Crop Disease Detection and Prediction Using Computer Vision and Environmental Data

Submitted by:

Ananya.K.Bijoy
Msci (Integrated) Data Science

Goa Business School,Goa University

Academic Year: 2025

Contents

Abstract

This project presents an integrated framework for early detection and prediction of crop diseases through the combined use of computer vision and environmental data. A Convolutional Neural Network (CNN) is employed to classify crop leaf images based on disease symptoms, while a Random Forest model utilizes weather data to predict conditions favorable for pest or disease outbreaks. The outputs from both models are integrated into a unified risk prediction pipeline, offering actionable insights to farmers and agricultural stakeholders. This approach aims to reduce yield losses, improve decision-making, and enhance food security through timely disease management and environmental forecasting.

—

1 Problem Statement

Can we detect and predict the onset of crop diseases using leaf images and environmental data before the damage becomes too severe?

Crop diseases are a major challenge to global food security. Many small-scale farmers experience heavy losses due to delayed diagnosis of infections caused by fungi, bacteria, or viruses. The lack of early detection mechanisms and reliable forecasting systems leads to reactive rather than preventive responses.

This project leverages **computer vision** and **environmental modeling** to build an intelligent system capable of both identifying visible symptoms and forecasting disease-prone conditions, enabling proactive agricultural management.

—

2 Objectives

1. Develop a computer vision model for accurate classification of crop diseases from leaf images.
2. Construct a weather-based predictive model to forecast the likelihood of disease occurrence.
3. Integrate both models into a single predictive system capable of real-time disease risk assessment.
4. Deploy the integrated system through a user-friendly dashboard for practical field use.

—

3 Methodology

The project is structured into five primary stages: **data collection, preprocessing, model development, integration, and visualization.**

3.1 Data Collection

- **Leaf Image Data:** Acquired from the *PlantVillage Dataset*, containing labeled images of healthy and diseased crop leaves.
- **Environmental Data:** Weather parameters (temperature, humidity, rainfall, wind speed) collected via the *OpenWeatherMap API*.

3.2 Data Preprocessing

- Image normalization, resizing, and augmentation (rotation, flipping, scaling) for better generalization.
- Removal of low-quality or mislabeled samples.
- Standardization of weather data and imputation of missing values.

3.3 Model Development

Computer Vision Model (by Ananya):

- Designed and implemented a **Convolutional Neural Network (CNN)** for detecting and classifying pest-affected crops.
- Used transfer learning via *ResNet50* and *MobileNetV2*.
- Achieved validation accuracy of approximately **89%**.

Weather-Based Prediction Model:

- Developed a **Random Forest** model trained on historical weather data.
- Predicted conditions favorable for pest or disease development.
- Analyzed feature importance for climatic influence on pest outbreaks.

3.4 Model Integration

- Combined the CNN and Random Forest models into a unified inference pipeline.
- Generated a composite **Disease Risk Index**.

3.5 Visualization and Deployment

- Developed an interactive **Streamlit dashboard**.
- Displayed disease classification results and risk forecasts.

—

4 Dataset Description

Dataset	**Description**	**Source**
PlantVillage	Labeled crop leaf images (healthy/diseased)	Kaggle
OpenWeatherMap	Historical weather parameters	API
Custom Field Data	Manually collected samples (optional)	Local

—

5 System Architecture

The architecture of the Integrated Crop Pest Detection System follows a modular, four-layer pipeline ensuring scalability and real-time usability:

- **Data Acquisition Layer:** Collects leaf images and live meteorological data from APIs.
- **Modeling Layer:** Contains two submodules – CNN for image analysis and Random Forest for environmental prediction.
- **Integration Layer:** Merges outputs to generate unified pest risk predictions.
- **Presentation Layer:** Streamlit-based dashboard enabling real-time visualization and predictions.

—

6 Results and Discussion

The CNN model achieved:

- Training Accuracy: **92%**

- Validation Accuracy: **89%**

The Random Forest model effectively linked pest outbreaks to humidity and temperature variations. Integration improved predictive reliability, enabling early warnings for potential disease outbreaks.

—

7 Limitations and Challenges

- Limited dataset diversity may affect model generalization.
- Environmental data accuracy depends on weather station coverage.
- Real-time deployment requires optimization for latency.

—

8 Future Work

1. Integration with IoT-based agricultural sensors.
2. Expansion to additional crops and disease types.
3. Explainable AI for transparent decision-making.
4. Development of a multilingual mobile app.

—

9 Tech Stack

Category	**Tools and Frameworks**
Programming Language	Python
Computer Vision	TensorFlow, PyTorch, OpenCV
Machine Learning	Scikit-learn, XGBoost
Visualization	Matplotlib, Seaborn, Streamlit
APIs	PlantVillage Dataset, OpenWeatherMap
Deployment	Streamlit, Flask, AWS/GCP

—

10 Team Contributions

Ananya — CNN Model and Image Processing

- Designed, trained, and fine-tuned CNN model.
- Handled preprocessing of pest-infected crop images.
- Documented the implementation in `cnn_model.ipynb`.

Cassiel — Weather-Based Model

- Created Random Forest model using OpenWeatherMap API data.

Aaliyah — Integration

- Integrated CNN and weather models into unified pipeline.

Pranita — Streamlit Dashboard & Preprocessing

- Designed frontend dashboard and assisted in image preprocessing.

—

11 Ethical Considerations

The system supports, not replaces, agricultural experts. Predictions are probabilistic and require expert validation. All datasets are ethically sourced and open access.

—

12 References

1. Mohanty, S.P., Hughes, D.P., & Salathé, M. (2016). Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in Plant Science.*
2. OpenWeatherMap API – `https://openweathermap.org/api`
3. PlantVillage Dataset – `https://www.kaggle.com/datasets/emmarex/plantdisease`
4. Singh, U., et al. (2020). Machine Learning Approaches for Crop Disease Prediction Using Environmental Data. *Journal of Agricultural Informatics.*

—

Acknowledgements

The author expresses sincere gratitude to the faculty of Goa Business School, Goa University, for their valuable guidance and mentorship throughout this project. Appreciation is also extended to open-source developers and researchers for providing datasets, APIs, and frameworks that made this study possible.