Leaf Disease Classifier using CNN

UNIVERSITY OF ENGINEERING
&
MANAGEMENT, JAIPUR

Leaf Disease Classifier using CNN

Submitted in the partial fulfilment of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE & ENGINEERING

Under

UNIVERSITY OF ENGINEERING & MANAGEMENT, JAIPUR BY

Name/Roll No.: -Abhay Prasad (12021002026024)
Arya Kumar Johary (12021002026028)

UNDER THE GUIDANCE OF

Asst.Prof. Hriday Banerjee

&

Asst.Prof. Sagarika Ghosh

COMPUTER SCIENCE & ENGINEERING

UNIVERSITY OF ENGINEERING & MANAGEMENT, JAIPUR

Approval Certificate

This is to certify that the project report entitled "**Leaf Disease Classifier using CNN**" submitted by **Candidate's Name/Roll No.: -Abhay Prasad (12021002026024) , Arya Kumar Johary (12021002026028)** in partial fulfilment of the requirements of the degree of **Bachelor of Technology** in **Computer Science & Engineering(AIML)** from **University of Engineering and Management, Jaipur** was carried out in a systematic and procedural manner to the best of our knowledge. It is a bona fide work of the candidate and was carried out under our supervision and guidance during the academic session of 2021-2025.

____

Asst.Prof. Hriday Banerjee & Asst.Prof.Sagarika Ghosh
Project Guide, Assistant Professor (CSE)
UEM, JAIPUR

____

Prof. Mrinal Kanti Sarkar
HOD (CSE)
UEM, JAIPUR

____

Prof. A Mukherjee
Dean
UEM, JAIPUR

ACKNOWLEDGEMENT

We would like to express my gratitude towards **Assr.Prof. Hriday Banerjee** and **Asst.Prof. Sagarika Ghosh** who has been a precious help in my voyage of finishing this undertaking. Without his assistance, direction, and support, making this project a reality would have been unimaginable. His opportune criticism was tremendously essential for me and helped me greatly to reach the targets set for the project tasks. Likewise, we would also like to express gratitude toward **Dr**. **Mrinal Kanti Sarkar** his commitment and supporting this project as my second reader. In conclusion, we would like to thank the personnel and staff of the Department of Computer Science Engineering at University of Engineering And Management Jaipur for their ceaseless regulatory help.

Abhay Prasad (12021002026024),
Arya Kumar Johary (12021002026028)

ABSTRACT

In our project, we focus on predicting leaf diseases in potato, tomato, and pepper plants, recognizing the significance of these crops in global agriculture. Potato, tomato, and pepper are among the key staples, each facing its set of challenges with diseases affecting their leaves. The demand for these crops has surged, particularly amidst the coronavirus pandemic, heightening the need for effective disease management strategies. We propose a system leveraging deep learning techniques, specifically a convolutional neural network architecture, to accurately classify diseases based on leaf conditions. Timely and precise identification of diseases is crucial to mitigate losses in both quality and quantity of harvests. Our model achieves impressive accuracies of 96% for potato, 79% for tomato, and 99.5% for pepper leaf disease classification. These results underscore the feasibility and effectiveness of employing deep neural networks in addressing disease challenges across different crops.

Table of Contents

List of Figures

List of Tables

CHAPTER 1 - INTRODUCTION

Agriculture is an essential sector in countries like India as those countries' economy directly or indirectly dependent on agriculture. It indicates the necessity of taking care of plants from seedling until the expected crop obtains. Through this process, the crop needs to cross a lot of phases to obtain the expected crop such as weather conditions, the survival of the crop from various diseases, and the survival of the crop from various animals. The next major issue is weather conditions which will not be in the control of humans, humans can only pray for better weather conditions to obtain a better crop. Finally, the major issue which is very crucial to protect the crop from various diseases as these diseases can impact the complete growth and yield of the crop. If this process of identification and classification of diseases able to digitalize which would be helpful for the agriculturists. It will decrease the time for the identification of disease and precision in classifying the diseases.

There are a lot of significant crops exist in India, one among them is Potato. More than three fourths of the population of India consumes potato daily at the same time it is one of the popular yielding crops in India. Yet, the yield of the potato crop can be diminished due to various diseases such as late blight and early blight. These diseases are also known as Phytophthora Infestans and Alternaria Solani respectively in scientific terms. Timely identification and classification of these diseases will lead to avoid the yield as well as financial losses. The popular way of identification of these diseases through the utilization of the human eye for decades. But, this methodology arises with certain infeasibilities such as overtime will be taken for processing and shortage of experts at fields in remote locations. Therefore, the image analysis turned out to be an efficient methodology that will play a vital role in monitoring as well as the identification of the plant disease conditions effectively. Because the visible patterns are available on the plant leaves and patterns will be identified using various image processing methodologies for obtaining a particular pattern corresponding to a disease which will create an impact in the identification of various diseases. Thus, the obtained features or patterns will be compared with the historical data and able to classify the disease which can be done by using various machine learning and deep learning methodologies. So, the combination of image processing machine learning and deep learning is very effective in the identification and classification of diseases.

1.2 BRIEF LITERATURE SURVEY

- Monzurul Islam et al. 2017[1] proposed an approach that combines the processing of images and machine learning to allow leaf image disease to be diagnosed. This has been an automated system that categorizes potato plant diseases as well as unaffected leaves from the public image, known as ‘Plant Village’. The mentioned segmentation process and classification process through support vector machine methodology displays classification of images about 300 and the accuracy of the proposed model about 95%. Thus, the proposed approach offers a way for the automatic diagnosis of plant diseases on a huge scale. The multiclass SVM image segmentation is used for designing a system that is automated and easy to use. For primary diseases in potatoes such as Late Blight and Early Blight, a little computational effort is identified. The approach would provide framers with viable, reliable, and successful methodology and time-saving processes for disease identification.
- Harshal Waghmare and Radha Kokare 2016[2] proposed technologies for plant disease detection analysis and pattern detection of the leaf texture. This work is based on the method of detection of grapes leaf disease. The device is used as input and segmentation on a single plant leaf after context removal is carried out. The image segmentation of the diseased component is then analyzed using a high-pass for the leaf. Special sectional leaf texture is obtained. Locally based fractal features nature invariant provides a good model of texture. The texture would be different for each independent illness. The texture pattern then extracted is graded by multiclass SVM. Multiclass SVM implementations are formulated to identify the diseases observed in grape plants for the processing of DSS (Decision Support Systems) automated and farmers, easily available. The Scheme performs segmentation and examination of a single leaf and the diseased portion of the leaf is observed by the high pass filter. A fractal-based retrieval of segmented leaf texture function that is invariant locally in nature and the strong texture module then provides. The texture removed pattern is then classified as an SVM designation for multiclass in groups who are healthy or ill classes respectively. The study concentrates on major widely encountered diseases grasses are downy mildew &

black red. The recommendation approach quickly provides farmers with guidance from agricultural experts with 96.6 percent of accuracy.

- Mrs. Shruthi U et al. 2019 [3] proposed machine learning methods to classify diseases and it applies mainly to data and It prioritizes itself and the performance of those tasks. This paper demonstrates the phases of general identification of plant diseases and the machine learning method on comparative research is for plant disease identification such as the acquisition of image-set, processing of the obtained data in the image-set, segmentation of the data in the image-set, feature extraction, and classification of images in image-set based on the extracted features and patterns identified. This survey shows that the Convolutional Neural Network high accuracy and more number of multiple disease crops.
- Rajleen Kaur and Dr. Sandeep Singh Kang 2015[4] proposed for automatic disease detection and disease part of the plant leaf images and also crop agriculture production. It's achieved with computer advancement technology that allows agriculture to develop production. SVM is the latest classifier of the neural network approach and problem for the detection of accuracy. SVM is introduced in this article contains two datasets; one is a data collection for training and dataset of the train. The original image is first taken and used for processing. It offers, secondly, image pixels in black and background and hue section and saturation section is separated. Third, disease diagnosis and the unhealthy component are identified and a stable portion is segmented from it. This work also provides a percentage of the region where diseases arise and give us the disease name. As the image results the region impacted is 5.56%. This work gives accuracy which is stronger than the proposed results of the algorithm.
- S.Ramesh and D.vydeki 2018[5] offers an algorithm for machine learning to find symptoms in rice plant about the disease. Automated plant identification, the Machine learning algorithm is used to carry out the disease. Images of safe leaves and blasting diseases damaged leaves are taken for the suggested method. The characteristics are removed safely and disease-prone portions of the rice leaf. The whole thing. Collection of 300 images for training and testing. The images are store under the proposed system and the leaf suggested were graded as either good or infected. The effects of the simulation give a 99 percent accuracy for bursts and 100 percent is during the training phase under normal images. The exams accuracy of the process for infected and healthy persons is 90 and 86 percent respectively. The above approach helps farmers in crops

protection from disease. This approach gives you a free of disease and quality development of the crop. Finally, it would be proposed that the Indian farmers should use this approach to prevent the disease spread and make choices in the crop at all times for improving crop production and improving profitability

- V. Suresh et al. in 2020 [6] proposed a website-based image processing and machine learning technological methodology. In this framework, firstly, the images will be gathered and then provided to the trained classifier through the website. Once the image is processed and extracted the features then the disease will be detected. Once the disease is detected, the website redirects the page into the page containing appropriate pesticides and chemicals with their usage and MRPs as well. It is one of the great work which identified so far as every researcher looking to detect and classify the diseases whereas this proposed methodology not only detecting the disease but also suggesting the appropriate pesticides suitable for it.
- Sukhvir Kaur et al. 2018[7] explain about various parts of a plant the signs of plant diseases are clear, but the most often found leaves are for infection detection. Researchers tried to simplify the method of diagnosis of plant diseases and leaf image classification. Many researchers successfully employed and made a major contribution to computer vision technology. This manuscript sums up the benefits and drawbacks of all these experiments to illustrate some critical topics. A conversation on common infections in various stages of a disease and a research situation is presented. To classify those who appear to be, the success of state-of-the-art techniques is analyzed to fit well through several types of crops. The manuscript highlights a variety of suitable strategies along with future research paths, some areas of concern. The survey will help scientists accomplish computer vision understanding systems for the identification of plant diseases

CHAPTER 2 – OBJECTIVES & METHODOLOGY

2.1 OBJECTIVES

- The general objectives of the study is to develop a reliable, convenient and accurate system. The study has the following specific objectives
- Identification of the causal agents responsible for potato diseases, including fungi, bacteria, viruses, and nematodes.

- Characterization of the symptoms and damage caused by each type of potato disease, including their impact on yield and economic outcomes.
- Development of a classification system based on the characteristics of potato diseases, including their mode of transmission, environmental conditions that favor their development, and the methods of disease diagnosis.
- Provision of guidance to growers, researchers, and other stakeholders on effective disease management and control strategies based on the classification system.

2.2 PROBLEM FORMULATION

- The plant diseases affect the leaves, stems, roots and fruits; it also affects the crop quality and quantity, which causes food deprivation and insecurity throughout the world
- The estimated annual crop yield loss due to crop disease is about 16% globally, which is the primary cause of food shortage and increased food production costs. According to the Food and Agriculture Organization Report (FAO), the world's population will reach approximately 9.1 billion by 2050. For a steady food supply, about 70% of food production growth is required
- The factors affecting the plants and their products are categorized as diseases and disorders. The biotic factors are the diseases caused by algae, fungi, or bacteria, whereas the biotic factors inducing disorders are rainfall, moisture, temperature and nutrient deficiency
- There exist many methods to diagnose plant diseases; one of the primary and straightforward approaches is a visual estimation. The traditional plant disease diagnosing techniques depend on the farmer's experience, which is most uncertain and unreliable

2.3 PROPOSED METHODOLOGY

Here's a high-level overview of the workflow used in our project:

Step-1: Image Acquisition: It deals with the acquisition of data from reliable sources to maintain the standard and stability so that it can be compared or extended for future studies.

Step-2: Image Pre-processing: It is a very essential phase of the framework. In this phase, mainly deals with the denoising of the image, enhancement of the image, and maintaining standard image size for all the images. Denoising and enhancement of images are essential to get a better result while segmenting the images.

Step-3: Image Segmentation: In this phase, the image will be segmented according to the region of interest. Here, in this case, the region of interest is the regions on the leaf which are affected by various diseases that need to be separated from the existing images.
Step-4: Extraction of Features: Depending on the obtained region of interest need to identify the patterns that exist. A different region of interests will have different patterns, from that scenario, one can able to extract features that are crucial in deciding the detection as well as classification.

Step-5: Evaluate the Affected Region: By comparing the region of interests and extraction features, one can able to evaluate the affected regions to obtain better accuracy in the model, otherwise there exist higher deviations.

Step-6: Processed Data: All the information related to processed image data by using the steps-1 to 5 will be gathered into a single location.

Step-7: Training Data: The training data will be obtained from the processed data. About 75% of the data with random indexing was considered to train the classifier model.
Step-8: Testing Data: The testing data will also be obtained from the processed data. About 25% of the data with random indexing was considered to test the classifier model.
Step-9: Classification: Test data will be provided to the trained classifier to classify the images into various categories such as Late Blight, Early Blight, and Healthy.
Step-10: Evaluation Metrics: Depending on the obtained results from the classifier model, the evaluation metrics such as accuracy will be obtained.

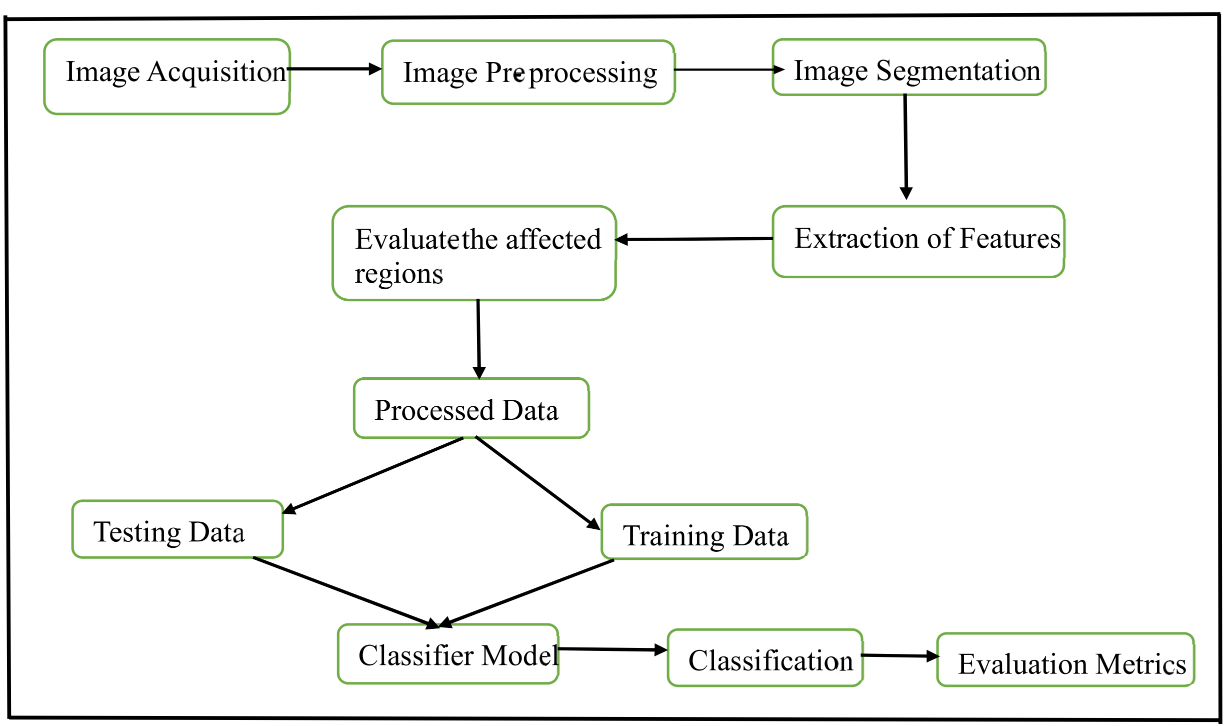


Figure2.3.1 Flow Chart of the Proposed Methodology

CHAPTER 3

MODULES REQUIRED FOR PROPOSED WORK

In this Project we have used CNN (convolution neural network), image processing, open cv 2, sequential modelling and deep learning methodologies. We have also utilized a number of other modules to implement various features and functionalities. These modules and libraries include the following:

TensorFlow: TensorFlow is an open source framework developed by Google researchers to run machine learning, deep learning and other statistical and predictive analytics workloads

Numpy: Numpy can be used to perform a wide variety of mathematical operations on arrays. It adds powerful data structures to Python that guarantee efficient calculations with arrays and matrices and it supplies an enormous library of high-level mathematical functions that operate on these arrays and matrices.

Sklearn: Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

Matplotlib: Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible. Create publication quality plots. Make interactive figures that can zoom, pan, update.

Keras: Keras is a simple-to-use but powerful deep learning library for Python. In this post, we'll build a simple Convolutional Neural Network (CNN) and train it to solve a real problem with Keras.

CNN: A convolutional neural network (CNN or ConvNet) is a network architecture for deep learning that learns directly from data. CNNs are particularly useful for finding patterns in images to recognize objects, classes, and categories. They can also be quite effective for classifying audio, time-series, and signal data.

Sequential modelling: Sequence models are the machine learning models that input or output sequences of data. Sequential data includes text streams, audio clips, video clips, time-series data and etc. A CNN can be instantiated as a Sequential model because each layer has exactly one input and output and is stacked together to form the entire network.

CHAPTER 4 – RESULTS & DISCUSSION

4.1 ACCURACY AND LOSS

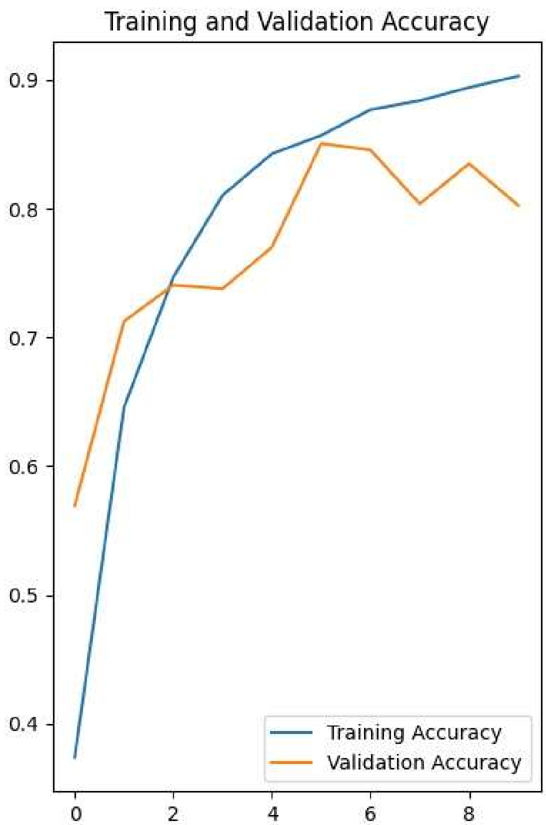


Fig 4.1.1 Tomato Model Accuracy

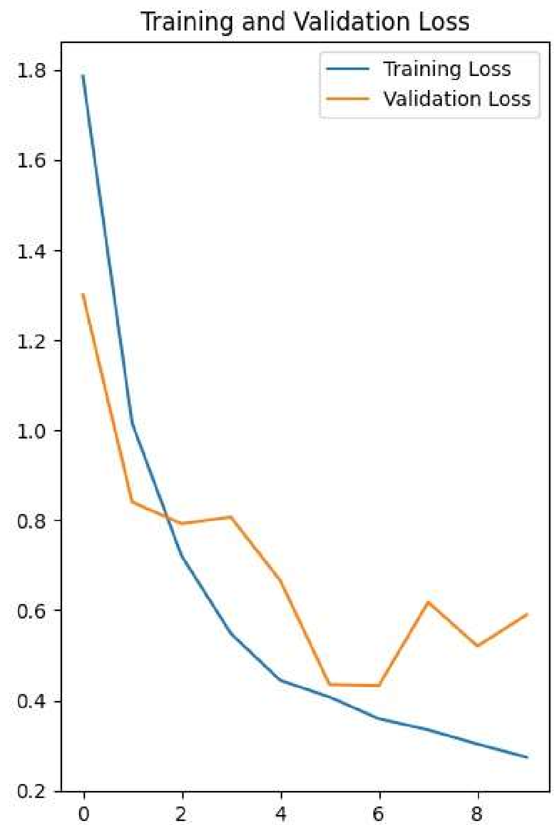
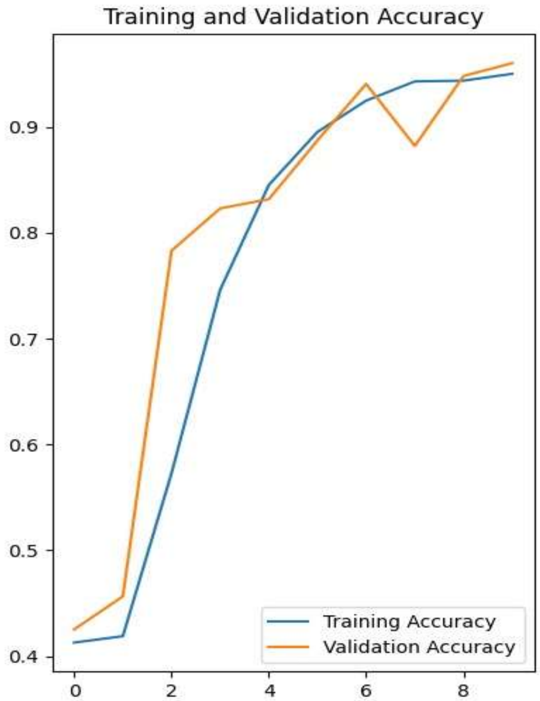


Fig 4.1.2 Tomato Model Loss



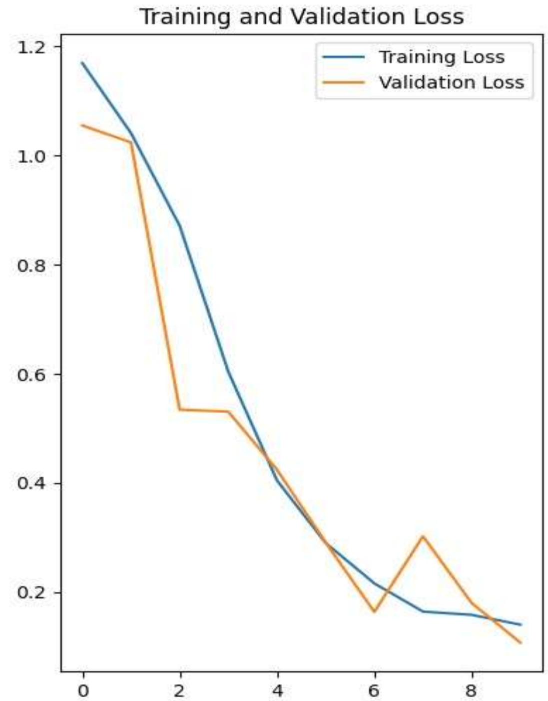
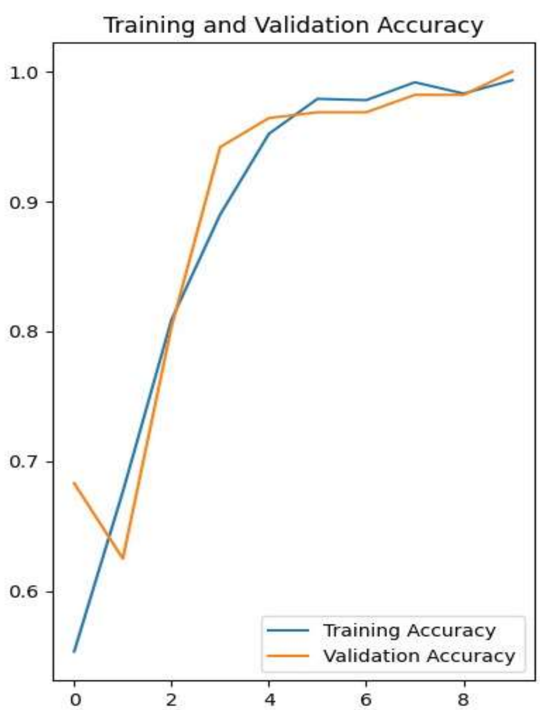


Fig4.1.3. Potato Model Accuracy

Fig4.1.4 Potato Model Loss



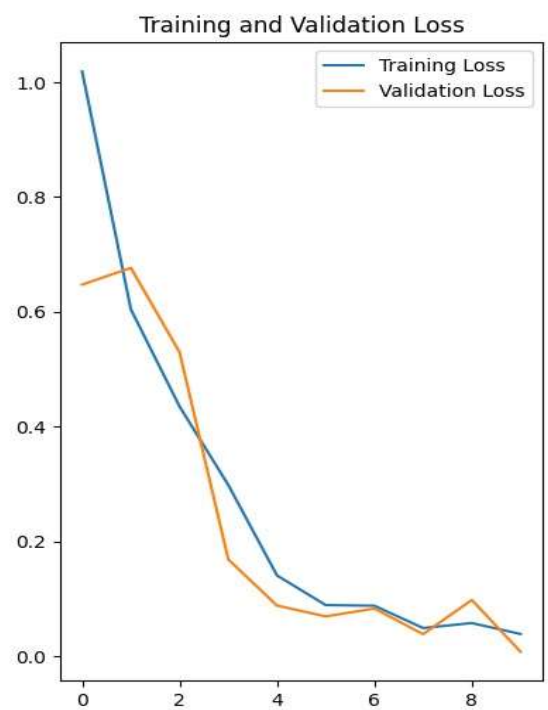
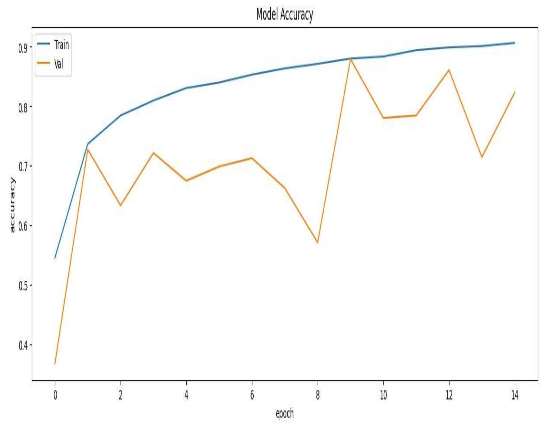
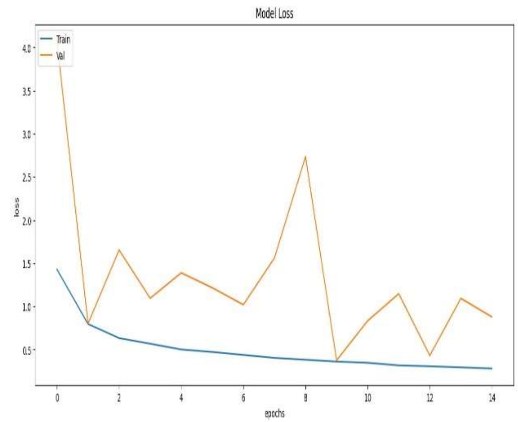


Fig4.1.5 Pepper Model Accuracy

Fig4.1.6 Pepper Model Loss

Comparison of Tomato Leaf Accuracy





Model by- Mr. Umair Shah Pirzada [4]

`loss: 0.8781 - accuracy: 0.8231`

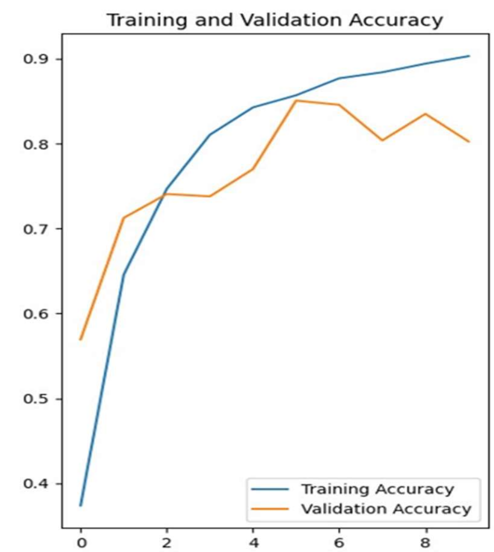


Fig 5.1.1 Tomato Model Accuracy

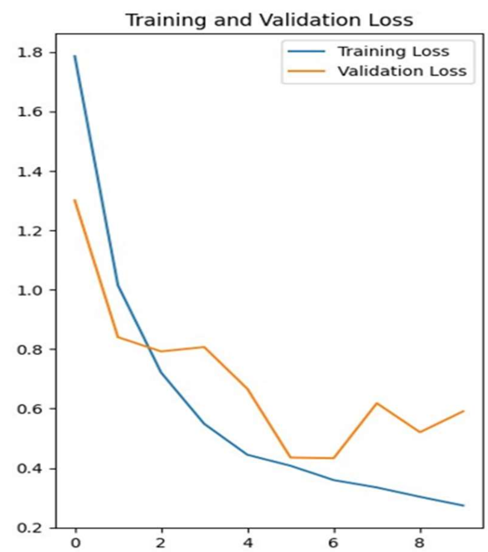


Fig 5.1.2 Tomato Model Loss

Comparison of Potato Leaf Accuracy

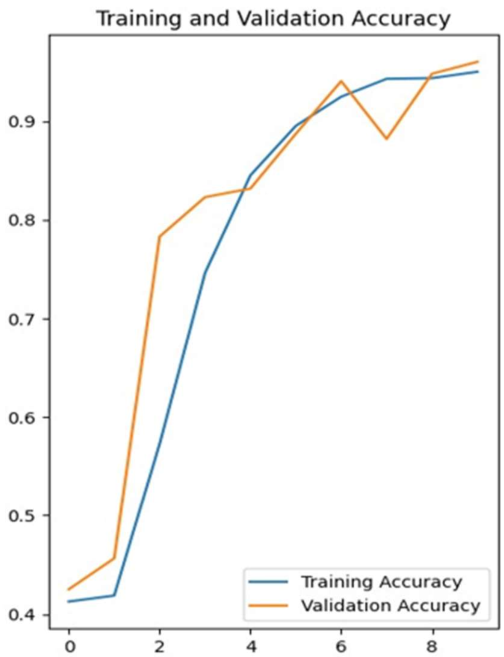


Fig5.1.3. Potato Model Accuracy

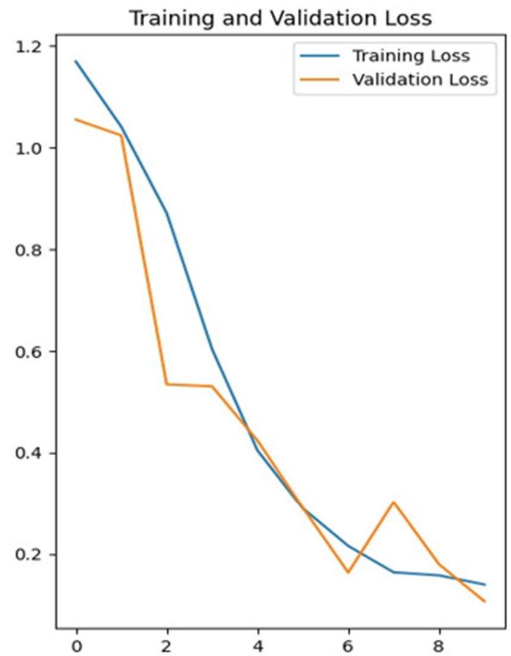
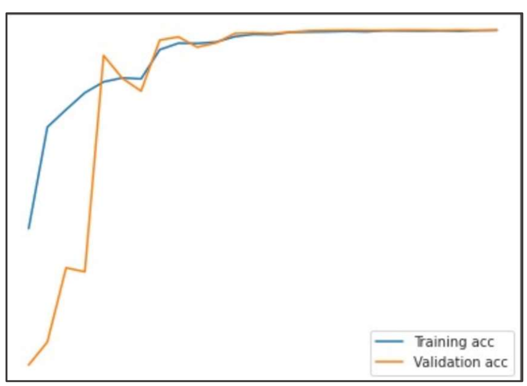
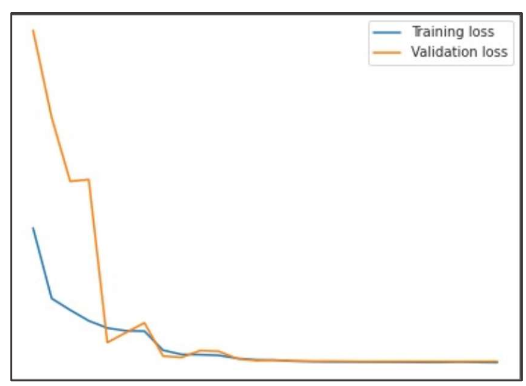


Fig5.1.4 Potato Model Loss





Model by – Amar Nassanlhajali [5]

`loss: 0.21 - accuracy: 0.92`

4.2 Leaf Diseases

Fig4.2.1. Tomato Leaf Diseases

sss

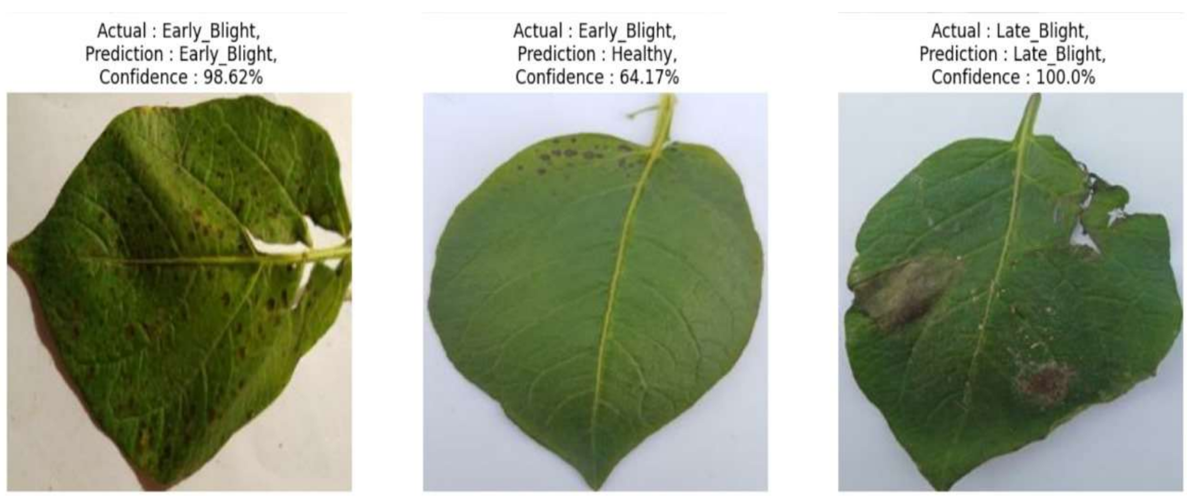


Fig4.2.2. Potato Leaf Diseases

Fig 4.2.3 Pepper Leaf Diseases

Here we have three leaves example on the basis of which we will have our evaluation metrics:

Leaf	Accuracy	Loss
Potato	0.9600	0.1114
Pepper	0.9958	0.0212
Tomato	0.7900	0.6650

Table 4.2.1 Evaluation Metrics of Proposed Methodology

CHAPTER 5 – CONCLUSION AND FUTURE SCOPE

5.1 CONCLUSION

We have used the concept of CNN and have developed a model to classify conditions in the potato leaves like early blight, late blight and healthy achieving classification accuracy of around 96.93%. Digitalization increasing across all the fields and it is high time to adopt digitalization into the field of agriculture as well to obtain better protection in terms of growth and yield. Keeping this intention as the motivation for the proposed model to detect and classify the affected and unaffected leaves of potato. The proposed framework able to achieve an accuracy of 90% and above. Yet, this accuracy needs to be improved. The existing work further can be extended by using artificial neural networks, particularly, convolutional neural networks. These days, a lot of research related to images is happening based on CNN methodologies to obtain better and reliable accuracy. The concept of activation functions, batch normalizations, convolutional layers, and fully connected layers are playing a key role in CNN architectures to attain better accuracy.

5.2 FUTURE SCOPE

- Improved Accuracy: Deep Learning and CNN models can be refined with advanced architectures to achieve higher accuracy in predicting potato leaf diseases.
- Real-time Prediction: Real-time prediction models using mobile devices or drones can quickly detect and treat potato leaf diseases in the field.
- Multi-class Classification: Multi-class classification can enable simultaneous prediction of different types of potato leaf diseases for comprehensive disease management.
- Transfer Learning: Leveraging pre-trained models from related datasets or crop types through transfer learning can improve the performance of the potato leaf disease prediction model.
- Explain ability and Interpretability: Developing interpretable and explainable models can provide insights into the important features or regions in potato leaf images for increased trust and adoption.

- Deployment in Agricultural Systems: Integrating prediction models into existing agricultural systems can enable data-driven decisions and timely actions for effective disease management.

CHAPTER 6 – BIBLIOGRAPHY

1. M. Islam, A. Dinh, K. Wahid, and P. Bhowmik, Detection of potato diseases using image segmentation and multiclass support vector machine 2017, Can. Conf. Electr. Comput. Eng., 8–11.

2. H. Waghmare, R. Kokare, and Y. Dandawate, Detection and classification of diseases of Grape plant using opposite color Local Binary Pattern feature and machine learning for automated Decision Support System 2016, 3rd Int. Conf. Signal Process. Integ. Networks

3. U. Shruthi, V. Nagaveni, and B. K. Raghavendra, A Review on Machine Learning Classification Techniques for Plant Disease Detection 2019, 2019 5th Int. Conf. Adv. Comput. Commun. Syst., 281–4.

4. Tomato Leaf Disease Classification Model by – Umar Pirzada
https://www.kaggle.com/code/umairshahpirzada/tomato-leaf-disease-classification-mobilenet

5. Leaf Disease Model by – Amar Nassanlhajali
https://www.kaggle.com/code/ammarnassanalhajali/tomato-leaf-disease-detection-0-998-training

6. Suresh V, Gopinath D, Hemavarthini M, Jayanthan K, and Mohana Krishnan, Plant disease using image processing 2020, IJERT, 9, 78-82.

7. S. Kaur, S. Pandey, and S. Goel, Plants Disease Identification and Classification Through Leaf Images: A Survey 2019, Arch. Comput. Methods Eng., 26, 507–30.