

P.E.S. COLLEGE OF ENGINEERING, MANDYA 571401

(An Autonomous & Govt. Aided Institution, Affiliated to VTU,
Belagavi)



A Project Report on

***“A MACHINE LEARNING BASED
PREDICTION OF TOMATO PLANT LEAF
DISEASE USING CNN”***

Submitted in partial fulfilment for the award of the degree of

BACHELOR OF ENGINEERING

in

INFORMATION SCIENCE & ENGINEERING

Submitted by

Manoj Gowda B K [4PS20IS019]

Under the Guidance of

Rakshith N

Assistant Professor

Dept of ISE PESCE, Mandya

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING

2022-23

P.E.S. COLLEGE OF ENGINEERING, MANDYA

(An Autonomous & Govt. Aided Institution, Affiliated to VTU, Belagavi)

Department of Information Science & Engineering



CERTIFICATE

This is to certify that, **Mr. Manoj Gowda B K, [4PS20IS019]** have been successfully completed the Project work (P18ISL66) entitled “**A machine Learning based Prediction of tomato Plant Leaf Disease using CNN**” in partial fulfilment for the award of Bachelor of Engineering in Information Science and Engineering of P.E.S. College of Engineering, Mandya (An Autonomous Institution, Affiliated to VTU, Belagavi) during the academic year 2022-2023. It is certified that, all corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the departmental library. The Project report has been approved as it satisfies the academic requirements in respect of Project work prescribed for the Degree in Bachelor of Engineering.

Signature of Guide

Rakshith N

Assistant professor

Dept. of IS&E

PESCE, Mandya

Signature of Head of the Department

Dr. Anitha M L

Professor and HoD

Dept. of IS&E

PESCE, Mandya

Details of Project work Viva Voce Examination

Sl. No.	Name of the Examiners	Signature	Date
1			
2			

ACKNOWLEDGEMENT

Engineering Project makes the transformation mere theoretical knowledge into visible reality. Our report acknowledges some guidance, some supervision, and a lot of inspiration. It is time now to acknowledge our obligation to all who have extended their co-operation all along with our study tenure of project work.

We take this pleasure in thanking, our beloved Principal **Dr. R M MAHALINGE GOWDA**, PES College of Engineering, Madya, who has always been a great source of inspiration and has encouraged us all through.

We wish to express our deep sense of gratitude to our project guide **Dr. ANITHA M L**, Professor and head of Department of Information Science & Engineering for her guidance and useful suggestions throughout our project work and we are thankful for his time, effort and also for her continued encouragement in preparing this report.

We thank one and all who have helped us directly or in-directly in completing our project work.

Manoj Gowda B K

[4PS20IS019]



P. E. S. COLLEGE OF ENGINEERING, MANDYA

(An Autonomous Institute under VTU, Belagavi)

Department of Information Science and Engineering

DECLARATION

I am, **Manoj Gowda B K**

students of 6th semester Bachelor of Engineering in Information Science and Engineering at P.E.S. College of Engineering, Mandya, hereby declare that the work being presented in the Project Work entitled “**A MACHINE LEARNING BASED PREDICTION OF TOMATO PLANT LEAF DISEASE USING CNN**” is an authentic record of the work that has been independently carried out by us and submitted in partial fulfilment of the requirement for the award of degree in Bachelor of Engineering in Information Science and Engineering, P.E.S. College of Engineering, affiliated to Visvesvaraya Technological University, Belagavi during the academic year 2022 – 2023. The work contained in this report has not been submitted in part or full to any other university or institution or professional body for the award of any other degree or any fellowship.

Place: Mandya

Manoj Gowda B K

[4PS20IS019]

Date :

TABLE OF CONTENTS

Chapter	Title	Page No.
1	Introduction	1
2	Literature Survey	2
3	Requirement	3
4	Design and Implementation	4
5	Results	9
6	Conclusion and future enhancement	16
7	References	17

CHAPTER 1

INTRODUCTION

Tomato plants are one of the most widely cultivated and economically important crops globally. However, they are highly susceptible to various diseases, which can cause substantial yield losses and affect crop quality. Early detection and accurate prediction of these diseases are crucial for implementing timely intervention strategies and minimizing the impact on tomato production.

In recent years, machine learning algorithms, particularly convolutional neural networks (CNNs), have emerged as powerful tools for image recognition and classification tasks. CNNs are designed to automatically learn and extract relevant features from images, making them well-suited for analyzing and diagnosing plant diseases based on leaf images. By leveraging the ability of CNNs to detect subtle patterns and variations in leaf characteristics, it becomes possible to develop accurate and efficient disease prediction models.

This study focuses on the application of CNNs in the context of tomato plant disease prediction. The primary objective is to leverage the power of CNNs to train a predictive model capable of accurately identifying and classifying tomato plant diseases based on leaf images. By analyzing a large dataset of healthy and diseased tomato plant leaf images, the CNN model can learn the intricate patterns and unique features associated with different diseases.

The utilization of machine learning algorithms, particularly CNNs, in the field of plant disease prediction holds great potential for revolutionizing disease management practices in agriculture. Accurate and early detection of diseases can lead to timely intervention, reducing the reliance on broad-spectrum pesticides and minimizing crop losses. Additionally, the automated and scalable nature of the CNN approach allows for efficient disease monitoring in large-scale tomato cultivation.

CHAPTER 2

LITERATURE SURVEY

[1] H.Al-Hiary et al, proposed a plant disease detection method, using the k-means clustering algorithm with Neural Network. Both the detection and classification of plant diseases can be identified by this trained model. It provides a precise accuracy. Merits include precise disease detection with less computational effort. Demerits are recognition rate is found to be declined

[2] Dheeb Al Bashish et al, proposed K-means clustering for segmentation and ANN for disease detection and classification of leaf disease. This proposed model recognized the diseases with moderate accuracy. Merits include, very effective in recognizing diseases. Demerits include, finer segmentation and feature extraction is required.

[3] Usama Mokhtar et al, proposed SVM with different kernel functions to detect the tomato leaves diseases with an excellent annotation. Merits include effective and reliable results. Demerits include a large scale of inputs that leads to a decrease in performance.

CHAPTER 3

REQUIREMENTS

3.1 HARDWARE REQUIREMENTS

- 4GB RAM or above
- Intel core i3/ AMD Ryzen 3 and above
- 32 bit system or above

3.2 SOFTWARE REQUIREMENTS

- Programming Language :Python v 3.10.8
- IDE :Jupyter notebook or Visual Studio Code
- Python packages :numpy==1.18.4, pandas==2.0.2, scikit learn ==1.1, tensor
flow,keras, flask

CHAPTER 4

IMPLEMENTATION

4.1 SELECTION OF METHODS AND METRICS:

Plant disease detection can be performed using different classifiers and a multitude of techniques have been used in the past for this purpose. In this project, the classifiers that were used for performing the detection were the Convolution Neural Network. The evaluation of the afore- mentioned models was done using the following metrics: accuracy, precision, recall and f1-score. classifier was evaluated using the same four evaluation metrics and the results of the classifiers was used to find the best performing model on the disease detection of tomatoleaves from the plant village dataset.

4.2 DATASET:

The dataset chosen for this paper contains 10,000 images of tomato leaves with 10 different categories. Each category contains the same number of images. The images of plant diseases are of the size 256x256 and the dataset does not contain any missing images. The dataset is commonly referred to as the plant village dataset and can be found on the public website, Mendeley data 1.











Bacterial Spot	Early Blight	Healthy	Late Blight	Leaf Mold
				
				
Septoria Leaf Spot	Spider Mites Two Spotted Spider Mite	Target Spot	Mosaic Virus	Yellow Leaf Curl Virus

Fig 1 - 10 different types of tomato leaf diseases

4.3 LIBRARIES USED:

Table 1 - Libraries used in the implementation of CNN.

NumPy	Is a library in python programming language that helps in managing large, multidimensional arrays and matrices. This library also enables the operation of mathematical functions on the created arrays or matrices
Cv2	Is an open source computer vision and machine learning library that is used to solve computer vision problems such as reading or resizing an image
os	It is a standard utility module that provides a portable way of using operating system-dependent functionality. The os and os.path modules include many functions to interact with the file system.
TnsorFlow-Keras	Is an open source neural network library that is designed to enable the fast implementation of deep neural networks with in the python interface itself
Matplotlib	Is a plotting library used in python for creating static and animated visualizations
Flask	Is a python framework to integrate the python code with front end website.

4.4 METHODOLOGY:

1. Importing the necessary libraries:

- The code starts by importing the required libraries, including TensorFlow.
- Then, it sets up the GPU configuration to allow for dynamic memory allocation during model training.

2. Creating the CNN model:

- The code initializes a sequential model using tensor flow function Sequential().
- The first layer added to the model is a convolutional layer (Conv2D) with 32 filters, a kernel size of 3x3, and the ReLU activation function.
- After the convolutional layer, a max-pooling layer (MaxPooling2D) with a pool size of 2x2 is added.
- Next, a second convolutional layer and max-pooling layer with the same

configurations are added.

- To prepare for the fully connected layers, the feature maps are flattened using Flatten().
- Two fully connected layers are added using Dense with 128 and 10 units, respectively. The ReLU activation function is used for the first dense layer, while the sigmoid activation function is used for the final layer.

3. Compiling the CNN model:

- The model is compiled using the Adam optimizer, categorical cross-entropy loss function, and accuracy as the evaluation metric.

4. Data preparation and augmentation:

- ImageDataGenerator is used to preprocess and augment the training and validation datasets.
- The rescale parameter scales the pixel values to the range [0, 1].
- Other augmentation techniques such as shear, zoom, and horizontal flip are applied to diversify the training data.

5. Loading the training and validation datasets:

- The flow_from_directory function is used to load the images from the specified directories. The images are resized to 128x128 pixels.
- The training and validation datasets are loaded separately, along with their corresponding labels.

6. Model training:

- The fit_generator function is used to train the model.
- The training_set is provided as the training data, and the valid_set is used for validation during training.
- The steps_per_epoch parameter defines the number of steps (batches) to be processed for each epoch.
- The epochs parameter specifies the number of training epochs.

7. Saving the trained model:

- The trained model is saved in two formats: as a JSON file (model1.json) and as a binary file (model.h5) that contains the weights and architecture.
- The saved model can be later loaded for making predictions on new images.

8. Importing the necessary libraries of flask:

- In leaf.py file the code starts by importing the Flask library, which is used for creating

the web application.

- Additionally, it imports other required libraries, including numpy and os.

9. Loading the trained model:

- The code loads the pre-trained model using the load_model function from TensorFlow.
- The filepath variable specifies the path to the saved model file ('model.h5').
- The loaded model is assigned to the model variable.
- The model is printed to confirm successful loading.

10. Flask web application setup:

- The code initializes the Flask application using Flask().
- The Flask application is responsible for handling web requests and rendering HTML templates.

11. Prediction function:

- The pred_tomato_disease function takes an image of a tomato plant as input for prediction.
- It loads the image using load_img and resizes it to (128, 128) pixels.
- The image is converted to a numpy array and normalized by dividing by 255.
- The model predicts the disease by calling model.predict on the preprocessed image.
- The predicted class index is obtained by finding the argmax of the prediction result.
- This prediction will be any one of the diseases; it can be "Tomato - Early Blight Disease", "Tomato - Healthy and Fresh", "Tomato - Late Blight Disease", "Tomato - Leaf Mold Disease", "Tomato - Septoria Leaf Spot Disease", "Tomato - Target Spot Disease", "Tomato - Tomato Yellow Leaf Curl Virus Disease", "Tomato - Tomato Mosaic Virus Disease", "Tomato - Two Spotted Spider Mite Disease".
- Based on the predicted class, the function returns the corresponding disease name and the HTML template associated with it.

12. Routing and HTML templates:

- The Flask application defines different routes for handling web requests.
- When a request is made to a specific route, the associated function is executed.
- Each route calls the pred_tomato_disease function with the uploaded image and renders the corresponding HTML template based on the returned disease name.

13. Explanation of prediction:

- The disease name and HTML template associated with the predicted class are returned to the client.⁷

- The disease name is printed to the console.
- The HTML template will be rendered on the web page to display the predicted disease and provide further information.

4.5 Flow chart

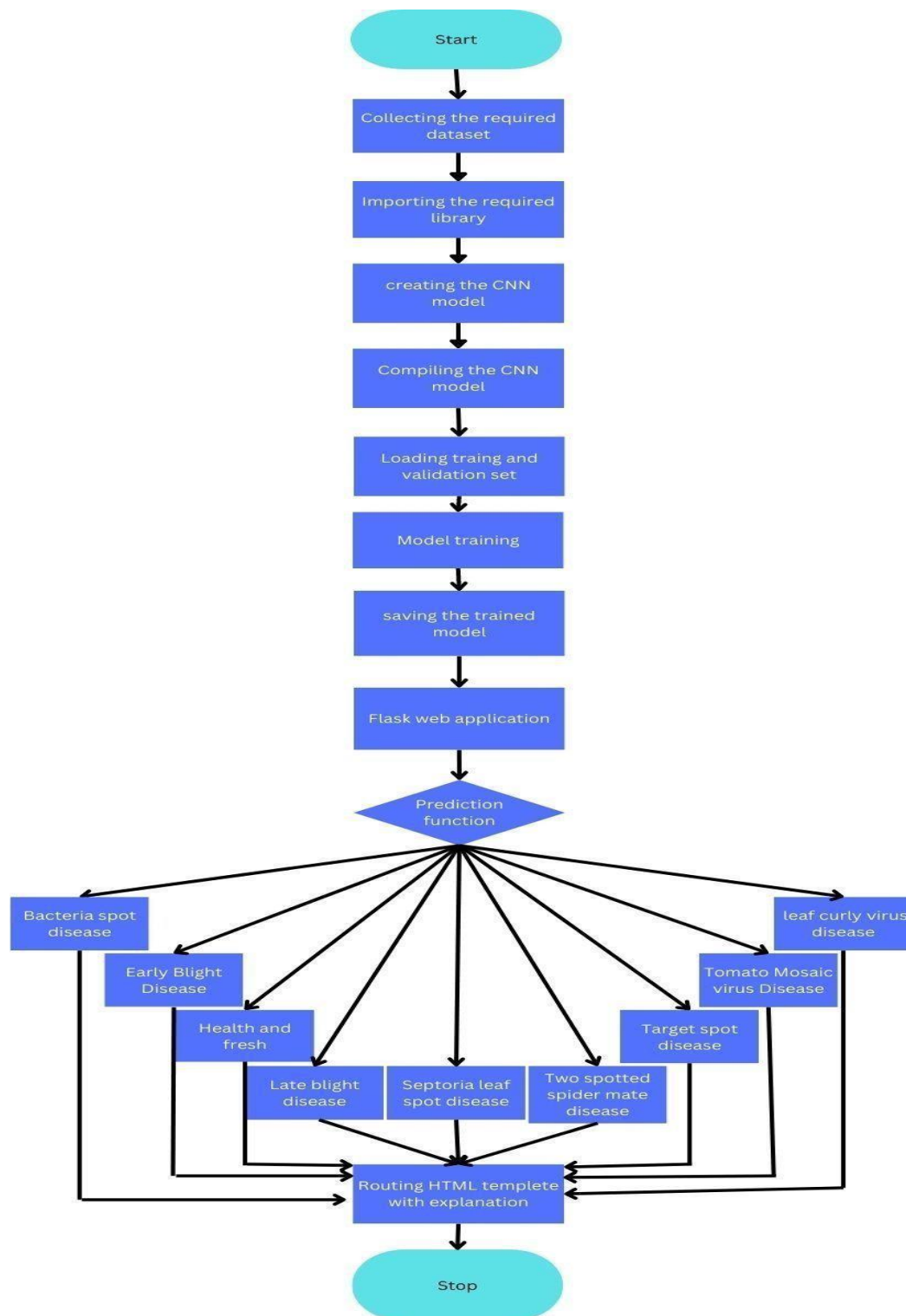


Fig 2: flow chart of the CNN model

CHAPTER 5

RESULTS

Our model was developed in Google Collaboratory, a free Python development environment that runs in the cloud. Google Collaboratory is widely used for the development of machine learning and deep learning projects. In our project, we use the following libraries: Tensorflow, an open-source library used for numerical computation and automated learning; Keras, a library used for the creation of neural networks; numpy, used for data analysis and mathematical calculations; matplotlib used for graph management and TensorBoard to visually inspect the different runs and graphs.

The model was trained with 200 epochs. We applied early stopping to monitor the performance of the model for the 200 epochs on a held-out validation set during the training to reduce overfitting and to improve the generalization of the neural network. For the evaluation of the model, the validation accuracy scheme allowed early stopping to be activated during the process.

Since our problem is a multi-class classification model, we use the Adam algorithm as the optimizing algorithm. In addition, the cross-entropy categorical loss function was used due to the nature of the multi-class classification environment. During the training process, we implemented checkpoints to save the model with the best validation accuracy, and thus be able to load it later to continue training from the saved state if necessary. Followings are the resultant page of the tomato plant leaf disease prediction using CNN application.

Project constraints:

1. only tomato plant leaf need to be given as input to the trained model.
2. if input image is other than tomato plant leaf, then model may predict wrongly.
3. prediction is not always correct.
4. only one tomato plant leaf image should be given as input.

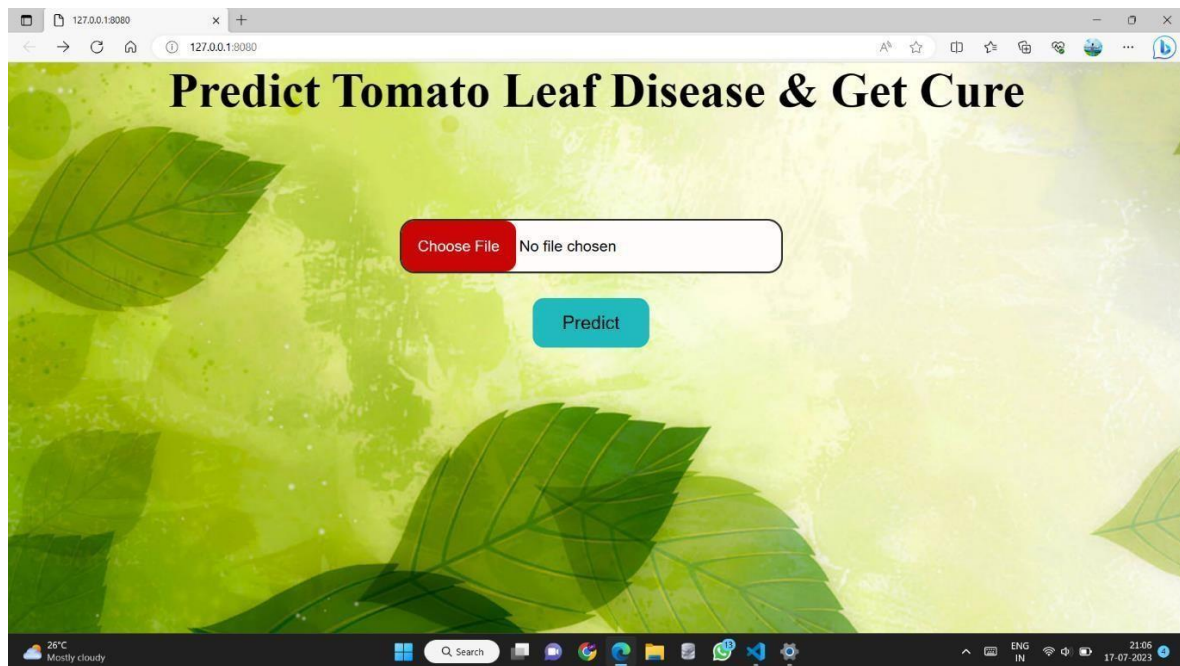


Fig 5.1: Home page of tomato plant leaf disease prediction application.

The tomato plant leaf disease prediction using CNN application is been developed by the HTML and CSS with using Flask framework, home page of the application is consisting header “Predict Tomato Leaf Disease & Get Cure” and also consist of input field to upload the photo of the diseased leaf of tomato to predict the kind of disease that leaf is been affected.

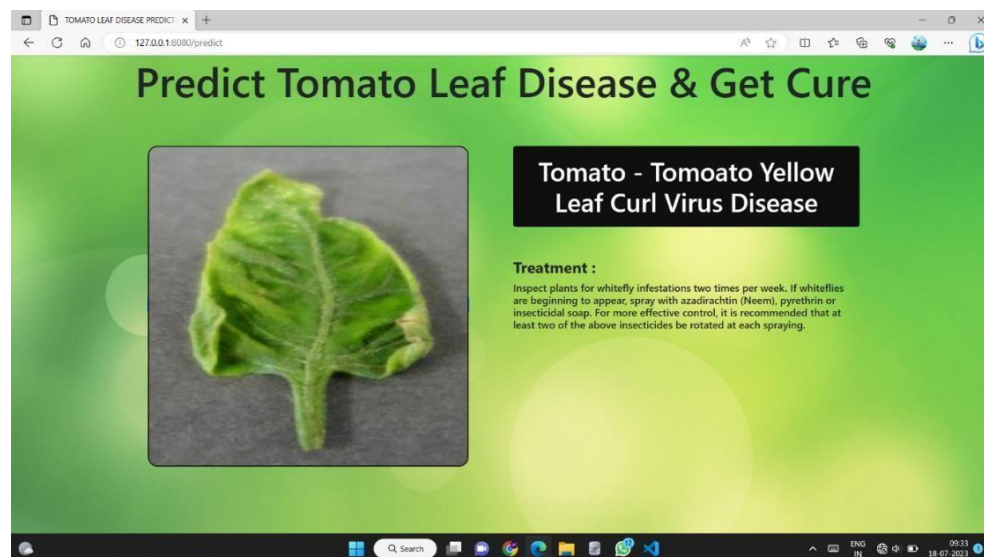


Fig5. 2: tomato yellow leaf curl virus disease

The Fig 5.2 shows the tomato yellow leaf curl virus disease, this page is consist of one sampleimage of the Tomato yellow leaf curl virus disease which is located left side of the page and to the right side it consisting full name of the disease “tomato yellow leaf curl virus disease”

along with this it also consist the treatment to be given to the plant to cure from the tomato yellow leaf virus disease.

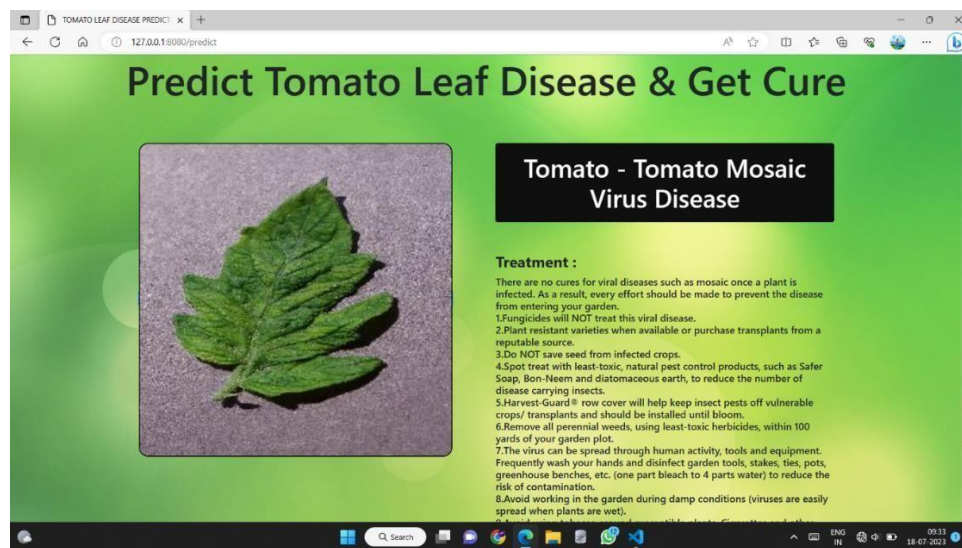


Fig 5.3: Tomato Mosaic virus disease

The Fig 5.3 shows the tomato Mosaic virus disease, this page is consist of one sample image of the Tomato Mosaic virus disease which is located left side of the page and to the right side it consisting full name of the disease “tomato Mosaic virus disease” along with this it also consist the treatment to be given to the plant to cure from the tomato mosaic virus disease.

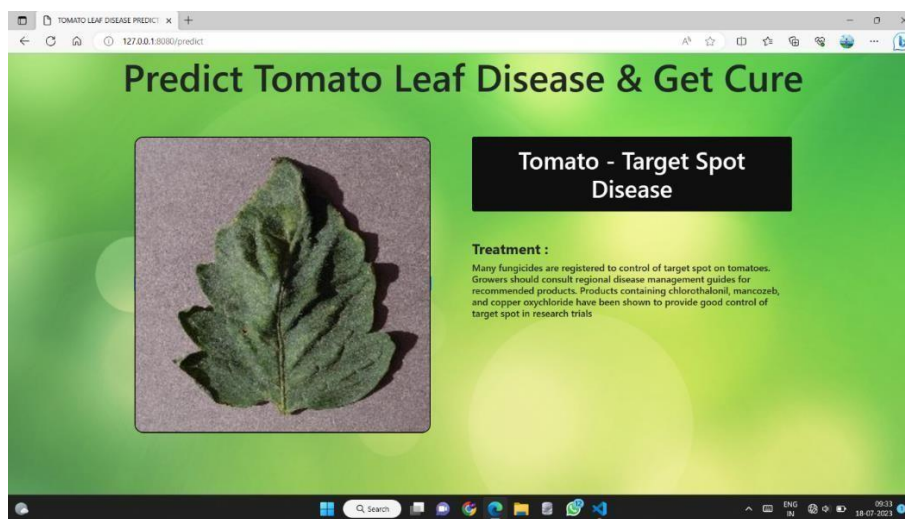


Fig 5.4: Tomato Target Spot disease

The Fig 5.4 shows the tomato Target Spot disease, this page is consist of one sample image of the Tomato Target Spot disease which is located left side of the page and to the right side it consisting full name of the disease “tomato Target Spot disease” along with this it also consist the treatment to be given to the plant to cure from the tomato Target Spot disease.

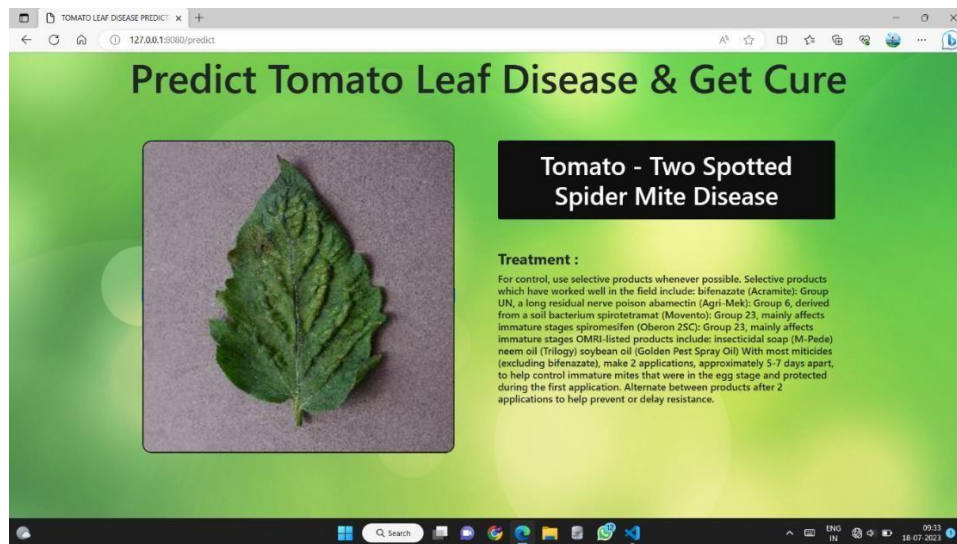


Fig 5.5: Tomato Two Spotted Spider Mite disease

The Fig 5.5 shows the tomato Two Spotted Spider Mite Disease, this page is consist of one sample image of the Tomato Two Spotted Spider Mite disease which is located left side of thepage and to the right side it consisting full name of the disease “tomato Two Spotted disease”along with this it also consist the treatment to be given to the plant to cure from the tomato Two Spotted Spider Mite disease.

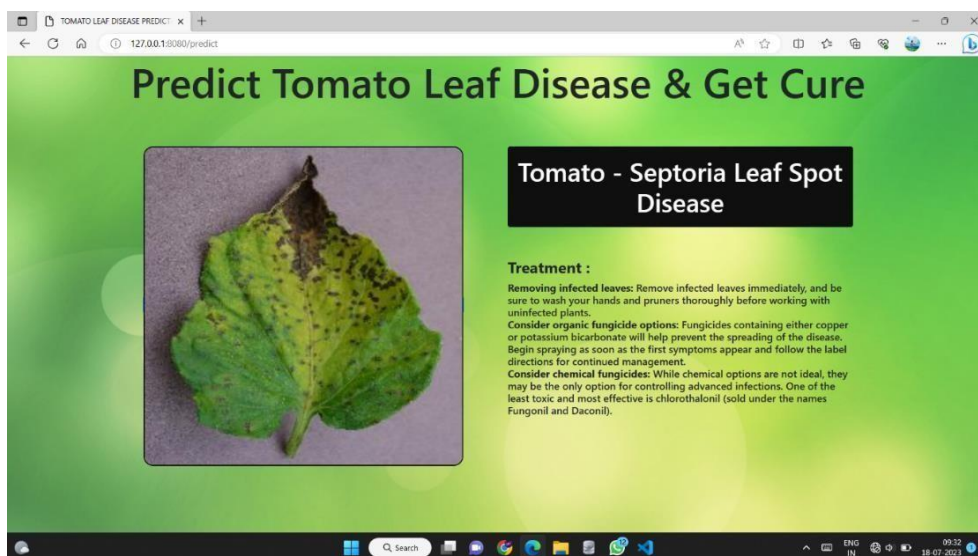


Fig 5.6: Tomato Leaf Spot Disease

The Fig 5.6 shows the Tomato Leaf Spot Disease, this page is consist of one sample image ofthe Tomato Leaf Spot Disease which is located left side of the page and to the right side it consisting full name of the disease “Tomato Leaf Spot Disease” along with this it also consistthe treatment to be given to the plant to cure from the Tomato Leaf Spot Disease.

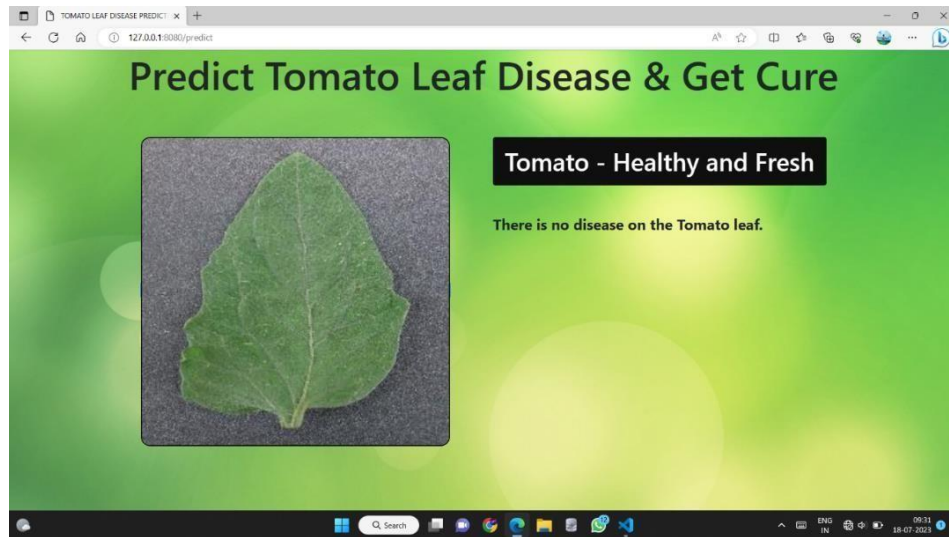


Fig 5.7: Healthy and Fresh leaf

The Fig 5.7 shows the Healthy and Fresh leaf, this page is consisting of one sample image of the Healthy and Fresh leaf which is located left side of the page and to the right side it consists full name Healthy and Fresh leaf along with this it also consists a sentence “There is no disease on the Tomato leaf”.

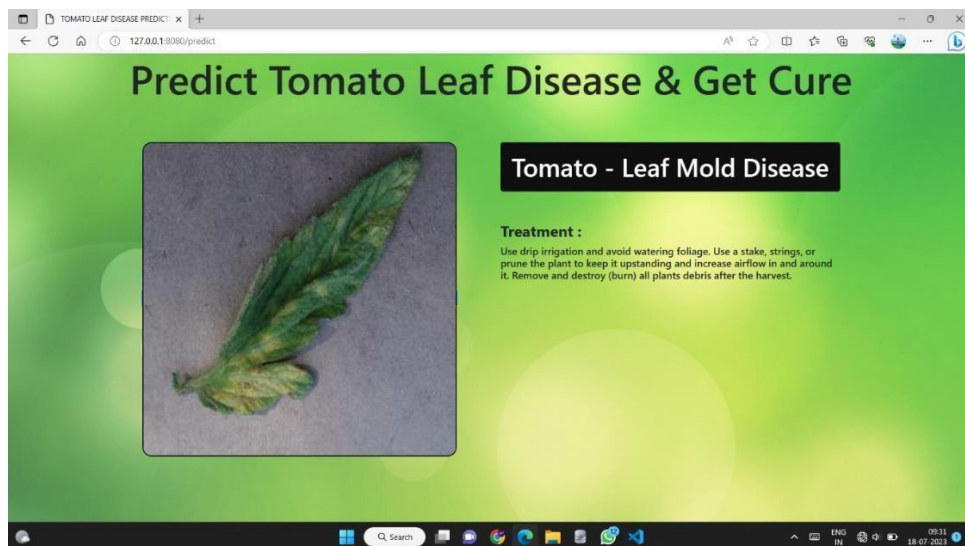


Fig 5.8: Tomato Leaf Mold Disease

The Fig 5.8 shows the Tomato Leaf Mold Disease, this page is consisting of one sample image of the Tomato Mold Spot Disease which is located left side of the page and to the right side it consists full name of the disease “Tomato Leaf Mold Disease” along with this it also consists the treatment to be given to the plant to cure from the Tomato Leaf Mold Disease.

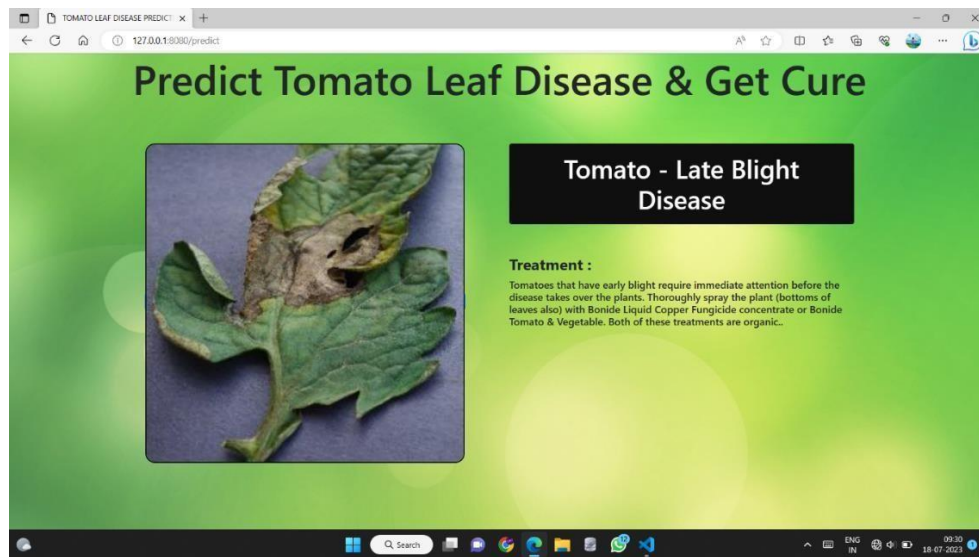


Fig 5.9: Tomato late Blight Disease

The Fig 5.9 shows the Tomato late Blight Disease, this page is consisting of one sample image of the Tomato Late Blight Disease which is located left side of the page and to the right side it consisting full name of the disease “Tomato Late Blight Disease” along with this it also consist the treatment to be given to the plant to cure from the Tomato late Blight Disease.

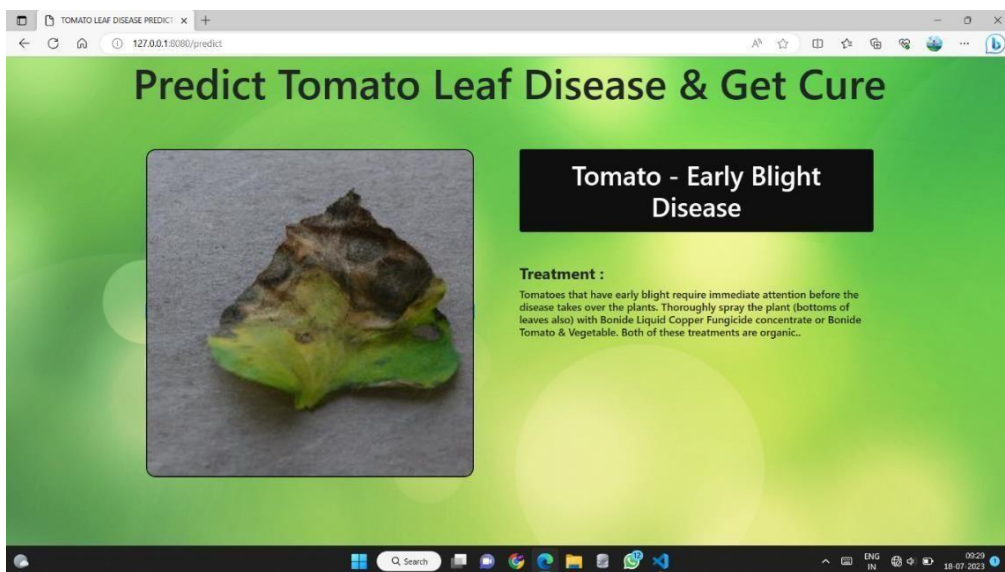


Fig 5.10: Tomato Early Blight Disease

The Fig 5.10 shows the Tomato Early Blight Disease, this page is consisting of one sample image of the Tomato Early Blight Disease which is located left side of the page and to the right side it consisting full name of the disease “Tomato Early Blight Disease” along with this it also consist the treatment to be given to the plant to cure from the Tomato Early Blight Disease.

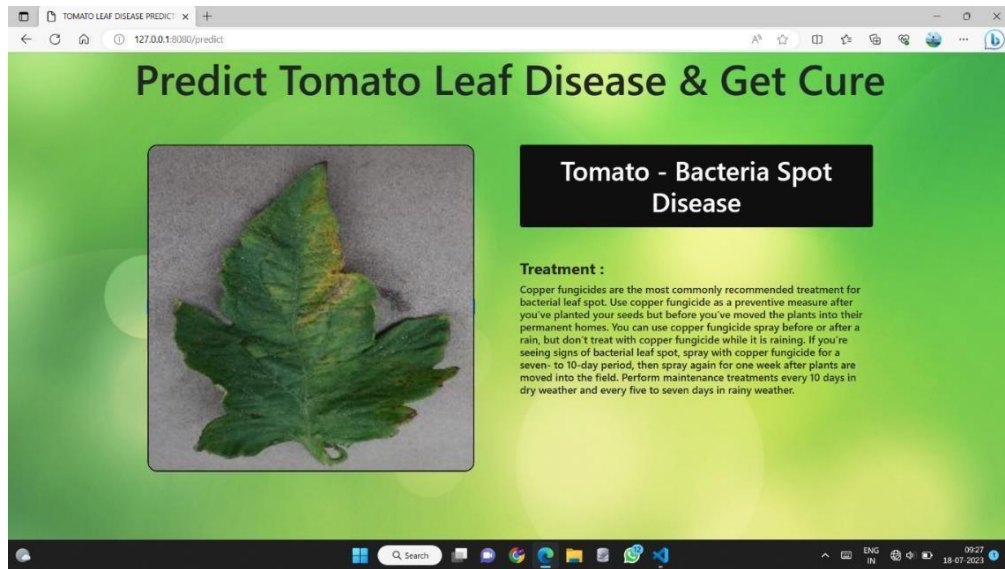


Fig 5.11: Tomato Bacteria Spot Disease

The Fig 5.11 shows the Tomato Bacteria Spot Disease, this page is consisting of one sample image of the Tomato Bacteria Spot Disease which is located left side of the page and to the right side it consisting full name of the disease “Tomato Bacteria Spot Disease” along with this it also consist the treatment to be given to the plant to cure from the Tomato Bacteria SpotDisease.

CHAPTER 6

CONCLUSIONS

In this research, we propose an architecture based on CNNs to identify and classify nine different types of tomato leaf diseases. The complexity in detecting the type of disease lies in the fact that the leaves deteriorate in a similar way in most of the tomato diseases. It means that it is necessary to develop a deep image analysis to judge the types of tomato leaf diseases with a proper accuracy level.

The CNN that we design is a high-performance deep learning network that allows us to have a complex image processing and feature extraction through four modules: the module dataset creation that makes an experimental dataset using public datasets and photographs taken in the fields of the country; model creation that is in charge of parameters configuration and layers definition; data distribution to train, validate and test data; and processing for the optimization and performance verification.

We evaluate the performance of our model via accuracy, precision, recall and the F1-score metrics. The results showed a training accuracy of more than 75% and a validation accuracy of more than 50% in the leaf disease classification.

6.1 FUTURE ENHANCEMENT

In the future, advancements in Convolutional Neural Networks (CNN) will greatly enhance the prediction of tomato plant leaf diseases. With increased computing power and more comprehensive datasets, CNN models will be able to extract intricate patterns and features from leaf images, leading to improved accuracy in disease diagnosis. Additionally, the integration of deep learning techniques with IoT (Internet of Things) devices will enable real-time monitoring of environmental conditions such as humidity, temperature, and soil moisture, providing valuable insights into disease progression.

To achieve 100% accuracy in tomato plant leaf disease prediction, several key enhancements will be crucial. Firstly, a large and diverse dataset comprising high-resolution images of healthy and diseased tomato leaves will be essential to train the CNN model. Gathering data from various regions and seasons will help capture the full range of disease symptoms and environmental factors. Secondly, continuous model optimization and fine-tuning will be necessary to refine the CNN architecture and improve its ability to detect subtle disease indicators. Regular updates and retraining of the model with new data will ensure it stays up-to-date with emerging diseases and variations. Finally, close collaboration between researchers, agronomists, and farmers will be vital to validate the accuracy of predictions and provide real-world feedback to further enhance the CNN model's performance.

CHAPTER 7

REFERENCES

- [1] Al-Hiary, Heba, et al. "Fast and accurate detection and classification of plant diseases." *International Journal of Computer Applications* 17.1 (2011): 31-38.
- [2] Al Bashish, Dheeb, Malik Braik, and Sulieman Bani-Ahmad. "Detection and classification of leaf diseases using K-means-based segmentation and." *Information technology journal* 10.2 (2011): 267-275.
- [3] Mokhtar, Usama, et al. "Tomato leaves diseases detection approach based on support vector machines." 2015 11th International Computer Engineering Conference (ICENCO).IEEE, 2015.