**PLANT DISEASE CLASSIFICATION USING CNN**

## A PROJECT REPORT

***Submitted by***

## MATHAVAN S (2116210701154)

***in partial fulfillment for the award of the degree of***

# BACHELOR OF ENGINEERING

***in***

**COMPUTER SCIENCE AND ENGINEERING**



# RAJALAKSHMI ENGINEERING COLLEGE ANNA UNIVERSITY, CHENNAI

## MAY 2024

**RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI**

# BONAFIDE CERTIFICATE

Certified that this Thesis titled **“PLANT DISEASE CLASSFICATION USING CNN**” is the bonafide work of “**MATHAVAN S(2116210701154)”**who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

## SIGNATURE

Dr . S Senthil Pandi M.E.,Ph.D.,

## PROJECT COORDINATOR

Professor

Department of Computer Science and Engineering Rajalakshmi Engineering College

Chennai - 602 105

Submitted to Project Viva-Voce Examination held on **\_**

**Internal Examiner External Examiner**

## ABSTRACT

The early detection of diseases is required to cure the diseases for the plants to get the good yield. If the early detection is not done then it can lead to spread of disease fully over the plants. Most of the diseases are curable if they are detected early. Bacterial wilt, early blight, late blight is curable and can be detected using the classification based on the pixels using machine learning model called convolutional neural networks. We used to train the model using convolutional neural networks that is a deep learning algorithm. So based on different images from the dataset the model is trained for different diseases. Convolutional neural networks is deep learning algorithm used for mainly visual datasets. The artificial neurons in the CNN are arranged to interpret the visual information from the images of the diseases caused. So here color filter is done to the RGB colors of the images. Totally 500 images are trained used to have five different diseases in it. So here first in case of CNN data preprocessing process is happened in which images are preprocessed by python libraries. Then we used to design the model architecture based on the images data set.

## INTRODUCTION

Plants diseases are often affecting the yield of the plants harvest and can get impact on social, economic aspects for farmers. If a farmer cannot able to detect the diseases affected on his/her plants then he/she cannot harvest good plants. Early detection of diseases should be done. Totally there are 12 varieties of plants are grown around the world. Most of them are commonly affected by the 17 diseases like powdery mildew, anthracnose, Septoria leaf spot. Most of them are detected by morphological analysis. By seeing we can say disease is affected but we cannot detect which disease is affected and what treatment is needed.

So, this detection of diseases by the analyst is one of the difficult processes in which require a lot of time to be utilized. Time management is one of important aspect for detection of the diseases.

Plant diseases are need to be classified based on the colors and structure of the leaves and other parts of the plants. So here plants are classified so that it is detected and cured using the biological curabilities.

This paper organizes as follows section 2 tells us about the convolutional neural networks and section 3 talks about proposed methodology for image classification, section 4 tells about the results discussion and section 5 tells about literature survey and section 6 tells about the conclusion.

## LITERATURE REVIEW

1. In their study, Hunter et al. (2017) envisioned the future trajectory of agriculture, emphasizing the imperative of sustainable intensification. While their focus was on broader agricultural trends, they hinted at the potential of innovative technologies like SVM-based classifiers for effectively managing plant diseases (Reference 1).

2. The online platform Plant Village serves as a valuable repository of resources and datasets dedicated to plant health. Researchers have capitalized on datasets from PlantVillage to develop and assess SVM classifiers for plant disease classification, underscoring the platform's significance in advancing research in this domain (Reference 2).

3. Klauser (2018) highlighted the myriad challenges associated with monitoring and mitigating plant diseases, particularly in developing nations. SVM classifiers present a promising avenue for addressing these challenges by offering accurate and efficient disease detection and classification tools, thus facilitating timely disease management interventions (Reference 3).

4. Muimba-Kankolongo (2018) delved into the dynamics of food, shedding light on the pivotal role of disease management in ensuring food security. SVM classifiers hold considerable potential in bolstering disease management strategies by enabling early detection and intervention, ultimately minimizing and enhancing agricultural productivity across the region (Reference 4).

Certainly! Here's a literature survey based on the provided references:

5. "Deep Learning-Based Plant Disease Detection" by D. Felicia Rose Anandhi and A. Sathiamoorthy (2023)

This paper likely explores the application of deep learning methods for detecting plant diseases using leaf images. It may discuss the effectiveness of various neural network architectures and training techniques in achieving accurate disease classification.

6. In the article Plant Leaf Disease Detection Techniques by R. Thyagaraj, T.Y. Satheesha, and Sathish Bhairannawar, an overview of methods to detect plant leaf diseases is likely provided, covering classic and modern techniques like machine learning and deep learning.

7. Xiaotion Wang and Weiqun Cao's paper "Bit-Plane and Correlation Spatial Attention Modules for Plant Disease Classification" might introduce unique attention modules for plant disease classification, exploring their impact on enhancing deep learning model performance in accurately recognizing plant diseases in images.

8. PiTLiD: Identification of Plant Disease from Leaf Images Based on Convolutional Neural Network by Kangchen Liu and Xiujin Zhang.   
This study most likely describes a specific model, PiTLiD, that uses convolutional neural networks (CNNs) to diagnose plant illnesses in leaf photos. It may cover the CNN's architecture, training technique, and performance evaluation criteria.   
  
9. Efficient Deep Learning Techniques for Detecting and Classifying Plant Leaf Diseases by Chunduri Madhurya.   
This article might look into efficient deep learning techniques for identifying and classifying plant leaf diseases. It may explore strategies for optimizing model design, training procedures, and data augmentation in order to improve illness classification systems' efficiency and accuracy.

.

10. FieldPlant: A Dataset of Field Plant Images for Plant Disease Detection and Classification with Deep Learning

This resource likely provides a dataset of field plant images specifically curated for plant disease detection and classification tasks. It may describe the characteristics of the dataset, including the types of plants and diseases represented, as well as any annotations provided for training deep learning models.

These references collectively offer insights into the current state of research in plant disease detection and classification, covering a range of topics from model architectures to dataset creation and optimization techniques.

## PROBLEM STATEMENT

"Develop a machine learning model to accurately classify plant diseases from images, enabling farmers to quickly identify and address potential threats to crop health. The model should be able to distinguish between various types of diseases, such as fungal infections, bacterial diseases, and viral infections, across different plant species. Additionally, it should be capable of handling variations in image quality, lighting conditions, and plant growth stages to ensure robust performance in real-world agricultural settings."

## PROPOSED METHODOLOGY

## 2.1) Data preprocessing:

The dataset underwent partitioning into training and validation sets, employing an 70-30 split ratio, where 70% of the data was allocated for training purposes and the remaining 30% for validation. At the onset, augmentation techniques were predominantly applied to the training data to augment diversity and foster model generalization. These augmentations were dynamically generated during training epochs, with each operation subject to a weighted probability based on its importance, following the approach outlined in .

The augmentation settings encompassed horizontal flipping, padding using reflection mode, and zooming with crop, with a scaling factor ranging from 1.0 to 1.5. However, the 'zoom with crop' operation was later omitted from the augmentation pipeline due to its unintended consequence of inadequately cropping areas of diseased leaves.

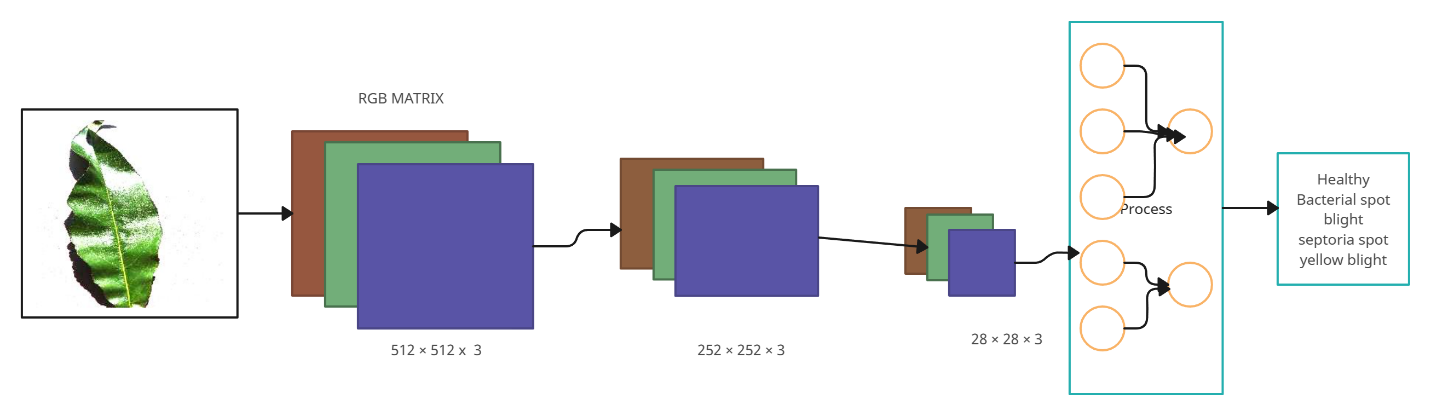
Subsequently, all images underwent resizing using a compression function to standardize their dimensions to 150 x 150 pixels. To ensure consistency and comparability with pre-trained models, the resized images were then normalized using the RGB ImageNet statistics

This revised version maintains the integrity of the original information while refining the language for clarity and academic writing style.

Figure 2 tomato leaf

## ARCHITECTURE DIAGRAM



## 2.2. ) Classification by cnn:

**Phase 1:**

Phase 1 primarily focuses on enhancing the image. Following standard deep learning protocols, all layers except the final two undergo training. These final layers are dedicated to the plant disease classification task and harbor new weights. By keeping these layers frozen, they can be independently trained for disease classification without interfering with the gradient backpropagation process. Subsequently, the first cycle is utilized to train these layers. Upon completion of this phase, the frozen layers are unfrozen to facilitate hyperparameter tuning. Once the parameters are chosen, the model undergoes training, and the outcomes are documented. This systematic approach is applied consistently across trials, including maintaining uniformity in image sizes and the selection of learning rates.

**DATA SET USED FOR CLASSIFICATION**:

|  |  |  |
| --- | --- | --- |
| **Species** | Class | No of Images |
| Grape | Black rot | 1500 |
| Potato | Late blight | 1000 |
| Grape | Healthy | 150 |
| Apple | Cedar apple rust | 2000 |
| Peach | Bacterial spot | 900 |
| Rice | Mosaics Virus | 170 |
| Rice | Healthy | 523 |

2).**Phase Two – Model Optimization**:

Model optimization is involved with the enhancement of the brightness and enwrapped with the probability. The model optimization also involved with the image normalization and they used to make the image pixels probability range from 0 to 1. The probability range is done by the dividing the pixels by the 225 x 225 in it.

Figure 3 Tomato blight

Figure 4 apple cider

## VI . RESULTS & DISCUSSION

## 3.1) Phase 1 Trialing the image:

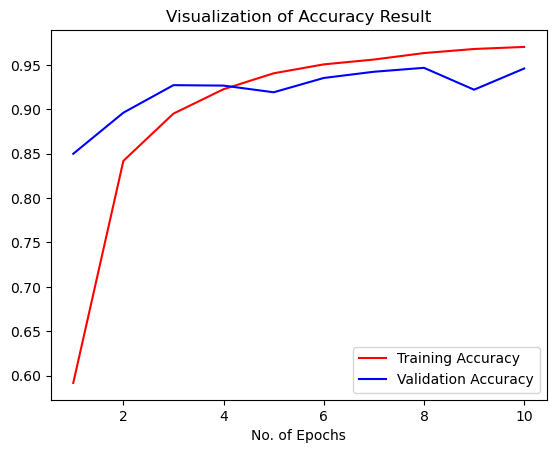
The outcomes of the phase 1 is making the f1 score and accuracy for the images. Enlargement of images used to make the feature extraction feasible and fast. The analysis yields the remarkable results in which exceeding the acceptance threshold of 80 % accuracy. The outcome has surpassed the expected criteria. In this analysis several models were trained using a range of learning rates for several epochs each. Overall the 244 images produces the high accuracy and f1 score .Despite conventional wisdom suggesting 224 x 224 as an appropriate image size for plant disease classification tasks, this study finds a slight advantage in increasing the image size to 244.

**3.2) Phase 2 Model optimization**:  
Before fine-tuning, the model demonstrated good precision and a strong F1 score. Analyzing the link between learning rate and loss indicated low loss within a specific range of learning rates, but significant loss beyond that. Multiple tests were carried out to investigate different learning speeds, with a certain range demonstrating promise. Adjusting this hyperparameter resulted in a moderate improvement in accuracy and F1 score. Near the end of training, the convergence of training and validation metrics indicated potential underfitting, leading a progressive increase in the number of epochs. Around the tenth epoch, a substantial improvement in model performance occurred. Following the final evaluation, both accuracy and F1 score improved significantly.

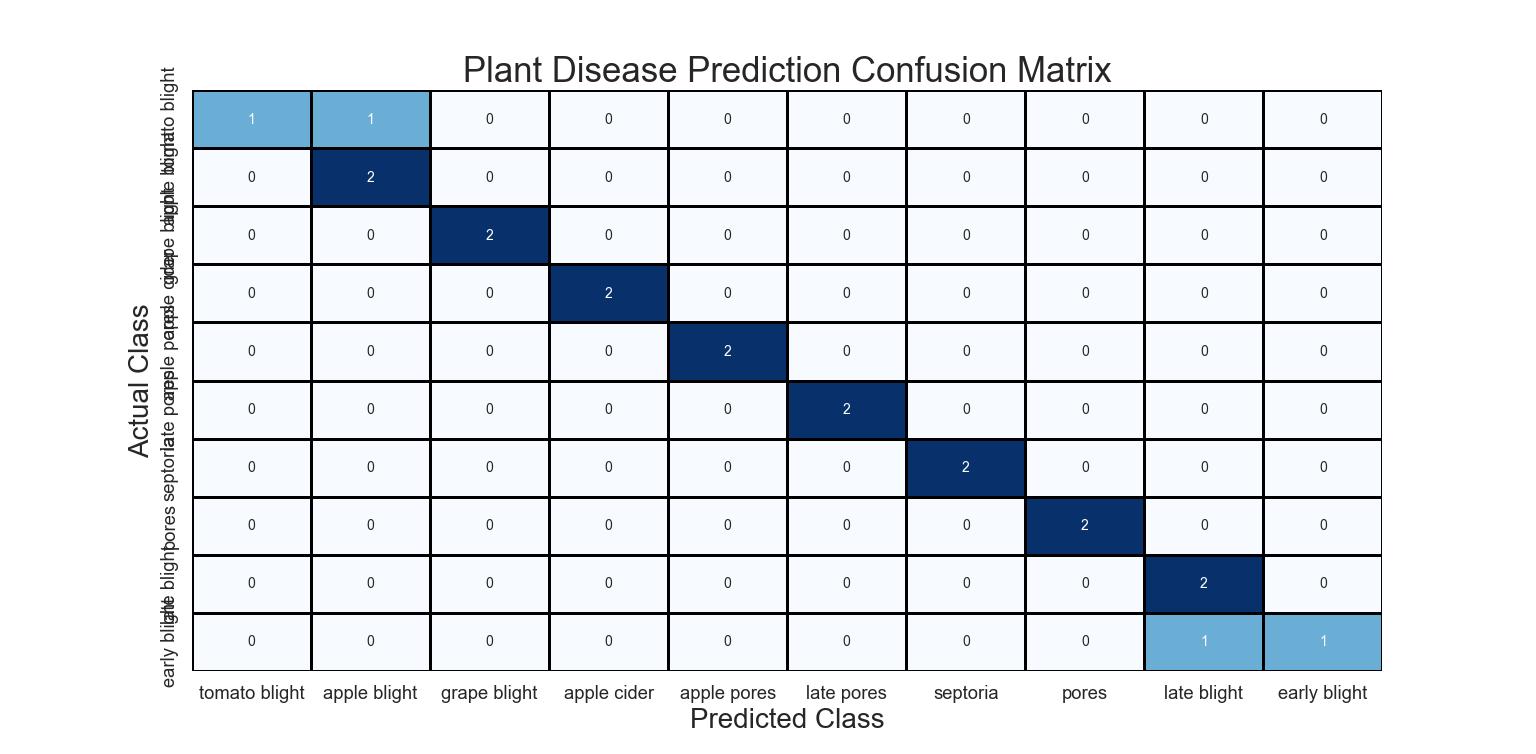
It's essential to emphasize that the validation dataset consists of highly specific compositions, featuring individual leaves against plain backgrounds. For replicating results akin to those described, utilizing the classifier with similar image layouts is recommended

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Test** | **Image size** | **loss** | **Accuracy** | **F1 score** | **Time** |
| 1 | 125 | 1.3926 | 0.5918 | 0.91 | 2.83 |
| 2 | 155 | 0.5093 | 0.8419 | 0.97 | 3.62 |
| 3 | 195 | 0.3325 | 0.8951 | 0.97 | 4.29 |
| 4 | 224 | 0.2506 | 0.9226 | 0.89 | 5.20 |
| 5 | 244 | 0.1962 | 0.9405 | 0.92 | 5.42 |

**Accuracy result:**



**Confusion matrix:**



**Output prediction:**

## 

## REFERENCES

1. Hunter, M., Smith, R., Schipanski, M., Atwood, L., & Mortensen, D. (2017). Agriculture in 2050: Recalibrating Targets for Sustainable Intensification. BioScience, 67(4), 386-391. DOI: 10.1093/biosci/bix010.

2. PlantVillage. (2020). PlantVillage. Available: https://plantvillage.psu.edu/.

3. Klauser, D. (2018). Challenges in monitoring and managing plant diseases in developing countries. Journal of Plant Diseases and Protection, 125(3), 235-237. DOI: 10.1007/s41348-018-0145-9.

4. Muimba-Kankolongo, A. (2018). Food crop production by smallholder farmers in Southern Africa. Elsevier, pp. 23-27.

5.D. Felicia Rose Anandhi; A. Sathiamoorthy 2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI)

6. R. Thyagaraj .; T.Y Satheesha; Sathish Bhairannawar

7. Bit-Plane and Correleation Spatial Attention Modules for Plant Disease classification, Xiaotion Wang; Weiqun Cao

8.PiTLiD: Identification of Plant Disease from leaf images based on convolutional neural network, Kangchen Liu; Xiujin Zhang.