Plant Disease Detection using Machine Learning*

*Note: Sub-titles are not captured in Xplore and should not be used

Goutham U R
IMT2021045
IIIT Bangalore
Goutham.UR@iiitb.ac.in

Adithya Nagaraja Somasalle
IMT2021054
IIIT Bangalore
Adithya.Somasalle@iiitb.ac.in

Meher Ashish Nori
IMT2021085
IIIT Bangalore
meher.ashish@iiitb.ac.in

***Abstract*—This report outlines our approach to automated plant disease classification using machine learning. Drawing upon the Plant Village dataset, we strategically implemented feature extraction methods, optimizing model performance. Training various machine learning models revealed significant accuracy and F1 score improvements, highlighting the efficacy of our approach for efficient disease classification.**

***Index Terms*—machine learining, feature extraction, classification**

I. Introduction

In response to the escalating need for efficient plant disease classification, our project focuses on developing a lightweight model using various machine learning techniques. Leveraging the Plant Village dataset, we meticulously navigated through three key project phases. Initially, we strategically selected and analyzed the dataset, implementing feature extraction methods to optimize our model. Subsequently, we trained machine learning models such as Decision Trees, Random Forests, KNN, and XgBoost using original images. In the final phase, we incorporated feature extraction, noting a significant improvement in accuracy and F1 score. This report outlines our methodology and findings, contributing valuable insights to the realm of automated plant disease classification.

II. Dataset

The Plant Village dataset stands as a comprehensive repository, featuring a diverse collection of images showcasing leaves from three prevalent plant species:

Plants in the Dataset:

- Tomato
- Potato
- Pepperbell

Diseases/Classes:

1) Tomato__Tomato_mosaic_virus
2) Tomato_Late_blight
3) Tomato_Leaf_Mold
4) Tomato__Tomato_YellowLeaf__Curl_Virus
5) Tomato_Bacterial_spot
6) Tomato_Septoria_leaf_spot
7) Tomato_healthy
8) Tomato_Spider_mites_Two_spotted_spider_mite
9) Tomato_Early_blight
10) Tomato__Target_Spot
11) Potato__Early_blight
12) Potato__Late_blight
13) Potato__healthy
14) Pepper__bell__healthy
15) Pepper__bell__Bacterial_spot

Significance: The dataset holds significant value for researchers, developers, and practitioners in the fields of computer vision and machine learning, particularly within the domain of plant pathology. The diverse range of plant species and diseases included allows for robust exploration and development of algorithms for plant disease detection, yield prediction, crop monitoring, and agricultural research.

Use Cases:

a) Plant disease detection
b) Yield prediction
c) Crop monitoring
d) Agricultural research and development

Availability: The Plant Village dataset is readily accessible through online repositories such as Kaggle or GitHub, providing a convenient resource for those engaging in research and development within the realm of plant pathology, computer vision, and machine learning.

III. Methodology

A. Feature Extraction Techniques

In the domain of image analysis dedicated to the classification of plant diseases, the selection of appropriate feature extraction methods is pivotal for the development of accurate models. This section delineates the methodologies employed for feature extraction in our codebase, encompassing a spectrum of techniques designed to capture essential information from images.

- **Canny Edge Detection**: Canny edge detection is applied to identify and emphasize edges within the image.
  Purpose: Recognizing structural details and outlining boundaries in the plant images for effective disease classification.
- **Color Histogram**:A color histogram is utilized to represent the distribution of pixel intensities for each color channel.

Identify applicable funding agency here. If none, delete this.

Fig. 1. Pepper with Bacterial Spot Disease

Fig. 2. Grey Scale Image

Purpose: Summarizing color information to differentiate between healthy and diseased plants based on distinct color patterns.

- **GLCM Texture Features**: The Gray Level Co-occurrence Matrix (GLCM) is utilized to quantify texture by analyzing the co-occurrence of pixel intensities.
  Purpose:Capturing textural nuances in the images, enhancing the discrimination between different types of plant diseases.

For Implementation we have used computer vision libraries like **cv2, mahotas**.(cv2 for Canny Edge and Color Histogram & mahotas for GLCM Texture Extraction).

The amalgamation of these diverse feature extraction techniques affords a comprehensive set of features, enhancing the depth of our image analysis for plant disease classification.

B. Models Used

- K-Nearest Neighbours Classifier

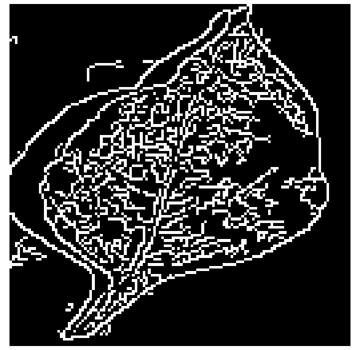

Fig. 3. Canny Edge Detection

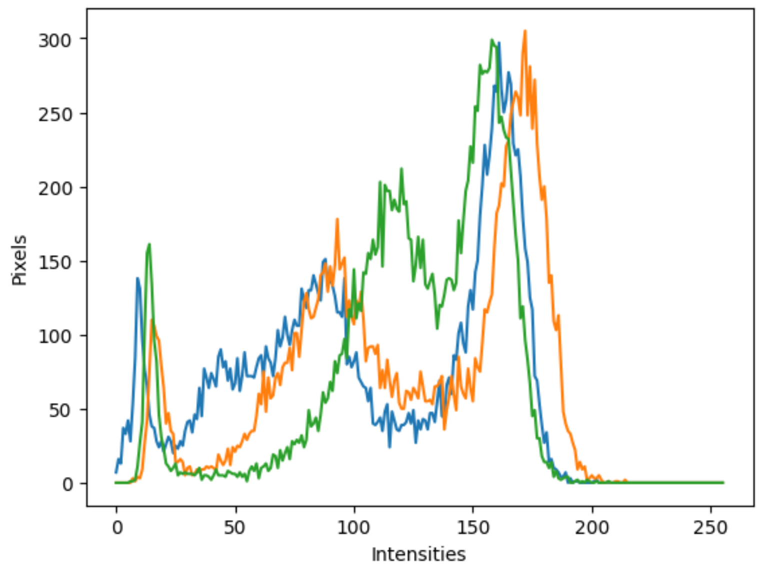

Fig. 4. Color Histograms (B-blue, G-green, R-orange) for Pepper with Bacterial Spot disease

- Decision Tree Classifier
- Random Forest Classifier
- XGBoost Classifier

Results and Conclusion

TABLE I
PLANT DISEASE CLASSIFICATION RESULTS

Plant	Model	Accuracy Score	F1 Score
Tomato	KNN	0.533	0.53
Tomato	Random Forest	0.69	0.68
Tomato	XgBoost	0.841	0.84
Tomato	Decision Tree	0.416	0.414
Potato	KNN	0.69	0.663
Potato	Random Forest	0.876	0.848
Potato	XgBoost	0.942	0.938
Potato	Decision Tree	0.740	0.739
PepperBell	KNN	0.715	0.70
PepperBell	Random Forest	0.882	0.861
PepperBell	XgBoost	0.945	0.945
PepperBell	Decision Tree	0.77	0.76

Considering Features: Feature extraction led to improved performance in every model when considering features.

Intermodal Comparisons XG-Boost outperformed KNN, Decision Trees, and Random Forests, delivering the most favorable results among the models.

Best Results Best results were observed when XG-Boost was used along with feature extraction, and the accuracy obtained was 94.5%.

REFERENCES

[1] S. C. Madiwalar, M. V. Wyawahare , ‘Plant Disease Identification: A Comparative Study‘, 2017.

[2] S. P. Mohanty,, D. Hughes, d M. Salathé , ‘Using Deep Learning for Image-Based Plant Disease Detection‘, 2016.

[3] E. C. Tetila, B B Machado ‘Automatic Recognition of Soybean Leaf Diseases Using UAV Images and Deep Convolutional Neural Networks‘, 2020.

[4] A. A. Alatawi, S. H. Alomani ‘ Plant Disease Detection using AI based VGG-16 Model‘, 2022.

[5] Y. Es-saady, I E Massi ‘Automatic recognition of plant leaves diseases based on serial combination of two SVM classifiers‘, 2016.