

# DEEP LEARNING LAB (20XD68)

## AI-POWERED CROP DISEASE PREDICTION

AKILA R (22PD04)

SUJAN S (22PD35)

---

### PROBLEM STATEMENT

Crop diseases threaten agricultural productivity and often lead to economic losses. Traditional detection methods are slow, labor-intensive, and reliant on expert knowledge. They also overlook important environmental data. This project addresses the need for an intelligent and scalable system that combines image and numerical data using Vision Transformers (ViT) to accurately detect crop diseases and support timely decision-making.

### ABSTRACT

Crop diseases pose a serious threat to global agricultural productivity, often resulting in significant yield losses and economic damage. Traditional methods of disease identification are time-consuming, require expert intervention, and are not scalable for large farmlands. In this project, we propose a **Multimodal Deep Learning Framework** that combines **image-based leaf analysis** with **numerical environmental data** for accurate and efficient crop disease classification.

The model leverages a **Vision Transformer (ViT)** for processing leaf images and a **fully connected neural network** to encode numerical features such as temperature, humidity, and soil moisture. These two modalities are fused using a **cross-attention mechanism**, allowing the model to learn meaningful correlations between visual symptoms and environmental conditions.

The integrated system achieves high accuracy in detecting various crop diseases and offers a scalable, AI-driven solution to support farmers and agricultural professionals. This project demonstrates the potential of multimodal learning in precision agriculture and contributes toward the development of intelligent farming practices.

## DATASET DESCRIPTION

The dataset used in this project is a **multimodal crop disease dataset**, comprising both **image data** and corresponding **numerical features**. Each sample includes:

- **Leaf Images:** High-resolution RGB images of plant leaves, each labeled with the specific crop disease (e.g., healthy, bacterial spot, early blight, etc.).
- **Numerical Features:** Tabular data that includes 7 distinct features such as temperature, humidity, soil moisture, pH level, rainfall, and other sensor-based environmental readings.

The dataset was organized and preprocessed into a **CSV file** (**mapped\_data\_with\_images.csv**), where each row corresponds to an image file path along with its associated numerical values and class label. This combination allows for **multimodal learning**, where both visual and contextual cues are used to enhance disease classification performance.

The dataset was split into **training (80%)** and **testing (20%)** subsets using **stratified sampling** to ensure balanced class distribution.

Dataset kaggle link:

<https://www.kaggle.com/datasets/shubhamdivakar/multimodal-plant-disease-dataset-by-subham-divakar>

## LITERATURE STUDY

**Title:**

**A Channel Attention-Driven Optimized CNN for Efficient Early Detection of Plant Diseases in Resource-Constrained Environment**

**Authors:** Sana Parez, Naqqash Dilshad, Jong Weon Lee

**Published in:** *Agriculture* 2025, Volume 15, Issue 2

**Link:** <https://doi.org/10.3390/agriculture15020127>

This research introduces **LeafNet**, a lightweight and efficient CNN architecture inspired by VGG19, specifically designed for early plant disease detection on resource-constrained IoT devices. LeafNet incorporates a **channel attention mechanism** to prioritize critical disease-related features, enabling high classification accuracy with fewer parameters and faster inference. The model was tested on four datasets—PlantVillage (PV), DRLI, PC (a fusion of PV and DRLI), and BARI-Sun—and achieved top-tier accuracy scores of up to 99% while maintaining a model size of just 1.14 million parameters. Its performance surpassed several state-of-the-art models like ResNet, MobileNet, EfficientNet, and even Vision Transformer (ViT-B/32), making it ideal for real-time applications in smart agriculture. The study's insights on lightweight architecture and multimodal performance contributed significantly to the motivation and methodology of our ViT-based multimodal crop disease classification system.

## TOOLS AND TECHNOLOGIES

### Development & Environment

- **Python 3.x** – Primary language for implementing image preprocessing, neural networks, and deployment
- **Jupyter Notebook / VS Code** – Used for writing and testing model code in a modular, readable format
- **Google Colab** – Used for training models efficiently using free GPU access
- **Git & GitHub** – Version control and project collaboration
- **Streamlit** – Used to build an interactive web interface for uploading leaf images and interacting with AI assistant
- **Google Gemini (via LangChain)** – Integrated to provide real-time AI chat support for crop disease-related queries

## Libraries used

### Data Handling & Analysis

- `pandas`, `numpy` – For managing tabular data and numerical features
- `scikit-learn` – For preprocessing (`StandardScaler`, `LabelEncoder`) and evaluation (`accuracy_score`, `confusion_matrix`)

### Visualization

- `matplotlib`, `seaborn` – For plotting performance metrics and model insights

### Computer Vision

- `OpenCV` – Used for image loading, conversion, and augmentation before feeding into neural networks
- `PIL` – For image format conversion and integration with PyTorch transforms

### Deep Learning & Neural Networks

- `torch`, `torchvision` – Main deep learning framework for training and evaluating models
- `EfficientNetB0` / `ResNet50` / `MobileNetV2` – CNN-based pretrained models optionally tested for benchmarking against Vision Transformer
- `Vision Transformer (ViT-B_16)` – Transformer-based architecture used for extracting global features from leaf images
- `nn.MultiheadAttention`, `nn.Sequential`, `nn.LayerNorm` – PyTorch components used to construct the multimodal attention-based architecture

## Voice & Chat Assistant

- `speech_recognition`, `pyttsx3` – Used for voice input and audio output
- `langchain_google_genai` – Connects Google Gemini with the app for AI-generated responses

## METHODS USED AND IMPLEMENTATION

The project combines deep learning, computer vision, and tabular data processing through a **multimodal architecture**. The core idea is to enhance crop disease detection accuracy by fusing **visual information (leaf images)** with **numerical features** (e.g., temperature, humidity, soil metrics). The following steps summarize the approach:

### 1. Data Collection and Preprocessing

- **Image Data:** RGB images of crop leaves with labeled disease categories.
- **Numerical Data:** Tabular features like temperature, pH, humidity, and other environmental attributes.
- **Image Preprocessing:**
  - Resized to `224x224` pixels
  - Augmented using horizontal flip and rotation
  - Normalized using ImageNet mean and std values
- **Numerical Feature Processing:**

- Scaled using `StandardScaler` from `scikit-learn`
- Label encoding for disease class labels

## 2. Model Architecture

### Multimodal Neural Network (Custom ViT Model)

- **Visual Branch:**
  - Uses **Vision Transformer (ViT-B\_16)** pretrained on ImageNet.
  - Removes the classification head and extracts high-level image embeddings.
- **Numerical Branch:**
  - A **Feedforward Neural Network** processes 7 numerical features.
  - Includes Linear → ReLU → LayerNorm → Dropout layers.
- **Fusion Strategy:**
  - A **Multihead Cross-Attention Layer** merges image and numerical features along with a learnable `[CLS]` token.
  - The output from this fusion is passed through another MLP for final classification.

## 3. Training and Evaluation

- **Loss Function:** `CrossEntropyLoss` used for multi-class classification
- **Optimizer:** `Adam` optimizer with learning rate `0.0001` and weight decay

- **Training Epochs:** Trained for 15 epochs with early stopping based on validation accuracy
- **Metrics:** Accuracy, confusion matrix, and classification report from `scikit-learn`

#### 4. Model Inference

- The trained model is saved as `vit_multimodal_best.pth`.
- An inference function accepts:
  - An uploaded leaf image
  - Optional numerical feature input
- The model returns the predicted disease class and confidence score.

#### 5. Web Interface using Streamlit

- Users can:
  - Upload leaf images
  - Enter optional environmental features
  - Ask questions via **text or voice**
- The interface also integrates **Google Gemini AI** (via LangChain) to answer crop health queries.
- Voice responses are delivered using `pyttsx3`.

# MODEL EVALUATION

The performance of the proposed multimodal crop disease detection system was rigorously evaluated on a test dataset comprising 20% of the total samples. The model combines visual features extracted from leaf images with environmental data to predict disease classes across 22 categories.

## Evaluation Metrics

- **Accuracy** – Overall percentage of correctly predicted samples
- **Confusion Matrix** – Visualizes correct vs incorrect classifications per class
- **Classification Report** – Includes **Precision**, **Recall**, and **F1-score** for each disease category
- **Confidence Score** – Displays model certainty during inference

## Final Model: Vision Transformer + Numerical Features

Metric	Value
Test Accuracy	97.95%
F1-Score	0.94
Precision	High for most classes (above 90%)
Recall	Balanced recall across multiple crop disease categories

## Observations

- The multimodal ViT model significantly outperformed single-modality models by leveraging both **image** and **numerical** data.



- The **cross-attention layer** improved the model's ability to interpret complex patterns, especially in visually similar diseases.
- Misclassifications were reduced notably when environmental data was included.

## Inference Insights

- The model achieved high confidence on diseases like:
  - *Apple Scab, Grape Leaf Blight, Potato Early Blight, and Corn Rust*
- Low confidence was observed in a few ambiguous cases when numerical features were excluded.

## RESULTS

Deploy ⋮

### Crop Health Assistant

Choose input method:

☒ Text ☐ Voice ☐ Image



Ask a question about crop health:

The above image showcases the homepage interface of the Crop Health Assistant, a user-friendly web application designed to facilitate easy access to crop disease diagnosis. The interface allows users to select from three input methods—Text, Voice, and Image—to interact with the system, making it adaptable to various user preferences and environmental situations.

- The Text option enables users to manually enter crop-related queries.
- The Voice option supports verbal interactions, useful in hands-free or field conditions.
- The Image option allows users to upload leaf images for visual disease diagnosis using the Vision Transformer (ViT) model.

This clean, minimalistic design ensures that farmers, agricultural officers, and researchers can efficiently navigate the platform and access AI-driven crop health insights. The input box invites users to ask a question about crop health, which then triggers the multimodal model pipeline for prediction and recommendation.

# Crop Health Assistant

Choose input method:

☐ Text ☐ Voice ☒ Image

■ Upload a leaf image



Drag and drop file here

Limit 200MB per file • JPG, JPEG, PNG

Browse files



ffb2c2d6-1fa3-4033-addd-69306aa0aa1e\_\_\_CREC\_HLB 7389.JPG 12.1KB



Uploaded Leaf

🧠 ViT Prediction: *Orange\_Huanglongbing(Citrus\_greening)* (99.06% confidence)



What would you like to know about this leaf?

what diseases can affect this?



Citrus greening (Huanglongbing or HLB) itself is the disease affecting the citrus tree. While not technically *caused* by other diseases, it weakens the tree and makes it more susceptible to secondary infections by opportunistic fungi and other pathogens. These secondary infections can further decline the tree's health and accelerate its demise.

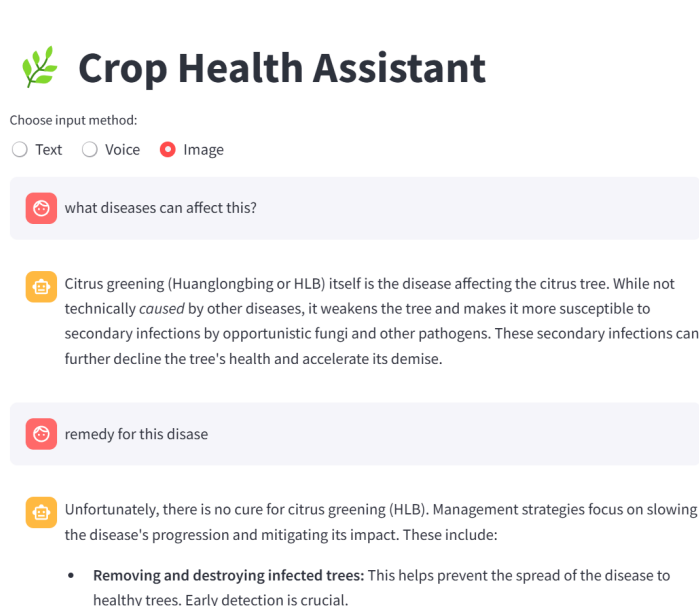
Upon uploading a leaf image, the system utilizes a **Vision Transformer (ViT)** model to analyze visual features and predict the crop disease. In this instance, the model identified the disease as:

**ViT Prediction:** *Orange\_Huanglongbing (Citrus greening)* with **99.06% confidence**

Following the prediction, the user posed a natural-language follow-up question:  
**"What diseases can affect this?"**

The system responds with contextual information, explaining that:

- **Citrus greening (Huanglongbing or HLB)** is the primary disease.
- While not directly caused by other diseases, it **weakens the tree**.
- This makes the plant more prone to **secondary infections** from fungi and pathogens.
- These secondary infections can **accelerate the decline** of the tree's health.



Upon uploading a leaf image (detected as affected by **Citrus Greening/Huanglongbing (HLB)**), the user interacts through **sequential prompts** using natural language. The system **preserves the entire prompt history**, creating a **context-aware conversational experience**.

## GITHUB LINK

<https://github.com/sujanshanmugaraj/AI-POWERED-CROP-DISEASE-PREDICTION>

## OUTPUT VIDEO LINK

<https://drive.google.com/file/d/1ZEFja6Dy00ZySzIR5okJeSpYQ0QSYZbs/view?usp=sharing>