🌿

Smart Tomato Disease Detection System using Deep Learning and Environmental Context

🔍 1. Background from Literature Review

Tomato crops are susceptible to various **leaf diseases** such as Early Blight, Leaf Mold, and Yellow Leaf Curl Virus, leading to massive yield loss and economic damage. Conventional detection methods relying on human observation are time-consuming and error-prone.

Recent literature has explored:

- 🧠 Deep Learning (DL) models like **CNN, InceptionNet, U-Net** for high-accuracy classification.

- 📊 Classical ML approaches (Random Forest, SVM, KNN) with texture features like **GLCM (Gray-Level Co-occurrence Matrix)** and **Color Moments**.
- 🖼️ Use of segmentation for localizing leaf lesions before classification.
- 📈 Dataset balancing via augmentation to prevent overfitting.

However, these approaches show **limitations**:

- Lack of **contextual factors** (e.g., environment).
- No disease **co-occurrence analysis**.
- Poor **generalization** to new conditions.
- Absence of **explainable predictions**.
- Focused only on **binary or single-label** classification.

💡 2. Our Novel Contribution

We aim to develop a **comprehensive, real-world-ready disease detection system** that builds on existing methods and fills critical research gaps.

✅ Our Innovations:

🌡️ A. Environmental Context-Aware Prediction

- **Why?** Disease outbreaks depend heavily on weather — high humidity triggers Leaf Mold, while dry heat increases Spider Mite risk.
- **How?** Integrate **real-time temperature and humidity** via APIs (e.g., OpenWeatherMap) or sensors into the model.
- **Impact:** Model adapts based on current weather, increasing prediction relevance and accuracy in-field.

📊 B. Multi-label Classification with Real-World Dataset

- **Why?** Leaves often show signs of **multiple diseases** simultaneously.
- **How?** Our custom dataset supports **multi-label output**, where each image can have multiple active disease labels.

Example row from CSV:

```
Copy code
filename, Early Blight, Healthy, Leaf Mold, Mosaic Virus, ...
Tomato_Leaf_12.jpg, 1, 0, 1, 0, ...
```

- **Impact:** Supports **realistic diagnosis** without forcing one-class-per-image assumption.

🔥 C. Grad-CAM Heatmap for Multi-label Explainability

- **Why?** Trust and validation are crucial for farmers or agronomists.
- **How?** Use **Grad-CAM** to generate a heatmap that visually shows **where the CNN is focusing** for each predicted disease label.
- **Impact:** Builds **transparency** and increases system credibility for real-world use.

🧬 D. Disease Co-Occurrence Analysis

- **Why?** Some diseases frequently occur together due to shared causes (e.g., Mosaic Virus & Yellow Curl Virus in viral seasons).
- **How?** Compute and visualize a **co-occurrence matrix** from the label CSV and generate a **heatmap**.
- **Impact:** Enables **early multi-disease diagnosis**, better pesticide planning, and targeted intervention.

📉 E. Stage-wise Severity Prediction (from multi-label probabilities)

- **Why?** Farmers need not only disease type but also **how advanced it is**.
- **How?** Use **multi-label probabilities** (e.g., `Mosaic Virus = 0.2 vs 0.9`) to infer **mild, moderate, severe** stages.
- **Impact:** Gives actionable insights for prioritizing treatment.

🧠 3. Technical Summary of the Model

Component	Description
CNN Backbone	Custom 3–4 layer CNN (Conv → BN → Pool), optionally InceptionV3
Input 1	Image (resized, Z-score normalized)
Input 2	Real-time temperature and humidity (via API)
Output	Multi-label sigmoid output (9 diseases)
Loss	`binary_crossentropy`
Explainability	Grad-CAM heatmaps per label
Extras	Data Augmentation (rotation, scaling, flipping), Disease Co-occurrence Matrix

---

📦 4. Dataset Structure

The dataset is divided into `train/`, `valid/`, and `test/` folders with:

- Image files (`.jpg`)
- A `_classes.csv` in each folder with columns:

```
filename, Early Blight, Healthy, Late Blight, Leaf Miner, Leaf Mold, Mosaic Viru
s, Septoria, Spider Mites, Yellow Leaf Curl Virus
Tomato_Leaf_001.jpg, 1, 0, 0, 1, 0, 0, 0, 0, 0
```

We map this into:

```
python
Copy code
df['labels'] = df[DISEASE_COLUMNS].values.tolist()
```

🎯 5. Project Goal

> To build a deep learning-based multi-label disease diagnosis system for tomato plants that:

- Uses **environmental conditions**
- Generates **explainable predictions**
- Detects **co-occurring diseases**
- Predicts **severity**
- Is robust enough for **real-field deployment**

📚 6. References (IEEE Format)

[1] R. Louis, “Deep learning-based classification, detection, and segmentation of tomato leaf diseases: A state-of-the-art review,” *Elsevier Computers and Electronics in Agriculture*, 2023.

[2] M. Sharma and A. Rani, “An experimental study of tomato viral leaf diseases detection using machine learning classification techniques,” *Procedia Computer Science*, vol. 173, pp. 211–220, 2020.

[3] A. Kumar et al., “Classification of Tomato Plant Diseases Through Leaf Using Gray-Level Co-occurrence Matrix and Color Moment with Convolutional Neural Network Methods,” *Springer Lecture Notes in Electrical Engineering*, vol. 937, 2023.

[4] M. S. Chowdhury et al., “Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease,” *Springer Soft Computing*, vol. 26, pp. 19831–19848, 2022.

[5] Y. Sevilla et al., “InceptionNet-based detection of leaf diseases in tomato using compound scaling,” *IEEE Access*, vol. 11, pp. 8762–8774, 2023.

[6] D. Gupta et al., “Plant disease detection and classification techniques: a comparative study of the performances,” *Materials Today: Proceedings*, vol. 46, pp. 389–396, 2021.

[7] OpenWeatherMap API, “Current Weather Data,” [Online]. Available: https://openweathermap.org/api