Pixel2Prescription

A Vision Foundational Model for Agricultural Intelligence

Aksshay Mathew, Jugal Alan, and Tanya Khanna

Department of Computer Science, IIT Ropar

Individual Contributions

Aksshay Mathew: Self-supervised pretraining and fine-tuning

Jugal Alan: Knowledge-base, website and model to prescription development

Tanya Khanna: Dataset curation, DINOv2+EfficientNet integration, model comparison

The Agricultural AI Challenge

The Problem

Global food security is at risk. Plant diseases account for 20-40% of annual crop losses, costing billions and causing food shortages. Traditional diagnosis relies on expert agronomists who are often unavailable in remote agricultural regions.

The Opportunity

Self-Supervised Learning (SSL), particularly DINOv2, can learn rich visual features from vast, unlabeled agricultural image datasets. This technology can bridge the expertise gap and provide timely, accessible diagnostics to farmers everywhere.

Dataset(I)

Dataset 1: Pretraining

A massive, custom dataset of **167,716 images**.
Crops: Apple, Rice, Chilli, Cotton, Cassava, Wheat, Cucumber, Corn, Tomato, and Sugarcane.
Sources: PlantDoc, CCMT, New Plant Diseases, DeepWeeds

Dataset 2: Fine-Tuning

The public **Plant Village** dataset.
Classes: 38 distinct plant-disease conditions.
Total Samples: 19,326 images (Train/Val/Test).

Dataset(II)

Crop	Images	Classes
Apple	32,297	8
Rice	18,679	15
Chili	8,622	5
Cotton	8,028	7
Cassava	7,003	4
Wheat	6,870	7
Cucumber	6,112	9
Corn	5,957	10
Tomato	5,311	9
Sugarcane	3,068	6

a) Pretraining

Pest	Images	Pest	Images	Pest	Images
Leaf folder	5,054	Green horned cat.	2,385	Aphid	1,817
Brown planthopper	4,858	Caterpillar	2,384	Yellow stem borer	1,750
Rootworm	4,721	Skipper	2,300	White backed hopper	1,682
Stem Borer	4,575	Swarming caterpillar	2,199	Gall midge	1,682
Weevil	4,240	Others	1,908	Case worm	1,605
Green leafhopper	4,207	Termites	1,883	Mealybug	1,441
Beetle	4,036	Spiny beetle	1,870	Grasshopper	1,297
Thrips	2,516	Earhead bug	1,870	Snail	1,030
Whorl-maggot	2,459				

b) Fine-tuning

Literature Survey

Model (Year)	Pretraining Technique	Domain	Key Idea
MoCo v2 (2020)	Self-Supervised Contrastive	General CV	Uses a momentum encoder for contrastive learning.
BYOL (2020)	Self-Supervised (No Negatives)	General CV	Learns by predicting a "target" network's output without using negative samples.
DINO (2021)	Self-Distillation (No Labels)	General CV	A "student" ViT is trained to match the output of a "teacher" ViT.
MAE (2021)	Masked Image Modeling (MIM)	General CV	A ViT encoder-decoder learns to reconstruct randomly masked-out image patches.
SimMIM (2022)	Masked Image Modeling (MIM)	Industrial Vision	A lightweight masked prediction head (simpler than MAE).
iBOT (2022)	Hybrid (MIM + Distillation)	General CV	Combines the patch-level MIM from MAE with the self-distillation from DINO.
AgriCLIP (2024)	Contrastive V-L Pretraining	Agriculture	CLIP-style alignment on agriculture-specific image-text pairs.
AgriMAE (2025)	Self-Supervised MAE	Agriculture	A domain-adapted MAE (masked autoencoder) for crop health monitoring.

Our Key Contributions

Agri-Foundation Model

Pretrained a DINOv2 foundation model on 167,716 agricultural images across ten major crops.

Hybrid Architecture

Developed a DINOv2-EfficientNet model, achieving state-of-the-art performance and efficiency.

Real-World Deployment

Launched a web app for real-time diagnosis and AI-powered treatment prescriptions.

A 4-Stage Methodology

Our pipeline moves from large-scale pretraining to a production-ready deployment.

Pipeline

Stage 1: Pretraining

DINOv2 trained on 167k unlabeled images to learn general agricultural features.

Stage 2: Fine-Tuning

Model adapted to classify 38 specific diseases from the Plant Village dataset.

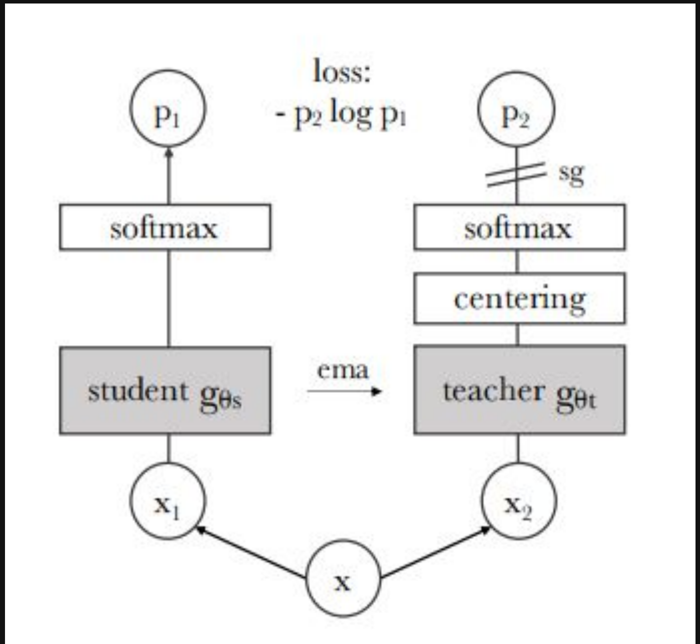
Stage 3: Hybrid Model

Fused DINOv2 (global) and EfficientNet (local) features for maximum accuracy.

Stage 4: Deployment

Integrated into a web app with a knowledge base for real-time prescriptions.

Pretraining



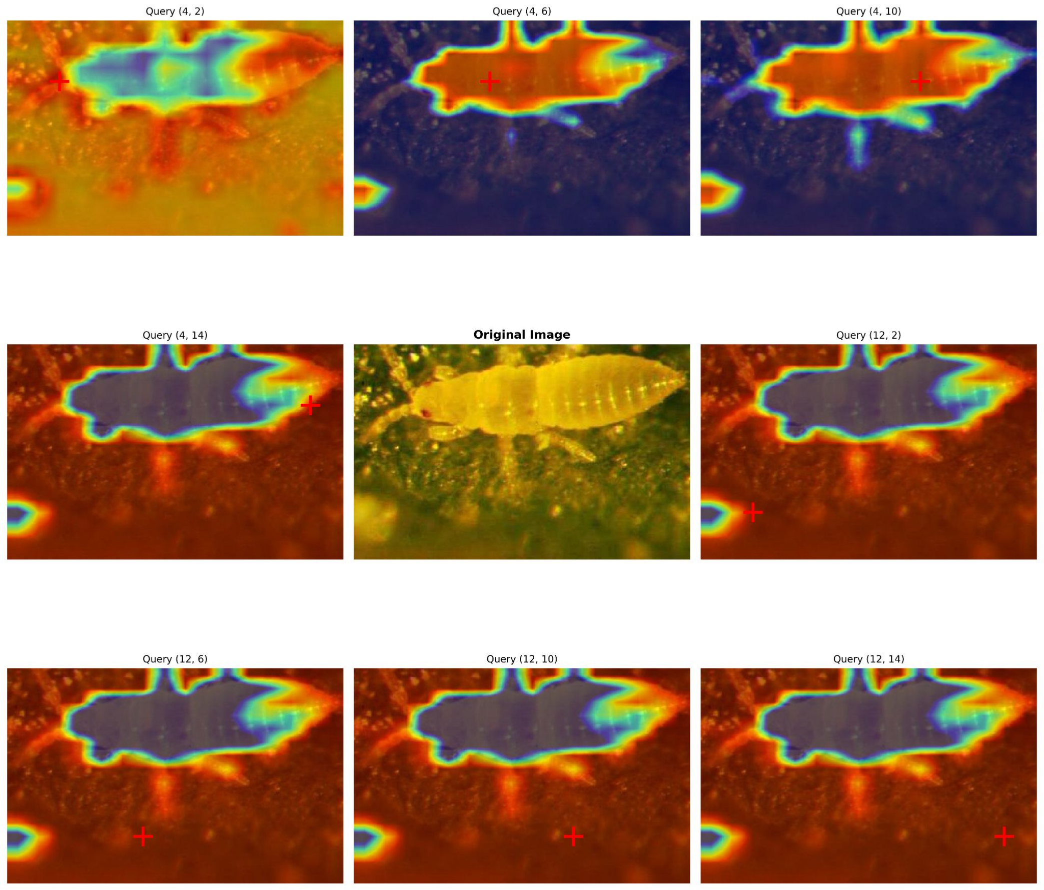
Training Overview

- **Total Epochs:** 15
- **Hardware:** Single NVIDIA H100 80GB GPU
- **Time per Epoch:** Approximately 1 hour
- **Total Training Time:** Approximately 15 hours

Loss & Convergence: The model learned to represent the data very quickly.

- **Epoch 1 (Initial Loss):** 0.2616
- **Epoch 2 (Rapid Drop):** 0.0005
- **Epoch 15 (Final Loss):** 0.0001

Cosine Similarity Heatmap



What Are We Seeing? This grid visualizes the "attention" of our pretrained DINOv2 model.

- **Center:** The original input image (a "Thrips" pest).
- **Outer Images:** These are **cosine similarity heatmaps**.
- **Red Cross +:** This is the "query patch." The heatmap shows how "similar" all other patches in the image are to this single query patch.

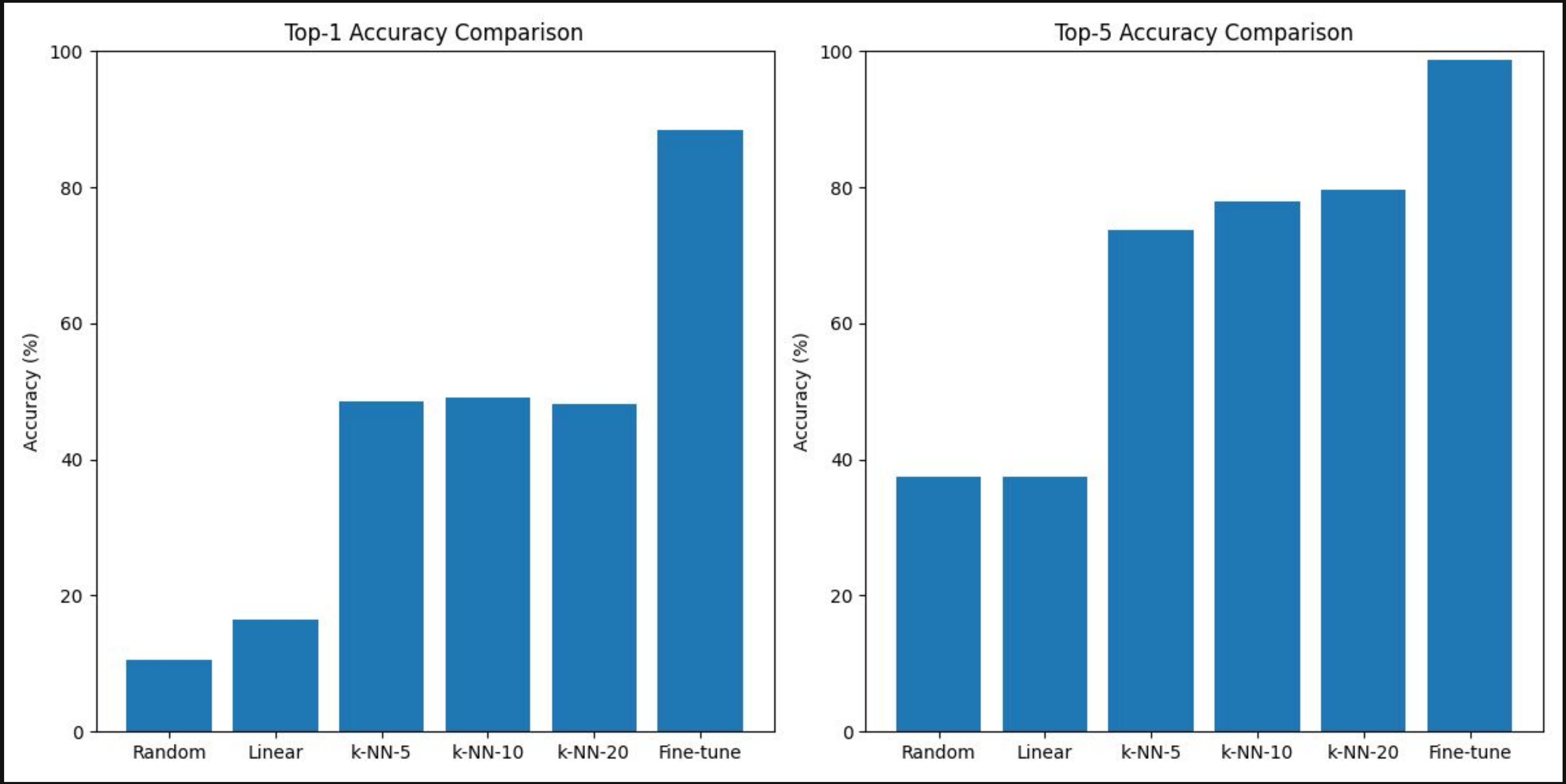
Fine-tuning

Task: Classify 38 different plant diseases from the standard PlantVillage dataset.

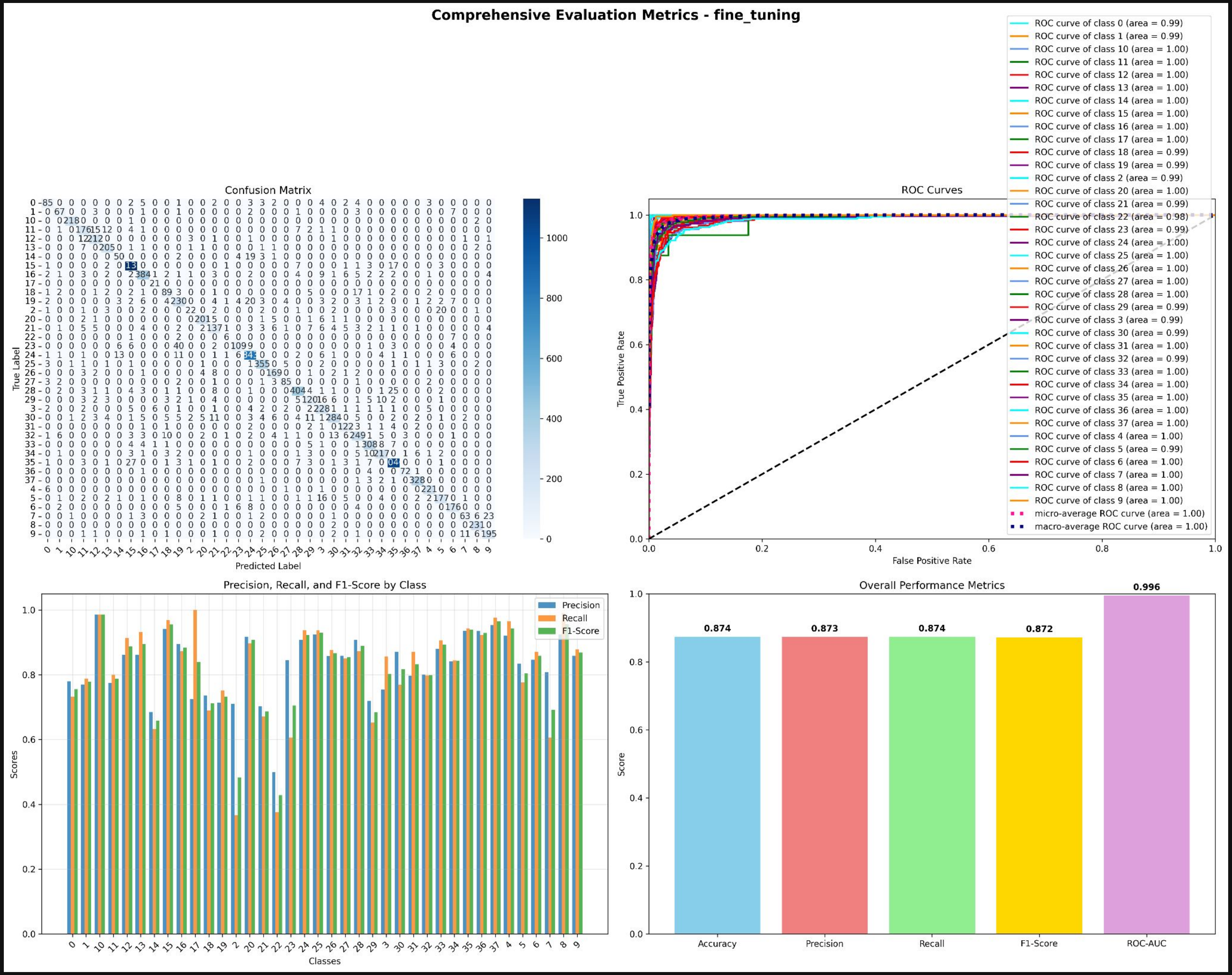
Methods Tested: We compared four different evaluation strategies.

- **Random Initialization (Baseline):** Training the model from scratch.
- **Linear Probing:** Freezing the 'Agri-ViT' backbone and training only a final linear layer.
- **k-Nearest Neighbors (k-NN):** Using the 'Agri-ViT' backbone as a feature extractor.
- **Fine-Tuning:** Unfreezing the entire 'Agri-ViT' model and training it on the new task.

Accuracy Comparison



Fine Tuning Evaluation Metrics Comparison

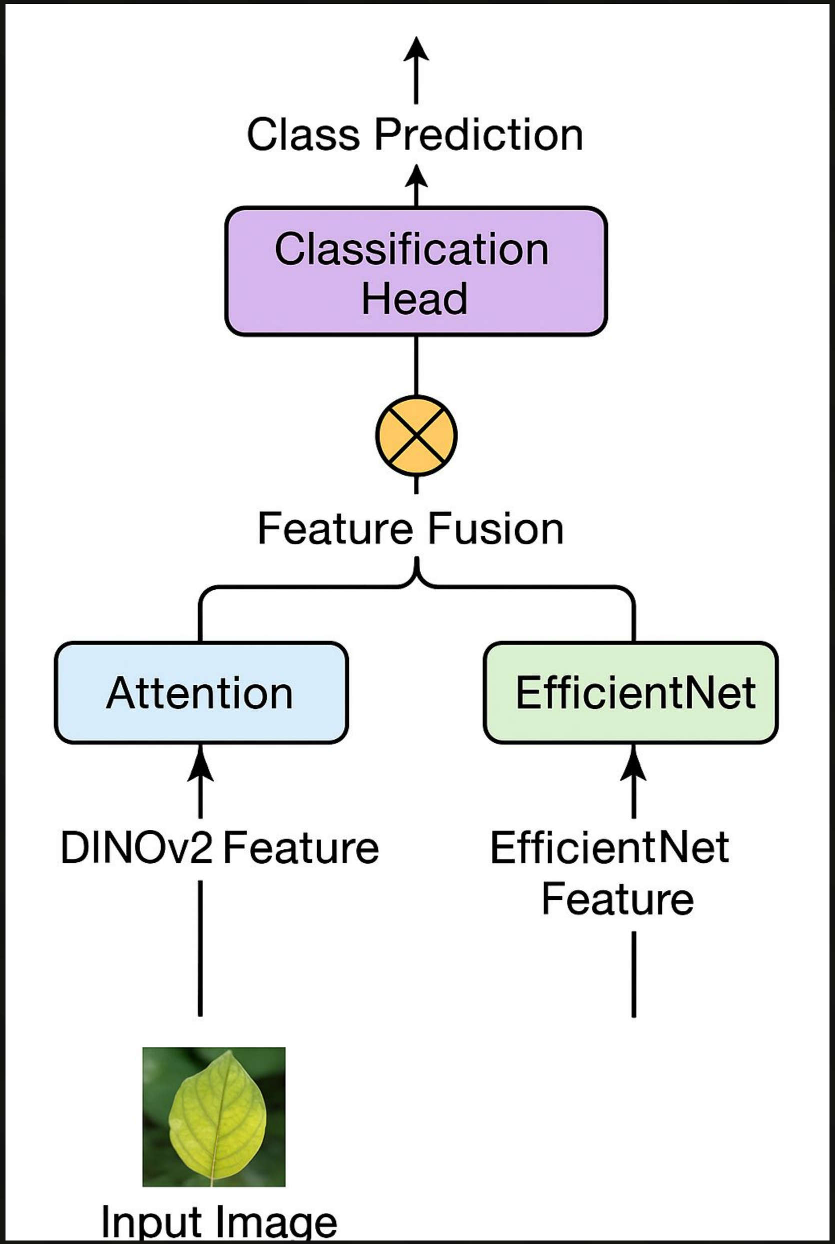


Hybrid Model Architecture

We fused two powerful architectures to get the best of both worlds:

- **DINOv2 (ViT):** Provides global semantic understanding and strong background invariance.
- **EfficientNet:** Excels at identifying local, discriminative patterns and subtle visual symptoms.

An attention-based mechanism combines these features, creating a comprehensive and highly accurate representation for diagnosis.



State-of-the-Art Performance

Our hybrid model achieves the highest Top-1 Accuracy for the 38-class Plant Village dataset.

Efficiency is Key

14.3M

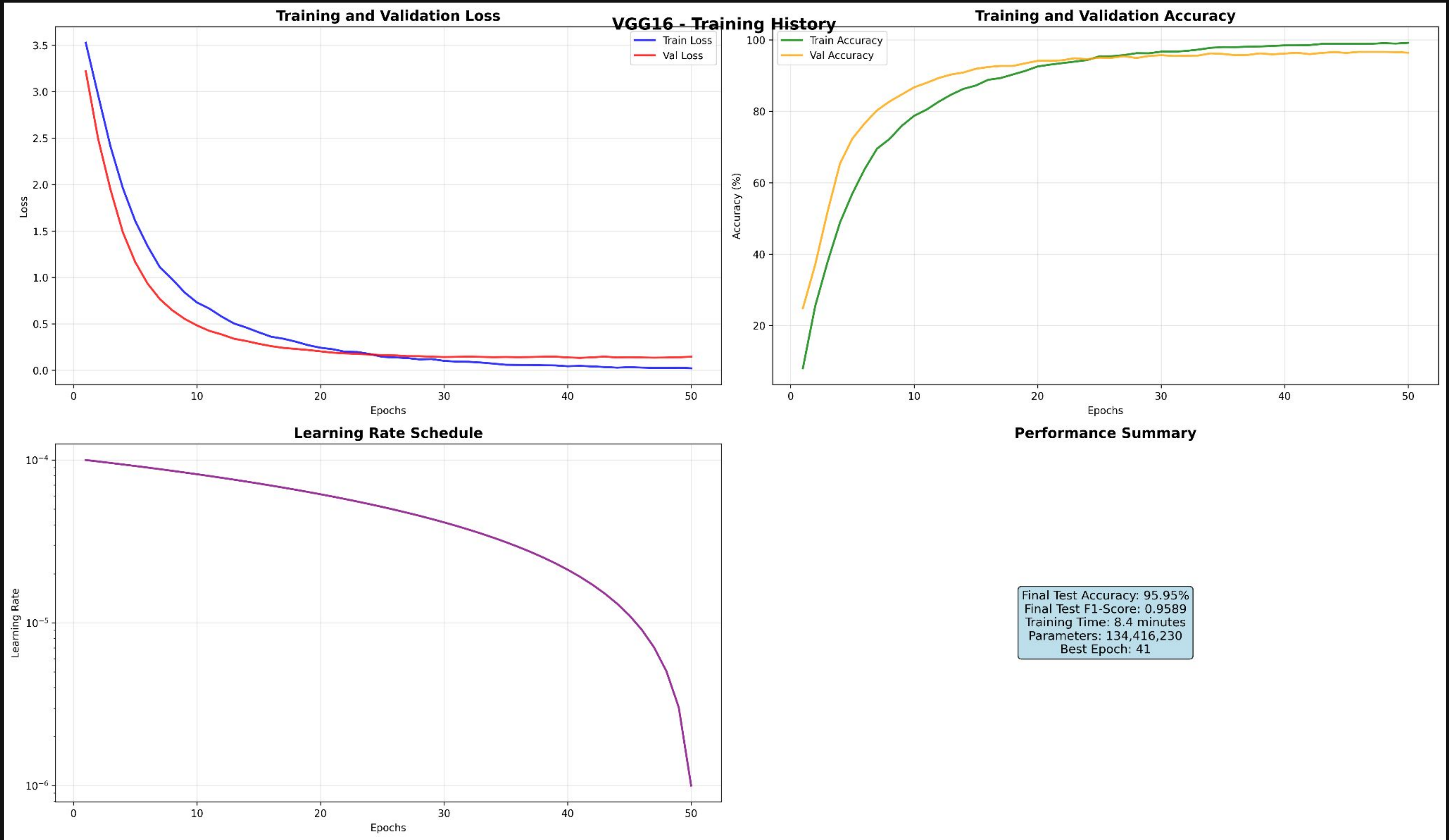
Total Parameters

Higher Accuracy, Smaller Model

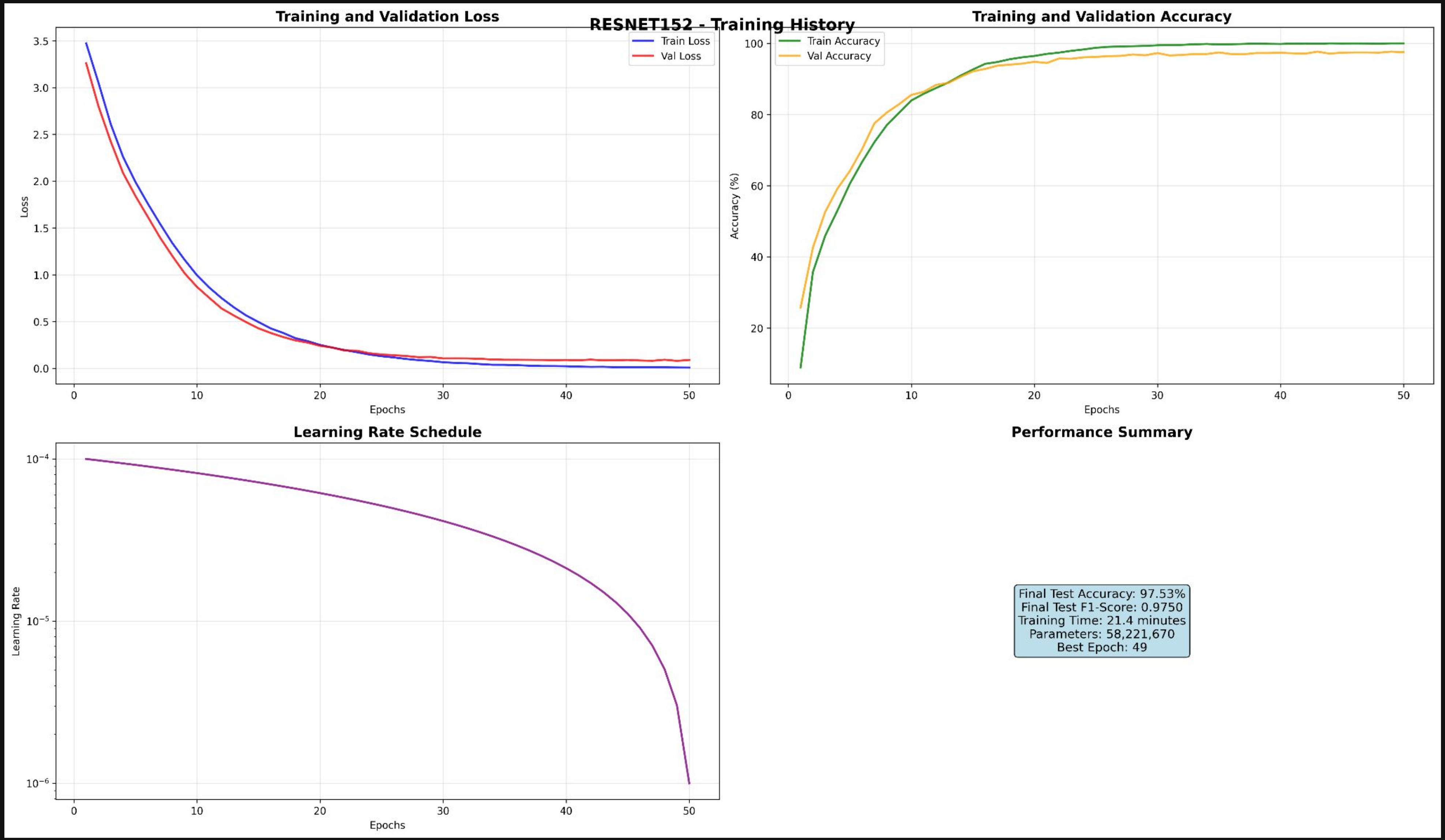
Our hybrid model is **9.4x smaller** than VGG16 (134.4M) and **4.1x smaller** than ResNet152 (58.2M), while achieving superior accuracy.

This low parameter count is critical for efficient real-world deployment on web and mobile devices.

VGG16

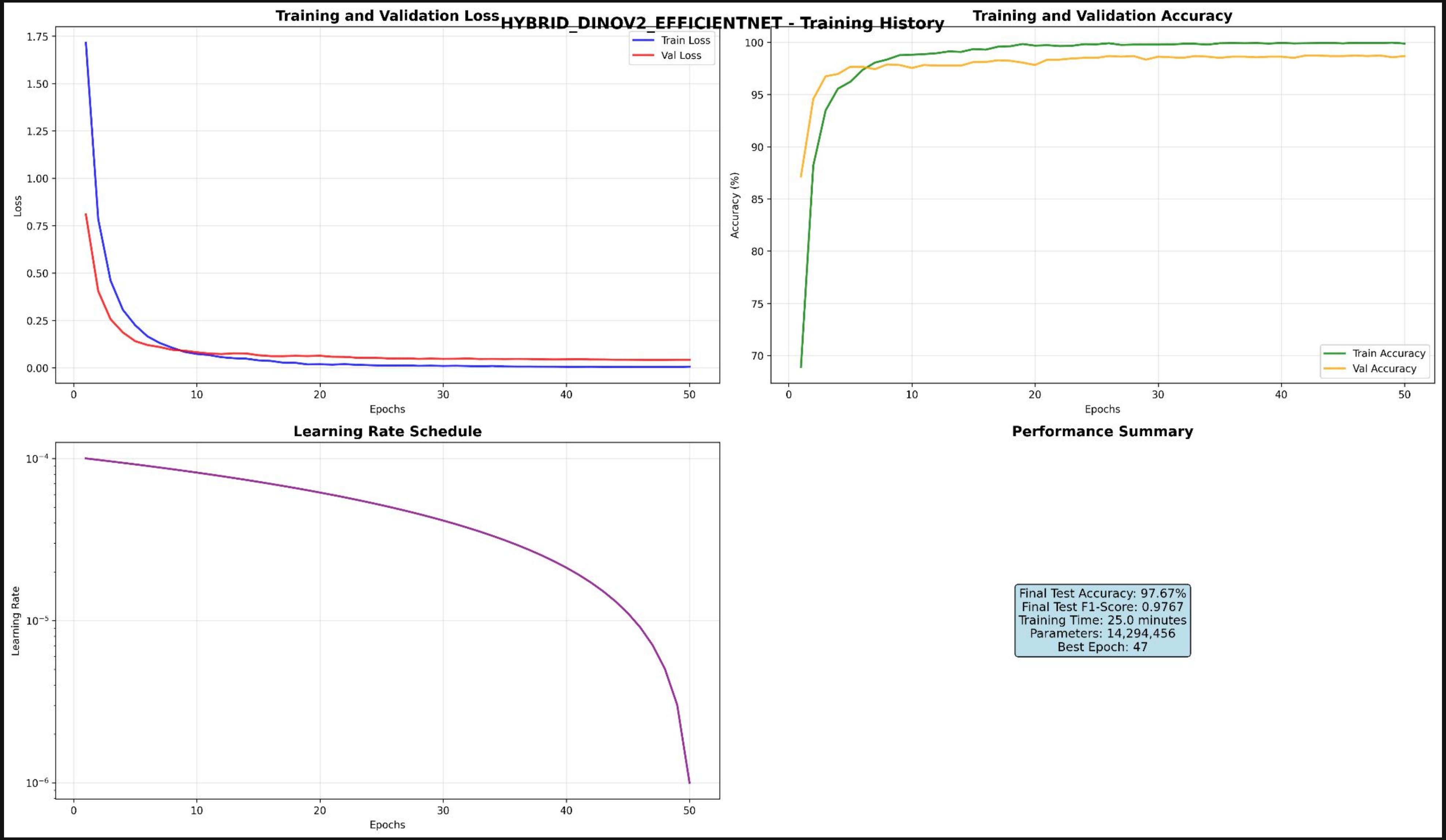

VGG16 - Training History
Training and Validation Loss
Train Loss
Val Loss
Loss
Epochs
Training and Validation Accuracy
Train Accuracy
Val Accuracy
Accuracy (%)
Epochs
Learning Rate Schedule
Learning Rate
Epochs
Performance Summary
Final Test Accuracy: 95.95%
Final Test F1-Score: 0.9589
Training Time: 8.4 minutes
Parameters: 134,416,230
Best Epoch: 41

RESNET152



RESNET152 - Training History
Training and Validation Loss
Train Loss
Val Loss
Loss
Epochs
Training and Validation Accuracy
Train Accuracy
Val Accuracy
Accuracy (%)
Epochs
Learning Rate Schedule
Learning Rate
Epochs
Performance Summary
Final Test Accuracy: 97.53%
Final Test F1-Score: 0.9750
Training Time: 21.4 minutes
Parameters: 58,221,670
Best Epoch: 49

Hybrid DINOv2



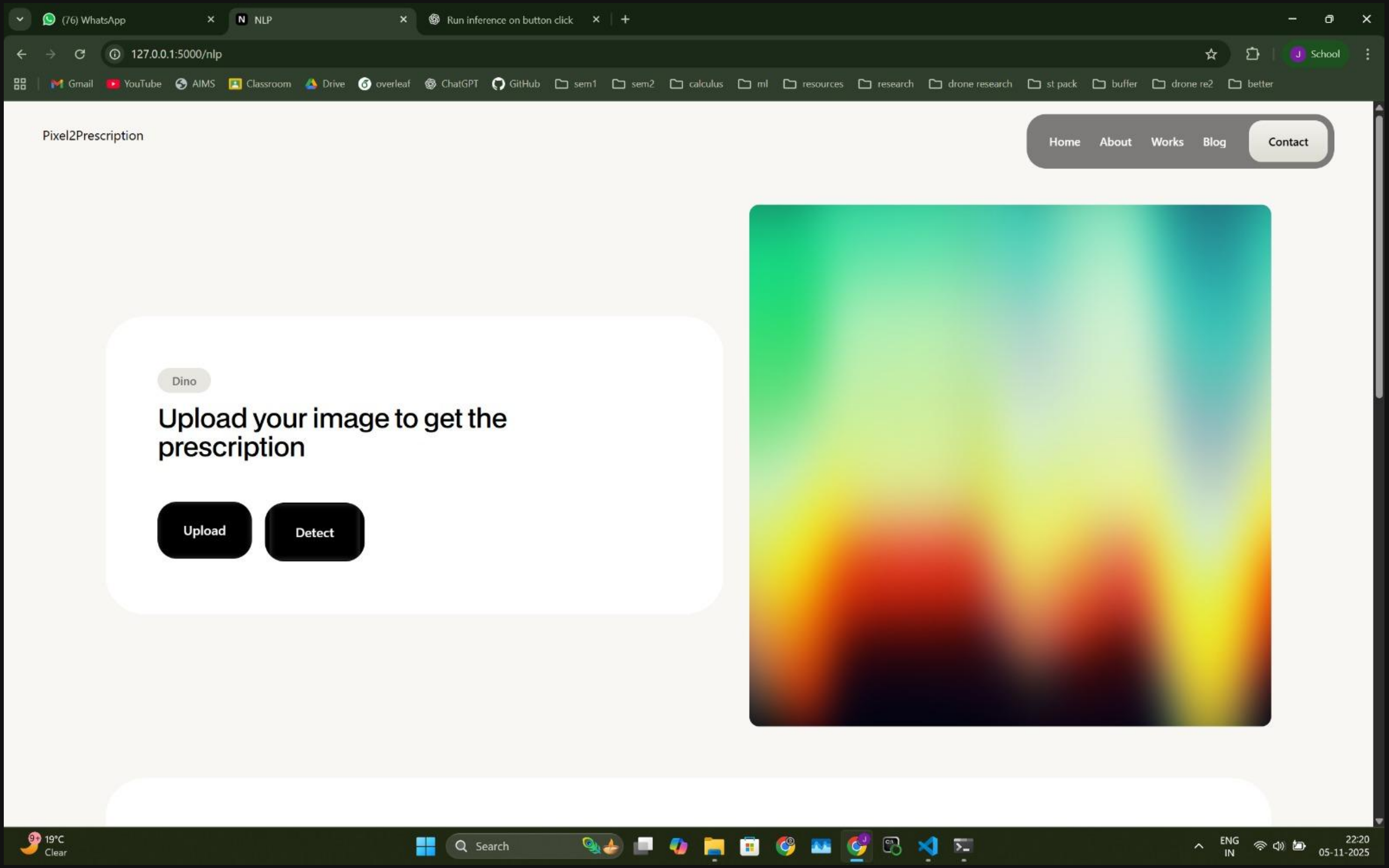
From Diagnosis to Prescription

Source: www.mdpi.com

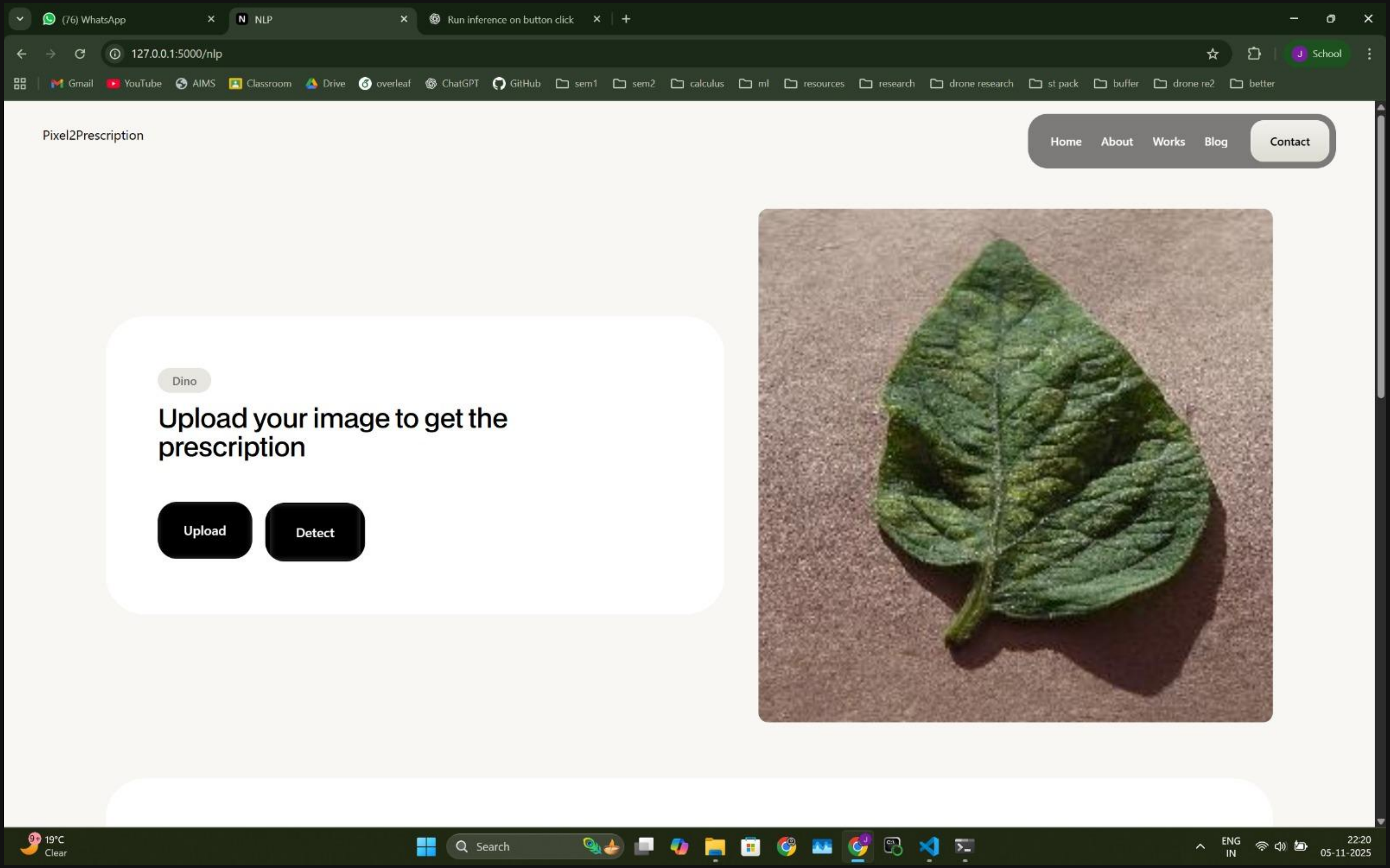
A Complete Solution

- **1. Diagnosis:** Identifies the plant disease with 97.67% accuracy.
- **2. Confidence:** Provides a confidence score for the prediction.
- **3. Prescription:** Retrieves actionable advice from an agricultural knowledge base (Organic, Chemical, and Prevention strategies).

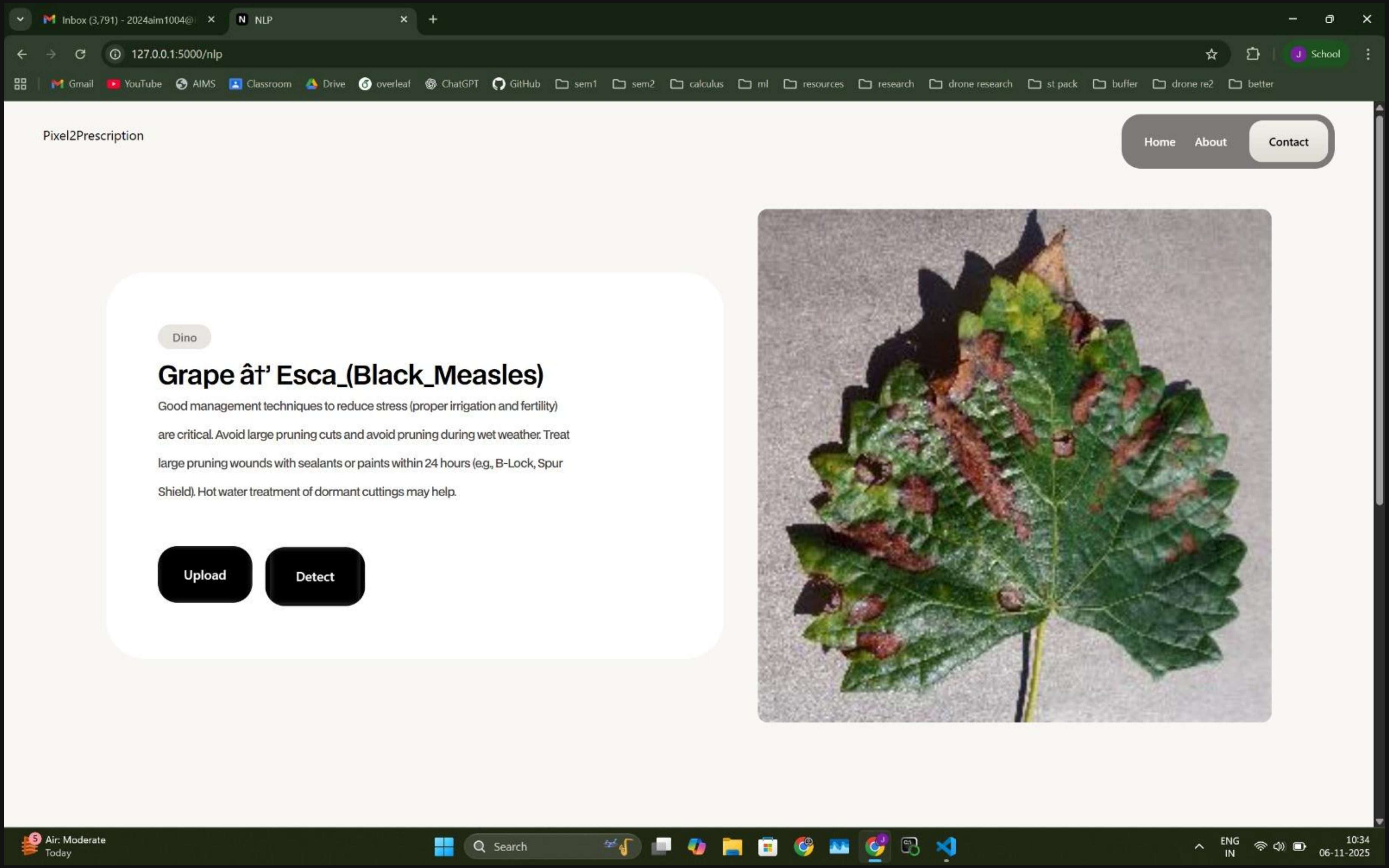
Prescription(I)



Prescription(II)



Prescription(III)



Conclusion & Key Achievements

- **Foundation Model:** Successfully created the first large-scale, self-supervised foundation model specifically for agriculture.
- **State-of-the-Art:** Achieved 97.67% accuracy, outperforming all existing CNN and ViT benchmarks on the Plant Village dataset.
- **Parameter Efficient:** Delivered SOTA results with only 14.3M parameters, making it highly efficient and deployable.
- **Real-World Impact:** Deployed a production-ready system that provides immediate, accessible, and actionable value to farmers.

Future Work

- **Model Expansion:** Extend the scalable framework to include a wider variety of crops, new geographical regions, and a broader range of agricultural challenges.
- **Address New Challenges:** Move beyond disease identification to tackle other critical issues like pest detection, nutrient deficiency analysis, and plant stress monitoring.
- **Improve Accessibility:** Continue to democratize agricultural AI by enhancing the web platform, developing mobile-friendly applications, and adding multi-lingual support for global reach.

Thank You

Questions?