Mini Project Report

on

"Plant Disease Detection System"

By:

Alekh Taori (Reg. No. 202000423)

Jeewan Sharma (Reg. No. 202000381)

Aman Kumar Gupta (Reg. No. 202000547)

Group Id – C16

In partial fulfillment of requirements for the award of degree in

Bachelor of Technology in Computer Science and Engineering

(2023)

Under the Project Guidance of

Mr. BIRAJ UPADHYAYA

ASSISTANT PROFESSOR-I

Sikkim Manipal Institute of Technology, Majitar

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SIKKIM MANIPAL INSTITUTE OF TECHNOLOGY

(A constituent college of Sikkim Manipal University)

MAJITAR, RANGPO, EAST SIKKIM – 737136

PROJECT COMPLETION CERTIFICATE

This is to certify that the below mentioned students of Sikkim Manipal Institute of Technology have worked under my supervision and guidance from **9th January 2023 to 29th April 2023** and successfully completed the Mini project entitled **"Plant Disease Detection System"** in partial fulfillment of the requirements for the award of Bachelor of Technology in Computer Science and Engineering.

University Registration No	Name of Student	Course
202000423	**Alekh Taori**	**B.Tech (CSE)**
202000381	**Jeewan Sharma**	**B.Tech (CSE)**
202000547	**Aman Kumar Gupta**	**B.Tech (CSE)**

Biraj

Mr. Biraj Upadhyaya

Assistant Professor-I

Department of Computer Science and Engineering

Sikkim Manipal institute of Technology

Majhitar, Sikkim – 737136

PROJECT REVIEW CERTIFICATE

This is to certify that the work recorded in this project report entitled **"Plant Disease Detection System"** has been jointly carried out by **Alekh Taori (Reg. 202000423), Jeewan Sharma (Reg. 202000381) and Aman Kumar Gupta (Reg. 202000547)** of Computer Science & Engineering Department of Sikkim Manipal Institute of Technology in partial fulfillment of the requirements for the award of Bachelor of Technology in Computer Science and Engineering. This report has been duly reviewed by the undersigned and recommended for final submission for Mini Project Viva Examination.

Biraj

Mr. Biraj Upadhyaya

Assistant Professor-I

Department of Computer Science and Engineering

Sikkim Manipal Institute of Technology

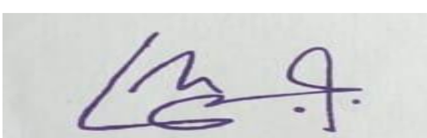
Majhitar, Sikkim – 737136

CERTIFICATE OF ACCEPTANCE

This is to certify that the below mentioned students of Computer Science & Engineering Department of Sikkim Manipal Institute of Technology (SMIT) have worked under the supervision of **Mr. Biraj Upadhyaya**, Assistant Professor-I, Department of Computer Science and Engineering from **9th January 2023 to 29th April 2023** on the project entitled **“Plant Disease Detection System”.**

The project is hereby accepted by the Department of Computer Science & Engineering, SMIT in partial fulfillment of the requirements for the award of Bachelor of Technology in Computer Science and Engineering.

University Registration No	Name of Student	Project Venue
202000423 **202000381** **202000547**	**Alekh Taori** **Jeewan Sharma** **Aman Kumar Gupta**	**SMIT**



Dr. Udit Kumar Chakraborty

Professor & Head of the Department

Computer Science & Engineering Department

Sikkim Manipal Institute of Technology

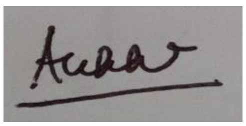
Majhitar, Sikkim – 737136

DECLARATION

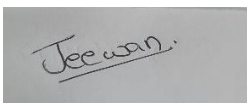
We, the undersigned, hereby declare that the work recorded in this project report entitled "**Plant Disease Detection System**" in partial fulfillment for the requirements of award of B.Tech (CSE) from Sikkim Manipal Institute of Technology (A constituent college of Sikkim Manipal University) is a faithful and bonafide project work carried out at "**SIKKIM MANIPAL INSTITUTE OF TECHNOLOGY**" under the supervision and guidance of **Mr. Biraj Upadhyaya,** Assistant Professor- I, Department of Computer Science and Engineering.

The results of this investigation reported in this project have so far not been reported for any other Degree or any other Technical forum.

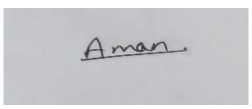
The assistance and help received during the course of the investigation have been duly acknowledged.



Alekh Taori (202000423)



Jeewan Sharma (202000381)



Aman Kumar Gupta (202000547)

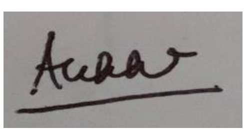
ACKNOWLEDGMENT

We take this opportunity to acknowledge indebtedness and a deep sense of gratitude to our guide **Mr. Biraj Upadhyaya** for his valuable guidance and supervision throughout the course which shaped the present work as it shows.

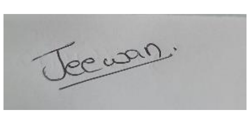
We pay our deep sense of gratitude to **Prof. (Dr.) Udit Kumar Chakraborty, HOD, Computer Science & Engineering Department, Sikkim Manipal Institute of Technology** for giving us the opportunity to work on this project and providing all support required.

We are obliged to our project coordinators **Dr. Sandeep Gurung** and **Mr. Biraj Upadhyaya** for elevating, inspiration and supervising in completion of our project.

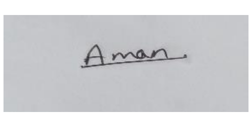
We would also like to thank any other staff of **Computer Science & Engineering Department, Sikkim Manipal Institute of Technology** for giving us continuous support and guidance that has helped us in completion of our project.



Alekh Taori (202000423)



Jeewan Sharma (202000381)



Aman Kumar Gupta (202000547)

DOCUMENT CONTROL SHEET

1	Report No	CSE/Mini Project/Internal/B.Tech/C16/2023
2	Title of the Report	Plant Disease Detection System
3	Type of Report	Technical
4	Author	Alekh Taori, Jeewan Sharma, Aman Kumar Gupta
5	Organizing Unit	Sikkim Manipal Institute of Technology
6	Language of the Document	English
7	Abstract	Plant disease detection using computer vision techniques is an important tool in agriculture to prevent crop losses and ensure food security.
8	Security Classification	General
9	Distribution Statement	General

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Plant disease detection using deep learning is a method of using artificial intelligence and machine learning techniques to identify and diagnose different types of diseases that affect plants. This approach typically involves the use of convolutional neural networks (CNNs) or other deep learning algorithms to analyze images of plant leaves, fruits, or stems and identify the presence of disease symptoms. The goal of this method is to improve the accuracy and speed of plant disease detection, while also reducing the need for human involvement in the process. Some of the benefits of this approach include the ability to detect diseases at an early stage, improved accuracy, and the ability to process large amounts of data quickly. Additionally, using deep learning techniques for plant disease detection enables the ability to detect diseases that are visually similar and hard to distinguish by humans, thus improving the overall efficiency of disease management in agriculture

1. INTRODUCTION

Plant Disease Detection is one of the mind boggling issue that exits when we talk about using Technology in Agriculture.Although researches has been done to detect weather a plant is healthy or diseased using Deep Learning and with the help of Neural Network, new techniquies are still being discovered. Plant diseases and pests are important factors determining the yield and quality of plants. Plant diseases and pests identification can be carried out by means of digital image processing. In recent years, deep learning has made breakthroughs in the field of digital image processing, far superior to traditional methods.

Plant diseases and pests detection is a very important research content in the field of machine vision. It is a technology that uses machine vision equipment to acquire images to judge whether there are diseases and pests in the collected plant images . At present, machine vision-based plant diseases and pests detection equipment has been initially applied in agriculture and has replaced the traditional naked eye identification to some extent.

For traditional machine vision-based plant diseases and pests detection method, conventional image processing algorithms or manual design of features plus classifiers are often used. This kind of method usually makes use of the different properties of plant diseases and pests to design the imaging scheme and chooses appropriate light source and shooting angle, which is helpful to obtain images with uniform illumination. Although carefully constructed imaging schemes can greatly reduce the difficulty of classical algorithm design, but also increase the application cost.

Deep learning is a type of machine learning that uses neural networks to model complex patterns and relationships in data. In the context of plant disease detection, deep learning can be used to analyze images of plants and identify the presence of specific diseases. This can be done by training a convolutional neural network (CNN) on a dataset of labeled plant images, where the labels indicate the presence or absence of a particular disease. The CNN can then be used to classify new images as healthy or diseased based on the patterns it has learned from the training data. One of the advantages of using deep learning for plant disease detection is that it can be highly accurate and can handle large amounts of data.

The work being propose is to detect whether a plant leaf is healthy or unhealthy by utilizing classical Machine Learning Algorithm , Pre-processing the data using Image Processing to achieve the maximum accuracy.

2. LITERATURE SURVEY

Sl. No.	**Author**	**Paper and Publication Details**	**Findings**	**Relevance**
1	Ramprasath, Muthukrishnan, M. Vijay Anand, and Shanmugasundaram Hariharan.	Image classification using convolutional neural networks *International Journal of Pure and Applied Mathematics, 119(17), 1307-1319. (2018)*	CNN models can achieve high accuracy in image classification tasks, such as the number of layers and filters.	The use of transfer learning when limited training data is available. Design and training of CNN models for plant disease detection.
2	Chohan **et. al.**	Plant disease detection using deep learning. *International Journal of Recent Technology and Engineering*, *9*(1), 909-914, *(2020).*	Transfer learning-based models outperformed CNN-based models in terms of accuracy and computational efficiency.	Effectiveness of deep learning techniques, particularly transfer learning-based models, can significantly improve the accuracy of the models.
3	Ahmad, Aanis, Dharmendra Saraswat.	A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. *Smart Agricultural Technology (2022): 100083.*	Ensemble learning, where multiple models are combined, Explainable AI techniques can help in interpreting the results of deep learning models.	Uses of techniques for plant disease diagnosis and selecting appropriate techniques, preprocessing strategies, and developing user-friendly interfaces.

Table 1.1

3. PROBLEM DEFINITION

- **Limited dataset:** One of the major challenges in plant disease detection is the availability of limited labeled data. The limited availability of labeled data can arise due to various factors such as the cost and time required to collect and annotate large amounts of data, the difficulty of identifying and diagnosing different diseases, and the lack of standardized datasets.

- **Complex feature extraction:** Plant disease detection requires identifying complex patterns and textures in the images, which can be challenging for traditional computer vision techniques. It is a challenging task that requires the identification of complex patterns and textures in images of plants, which can be difficult to achieve using traditional computer vision techniques. This is because the symptoms of plant diseases can manifest in various ways, and the visual appearance of these symptoms can be highly variable depending on factors such as plant species, disease type, and environmental conditions.

4. SOLUTION STRATEGY

- **Data Loading :** Loading the image dataset into a computer and converting it into a format that is compatible with the machine learning framework being used. The dataset is then divided into three classes based on the type of plant disease being diagnosed.

- **Splitting Dataset :** The dataset is further split into training, validation, and testing subsets. The training subset is used to teach the CNN to recognize patterns and features associated with the different classes, while the validation subset is used to fine-tune the model and prevent overfitting. The testing subset is used to evaluate the accuracy of the model on new, unseen data.

- **Preprocessing :** It is the process of generating new training samples by applying various transformations to the original images, such as resizing and rescaling. This increases the diversity and quantity of the training data, which can help improve the robustness and generalization of the model.

- **Building Model :** CNN model is constructed using a combination of Conv2D and MaxPooling layers. Conv2D layers perform feature extraction by applying a set of filters to the input image, while MaxPooling layers downsample the feature maps to reduce the spatial dimensions and preserve the most salient features.

- **Compiling Model :** the model is compiled using a specific optimizer, such as Adam or SGD, a loss function, such as categorical cross-entropy, and metrics, such as accuracy or F1-score, to assess the model's performance.

- **Training Model :** the training and validation subsets using backpropagation and stochastic gradient descent to adjust the model's parameters and minimize the loss function.

- **UI implementation and I/O:** Finally the development of a user-friendly interface that enables users to input images of plant leaves and receive a corresponding diagnosis of whether the leaves are healthy or affected by a disease. The interface should allow users to upload images in various formats, such as JPEG, PNG, and provide feedback on the diagnosis, such as the confidence score and the type of disease.

5. DESIGN

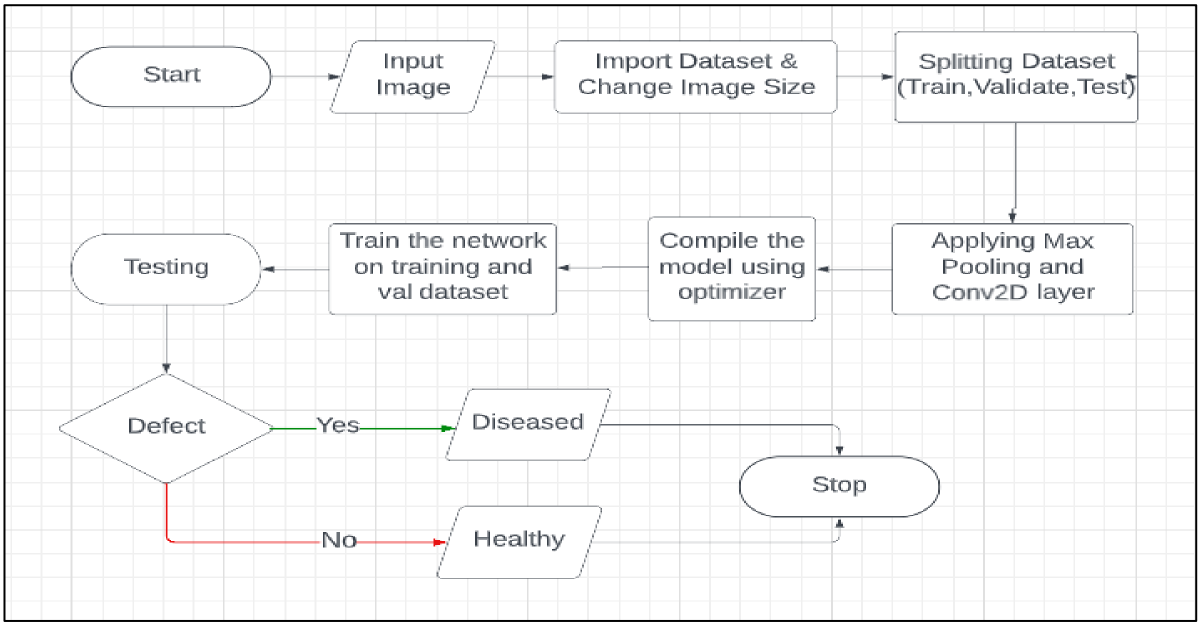


Fig 5.1: Block diagram of the model

- **Input Image:** The image of a plant that needs to be tested for disease detection.
- **Import dataset and change image size:** Importing the dataset and resizing the images to a standard size for consistency in input.
- **Splitting dataset:** Dividing the dataset into training, validation, and testing sets for training, hyperparameter tuning, and evaluation purposes.
- **Applying max pooling and conv2d layer:** Using convolutional and max pooling layers in a CNN to extract features from images and reduce computational cost.
- **Compile model using optimizer:** Compiling the network with an optimizer to minimize the loss function during training.
- **Train the network:** Training the compiled network on the training set to adjust the weights and learn the features of healthy and diseased plants.
- **Testing:** Evaluating the trained model on the testing set to check its performance on unseen data.
- **Detect:** Using the trained and tested model to predict whether a new plant is healthy or diseased based on its image.

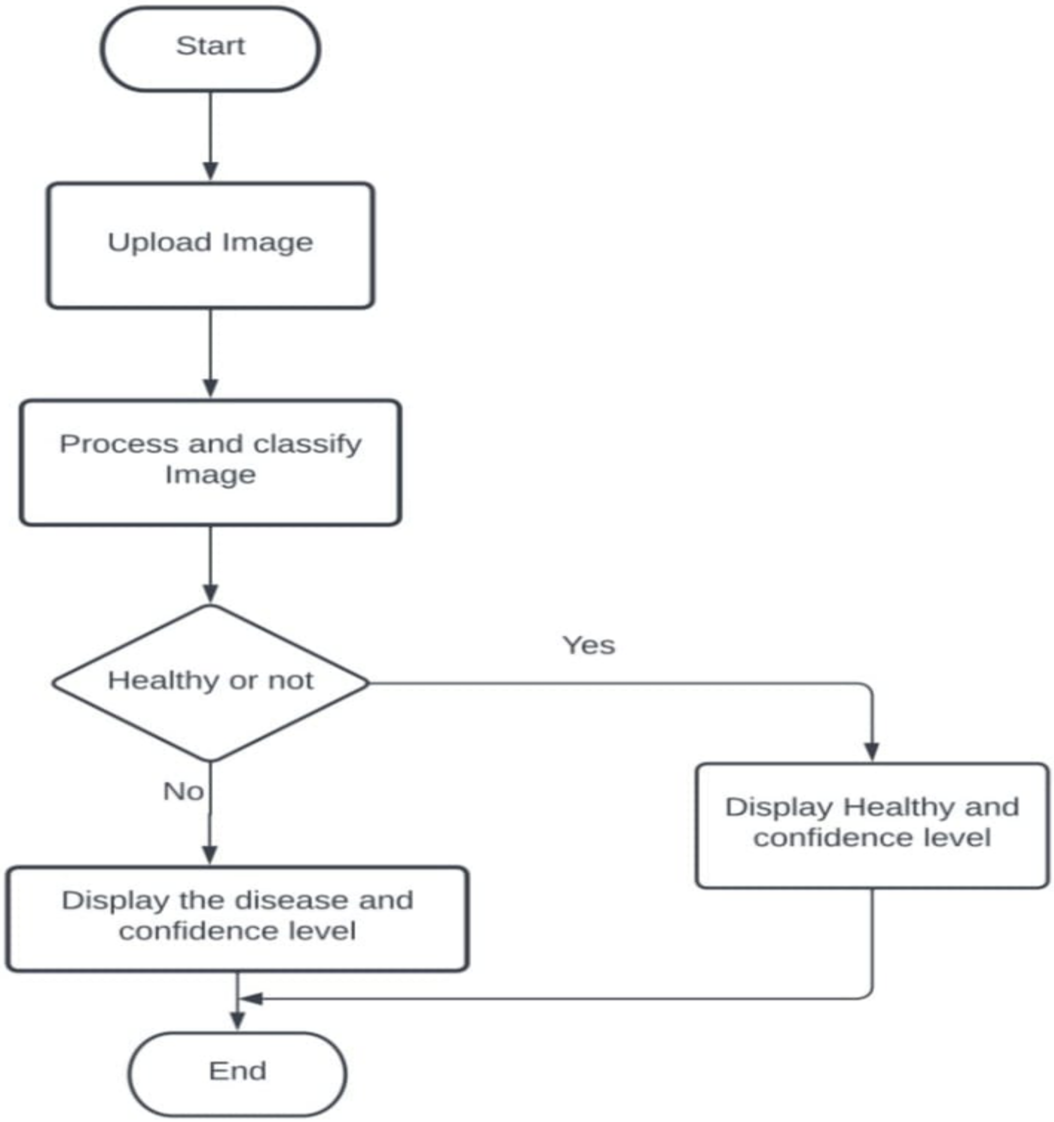


Fig 5.2: Flowchart for the proposed system

6. IMPLEMENTATION DETAILS

Our project aimed to develop a plant disease detection system using convolutional neural networks (CNN). We started with a dataset of images that were labeled as healthy or diseased with either early blight or late blight. To preprocess the data, we applied various data augmentation techniques, such as resizing, rescaling, and horizontal and vertical flipping. This helped to increase the dataset size and balance the number of images in each class.

Next, we designed a CNN architecture with multiple convolutional and pooling layers, followed by dense layers. We used ReLU activation functions and applied dropout regularization to prevent overfitting. We also used a softmax activation function in the output layer to classify the images into the respective classes.

After designing the architecture, we compiled the model using appropriate optimizers, metrics, and loss functions. We then trained the network on the training dataset and evaluated its performance on the validation dataset. We repeated this process several times, adjusting the hyperparameters as needed, until we achieved satisfactory performance.

Finally, we ran the trained model on the test dataset and evaluated its accuracy, precision, recall, and F1 score. Our results showed that the CNN-based plant disease detection system achieved high accuracy and could correctly classify the images into the respective classes. Our project demonstrates that using CNNs and appropriate preprocessing techniques can improve the accuracy and reliability of plant disease detection systems, which can have significant implications for the agriculture industry.

Fig 6.1: Early Blight dataset

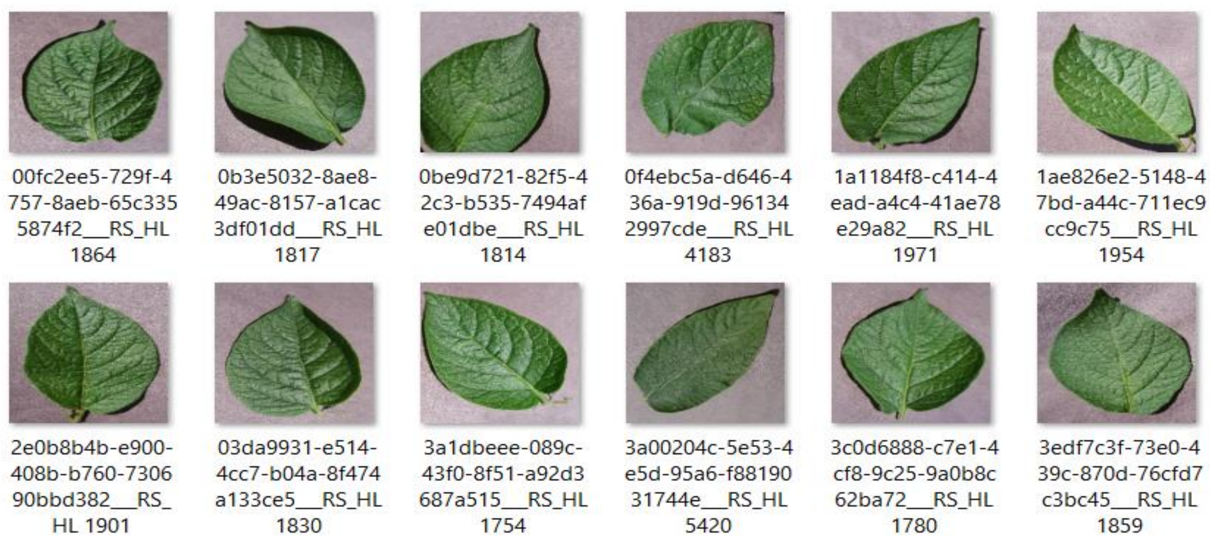


Fig 6.2: Healthy dataset

Fig 6.3: Late Blight dataset

```
In [38]: def get_dataset_partitions_tf(ds, train_split=0.8, val_split=0.1, test_split=0.1, shuffle=True, shuffle_size=10000):
             assert (train_split + test_split + val_split) == 1

             ds_size = len(ds)

             if shuffle:
                 ds = ds.shuffle(shuffle_size, seed=12)

             train_size = int(train_split * ds_size)
             val_size = int(val_split * ds_size)

             train_ds = ds.take(train_size)
             val_ds = ds.skip(train_size).take(val_size)
             test_ds = ds.skip(train_size).skip(val_size)

             return train_ds, val_ds, test_ds

In [39]: train_ds,val_ds,test_ds = get_dataset_partitions_tf(dataset)

In [40]: len(train_ds)
Out[40]: 54

In [41]: len(val_ds)
Out[41]: 6

In [42]: len(test_ds)
Out[42]: 8

In [43]: len(train_ds)
Out[43]: 54
```

Fig 6.4: Code to Split the dataset

```
In [47]: input_shape = (BATCH_SIZE, IMAGE_SIZE, IMAGE_SIZE, CHANNELS)
         n_classes = 3

         model = models.Sequential([
             resize_and_rescale,
             layers.Conv2D(32, kernel_size = (3,3), activation='relu', input_shape=input_shape),
             layers.MaxPooling2D((2, 2)),
             layers.Conv2D(64,  kernel_size = (3,3), activation='relu'),
             layers.MaxPooling2D((2, 2)),
             layers.Conv2D(64,  kernel_size = (3,3), activation='relu'),
             layers.MaxPooling2D((2, 2)),
             layers.Conv2D(64, (3, 3), activation='relu'),
             layers.MaxPooling2D((2, 2)),
             layers.Conv2D(64, (3, 3), activation='relu'),
             layers.MaxPooling2D((2, 2)),
             layers.Conv2D(64, (3, 3), activation='relu'),
             layers.MaxPooling2D((2, 2)),
             layers.Flatten(),
             layers.Dense(64, activation='relu'),
             layers.Dense(n_classes, activation='softmax'),
         ])

         model.build(input_shape=input_shape)
```

Fig 6.5: Code to Build the model

Epoch	Loss	Accuracy
1	0.9066	0.4919
2	0.6552	0.6956
3	0.3942	0.8490
4	0.2941	0.8693
5	0.2167	0.8764
6	0.1787	0.8788
7	0.1745	0.8793
8	0.1698	0.8832
9	0.1341	0.8846
10	0.0754	0.8863
18	0.0137	0.8921
19	0.0040	0.8949
20	0.0011	0.9041

Table-6.1: Loss and Accuracy

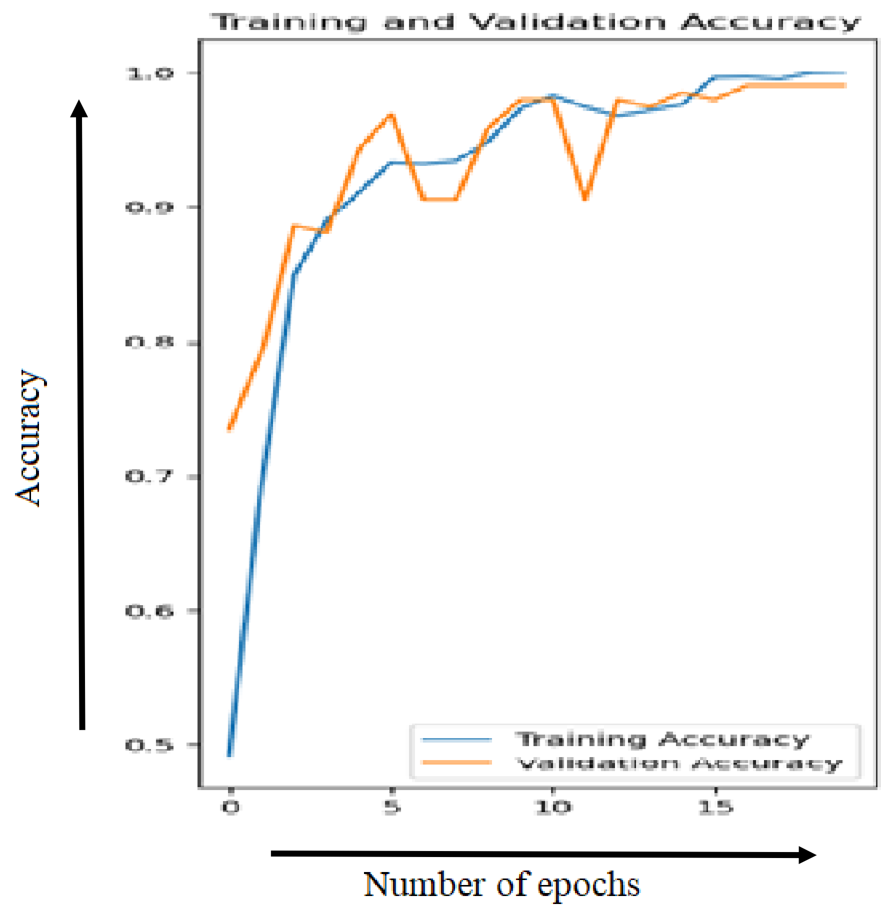
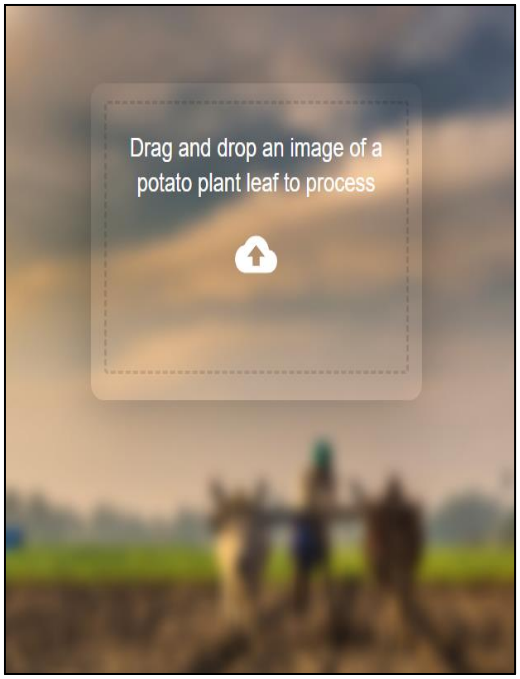


Fig-6.6: Training and Validation Accuracy



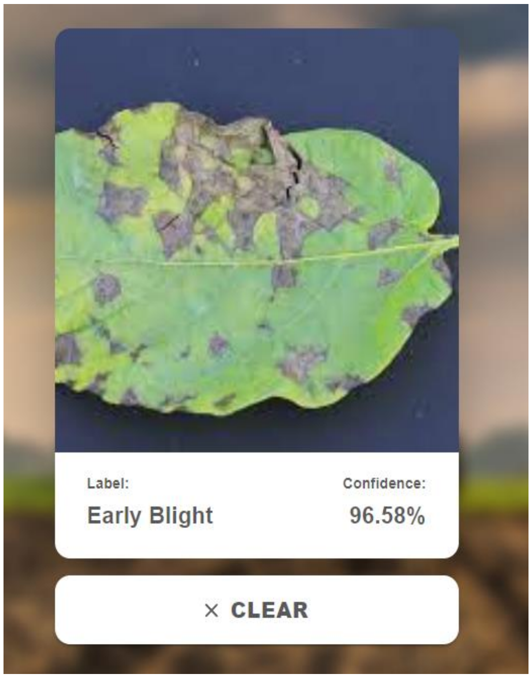
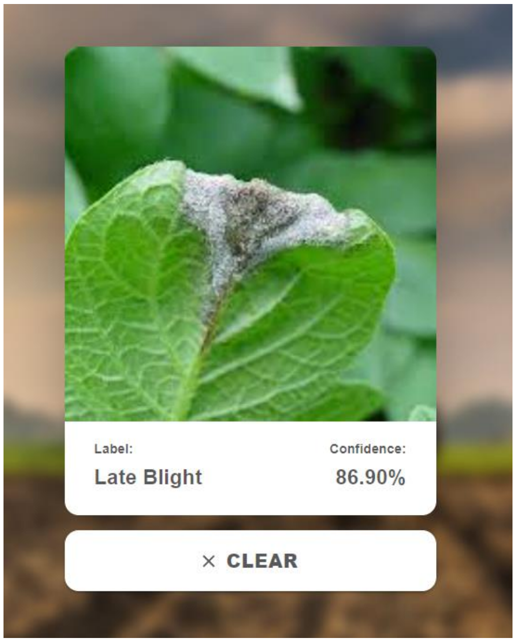


Fig-6.7: Input From User

Fig-6.8: Output Showing Early Blight



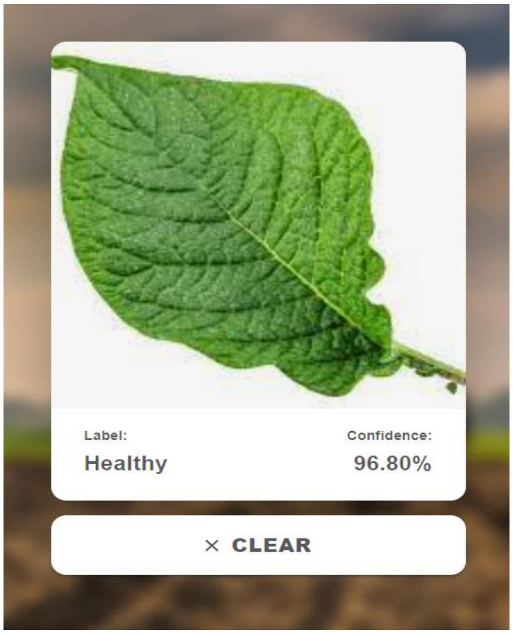


Fig-6.9: Output Showing Late Blight

Fig-6.10: Output Showing Healthy

7. RESULTS AND DISCUSSIONS

The system implements a convolutional neural network (CNN) for image classification using TensorFlow and Keras. The model is trained to classify plant diseases using the PlantVillage dataset. The dataset consists images of diseased and healthy plants, belonging to 3 classes. The model architecture is defined using the Sequential API of Keras, which allows the easy stacking of layers. The model consists of several convolutional layers with increasing depth, followed by max-pooling layers to reduce the spatial dimensions of the feature maps. The output of the last convolutional layer is flattened and fed into two fully connected (dense) layers, with ReLU activation functions, followed by a final output layer with a softmax activation function to produce the probability distribution over the 3 classes.

The input images are preprocessed using a data augmentation pipeline consisting of random horizontal and vertical flips, and random rotations of up to 20 degrees. The images are also resized to a uniform size of 256x256 pixels and rescaled to have pixel values in the range [0, 1]. The model is trained using the Adam optimizer and the sparse categorical cross-entropy loss function. The performance of the model is evaluated using accuracy and loss metrics, computed over the validation and test sets.

The training history of the model shows that it achieves a training accuracy of over 85% and a validation accuracy of around 90%, after 20 epochs of training. The training loss decreases rapidly in the first few epochs, but then plateaus, indicating that the model may have reached its optimal performance. The validation loss shows a similar pattern, indicating that the model may be slightly overfitting the training data. However, the validation accuracy is consistently high, indicating that the model generalizes well to new, unseen data.

8. CONCLUSION

In conclusion, the implemented Convolutional Neural Network (CNN) model for plant disease classification using the Plant Village dataset demonstrates promising results. During training, the model achieves high accuracy on both the training and validation sets, with a training accuracy of over 85% and a validation accuracy of around 90%. This indicates that the model successfully learns the patterns and characteristics of different plant diseases, enabling accurate classification. However, it is worth noting that the model's performance on the validation set plateaus after a certain number of epochs, suggesting that further training may not significantly improve its performance.While the model demonstrates good generalization to unseen data, there is a slight indication of overfitting, as the training and validation loss start to diverge after a few epochs. This implies that the model may be too complex for the given dataset, and regularization techniques such as dropout or weight decay could be beneficial in mitigating overfitting and improving generalization performance.Overall, the implemented CNN model serves as a solid foundation for plant disease classification. With continued refinement, including fine-tuning hyperparameters, applying regularization techniques, expanding the dataset, and exploring transfer learning, the model's accuracy and generalization capabilities can be further improved, thereby contributing to more accurate and efficient plant disease diagnosis and management systems.

9. LIMITATIONS AND FUTURE SCOPE OF THE PROJECT

The future scope of plant disease detection using CNN is vast as it can help farmers and researchers to detect diseases in plants accurately and in a timely manner. With the increasing demand for food production, the use of advanced technology in agriculture is crucial to prevent crop losses due to diseases.

However, there are certain limitations to the model presented above. One of the limitations is the lack of diversity in the dataset. The model is trained on a specific dataset "PlantVillage" which contains only three classes of plant diseases. To make the model more robust and generalizable, a larger and diverse dataset should be used that includes various types of plants and diseases.

Another limitation is the computational power required to train the model. The presented model requires a lot of computing resources to train, and this can be a challenge for small-scale farmers and researchers who may not have access to high-end machines. Therefore, more research is needed to develop lightweight models that can be trained and deployed on low-end devices such as mobile phones.

In conclusion, the system proposed is a good starting point for detecting plant diseases using CNN, but further research is needed to overcome the limitations and to develop a more robust and generalizable model.

10. GANTT CHART

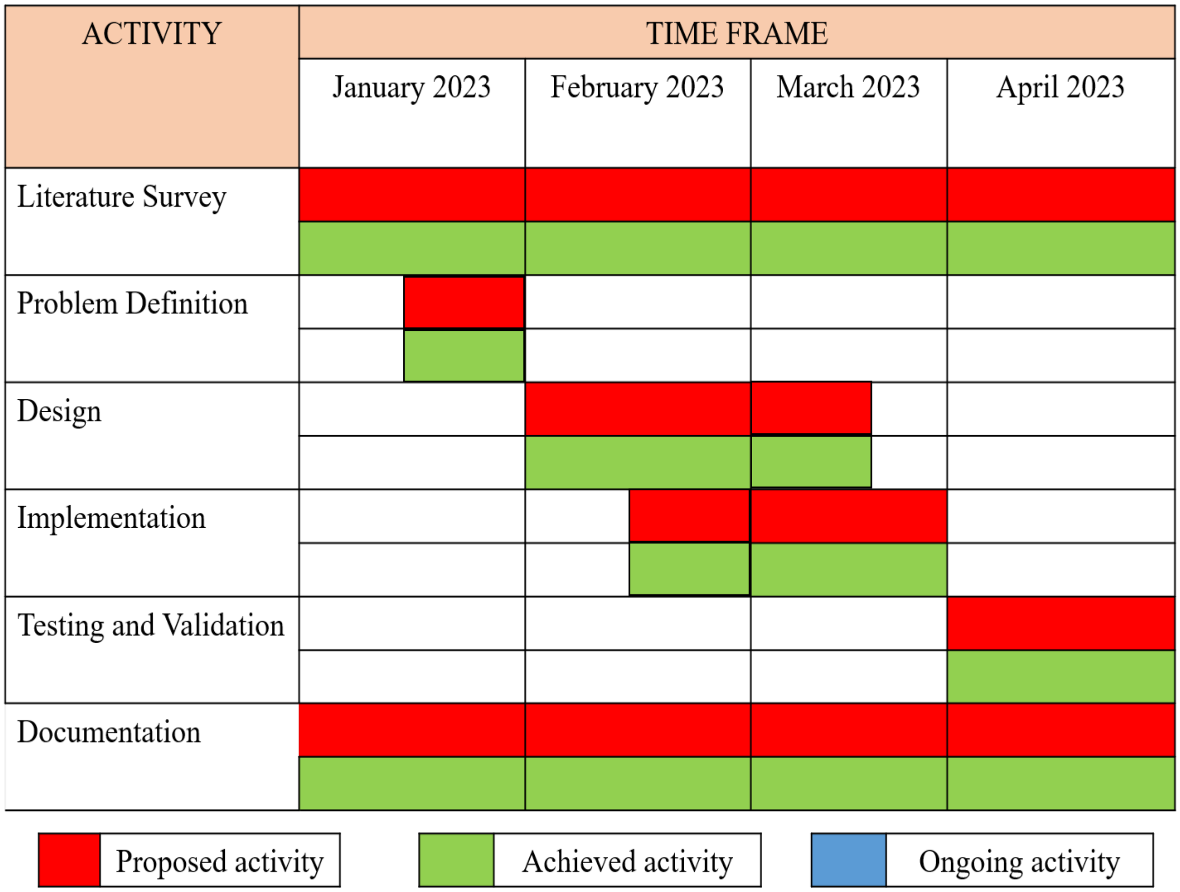


Fig 10.1: Gantt Chart

11.REFERENCES

[1] Chohan, Murk, et al. "Plant disease detection using deep learning." *International Journal of Recent Technology and Engineering* 9.1 (2020): 909-914.

[2] Ramprasath, Muthukrishnan, M. Vijay Anand, and Shanmugasundaram Hariharan. "Image classification using convolutional neural networks." *International Journal of Pure and Applied Mathematics* 119.17 (2018): 1307-1319.

[3] Ahmad, Aanis, Dharmendra Saraswat, and Aly El Gamal. "A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools." *Smart Agricultural Technology* (2022): 100083.

[4] Ferentinos, Konstantinos P. "Deep learning models for plant disease detection and diagnosis." *Computers and electronics in agriculture* 145 (2018): 311-318.

[5] Shruthi, U., V. Nagaveni, and B. K. Raghavendra. "A review on machine learning classification techniques for plant disease detection." *2019 5th International conference on advanced computing & communication systems (ICACCS)*. IEEE, 2019.

[6] Mahum, Rabbia, et al. "A novel framework for potato leaf disease detection using an efficient deep learning model." *Human and Ecological Risk Assessment: An International Journal* 29.2 (2023): 303-326.

[7] Binnar, Vikki, and Sanjeev Sharma. "Plant Leaf Diseases Detection Using Deep Learning Algorithms." *Machine Learning, Image Processing, Network Security and Data Sciences: Select Proceedings of 3rd International Conference on MIND 2021*. Singapore: Springer Nature Singapore, 2023.

Mini Project C-16

ORIGINALITY REPORT

13%	8%	8%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	"International Conference on Innovative Computing and Communications", Springer Science and Business Media LLC, 2022 Publication	2%
2	Submitted to University of Westminster Student Paper	2%
3	www.philstat.org.ph Internet Source	1%
4	Submitted to Harrisburg University of Science and Technology Student Paper	1%
5	Submitted to Berlin School of Business and Innovation Student Paper	1%
6	ijournalse.org Internet Source	1%
7	www.mdpi.com Internet Source	1%
8	Submitted to University of Florida Student Paper	1%

9	Submitted to Saveetha Dental College and Hospital, Chennai Student Paper	1 %
10	studenttheses.uu.nl Internet Source	1 %
11	www.slideshare.net Internet Source	<1 %
12	L Jason Anastasopoulos, Andrew B Whitford. "Machine Learning for Public Administration Research, With Application to Organizational Reputation", Journal of Public Administration Research and Theory, 2019 Publication	<1 %
13	Pallavi M, Thivakaran T K, Chandankeri Ganapathi. "Evaluation of Land Use/Land Cover Classification based on Different Bands of Sentinel-2 Satellite Imagery using Neural Networks", International Journal of Advanced Computer Science and Applications, 2022 Publication	<1 %
14	E. Saraswathi, J. FarithaBanu. "A Novel Ensemble Classification Model for Plant Disease Detection Based on Leaf Images", 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), 2023 Publication	<1 %