UF | UNIVERSITY of FLORIDA

Kuan-Chen,Chen
Ku.chen@ufl.edu

LeafInspire

A Dual-Application Framework for Tomato Disease Detection and Biomimetic Surface Analysis Using Deep Learning

Introduction

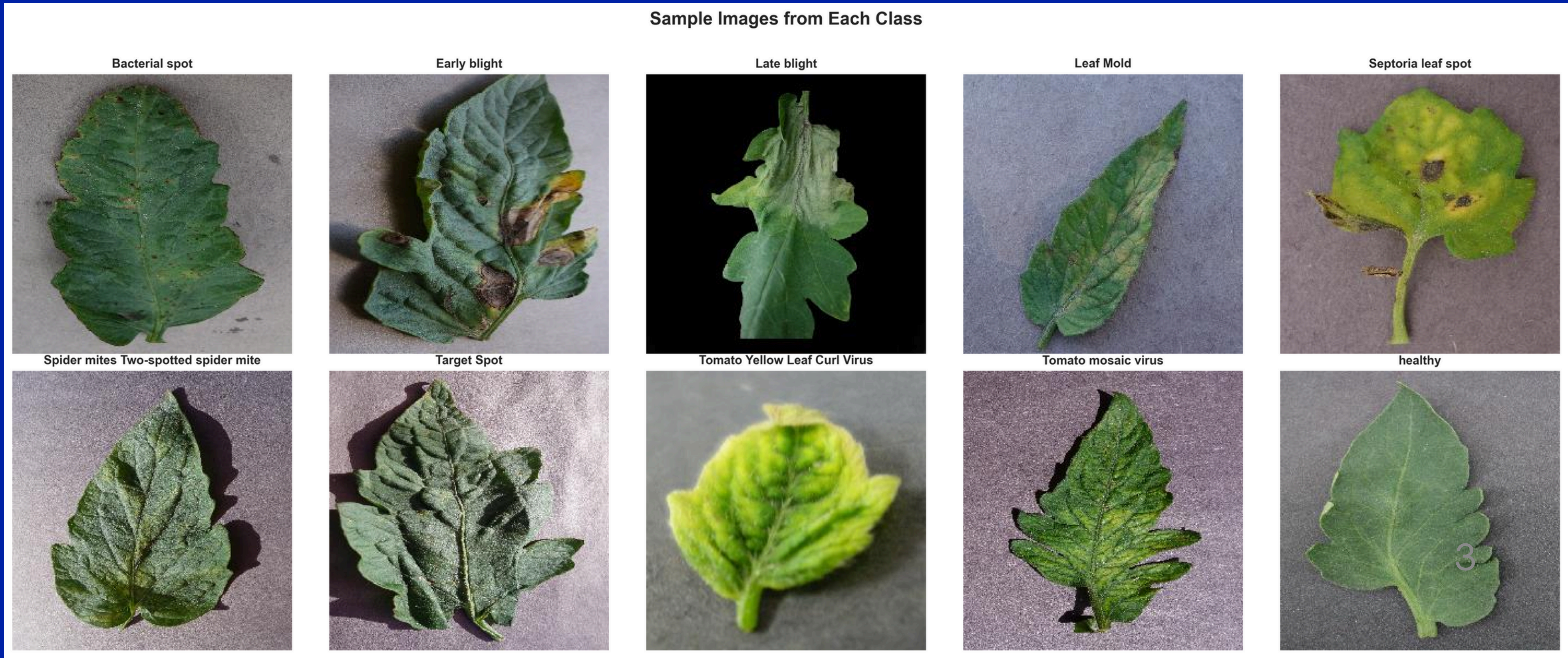
- Plant diseases significantly reduce crop yield; farmers need fast and reliable detection tools
- Natural organisms (like leaf textures) often inspire biomimetic material design
- Most research uses image datasets for only a single purpose
- Question: Can one dataset serve multiple domains at once?
- This project demonstrates that plant imagery can support both agricultural diagnosis and engineering texture modeling

Dataset Overview

Tomato Leaf Dataset (PlantVillage) Kaggle

- 10 classes: 9diseases + healthy
- ~12,000 labeled high-resolution images
- Images captured under controlled lighting and uniform background
- Covers a wide variety of biological symptoms:
- Clear visual diversity allows robust feature learning across multiple tasks
- Curling, blistering, and mosaic patterns
- Color chlorosis
- Necrotic lesions
- Mold-like textures

1. Bacterial spot
2. Early blight
3. Late blight
4. Leaf Mold
5. Septoria leaf spot
6. Spider mites
7. Target Spot
8. Yellow Leaf Curl Virus
9. Tomato mosaic virus
10. Healthy



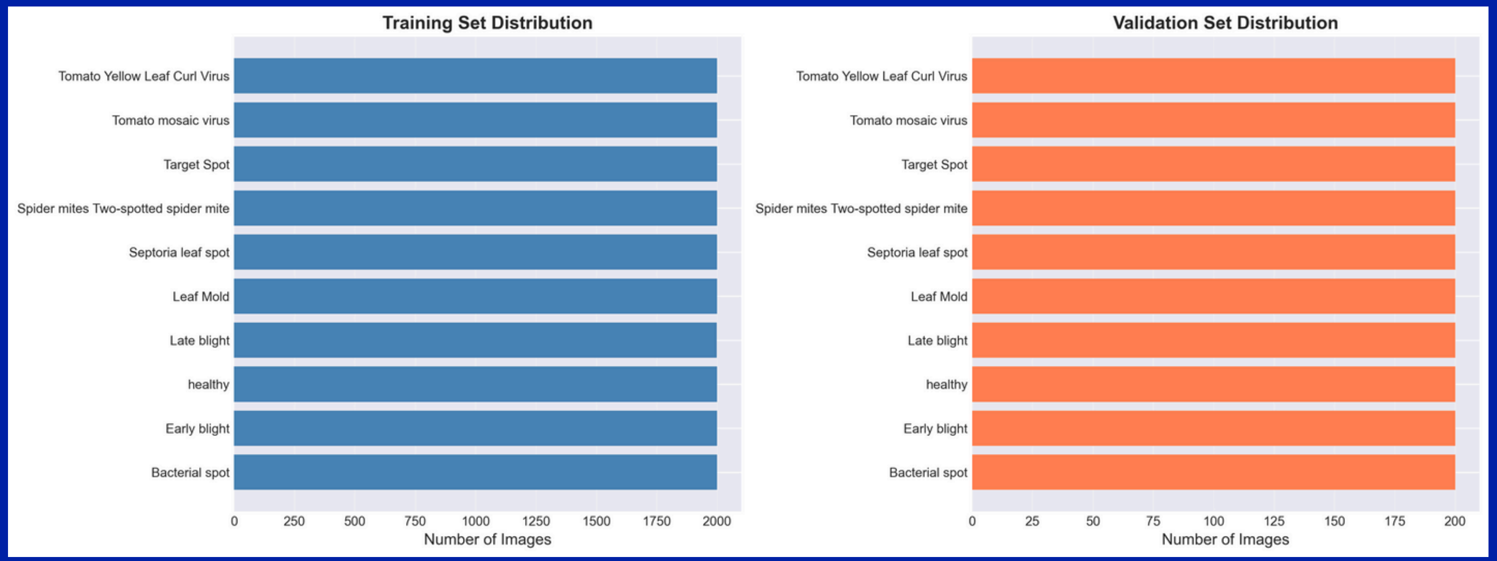
Dataset Statistics & Properties

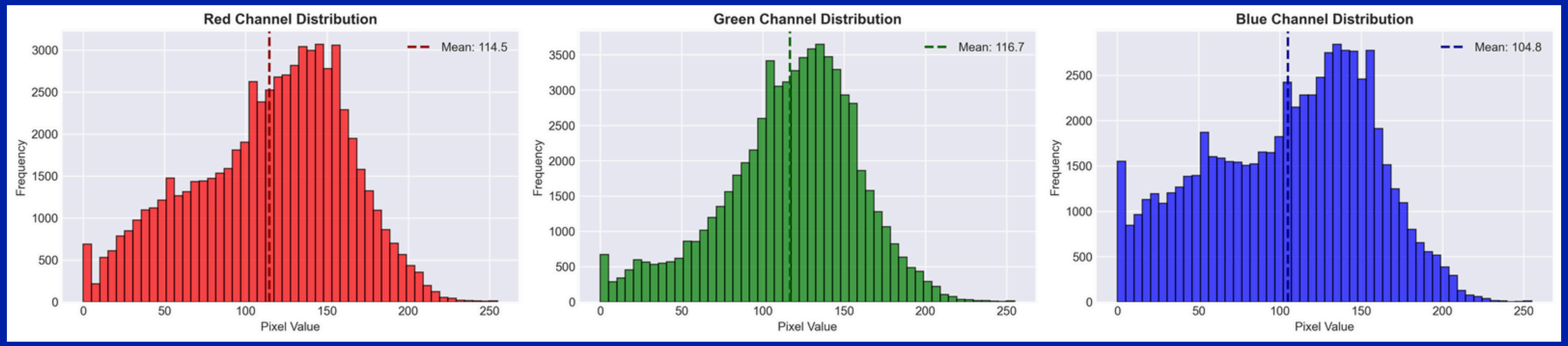
Balanced Class Distribution

- Training and validation sets maintain equal distribution
- Prevents class dominance and reduces label imbalance bias
- Ensures stable optimization during CNN training

Color Distribution Analysis

- RGB histograms show consistent illumination
- Helps confirm preprocessing pipeline reliability
- Green channel slightly dominant → aligned with leaf pigmentation





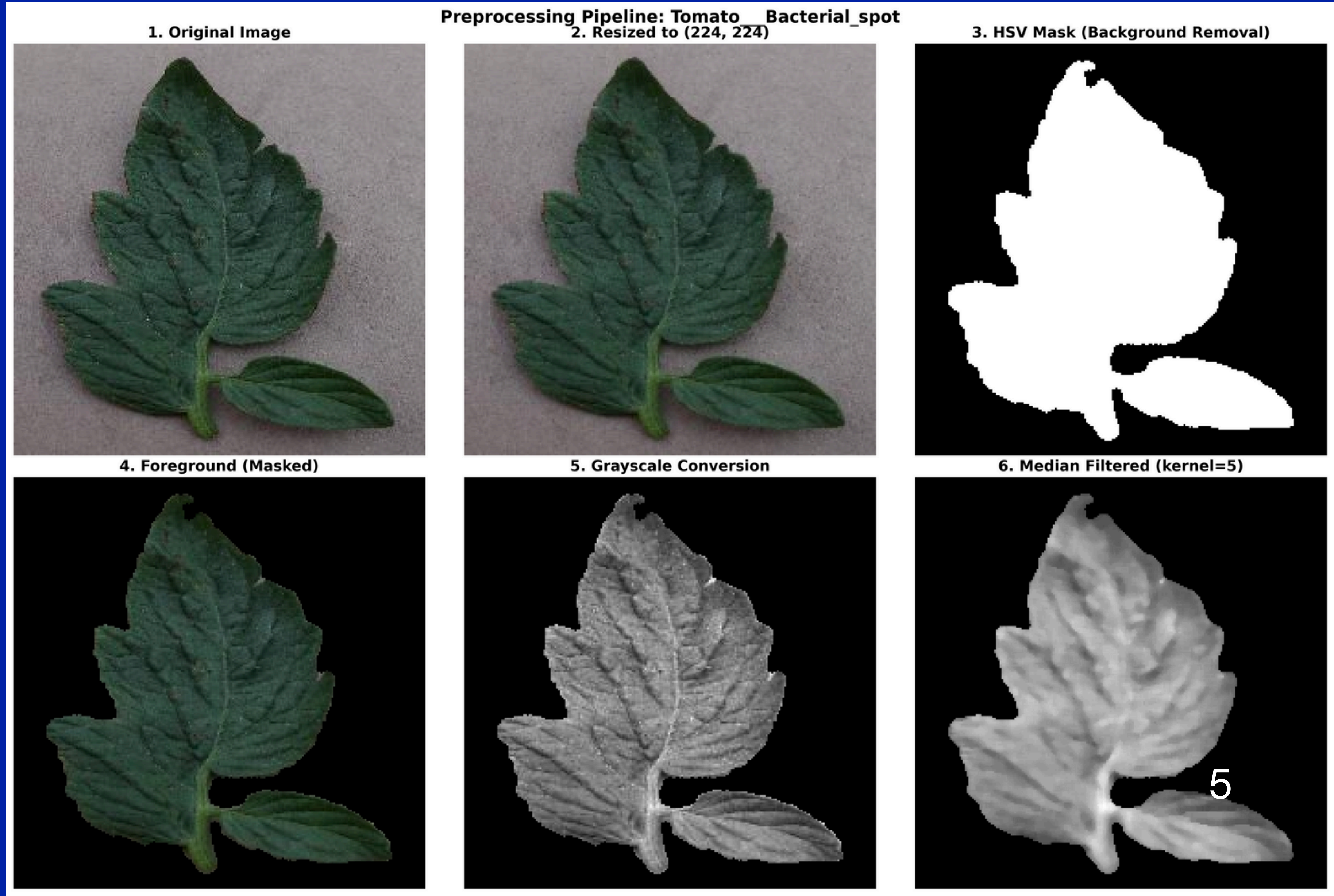
Preprocessing Pipeline

Six-Step Image Processing Workflow

- Image resizing for uniformity (224×224)
- Median filtering to remove noise while maintaining structural edges
- Conversion to HSV for more stable color-based segmentation
- Background removal using saturation thresholding
- Extraction of leaf foreground mask
- Grayscale transformation for texture extraction

Purpose

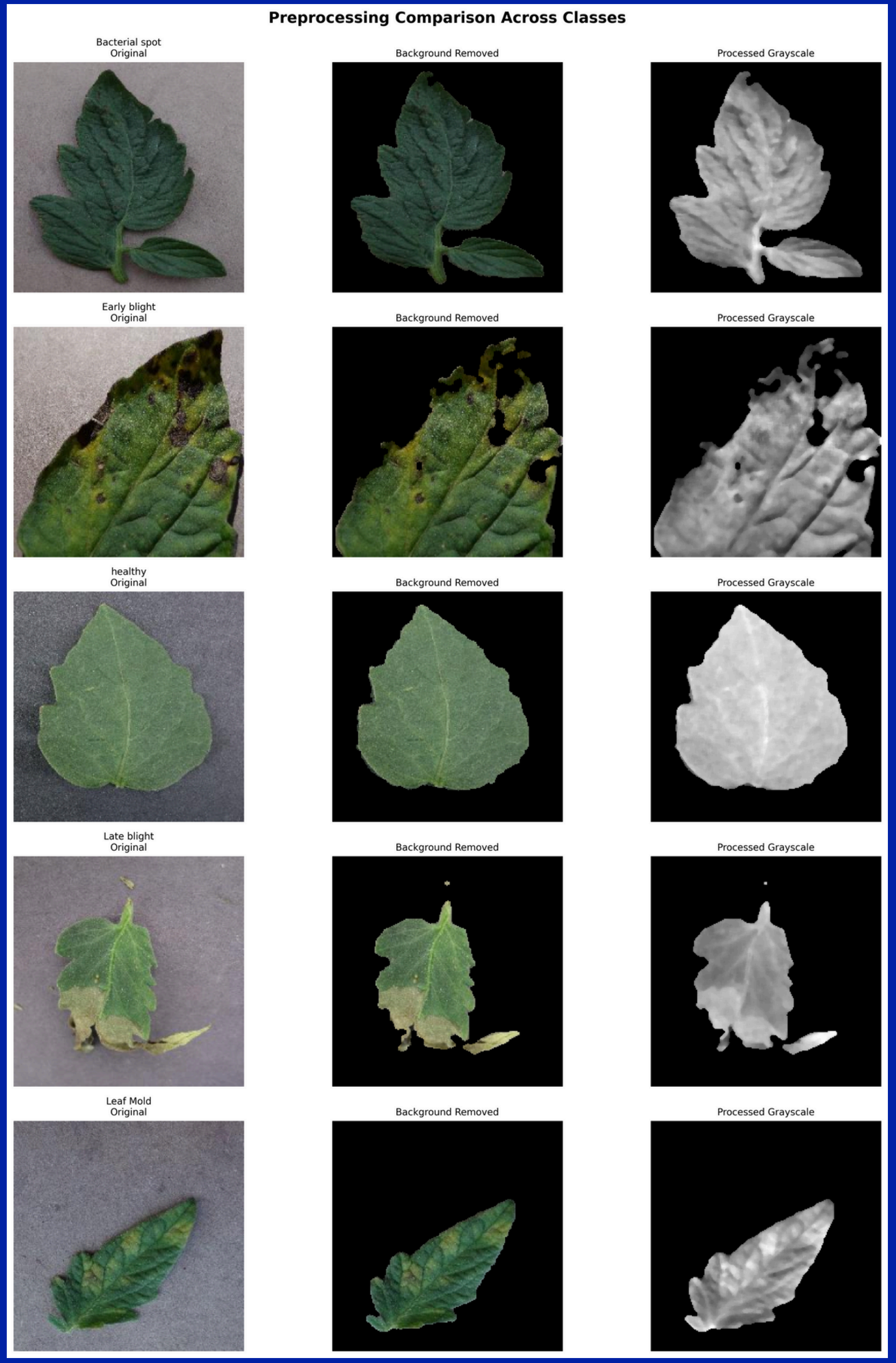
- Standardize input for both CNN and texture analysis
- Remove irrelevant background textures
- Improve GLCM and fractal feature accuracy



Preprocessing Comparison Across Classes

Cross-Class Robustness

- Segmentation works even on difficult cases (light-colored backgrounds, damaged leaf edges)
- Mold-like diseases maintain shape integrity after filtering
- Background removal isolates disease-related texture patterns
- Essential for stable feature computation across all disease types



Feature Extraction: Multi-Source Features

Total Feature Vector 1351 Dimensions

1. Classical Texture Features

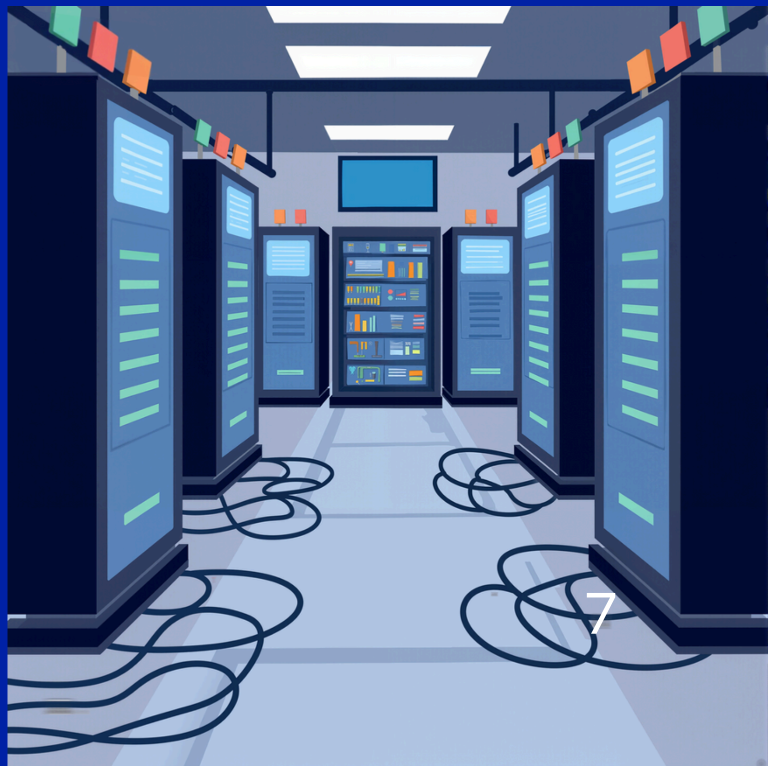
GLCM (60D):
Captures spatial co-occurrence of pixel intensities
Sensitive to roughness, lesion granularity, and patch irregularity

Fractal Dimension (1D):
Measures multi-scale complexity
Higher values correspond to irregular edge structures

Vein Features (10D):
Vein density, geometric compactness, Hu moments
Useful for structural abnormalities

2. Deep Learning Features

MobileNetV2 (1280D):
Extracts semantic + texture features
Pretrained on ImageNet → strong general feature base
Lightweight and efficient



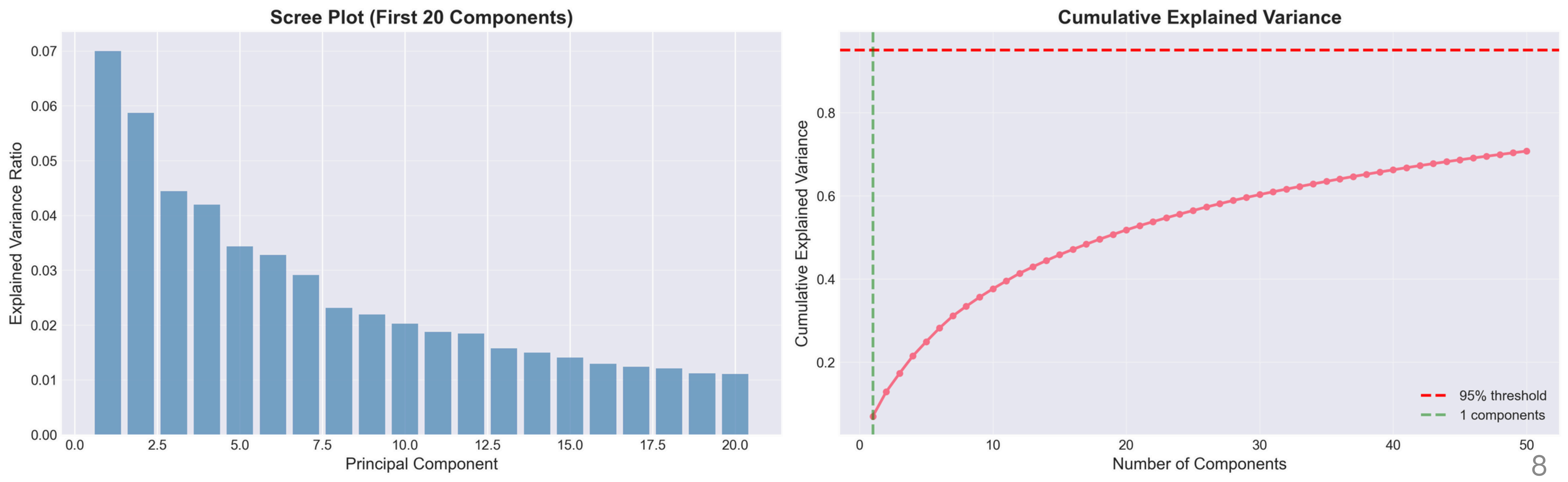
PCA Dimensionality Reduction

Dimensionality Optimization

- StandardScaler applied prior to PCA
- PCA reduces 1351D → 50 principal components
- Retains ~95% of cumulative variance
- Eliminates redundant features
- Enhances clustering stability

Interpretation

- First few components capture dominant texture patterns
- PCA projection reveals natural groupings among leaf textures



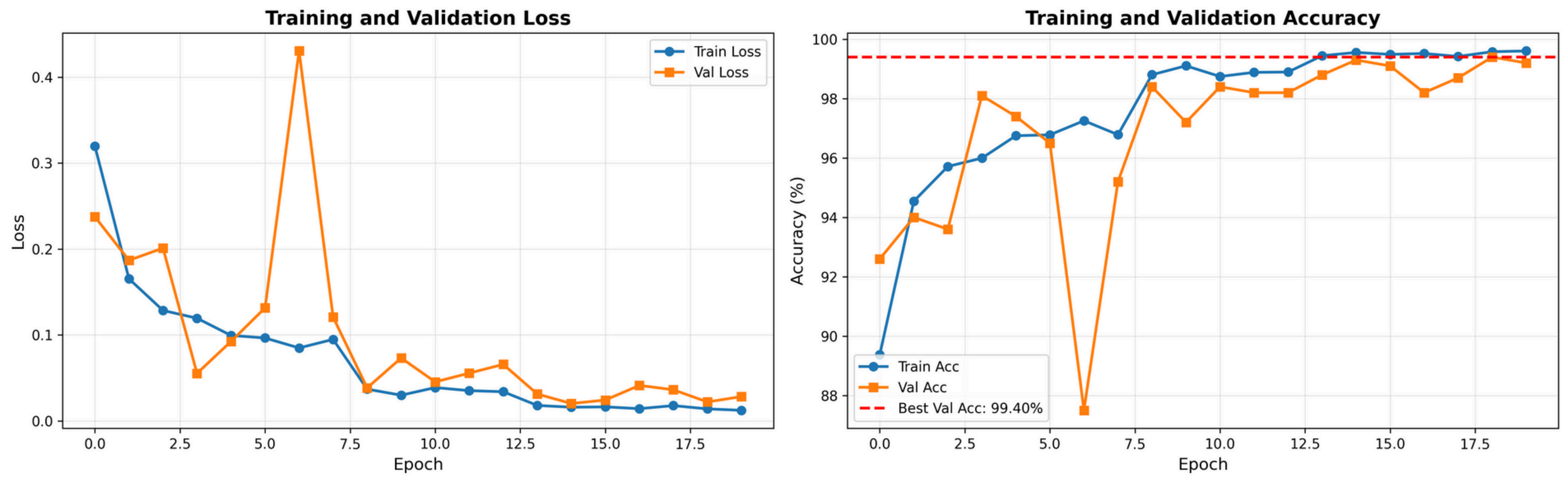
Training Progress (Classification)

MobileNetV2 Training Dynamics

- Rapid decrease in training loss within first 5 epochs
- Validation loss stable → minimal overfitting
- Accuracy quickly converges above 95%

Observations

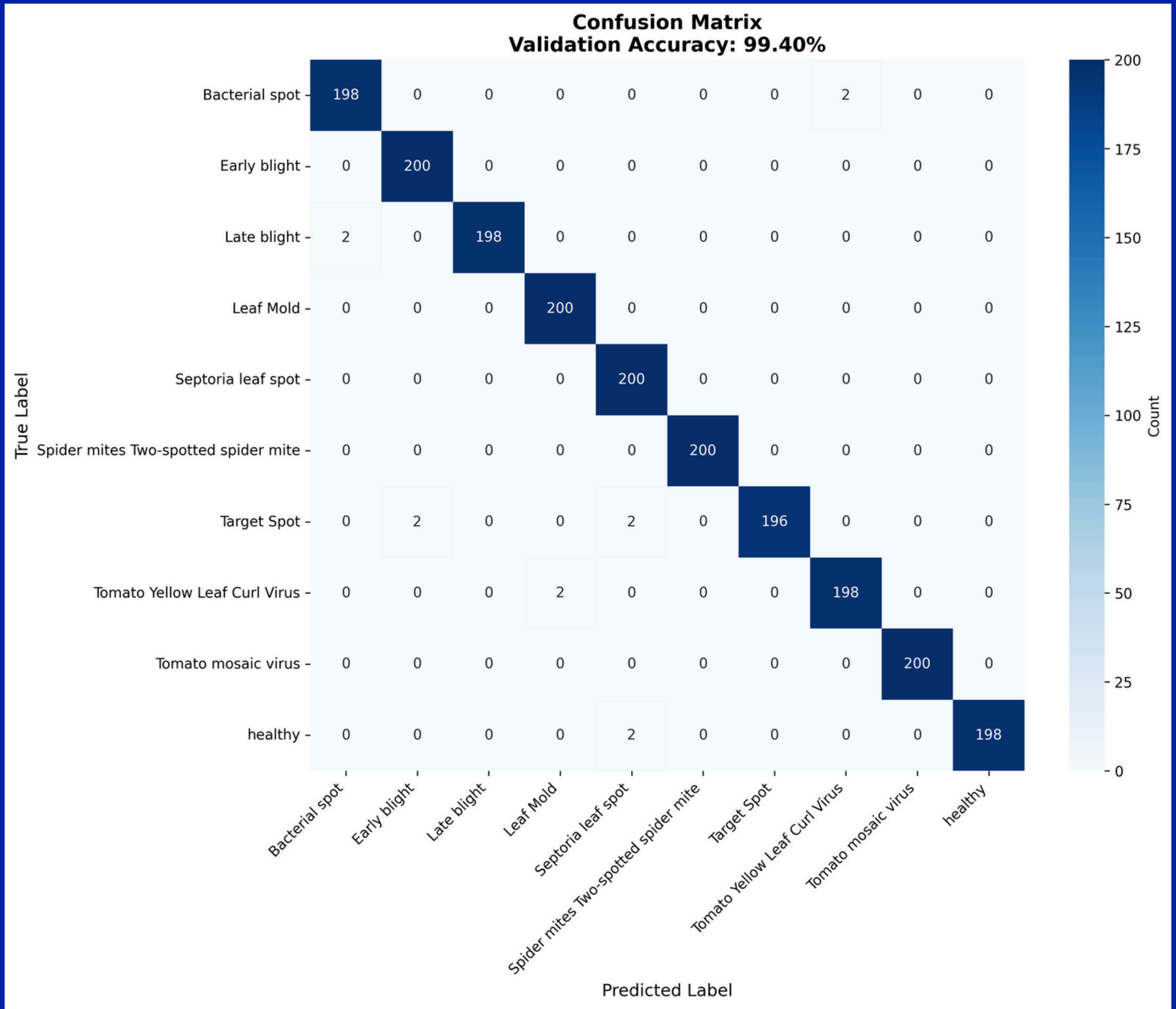
- Sharp accuracy dip at epoch 6 suggests temporary over-regularization
- Model recovers and stabilizes → final accuracy near 99%
- Demonstrates strong generalization ability



Final Classification Performance

Confusion Matrix Analysis

- Validation Accuracy: 99.4%
- Nearly perfect separation across all 11 categories
- Misclassifications extremely rare (<0.5%)
- Diseases with similar early-stage symptoms (e.g., Target Spot vs Late blight) still classified correctly
- Confirms CNN’s ability to extract discriminative features



Selecting k for Texture Clustering

Elbow Method

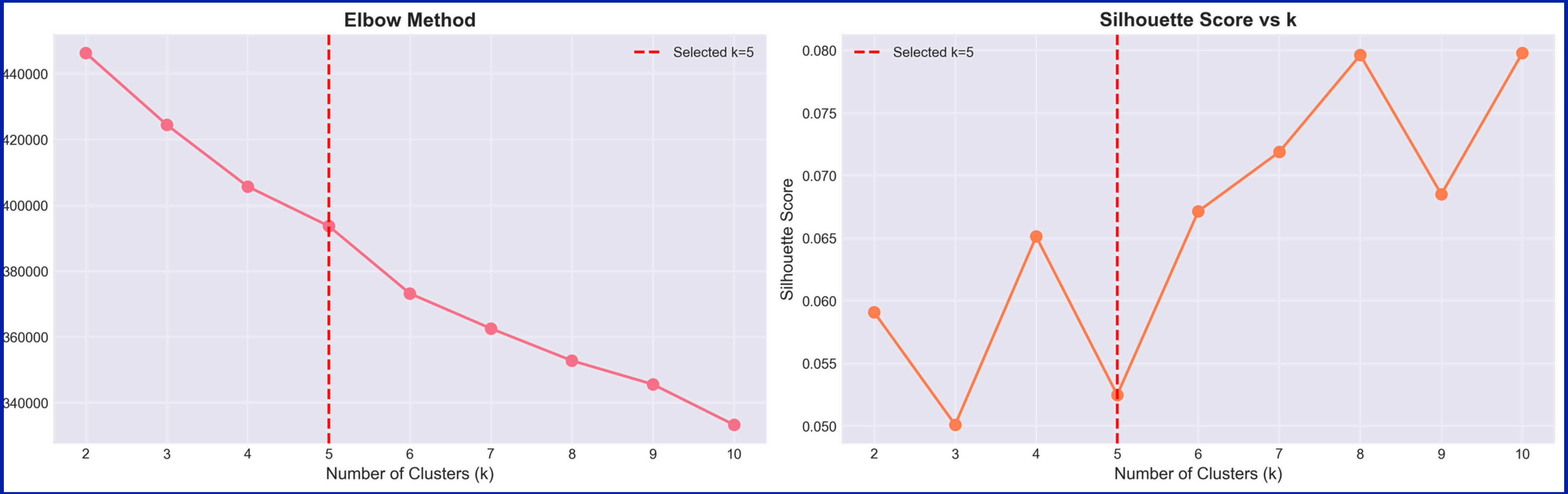
Elbow observed around k=5
Inertia drops quickly between k=2–5, then plateaus

Silhouette Score

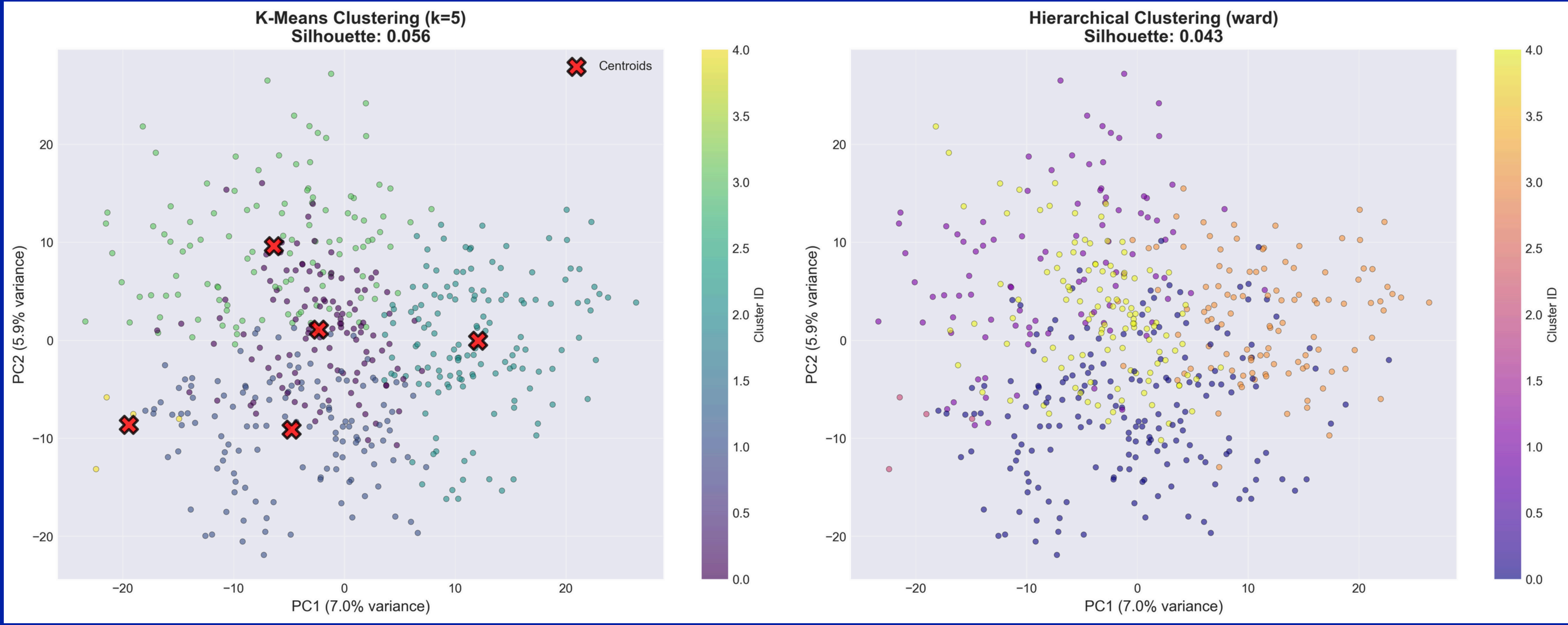
Highest local score near k=5
Indicates moderate cluster separation in texture space

Conclusion

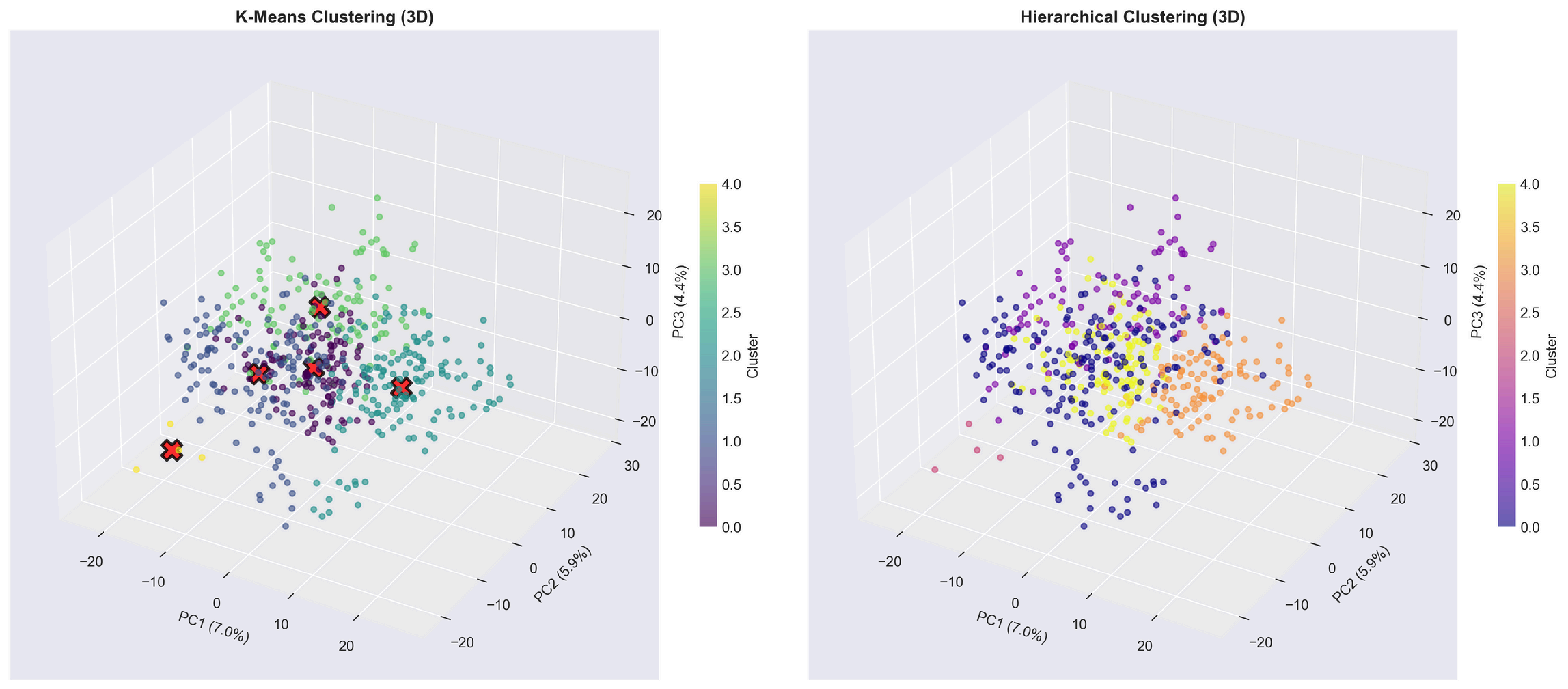
k = 5 chosen for best balance between compactness and interpretability



Texture Space Visualization (2D)



Texture Space Visualization (3D)



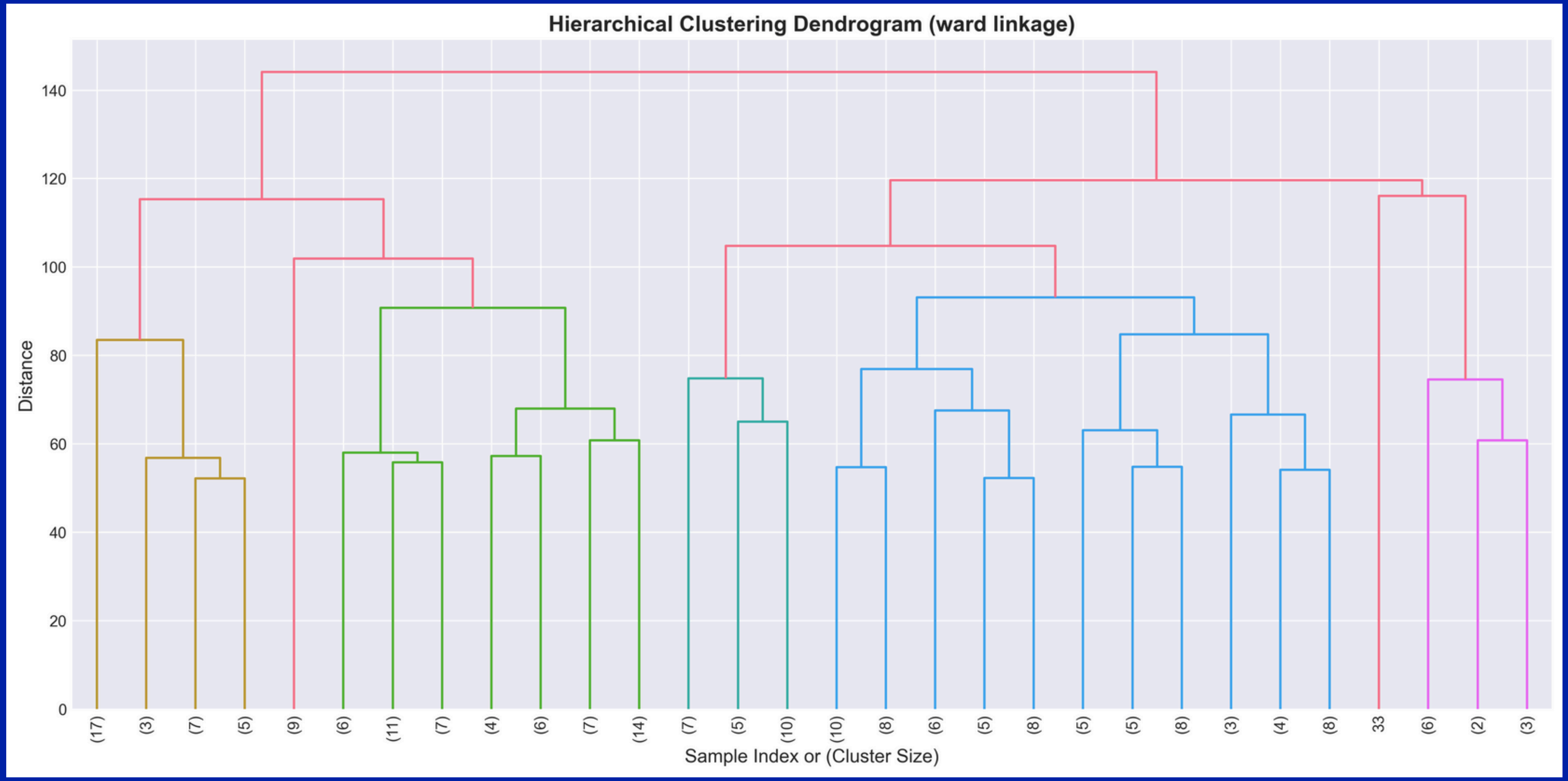
Disease–Cluster Distribution

Cluster vs. Disease Heatmap

Diseases strongly aligned with specific texture groups

Example:

Late blight → Cluster 2 (high complexity)
Leaf Mold → Cluster 3 (low vein density)
Healthy → Cluster 1 (smooth, consistent texture)

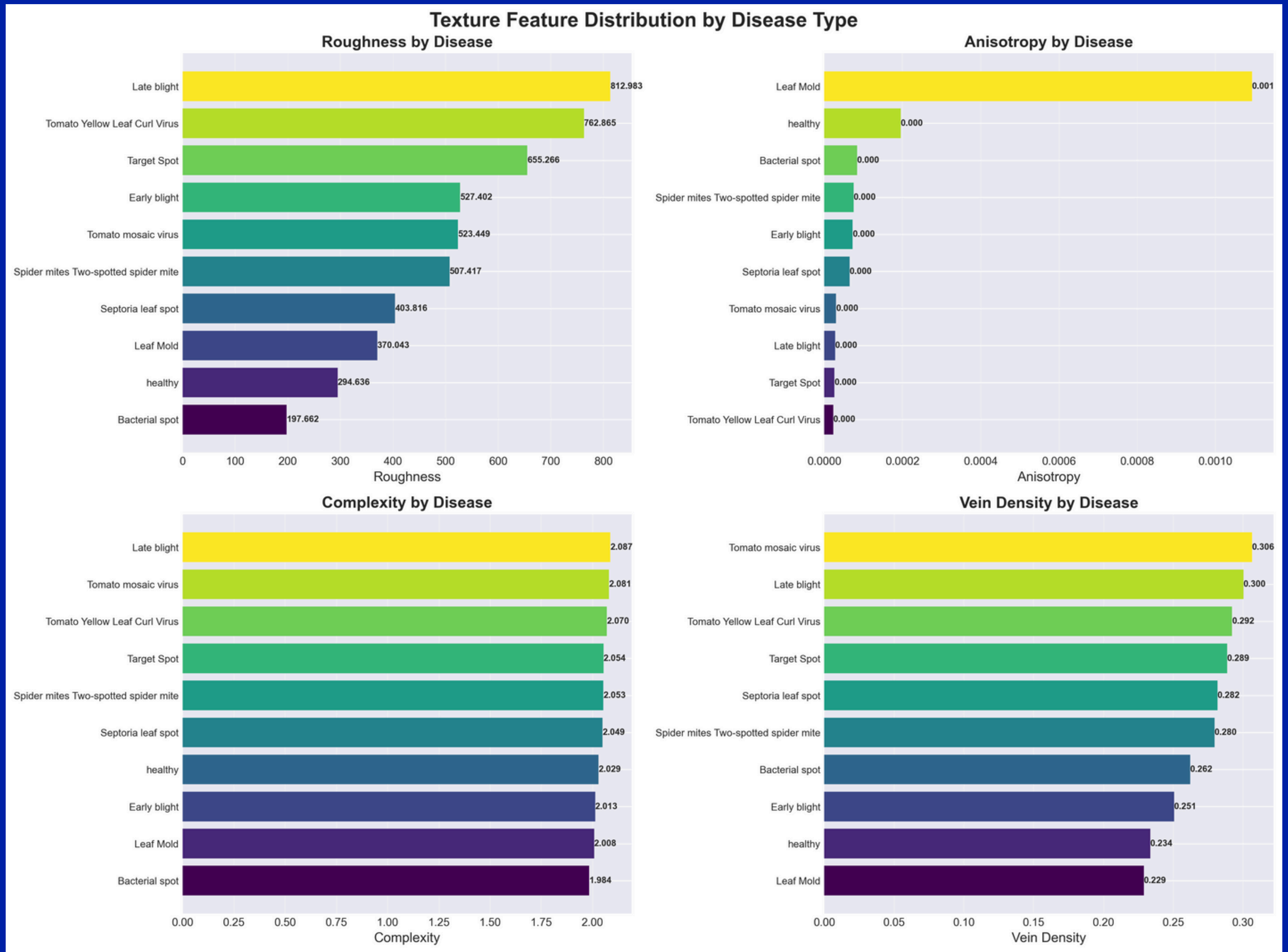


Insight

Even without labels, texture features alone reflect biological patterns

Texture Feature Distribution

Key Texture Features
Four engineered texture metrics analyzed across diseases:
Roughness, Anisotropy, Complexity, Vein Density.



Main Observations

Roughness:
– Highest in Late blight (812.98); lowest in Bacterial spot (197.66).
– High roughness correlates with severe lesion formation.

Anisotropy:
– Leaf Mold shows the strongest directional texture (0.00109).
– Indicates highly organized fungal growth patterns.

Complexity (Fractal Dimension):
– Highest in Late blight (2.087) and Tomato mosaic virus (2.081).
– Reflects irregular disease progression patterns.

Vein Density:
– Highest in Tomato mosaic virus (0.306); lowest in Leaf Mold (0.229).
– Viral infections preserve vein structure better than fungal diseases.

Takeaway
Each disease exhibits a distinct texture signature, supporting meaningful clustering and cross-domain interpretation

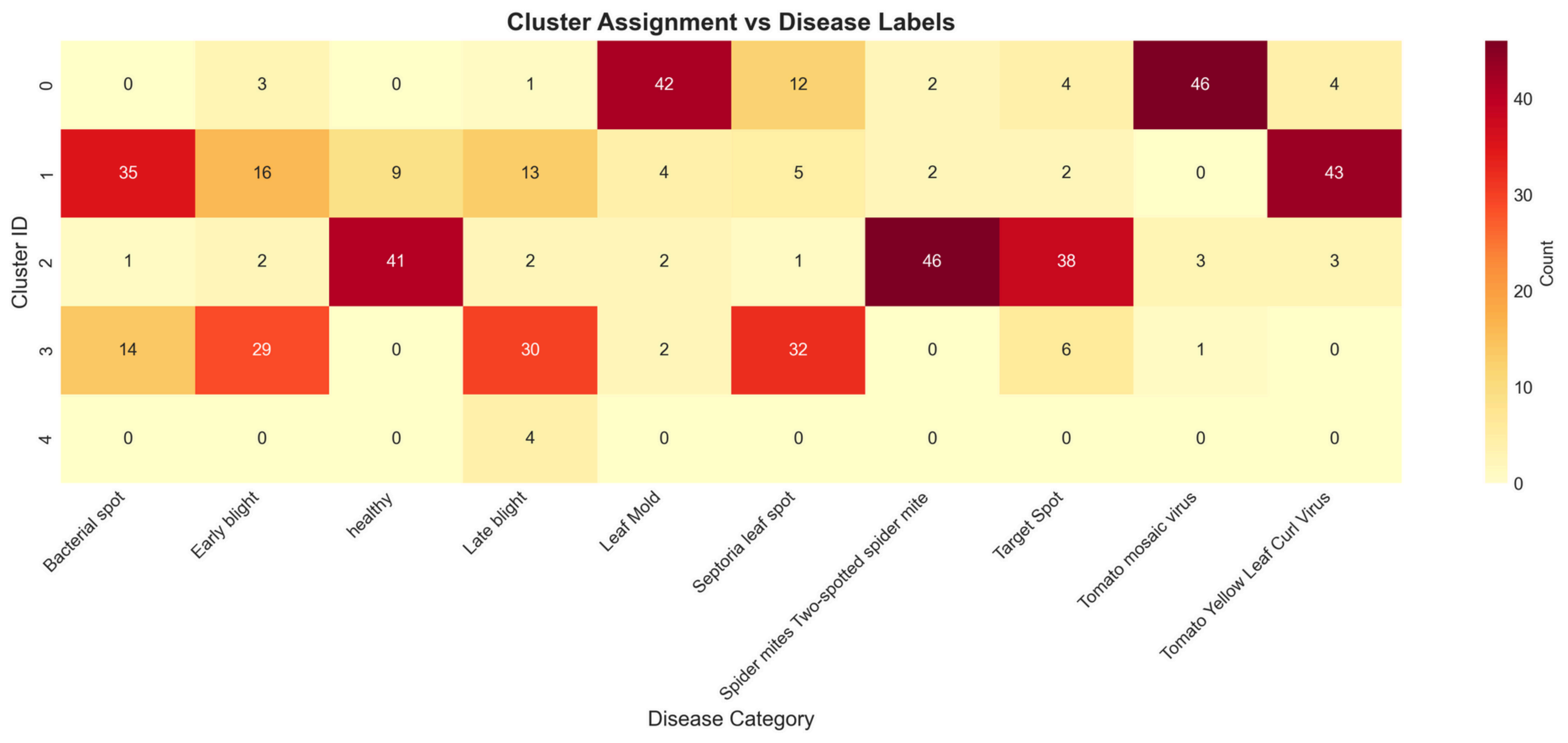
Disease–Cluster Relationships

Cross-Domain Patterns

- Each disease has a signature texture "profile"
- Strong correlations suggest:
  - Potential for disease pre-screening via texture alone
  - Texture-based engineering applications inspired by biological surfaces

Engineering Relevance

Rough surfaces → friction enhancement
Dense vein patterns → micro-channel flow modeling
Complex surfaces → light scattering / diffusion applications



Full Dual-Application Demonstration

Integrated Example Analysis

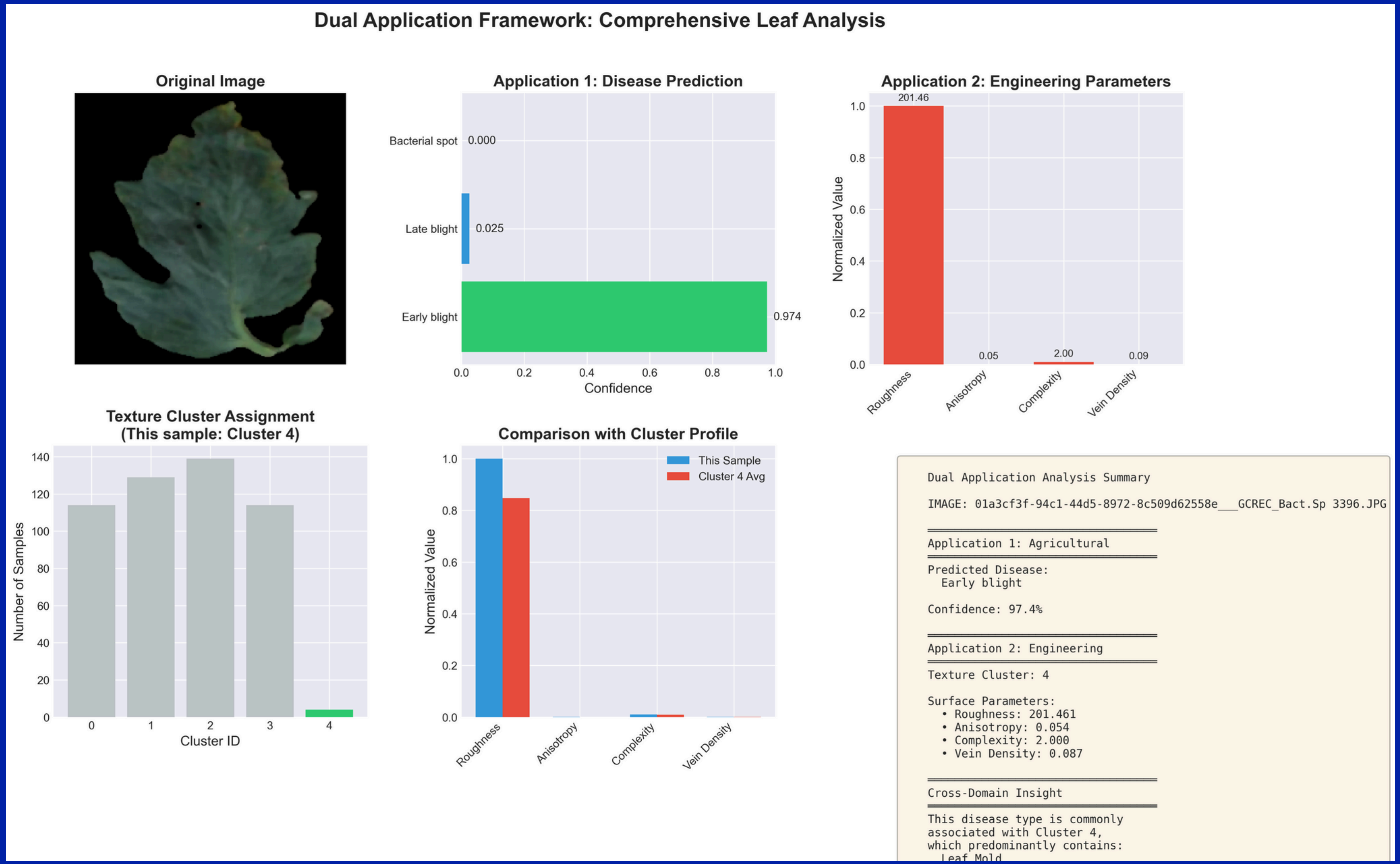
Given a single leaf image:

Disease prediction (Application 1)

Texture parameter extraction (Application 2)

Cluster assignment

Comparison with typical cluster statistics



Outcome

System provides both biological and engineering interpretations

Demonstrates the power of a unified multi-domain pipeline

Q&A

Thank You For listening

GITHUB

https://github.com/Ckcinnabar/term-project.git