# Plant Leaf Image Classification through SAM-based Segmentation in Vision LLM

## Making AI an Expert of Recognizing Plants Disease

*University of Southern California*

*Xiao Bai*

*Rongyi Shen*

*Wenjing Huang*

USC Viterbi
School of Engineering
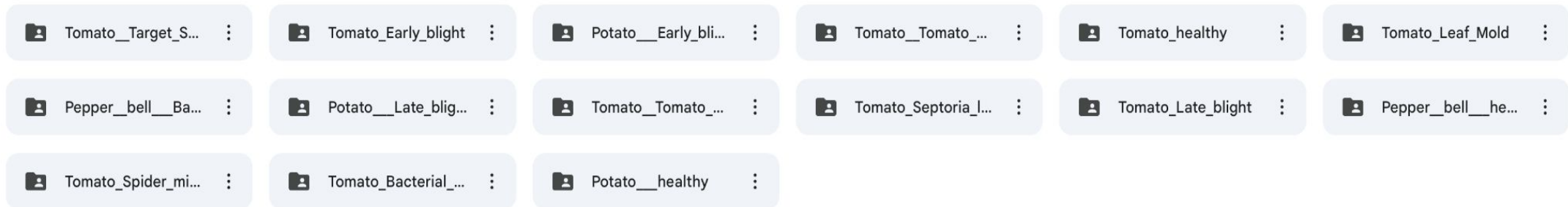
University of Southern California

# Main Project Idea

**Leaf Disease Detection** through the development and evaluation of a Vision-Language model integrated with segmentation techniques.

- ViT (Vision Transformer) for image classification
- SAM (Segment Anything Model) for precise segmentation

Dataset → ViT → SAM → Augmentation → Results Plan

**Dataset Overview:** 20,000

| | | | |
|---|---|---|---|
| Tomato_Target_S... | Tomato_Early_blight | Potato__Early_bli... | Tomato__Tomato_... |
| Pepper__bell___Ba... | Potato___Late_blig... | Tomato_Tomato_... | Tomato_Septoria_l... |
| Tomato_Spider_mi... | Tomato_Bacterial_... | Potato___healthy | |

Tomato_healthy | Tomato_Leaf_Mold
Tomato_Late_blight | Pepper__bell___he...

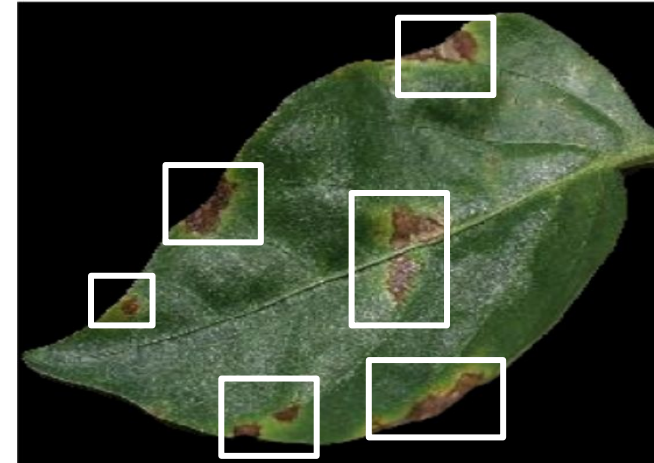# Model Training for ViT Model



```
Zero-Shot Classification Accuracy on 300 images: 0.00%
```

**Setup**

- Train:Test = 80:20
- Optimizer: AdamW.
- Loss Function: Cross-Entropy Loss.
- Scheduler: Linear learning rate scheduler.

**Initial Performance:**

- Fine-tuning:
  - After 5 epochs, Training Accuracy: 99.99%, Testing Accuracy: 99.73%.

```
Epoch 1/5
Epoch 1: 100%|          | 516/516 [02:37<00:00, 3.28it/s, accuracy=93, loss=0.147]
Epoch 1 completed. Loss: 0.5701, Accuracy: 93.01%
Epoch 2/5
Epoch 2: 100%|          | 516/516 [02:35<00:00, 3.31it/s, accuracy=99.5, loss=0.0417]
Epoch 2 completed. Loss: 0.0809, Accuracy: 99.54%
Epoch 3/5
Epoch 3: 100%|          | 516/516 [02:35<00:00, 3.31it/s, accuracy=99.8, loss=0.0295]
Epoch 3 completed. Loss: 0.0388, Accuracy: 99.78%
Epoch 4/5
Epoch 4: 100%|          | 516/516 [02:35<00:00, 3.32it/s, accuracy=100, loss=0.0179]
Epoch 4 completed. Loss: 0.0228, Accuracy: 99.96%
Epoch 5/5
Epoch 5: 100%|          | 516/516 [02:35<00:00, 3.32it/s, accuracy=100, loss=0.018]Epoch 5 completed. Loss: 0.0175, Accuracy: 99.99%
```
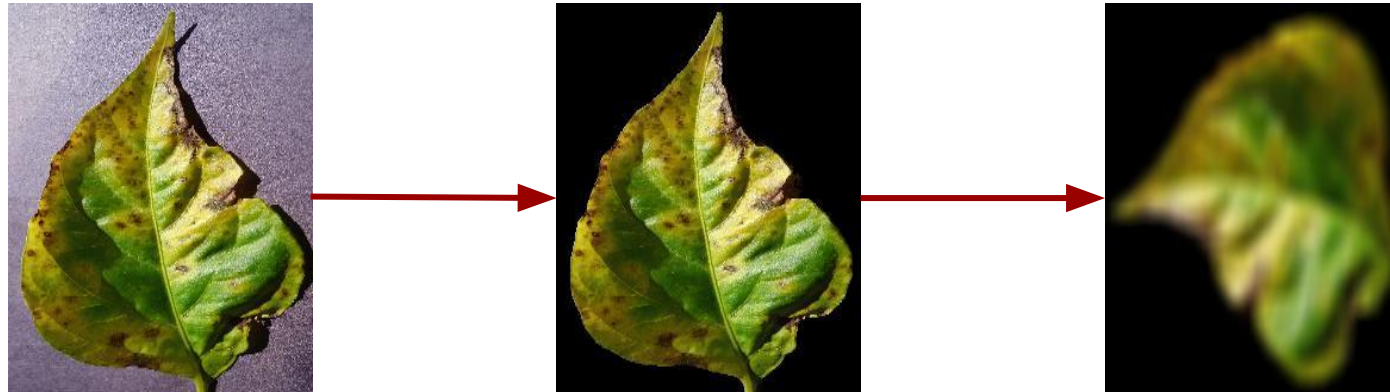
**Key Observation:**

- Model learns dataset-specific patterns and lacks robustness on unseen cases

# Enhancing Dataset Diversity with SAM-Based Segmentation

Segmented main regions of the image. Flipping, Rotating and Gaussian blur, combined SAM-segmented images with augmented images to create a new dataset.



**Model Performance after SAM**
- Fine-tuned ViT on SAM-Segmented Data:
    - Testing Dataset size: 5,000.
    - Accuracy: Dropped to 78.9%.

Model is focusing on more generalizable, meaningful features.

# Future Plan

**Next Steps**:

- Train with total Augmented and Segmented Datasets (~40,000)

- Investigate additional strategies to improve robustness.

**Future Applications**:

- Generalizable models for real-world plant disease detection.

# Thank you!

**CSCI 566: Deep Learning and its Applications**

*Segment Everything:*
*Xiao Bai*
*Rongyi Shen*
*Wenjing Huang*