor the model's learning progress during training.

**Accuracy Over Epochs:** Accuracy represents the ratio of correctly predicted instances to the total number of instances. It is a measure of the model's overall correctness. The accuracy over epochs plot shows how the model's accuracy evolves during training. As epochs progress, an increasing accuracy indicates improved model performance.

**Loss Over Epochs**: Loss is a measure of the model's prediction error during training. It quantifies the difference between the predicted and actual values. The loss over epochs plot displays how the loss decreases over time. The goal is to minimize the loss, indicating that the model is learning and making more accurate predictions.

In crop disease detection, the training and validation of the Convolutional Neural Network (CNN) were meticulously examined over 20 epochs. The accuracy and loss curves depicted the model's learning trends throughout the training process

|  |  |  |
| --- | --- | --- |
| Dataset | Accuracy over Epochs | Loss over Epochs |
| Plant  Pathology |  |  |
| Apple |  |  |
| Potato |  |  |
| Corn |  |  |
| Grape |  |  |
| Tomato |  |  |

Table 3. Accuracy and loss graphs of different datasets

**ROC Curve:**

The Receiver Operating Characteristic (ROC) curve serves as a crucial tool for assessing the performance of our machine learning models across multiple classes. The ROC curve is a graphical representation of the trade-off between true positive rate (sensitivity) and false positive rate (1-specificity) at various thresholds. For our multi-class classification task, we have employed a SVM model on the predictions generated by a CNN model.

The presented ROC curves showcase the SVM model's discrimination ability for each class, providing insights into the model's performance across diverse disease categories. Each curve is associated with an Area Under the Curve (AUC) value, quantifying the model's effectiveness in distinguishing positive instances from negatives. A higher AUC indicates superior classification performance. The diagonal dashed line represents the ROC curve for a random classifier, serving as a baseline reference. Our goal is to surpass this baseline and achieve ROC curves that significantly deviate from the diagonal, reflecting robust disease classification. The legends accompanying each curve specify the corresponding class and the associated AUC, facilitating a comprehensive understanding of the SVM model's discriminative capabilities.

|  |  |
| --- | --- |
| Datasets | ROC Curve |
| Plant  Pathology |  |
| Apple |  |
| Potato |  |
| Corn |  |
| Grape |  |
| Tomato |  |

Table 4. ROC curves of different datasets in CNN+SVM model

1. **Conclusion**

A considerable portion of the global populace makes their living from agriculture, and it is essential to both the global economy and the continuation of human life. Farmers suffer large losses because of crop diseases severe negative effects on agricultural output. We can enable farmers to closely monitor crop health and increase yields by using deep learning and ML approaches. We have seen encouraging results in diagnosis of crop diseases by using classification algorithms such as Random Forest, CNN, and SVM. Additionally, the outcomes showed that CNN+SVM had the highest accuracy across all datasets.

1. **Future Work**

In the future, there are numerous ways to improve the field of agricultural disease detection. The first and most important task is to keep adding images to the datasets by gathering more detailed and varied visual data of different crops. Integrating (Internet of Things) IoT can further help in collecting live images of the crop. We can also build an app which helps people to use our model from their own crop fields. In conclusion, a major development in agricultural technology is the use of ML and deep learning techniques for identification of crop disease. It has the power to transform agriculture, guarantee food security for all people, and lessen the negative effects that crop diseases have on the environment and the economy. Through sustained enhancement of these models and increased availability to farmers, we may cultivate a more prosperous 9 sustainable agriculture in the future.

# REFERENCES

# Alexandre Meybeck, Vincent Gitz, Leslie Lipper, Susan Braatz, and Cassandra De Young, “Climate change and food security: risks and responses”, 2015, Food and agriculture organization of the UN.

# Albert Khakimov, Alisher Omonlikov, Samad Utaganov, and IIkhom Salakhutdinov, “Traditional and current-prospective methods of agricultural plant diseases detection: A review,” 2022, DOI:10.1088/1755-1315/951/1/012002.

# Role of Modern Technology in Agricultural, 2023.

# Jehoon Sung, “The Fourth Industrial Revolution and Precision Agriculture,” 2018, DOI: 10.5772/intechopen.71582.

# “Plant Village Dataset”, Abdallah Ali.

# “Plant Pathology 2020-fgvc7” ,2020, kaggle competition.

# Shujuan Zhang, Bin Wang, and Lili Li,” Plant Disease Detection and classification by deep learning: Review,” 2021, Shanxi Agricultural University, Jinzhong, China, DOI: 10.1109/ACCESS.2021.3069646.

# Niveditha M, Ramachandra Hebbar, Shashank N, Pooja R, Shima Ramesh, Prasad Bhat N, MVJ Engineering college, Karnataka, India, DOI: 10.1109/ICDI3C.2018.00017.

# Srdjan Sladojevic, Marko Arsenovic, Dubravko Culibrk, Darko Stefanovic,” Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification,” 2016 University of Novi Sad, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia, University of Trento, Via Sommarive 9, Povo, 38123 Trento, Italy.

# Pranesh Kulkarni, Atharva Karwande, Tejas Kolhe, Soham Kamble, Akshay Joshi, and Medha Wyawahare, “Plant Disease Detection Using Image Processing and Machine Learning,” Department of Electronics and Telecommunication, Vishwakarma Institute of Technology, Pune, India

# Sunil S. Harakannanavar,jayasri M. Rudagi, Veena I Puranikmath, Ayesha Siddiqua, and R Pramodhini,”Plant leaf disease detection using computer vision and machine learning algorithms,” Nitte Meenakshi Institute of Technology, Yelahanka, Bangalore, Karnataka, Department of Electronics and Communication Engineering, S. G. Balekundri Institute of Technology, Shivabasava Nagar, Belagavi, Karnataka, India.

# Gnanavel Sakkarvarthi, Godfrey Winster Sathianesan, Vetri Selvan Murugan, Avulapalli Jayaram, Prabhu Jayagopal, Mahmoud Elsisi, “Detection and Classification of Tomato Crop Disease Using Convolutional Neural Network,” 2022, SRM Institute of Science and Technology, Panimalar Engineering College, Vellore Institute of Technology, India, Benha University, Egypt, National Kaohsiung University of Science and Technology, Taiwan.

# C Jackulin, and S Murugavalli, “A comprehensive review on detection of plant disease using machine learning and deep learning approaches,” 2022, Department of CSE, Panimalar Engineering College, Tamilnadu, India, DOI: 10.1016/j.measen.2022.100441.

# Bala Murugan MS, Manoj Kumar Rajagopal, and Diproop Roy, “IoT Based Smart Agriculture and Plant Disease Prediction”, 2021, School of Electronics Engineering, Vellore Institute of Technology, Chennai, India.

# Sharada P. Mohanty, David P. Hughes, and Marcel Salathe, “Using Deep Learning for Image-Based Plant Disease Detection”, 2016, Penn State University, State College, PA, USA.

# Murk Chohan, Adil Khan, Saif Hassan Katper, and Muhammad Saleem Mahar, “Plant Disease Detection using Deep Learning”, 2020, Begum Nusrat Bhutto Women University Sukkur, Sukkur Institute of Business Administration.

# Vishnu S, and Ranjith Ram, “Plant Disease Detection Using Leaf Pattern: A Review,” 2015, ECE Department, Govt College of Engineering, Kannur, Kerala, India.