Plant AI Doctor

INTERIM PROJECT PRESENTATION

Motivation

Farmers struggle to distinguish chemical damage from biological diseases in tomato leaves. Visual similarity leads to wrong treatment and yield loss.

The Pivot

Shift from multiclass to binary classification: chemical vs. disease. This focus is clearer and technically more feasible for data limitations.

Novelty

One of the first datasets for chemical vs. biological distinction, leveraging Generative AI for advanced synthetic data generation.

Scientific Literature Review

Paper (Year)	Task	Methods	Data	Results	Relation to Project
Tomato Leaf Disease Segmentation (2022)	Disease classification + segmentation	U-Net segmentation + CNN classifier	18K tomato leaf images	High accuracy: segmentation improves performance	Works only on diseases, not chemical vs disease (our novelty)
Efficient Tomato Disease Detection (2025)	Disease detection (efficient model)	Custom CNN model (E-TomatoDet)	Tomato leaf dataset	High detection performance (mAP ≈ 97.2%)	Shows SOTA disease detection, but no chemical damage classification
Review of Tomato Leaf DL Methods (2025)	Review of DL methods for tomato leaf diseases	Survey of CNNs, segmentation, preprocessing	Multiple tomato datasets (e.g., PlantVillage)	Identifies major limitations in datasets	Highlights gap: no dataset for chemical vs disease

Dataset Analysis & Visual Samples

Dataset Components

- **Real Images:** 157 total images (RGB).
- **Chemical damage:** 67 images.
- **Biological disease:** 90 images.
- **Labels:** assigned at the image level based on damage type

EDA & Strategy

- **Class Imbalance:** Slight (chemical < biological).
- **Split:** Train / Val / Test (no overlap).
- **Format:** RGB images, varying resolutions and backgrounds.

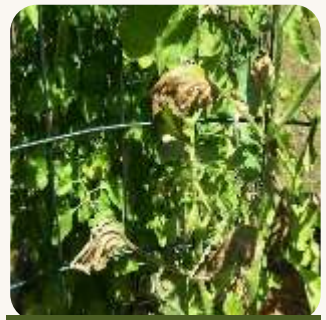
Future Strategy:

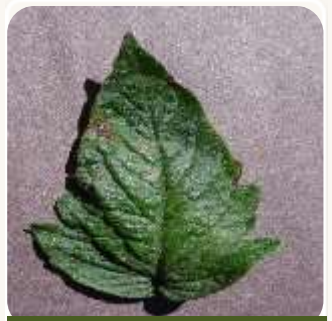
Synthetic data generation via Stable Diffusion and ControlNet to address data scarcity and balance classes.


Chemical Damage


Chemical Damage

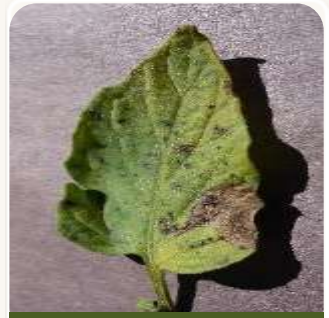

Chemical Damage


Chemical Damage


Biological Disease


Biological Disease


Biological Disease


Biological Disease


Synthetic Chemical Damage


Synthetic Chemical Damage


Synthetic Chemical Damage


Synthetic Chemical Damage

Baseline Solution and Results

Baseline Approach: Pretrained ResNet18 CNN adapted for binary classification. Uses minimal modifications. No synthetic data used at this stage

Dataset	Accuracy
Validation (real)	100%
Test (real)	100%

No classification errors were observed on the validation and test sets.

Performance Assessment

The perfect accuracy is likely due to:

- Small size of current real dataset.
- Powerful pretrained ResNet model.
- Strong visual cues in the dataset.

Conclusion: Results suggest potential overfitting. Accuracy alone is insufficient to assess model robustness.

Execution Roadmap

Step	Scope / Description	Models & Techniques	Expected Outcome
1	Synthetic chemical data generation	Stable Diffusion	Balanced chemical image dataset
2	Realistic damage pattern refinement	Inpainting, ControlNet	Localized, realistic damage areas patterns
3	Dataset Expansion	Merge Real + Synthetic (train only)	Robust training set for higher generalization
4	Updated Model Retraining	ResNet18 (Fine-tuning)	Enhanced classifier performance
5	Robustness Validation	Crop / Illumination changes	Verified stability in varying conditions
6	Final Submission	Results summary and presentation	**Completion of Final Presentation**