AgroSmart

The Smart Agriculture

Mini Project Team-12
Dept. of Computer Science Engineering
Rajiv Gandhi University of Knowledge Technologies
Nuzvid

Our Team Members

Guide: Mr. T. Chandrasekhar

Boddu Sri Pavan

N180606

Bandi Vikram Kumar

N181167

Mannem Pavan Kumar

N180520

Mallidi Sandhyarani

N181108

Pragada Padma Priya

N180272

Acknowledgement

We express our sincere gratitude towards our guide, **Mr. T. Chandrasekhar**, Computer Science Engineering for his valuable guidance, suggestions, and supervision throughout the work. Without his patronage and guidance, the project would not have taken shape. We owe our sincere thanks to the **Dept.of Computer Science Engineering, Rajiv Gandhi University of Knowledge Technologies** for providing support to our project. We are grateful to the **Eswara** for the abundant grace and blessings bestowed upon us, which enabled us to successfully complete this project.

Timeline Chart

WEEK- 1
Revised numpy, pandas, matplotlib libraries

WEEK- 2
Crop prediction literature survey

WEEK- 3
Crop prediction model is trained

WEEK- 4
Plant disease dataset collection

WEEK- 5
Mock PPT presentation

WEEK- 6
Worked on deep learning

WEEK- 7
Literature survey for plant disease detection

WEEK- 8
CNN Prototype model is trained

WEEK- 9
Worked on Python Flask

WEEK- 10
Crop disease detection model is trained

WEEK- 11
Documentation Work is in progress

WEEK- 12
Documentation completed, Mockpanel presentation

Abstract

Agriculture is the primary source in providing food for entire world. Greater than 45% of the world and 70% of the Indian Population relies on agriculture for its livelihood. Around 50% of loss in crop yield is reported due to pests and diseases. Wrong selection of crop leads to soil infertility as well as crop failure. Early prediction of diseases can save crops. Our project presents state-of-the-art models to predict suitable crop from minimum number of environmental features with an accuracy of 97.05% and to predict disease with minimum number of parameters with a maximum accuracy of 93.61%.

Introduction

Agriculture sector is the backbone of the Indian society, employing more than 70% of the population. It is primarily engaged in growing crops and livestock management. There are many challenges associated with agriculture includes irrigation, soil fertility, temperature, natural calamities, diseases and pest attacks. All these factors result in the decrease of crop yield. They severely affect people and causes famines. Machine Learning and deep learning transforms data into models which can be used for futuristic predictions.

Problem Statement

- To predict the best suitable crop for a field with as minimum number of input features as possible.

- To predict the disease from the image of a plant leaf with optimized CNN architecture.

Literature Review

Sl.No	Paper Description	Proposed System	Accuracy
1	Crop Prediction System using Machine Learning Algorithms by Pavan Patil, et al.	KNN	89.4%
2	Crop Recommendation System and Plant Disease Classification using Machine Learning for Precision Agriculture	RF	93%
3	Plant Leaf Detection and Disease REcognition using Deep Learning by Sammy V. Militante, et al.	GNB & SVM	90%
4	Crop Prediction using Machine Learning by M Kalimuthu, et al.	GNB	97%
5	Crop Recommendation System using Machine Learning by Aman Sinha, et al.	GNB	94.73%
6	An artificial intelligence solution for crop Recomendation by Varshitha. D. N, et al.	Neural Networks	87%

Sl.No	Paper Description	Proposed System	Accuracy
7.	Plant Leaf Detection and Disease Recognition using Deep Learning By Sammy V.Militante, et al.	CNN	96.5%
8	A Deep CNN Approach for Plant Disease Detection By Fatma MARZOUGUI, et al.	CNN	94.80%
9	Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm by Melike Sardogan, et al.	CNN	86%
10	Transfer Learning Approach for Plant Disease Detection Using CNN with Pre-trained Feature Extraction Method Mobilnetv2 by Raida Moyazzoma, et al.	CNN	90.38%
11	Plant Leaf Diseases Detection and Classification Using Image Processing and Deep Learning Techniques by Marwan Adnam Jasim, et al.	CNN	98.29%
12	Detection of Diseases in Tomato Plant using Machine Learning by A. Chadak, et al.	ResNet Pre-trained model	99.92%
13	Plant Disease Detection Using Image Processing and Machine Learning by Pranesh Kulkarni, et al.	Random Forest Classifier	93%

Need of the study

- More number of features[1][2][3][4][5][6]
- Underfitting[1]
- Less varieties of crops[8] [9][11][12][13]
- Small dataset[8][9][10][3]
- Complex CNN architecture[8][11]

Methodology for Crop Recommendation

Data Collection

↓

Exploratory Data Analysis

↓

AEFS Technique

↓

Model Selection

↓

Training (80% of data)

↓

HyperParameter Tuning

↓

Testing (20% of data)

↓

Model Evaluation

Methodology for plant disease prediction

Data Collection

↓

Exploratory Data Analysis

↓

Design CNN Architecture

↓

Training

- Training (80% of data)
- ↓
- Validation (10% of data)

↓

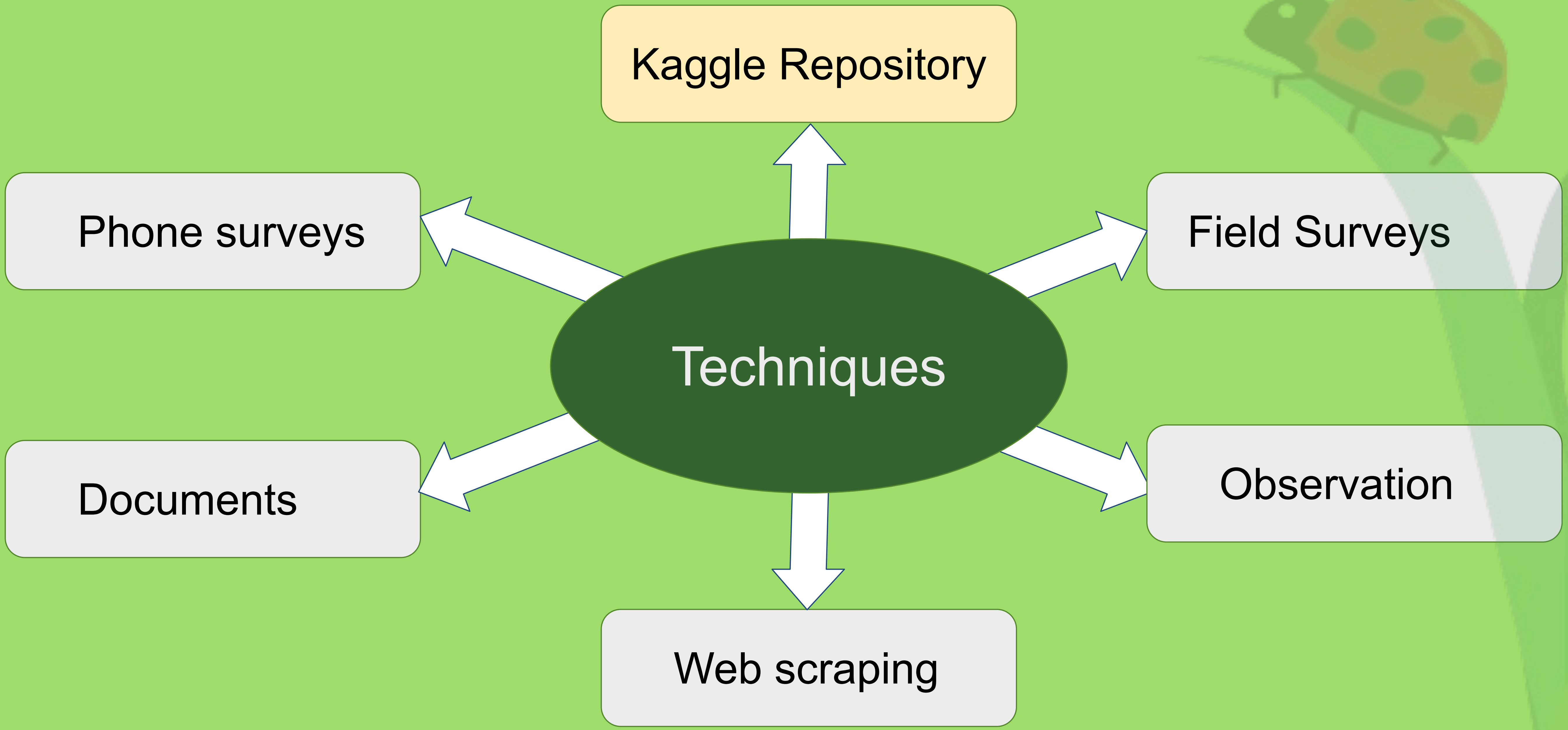
Testing (10% of data)

↓

Model Evaluation

Data Collection

Data collection is the systematic process of gathering required and relevant information for particular study.



Exploratory Data Analysis

Exploratory data analysis is an approach to analyzing datasets to summarize their main characteristics often using statistical graphics and other data visualization methods to reveal insights hidden in the data.

Crop recommendation dataset :

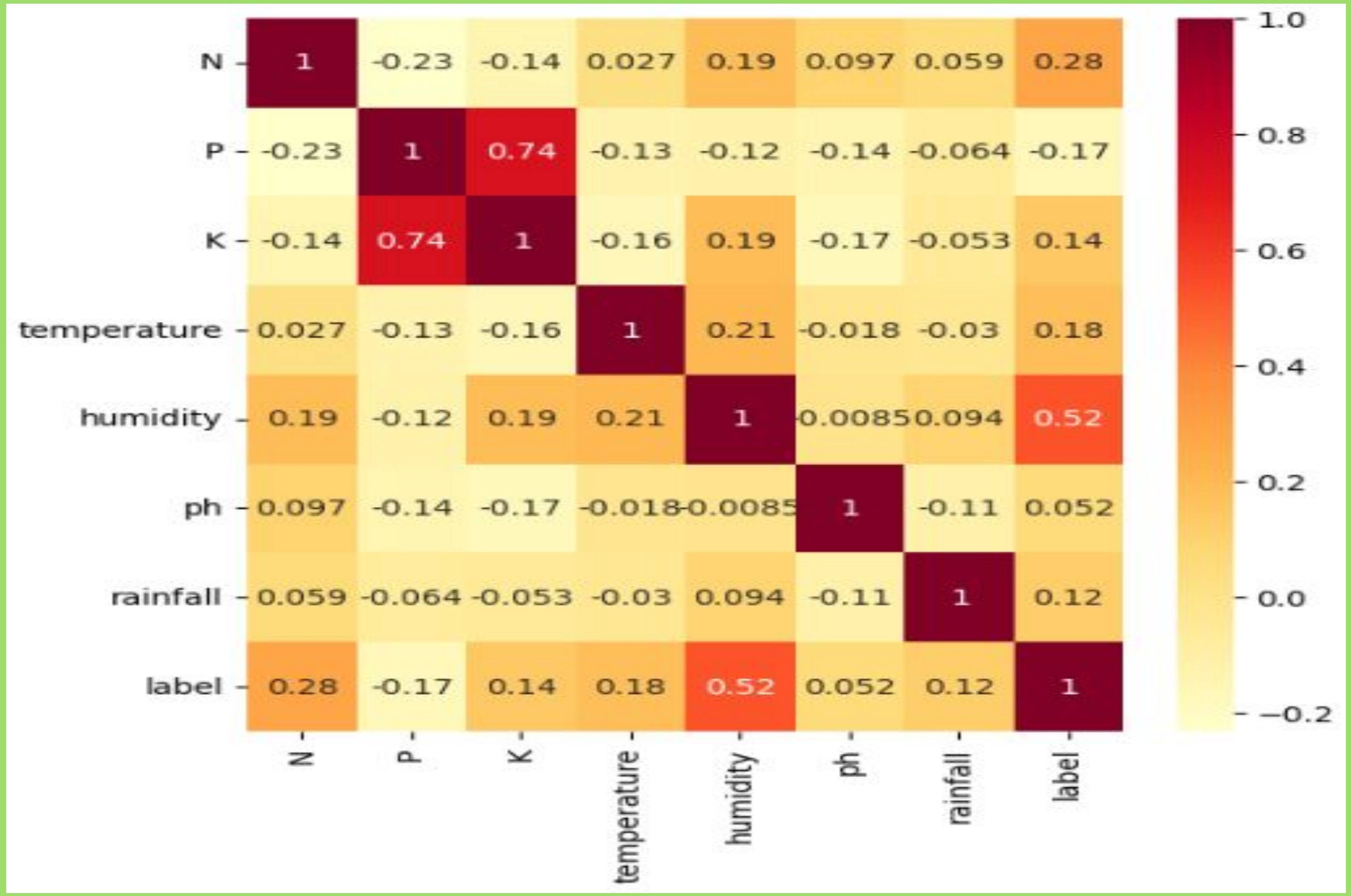
- 2200 samples
- 7 numeric attributes
- 22 different crops (100/ crop)
- No null, missing, or duplicate entries.
- *train:test= 80:20*

Plant Disease dataset:

- 87,867 RGB images
- 38 classes
- Batch size =64
- Image size=100 X 100px
- *train:validation:test= 80:10:10*

Correlation analysis on crop recommendation dataset

Correlation expresses the extent to which two features are linearly related.



- Fails to capture non-linear relationships.

Crop recommendation dataset distributions

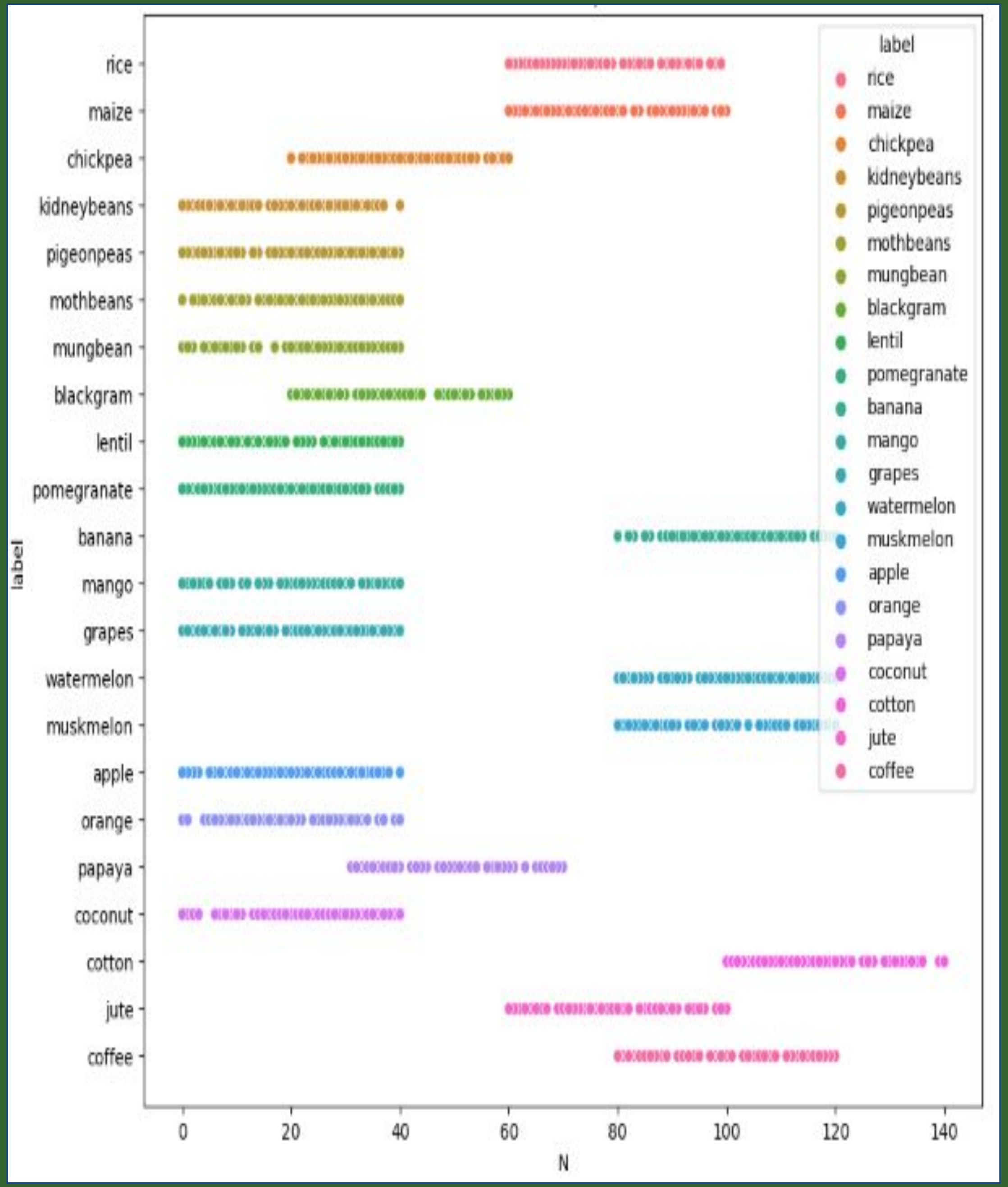


Figure-1(a): N vs. Crop

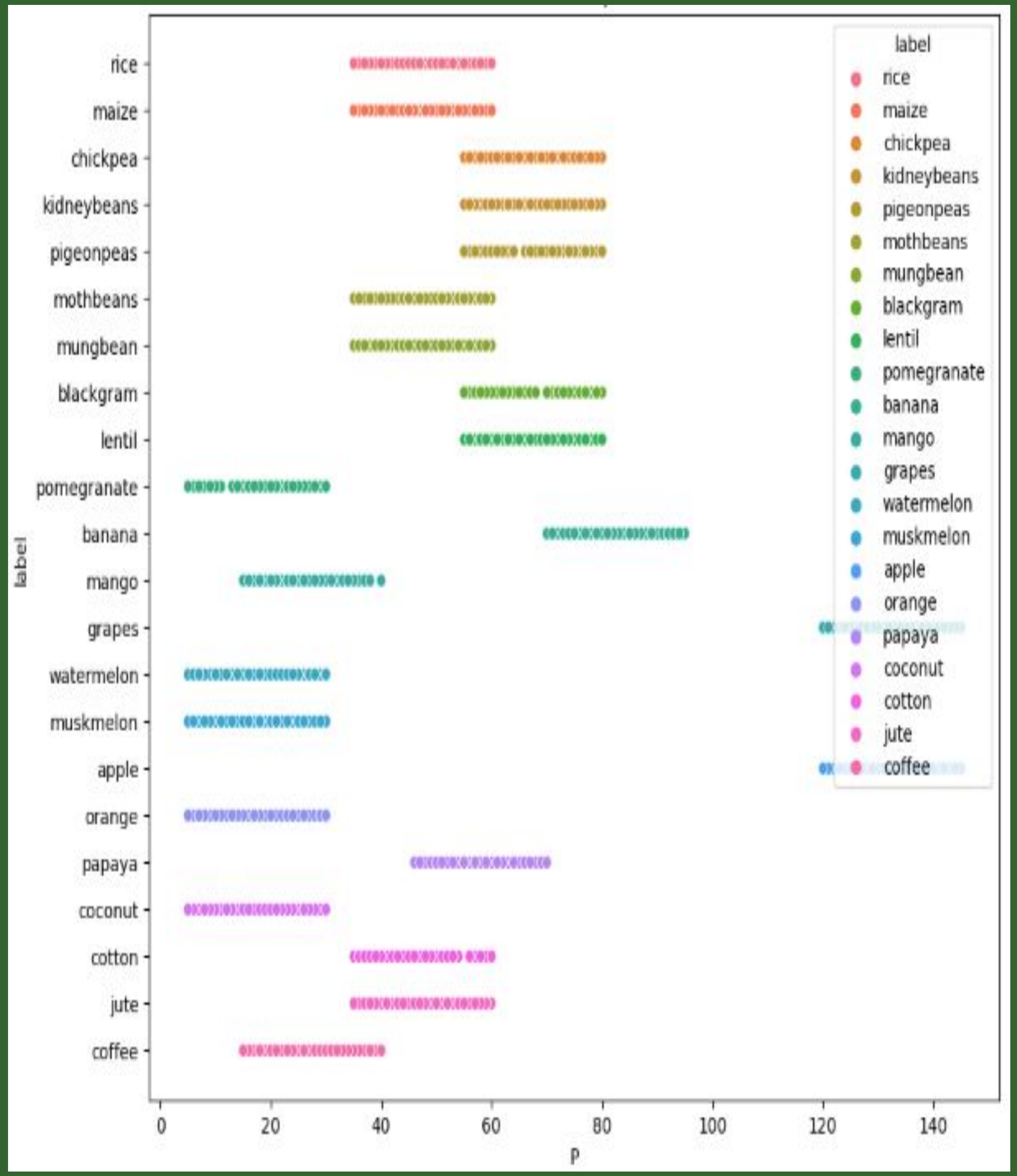


Figure-1(b): P vs. Crop

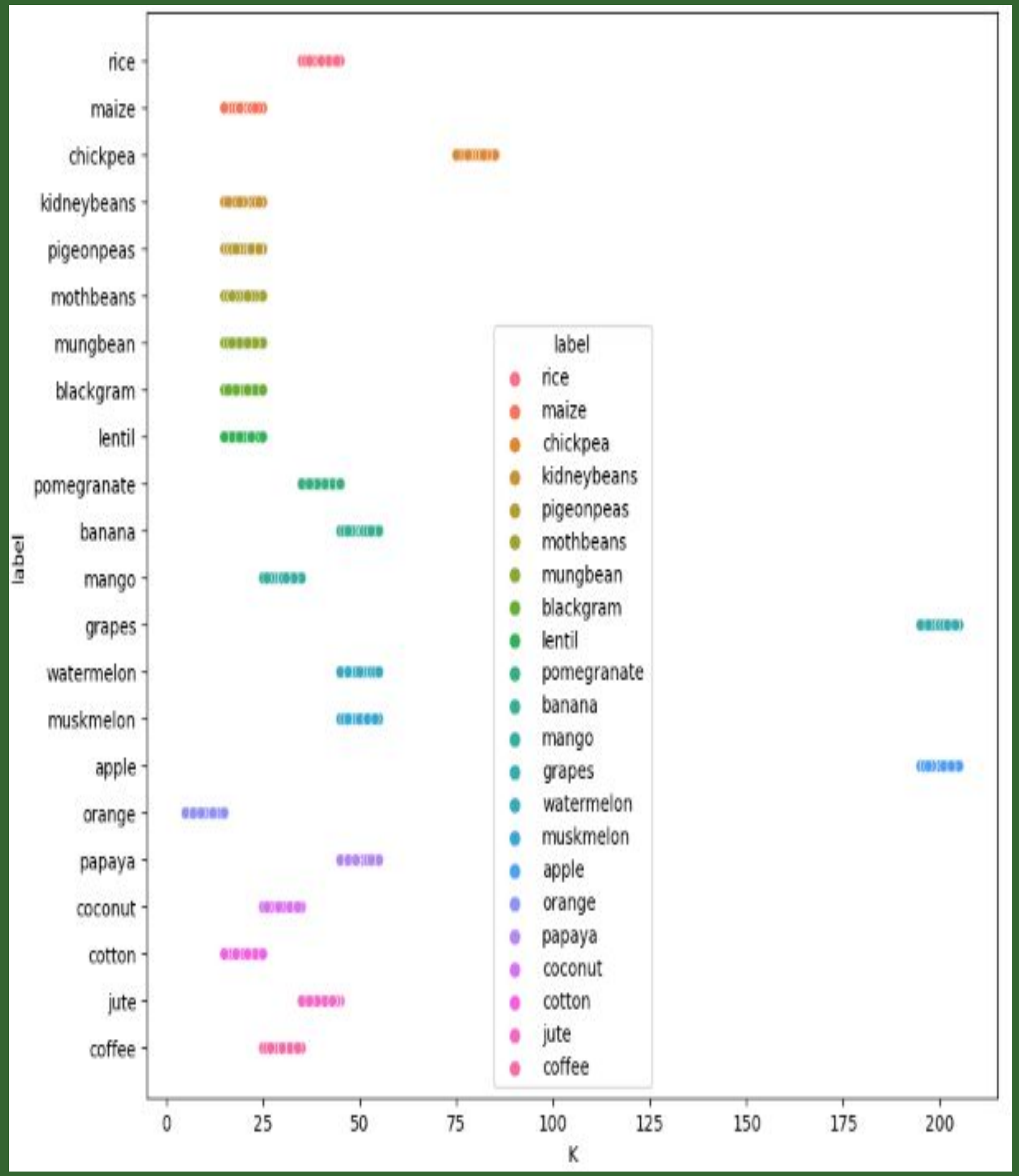


Figure-1(a): K vs. Crop

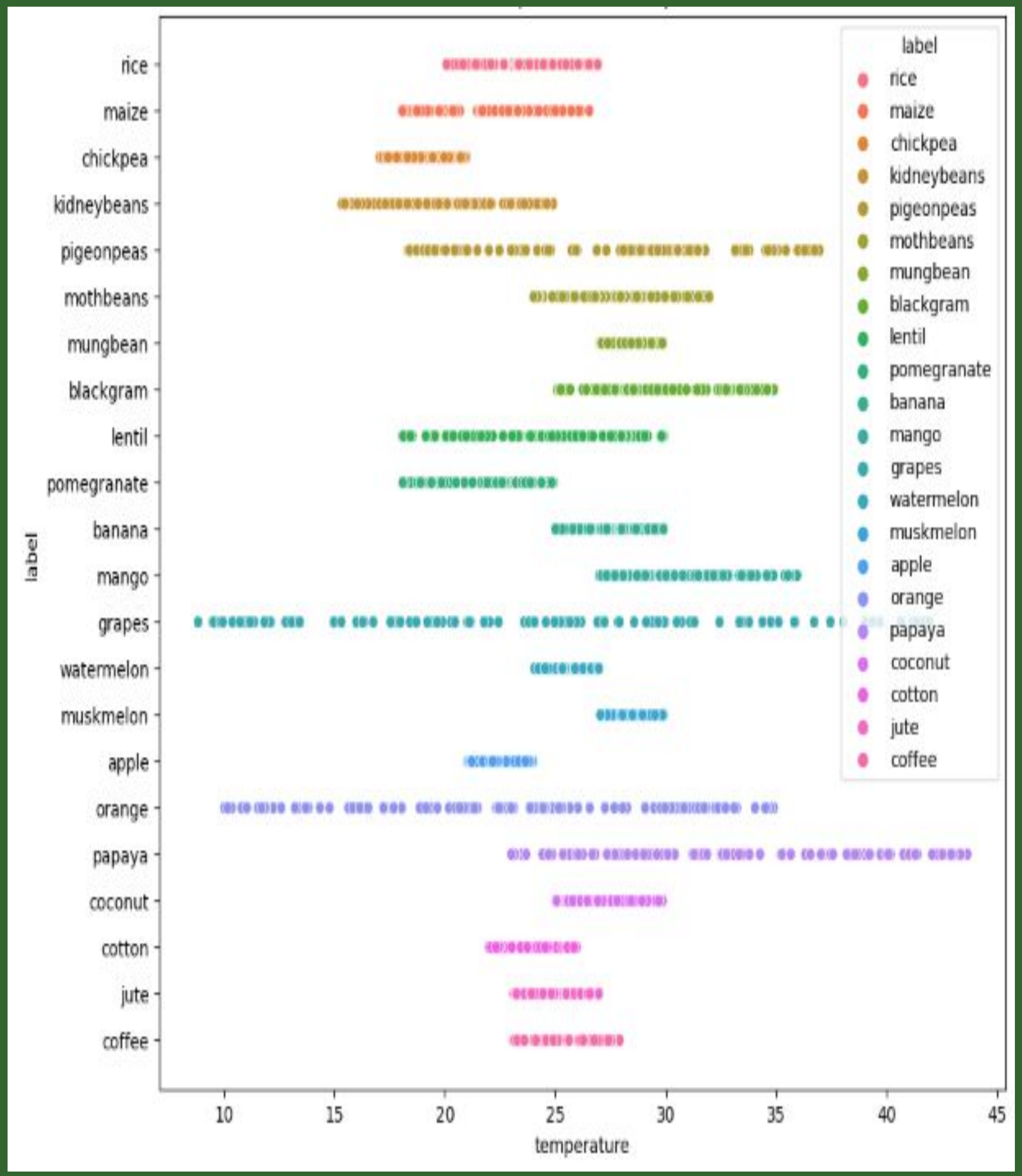


Figure-1(d):Temperature vs.Crop

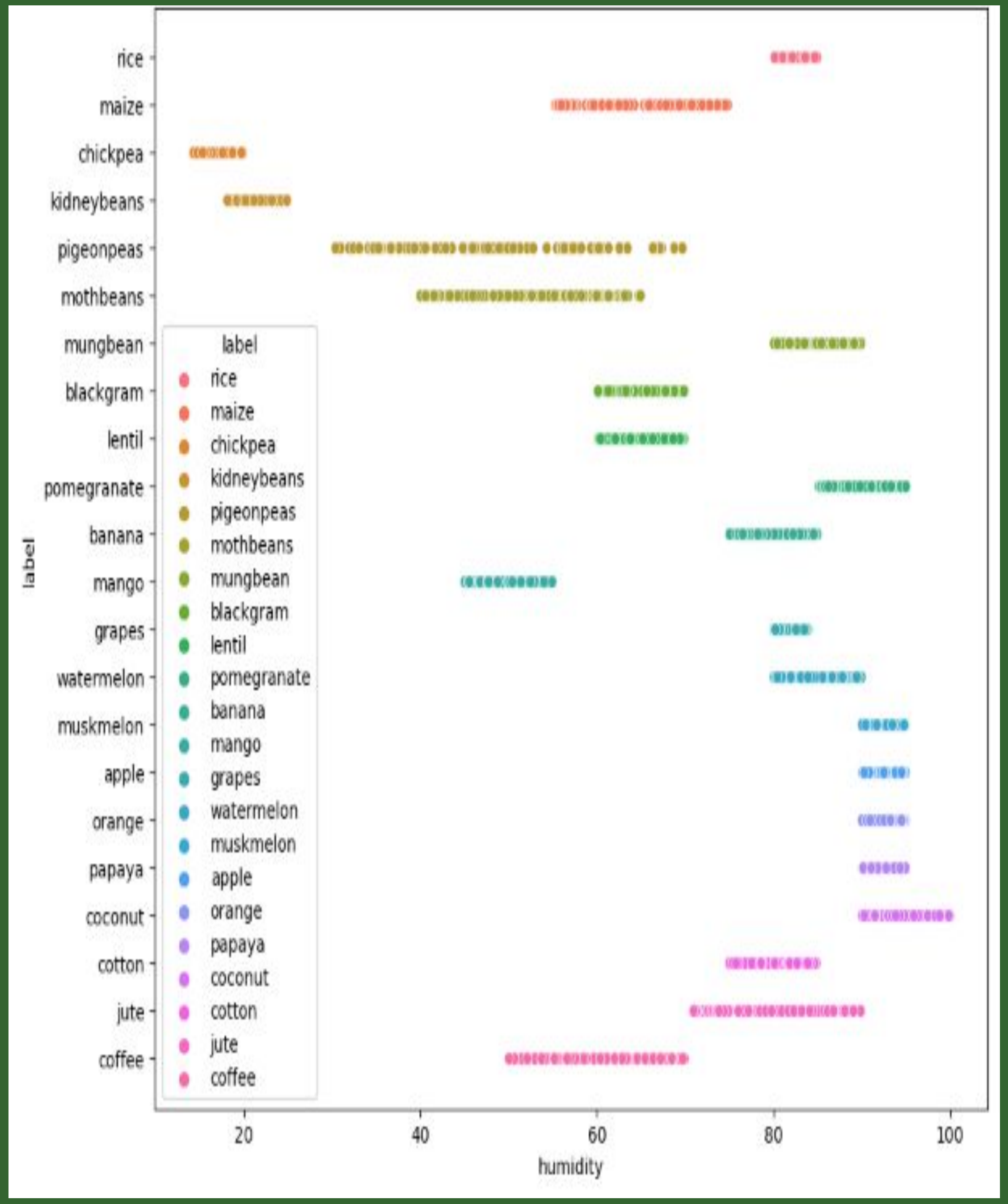


Figure-1(e): Humidity vs. Crop

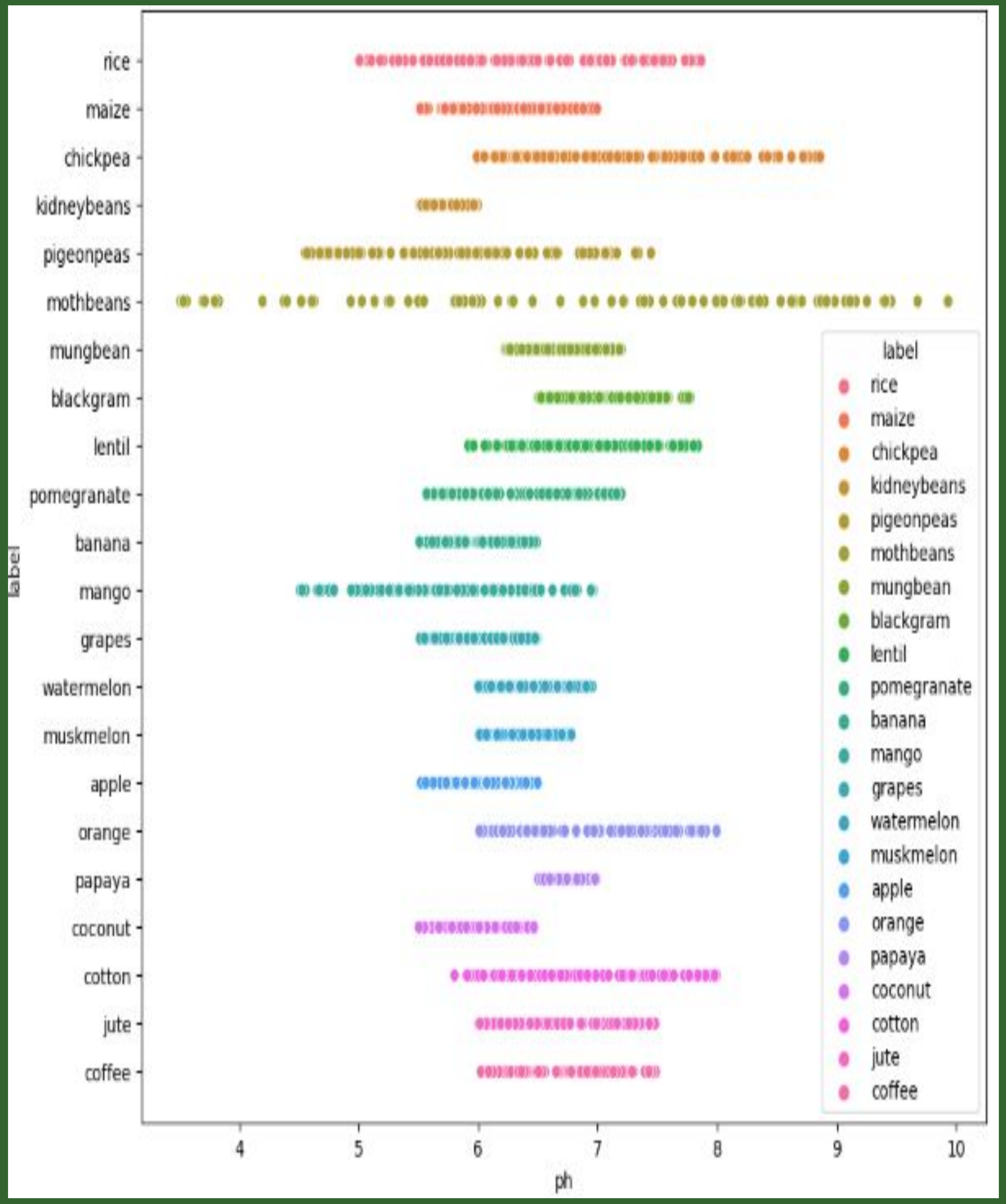


Figure-1(f): pH vs. Crop

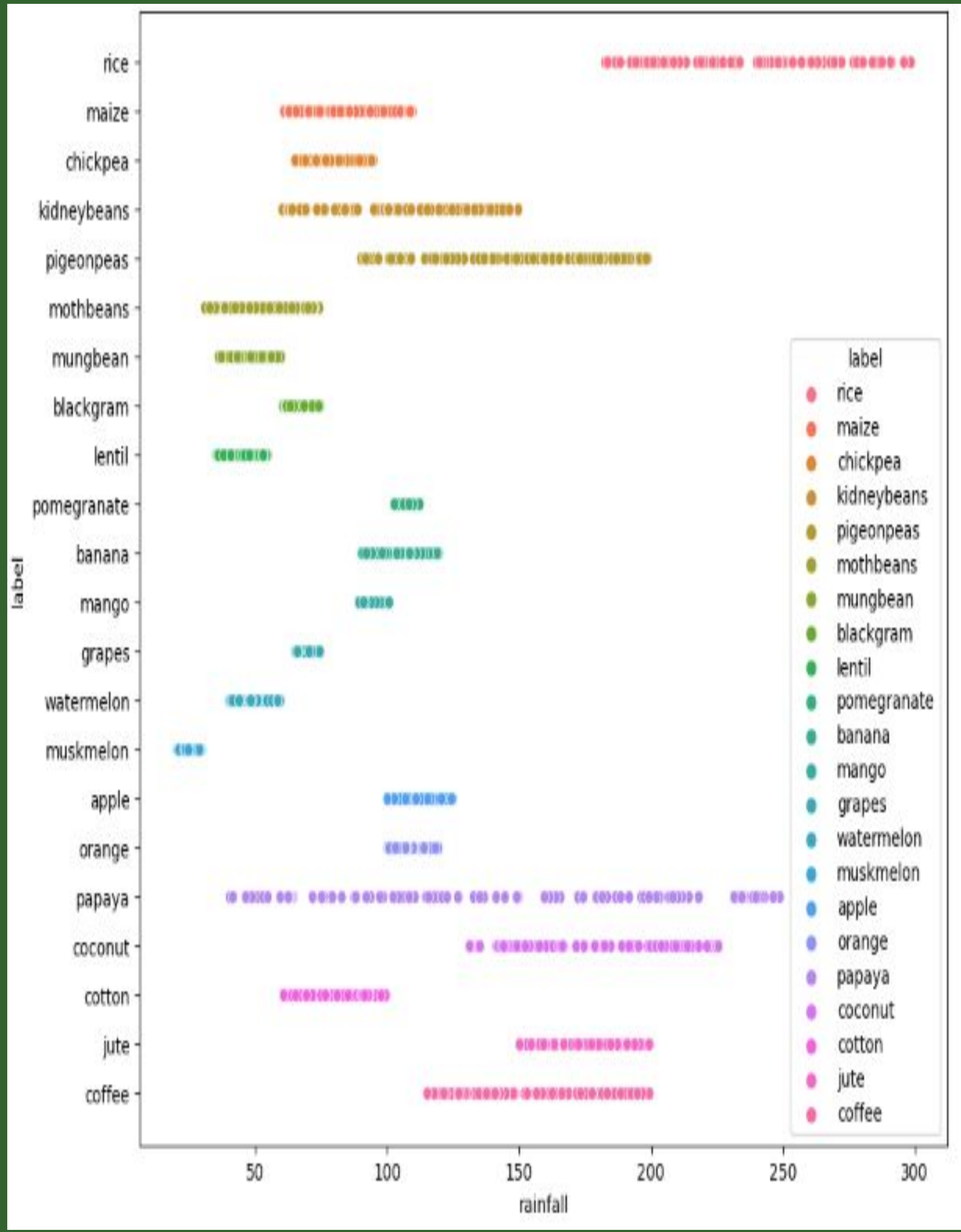


Figure-1(g): Rainfall vs. Crop

Plant Disease dataset distributions

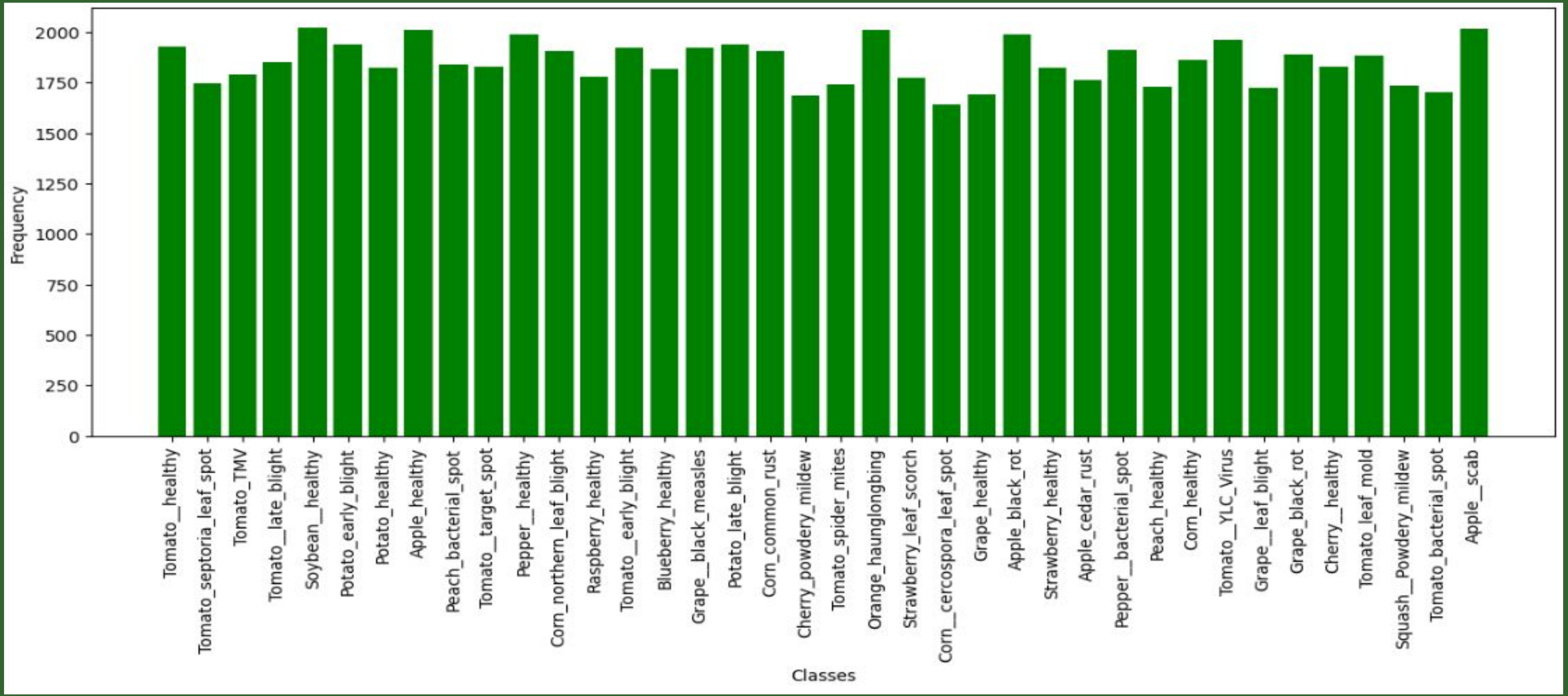


Figure-2(a): Training data distribution

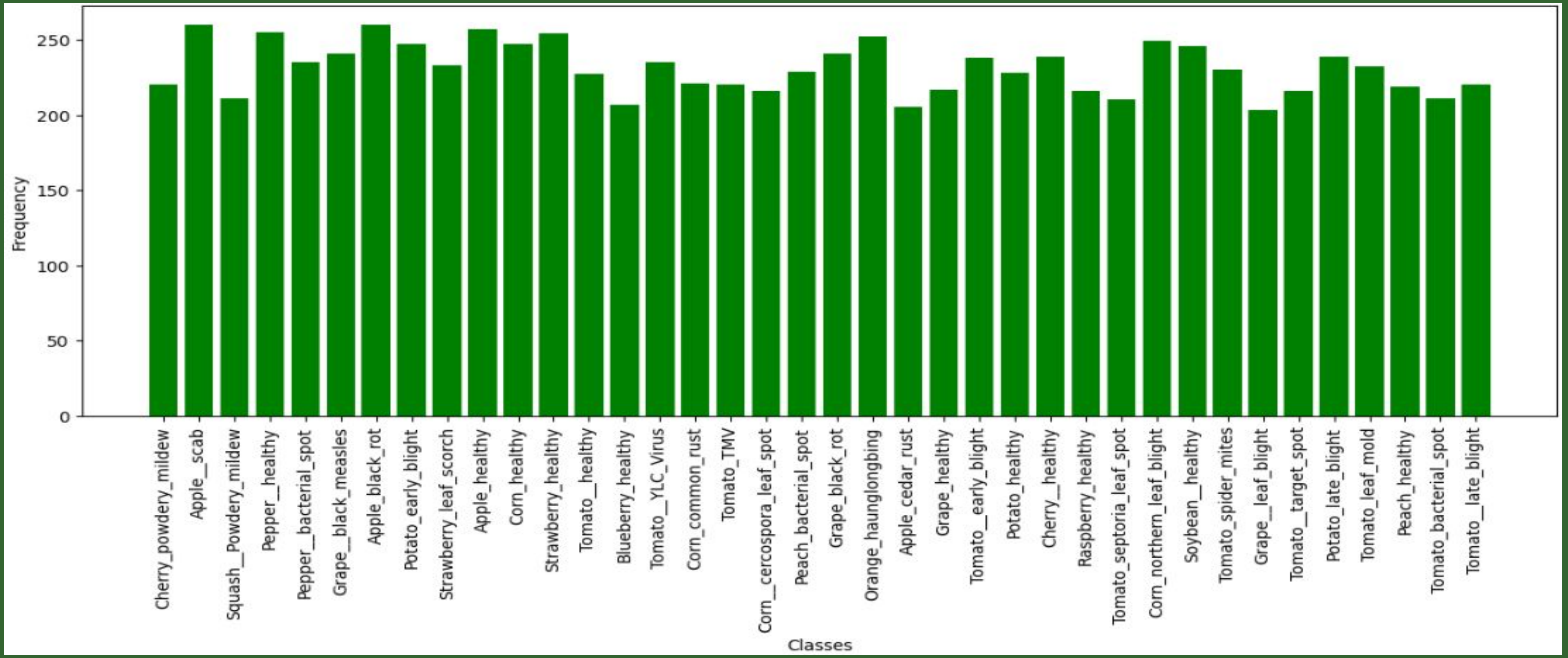


Figure-2(b): Validation data distribution

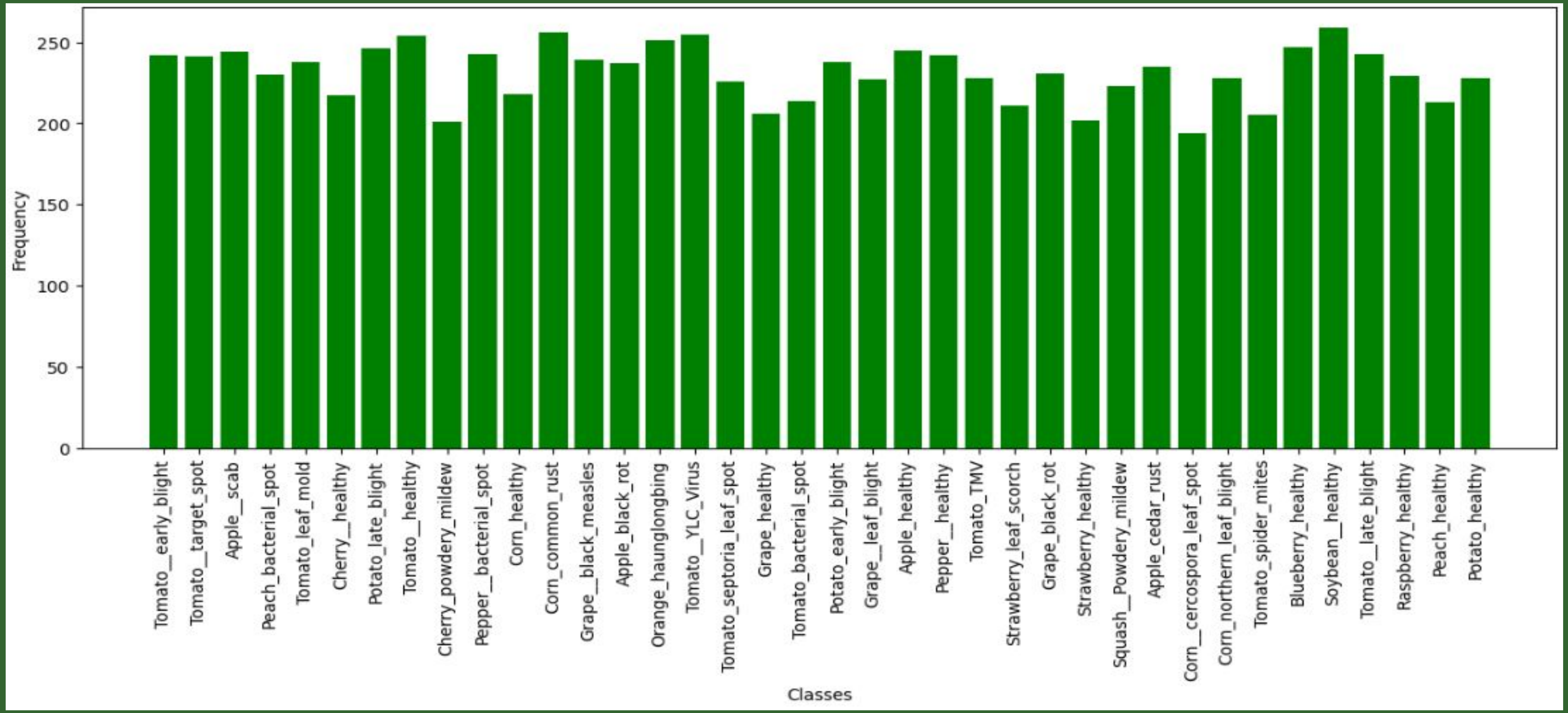


Figure-2(c): Testing data distribution

Ascending Exhaustive Feature Selection (AEFS)

- Modified exhaustive feature selection
- Finds out the smallest possible subset of features yielding maximum accuracy threshold.
- Brute force technique
- Efficient for data sets with a small number of attributes.
- Baseline model: Random Forest Classifier
- Accuracy threshold: 95%

Resultant feature set: **{ K, Humidity, Rainfall }**

ALGORITHMS

- ***K-Nearest Neighbors*** algorithm finds K closest data points to a query point and makes predictions based on their labels/values for classification/regression tasks.
- ***Support Vector Machines*** algorithm finds the optimal hyperplane to separate data points into classes in a higher dimensional feature space.
- ***Decision Tree*** algorithm recursively partitions the feature space into subsets based on the most informative features, creating a tree-like structure for classification or regression tasks.
- ***Random Forest*** algorithm is an ensemble learning method that constructs multiple decision trees and combines their predictions for improved accuracy in classification and regression tasks.

- ***Extra Trees Classifier*** algorithm is an ensemble learning method that constructs multiple decision trees with random splits and combines their predictions for improved accuracy in classification tasks.

- ***Gradient Boosting*** algorithm is an ensemble learning method that builds a strong predictive model by iteratively adding weak models in a gradient descent fashion, minimizing the loss function for regression or classification tasks.

- ***Logistic Regression*** algorithm is a statistical model that uses a logistic function to estimate the probability of a binary or categorical outcome based on input features.

- ***Naive Bayes*** algorithm is a probabilistic method that applies Bayes' theorem with the "naive" assumption of independence between features to make predictions in classification tasks.

Random Forest Algorithm

Random Forest classifier is a non-parametric, eager learner, which is an ensemble of randomly subsampled decision trees.

Algorithm :

Step-1 : Consider no.of decision trees to be built as n_estimators.

Step-2 : Bootstrap sampling : generate randomized samples with replacement from the original dataset.

Step-3 : Randomized feature splits : random subset of features are chosen.

Step-4 : Generate decision tree and iteratively construct n_estimators no.of decision trees.

Step-5 : Predict test sample from ensemble of decision trees and take result of majority voting as the final prediction.

Hyperparameter Tuning

Hyperparameter tuning is the process of choosing a set of optimal hyperparameters for a model.

n_estimators **:** The number of trees in the forest. (value= 500).

max_features **:** The number of features to consider when looking for the best split (value= sqrt).

max_depth **:** Maximum depth of the tree (value= 10).

criterion **:** Function to measure the quality of a split (value= gini).

random_state **:** To generate reproducible results (value= 111).

Model Selection

Sl. No	Algorithm	Training Accuracy	Testing Accuracy	F1- Score
1.	KNN	95.80%	92.95%	92.95%
2.	SVM	90.74%	89.55%	89.55%
3.	Decision Tree	100%	93.41%	93.41%
4.	Random Forest	97.73%	97.05%	97.05%
5.	Extra trees	100%	95.23%	95.23%
6.	Gradient Boosting	100%	96.59%	96.59%
7.	Logistic Regression	96.25%	94.77%	94.77%
8.	Naive Bayes	68.58%	68.18%	68.18%

Confusion Matrix

The Confusion matrix visualizes and summarizes the performance of a classification algorithm.

		Actual	
		Positive	**Negative**
Predicted	**Positive**	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Metrics

$$Accuracy = \frac{TrueNegatives + TruePositive}{TruePositive + FalsePositive + TrueNegative + FalseNegative}$$

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall\ = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)}$$

$$\text{F1} = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

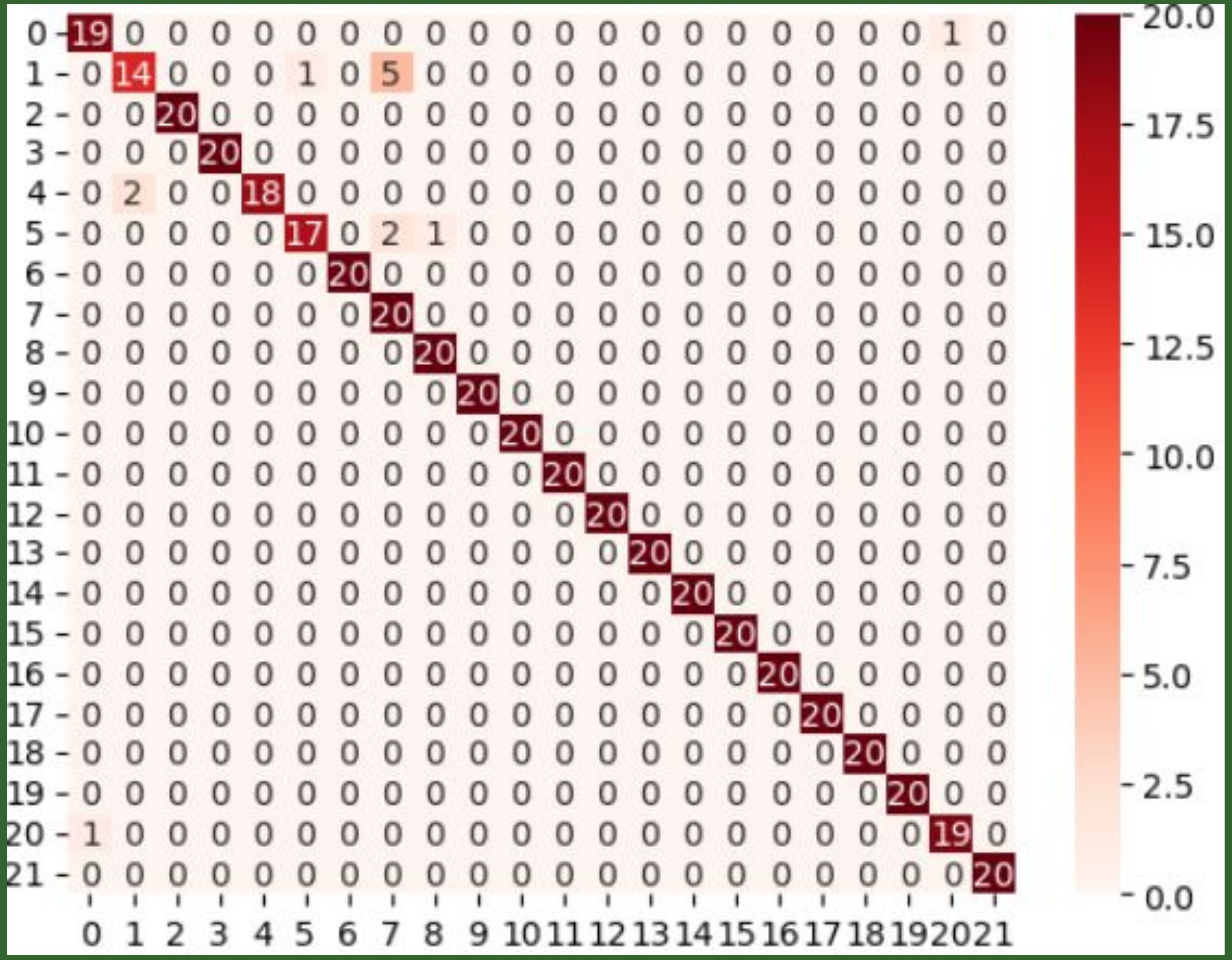


Figure-3: Confusion matrix for Random Forest Classifier

Convolutional Neural Network (CNN)

CNN is a network architecture for deep learning which is composed of several layers of artificial neurons.

1. Convolution layer: The convolution layer acts as a filter to extract a feature using the back-propagation technique. The sliding window moves over the entire image and obtains feature maps. Activation functions are used to introduce non-linearly into the output of neurons. Regularization functions are used to control overfitting. Images are normalized before they are given as input to this layer.

2. Pooling layer: Pooling layer is used to reduce the size of the feature map using sliding window technique.

3. Dense layer: Dense layer contains a single layer of artificial neurons that act as a hidden layer. Images are flattened into 1-D before this layer.

4. Output layer: The final layer of neurons is called "Output layer".This layer should contain the total number of neurons equal to the total number of classes.

CNN Architecture

Resizing Layer
Min-Max Normalization Layer
Convolution Layer-1: 64 features, kernel size=2X2, padding = “same”, activation_function =“ReLU”, kernel_regularizer = l2(0.005)
Pooling Layer-1: MaxPooling with pool_size = 3X3
BatchNormalization layer
Convolution Layer-2: 64 features, kernel_size = 3x3, activation_function= “ReLU”, kernel_regularizer = l2(0.003)
Pooling Layer-2: MaxPooling technique with pool_size = 3x3
BatchNormalization layer
Dropout layer with 50% dropout neurons
Flatten layer
Fully Connected Dense layer with 100 neurons, activation_function= “ReLU”
Output layer with 38 neurons, activation_function = “softmax”

Regularization Techniques

Regularization techniques are used to calibrate the model in order to minimize the adjusted loss function and present overfitting or underfitting.

1. **l2 Regularization**: It forces the weights to decay towards zero but not exactly to zero.
2. **Batch Normalization:** It is a normalization technique done between the layers of the neural network along min-batches instead of the full dataset to speed up training and make learning easier by using higher learning rates.
3. **Dropout Regularization:** It is the practice of disregarding certain neurons in a layer at random during the training phase. It breaks circumstances committed by prior layers making the model more robust. It avoids the overfitting problem.

Accuracy Curves

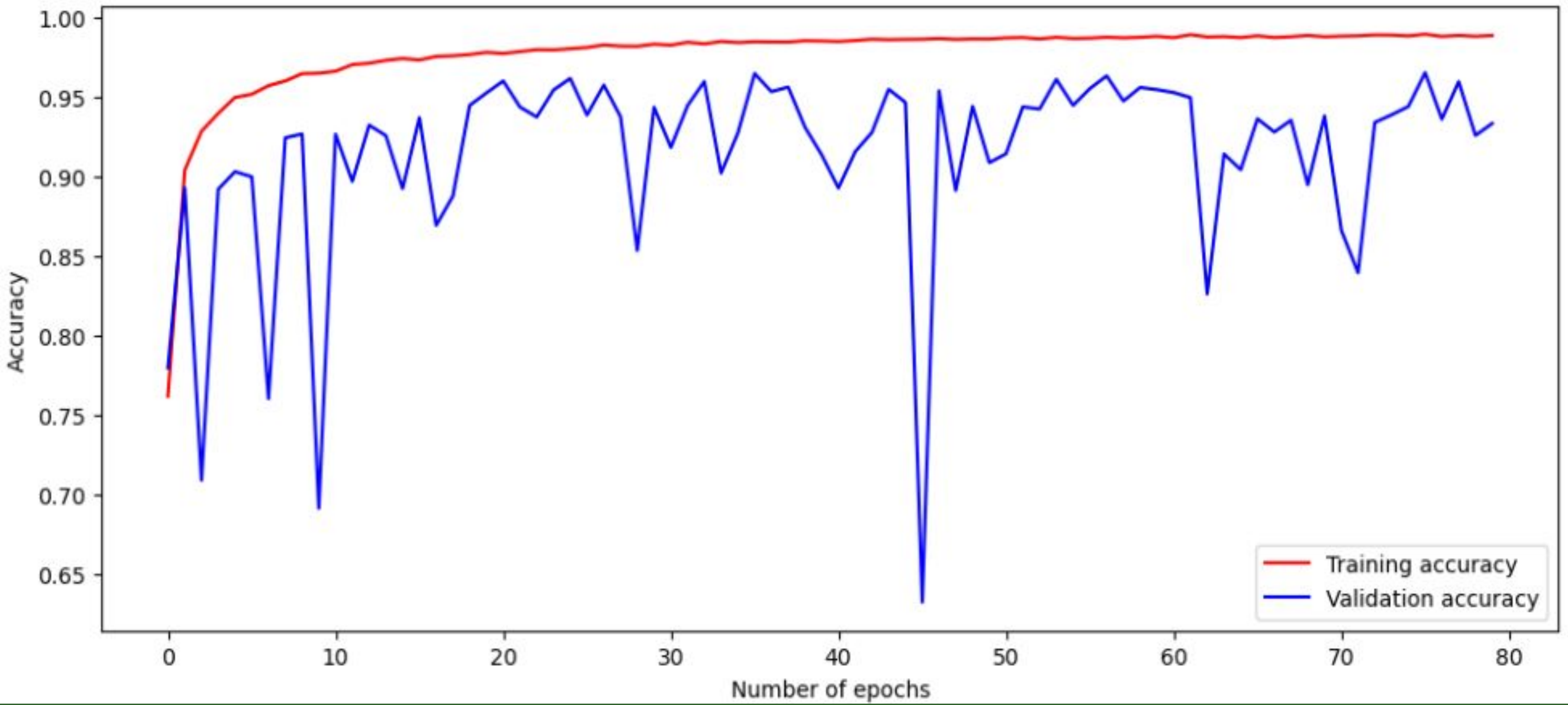


Figure-(a): Training vs. Validation accuracy

Loss Curves

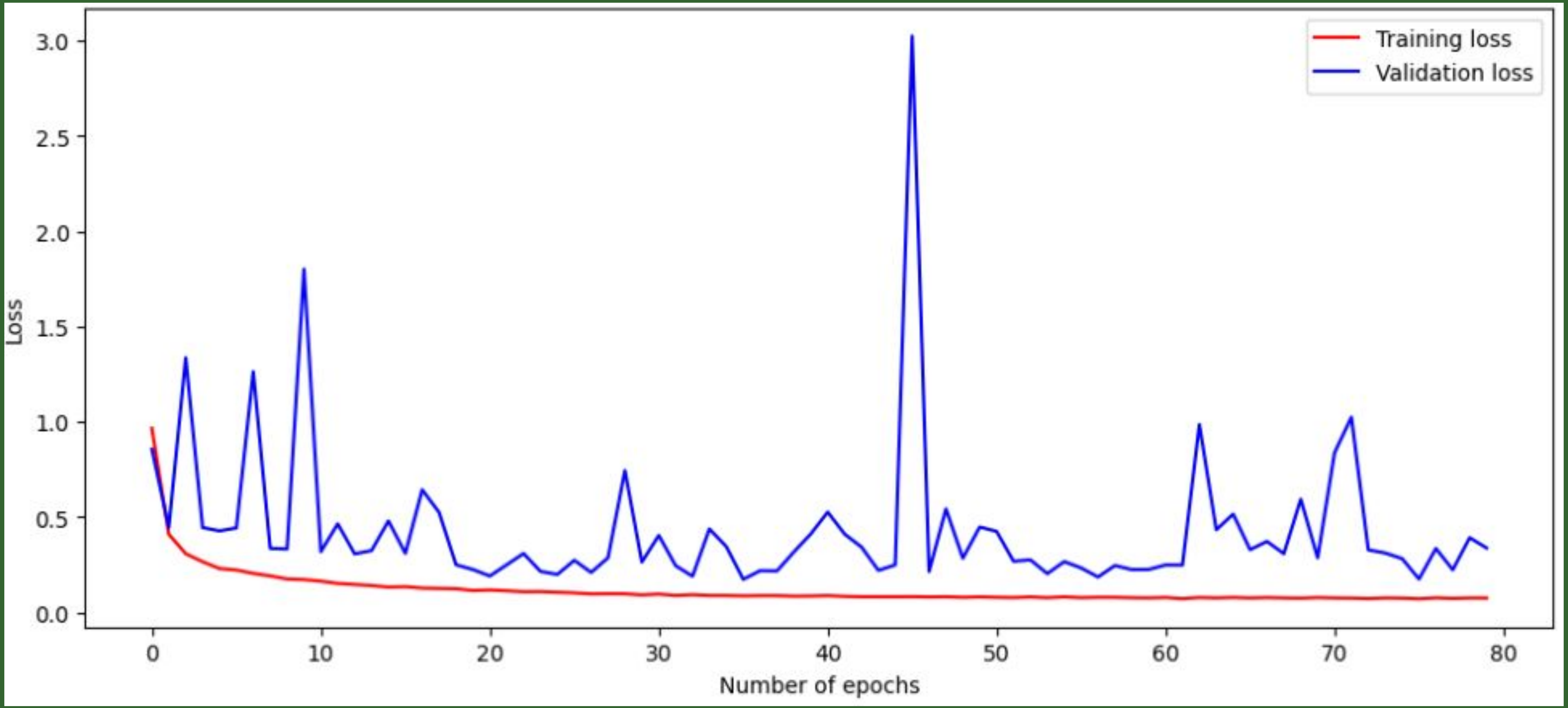


Figure-2(b): Training vs. Validation loss

CNN Performance Report

	Training	Validation	Testing
Accuracy	98.85%	93.32%	93.61%
Loss	0.075	0.338	0.300
Sparse Categorical Cross Entropy	0.036	0.299	0.262

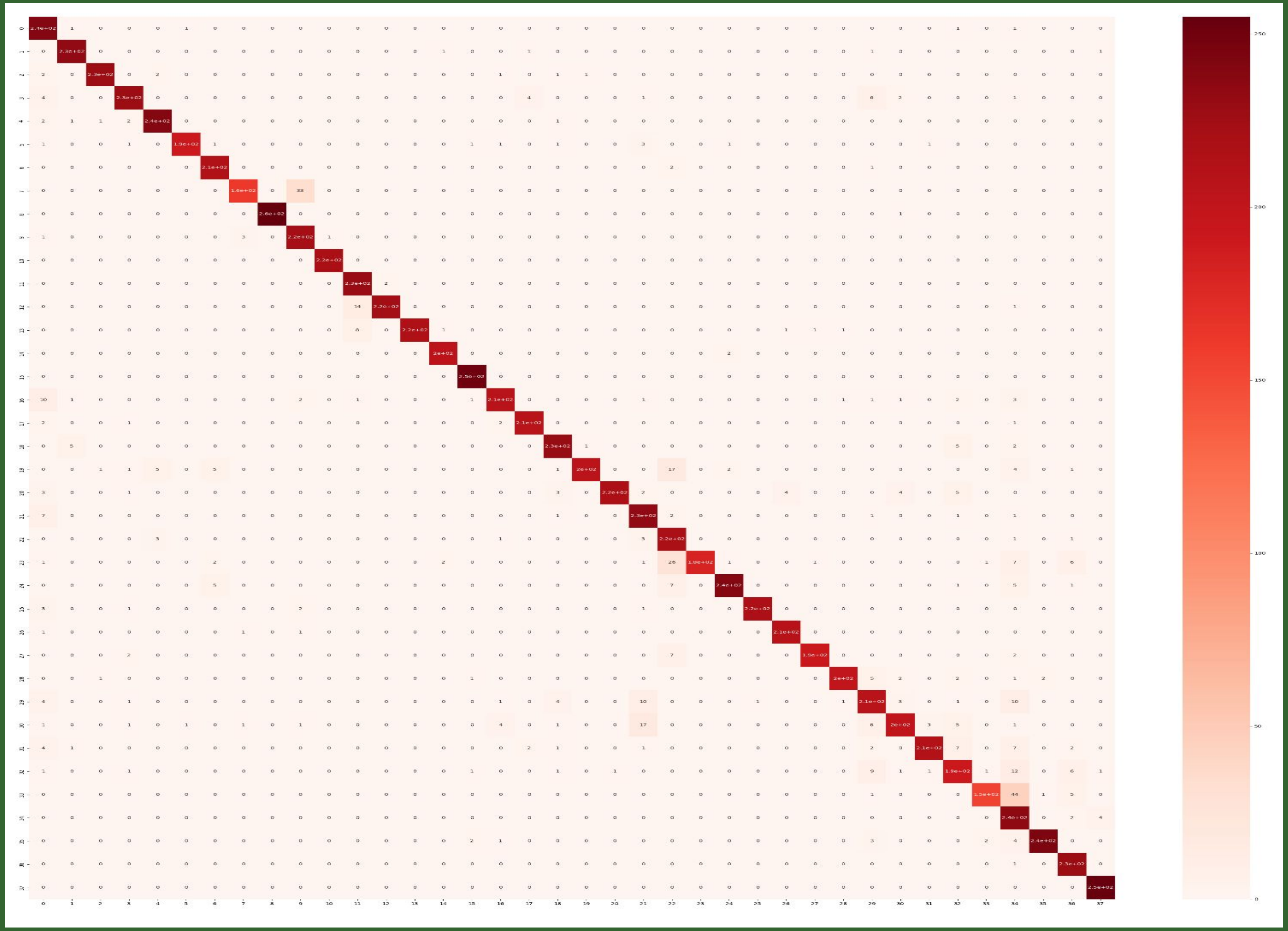
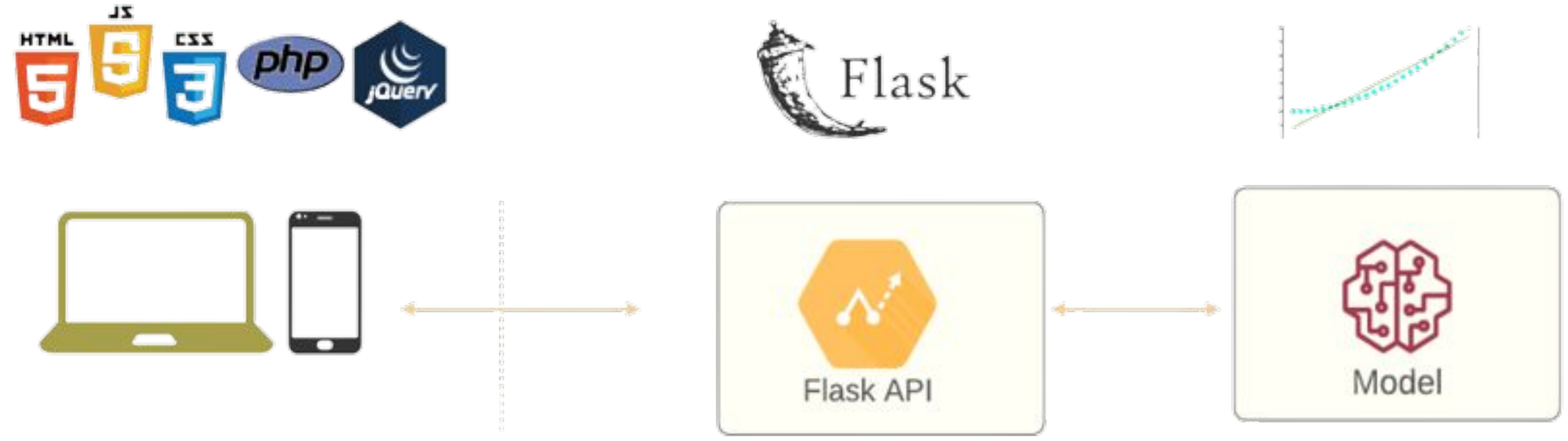


Figure-3: Confusion matrix for CNN

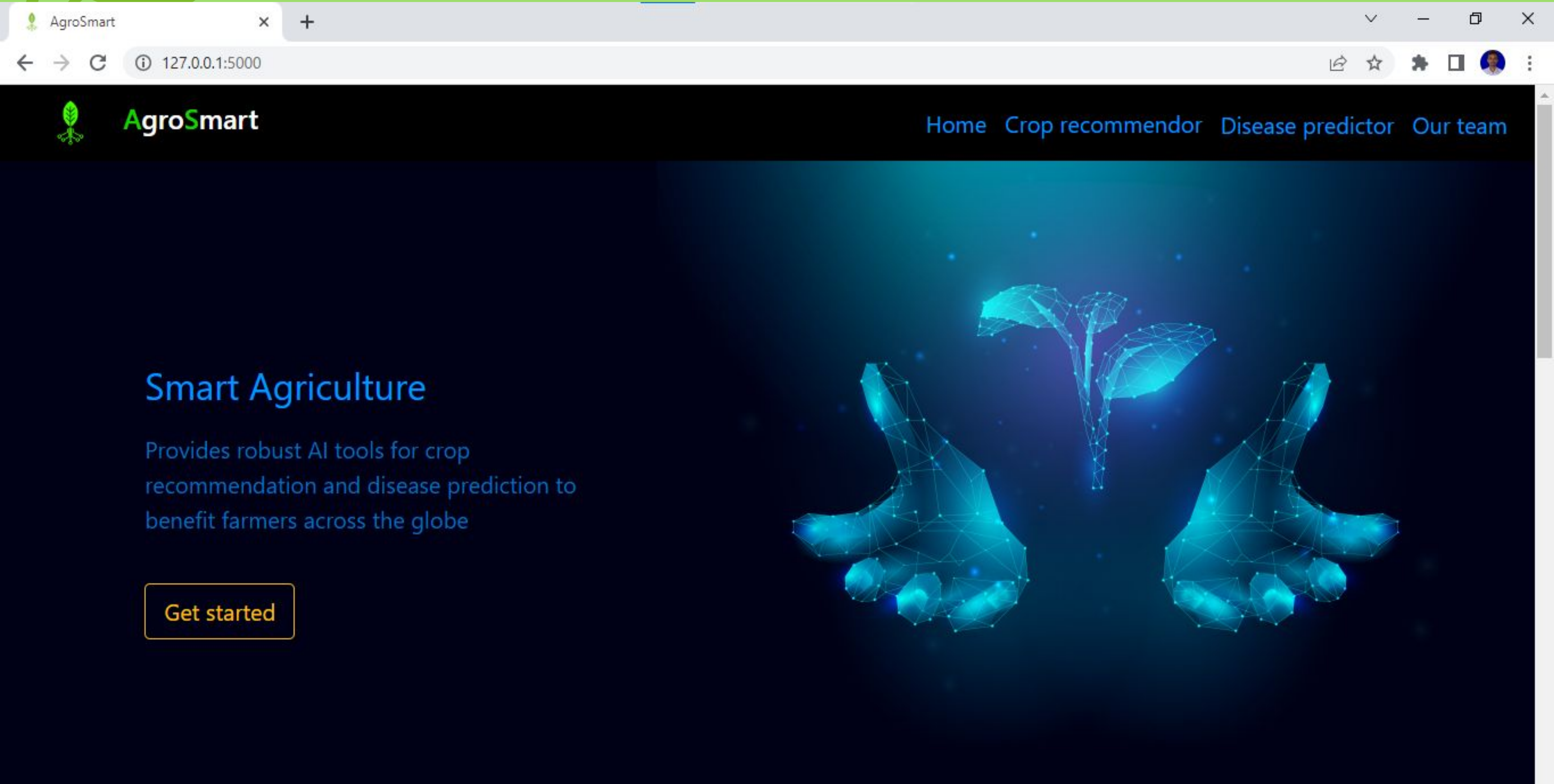
Web Application Development



Technologies Used

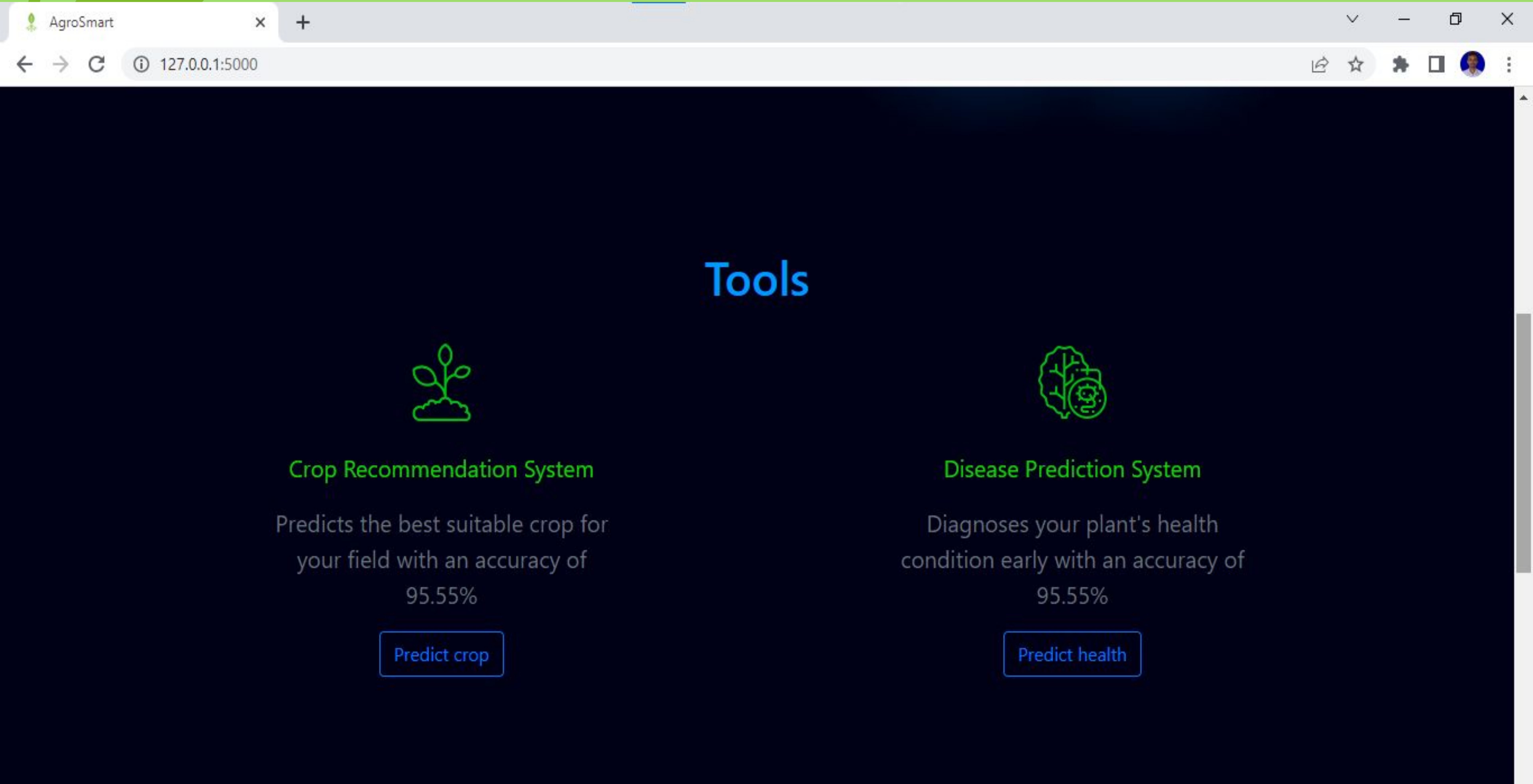
1. **HTML** : Hyper Text Markup Language, provides various tags to create web interfaces.
2. **CSS** : Cascading Style Sheets, adds look and feel to html pages
3. **Bootstrap** : Library built on CSS, provides various classes to simplify raw-CSS implementation.
4. **JQuery** : JavaScript library, injects dynamicness in web pages.
5. **Flask** : Light-weight Python’s backend engine

Home page

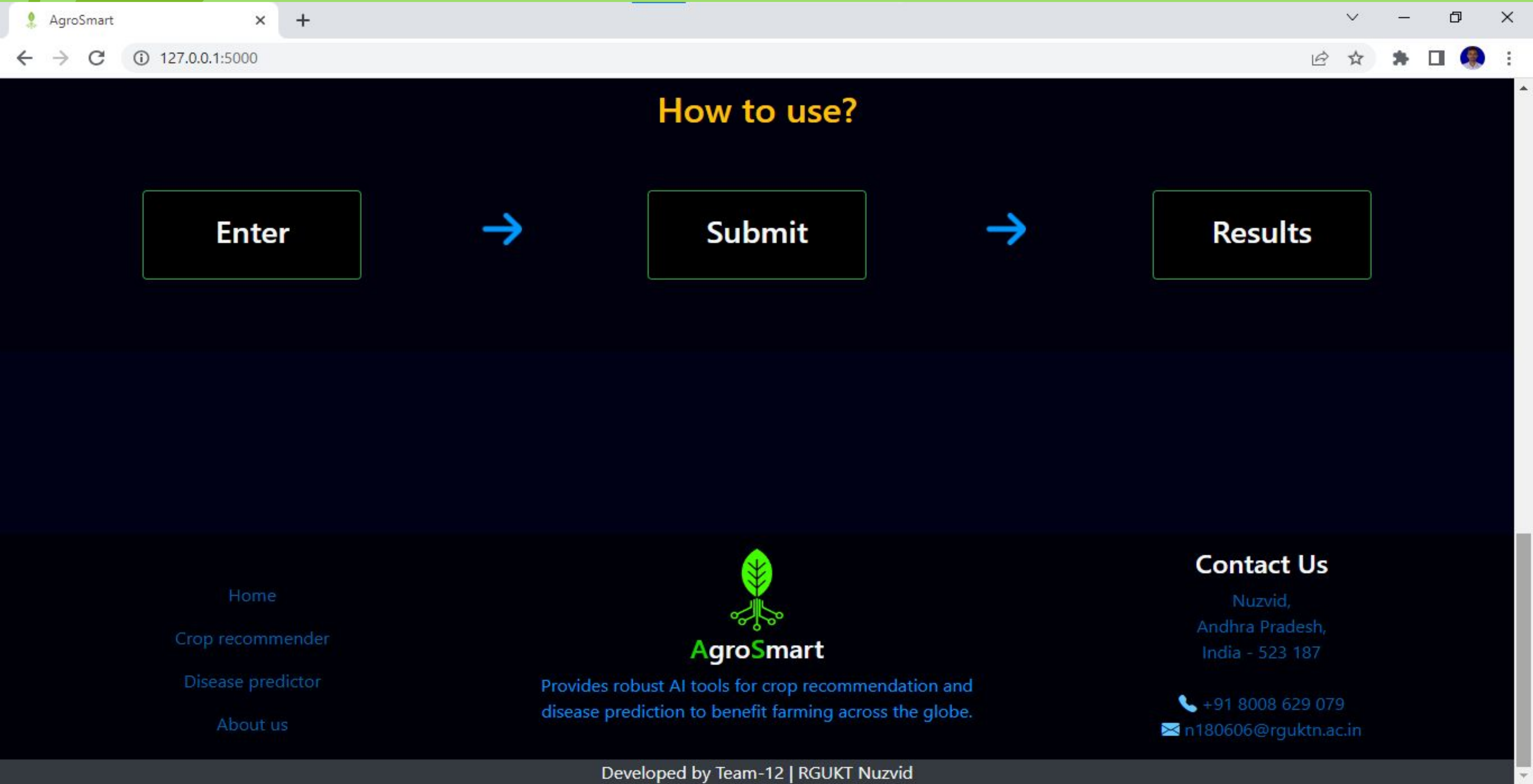


AgroSmart
127.0.0.1:5000
AgroSmart
Home
Crop recommendor
Disease predictor
Our team
Smart Agriculture
Provides robust AI tools for crop recommendation and disease prediction to benefit farmers across the globe
Get started

Home page

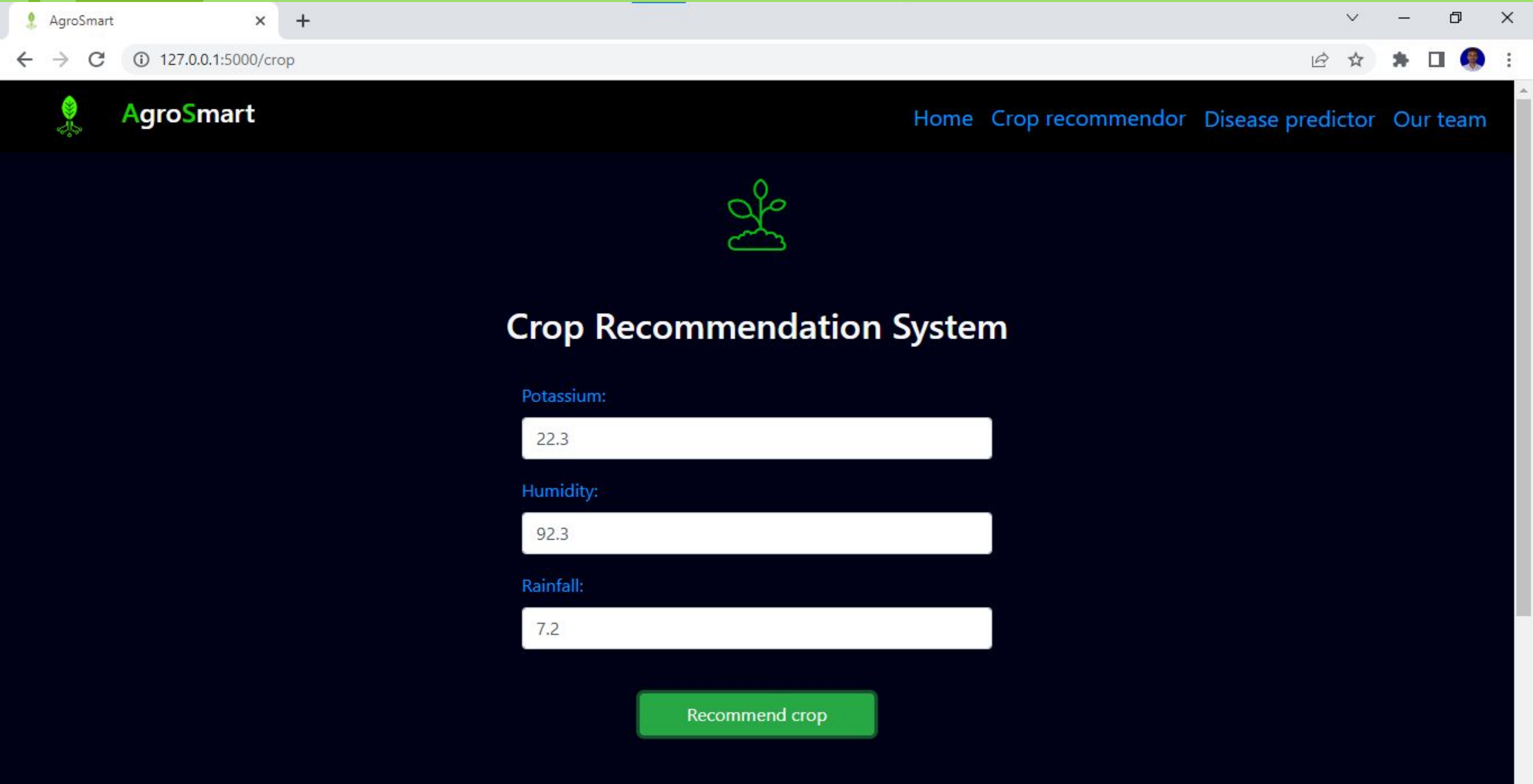

AgroSmart
127.0.0.1:5000
Tools
Crop Recommendation System
Predicts the best suitable crop for your field with an accuracy of 95.55%
Predict crop
Disease Prediction System
Diagnoses your plant's health condition early with an accuracy of 95.55%
Predict health

Home page



AgroSmart
127.0.0.1:5000
How to use?
Enter
Submit
Results
Home
Crop recommender
Disease predictor
About us
AgroSmart
Provides robust AI tools for crop recommendation and disease prediction to benefit farming across the globe.
Contact Us
Nuzvid,
Andhra Pradesh,
India - 523 187
+91 8008 629 079
n180606@rguktn.ac.in
Developed by Team-12 | RGUKT Nuzvid

Crop recommendation page



Crop recommendation page

AgroSmart

127.0.0.1:5000/crop

Crop recommended

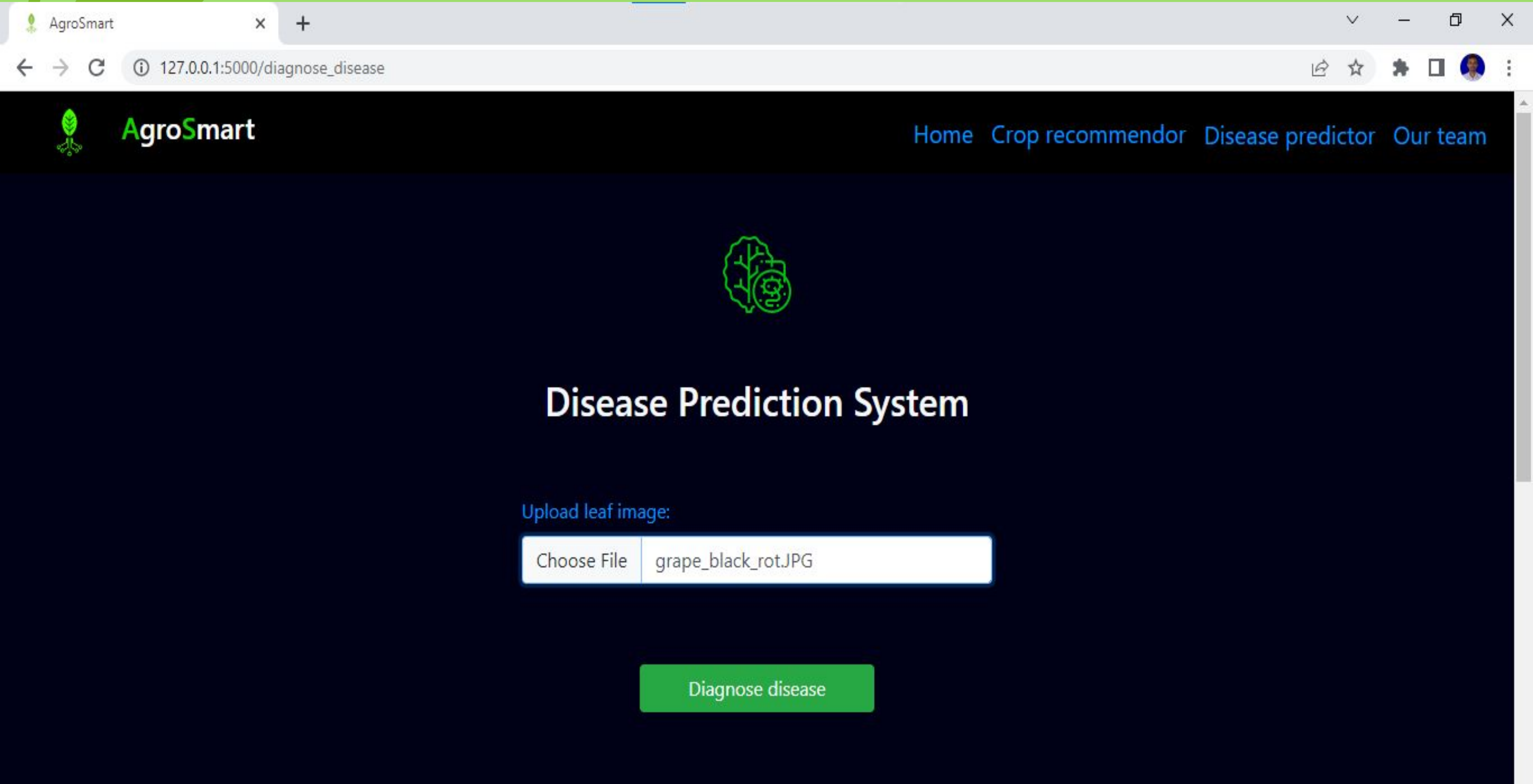
Muskmelon

Crop period	Season	Yield
110-120 days	Zaid	65 q/acre

Crop description

Muskmelon, also known as cantaloupe, is a flavorful fruit crop.It is characterized by its round shape, ribbed skin, and sweet aromatic flesh.Muskmelons come in different varieties, with varying colors and textures.They require warm weather and well-drained soil for optimal growth.Muskmelons are consumed fresh and are often used in fruit salads and desserts.

Disease prediction page



Disease prediction page

AgroSmart

127.0.0.1:5000/diagnose_disease

Disease diagnosed

Grape Black Rot

Treatment measures

Remove and destroy infected grapes. Apply fungicides as directed. Prune for proper airflow and sunlight exposure. Clean up and dispose of fallen leaves and debris. Practice good vineyard hygiene.

System Specifications

Processor	:	12th Gen Intel(R),Core(TM), i7-12700 2.10GHz
Installed RAM	:	16.0GB (15.7 GB usable),
System type	:	64-bit operating system, x64-based processor
Operating System	:	Windows 11
Programming Language	:	Python
Tool(s)	:	Jupyter Notebok

Open-Source Contributions

- Open-source contribution of optimized models in GitHub
- Developed new plotting technique “IQR Plot” for statistical summary
- Proposed an approach to generate confusion matrix using TensorFlow

Future Scope

We have developed web application for crop recommendation and disease prediction. As a future work, we are planning to design a full-fledged mobile application where farmers can enter, promote their products and sell their products directly to customers. We want to increase crop recommendation for at least 250 crops and disease prediction for at least 500 classes.

Conclusion

Agriculture is a vast sector. Profits earned mostly depend on the produced yield. Although there are many challenges in farming, it is the amount and quality of yield which benefits farmers. Our study mainly focuses on two main challenges of selecting suitable crop for a field and early detection of plant diseases. We have built robust machine learning models for crop recommendation with an accuracy of 97.05%, f1-score of 97.05%, and disease prediction with an accuracy of 93.61%, f1-score of 97.05%. The models are guaranteed with cross validation and regularization techniques. As it is ideal to nullify loss in CNN, we have used regularization techniques to minimize it to a greater extent.

Thank you