Saathi - An ML based crop recommendation and plant disease identification website

Sahil Gangurde(2019IMT-034)
Under supervision of: Dr. Pinku Ranjan

3rd Year Student - I.T. Department
ABV-IIITM Gwalior
Gwalior-474 010, MP, India

May 12, 2022

Outline

- Introduction
- System Architecture
- Dataset exploration
- Model - Training, Discussion and Designing
- Tools and technologies used
- Results
- Future work
- Conclusion
- References

Introduction

The project can be divided into three major parts

- Crop recommendation
- Plant Disease Detection
- Deployment over web server

The first two tasks are ML related while the last task is related to web development.

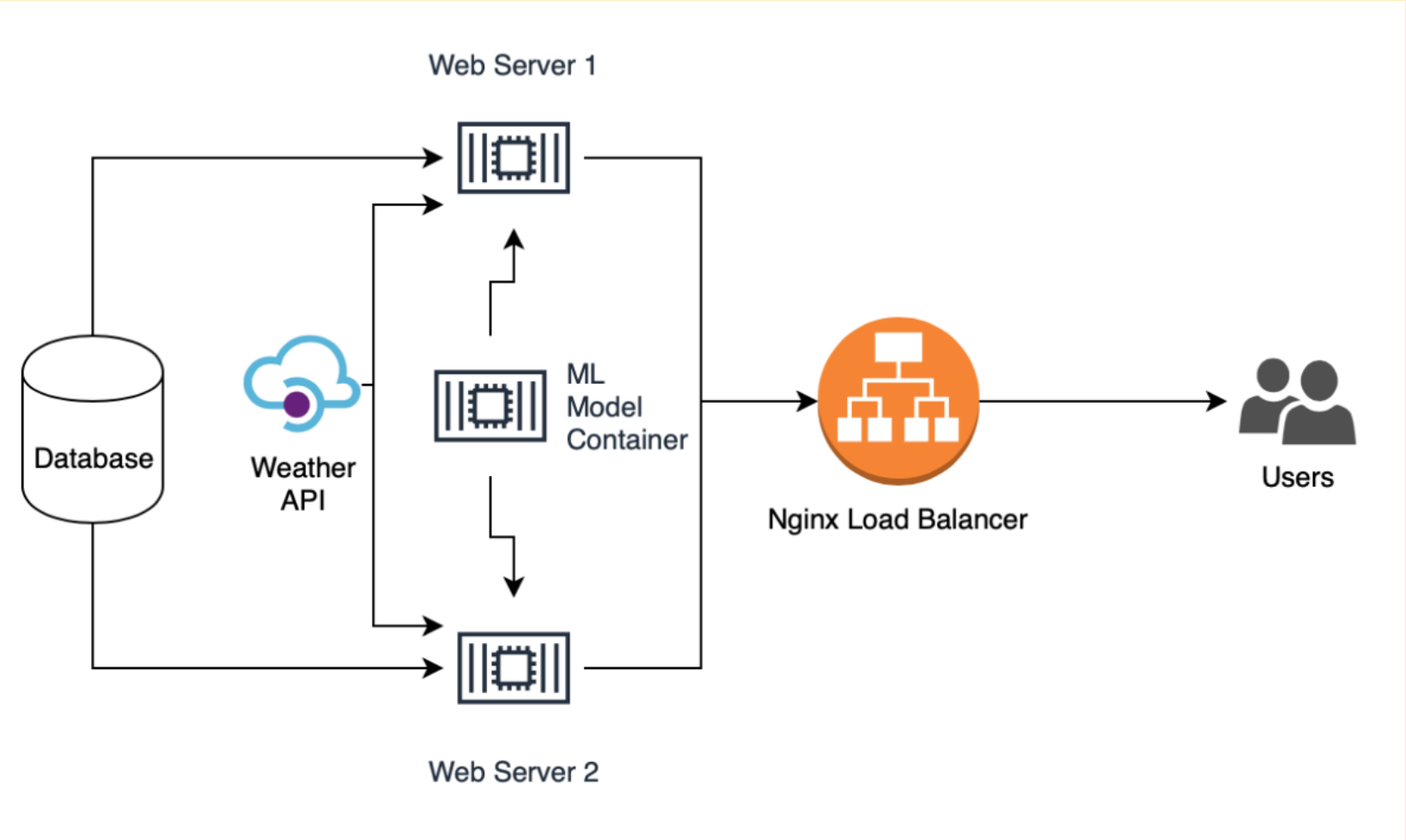
Contd..

This project aims at building a website focused on the agriculture sector solving two major issues crop recommendation and crop disease identification.

The method used to solve these problems is by training models on datasets available over the internet and comparing them.

Models with good accuracy are embedded into the website which is then deployed on cloud.

Overall System Architecture



Overall System Architecture

- Docker: Docker helps in containarization process. It creates a whole development environment whoch is easy to use and easily deployable across any server
- NginX: Nginx is used to create a load balancer. A load balancer help distrubute the traffic among multiple available servers using different algorithms.
- Tensorflow: It is an open-dource machine learning framework provided by Google.
- Keras: Keras is a deep learning library which provides powerful deep learning solutions
- MySQL: A relational database system
- Flask: A light-weight python web server

Crop Recommendation

For the crop recommendation the dataset used had 2200 rows and had 7 features and 20 output labels.

The task of recommendation was a simple classification task.

Five algorithms were chosen to do a comparative study.

Contd..

The five algorithms included

- Logistic Regression
- Decision Tree
- Support Vector Machine (SVM)
- Multi Layer Perceptron
- Random Forest

Contd..

The dataset for the crop recommendation had following data

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

Contd..

The columns available were as follows

```
RangeIndex: 2200 entries, 0 to 2199
Data columns (total 8 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   N            2200 non-null   int64
 1   P            2200 non-null   int64
 2   K            2200 non-null   int64
 3   temperature  2200 non-null   float64
 4   humidity     2200 non-null   float64
 5   ph           2200 non-null   float64
 6   rainfall     2200 non-null   float64
 7   label        2200 non-null   object
dtypes: float64(4), int64(3), object(1)
memory usage: 137.6+ KB
```

Plant Disease Detection

The task of disease detection using leafs of images is related to CNNs.

CNN can perform better on image classification tasks than the traditional ML algorithms.

3 different CNN architectures are been implemented and a comparative study between them is shown.

Contd..

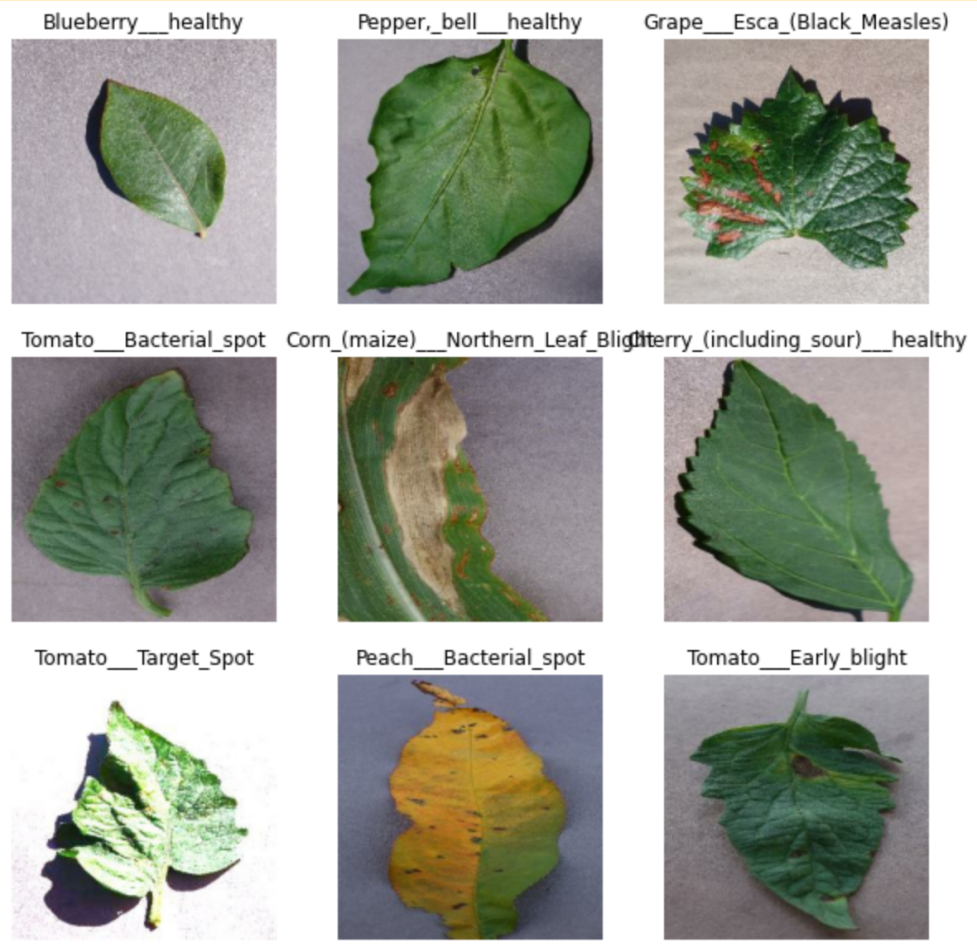
There are 3 different CNN architecture available for the image classification task.

- VGG16
- ResNet50
- EfficientNetV2S

Each model is trained with the dataset and a comparative study is shown.

Contd..

Sample images form the plant disease detection dataset



Contd..

The dataset consists of 70,000 plant images having various diseases.

All the images are of resolution 256x256 and are in RGB color mode.

There are in total 38 classes available where there are 14 different plants and 26 diseases to be identified.

Tools and Technology used

The libraries used for the machine learning model training are as follows

- Tensorflow
- Keras
- SkLearn

Results

For the first task, 5 classification algorithms were taken and then the accuracy over each of them were measured. For the task 2 three CNN architectures VGG16, ResNet50 and EfficientNetV2 are trained.

Results

Algorithm	Accuracy	Precision	Recall	F1-Score
Logistic Regression	94.54	0.95	0.95	0.94
Decision Tree	97.72	0.98	0.98	0.98
Support Vector Machine	9.09	0.59	0.09	0.11
Multi-Layer Percep-tron	95.22	0.96	0.95	0.95
Random Forest	99.31	0.99	0.99	0.99

Table: Crop recommendation task accuracy over various algorithms. This table concludes that the random forest outperforms every other algorithm achieving 99.31% accuracy

Contd...

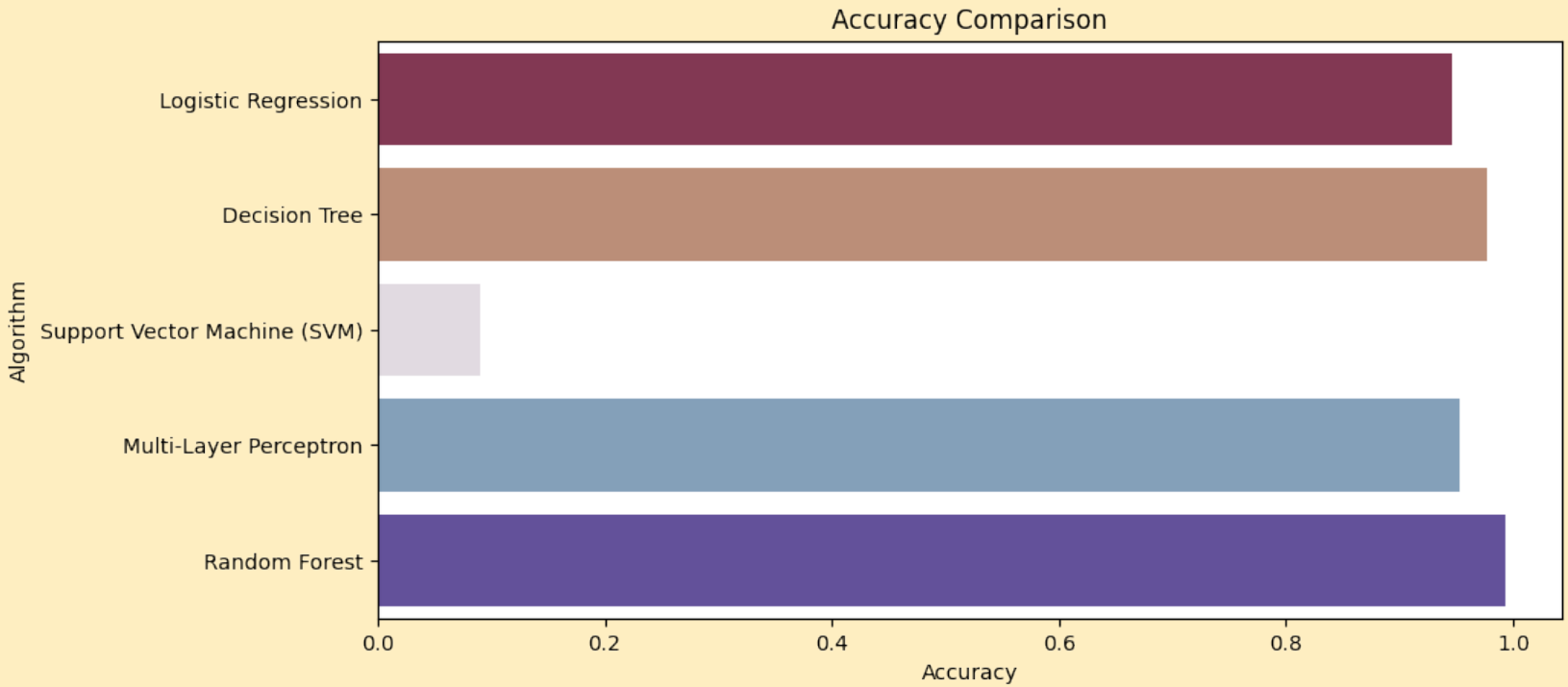


Figure: Histogram representation of the above table

Contd...

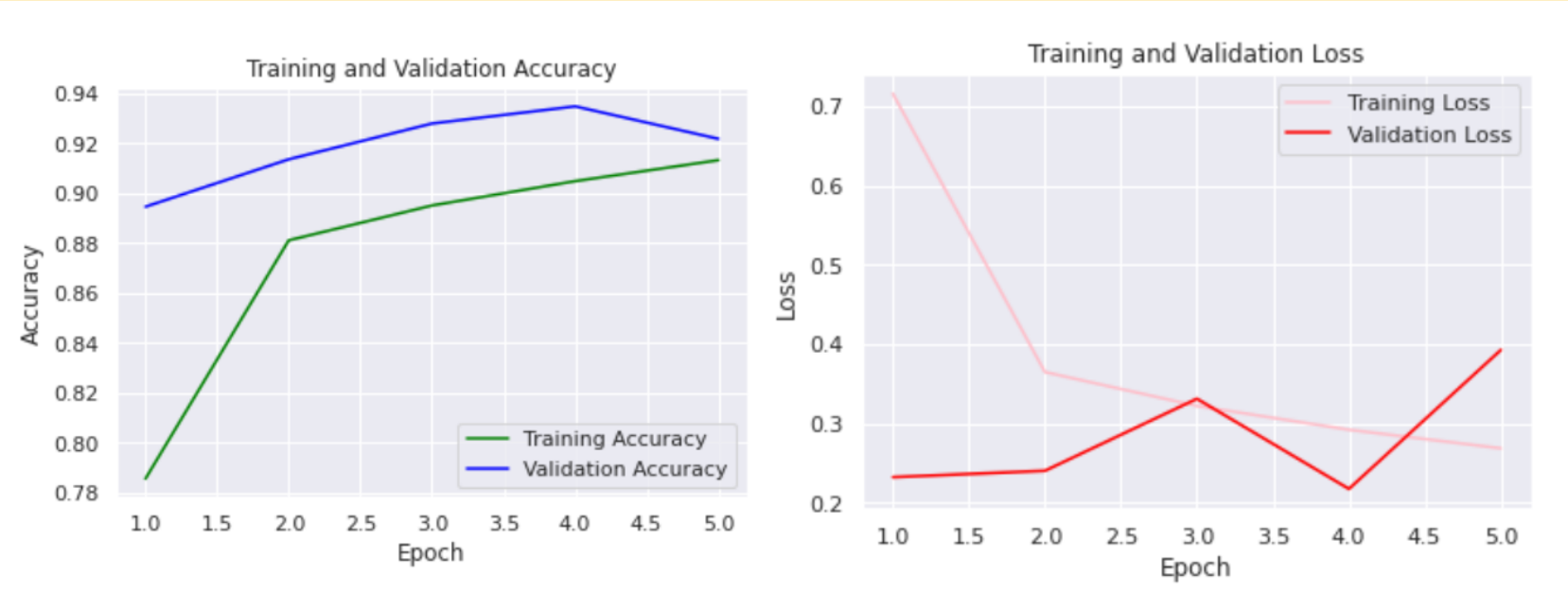


Figure: Accuracy and Loss graphs of the VGG model

Contd..

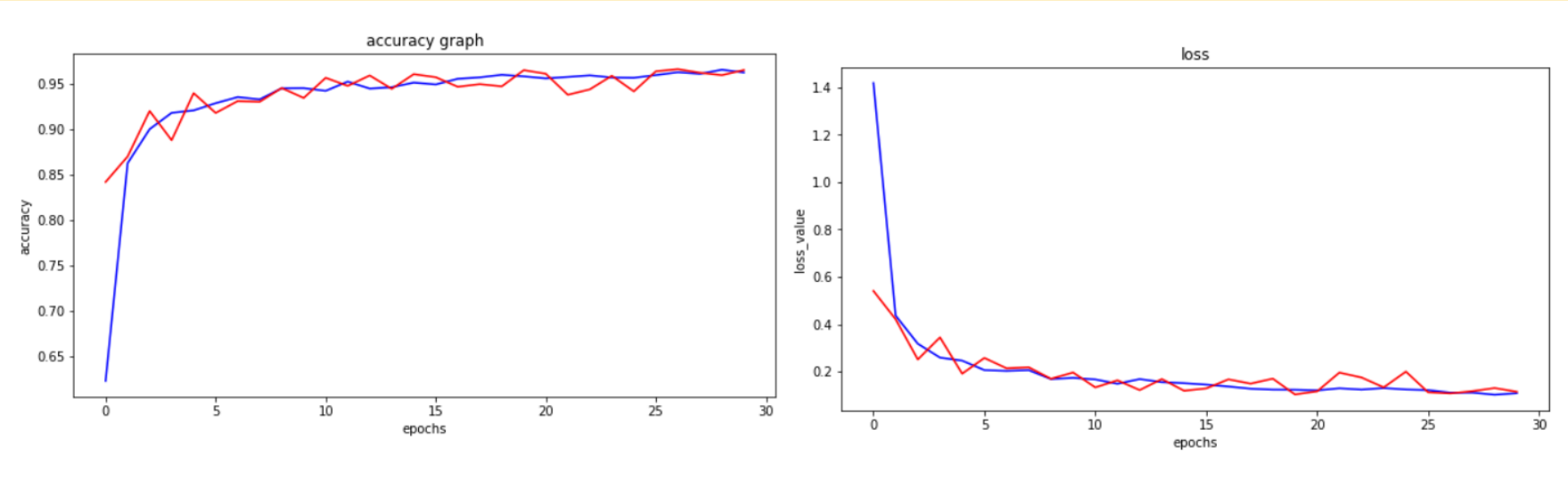


Figure: Accuracy and Loss graphs of the ResNet50 model

Contd..

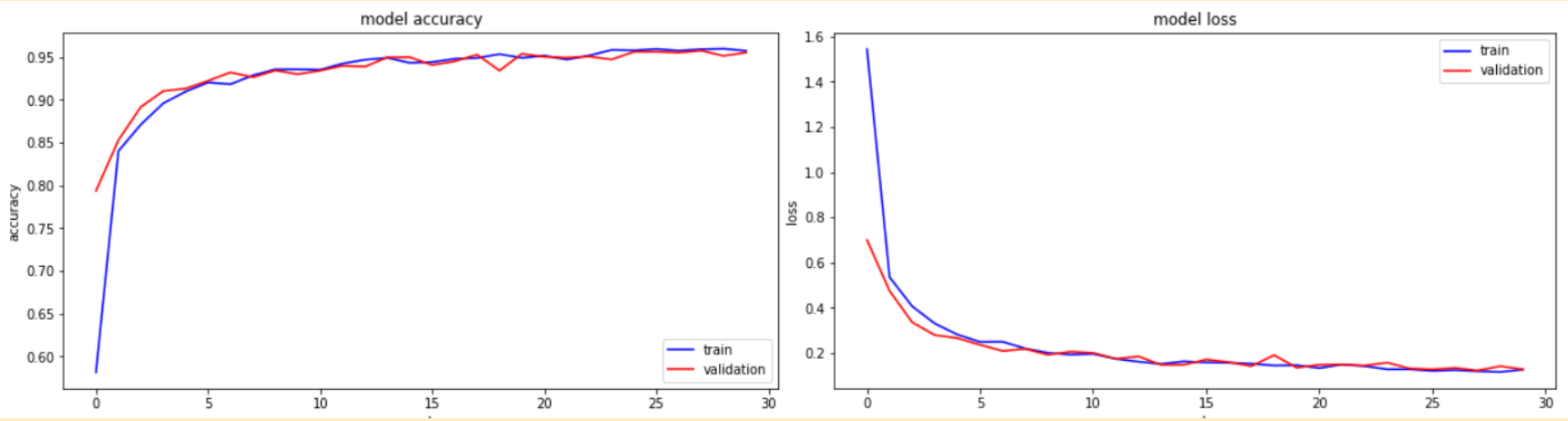


Figure: Accuracy and Loss graphs of the EfficientNetV2s model

Contd..

Architecture	Training Acc.	Validation Acc.	Test Acc.
VGG16	92.18	91.33	91.78
ResNet50	96.02	95.41	95.53
EfficientNetV2	96.06	95.53	95.83

Table: Plant Disease Classification task accuracy over various architectures. This table concludes that EfficientNetV2 is better.

Website

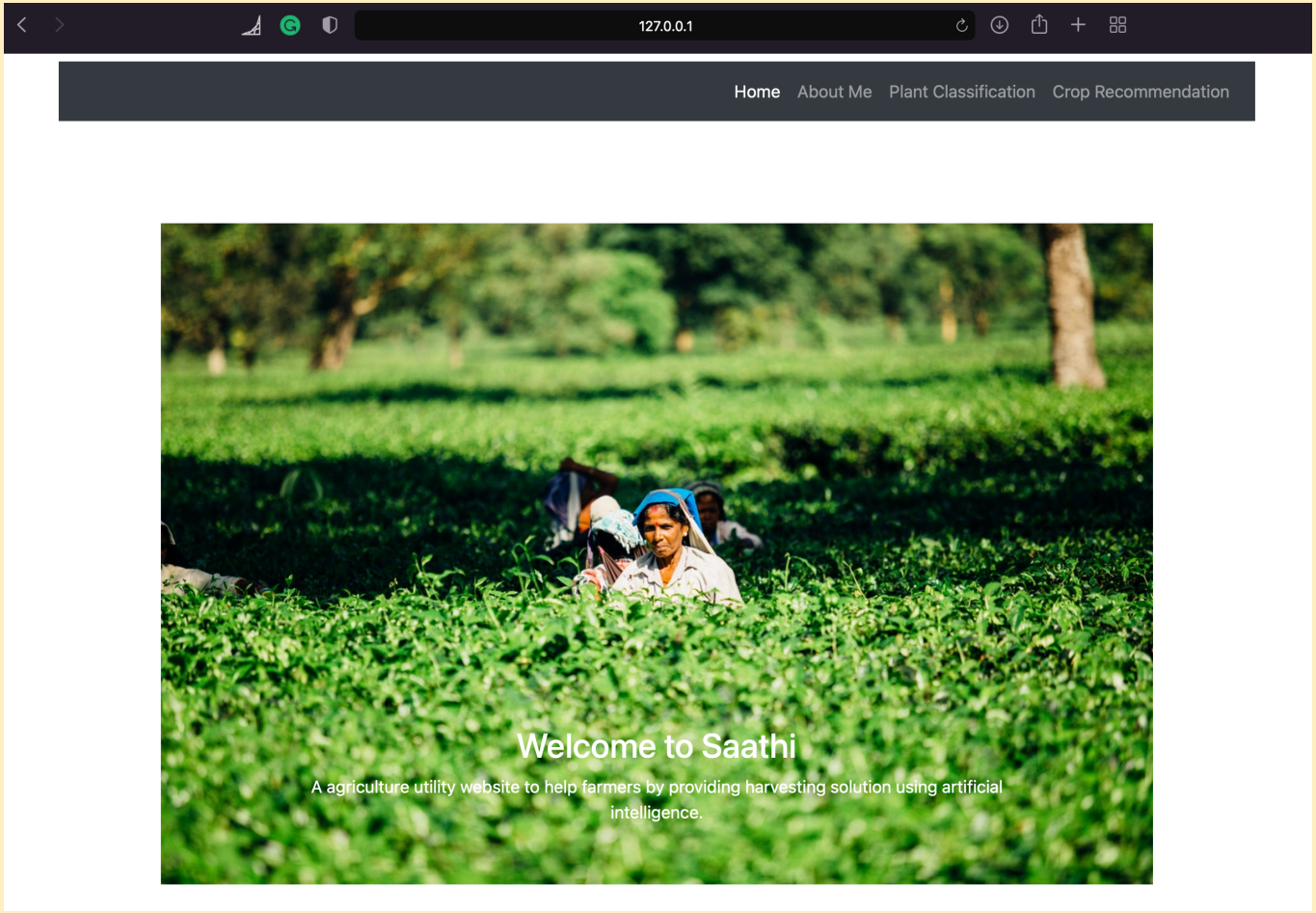


Figure: The first landing of the website

Contd..

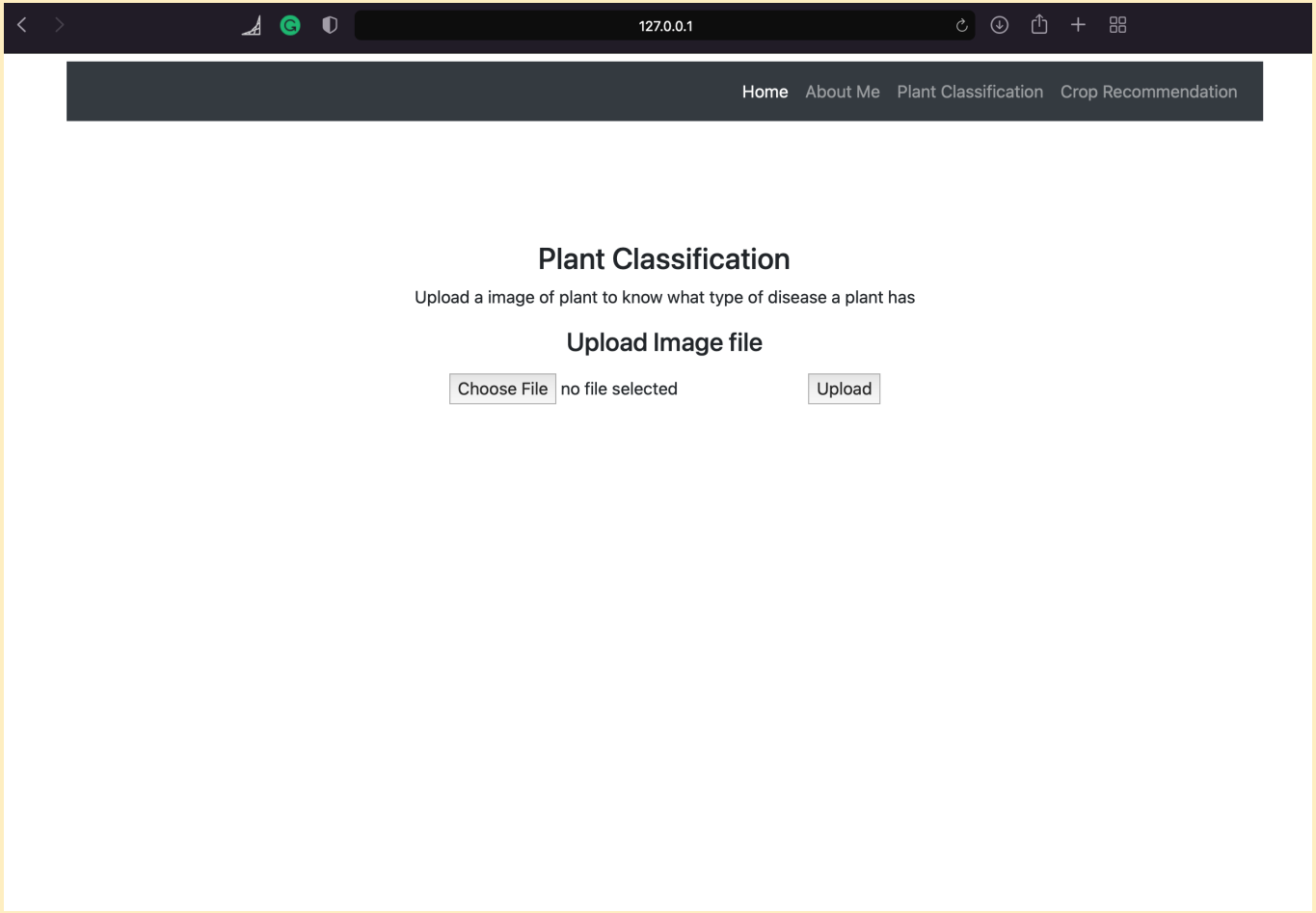


Figure: Plant disease detection landing page

Contd..

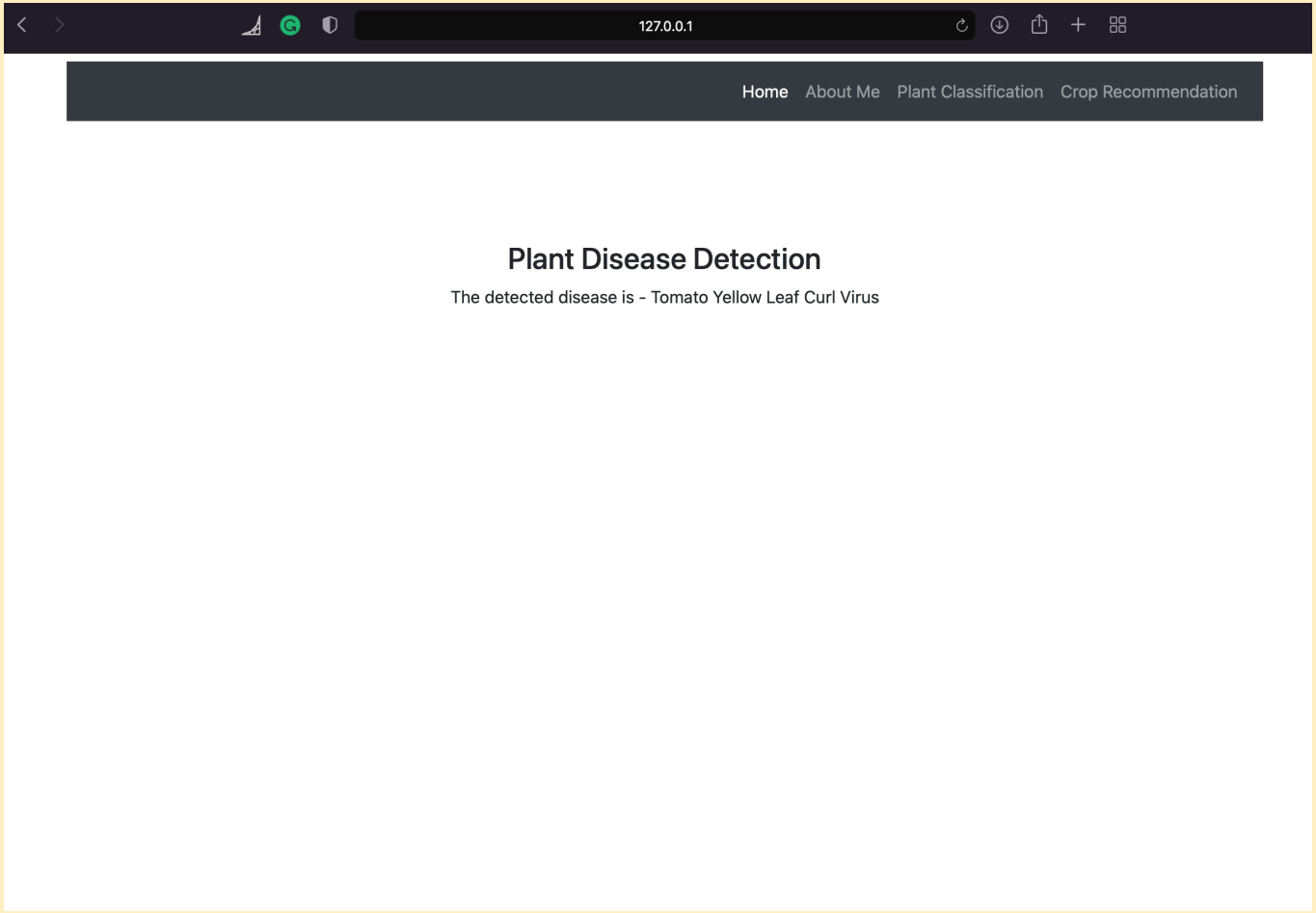


Figure: Plant disease detection prediction on website

Contd..

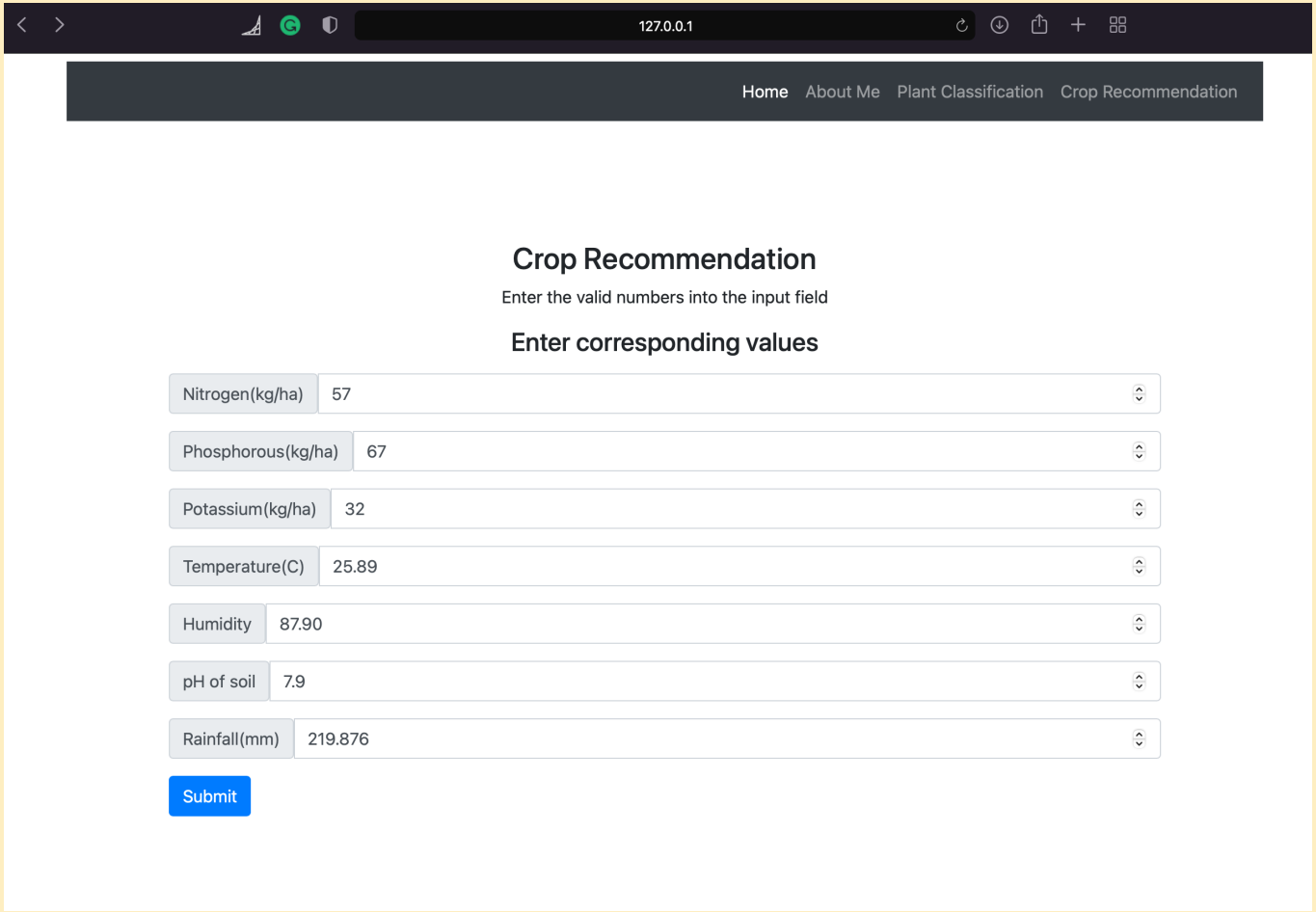


Figure: Crop Recommendation landing page

Contd..

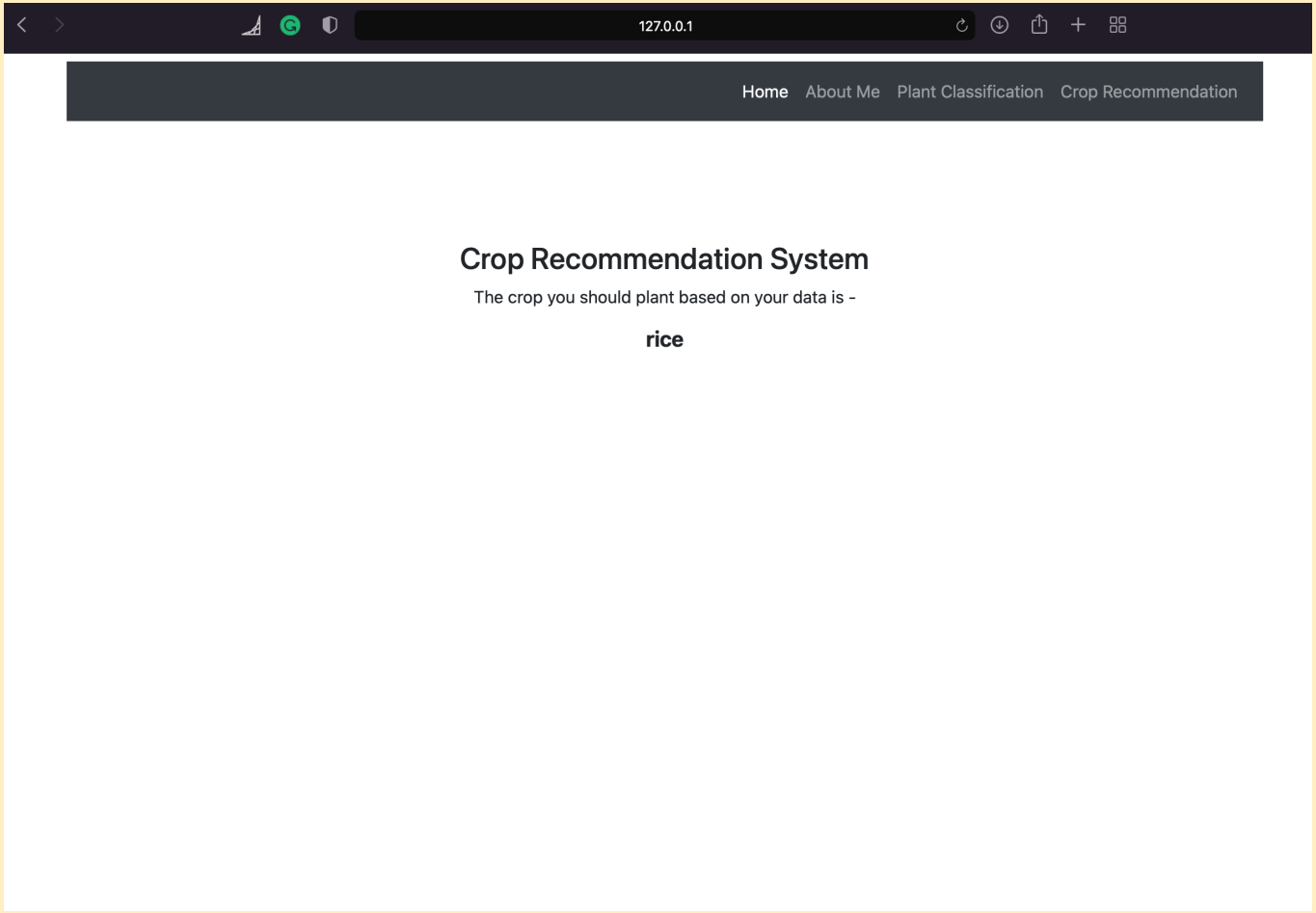


Figure: Crop recommendation prediction on website

Future Work

Transformer neural networks can be trained to achieve higher accuracy for plant disease detection.

For crop recommendation system the dataset can be increased to cover more features.

A device can be created which can directly detect the disease of the plant.

Conclusion

This project solves the problem of agricultural industry by providing the solution to a major problem of harvesting. We studied 5 different algorithms for the task 1 and reached to a conclusion that Random Forest is the best suited for the selected dataset.Random Forest achieved an overall accuracy of 99.3%. For the task 2 a comparative study was shown between VGG16, ResNet50 and EfficientNetV2S. EfficientNetV2 outperformed VGG16 and ResNet50 by achieving an overall accuracy of 96.06%. ResNet50 performed better than VGG16 gaining an overall accuracy of 95.53%. These two models were then deployed on web to be accessed by people.

Thank You

References

- Kulkarni, P., Karwande, A., Kolhe, T., Kamble, S., Joshi, A. and Wyawahare, M., 2021. Plant Disease Detection Using Image Processing and Machine Learning. arXiv preprint arXiv:2106.10698.
- G. Chauhan and A. Chaudhary, "Crop Recommendation System using Machine Learning Algorithms," 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), 2021, pp. 109-112, doi: 10.1109/SMART52563.2021.9676210.
- Sharada Prasanna Mohanty, David Hughes, Marcel Salathe, 2016, Using Deep Learning for Image-Based Plant Disease Detection. arXiv preprint arXiv:2106.10698.
- Hassan SM, Maji AK, Jasin ski M, Leonowicz Z, Jasin ska E. Iden- tification of Plant-Leaf Diseases Using CNN and Transfer-Learning Approach. Electronics. 2021; 10(12):1388