Plant Disease Prediction System Using CNN

Vishwa Pathak
Computer Scinece and Engineering
R.N.G. Patel Institute Of Technology
Bardoli,Navsari
cse.230843131019@gmail.com

Faaizabegam Saiyad
Computer Scinece and Engineering
R.N.G. Patel Institute Of Technology
Bardoli,Navsari
cse.220840131126@gmail.com

***Abstract*—Detecting plant diseases early can help prevent crop loss and improve agricultural productivity. In this work, we introduce a smart and easy-to-use system that identifies plant leaf diseases using Convolutional Neural Networks (CNNs). The model is trained on the well-known PlantVillage dataset and achieves accuracy in classifying different plant diseases. This solution is designed to support farmers, gardeners, and agricultural professionals by providing fast, reliable, and affordable plant disease detection as part of modern smart farming practices.**

Keywords—Plant Disease Detection, Convolutional Neural Network (CNN), Deep Learning, Image Classification

I. INTRODUCTION

Agriculture remains a cornerstone of global food security. However, plant diseases can devastate crop yield and quality, often going undetected until significant damage occurs. Traditional methods of detection rely heavily on expert analysis and field visits, which are costly, time-consuming, and subject to human error.

With the recent surge in AI and Deep Learning, particularly CNNs, computer vision techniques can now be employed to automate plant disease detection [1]. This paper presents a CNN-based approach, integrated into a Streamlit web app and deployed using Docker, to offer a real-time, scalable, and easy-to-use solution for identifying plant diseases from images.

II. LITERATURE REVIEW

Over the past few years, many researchers have explored the use of deep learning to detect plant diseases from images. Traditional machine learning methods like Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) have been used with handcrafted features, but they often lack accuracy and scalability .

Recent work shows that Convolutional Neural Networks (CNNs) can outperform these older techniques by learning features directly from images. Mohanty et al. [2] trained several CNN architectures using the PlantVillage dataset and achieved high accuracy in disease classification.

Other efforts used transfer learning and ensemble methods to improve results [3], but often lacked deployment focus. Unlike prior works, our system prioritizes both accuracy and usability, combining efficient CNN architecture, an intuitive UI, and full Dockerization.

III. METHODOLOGY

A. Dataset

We used the PlantVillage dataset [4], which includes over 50,000 labeled images across 38 classes, including healthy and diseased leaves of crops such as tomato, potato, and grape. The images were resized to 224×224 pixels and categorized by folder name.

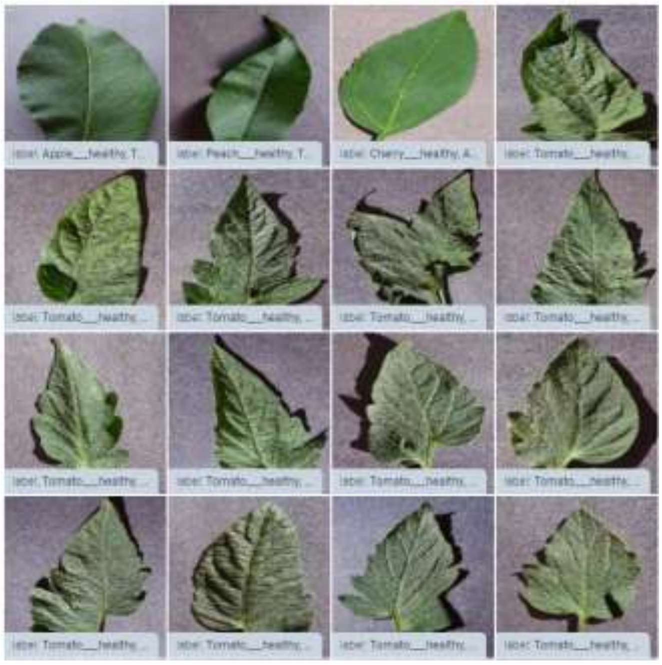

Fig. 1 – Sample Images from PlantVillage Dataset

B. Data Preprocessing

To ensure accurate predictions, the raw plant leaf images required several preprocessing steps before being used to train the CNN model. These steps helped improve image consistency and model performance, especially since the data came from diverse sources and varied in quality [5].

The preprocessing involved the following techniques:

- Image Resizing: All images were resized to 128×128 pixels to ensure uniform input dimensions and reduce computational load.
- Normalization: Pixel values were scaled to the range [0, 1] by dividing by 255, enabling faster and more stable model training.
- Data Augmentation: To increase dataset diversity and reduce overfitting, we applied transformations such as random rotation, horizontal/vertical flips, zooming, and brightness adjustment.
- Noise Reduction: Basic filtering techniques like Gaussian blur were used in some cases to reduce background noise and improve feature visibility.
- Label Encoding: Disease labels were converted into one-hot encoded vectors suitable for multiclass classification.

These preprocessing steps helped the model learn important features more effectively, ultimately leading to better accuracy when identifying plant diseases from new images.

C. *CNN Architecture*

Our CNN model consists of three convolutional blocks, followed by fully connected layers and a softmax output. It is compact, efficient, and tailored for image classification tasks.

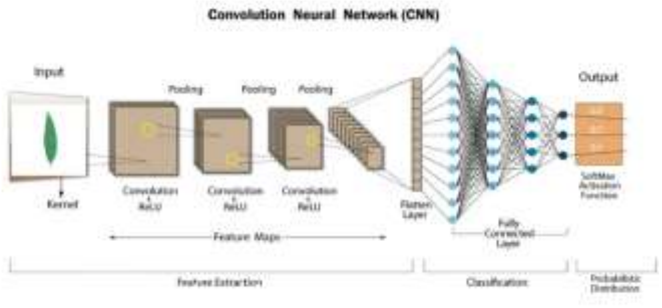

Fig. 2 – CNN Model Architecture (Flowchart)

Mathematical Formula:

The convolution operation is defined as:

$$f(x,y) = \sum_{i=0}^{m} \sum_{j=0}^{n} I(x+i, y+j), K(i,j) \qquad (1)$$

Where:

I = input image

K = kernel

f = output feature map

In our architecture, the first block uses 32 filters of size 3×3 with a ReLU activation, followed by a max pooling layer that reduces the spatial dimensions. The second and third blocks progressively increase the number of filters to 64 and 128, respectively, maintaining the same structure. These blocks extract low- to high-level features such as edges, textures, and disease-specific patterns.

After the convolutional layers, the output is flattened into a one-dimensional vector, which is then passed through a dense (fully connected) layer with 128 neurons and a ReLU activation. To reduce overfitting, a dropout layer is applied, randomly disabling 50% of the neurons during training. Finally, a softmax layer maps the output to a probability distribution across multiple plant disease classes.

This layered design enables the model to automatically learn hierarchical representations of leaf images, making it highly effective for disease classification even under variable field conditions.

D. *Training the Model*

To train our CNN model, we selected Adam as the optimizer due to its adaptive learning rate capabilities, which help accelerate convergence and improve training stability. The categorical cross-entropy loss function was used, as it is well-suited for multi-class classification problems where each input belongs to exactly one category.

The model was trained for 5 epochs with a batch size of 32. This configuration balanced training time with performance, ensuring that the model had enough iterations to learn features without overfitting. The training dataset consisted of labeled plant leaf images, while a separate validation set was used to monitor generalization performance.

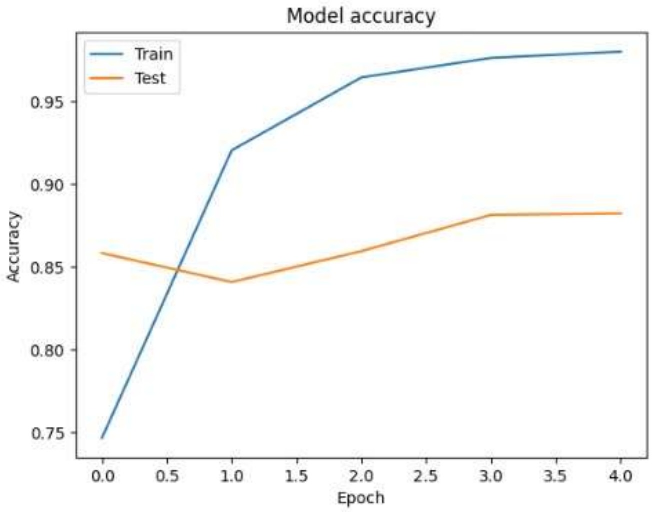

Fig. 3 – Accuracy vs. Epoch Graph

Throughout training, both training and validation accuracy gradually improved, indicating that the model was effectively learning the distinguishing features of various plant diseases. The performance was tracked using accuracy as the primary metric, providing insight into the model's ability to correctly classify unseen images.

IV. RESULTS

After training the Convolutional Neural Network (CNN) model on the PlantVillage dataset, with a 20% validation split, the model demonstrated impressive classification accuracy. The architecture consisted of three convolutional layers followed by dense layers, and the Adam optimizer was used along with categorical cross-entropy loss to train the model. The training process showed steady improvement, with accuracy increasing over five epochs. The accuracy and loss plots revealed that the model was successfully learning and generalizing to unseen data. When tested with sample leaf images, the model was able to accurately identify diseases across various plant species, showcasing its potential for real-world applications.

V. CONCLUSION AND FUTURE WORK

In this study, we developed a CNN-based system for detecting plant diseases using leaf images. The model achieved high accuracy during training and testing, demonstrating its reliability in identifying various plant diseases. By using a well-structured CNN architecture and the PlantVillage dataset, we were able to build a solution that could support early diagnosis of crop issues. This approach highlights how deep learning can be applied effectively in agriculture to assist farmers and researchers in making timely and informed decisions.

Future enhancements may include:

- Expanding to more plant species
- Integrating GPS/location for disease spread analytics

- Deploying on cloud platforms with real-time notifications
- Collecting real-world images to improve generalization

Acknowledgment

We would like to express our sincere gratitude to everyone who supported us throughout this project. Special thanks to our mentors and teachers for their constant guidance and encouragement. We also acknowledge the creators of the PlantVillage dataset and the open-source community for providing valuable tools like TensorFlow which made this project possible. Lastly, we are grateful to our friends and family for their continuous motivation and support during this journey.

References

[1] J. Brownlee, Deep Learning for Computer Vision, Machine Learning Mastery, 2019.

[2] S. Mohanty, D. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, p. 1419, 2016.

[3] H. Too et al., "A Comprehensive Review of Machine Learning Approaches for Plant Disease Detection," Computers and Electronics in Agriculture, vol. 163, 2019.

[4] Kaggle, “PlantVillage Dataset,” [Online]. Available: https://www.kaggle.com/datasets/emmarex/plantdisease

[5] F. Chollet, Deep Learning with Python, Manning Publications, 2017.